Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP04/2016

Asymptotic Theory for Aggregate Efficiency
Leopold Simar, Valentin Zelenyuk

Date: May 2016

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Asymptotic Theory for Aggregate Efficiency

Léopold Simar* Valentin Zelenyuk§
leopold.simar@uclouvain.be v.zelenyuk@uq.edu.au

May 12, 2016

Abstract

Applied researchers in the field of efficiency and productivity analysis often need to estimate and inference about aggregate efficiency, such as industry efficiency or aggregate efficiency of a group of distinct firms within an industry (e.g., public vs. private firms, regulated vs. unregulated firms, etc.). While there are approaches to obtain point estimates for such important measures, no asymptotic theory have been derived for it–the gap in the literature that we fill with this paper. Specifically, we develop full asymptotic theory for aggregate efficiency measures when the individual true efficiency scores being aggregated are observed as well as when they are unobserved and estimated via DEA or FDH. As a result, the developed theory opens a path for more accurate and theoretically better grounded statistical inference on aggregate efficiency estimates such as industry efficiency, etc.

Keywords: DEA, FDH, Efficiency, Aggregation, Industry Efficiency, Asymptotics, Limiting distribution, Consistency, Convergence, Jackknife, Bias correction.

*Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium. Research supported by IAP Research Network P7/06 of the Belgian State (Belgian Science Policy).

§School of Economics and Centre for Efficiency and Productivity Analysis (CEPA), University of Queensland, Colin Clark Building (39), St Lucia, Brisbane, Qld 4072, Australia

1 Introduction

Researchers performing applied work in the field of efficiency and productivity analysis often need to estimate aggregate efficiency or efficiency of a group consisting of individual decision making units (DMUs). This could be, for example, the industry efficiency, or the aggregate efficiency of some particular groups of firms within an industry (e.g., public firms vs. private firms, or regulated vs. unregulated firms, etc.). While there are approaches to obtain point estimates for such practically important measures, no asymptotic theory have been derived for them except for the cases of equally weighted means–this is the gap in the literature that we fill with our paper. Specifically, here we develop full asymptotic theory for aggregate efficiency measures when the individual true efficiency scores being aggregated are observed as well as when they are unobserved and estimated using the data envelopment analysis (DEA) and the free disposal hull (FDH) estimators.

The envelopment estimators that we consider here have proven to be very useful tools for performance analysis. This is especially the case for DEA–a very powerful linear programming technique developed by Charnes, Cooper and Rhodes (1978) who elaborated on the seminal work of Farrell (1957). These methods have reached practically uncountable number of applications in many areas of human activity, both in public sector (e.g., see Ruggiero, 1996) and in private sector (e.g., see Cook and Zhu, 2006) of many economies.[1]

In this work we will particularly focus on developing statistical theory for the aggregate efficiency estimated with standard DEA and FDH estimators. Our approach can also be extended by analogy to other variations, such as the IDEA (Zhu (2004)), CAR-DEA (Cook and Zhu (2008)), DEA with non-homogeneous firms (Cook et al. (2013)), the related to DEA nonparametric models of optimizing behavior (Cherchye et al (2008)), Stochastic DEA and Stochastic FDH (e.g., Simar and Zelenyuk (2011)), to mention just a few popular variations.

To develop our theory, we first give a new interpretation to the existing aggregate efficiency measures–not as the weighted averages (as usually presented), but as the ratios of simple averages. We then point out that for a hypothetical case when the individual true efficiency scores being aggregated are observed, a proper application of existing central limit theory for ratios of means may solve the task. However, in the realistic case when the individual true efficiency scores being aggregated are not observed but estimated with DEA or FDH, an application of standard central limit theory is problematic due to inherent bias in DEA and FDH estimators, as was pointed out by Kneip, Simar and Wilson (2015a,b) and we used these two recent works to develop a related asymptotic theory, and

[1] E.g., see Cooper, Seiford and Zhu (2011) for an extensive discussion of DEA and its many variations.

procedures for correcting the bias, for the case of aggregate efficiency measures that were developed by Färe and Zelenyuk (2003) and Simar and Zelenyuk (2007). In fact, the latter work also has proposed a practical way of making statistical inference on the aggregate efficiency and correcting for bias, using sub-sampling bootstrap, but an asymptotic theory was not developed there. In addition the choice of the optimal subsample size is still an unresolved problem. Thus, the current paper makes substantial improvement upon Simar and Zelenyuk (2007).

The strategy we use to derive asymptotic properties consists of three steps. In the first step, we convert the weighted aggregate efficiencies described above into ratios of the non-weighted means. In the second step we adapt statistical theory of the ratios of sample means to our aggregate efficiency scores assuming the individual efficiency scores are known. In the third step we extend the results to the case when individual efficiency scores are unknown but estimated, accounting for the characteristics pertinent to particular efficiency estimators at use.

The asymptotic theory developed in this paper paves the way to more accurate and better theoretically grounded statistical inference on aggregate efficiency estimates, such as industry efficiency or aggregate efficiency of a group of firms within an industry or efficiency of group of countries, etc.

The paper is structured as following: Section 2 briefly outlines the economic theory characterization of technology, efficiency measures for individual units (firms, banks, hospitals, etc.) and of their theoretical aggregates traditionally represented in the form of weighted averages of the individual scores. Section 3 gives an alternative representation of the aggregate efficiencies–as ratios of simple averages, which is more convenient for statistical analysis because some well-known theorems can be applied to derive the asymptotic results when the true efficiency is known, as is done in Section 4. Because in reality the true efficiency is typically unknown, and one has to use its estimates, proper asymptotic theory needs to be derived for the case of specific estimators because the standard asymptotic theory generally fails, except for very restricted cases. Section 5 briefly describes the estimators we focus on here (DEA and FDH) and summarizes the key asymptotic results from Kneip, Simar and Wilson (2015a,b) about averages of sample means, which are needed for us to derive our results, in Sections 6, 7, 8, and 9, for the context of aggregate efficiencies represented as ratios of averages. Specifically, Section 6 outlines the key asymptotic results, Section 7 describes estimation of the bias, Section 8 outlines two main central limit theorems and Section 9 describes how to use these theorems to estimate confidence intervals for the aggregate efficiency. Finally, Section 10 makes concluding remarks on the value from this paper and on paths for future research it opens.

2 Individual and Aggregate Efficiency

While considering modeling and estimation of a group of economic agents (plants, departments, firms, countries, regions, etc.) we will conventionally call them as decision making units (hereafter DMUs). Throughout the paper we will assume that a DMU i uses p inputs $x \in \mathbb{R}^p_+$ to produce a vector of q outputs, denoted by $y \in \mathbb{R}^q_+$. We also assume that technology of a DMU i can be characterized by the technology set, Ψ, defined, in general terms, 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 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 usual regularity conditions on technology in production theory (e.g., see Färe and Primont, 1995), and in particular we need the following two assumptions to hold:

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

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

The technical efficiency is usually viewed as a gap between the actual observation and the frontier Ψ^∂ and is usually be measured via the so-called Farrell-Debreu output oriented measure of technical efficiency, which for a DMU with an input-output allocation (x, y) is defined as

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

Another popular measure of efficiency used in practice that we will involve is known as the overall output efficiency or the revenue efficiency, which for a DMU with an input-output allocation (x, y) that face output prices $w \in \mathbb{R}^q_+$ is defined as

$$RE(x, y, w) := \frac{RF(x, w)}{w^T y} \tag{2.4}$$

where $RF(x, \omega)$ is the classical revenue function from economic theory, defined as

$$RF(x, w) := \max_y \left\{w^T y \ : \ (x, y) \in \Psi\right\} \tag{2.5}$$

and throughout the paper we assume that $w^T y \neq 0$.

The two efficiency measures are known to relate to each other via the Mahler's inequality, as

$$RE(x, y, w) \geq \lambda(x, y),\ \forall (x, y) \in \Psi,\ w \in \mathbb{R}^q_{++} \tag{2.6}$$

which can be used to introduce the so-called allocative efficiency measure, defined as the multiplicative residual that closes inequality (2.6), i.e., $AE(x, y, w) = RE(x, y, w)/\lambda(x, y)$. This, in turn, ensures the following decomposition of efficiencies at the individual level, for each DMU,

$$RE(x, y, w) = \lambda(x, y) \times AE(x, y, w), \forall (x, y) \in \Psi,\ w \in \mathbb{R}^q_{++}. \tag{2.7}$$

Of course, in practice neither Ψ nor Ψ^∂ is observed and so must be estimated from a sample of of data on inputs and outputs, denoted by $\mathcal{X}_n = \{(X_i, Y_i) \,|\, i = 1, \ldots, n\}$, and we will come back to this matter after outlining important results from theory of production, efficiency and aggregation.

Given certain assumptions on the aggregate technology and the law of one price on output markets (i.e., $w^i = w, \forall i$) and using the analogue to Koopmans (1957) theorem, Färe and Zelenyuk (2003) derived an economic theory justified aggregation scheme for obtaining the aggregate efficiency scores from the individual efficiencies outlined above. Specifically, from their approach, the aggregate revenue efficiency for a group of some n DMUs is given by

$$\overline{RE}_n := \sum_{i=1}^{n} RE(X_i, Y_i, w) \times S^i, \quad S^i = \frac{w^T Y_i}{\sum_{i=1}^n w^T Y_i} \tag{2.8}$$

while the aggregate technical efficiency is given by

$$\overline{TE}_n := \sum_{i=1}^{n} \lambda(X_i, Y_i) \times S^i, \tag{2.9}$$

and the aggregate allocative efficiency is given by

$$\overline{AE}_n := \sum_{i=1}^{n} AE^i \times S^i_a, \quad S^i_a = \frac{w^T Y_i^\partial}{\sum_{i=1}^n w^T Y_i^\partial}, \tag{2.10}$$

where Y_i^∂ is the counterfactual output vector for DMU i obtained as a radial expansion of Y_i by the technical efficiency score of DMU i, i.e.,

$$Y_i^\partial := Y_i \lambda(X_i, Y_i). \tag{2.11}$$

Here, n can be the number of all DMUs in a population (e.g., firms in an industry) or a sub-population (e.g., certain type of firms in an industry) or of a sample from a population or a sub-population, while from the next section we will treat n as the sample size. Such aggregation can be also implemented for different layers, first aggregating within sub-groups then aggregating between sub-groups into larger groups using between the group weights, as derived by Simar and Zelenyuk (2007), but at higher notational complexity. For the sake of simplicity, here we will focus on aggregation into one group only, e.g., when the focus is on obtaining the *industry efficiency*.

A particular feature of such aggregation scheme is that the three aggregate efficiency scores, (2.8), (2.9) and (2.10), are related to each other in a similar fashion as the individual efficiency scores. That is we have an analogue of decomposition (2.7), at the aggregate level, given by

$$\overline{RE}_n = \overline{TE}_n \times \overline{AE}_n \tag{2.12}$$

In the sections that follow, we take this aggregation scheme above as given and derive asymptotic properties of the estimates of these aggregate efficiencies.

3 Aggregate Efficiency as Ratios of Simple Averages

While the form of aggregates given in (2.8)–(2.10) is intuitive in itself, indicating that each efficiency score is accounted by its relative weight in the group in the aggregation, these measures turn out to have another useful economic interpretation which appears to have been overlooked so far—they can be represented as ratios of totals or, equivalently, of simple means. This form turns out to be more convenient for our statistical analysis. Specifically, the aggregate revenue efficiency of a group given in (2.8) can be transformed into

$$\overline{RE}_n = \frac{w^T \sum_{i=1}^n Y_i^*}{w^T \sum_{i=1}^n Y_i} = \frac{\frac{1}{n}\sum_{i=1}^n w^T Y_i^*}{\frac{1}{n}\sum_{i=1}^n w^T Y_i} \tag{3.1}$$

where Y_i^* can be understood as the revenue-optimal counterfactual output vector for DMU i, which can be obtained as a radial expansion of Y_i by the revenue efficiency score of DMU i, i.e.,

$$Y_i^* := RE(X_i, Y_i, w) Y_i \tag{3.2}$$

In this sense, $w^T Y_i^*$ is the counterfactual revenue value for DMU i which would be equal to the maximal revenue for this DMU when the latter is obtained according to

optimization problem (2.5). Thus, by this simple transformation, we obtained another interesting economic interpretation of the aggregate revenue efficiency measure—as the ratio of the sum of maximal (counterfactual) revenues of the individuals in the group to the sum of actual revenues of the individuals in this group. Furthermore, dividing numerator and denominator by $1/n$ gives the analogous interpretation in terms of the equally-weighted means, which is particularly convenient for statistical analysis.

In a similar fashion, we can convert the aggregate technical efficiency given in (2.9) into

$$\overline{TE}_n = \frac{w^T \sum_{i=1}^n Y_i^\partial}{w^T \sum_{i=1}^n Y_i} = \frac{\frac{1}{n}\sum_{i=1}^n w^T Y_i^\partial}{\frac{1}{n}\sum_{i=1}^n w^T Y_i} \tag{3.3}$$

where Y_i^∂ is the counterfactual output vector for DMU i obtained as a radial expansion of Y_i by the technical efficiency score of DMU i, defined in (2.11).

Thus, the aggregate technical efficiency is represented as the ratio of simple average (or the sum) of the counterfactual revenues (obtained by correcting only for technical inefficiency of each DMU) of the group to the simple average (or the sum) of the actual revenues of this group.

Finally, we also can convert the aggregate allocative efficiency score given in (2.10), into

$$\overline{AE}_n = \frac{w^T \sum Y_i^*}{w^T \sum_{i=1}^n Y_i^\partial} = \frac{\frac{1}{n}\sum_{i=1}^n w^T Y_i^*}{\frac{1}{n}\sum_{i=1}^n w^T Y_i^\partial} \tag{3.4}$$

i.e., the aggregate allocative efficiency can also be represented as the ratio of simple average (or the sum) of the counterfactual revenues (obtained by correcting for revenue inefficiency of each DMU in the group) to the simple average (or the sum) of the counterfactual revenues obtained by correcting only for the technical inefficiency of each DMU of this group.

Given that the aggregate efficiency scores can be interpreted as ratios of simple means, we can adapt statistical theory of the ratio of sample means and in the next section we summarize some of the main points of it. For the sake of brevity, in the rest of the paper we will focus only on the aggregate technical efficiency, because it appears to be the most popular in practice, and note that similar theory can be developed for the revenue efficiency and the (output oriented) allocative efficiency. It is also worth noting that this same ideas can be adapted to the case of aggregation of other measures based on the Farrell-type efficiency measures, e.g., input oriented technical, allocative and cost efficiency measures, the Malmquist Productivity Indexes, the Hicks-Moorsteen

Productivity Indexes, the scale efficiency measures, etc. (See related aggregation theory results in Zelenyuk (2006), Mayer and Zelenyuk (2014a), Zelenyuk (2015)).

4 Asymptotics when True Efficiency is Known

We have seen that when analyzing aggregate efficiencies, we are interested in a ratio of two means, both being estimated by appropriate empirical means. For the case of aggregate technical efficiencies, the two means are the expectations of the two elements of the random vector

$$\left[\begin{array}{c} Z_i^{\partial} \\ Z_i \end{array} \right] = \left[\begin{array}{c} \lambda(X_i, Y_i) Z_i \\ Z_i \end{array} \right], \tag{4.1}$$

where Z_i is a scalar valued function (typically a linear combination) of the vector Y_i, e.g., $Z_i = w^T Y_i$ where w is a given q-dimensional vector of prices. So, Z_i^{∂} can be viewed as the monetary value (under prices w^T) of the projection of the observation Y_i on the boundary of Ψ: $Z_i^{\partial} \equiv w^T Y_i^{\partial} = \lambda(X_i, Y_i) w^T Y_i$.[2]

We denote the first two moments of these two variables as

$$\mu = \left[\begin{array}{c} \mathbb{E}(Z_i^{\partial}) \\ \mathbb{E}(Z_i) \end{array} \right] = \left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right], \; \Sigma = \left[\begin{array}{cc} \mathrm{VAR}(Z_i^{\partial}) & \mathrm{COV}(Z_i^{\partial}, Z_i) \\ \mathrm{COV}(Z_i^{\partial}, Z_i) & \mathrm{VAR}(Z_i) \end{array} \right] = \left[\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{array} \right] \tag{4.2}$$

with implicit assumption that all these moments are finite.

Our parameter of interest is $\tau = \mu_1/\mu_2$ and if the λ function would be known, a natural estimator would be given by the ratio of the sample means, i.e.,

$$\widehat{\tau}_n = \frac{\widehat{\mu}_{1,n}}{\widehat{\mu}_{2,n}} = \frac{(1/n)\sum_{i=1}^n Z_i^{\partial}}{(1/n)\sum_{i=1}^n Z_i}. \tag{4.3}$$

By standard arguments, based on Central Limit Theorem (CLT) and the delta method (first order Taylor expansion) we obtain the well-known results (e.g., see Kendall and Stuart, 1977)

$$\sqrt{n}\left(\frac{\widehat{\tau}_n - \tau}{\sqrt{V(\widehat{\tau}_n)}} \right) \xrightarrow{\mathcal{L}} N(0,1), \tag{4.4}$$

where

$$V(\widehat{\tau}_n) = nVAR(\widehat{\tau}_n) = \tau^2 \left(\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} - 2\frac{\sigma_{12}}{\mu_1 \mu_2} \right). \tag{4.5}$$

[2]Clearly the boundary of Ψ is a non-random object, but Z_i^{∂} is random as a function of the random (X_i, Y_i).

This result can be used for doing inference on τ. In practice, applying Slutsky Theorem, we replace the unknown moments appearing in $V(\widehat{\tau}_n)$ by some consistent estimators (their empirical versions), providing $\widehat{V}(\widehat{\tau}_n) \xrightarrow{p} V(\widehat{\tau}_n)$.

Of course, we do not observe Z_i^{∂} because the values of the function λ are unknown. However we have estimates of λ, like the nonparametric envelopment estimators, discussed in more detail in the next section. So one would be tempted to replace $\widehat{\mu}_{1,n}$, the unavailable numerator of $\widehat{\tau}_n$, by its estimator $\widehat{\widehat{\mu}}_{1,n} = n^{-1}\sum_{i=1}^n \widehat{Z}_i^{\partial}$, where $\widehat{Z}_i^{\partial} = \widehat{\lambda}(X_i, Y_i)Z_i$, providing

$$\widehat{\widehat{\tau}}_n = \frac{\widehat{\widehat{\mu}}_{1,n}}{\widehat{\mu}_{2,n}}. \tag{4.6}$$

Unfortunately, the envelopment estimators suffer from an inherent bias and averaging in the numerator of $\widehat{\widehat{\tau}}_n$ drives its variance to zero faster than convergence of the bias to zero. So usual central limit theorems fail, unless, as explained below, we use an appropriate correction for the bias and, when the dimension of sample space $p+q$ increases, we use an appropriate rescaling of the test statistics to control for convergence rates of both bias and variance.

5 Asymptotics for Estimators of Individual Efficiency

To estimate the technology frontier Ψ^{∂} and the related values of efficiency scores, λ, many methods were suggested in the literature, and here we focus on the nonparametric envelopment estimators such as DEA and FDH. Specifically, if one is ready to assume Constant Returns to Scale (CRS) for Ψ^{∂}, then the CRS-DEA estimator is given by the following linear programming (LP) problem

$$\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) \equiv \max_{\xi_1,\ldots,\xi_n,\theta} \{\theta \ : \ \sum_{k=1}^n \xi_k Y_k \geqq \theta\, Y_i, \ \sum_{k=1}^n \xi_k X_k \leqq X_i, \ \theta \geq 0, \forall \xi_k \geq 0\}. \tag{5.1}$$

which is also known as CCR approach (due to Charnes, Cooper and Rhodes (1978)).

Alternatively, if one is ready to assume Variable Returns to Scale (VRS) for Ψ^{∂}, then the VRS-DEA estimator is given by the following LP problem

$$\begin{aligned}\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) \equiv \max_{\xi_1,\ldots,\xi_n,\theta} \{\theta \ : \ & \sum_{k=1}^n \xi_k Y_k \geqq \theta\, Y_i, \ \sum_{k=1}^n \xi_k X_k \leqq X_i, \\ & \theta \geq 0, \ \forall \xi_k \geq 0, \ \sum_{k=1}^n \xi_k = 1\}.\end{aligned} \tag{5.2}$$

i.e., constraint $\sum_{k=1}^{n} \xi_k = 1$ is added to the CCR specification. Moreover, if convexity of Ψ is not assumed then one can use the FDH estimator (due to Deprins, Simar and Tulkens (1984)) given by

$$\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) \equiv \max_{\theta}\{\theta \ : \ Y_k \geqq \theta\, Y_i, \ X_k \leqq X_i, \ \theta \geq 0\}. \tag{5.3}$$

These DEA and FDH estimators for λ have well established statistical properties when evaluated at fixed points (x, y), e.g., see Simar and Wilson, (2015) for a recent survey. Furthermore, a recent theoretical results of Kneip, Simar and Wilson (2015a), hereafter KSW, extend these to the case where, like in our case, one evaluates the estimators at random points (X_i, Y_i) to build some statistics of interest (here means). The basic results established in KSW can be summarized as follows: for the FDH and DEA estimators, under some regularity conditions and as $n \to \infty$,

$$E\left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) - \lambda(X_i, Y_i)\right) = C_0 n^{-\kappa} + R_{n,\kappa} \tag{5.4}$$

$$E\left(\left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) - \lambda(X_i, Y_i)\right)^2\right) = o\big(n^{-\kappa}\big), \tag{5.5}$$

$$\left|\mathrm{COV}\left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) - \lambda(X_i, Y_i),\ \widehat{\lambda}(X_j, Y_j \mid \mathcal{X}_n) - \lambda(X_j, Y_j)\right)\right| = o\big(n^{-1}\big) \tag{5.6}$$

Also, we make it explicit in this notation that the estimator is computed with respect to the sample $\mathcal{X}_n$ because it will be important for subsequent developments to be explicit on this.

Importantly, note that the values of the constant C_0, the rate κ, and the remainder term $R_{n,\kappa}$ depends on which estimator is used. Specifically, we consider the following cases:

- When we assume constant returns to scale, then the CRS-DEA estimator (5.1)achieve the rate

$$\kappa = 2/(p+q) \quad \text{and} \quad R_{n,\kappa} = O(n^{-3\kappa/2}(\log n)^{\alpha_1})$$

- When we assume variable returns to scale, then it is more appropriate to use the VRS-DEA estimator (5.2), which achieves the rate

$$\kappa = 2/(p+q+1) \quad \text{and} \quad R_{n,\kappa} = O(n^{-3\kappa/2}(\log n)^{\alpha_2})$$

- When we do not assume convexity but maintain the free disposability (monotonicity) assumption (see Assumption 2.2), then appropriate estimator is the FDH estimator (5.3). It achieves the rate

$$\kappa = 1/(p+q) \quad \text{and} \quad R_{n,\kappa} = O(n^{-2\kappa}(\log n)^{\alpha_3})$$

In all these three cases, the values of $\alpha_j > 1$, $j = 1, 2, 3$ are given in KSW; for purposes of the results needed here, the $\log n$ factor contained in $R_{n,\kappa}$ does not play a role and can be ignored. The results outlined here are valid under a set of corresponding regularity assumptions (see Theorems 3.1, 3.2 and 3.3 in KSW), typically smoothness of the frontier and of the density $f(x, y)$ of (X, Y) on Ψ. In addition, the latter density has to be strictly positive in a neighborhood or Ψ^{∂}. These regularity conditions are summarized in Appendix A.

6 Asymptotics when True Efficiency is Unknown

To start the analysis, we first have to restate the basic results from KSW summarized in (5.4)–(5.6). The basic quantity of interest here is the difference between the true projection of the random variable $Z_i = w^T Y_i$ on the frontier Z_i^{∂} and its estimates obtained from a sample $\mathcal{X}_n$, that we now explicitly write as $\widehat{Z}_i^{\partial}(\mathcal{X}_n)$ to stress the fact that the efficiency score is estimated with the full sample of data $\mathcal{X}_n$. So we have

$$\widehat{Z}_i^{\partial}(\mathcal{X}_n) - Z_i^{\partial} = \left(\widehat{\lambda}(X_i, Y_i|\mathcal{X}_n) - \lambda(X_i, Y_i)\right) Z_i, \tag{6.1}$$

from which it appears that, under the mild assumption that the first two moments of Z_i are finite, the order of the first moments of the left hand side of (6.1) inherits those from the first factor of the right hand side. So we have the following analogous results, under the same regularity conditions.

Corollary 6.1. *Under the appropriate set of Assumptions described in Theorem 3.1, 3.2 or 3.3 of Kneip et al. (2015a), and as $n \to \infty$*

$$E\left(\widehat{Z}_i^{\partial}(\mathcal{X}_n) - Z_i^{\partial}\right) = C_1 n^{-\kappa} + R_{n,\kappa} \tag{6.2}$$

$$E\left(\left(\widehat{Z}_i^{\partial}(\mathcal{X}_n) - Z_i^{\partial}\right)^2\right) = o\big(n^{-\kappa}\big), \tag{6.3}$$

$$\left|COV\left(\widehat{Z}_i^{\partial}(\mathcal{X}_n) - Z_i^{\partial},\ \widehat{Z}_j^{\partial}(\mathcal{X}_n) - Z_j^{\partial}\right)\right| = o\big(n^{-1}\big) \tag{6.4}$$

for all $i, j \in \{1, \ldots, n\}$, $i \neq j$ and where C_1 is some finite constant and $R_{n,\kappa} = o\left(n^{-\kappa}\right)$ has the same order as above, according the type of estimator used.

Now we can follow the same arguments as in KSW, adding where it is necessary the additional results we need for our setup. First, we have the following lemma summarizing important properties for the estimator $\widehat{Z}_i^{\partial}(\mathcal{X}_n)$ itself.

Lemma 6.1. *Under the appropriate set of Assumptions described in Theorem 3.1, 3.2 or 3.3 of Kneip et al. (2015a), we have as* $n \to \infty$

$$(i) \qquad E\left(\widehat{Z}_i^{\partial}(\mathcal{X}_n)\right) = \mu_1 + C_1 n^{-\kappa} + R_{n,\kappa} \tag{6.5}$$

$$(ii) \qquad VAR\left(\widehat{Z}_i^{\partial}(\mathcal{X}_n)\right) = \sigma_1^2 + o\left(n^{-\kappa/2}\right) \tag{6.6}$$

$$(iii) \qquad COV\left(\widehat{Z}_i^{\partial}(\mathcal{X}_n), Z_i\right) = \sigma_{12} + o\left(n^{-\kappa/2}\right) \tag{6.7}$$

Proof. The proofs of (i) and (ii) follow directly the proofs in Lemma 1 of KSW. The point (iii) goes as follows:

$$\begin{aligned}
\mathrm{COV}\left(\widehat{Z}_i^{\partial}(\mathcal{X}_n), Z_i\right) &= E\left([\widehat{Z}_i^{\partial}(\mathcal{X}_n) - E(\widehat{Z}_i^{\partial}(\mathcal{X}_n))](Z_i - \mu_2)\right) \\
&= E\left([\widehat{Z}_i^{\partial}(\mathcal{X}_n) - (Z_i^{\partial} - \mu_1) + (Z_i^{\partial} - \mu_1) - E(\widehat{Z}_i^{\partial}(\mathcal{X}_n))](Z_i - \mu_2)\right) \\
&= \sigma_{12} + E\left([\widehat{Z}_i^{\partial}(\mathcal{X}_n) - E(\widehat{Z}_i^{\partial}(\mathcal{X}_n)) - (Z_i^{\partial} - \mu_1)](Z_i - \mu_2)\right).
\end{aligned}$$

By the Cauchy-Schwartz inequality the latter term is bounded by $\sqrt{\mathrm{VAR}(\widehat{Z}_i^{\partial}(\mathcal{X}_n) - Z_i^{\partial})\sigma_2^2}$ which, due to (6.2) and (6.3), is of order $o\left(n^{-\kappa/2}\right)$, completing the proof. □

The following theorem establishes the key properties of $\widehat{\widehat{\mu}}_{1,n}$, the numerator of $\widehat{\widehat{\tau}}_n$, and provides consistent estimators of σ_1^2 and of σ_{12}.

Theorem 6.1. *Let* $\widetilde{\mu}_{1,n} = E(\widehat{\widehat{\mu}}_{1,n})$. *Under the assumptions of Lemma 6.1, we have*

$$(i) \qquad \widetilde{\mu}_{1,n} = \mu_1 + C_1 n^{-\kappa} + R_{n,\kappa} \tag{6.8}$$

$$(ii) \qquad \widehat{\widehat{\mu}}_{1,n} - \widetilde{\mu}_{1,n} = \widehat{\mu}_{1,n} - \mu_1 + o_p\left(n^{-1/2}\right) \tag{6.9}$$

$$(iii) \qquad \sqrt{n}\big(\widehat{\widehat{\mu}}_{1,n} - \widetilde{\mu}_{1,n}\big) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2) \tag{6.10}$$

$$(iv) \qquad \widehat{\widehat{\sigma}}_{1,n}^2 = n^{-1} \sum_{i=1}^{n} \big(\widehat{Z}_i^{\partial}(\mathcal{X}_n) - \widehat{\widehat{\mu}}_{1,n}\big)^2 \xrightarrow{p} \sigma_1^2 \tag{6.11}$$

$$(v) \qquad \widehat{\widehat{\sigma}}_{12,n} = n^{-1} \sum_{i=1}^{n} \big(\widehat{Z}_i^{\partial}(\mathcal{X}_n) - \widehat{\widehat{\mu}}_{1,n}\big)\big(Z_i - \widehat{\mu}_2\big) \xrightarrow{p} \sigma_{12}. \tag{6.12}$$

Proof. We know (from usual central limit theorem) that $\sqrt{n}\big(\widehat{\mu}_{1,n} - \mu_1\big) \xrightarrow{\mathcal{L}} N(0, \sigma_1^2)$. Now, consider $\zeta_n = n^{-1} \sum_{i=1}^{n} \big(\widehat{Z}_i^{\partial}(\mathcal{X}_n) - Z_i^{\partial}\big)$. From (6.2) we can conclude that $E(\zeta_n) = C_1 n^{-\kappa} + R_{n,\kappa}$ and from (6.3) with (6.4) we conclude that $\mathrm{VAR}(\zeta_n) = o\big(n^{-1}\big)$, so that $\zeta_n - E(\zeta_n) = o_p\big(n^{-1/2}\big)$. These, in turn, help establishing (i), (ii) and (iii). In particular, we also have $\widehat{\widehat{\mu}}_{1,n} \xrightarrow{p} \mu_1$.

Meanwhile, for the estimators of the variance, note that we have the following

$$\widehat{\widehat{\sigma}}_{1,n}^2 = n^{-1} \sum_{i=1}^{n} \big(\widehat{Z}_i^{\partial}(\mathcal{X}_n)\big)^2 - \widehat{\widehat{\mu}}_{1,n}^2 \xrightarrow{p} E\big(\widehat{Z}_i^{\partial}(\mathcal{X}_n)\big)^2 - \mu_1^2,$$

and by (i) and (ii) of Lemma 6.1, we have $E\big(\widehat{Z}_i^{\partial}(\mathcal{X}_n)\big)^2 - \mu_1^2 = \sigma_1^2$ when $n \to \infty$. The argument is similar for the covariance, by using (iii) of Lemma 6.1. This completes the proof. □

Now it is easier to understand why there is a problem in trying to extend the CLT, available for the unobservable $\widehat{\tau}_n$ in (4.4), to the estimator $\widehat{\widehat{\tau}}_n$. Indeed, we have

$$\widehat{\widehat{\tau}}_n = \widehat{\tau}_n + \frac{\widehat{\widehat{\mu}}_{1,n} - \widehat{\mu}_{1,n}}{\widehat{\mu}_{2,n}}, \tag{6.13}$$

and so one obtains formally

$$\sqrt{n}\frac{\widehat{\tau}_n - \tau}{\sqrt{V(\widehat{\tau}_n)}} = \frac{\sqrt{n}}{\sqrt{V(\widehat{\tau}_n)}}\left\{\widehat{\widehat{\tau}}_n - \tau - \frac{\widehat{\widehat{\mu}}_{1,n} - \widehat{\mu}_{1,n}}{\widehat{\mu}_{2,n}}\right\} \xrightarrow{\mathcal{L}} N(0,1), \tag{6.14}$$

where the bias introduced by replacing the unobserved Z_i^{∂} by their estimates $\widehat{Z}_i^{\partial}$ is

$$\begin{aligned}\frac{\widehat{\widehat{\mu}}_{1,n} - \widehat{\mu}_{1,n}}{\widehat{\mu}_{2,n}} &= \frac{\widetilde{\mu}_{1,n} - \mu_{1,n}}{\mu_{2,n}} + o_p\big(n^{-1/2}\big) \\ &= \frac{C_1 n^{-\kappa} + R_{n,\kappa}}{\mu_2} + o_p\big(n^{-1/2}\big),\end{aligned} \tag{6.15}$$

where $R_{n,\kappa} = o\big(n^{-\kappa}\big)$ has a particular value depending on the type of estimator which is used.

Now we see clearly that in most of the practical cases, we have a problem to use this results:

(i) If $\kappa > 1/2$ then the bias term can be ignored because it vanishes even when multiplied by $\sqrt{n}$. Note that this translates into the cases when $p+q \leq 3$ for CRS-DEA, $p+q \leq 2$ for VRS-DEA and $p+q \leq 1$ (i.e. $p=0$ and $q=1$), for the FDH case, which are quite limited for practice.

(ii) If $\kappa = 1/2$ the bias converges to a unknown constant and the result cannot be used;

(iii) If $\kappa < 1/2$, the bias explodes to infinity when $n \to \infty$. Here, it might be tempting to replace the factor $\sqrt{n}$ in (6.14) by n^{γ} with $\gamma < \kappa \leq 1/2$, but it is not a solution because the variance would then converge to zero.

In a similar situation, KSW suggest to introduce a suitable estimator of the bias that will allow to use the CLT for slightly smaller values of κ and in more general situation (bigger values of $p+q$), they suggest also to rescale the estimator of τ for controlling both the bias and the variance and we follow their strategy here as well.

7 Estimation Procedure for the Bias

To deal with the bias problem, we use a type of generalized jackknife technique aiming to reduce the bias of $\widehat{\widehat{\mu}}_{1,n}$. It is based on the principle of splitting randomly the full sample $\mathcal{X}_n$ into two parts of size $n_1 = \lfloor n/2 \rfloor$, where $\lfloor a \rfloor$ denotes integer part of a, and $n_2 = n - n_1$ respectively and from these two subsamples, build a bias estimate. To reduce the variance of the bias estimate, as suggested in Kneip, Simar and Wilson (2015b), the operations below will be repeated a large number of times by shuffling the observations before each split of the sample $\mathcal{X}_n$, for $\ell = 1, \ldots, L$ where $L \ll \binom{n}{n_1}$. Set $\ell = 1$. Then $\mathcal{X}^{(1)}_{n_1,\ell}$ denote a random subset of n_1 observations in $\mathcal{X}_n$ and $\mathcal{X}^{(2)}_{n_2,\ell}$ is the set of remaining points. So the sample is split evenly if n is even, or almost evenly (with a difference of one observation) where n is odd. For simplicity, we assume n is even, so $n_1 = n_2 = n/2$. Now, for $j = 1, 2$, let

$$\begin{aligned} \widehat{\widehat{\mu}}^{(j)}_{1,n_j,\ell} &= n_j^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}^{(j)}_{n_j,\ell}} \widehat{Z}^{\partial}_i(\mathcal{X}^{(j)}_{n_j,\ell}) \\ &= n_j^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}^{(j)}_{n_j,\ell}} \widehat{\lambda}(X_i, Y_i|\mathcal{X}^{(j)}_{n_j,\ell}) Z_i, \end{aligned} \tag{7.1}$$

where the notation introduced above indicates that the nonparametric estimators of the efficiency scores are obtained by using only the corresponding subsamples for $j = 1, 2$.

Now, note that by (ii) in Theorem 6.1, we have

$$\widehat{\widehat{\mu}}^{(j)}_{1,n_j,\ell} - \widetilde{\mu}_{1,n_j} = n_j^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}^{(j)}_{n_j,\ell}} (Z^{\partial}_i - \mu_1) + o_p\big(n^{-1/2}\big), \text{for } j = 1, 2. \tag{7.2}$$

Now we define

$$\widehat{\widehat{\mu}}^{*}_{1,n,\ell} = \frac{1}{2}\left(\widehat{\widehat{\mu}}^{(1)}_{1,n_1,\ell} + \widehat{\widehat{\mu}}^{(2)}_{1,n_2,\ell}\right), \tag{7.3}$$

We have $\widetilde{\mu}_{1,n_1} = \widetilde{\mu}_{1,n_2}$ (if n is odd the difference is negligible) and so it is easy to check that

$$\widehat{\widehat{\mu}}^{*}_{1,n,\ell} - \widetilde{\mu}_{1,n_1} = n^{-1} \sum_{i=1}^{n} \big(Z^{\partial}_i - \mu_1\big) + o_p\big(n^{-1/2}\big). \tag{7.4}$$

By subtracting (ii) in Theorem 6.1 from (7.4), we obtain

$$\begin{aligned} \widehat{\widehat{\mu}}^{*}_{1,n,\ell} - \widehat{\widehat{\mu}}_{1,n} &= \widetilde{\mu}_{1,n_1} - \widetilde{\mu}_{1,n} + o_p\big(n^{-1/2}\big) \\ &= C_1(2^{\kappa} - 1)n^{-\kappa} + R_{n,\kappa} + o_p\big(n^{-1/2}\big), \end{aligned}$$

where the remainder has the same order as the original $R_{n,\kappa}$. Therefore

$$(2^{\kappa} - 1)^{-1}\big(\widehat{\widehat{\mu}}^{*}_{1,n,\ell} - \widehat{\widehat{\mu}}_{1,n}\big) = C_1 n^{-\kappa} + R_{n,\kappa} + o_p\big(n^{-1/2}\big), \tag{7.5}$$

provides an estimator of the bias term of $\widehat{\widehat{\mu}}_{1,n}$, and the error is of order smaller than $n^{-1/2}$.

Repeating the operations above for $\ell = 1, \ldots, L$ and averaging reduces the variance of the estimate by a factor L^{-1}, as is usual with jackknife estimators.

Finally the bias estimator of $\widehat{\widehat{\tau}}_n$ we introduced in (6.15) is defined by

$$\widehat{B}_{n,\kappa} = \frac{L^{-1}\sum_{\ell=1}^{L}(2^{\kappa}-1)^{-1}\big(\widehat{\widehat{\mu}}^{*}_{1,n,\ell} - \widehat{\widehat{\mu}}_{1,n}\big)}{\widehat{\mu}_2}, \tag{7.6}$$

which is an estimate of $\mu_2^{-1}C_1 n^{-\kappa}$. Also note that using $\widehat{\mu}_2$, the empirical mean of Z_i, in place of μ_2 does not affect the results because this add a term of smaller order.

8 Central Limit Theorems for $\widehat{\widehat{\tau}}_n$

Using the bias correction $\widehat{B}_{n,\kappa}$ outlined in the previous section, we are now able to obtain two important for practice results, summarized in the two theorems below.

Theorem 8.1. *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*

$$\sqrt{n}\,\frac{\widehat{\widehat{\tau}}_n - \widehat{B}_{n,\kappa} - \tau + R_{n,\kappa}}{\sqrt{\widehat{\widehat{V}}(\widehat{\tau}_n)}} \xrightarrow{\mathcal{L}} N(0,1), \tag{8.1}$$

where $\widehat{B}_{n,\kappa}$ is the bias estimate described above, and

$$\widehat{\widehat{V}}(\widehat{\tau}_n) = \widehat{\widehat{\tau}}_n^{2}\left(\frac{\widehat{\widehat{\sigma}}^{2}_{1,n}}{\widehat{\widehat{\mu}}^{2}_{1,n}} + \frac{\widehat{\sigma}^{2}_{2,n}}{\widehat{\mu}^{2}_{2,n}} - 2\,\frac{\widehat{\widehat{\sigma}}_{12,n}}{\widehat{\widehat{\mu}}_{1,n}\widehat{\mu}_{2,n}}\right) \xrightarrow{p} V(\widehat{\tau}_n), \tag{8.2}$$

with $\widehat{\widehat{\sigma}}^{2}_{1,n}$ (resp. $\widehat{\sigma}^{2}_{2,n}$) is the empirical variance of the $\widehat{Z}_i^{\partial}$ (resp. Z_i) and $\widehat{\widehat{\sigma}}_{12,n}$ is the empirical covariance between $\widehat{Z}_i^{\partial}$ and Z_i.

Proof. The theorem is a direct consequence of the developments above. Note that for the values of κ as stated in the theorem, it can easily checked that $\sqrt{n}R_{n,\kappa}$ converges to zero, where the particular order $R_{n,\kappa}$ is different for each selected estimator (CRS-DEA, VRS-DEA or FDH), as was discussed in the last paragraph of Section 5. So the result is useful for practical inference. The consistency of the variance estimator is immediate using, in particular, the results of Theorem 6.1 above. □

In the case $p+q$ is bigger than the values given in the preceding theorem, then another central limit theorem has to be advocated and we summarize it below.

Theorem 8.2. *Under the appropriate set of Assumptions described in Theorem 3.1, 3.2 or 3.3 of Kneip et al. (2015), for $p+q>5$ if CRS-DEA estimator is used and Ψ^{∂} is globally CRS and convex, for $p+q>4$ if VRS-DEA estimator is used and Ψ^{∂} is convex and for $p+q>3$ if FDH estimator is used and Ψ^{∂} satisfies free disposability of inputs and outputs, when $n\to\infty$ we have*

$$n^{\kappa}\,\frac{\widehat{\widehat{\tau}}_{n_\kappa}-\widehat{B}_{n,\kappa}-\tau+R_{n,\kappa}}{\sqrt{\widehat{\widehat{V}}(\widehat{\tau}_n)}}\xrightarrow{\mathcal{L}}N(0,1), \tag{8.3}$$

with $\widehat{\widehat{\tau}}_{n_\kappa}$ being a subsample version of $\widehat{\widehat{\tau}}_n$, 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

$$\widehat{\widehat{\tau}}_{n_\kappa}=\frac{n_\kappa^{-1}\sum_{\{j|(X_j,Y_j)\in\mathcal{X}^*_{n_\kappa}\}}\widehat{Z}_j^{\partial}}{n_\kappa^{-1}\sum_{\{j|(X_j,Y_j)\in\mathcal{X}^*_{n_\kappa}\}}Z_j}, \tag{8.4}$$

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

Proof. Here we simply use the fact that the standard CLT theorem holds also for $\widehat{\tau}_{n_\kappa}$ since $n_\kappa\to\infty$ as $n\to\infty$. Subsampling deteriorates the rate of convergence but the CLT is still valid. Since in all the cases, $R_{n,\kappa}=o\big(n^{-\kappa}\big)$, these results are useful for practical inference.

Note that in (8.4), the averages are computed over the $j=1,\ldots,n_\kappa$ random elements of the subsample $\mathcal{X}^*_{n_\kappa}$, but the estimates $\widehat{\lambda}(X_j,Y_j|\mathcal{X}_n)$ at the numerator are relative to all the data points in $\mathcal{X}_n$, as the explicit notation indicates. Note also that we use here the same estimator for the bias and the variance as in Theorem 8.1. □

Remark 8.1. It should be noticed that, as pointed in KSW for the averages of efficiency scores, Theorem 8.1 could be stated as being valid for $\kappa\geq 1/2$ and Theorem 8.2 for $\kappa<1/2$, where κ is determined by the dimension of the input-output space and by the chosen estimator. This is true since in all the cases, $R_{n,\kappa}=o\big(n^{-\kappa}\big)$. In the two theorems, the values of κ are tuned depending on the more specific value of $R_{n,\kappa}$. It is easy to check that for some estimators and some values of κ both approximations will be applicable. For the VRS-DEA and CRS-DEA estimators, if $\kappa=2/5$, so $n_\kappa<n$. Theorem 8.1, using the scaling $\sqrt{n}$ has a remainder term $\sqrt{n}R_{n,\kappa}=O\big(n^{-1/10}\big)$, whereas Theorem 8.2, using the scaling n^{κ} has a remainder term $n^{\kappa}R_{n,\kappa}=O\big(n^{-1/5}\big)$. Thus, for $\kappa=2/5$, we may expect better approximation by using Theorem 8.2 rather than Theorem 8.1. The same is true for the FDH case when $\kappa=1/3$ and $n_\kappa<n$, Theorem 8.1 has a remainder of order $O\big(n^{-1/6}\big)$ after scaling by $\sqrt{n}$, whereas, the remainder term in Theorem 8.2 is only of order $O\big(n^{-1/3}\big)$ because the scaling factor is only n^{κ}.

9 Confidence Intervals for τ

Using the results of the above theorems, confidence intervals for τ can be easily built by using standard normal quantiles. 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 and for $p+q \leq 3$ if FDH estimator is used and Ψ^{∂} satisfies free disposability of inputs and outputs, an asymptotically correct $(1-\alpha)$-confidence interval for τ, is given by

$$\left[\widehat{\widehat{\tau}}_n - \widehat{B}_{n,\kappa} \pm z_{1-\alpha/2}\sqrt{\frac{\widehat{\widehat{V}}(\widehat{\tau}_n)}{n}}\right], \tag{9.1}$$

where $z_{1-\alpha/2}$ is the corresponding quantile of the standard normal distribution.

For larger values of $p+q$, we should rather use the asymptotically correct $(1-\alpha)$-confidence interval for τ, given by

$$\left[\widehat{\widehat{\tau}}_{n_\kappa} - \widehat{B}_{n,\kappa} \pm z_{1-\alpha/2}\sqrt{\frac{\widehat{\widehat{V}}(\widehat{\tau}_n)}{n_\kappa}}\right], \tag{9.2}$$

where we remind that $n_\kappa = \lfloor n^{2\kappa} \rfloor$.

Of course for the DEA cases, when $\kappa = 2/5$ and for the FDH case when $\kappa = 1/3$, the two intervals will be different but are asymptotically equivalent (see Remark 8.1).

10 Concluding Remarks

In this paper we developed asymptotic theory for aggregate efficiency measures (e.g., such as industry efficiency, etc.) that are constructed as the weighted averages of individual efficiency measures.

The key difficulty of the matter here is that when the individual true efficiency scores being aggregated are not observed but estimated with DEA or FDH, an application of standard central limit theory is problematic due to inherent bias in DEA and FDH estimators. This was first pointed out by Kneip, Simar and Wilson (2015a,b) and here we used these two recent works to develop similar asymptotic theory for the case of aggregate efficiency measures. In a nutshell, the strategy we used to derive the asymptotic theory followed three steps:

(i) converting the weighted aggregate efficiencies described above into ratios of the simple (non-weighted) means,

(ii) adapting statistical theory of the ratios of sample means to our aggregate efficiency scores assuming the individual efficiency scores are known,

(iii) extending the asymptotics to the case when individual efficiency scores are unknown but estimated, accounting for the characteristics pertinent to particular efficiency estimators at use and leveraging on results from Kneip, Simar and Wilson (2015a,b).

The main novelty of the paper is in the third step, where after deriving the key asymptotic results for the aggregate efficiency scores, we described the estimation procedure for the bias, derived two related central limit theorems and explained how use these results to estimate confidence intervals for the aggregate efficiency.

All in all, the asymptotic theory developed in this paper opens a path for more accurate and theoretically grounded statistical inference on aggregate efficiency estimates, such as industry efficiency or aggregate efficiency of distinct groups of firms within an industry, etc.

This paper also lays out the statistical foundation for many related theories and practical procedures to be developed. For example, the two central limit theorems we derived in this paper can also be used for developing statistical tests, e.g., for testing the equality of aggregate efficiencies of two independent groups of observations. A natural way to do so would be to adapt the procedure described in Kneip et al. (2015b), for testing equality of the mean efficiencies of two groups and use the two central limit theorems developed in this paper.

Furthermore, while for the sake of brevity in this paper we focused only on the aggregate output oriented technical efficiency, similar theories can be also developed for the revenue efficiency and the (output oriented) allocative efficiency. The same ideas can be also be adapted to the case of aggregation of other measures based on the Farrell-type efficiency measures, such as input oriented technical, allocative and cost efficiency measures, the input and output oriented scale efficiency measures, the input and output oriented Malmquist Productivity Indexes, the Hicks-Moorsteen Productivity Index, etc.

Finally, our approach can also be extended by analogy to other variations of envelopment estimators, such as the IDEA (Zhu (2004)), CAR-DEA (Cook and Zhu (2008)), DEA with non-homogeneous firms (Cook et al. (2013)), the related to DEA nonparametric models of optimizing behavior (Cherchye et al (2008)), Stochastic DEA and Stochastic FDH (e.g., Simar and Zelenyuk (2011)), to mention just a few.

A Appendix A: Regularity Assumptions

In this appendix we summarize the regularity Assumptions for Theorems 3.1, 3.2 or 3.3 of Kneip et al. (2015) that are advocated in all our theorems. In addition to Assumptions 2.1–2.2, the two following assumptions are required for the FDH, VRS-DEA, and CRS-DEA estimates.

Assumption A.1. *(i) The random variables* (X, Y) *possess a joint density* f *with support* $\mathcal{D} \subset \Psi$*; and (ii)* f *is continuously differentiable on* $\mathcal{D}$*.*

Assumption A.2. *(i)* $\mathcal{D}^* := \{(x, \lambda(x,y)y) \mid (x,y) \in \mathcal{D}\} \subset \mathcal{D}$*; (ii)* $\mathcal{D}^*$ *is compact; and (iii)* $f(x, \lambda(x,y)y) > 0$ *for all* $(x,y) \in \mathcal{D}$*.*

In the case we use the FDH estimators, we need the next assumption.

Assumption A.3. *(i)* $\lambda(x,y)$ *is twice continuously differentiable on* $\mathcal{D}$*; and (ii) all the first-order partial derivatives of* $\lambda(x,y)$ *with respect to* x *and* y *are nonzero at any point* $(x,y) \in \mathcal{D}$*.*

Recalling that the free disposability assumed in Assumption 2.2 implies that the frontier is weakly monotone, Assumption A.3 strengthens this by requiring the frontier to be strictly monotone with no constant segments.

Stronger assumptions are required by the DEA estimators. Both the VRS-DEA and CRS-DEA estimators require the next assumption.

Assumption A.4. $\lambda(x,y)$ *is three times continuously differentiable on* $\mathcal{D}$*.*

In addition, the VRS-DEA estimator requires the following assumption.

Assumption A.5. $\mathcal{D}$ *is* almost strictly convex*; i.e., for any* $(x,y), (\widetilde{x},\widetilde{y}) \in \mathcal{D}$ *with* $(\frac{x}{\|x\|}, y) \neq (\frac{\widetilde{x}}{\|\widetilde{x}\|}, \widetilde{y})$*, the set* $\{(x^*,y^*) \mid (x^*,y^*) = (x,y) + \alpha((\widetilde{x},\widetilde{y}) - (x,y))$ *for some* $0 < \alpha < 1\}$ *is a subset of the* interior *of* $\mathcal{D}$*.*

For the case of the CRS-DEA estimator, Assumption A.5 must be replaced by the following condition.

Assumption A.6. *(i) For any* $(x,y) \in \Psi$ *and any* $a \in [0,\infty)$*,* $(ax, ay) \in \Psi$*; (ii) the support* $\mathcal{D} \subset \Psi$ *of* f *is such that for any* (x,y)*,* $(\widetilde{x},\widetilde{y}) \in \mathcal{D}$ *with* $(\frac{x}{\|x\|}, \frac{y}{\|y\|}) \neq (\frac{\widetilde{x}}{\|\widetilde{x}\|}, \frac{\widetilde{y}}{\|\widetilde{y}\|})$*, the set* $\{(x^*,y^*) \mid (x^*,y^*) = (x,y) + \alpha((\widetilde{x},\widetilde{y}) - (x,y))$ *for some* $0 < \alpha < 1\}$ *is a subset of the* interior *of* $\mathcal{D}$*; (iii)* $\mathcal{D}$ *is a connected set; and (iv)* $(x,y) \notin \mathcal{D}$ *for any* $(x,y) \in \mathbb{R}^p_+ \times \mathbb{R}^q$ *with* $y^1 = 0$*, where* y^1 *denotes the first element of the vector* y*.*

References

[1] Charnes, A., Cooper, W.W. and E. Rhodes (1978) "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, 2, 6, 429-444.

[2] Cherchye, L. B. De Rock, F. Vermeulen (2008) "Analyzing Cost-Efficient Production Behavior Under Economies of Scope: A Nonparametric Methodology," *Operations Research* 56(1):204-221.

[3] Cook, W.D., J. Harrison, R. Imanirad, P. Rouse, and J. Zhu (2013) "Data Envelopment Analysis with Nonhomogeneous DMUs*" Operations Research* 61(3):666-676.

[4] Cook, W.D. and J. Zhu (2006) "Incorporating Multiprocess Performance Standards into the DEA Framework," *Operations Research* 54(4):656-665.

[5] Cook, W.D. and J. Zhu (2008) "CAR-DEA: Context-Dependent Assurance Regions in DEA," *Operations Research* 56(1):69-78.

[6] Cooper, W.W. L.M. Seiford and J. Zhu, editors (2011) *Handbook on Data Envelopment Analysis*, Second Edition, International Series in Operations Research & Management Sciences, Springer, New-York.

[7] Deprins, D., Simar, L. and H. Tulkens (1984), "Measuring Labor Efficiency in Post Offices," In M. Marchand, P. Pestieau and H. Tulkens (eds.), The Performance of Public Enterprises: Concepts and Measurements. Amsterdam: North Holland.

[8] Farrell, M.J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society. Series A (General), 120(3), 253-290.

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

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

[11] Kendall, M. and A. Stuart (1977), *Advanced Theory of Statistics*, Charles Griffin & Company, London.

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

[13] Kneip, A., Simar, L. and P.W. Wilson (2015b), Testing Hypothesis in Nonparametric Models of Production, in press *Journal of Business and Economic Statistics*, DOI: 10.1080/07350015.2015.1049747

[14] Koopmans, T.C. (1957) *Three Essays on the State of Economic Analysis*, McGraw-Hill, New York.

[15] Mayer, A. and V. Zelenyuk (2014), Aggregation of Malmquist productivity indexes allowing for reallocation of resources, *European Journal of Operational Research* 238, 3, 774–785.

[16] Simar, L. and P.W. Wilson (2015), Statistical Approaches for Nonparametric Frontier Models: A Guided Tour, *International Statistical Review*, 83, 1, 77–110.

[17] Simar, L. and V. Zelenyuk (2007), Statistical Inference for Aggregates of Farrell-type Efficiencies, *Journal of Applied Econometrics,* 22, 7, 1367-1394.

[18] Simar, L. and V. Zelenyuk (2011) "Stochastic FDH/DEA estimators for Frontier Analysis" *Journal of Productivity Analysis* 36:1, pp. 1-20.

[19] Zelenyuk, V. (2006) Aggregation of Malmquist Productivity Indexes, *European Journal of Operational Research* 174, 2, 1076-1086.

[20] Zelenyuk, V. (2015) Aggregation of Scale Efficiency, *European Journal of Operational Research* 240, 1, 269-277.

[21] Zhu, J. (2004) "Imprecise DEA via Standard Linear DEA Models with a Revisit to a Korean Mobile Telecommunication Company," *Operations Research* 52(2):323-329.