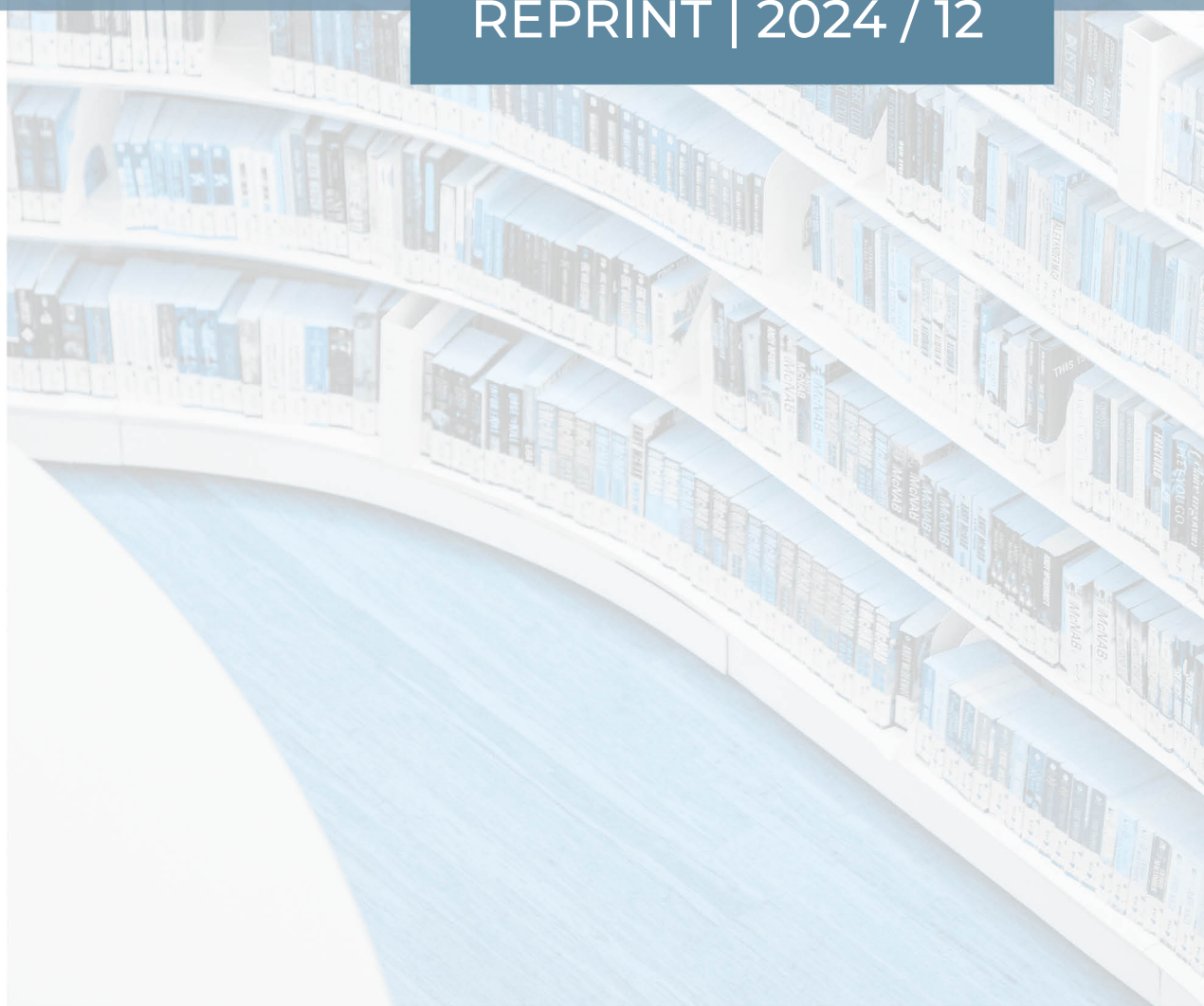
INFERENCE FOR AGGREGATE EFFICIENCY: THEORY AND GUIDELINES FOR PRACTITIONERS

Léopold Simar, Valentin Zelenyuk, Shirong Zhao

REPRINT | 2024 / 12



UCLouvain

ISBA | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

European Journal of Operational Research 316 (2024) 240–254

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Decision Support

Inference for aggregate efficiency: Theory and guidelines for practitioners

Léopold Simar [a], Valentin Zelenyuk [b,*], Shirong Zhao [c]

[a] Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA), LIDAM, Université Catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium

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

[c] School of Finance, Dongbei University of Finance and Economics, Dalian, Liaoning 116025, China

ARTICLE INFO

Keywords:
Data envelopment analysis
Efficiency
Non-parametric efficiency estimators
Free disposal hull
Aggregate efficiency

ABSTRACT

We expand and develop further the recently developed framework for the inference about the aggregate efficiency, extending the existing theory and providing guidelines for practitioners. In Monte Carlo simulations, we thoroughly examine the performance of the various improvement methods (compared with the original CLT results) for the aggregate input-oriented and output-oriented efficiency for different ranges of small samples and different dimensions of the production model. From the simulations, we conclude that: (i) when the sample sizes are relatively small (around 200 and less), the full variance correction method (adapted from Simar et al., 2023) with the data sharpening method (adapted from Nguyen et al., 2022) generally provides a better performance; (ii) when the sample sizes are relatively large, the full variance correction method without the data sharpening method is expected to perform better than the other suitable methods known to date. Finally, we use two well-known empirical data sets to illustrate the practical implementations and the differences across the existing methods to facilitate their use by practitioners.

1. Introduction

Nonparametric efficiency estimators are widely used tools to benchmark the performance of economic agents (firms, hospitals, banks, farms, etc.). The basic idea of these estimators is to use a sample of observed input–output pairs to estimate the production possibility set, i.e., the set describing how economic agents could use inputs to produce outputs, and then benchmark each producer relative to the upper boundary or frontier of the set.[1]

The nonparametric efficiency estimators that we focus on here include data envelopment analysis (DEA) (Banker et al., 1984; Charnes et al., 1978; Farrell, 1957) which imposes convexity on the production set, and free disposal hull (FDH) (Deprins et al., 1984) which relaxes the convexity assumption and assumes only free disposability on inputs and outputs. Based on the different assumptions on the returns to scale of the frontier, DEA estimators are further divided into constant returns to scale (CRS-DEA) and variable returns to scale (VRS-DEA) among others. Currently, VRS-DEA, CRS-DEA and FDH are the three widely used tools in the literature. Their theoretical and statistical properties are well-documented (e.g., see Kneip et al., 2015, 2016).

Recently, Kneip et al. (2015) established the central limit theorems (CLTs) results for the simple mean (or unweighted) technical efficiency estimates, which were further extended to the context of aggregate (or weighted) efficiency (Simar & Zelenyuk, 2018), conditional efficiency (Daraio et al., 2018), sources of Malmquist productivity change (Simar & Wilson, 2019), overall and allocative efficiency (Simar & Wilson, 2020), Malmquist productivity indices (Kneip et al., 2021) and aggregate Malmquist productivity indices (Pham et al., 2023), to name a few.

Compared to the simple mean efficiency, aggregate efficiency takes an economic weight of each DMU into account and can yield very different results and related conclusions. For example, according to the latest data from the World Bank, in 2022, the U.S. and Turkey made up about 25% and 1% of the world GDP, respectively.[2] Hence, when measuring the overall performance of the global economy, it is not appropriate to assume that the U.S. and Turkey have the same weight. Researchers need to carefully consider the aggregation methods and weights, which have been well studied by several papers (e.g., Färe et al., 2004; Färe & Zelenyuk, 2003; Mayer & Zelenyuk, 2019; Walheer, 2019; etc.). Specifically, Färe and Zelenyuk (2003) used Koopmans's aggregation theorem (Koopmans, 1957) to derive an aggregation method which is consistent with the grounds of economic

* Corresponding author.
E-mail addresses: leopold.simar@uclouvain.be (L. Simar), v.zelenyuk@uq.edu.au (V. Zelenyuk), shironz@163.com (S. Zhao).

[1] See Sickles and Zelenyuk (2019) for a comprehensive review and examples.

[2] See the GDP by countries/regions in 2022. Retrieved September 13, 2023, from https://databankfiles.worldbank.org/public/ddpext_download/GDP.pdf.

https://doi.org/10.1016/j.ejor.2024.01.028
Received 16 November 2023; Accepted 19 January 2024
Available online 1 February 2024

theory. This aggregation method is also adopted in this paper (see Section 2.2 for more details). The inference between the simple mean and aggregate efficiency is generally similar except that, compared to the simple mean efficiency, the variance of aggregate efficiency is more complicated, involving several moments and the covariance term in a nonlinear relationship (Simar & Zelenyuk, 2018).

It is worth noting that Simar and Zelenyuk (2018) only focused on aggregate output-oriented efficiency. In many contexts, researchers may need to use input-oriented efficiency, which is a similar yet different measure and which also implies different aggregation weights. For example, if the researcher is to identity the percentages of resources that have been over-utilized for the DMUs, then the input-oriented approach would be more relevant. See Cook et al. (2014) for more discussions on the choice of orientations for DEA methods. In principle, the CLT results for aggregate input-oriented efficiency are "analogous" to the output-oriented framework developed by Simar and Zelenyuk (2018) and so we make a very modest contribution from a pure theoretical perspective.

For the simple mean efficiency (Kneip et al., 2015) and aggregate efficiency (Simar & Zelenyuk, 2018), their extensive Monte Carlo (MC) experiments found that when the sample sizes increase, the coverages of the estimated confidence intervals based on the corresponding CLT results approximate the nominal coverage, which supports the established theoretical results. However, these authors also found that the estimated confidence intervals often under-cover the true values, especially for small sample sizes and high dimensions measured by the number of inputs and outputs specified. To further improve the finite sample performance of these CLT results, Simar and Zelenyuk (2020) proposed a *partial variance correction* method by adding squared bias of mean efficiency into the variance estimator, thus making the variance estimate slightly larger than for the case without adding the squared bias. It is worth noting that Simar and Zelenyuk (2020) only focused on the output orientation context for the simple mean and aggregate efficiency and left the input orientation to the readers. *Inter alia*, we will fill this gap as well.

To further improve the approximation, Nguyen et al. (2022) then proposed the *data sharpening* method through smoothing out the observations near the boundary, i.e., the so-called "spurious ones" and those with efficiency estimates close to one, where the magnitude of the closeness to the boundary is determined based on the sample size and the convergence rate of the estimators. Nguyen et al. (2022) found that the Simar and Zelenyuk (2020) partial variance correction method combined with the data sharpening method results in a better performance compared to the initial approaches of Kneip et al. (2015) and Simar and Zelenyuk (2020). It is also worth noting that Nguyen et al. (2022) only focused on the simple mean output-oriented efficiency and did not investigate whether the data sharpening method is also effective for the aggregate input-oriented and output-oriented efficiency. *Inter alia*, we will also fill this gap.

More recently, Simar et al. (2023) proposed a *full variance correction* method by using the bias-corrected individual efficiency estimates instead of the original individual efficiency estimates, to obtain the variance estimator for the simple mean efficiency. The simulation results in Simar et al. (2023) showed the effectiveness of their proposed method on improving the finite sample approximation of CLT results for the simple mean efficiency. However, similar as Nguyen et al. (2022), Simar et al. (2023) only focused on the simple mean output-oriented efficiency and did not investigate whether this method is also useful for the aggregate input-oriented and output-oriented efficiency. *Inter alia*, we will also fill this gap.

Hence, in this paper, we try to make a few small (yet important for practitioners) contributions to the literature by evaluating whether the various methods, adapted from the partial variance correction method in Simar and Zelenyuk (2020), the data sharpening method in Nguyen et al. (2022) and the full variance correction method in Simar et al. (2023), are effective for the aggregate input-oriented and output-oriented efficiency.

Another important goal of this paper is to provide the practitioners with information about the performance of various improvement methods for the aggregate input-oriented and output-oriented efficiency in small samples, for different ranges of samples (10, 20, 50, 100, 200, 300, 500, 1000) and different dimensions of the models. Such information is very important for practitioners to understand the levels of accuracy one can expect (or hope) for the samples in hand and for the models they intend to use. For example, such information can suggest to a practitioner that for the sample they have (say about 300 observations) they need to use some adequate dimension reduction approaches to reduce the model dimension to no more than four, if they want to achieve the desired level of accuracy.[3] Such information can be obtained from Monte Carlo (MC) simulations, for various scenarios, and we perform them in this work.

Interestingly, performing such simulations nudged us to discover further theoretical improvements that can be made to the current theory found in Simar and Zelenyuk (2018, 2020). These improvements are based on the adaptation and extension of ideas of data sharpening for DEA from Nguyen et al. (2022) and an alternative estimator of the variance from Simar et al. (2023), and we improve upon these two works as well. Again, these improvements are modest from the perspective of high-level mathematics, yet they turn out to provide substantial improvements for practitioners in terms of application of the theory for relatively small samples and especially for relatively high dimensions of the DEA models, as confirmed by the evidence from the simulations.

More importantly, we provide a practical implementation algorithm and the free tools (the programmed and tested in many scenarios *R* codes) to simplify the life of practitioners interested in implementing this framework for the aggregate input-oriented and output-oriented efficiency. Finally, we use the well-known Philippines rice data set and Penn World Table data set as two examples to illustrate the differences between all these methods in the estimated variance as well as the estimated confidence intervals for the aggregate input-oriented and output-oriented (respectively) efficiency.

The rest of the paper is organized as follows. Section 2 briefly introduces the theoretical background on the technical efficiency and aggregate efficiency, and then briefly summarizes the main theoretical results for the aggregate input-oriented and output-oriented efficiency. Section 3 introduces the various methods to improve the finite sample performance of CLT results for the aggregate efficiency, by adapting the partial variance correction method in Simar and Zelenyuk (2020), a variant of the data sharpening method in Nguyen et al. (2022), and the full variance correction method in Simar et al. (2023). Section 4 conducts MC experiments to compare the performance of all these methods. In Section 5, we provide guidelines for practitioners and use two empirical examples to illustrate the differences between all these methods. Section 6 concludes and recommends future research directions. Additional results are provided in separate Appendices.

2. The theoretical background

2.1. The production economics

The economic agents or decision making units (DMUs) are assumed to use p number of inputs $x \in \mathbb{R}^p_+$ to produce q number of outputs

[3] Wilson (2018) has shown the usefulness of dimension reduction through simulations for the simple mean efficiency. We think that similar analysis can be conducted for the case of aggregate efficiency—it would be a subject in itself (to do it thoroughly), requiring a new paper and we thank one of the anonymous referees for raising this point.

$y \in \mathbb{R}^q_+$. Suppose the best-practice technology available to all DMUs is characterized by technology set Ψ, defined as

$$\Psi := \{(x, y) \mid x \text{ can produce } y\}. \tag{2.1}$$

That is, Ψ includes all the feasible combinations of inputs and outputs given the current technology. The standard regularity assumptions of production theory are assumed to hold.[4] The *frontier* or *technology* contained in Ψ is defined as

$$\Psi^{\partial} := \{(x, y) \mid (x, y) \in \Psi,\ (x/\omega, \omega y) \notin \Psi \text{ for any } \omega \in (1, \infty)\}. \tag{2.2}$$

Intuitively, the technology frontier represents the set of technically efficient input–output allocations.

The gap or the distance between an input–output allocation in Ψ and the frontier of Ψ is deemed as technical inefficiency, which can be measured in different ways, e.g., in different directions or with different types of measurements. Farrell-type measures of efficiency (input-oriented and output-oriented) appear to be the most popular in practice. The Farrell input-oriented efficiency measure is defined as

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

which calculates the proportion of inputs that can be scaled downward by the same scalar, holding outputs and technology fixed. Analogously, the Farrell output-oriented efficiency measure is defined as

$$\lambda(x, y \mid \Psi) := \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi\}, \tag{2.4}$$

which calculates the proportion of outputs that can be scaled upward by the same scalar, holding inputs and technology fixed. By construction, $0 \le \theta(x, y \mid \Psi) \le 1$ and $\lambda(x, y \mid \Psi) \ge 1$ for any $(x, y) \in \Psi$.

Other useful measures of efficiency are the cost and revenue efficiency (or the overall input and output efficiency). The cost efficiency for a DMU (x, y) facing input prices $w^x \in \mathbb{R}^p_+$ can be defined as

$$C(x, y \mid w^x, \Psi) := \frac{C_{min}}{(w^x)^T x}, \tag{2.5}$$

where C_{min} is the minimum cost of producing a specific output vector of y given input prices w^x and production set Ψ, defined as

$$C_{min} := \min_x \left\{(w^x)^T x \mid (x, y) \in \Psi\right\} = (w^x)^T x^*, \tag{2.6}$$

where x^* is the solution to the optimization problem in (2.6). Thus, cost efficiency can also be written as

$$C(x, y \mid w^x, \Psi) = \frac{(w^x)^T x^*}{(w^x)^T x}. \tag{2.7}$$

By construction, $0 \le C(x, y \mid w^x, \Psi) \le 1$. The input technical efficiency and the cost efficiency are well known to be closely related by Mahler's inequality, as

$$C(x, y \mid w^x, \Psi) \le \theta(x, y \mid \Psi). \tag{2.8}$$

This inequality can be used to define the input allocative efficiency (Färe et al., 1985), which is

$$\mathcal{A}_x(x, y \mid w^x, \Psi) = \frac{C(x, y \mid w^x, \Psi)}{\theta(x, y \mid \Psi)}. \tag{2.9}$$

By construction, $0 \le \mathcal{A}_x(x, y \mid w^x, \Psi) \le 1$. Consequently, the input technical, cost and input allocative efficiency are related to each other in the following way,

$$C(x, y \mid w^x, \Psi) = \theta(x, y \mid \Psi) \times \mathcal{A}_x(x, y \mid w^x, \Psi). \tag{2.10}$$

Alternatively, the revenue efficiency for a DMU (x, y) facing output prices $w^y \in \mathbb{R}^q_+$ can be defined as

$$\mathcal{R}(x, y \mid w^y, \Psi) := \frac{R_{max}}{(w^y)^T y}, \tag{2.11}$$

where R_{max} is the maximum revenue of using a specific input vector of x given output prices w^y and production set Ψ, defined as

$$R_{max} := \max_y \left\{(w^y)^T y \mid (x, y) \in \Psi\right\} = (w^y)^T y^*, \tag{2.12}$$

where y^* is the solution to the optimization problem in (2.12). Thus, revenue efficiency can also be written as

$$\mathcal{R}(x, y \mid w^y, \Psi) = \frac{(w^y)^T y^*}{(w^y)^T y}. \tag{2.13}$$

By construction, $\mathcal{R}(x, y \mid w^y, \Psi) \ge 1$. The output technical efficiency and the revenue efficiency are also closely related by Mahler's inequality, as

$$\mathcal{R}(x, y \mid w^y, \Psi) \ge \lambda(x, y \mid \Psi). \tag{2.14}$$

This inequality can be used to define the output allocative efficiency (Färe et al., 1985), which is

$$\mathcal{A}_y(x, y \mid w^y, \Psi) = \frac{\mathcal{R}(x, y \mid w^y, \Psi)}{\lambda(x, y \mid \Psi)}. \tag{2.15}$$

By construction, $\mathcal{A}_y(x, y \mid w^y, \Psi) \ge 1$. Consequently, the output technical, revenue and output allocative efficiency are related to each other in the following way,

$$\mathcal{R}(x, y \mid w^y, \Psi) = \lambda(x, y \mid \Psi) \times \mathcal{A}_y(x, y \mid w^y, \Psi). \tag{2.16}$$

However, all the above introduced measures, Ψ, Ψ^{∂}, $\theta(x, y \mid \Psi)$, $C(x, y \mid w^x, \Psi)$, $\mathcal{A}_x(x, y \mid w^x, \Psi)$, $\lambda(x, y \mid \Psi)$, $\mathcal{R}(x, y \mid w^y, \Psi)$, and $\mathcal{A}_y(x, y \mid w^y, \Psi)$ are not observed and hence must be estimated from the data on input quantities $X_i \in \mathbb{R}^p_+$, output quantities $Y_i \in \mathbb{R}^q_+$, input prices $w^x_i \in \mathbb{R}^p_+$, and output prices $w^y_i \in \mathbb{R}^q_+$, which we will denote as $\mathcal{S}_n = \{(X_i, Y_i)\}_{i=1}^n$ and $\mathcal{W}_n = \{(w^x_i, w^y_i)\}_{i=1}^n$.[5] After introducing the aggregate efficiency, we will come back later to discuss the estimators.

2.2. Aggregate efficiency

Researchers and policy makers are often interested in the efficiency of a group of producers (such as the efficiency of an industry or a group within it), usually called *aggregate efficiency*. This is especially the case for studies with many individuals, where reporting all the estimates may overwhelm the reader and, therefore, some aggregate measures are needed to adequately summarize the results. How to adequately aggregate the individual efficiency into a group efficiency is a question by itself that has been explored in the literature quite substantially, in the past seven decades.[6]

Here we will follow the approach based on using Koopmans aggregation theorem, proposed by Färe and Zelenyuk (2003) in the output-oriented context and extended it further in Färe et al. (2004), and elaborated more recently in Mayer and Zelenyuk (2019). Specifically, their aggregate input-oriented technical efficiency for a group of n producers is given by

$$\bar{\theta} := \sum_{i=1}^n \theta(X_i, Y_i \mid \Psi) S^x_i, \tag{2.17}$$

where S^x_i is the ith observation's cost weight, defined as

$$S^x_i = \frac{(w^x_i)^T X_i}{\sum_{i=1}^n (w^x_i)^T X_i}. \tag{2.18}$$

The aggregate cost efficiency is given by

$$\bar{C} := \sum_{i=1}^n C(X_i, Y_i \mid w^x_i, \Psi) S^x_i, \tag{2.19}$$

[4] See Appendix A for more details.

[5] The economic theory based aggregation framework (Färe and Zelenyuk (2003), etc.) to arrive to (2.19) and (2.25) requires that all individuals ($i = 1, ..., n$) are price takers that face the same prices in their optimization behavior. The statistical theory developed below allows prices to vary across i, as is reflected in our notation.

[6] For a recent review, see Zelenyuk (2020).

while the aggregate input allocative efficiency is given by

$$\bar{\mathcal{A}}_x := \sum_{i=1}^{n} \mathcal{A}_x(X_i, Y_i \mid w_i^x, \Psi) S_i^{\partial x}, \tag{2.20}$$

where

$$S_i^{\partial x} = \frac{(w_i^x)^T X_i^\partial}{\sum_{i=1}^n (w_i^x)^T X_i^\partial}, \tag{2.21}$$

and where $X_i^\partial = \theta(X_i, Y_i \mid \Psi) X_i$ is the hypothetical efficient input vector (given output vector Y_i) for the ith producer by projecting the ith producer to the frontier Ψ^∂ through the input orientation. Similar to the relationship between the individual efficiency defined in (2.10), for aggregate input-oriented technical, cost and allocative efficiency, we also have

$$\bar{C} = \bar{\theta} \times \bar{\mathcal{A}}_x. \tag{2.22}$$

Similarly, the aggregate output-oriented technical efficiency for a group of n producers is given by

$$\bar{\lambda} := \sum_{i=1}^{n} \lambda(X_i, Y_i \mid \Psi) S_i^y, \tag{2.23}$$

where S_i^y is the ith observation's revenue weight, defined as

$$S_i^y = \frac{(w_i^y)^T Y_i}{\sum_{i=1}^n (w_i^y)^T Y_i}. \tag{2.24}$$

The aggregate revenue efficiency is given by

$$\bar{\mathcal{R}} := \sum_{i=1}^{n} \mathcal{R}(X_i, Y_i \mid w_i^y, \Psi) S_i^y, \tag{2.25}$$

while the aggregate output allocative efficiency is given by

$$\bar{\mathcal{A}}_y := \sum_{i=1}^{n} \mathcal{A}_y(X_i, Y_i \mid w_i^y, \Psi) S_i^{\partial y}, \tag{2.26}$$

where

$$S_i^{\partial y} = \frac{(w_i^y)^T Y_i^\partial}{\sum_{i=1}^n (w_i^y)^T Y_i^\partial}, \tag{2.27}$$

and where $Y_i^\partial = \lambda(X_i, Y_i \mid \Psi) Y_i$ is the hypothetical efficient output vector (given input vector X_i) for the ith producer by projecting the ith producer to the frontier Ψ^∂ through the output orientation. Similar to the relationship between the individual efficiency defined in (2.16), for aggregate output-oriented technical, cost and allocative efficiency, we also have

$$\bar{\mathcal{R}} = \bar{\lambda} \times \bar{\mathcal{A}}_y. \tag{2.28}$$

The above mentioned aggregate efficiency measures can also be expressed as ratios of simple means (Simar & Zelenyuk, 2018). Specifically,

$$\bar{\theta} = \sum_{i=1}^{n} \theta(X_i, Y_i \mid \Psi) S_i^x = \frac{(1/n)\sum_{i=1}^n (w_i^x)^T X_i^\partial}{(1/n)\sum_{i=1}^n (w_i^x)^T X_i}, \tag{2.29}$$

$$\bar{C} = \sum_{i=1}^{n} C(X_i, Y_i \mid w_i^x, \Psi) S_i^x = \frac{(1/n)\sum_{i=1}^n (w_i^x)^T X_i^*}{(1/n)\sum_{i=1}^n (w_i^x)^T X_i}, \tag{2.30}$$

$$\bar{\mathcal{A}}_x = \sum_{i=1}^{n} \mathcal{A}_x(X_i, Y_i \mid w_i^x, \Psi) S_i^{\partial x} = \frac{(1/n)\sum_{i=1}^n (w_i^x)^T X_i^*}{(1/n)\sum_{i=1}^n (w_i^x)^T X_i^\partial}, \tag{2.31}$$

$$\bar{\lambda} = \sum_{i=1}^{n} \lambda(X_i, Y_i \mid \Psi) S_i^y = \frac{(1/n)\sum_{i=1}^n (w_i^y)^T Y_i^\partial}{(1/n)\sum_{i=1}^n (w_i^y)^T Y_i}, \tag{2.32}$$

$$\bar{\mathcal{R}} = \sum_{i=1}^{n} \mathcal{R}(X_i, Y_i \mid w_i^y, \Psi) S_i^y = \frac{(1/n)\sum_{i=1}^n (w_i^y)^T Y_i^*}{(1/n)\sum_{i=1}^n (w_i^y)^T Y_i}, \tag{2.33}$$

and

$$\bar{\mathcal{A}}_y = \sum_{i=1}^{n} \mathcal{A}_y(X_i, Y_i \mid w_i^y, \Psi) S_i^{\partial y} = \frac{(1/n)\sum_{i=1}^n (w_i^y)^T Y_i^*}{(1/n)\sum_{i=1}^n (w_i^y)^T Y_i^\partial}, \tag{2.34}$$

where $(w_i^x)^T X_i^*$ is the minimum cost in (2.6) for a DMU with an allocation (X_i, Y_i) facing input prices w_i^x and X_i^* is the argmin of the cost function; $(w_i^y)^T Y_i^*$ is the maximum revenue in (2.12) for a DMU with an allocation (X_i, Y_i) facing output prices w_i^y and Y_i^* is the argmax of the revenue function. These expressions are useful as they can be used to develop the theoretical results for the aggregate efficiency by utilizing the statistical theory of the ratio of sample means. For the sake of conciseness, we will focus only on the aggregate input-oriented and output-oriented technical efficiency.

2.3. The estimators of individual efficiency and their asymptotic properties

Given a random sample $\mathcal{S}_n = \{(X_i, Y_i)\}_{i=1}^n$, if one assumes the production technology set Ψ is convex and exhibits CRS, one could use the CRS-DEA estimator (Charnes et al., 1978), given by

$$\widehat{\theta}_{\text{CRS}}(x, y \mid \mathcal{S}_n) = \min_{\theta, s_1, \ldots, s_n} \left\{ \theta \mid y \leqq \sum_{i=1}^{n} s_i Y_i,\ \theta x \geqq \sum_{i=1}^{n} s_i X_i,\ \forall\ s_i \geq 0 \right\}, \tag{2.35}$$

and

$$\widehat{\lambda}_{\text{CRS}}(x, y \mid \mathcal{S}_n) = \max_{\lambda, s_1, \ldots, s_n} \left\{ \lambda \mid \lambda y \leqq \sum_{i=1}^{n} s_i Y_i,\ x \geqq \sum_{i=1}^{n} s_i X_i,\ \forall\ s_i \geq 0 \right\}. \tag{2.36}$$

Alternatively, if one assumes Ψ is convex but exhibits VRS, one could use the VRS-DEA estimator (Banker et al., 1984) given by

$$\widehat{\theta}_{\text{VRS}}(x, y \mid \mathcal{S}_n) = \min_{\theta, s_1, \ldots, s_n} \left\{ \theta \mid y \leqq \sum_{i=1}^{n} s_i Y_i,\ \theta x \geqq \sum_{i=1}^{n} s_i X_i,\ \sum_{i=1}^{n} s_i = 1,\ \forall\ s_i \geq 0 \right\}, \tag{2.37}$$

and

$$\widehat{\lambda}_{\text{VRS}}(x, y \mid \mathcal{S}_n) = \max_{\lambda, s_1, \ldots, s_n} \left\{ \lambda \mid \lambda y \leqq \sum_{i=1}^{n} s_i Y_i,\ x \geqq \sum_{i=1}^{n} s_i X_i,\ \sum_{i=1}^{n} s_i = 1,\ \forall\ s_i \geq 0 \right\}. \tag{2.38}$$

Alternatively, if one allows non-convexity and only assumes strong disposability of all inputs and all outputs, the FDH estimator proposed by Deprins et al. (1984) could be used, which is given by VRS-DEA with additional constraint $s_i \in \{0, 1\}$ added, or by computing

$$\widehat{\theta}_{\text{FDH}}(x, y \mid \mathcal{S}_n) = \min_{i \in \mathcal{D}_{x,y}} \max_{j=1,\ldots,p} \left(\frac{X_{ij}}{x_j} \right), \tag{2.39}$$

and

$$\widehat{\lambda}_{\text{FDH}}(x, y \mid \mathcal{S}_n) = \max_{i \in \mathcal{D}_{x,y}} \min_{k=1,\ldots,q} \left(\frac{Y_{ik}}{y_k} \right), \tag{2.40}$$

where for a vector a, a_j denotes its jth component and $\mathcal{D}_{x,y} = \{i \mid (X_i, Y_i) \in \mathcal{S}_n, X_i \leqq x, Y_i \geqq y\}$.

The FDH, VRS, and CRS estimators of $\theta(x, y \mid \Psi)$ and $\lambda(x, y \mid \Psi)$ have well-developed statistical properties.[7] To summarize, all the estimators are consistent with a convergence rate n^κ where $\kappa = 2/(p+q)$, $2/(p+q+1)$, and $1/(p+q)$ for the CRS, VRS, and FDH estimators, respectively. Under an appropriate set of assumptions, all these estimators also have non-degenerate limiting distributions. Moreover, as recently pointed out by Kneip et al. (2015), all the estimators are biased, where the order of the bias is $O(n^{-\kappa})$. If $\kappa \leq 1/2$, the bias term of their average does not vanish quickly enough (relative to the variance) to zero, and the

[7] See Kneip et al. (1998), Park et al. (2000), Kneip et al. (2008), Park et al. (2010), Kneip et al. (2015), and a recent review by Simar and Wilson (2015).

usual CLT does not apply here. Kneip et al. (2015) also established new CLTs for the simple mean efficiency. The bias term for the simple mean (as well as individual efficiency) can be estimated using a generalized jackknife-type approach proposed by Kneip et al. (2015).

2.4. The estimators of aggregate efficiency and their asymptotics

We will focus on the variant form of the aggregate input-oriented efficiency in (2.29) and the aggregate output-oriented efficiency in (2.32), denoted as φ_n and τ_n, respectively, namely

$$\varphi_n = \frac{n^{-1}\sum_{i=1}^n (w_i^x)^T X_i^\partial}{n^{-1}\sum_{i=1}^n (w_i^x)^T X_i}, \tag{2.41}$$

and

$$\tau_n = \frac{n^{-1}\sum_{i=1}^n (w_i^y)^T Y_i^\partial}{n^{-1}\sum_{i=1}^n (w_i^y)^T Y_i}. \tag{2.42}$$

To simplify the notation, denote $Z_i = (w_i^x)^T X_i$, $Z_i^\partial = (w_i^x)^T X_i^\partial = (w_i^x)^T \theta(X_i, Y_i \mid \Psi) X_i$, $U_i = (w_i^y)^T Y_i$, and $U_i^\partial = (w_i^y)^T \lambda(X_i, Y_i \mid \Psi) Y_i$. Note that φ_n and τ_n are the estimates of $\varphi = \mu_1/\mu_2$ and $\tau = \mu_3/\mu_4$ (respectively), where $\mu_1 = E(Z_i^\partial)$, $\mu_2 = E(Z_i)$, $\mu_3 = E(U_i^\partial)$ and $\mu_4 = E(U_i)$.

However, $\theta(X_i, Y_i \mid \Psi)$ and $\lambda(X_i, Y_i \mid \Psi)$ are not observed, making φ_n and τ_n unobserved, and the best we can hope for are estimates of $\theta(X_i, Y_i \mid \Psi)$ and $\lambda(X_i, Y_i \mid \Psi)$ from a random sample of data $\mathcal{S}_n = \{(X_i, Y_i)\}_{i=1}^n$, e.g., using FDH, or VRS-DEA or CRS-DEA estimators as discussed above in Section 2.3, to obtain $\widehat{\theta}(X_i, Y_i \mid \mathcal{S}_n)$ and $\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n)$. Then estimates of φ_n and τ_n can be obtained by plugging $\widehat{\theta}(X_i, Y_i \mid \mathcal{S}_n)$ into (2.41) and $\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n)$ into (2.42) (respectively), i.e., via

$$\widehat{\varphi}_n = \frac{\widehat{\mu}_{1,n}}{\widehat{\mu}_{2,n}} = \frac{n^{-1}\sum_{i=1}^n \widehat{Z}_i^\partial}{n^{-1}\sum_{i=1}^n Z_i}, \tag{2.43}$$

and

$$\widehat{\tau}_n = \frac{\widehat{\nu}_{1,n}}{\widehat{\nu}_{2,n}} = \frac{n^{-1}\sum_{i=1}^n \widehat{U}_i^\partial}{n^{-1}\sum_{i=1}^n U_i}, \tag{2.44}$$

where $\widehat{Z}_i^\partial = (w_i^x)^T \widehat{\theta}(X_i, Y_i \mid \mathcal{S}_n) X_i$ and $\widehat{U}_i^\partial = (w_i^y)^T \widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) Y_i$.

The asymptotic properties of the aggregate output-oriented efficiency $\widehat{\tau}_n$ are established by Simar and Zelenyuk (2018) while those for aggregate input-oriented efficiency $\widehat{\varphi}_n$ are established in Appendix A of this paper. The main results are summarized in the following theorem.

Theorem 2.1. *Under the appropriate set of assumptions, for $\kappa \geq 2/5$ for CRS-DEA or VRS-DEA cases, or for $\kappa \geq 1/3$ for FDH case, as $n \to \infty$, we have*

$$\sqrt{n}\left(\widehat{\varphi}_n - \widehat{B}_{\varphi,n} - \varphi + o(n^{-\kappa})\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\varphi^2), \tag{2.45}$$

and

$$\sqrt{n}\left(\widehat{\tau}_n - \widehat{B}_{\tau,n} - \tau + o(n^{-\kappa})\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\tau^2), \tag{2.46}$$

and if $\kappa < 1/2$, we have

$$\sqrt{n_\kappa}\left(\widehat{\varphi}_{n_\kappa} - \widehat{B}_{\varphi,n} - \varphi + o(n^{-\kappa})\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\varphi^2), \tag{2.47}$$

and

$$\sqrt{n_\kappa}\left(\widehat{\tau}_{n_\kappa} - \widehat{B}_{\tau,n} - \tau + o(n^{-\kappa})\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\tau^2), \tag{2.48}$$

where $\widehat{B}_{\varphi,n}$ and $\widehat{B}_{\tau,n}$ are the estimated bias using a generalized jackknife-type approach proposed by Kneip et al. (2015); σ_φ^2 and σ_τ^2 are the true variance of aggregate input-oriented and output-oriented efficiency; $\widehat{\varphi}_{n_\kappa}$ and $\widehat{\tau}_{n_\kappa}$ are random subsample versions, with size $n_\kappa = \lfloor n^{2\kappa} \rfloor < n$, of $\widehat{\varphi}_n$ and $\widehat{\tau}_n$, respectively.[8] *Formally,*

$$\widehat{\varphi}_{n_\kappa} = \frac{n_\kappa^{-1}\sum_{\{i|(X_i,Y_i)\in\mathcal{S}_{n_\kappa}\}} \widehat{\theta}(X_i, Y_i|\mathcal{S}_n) Z_i}{n_\kappa^{-1}\sum_{\{i|(X_i,Y_i)\in\mathcal{S}_{n_\kappa}\}} Z_i}, \tag{2.49}$$

and

$$\widehat{\tau}_{n_\kappa} = \frac{n_\kappa^{-1}\sum_{\{i|(X_i,Y_i)\in\mathcal{S}_{n_\kappa}\}} \widehat{\lambda}(X_i, Y_i|\mathcal{S}_n) U_i}{n_\kappa^{-1}\sum_{\{i|(X_i,Y_i)\in\mathcal{S}_{n_\kappa}\}} U_i}, \tag{2.50}$$

where $\mathcal{S}_{n_\kappa}$ is a random subsample, with the sample size n_κ, of $\mathcal{S}_n$.[9]

Further, σ_φ^2 and σ_τ^2 are not observed and must be estimated. To estimate σ_τ^2, Simar and Zelenyuk (2018) suggest plugging the corresponding empirical estimates of these components in σ_τ^2, i.e.,

$$\widehat{\sigma}_{\tau,n}^2 = \widehat{\tau}_n^2 \left[\frac{\widehat{\sigma}_{3,n}^2}{\widehat{\mu}_{3,n}^2} + \frac{\widehat{\sigma}_{4,n}^2}{\widehat{\mu}_{4,n}^2} - 2\frac{\widehat{\sigma}_{34,n}}{\widehat{\mu}_{3,n}\widehat{\mu}_{4,n}}\right], \tag{2.51}$$

where $\widehat{\sigma}_{3,n}^2 = \widehat{\mathrm{Var}}(\widehat{U}_i^\partial)$, $\widehat{\sigma}_{4,n}^2 = \widehat{\mathrm{Var}}(U_i)$, $\widehat{\sigma}_{34,n} = \widehat{\mathrm{Cov}}(\widehat{U}_i^\partial, U_i)$, $\widehat{\mu}_{3,n} = \widehat{E}(\widehat{U}_i^\partial)$, and $\widehat{\mu}_{4,n} = \widehat{E}(U_i)$. Similarly, for the input-oriented case, the estimate for σ_φ^2 can be obtained from

$$\widehat{\sigma}_{\varphi,n}^2 = \widehat{\varphi}_n^2 \left[\frac{\widehat{\sigma}_{1,n}^2}{\widehat{\mu}_{1,n}^2} + \frac{\widehat{\sigma}_{2,n}^2}{\widehat{\mu}_{2,n}^2} - 2\frac{\widehat{\sigma}_{12,n}}{\widehat{\mu}_{1,n}\widehat{\mu}_{2,n}}\right], \tag{2.52}$$

where $\widehat{\sigma}_{1,n}^2 = \widehat{\mathrm{Var}}(\widehat{Z}_i^\partial)$, $\widehat{\sigma}_{2,n}^2 = \widehat{\mathrm{Var}}(Z_i)$, $\widehat{\sigma}_{12,n} = \widehat{\mathrm{Cov}}(\widehat{Z}_i^\partial, Z_i)$, $\widehat{\mu}_{1,n} = \widehat{E}(\widehat{Z}_i^\partial)$, and $\widehat{\mu}_{2,n} = \widehat{E}(Z_i)$.

3. Improving the approximation

For the simple mean efficiency (Kneip et al., 2015) and aggregate efficiency (Simar & Zelenyuk, 2018), their MC experiments found that when the sample sizes increase, the coverages of the estimated confidence intervals based on the corresponding CLT results approximate the nominal coverage, which supports the established theoretical results. However, these authors also found that the estimated confidence intervals often under-cover the true values, especially for small sample sizes and high dimensions measured by the number of inputs and outputs specified. This under-covering phenomenon in relatively small sample sizes was also observed for the aggregate input-oriented efficiency with MC results presented in Section 4. Some of the improvements were made through Simar and Zelenyuk (2020), Nguyen et al. (2022), and Simar et al. (2023) for the simple mean output-oriented efficiency, and through Simar and Zelenyuk (2020) for the aggregate output-oriented efficiency. In this section, we will adapt these methods to further improve the finite sample performance of the CLT results for the aggregate efficiency with a focus on the input orientation. The following results can also be applied to the output orientation.

3.1. Adapting the improvements from Simar and Zelenyuk (2020)

For the aggregate input-oriented efficiency, following Simar and Zelenyuk (2020), we use a corrected version for the variance estimate,

$$\ddot{\sigma}_{\varphi,n}^2 = \widehat{\varphi}_n^2 \left[\frac{\ddot{\sigma}_{1,n}^2}{\widehat{\mu}_{1,n}^2} + \frac{\widehat{\sigma}_{2,n}^2}{\widehat{\mu}_{2,n}^2} - 2\frac{\widehat{\sigma}_{12,n}}{\widehat{\mu}_{1,n}\widehat{\mu}_{2,n}}\right], \tag{3.1}$$

where $\ddot{\sigma}_{1,n}^2 = \widehat{\sigma}_{1,n}^2 + \widehat{B}_{\mu_1,n}^2$ and $\widehat{B}_{\mu_1,n}$ is the estimated bias for $\widehat{\mu}_{1,n}$. That is, we add the estimated squared bias of mean efficiency estimate for $\mu_{1,n}$ into its variance estimator $\widehat{\sigma}_{1,n}^2$, thus making the variance estimate slightly larger than for the case where no squared bias was added, i.e., the original method in (2.52).

[8] $\lfloor n^{2\kappa} \rfloor$ denotes the largest integer that is less or equal to $n^{2\kappa}$.

[9] Note that $\widehat{\theta}(X_i, Y_i|\mathcal{S}_n)$ and $\widehat{\lambda}(X_i, Y_i|\mathcal{S}_n)$ are computed relative to all the data points in $\mathcal{S}_n$ rather than $\mathcal{S}_{n_\kappa}$.

3.2. Adapting the improvements from Nguyen et al. (2022)

Although the partial variance correction method in Simar and Zelenyuk (2020) increases the performance of the CLT results for finite sample sizes, it is also observed from the simulations that the coverages of the Simar and Zelenyuk (2020) method are often still smaller than the nominal coverages. Recently, Nguyen et al. (2022) proposed the data sharpening method to further improve finite sample approximation of CLT results for the simple mean output-oriented technical efficiency. This data sharpening method is inspired by Simar and Zelenyuk (2006) and Kneip et al. (2011). The two latter papers used the sharpening to derive simple consistent bootstrap approximations of distribution of efficiency scores.

Specifically, we adapt the idea of the data sharpening method in Nguyen et al. (2022) to the input-oriented technical efficiency as follows,

$$\widehat{\widehat{\theta}}(X_i,Y_i \mid \mathcal{S}_n) = \begin{cases} \widehat{\theta}(X_i,Y_i \mid \mathcal{S}_n), & \text{if } \widehat{\theta}(X_i,Y_i \mid \mathcal{S}_n) < 1-\delta, \\ \widehat{\theta}(X_i,Y_i \mid \mathcal{S}_n)\times \varepsilon_i, & \text{otherwise}, \end{cases} \tag{3.2}$$

where the sharpening parameter δ needs to be in $(0,1)$ but also small enough, such that it preserves the properties of the CLTs and provides robust and good performances in the MC experiments. Later, we will derive the theoretical range for δ. Further, ε_i is a random independent number drawn from a uniform distribution on the interval $[1-\delta, 1]$. It can be shown that $\widehat{\widehat{\theta}}(X_i,Y_i \mid \mathcal{S}_n) = \widehat{\theta}(\widetilde{X}_i,Y_i \mid \mathcal{S}_n)$, where

$$\widetilde{X}_i = \begin{cases} X_i, & \text{if } \widehat{\theta}(X_i,Y_i \mid \mathcal{S}_n) < 1-\delta, \\ X_i/\varepsilon_i, & \text{otherwise}. \end{cases} \tag{3.3}$$

Then the regular nonparametric efficiency estimators are applied for the sharpened sample points $\{(\widetilde{X}_i,Y_i)\}_{i=1}^n$, but with the reference set being the original sample $\mathcal{S}_n = \{(X_i,Y_i)\}_{i=1}^n$. It is worth noting that the data sharpening method will have an effect on the estimates of efficiency and on its bias and variance, yet it does not have an effect on other quantities, such as $\widehat{\mu}_2$ and $\widehat{\sigma}^2_{2,n}$.

Now, we will formally derive the theoretical range for δ that preserves the properties of the CLTs. What we need is to keep δ small enough so that the CLTs expressed in (A.9) and (A.10) are still valid for the sharpened estimates. For notational simplicity, in what follows, we denote $\widehat{\theta}_i = \widehat{\theta}(X_i,Y_i|\mathcal{S}_n)$ and $\widehat{\widehat{\theta}}_i = \widehat{\widehat{\theta}}(X_i,Y_i|\mathcal{S}_n)$. Our data sharpening method in (3.2) then can be expressed as,

$$\widehat{\widehat{\theta}}_i = \begin{cases} \widehat{\theta}_i & \text{if } \widehat{\theta}_i < 1-\delta, \\ \widehat{\theta}_i\varepsilon_i & \text{if } 1-\delta \le \widehat{\theta}_i \le 1, \end{cases} \tag{3.4}$$

where $\varepsilon_i = 1-U_i$ and $U_i \sim \text{Unif}(0,\delta)$ is independent of $\widehat{\theta}_i$, so that $\text{Prob}(\widehat{\theta}_i - \widehat{\widehat{\theta}}_i \ge 0) = 1$. The consequence of this is that

$$\widehat{\varphi}_n = \frac{n^{-1}\sum_{i=1}^n w_i^T \widehat{\theta}_i X_i}{n^{-1}\sum_{i=1}^n w_i^T X_i}, \tag{3.5}$$

and

$$\widehat{\varphi}_{n_\kappa} = \frac{n_\kappa^{-1}\sum_{\{i|(X_i,Y_i)\in\mathcal{S}_{n_\kappa}\}} w_i^T \widehat{\theta}_i X_i}{n_\kappa^{-1}\sum_{\{i|(X_i,Y_i)\in\mathcal{S}_{n_\kappa}\}} w_i^T X_i}, \tag{3.6}$$

in (A.9) and (A.10) will be respectively replaced by

$$\widehat{\widehat{\varphi}}_n = \frac{n^{-1}\sum_{i=1}^n w_i^T \widehat{\widehat{\theta}}_i X_i}{n^{-1}\sum_{i=1}^n w_i^T X_i}, \tag{3.7}$$

and

$$\widehat{\widehat{\varphi}}_{n_\kappa} = \frac{n_\kappa^{-1}\sum_{\{i|(X_i,Y_i)\in\mathcal{S}_{n_\kappa}\}} w_i^T \widehat{\widehat{\theta}}_i X_i}{n_\kappa^{-1}\sum_{\{i|(X_i,Y_i)\in\mathcal{S}_{n_\kappa}\}} w_i^T X_i}. \tag{3.8}$$

So, to keep the CLTs in (A.9) and (A.10), we need in all the cases that $(\widehat{\theta}_i - \widehat{\widehat{\theta}}_i) = o_p\left(n^{-\tilde{\kappa}}\right)$ where $\tilde{\kappa} = \min(\kappa, 1/2)$.

Now by Markov Inequality, we have that for all $a > 0$,

$$\text{Prob}\left(\widehat{\theta}_i - \widehat{\widehat{\theta}}_i \ge a\right) \le \frac{\text{E}(\widehat{\theta}_i - \widehat{\widehat{\theta}}_i)}{a}. \tag{3.9}$$

Since $\widehat{\theta}_i - \widehat{\widehat{\theta}}_i = \widehat{\theta}_i U_i \mathbf{1}(\widehat{\theta}_i \ge 1-\delta)$ where $\mathbf{1}(.)$ is the indicator function and given that U_i is independent of $\widehat{\widehat{\theta}}_i$, we have

$$\text{E}\left(\widehat{\theta}_i U_i \mathbf{1}(\widehat{\theta}_i \ge 1-\delta)\right) = \text{E}\left(\widehat{\theta}_i \mathbf{1}(\widehat{\theta}_i \ge 1-\delta)\right)\text{E}(U_i) \tag{3.10}$$

$$= \frac{\delta}{2}\int_{1-\delta}^1 \hat{\theta} f(\hat{\theta}) d\hat{\theta} \tag{3.11}$$

$$= \frac{\delta}{2}\delta\bar{\theta} f(\bar{\theta}), \tag{3.12}$$

where $f(\hat{\theta})$ is the density of $\widehat{\theta}_i$ and $\bar{\theta}$ is, by the Mean Value Theorem, some number in the interval $(1-\delta, 1)$. So we end up with

$$\text{E}(\widehat{\theta}_i - \widehat{\widehat{\theta}}_i) = c_\theta \delta^2, \tag{3.13}$$

for some finite $c_\theta > 0$. So by the Markov inequality in (3.9), we have for all $a > 0$,

$$\text{Prob}\left(n^{\tilde{\kappa}}(\widehat{\theta}_i - \widehat{\widehat{\theta}}_i) \ge a\right) \le \frac{c_\theta \delta^2}{a n^{-\tilde{\kappa}}}. \tag{3.14}$$

It is clear that if $\delta = n^{-\gamma}$ for $\gamma > \tilde{\kappa}/2$ this probability converges to zero as $n \to \infty$, so that $(\widehat{\theta}_i - \widehat{\widehat{\theta}}_i) = o_p\left(n^{-\tilde{\kappa}}\right)$, as required.[10] Therefore, to keep the CLTs properties for aggregate input-oriented efficiency, we only need $\delta = n^{-\gamma}$ and $\gamma > \tilde{\kappa}/2$.

3.3. Adapting the improvements from Simar et al. (2023)

More recently, Simar et al. (2023) proposed a full variance correction method by using the bias-corrected individual efficiency estimates instead of the original individual efficiency estimates, to obtain the variance estimator for the simple mean efficiency. We adapt the full variance correction method in Simar et al. (2023) to the case of the aggregate efficiency. Specifically, for the original variance estimator of aggregate efficiency in (2.52), we propose replacing $\widehat{\theta}(X_i,Y_i|\mathcal{S}_n)$ by $\widetilde{\theta}(X_i,Y_i|\mathcal{S}_n) = \widehat{\theta}(X_i,Y_i|\mathcal{S}_n) - \widehat{B}_i$ at every place (the relevant quantities are $\widehat{\varphi}_n$, $\widehat{\sigma}^2_{1,n}$, $\widehat{\mu}_{1,n}$ and $\widehat{\sigma}_{12,n}$), where $\widehat{B}_i$ is the estimated individual bias using the generalized jackknife method of Kneip et al. (2015). That is, we estimate σ^2_φ as follows:

$$\widetilde{\sigma}^2_{\varphi,n} = \widetilde{\varphi}^2_n \left[\frac{\widetilde{\sigma}^2_{1,n}}{\widetilde{\mu}^2_1} + \frac{\widehat{\sigma}^2_{2,n}}{\widehat{\mu}^2_{2,n}} - 2\frac{\widetilde{\sigma}_{12,n}}{\widetilde{\mu}_1\widehat{\mu}_{2,n}}\right], \tag{3.15}$$

where

$$\widetilde{\varphi}_n = \widehat{\varphi}_n - \widehat{B}_{\varphi,n}, \tag{3.16}$$

$$\widetilde{\mu}_1 = \widehat{\mu}_{1,n} - \widehat{B}_{\mu_1,n}, \tag{3.17}$$

$$\widetilde{\sigma}^2_{1,n} = \widehat{\text{Var}}(\widetilde{\theta}(X_i,Y_i|\mathcal{S}_n)Z_i), \tag{3.18}$$

and

$$\widetilde{\sigma}_{12,n} = \widehat{\text{Cov}}(\widetilde{\theta}(X_i,Y_i|\mathcal{S}_n)Z_i, Z_i). \tag{3.19}$$

Note that, comparing (3.15) with (2.52), we not only use the bias-corrected individual efficiency to obtain estimates for σ^2_1 but also for φ, μ_1 and σ_{12}. The full variance correction here is different from the simple mean case in Simar et al. (2023), as the variance of aggregate efficiency is more complicated, involving several moments and the covariance term in a non-linear relationship.

[10] Nguyen et al. (2022) use the same property for deriving a lower bound to γ, but without a formal proof using only an intuitive (and correct) argument. Moreover, the upper bound for γ is generally not needed for our purposes of data sharpening in the context of approximation of the CLTs.

Compared with (2.52) and (3.1), (3.15) has the potential to improve, since (3.15) replaces the $\widehat{\theta}(X_i, Y_i|S_n)$ in (2.52) by its bias-corrected estimate $\widetilde{\theta}(X_i, Y_i|S_n)$ at every relevant place. To see this, we may view σ^2_φ as a function of only μ_1, σ^2_1 and σ_{12} since the replacement only has effects on these three terms in the variance estimator. Hence, the Taylor expansion of $\widehat{\sigma}^2_{\varphi,n}$ in (2.52) around σ^2_φ is given by

$$\widehat{\sigma}^2_{\varphi,n} = \sigma^2_\varphi + A_1(\widehat{\mu}_1 - \mu_1) + A_2(\widehat{\sigma}^2_1 - \sigma^2_1) + A_3(\widehat{\sigma}_{12} - \sigma_{12}) + R_n, \tag{3.20}$$

while the Taylor expansion of the proposed estimator $\widetilde{\sigma}^2_{\varphi,n}$ in (3.15) around σ^2_φ is given by

$$\widetilde{\sigma}^2_{\varphi,n} = \sigma^2_\varphi + A_1(\widetilde{\mu}_1 - \mu_1) + A_2(\widetilde{\sigma}^2_{1,n} - \sigma^2_1) + A_3(\widetilde{\sigma}_{12,n} - \sigma_{12}) + R_n, \tag{3.21}$$

where $A_1 = \frac{2\mu_1\sigma^2_2}{\mu^4_2} - \frac{2\sigma_{12}}{\mu^3_2}$, $A_2 = \frac{1}{\mu^2_2}$, and $A_3 = -\frac{2\mu_1}{\mu^3_2}$, while R_n is the remainder of a smaller order, i.e., $o(n^{-\kappa})$, and can be ignored for our purposes here.

According to Lemma 1 in Simar and Zelenyuk (2018), we have $\widehat{\sigma}^2_1 - \sigma^2_1 = o(n^{-\kappa/2})$, $\widehat{\sigma}_{12} - \sigma_{12} = o(n^{-\kappa/2})$, $\widehat{\mu}_1 - \mu_1 = O(n^{-\kappa})$. Thus, the first order error in (3.20) is given by

$$A_2(\widehat{\sigma}^2_1 - \sigma^2_1) + A_3(\widehat{\sigma}_{12} - \sigma_{12}) = o(n^{-\kappa/2}), \tag{3.22}$$

while the second order error is given by

$$A_1(\widehat{\mu}_1 - \mu_1) = O(n^{-\kappa}). \tag{3.23}$$

Similarly, for (3.21), we also have $\widetilde{\sigma}^2_1 - \sigma^2_1 = o(n^{-\kappa/2})$ and $\widetilde{\sigma}_{12} - \sigma_{12} = o(n^{-\kappa/2})$, but $\widetilde{\mu}_1 - \mu_1 = o(n^{-\kappa})$. Thus, the first order error in (3.21) is given by

$$A_2(\widetilde{\sigma}^2_{1,n} - \sigma^2_1) + A_3(\widetilde{\sigma}_{12,n} - \sigma_{12}) = o(n^{-\kappa/2}), \tag{3.24}$$

while the second order error is given by

$$A_1(\widehat{\mu}_1 - \mu_1) = o(n^{-\kappa}). \tag{3.25}$$

That is, while using the bias-corrected individual efficiency estimate does not reduce the first order error (it is still of the order $o(n^{-\kappa/2})$), it turns out we are able to reduce the second order term from $O(n^{-\kappa})$ to $o(n^{-\kappa})$, which is an improvement.[11]

How important is this reduction in the theoretical order from $O(n^{-\kappa})$ to $o(n^{-\kappa})$ in practice? By and large, it depends on the actual sample: the actual data generating process it came from, its size and, as usual, the luck of the particular draw. The only way to sense it is to investigate it in various scenarios, as we do in the next section. We will see that, although (3.20) and (3.21) are asymptotically equivalent, the proposed method can be a substantial improvement for relatively small samples. Indeed, it is possible that for a particular sample the second order error $O(n^{-\kappa})$ is substantially larger than the first order error $o(n^{-\kappa/2})$, i.e., similarly as was pointed out and confirmed in simulations for a simpler case of the simple mean efficiency by Simar et al. (2023), and as we can see it in our simulations for the aggregate efficiency reported below.

4. Monte-Carlo evidence

4.1. Details on Monte-Carlo simulations

For the sake of comparability, our Monte-Carlo experiments follow a similar strategy as in Simar and Zelenyuk (2018, 2020) and Simar et al. (2023). We also try a few scenarios in the multiple-outputs setting for DGP from Nguyen et al. (2022) and Kneip et al. (2015), and obtain similar conclusions.[12] Formally,

$$y_i = \prod_{j=1}^{p} (x^\partial_{ji})^{\beta_j}, \tag{4.1}$$

where $1\times p$ vector $x^\partial_i = (x^\partial_{1i}, x^\partial_{2i}, \dots, x^\partial_{pi})$ and $x^\partial_{ji} \overset{\text{iid}}{\sim} \text{Unif}(0,1), \forall j \in 1, \dots, p$. In addition, the ith true input-oriented efficiency score is generated from $\theta_i = 1/\lambda_i$, where $\lambda_i \sim |N(0, 0.42)| + 1$, i.e., we generate λ_i from the half normal distribution and then shift it up by 1 so that $\lambda_i \geq 1$.[13] Upon obtaining the frontier points $\{(x^\partial_i, y_i)\}^n_{i=1}$, we then set the ith observed input as $x_i = x^\partial_i/\theta_i$, i.e., by projecting (x^∂_i, y_i) from the frontier into the production set (x_i, y_i). Thus, a simulated sample $\{(x_i, y_i)\}^n_{i=1}$ is created. As the production technology exhibits VRS, we use VRS-DEA estimators to evaluate the methods. For each experiment, represented as one combination of (p, q, n), we perform 1000 trials of MC. For each trial, we check whether the estimated confidence intervals cover the true aggregate input efficiency. We then report the average of each trial's coverage over 1000 trials.

Table A.1 in Subsection B.1 of the Appendix B of the separate Appendices lists the values of β_j's and w_j's for each dimension p, where the β_j's are identical as in Simar and Zelenyuk (2020) and the values of w_j's are randomly chosen by the authors. The true aggregate input efficiency is obtained through simulating $10,000,000$ random realizations (x^∂_i, x_i) and then computing it using (2.41).

4.2. Simplified nomenclature

In the simulations, we are going to thoroughly compare the performance of the initial method in Theorem 2.1, the partial variance correction method adapted from Simar and Zelenyuk (2020), and the full variance correction method adapted from Simar et al. (2023). Further, we will also compare the above three methods for the cases—with and without the data sharpening adapted from Nguyen et al. (2022). Before presenting our MC results, to simplify the notation, we denote:

- Sol1: the original method in Theorem 2.1, i.e., the method in (2.52).
- Sol2: the partial variance correction method adapted from Simar and Zelenyuk (2020), i.e., the method in (3.1).
- Sol3: the full variance correction method adapted from Simar et al. (2023), i.e., the method in (3.15).
- Sol4: Sol1 combined with the data sharpening in (3.3).
- Sol5: Sol2 combined with the data sharpening in (3.3).
- Sol6: Sol3 combined with the data sharpening in (3.3).

4.3. Monte-Carlo results

4.3.1. General remarks

As discussed earlier, to maintain the consistency of the density estimator after the data sharpening, the values of γ need to be no less than $\tilde{\kappa}/2$ where $\tilde{\kappa} = \min(\kappa, 1/2)$. In our simulations, we first consider various values of γ from 0.55κ, 0.65κ, 0.75κ, 0.85κ, 0.95κ, κ, 1.2κ and 1.5κ and then select the γ with the observed best performance. Similar as Nguyen et al. (2022) for the simple mean efficiency, we find that the data sharpening can yield better performance for aggregate efficiency. However, unlike in Nguyen et al. (2022), for our context of input orientation we find that the level of γ near κ seems to yield the best performance. Consequently, in the following section we report the simulation results with $\gamma = \kappa$. The empirical coverages from the simulation results for 0.55κ, 0.65κ, 0.75κ, 0.85κ, 0.95κ, 1.2κ and 1.5κ are reported in the Subsection B.2 of the Appendix B of the separate Appendices.

[11] This theoretical justification is similar to that in Simar et al. (2023) and is adapted to the more complicated case of the weighted efficiency.

[12] See the Appendix D–E of the separate Appendices for more details.

[13] The selection of the parameter value of 0.42 is to ensure that the aggregate input-oriented efficiency is approximately 0.75; we tried many other values and obtained analogous results and conclusions.

4.3.2. Main results

Fig. 1 reports the empirical coverages for the aggregate input-oriented technical efficiency across dimensions and across finite sample sizes for Sol1–Sol6 when the nominal coverage is 0.95. The Monte Carlo simulations from Fig. 1 confirm the results from Simar et al. (2023): across different sample sizes and across different dimensions, in increasing order of performance are Sol1, Sol2, Sol3 for the cases without data sharpening; and in increasing order of performance are Sol4, Sol5, Sol6 for the cases with data sharpening.[14] For instance, when $n = 100$, $p = 4$, $q = 1$, the empirical coverage under Sol1, Sol2, and Sol3 is 0.487, 0.612, and 0.874, respectively, and under Sol4, Sol5, and Sol6 it is 0.772, 0.902, and 0.992, respectively. These results confirm that the proposed method in (3.15) further improves upon previous approaches for the approximation of the CLTs from Simar and Zelenyuk (2018) and their input oriented versions which we have presented here. In other words, for relatively small sample sizes, it is indeed useful to utilize the bias-corrected individual efficiency estimates (instead of the individual efficiency estimates) at every place they appear in the variance estimator of the aggregate efficiency.

Comparing the results without and with the data sharpening (i.e., comparing Sol1 vs. Sol4, Sol2 vs. Sol5, Sol3 vs. Sol6), we see that the data sharpening methods (Sol4, Sol5 and Sol6), in general, have better performance when the sample size is smaller than 100. Further, when the sample size is smaller than 200, the improvements of Sol6 are even more substantial and persistent relative to all the other five methods. For instance, when $n = 200$, $p = 5$, $q = 1$, the empirical coverage under Sol1, Sol2, Sol3, Sol4, Sol5, and Sol6 is 0.587, 0.721, 0.946, 0.765, 0.924, and 0.996 respectively. The empirical coverage under Sol6 is 0.996, i.e., higher than all the other five methods and higher than the nominal coverage of 0.95.

However, it is also important to note that when the sample size is larger than 200, the data sharpening methods in this context of input-oriented aggregate efficiency may perform worse than the corresponding methods without data sharpening methods, implying that the improvements of data sharpening methods might be limited to relatively small sample sizes (e.g., up to about 200 for the dimensions we have considered here).[15]

Furthermore, it is also important to note that Sol6 may slightly overshoot sometimes (e.g., see Fig. 1), especially when the dimensions are high. For instance, when $n = 100$, $p = 7$ and $q = 1$, the empirical coverage under Sol6 is 1.000 when the nominal coverage is 0.95 (see Fig. 1). Such overshooting phenomenon is not unique for Sol6, and is also observed for the other methods. Importantly, and as pointed out by Simar et al. (2023) who observed similar phenomena in their simulations, the error coming from such overshooting might be more preferable than the error from undershooting (provided that the length of the confidence intervals is similar). This is because the overshooting means a more conservative approach (i.e., *not rejecting* the *correct*).

We also try various values of the tuning constant γ, where those results are reported in the Subsection B.2 of the Appendix B of the separate Appendices. We find that the level of γ near κ seems to yield the best performance for our context (input-oriented aggregate efficiency). We also present simulation results for the aggregate output-oriented efficiency in the Appendix C of the separate Appendices, which also shows that the level of $\gamma = \kappa$ seems to yield better performance. This finding is slightly different from Nguyen et al. (2022) who found the level of γ near 0.75κ seems to yield the best performance in a different though similar context (about the simple mean output-oriented efficiency). This suggests that the choice of the tuning constant seems to play a more important role than we had thought earlier, and hence this topic (selection of γ) may constitute a fruitful avenue for future research.

To conclude for this section, our general conclusion is that: when the sample sizes are relatively small (around 200 or less for the dimensions we have considered here), the full variance correction method combined with the data sharpening method (Sol6) provides a better performance; for the larger sample sizes, the full variance correction method without the data sharpening method (Sol3) provides a better or the same performance as other suitable methods known to date.

5. Empirical illustrations and guidance for practitioners

In this section, we use two real data sets to illustrate the differences between the six methods (Sol1–Sol6) discussed above in the estimates of variance as well as the confidence intervals for the aggregate efficiency. To provide different perspectives of the usefulness of the full variance correction method in different estimators and different orientations, we use VRS-DEA estimators in the input orientation for one data set in Section 5.2.1, while using CRS-DEA estimators in the output orientation for another data set in Section 5.2.2. Before presenting the two empirical illustrations, an example of a practical implementation algorithm is outlined in the following Section 5.1, as a guide for practitioners.

5.1. A practical implementation algorithm

While there is hardly "one way to fit all" approach to the empirical studies, it still seems useful to sketch a plan or an algorithm consisting of inter-related steps for how such studies can be conducted. This is especially helpful for cases where many different and relatively complex methods exist for studying a research question. Here we will present one such algorithm, acknowledging that other algorithms might work as well or better. The steps of a practical algorithm are visualized in Fig. 2 and briefly discussed in the following steps. The technical details for Steps 3–7 are provided in Appendix F of the separate Appendices.

Step 1. Identify the research questions of interest, and formulate them as hypotheses to be explored and tested. Learn the necessary background of the context, observe the related phenomena of interest and obtain the relevant data.

Step 2. Based on the information from the Step 1, build an appropriate production model: select the relevant inputs and outputs, and decide which assumptions on technology suggested by the theory are to be maintained and which are to be tested. (Here, one may need some iterations with the previous step, to re-tune the existing models to be coherent with the data at hand, or to get additional data, etc.) Also, the researcher needs to choose the relevant reference frontier and the orientation for measuring the efficiency, and then select an appropriate estimator. In this paper we are focusing on measuring efficiency in either input or output orientation, with respect to the unconditional technology frontier, using DEA with CRS or VRS assumptions and free disposability for all inputs and for all outputs, though many other variations are possible, which would require adequate modifications.

Step 3. Estimate the individual and aggregate efficiency scores for the cases without and with the data sharpening method (as discussed in Section 3.2).

Step 4. Use the generalized jackknife-type approach to estimate the bias for the individual and aggregate efficiency estimates (as discussed in Subsection A and Appendix F of the separate Appendices).

Step 5. Estimate the variance proposed by various methods discussed in Sections 3 and 2.4.

Step 6. Compute the estimated confidence intervals using Sol1–Sol6 denoted in Section 4.2.

[14] For the values of the empirical coverages, see Table A.2 in the Appendix B of the separate Appendices. Further, when $\kappa = 2/5$ for VRS-DEA, Simar and Zelenyuk (2018) finds that both Theorem 2 and Theorem 3 in Simar and Zelenyuk (2018) are applicable, while Theorem 3 is preferred. Thus, for $p = 3$ and $q = 1$, we only include the results from the preferred method.

[15] The advantage of data sharpening is more pronounced for data sets where the estimator yields many efficiency scores equal to 1. This happens more often for relatively small samples, especially with large dimensions, and especially for FDH.

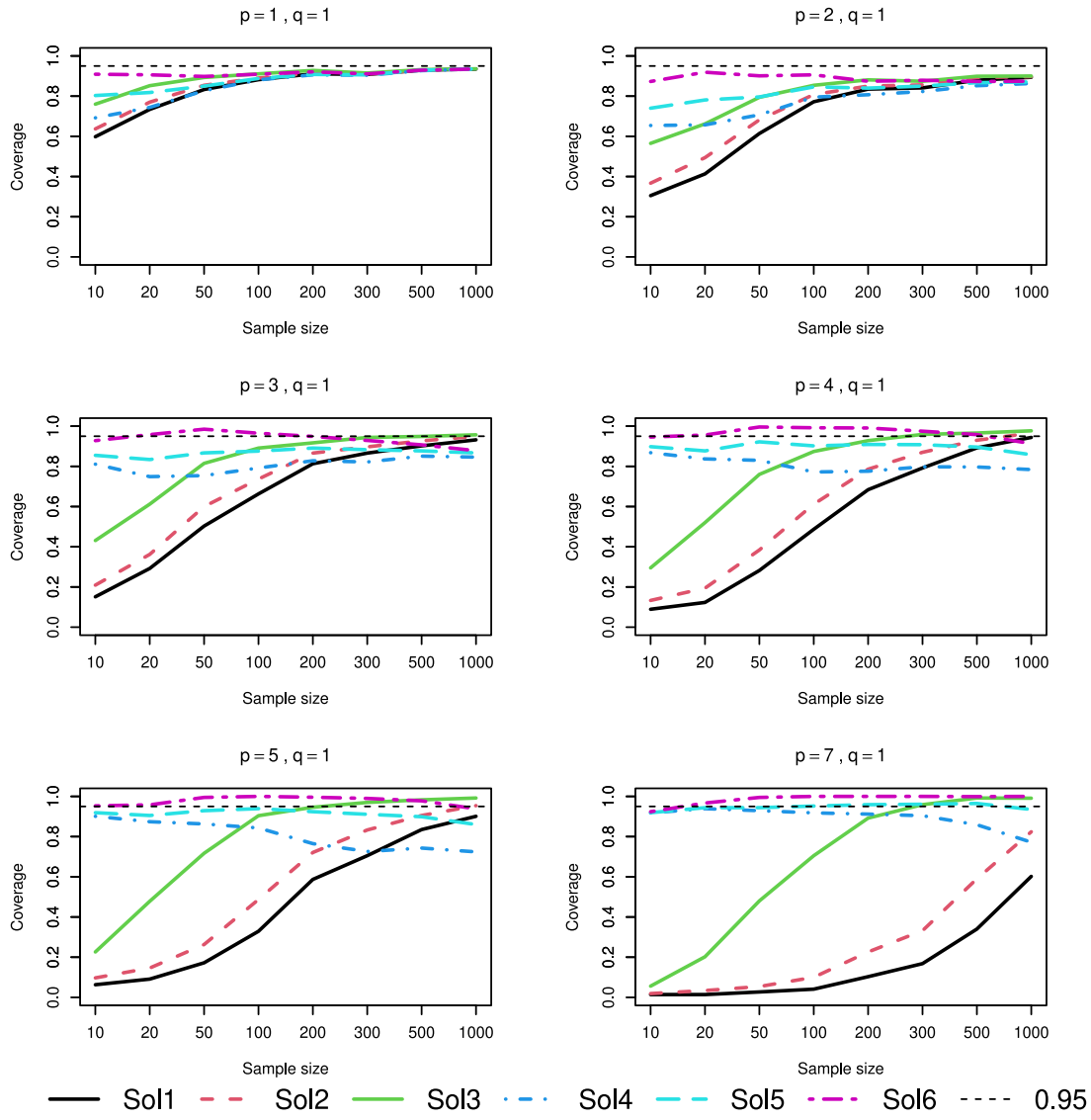


Fig. 1. Empirical Coverages for the Aggregate Input-Oriented Efficiency when $\gamma = \kappa$.

Step 7. Use the estimated confidence intervals from Step 6 to test the hypotheses formulated in Step 1. Applied researchers could compare the differences and similarities among Sol1–Sol6. However, as evident in our MC experiments, we suggest that when the sample sizes are relatively small (around 200 or especially less), Sol6 provides a better performance, while for the larger sample sizes, Sol3 is expected to perform the best.

Step 8. Draw conclusions, acknowledging the caveats/limitations and challenges for future research.

5.2. Empirical illustration

5.2.1. Inference for aggregate input-oriented efficiency with philippine rice data

We have two goals for the first illustration. First, we compare the differences and similarities among Sol1–Sol6 for the aggregate input-oriented efficiency. Our second goal is to verify the general conclusion from the MC experiments: when the sample sizes are relatively small (around 200 or especially less), Sol6 provides a better performance, while for the larger sample sizes, Sol3 is expected to perform the best.

The first illustration uses the same data set as Simar and Zelenyuk (2020) and Nguyen et al. (2022). This data set is a balanced panel data set, consisting of 43 rice producers in the Tarlac region of the Philippines over 1990–1997.[16] The rice producers are modeled to use three inputs (area planted x_1, labor used x_2 and fertilizers x_3) and their prices (w_1, w_2 and w_3) to produce one output (rice y). The total cost is defined as the sum of expenses on inputs, i.e., $cost = w_1x_1+w_2x_2+w_3x_3$.

Same as Simar and Zelenyuk (2020), we first obtain the results using only 43 observations for each year separately and then using all the 344 observations for the pooled data. Table 1 summarizes the results using VRS-DEA estimators in the input orientation. For Panel B with data sharpening methods, we let $\gamma = \kappa$ as indicated from MC simulations in the previous Section, and $\kappa = n^{-2/5}$ as we use VRS-DEA estimators and $p = 3$, $q = 1$.

First, for Panel A without data sharpening, for each year, in increasing width of confidence intervals (CIs) lie Sol1, Sol2 and Sol3; this result is due to the fact that, while Sol1–Sol3 yield the same aggregate efficiency estimates, the estimates of σ_φ are typically the smallest for

[16] This data set was compiled by International Rice Research Institute and widely used in Coelli et al. (2005) for various illustrations of efficiency and productivity estimations. The data are available at http://www.uq.edu.au/economics/cepa/crob2005/software/CROB2005.zip.

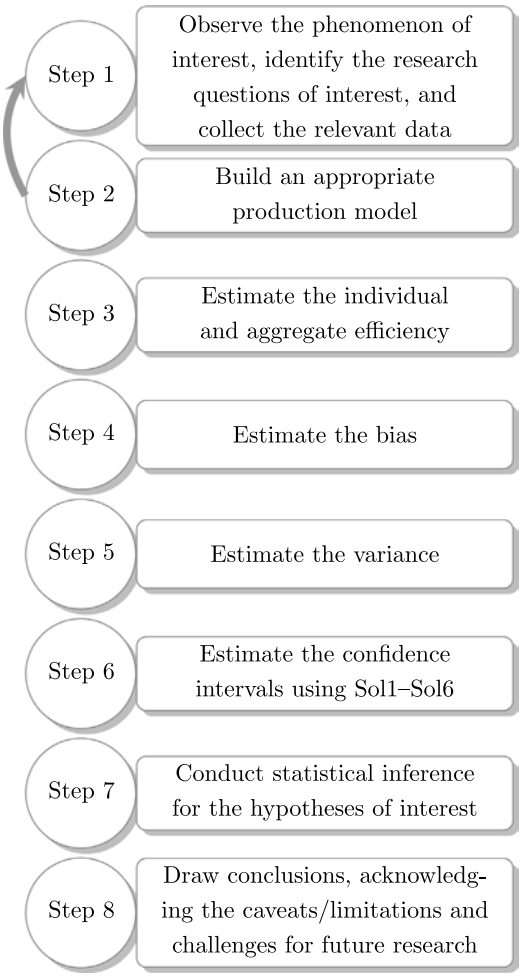


Fig. 2. The Steps of the Algorithms.

Sol1 and the largest for Sol3, with Sol2 being between these two. For Panel B with data sharpening, for each year, in increasing width of CI lie Sol4, Sol5 and Sol6. Again, this result is due to the fact that, while Sol4–Sol6 yield the same aggregate efficiency estimates, the estimates of σ_{φ} are usually the smallest for Sol4 and the largest for Sol6. These results are consistent with the findings in the MC simulations that the full variance correction method yields slightly larger coverages than the previous methods, with and without the data sharpening method, suggesting a potential improvement of the full variance correction method for the finite sample sizes.

Second, the addition of the data sharpening method has two effects: (1) decreasing the estimates of aggregate efficiency; (2) decreasing the estimates of σ_{φ}. Due to the effect (i), the centering of the estimated CI with the data sharpening method will be slightly further away from 1 (toward 0) than the corresponding case without the data sharpening method, as the data sharpening method "smoothes out" the observations with estimated efficiency that equal 1 and within the neighborhood of 1. Meanwhile, due to the effect (ii), the data sharpening method yields a narrower CI than the corresponding case without the data sharpening method. Note that a narrower CI for the data sharpening method does not mean it will perform worse than the corresponding case without the data sharpening method, since the data sharpening method might have a more accurate estimate for the aggregate efficiency by smoothing out the spurious observations. In other words, the data sharpening method might have a more accurate estimate for the centering of CIs. This is mainly confirmed in our previous Monte Carlo evidence, where we find that: Sol6 persistently gave better coverage when the sample sizes are relatively small (up to 200), meanwhile when the sample size was relatively large (> 200), Sol3 usually performed better.

All in all, combining the above first and second discussions, for this specific empirical data set, we anticipate that Sol6 is likely to be more accurate for each year (with the sample size 43), while for the pooled sample (with the sample size 344), Sol3 is likely to be more accurate. And indeed our results show that for the pooled sample the CI for Sol3 ($[0.3423, 0.5147]$) is expected to have a better coverage than that for Sol6 ($[0.3447, 0.5093]$) as the CI of Sol6 is contained by the CI of Sol3. Although, and importantly, note that in the latter case, all the methods give very similar estimates of the CIs, supporting the theory that says they are all asymptotically equivalent. Meanwhile, the difference is fairly substantial among the methods for the individual years where the samples are relatively small, emphasizing the importance to select a method that promises higher accuracy. Thus, based on our discussions in the theoretical and simulation sections, we would recommend relying more on Sol6 for these relatively small samples.

5.2.2. Inference for aggregate output-oriented efficiency with penn world table data

For the second illustration, we further compare the differences and similarities among Sol1–Sol6 for both the simple mean and aggregate output-oriented efficiency to illustrate the practical usefulness of the full variance correction method. Moreover, we consider both CRS-DEA and VRS-DEA estimators while focusing on the results obtained from CRS-DEA estimators.[17]

Our second illustration uses the publicly available data from Penn World Table (PWT 10.0, Feenstra et al., 2015), which provides information on relative levels of aggregate input, output, income, etc, for many countries across the world from 1950 to 2019.[18] Following Färe et al. (1994), Kumar and Russell (2002), and Badunenko et al. (2008), we model the production of countries using labor and capital stock to produce GDP. The labor is measured using the number of persons engaged (in millions), i.e., the variable emp in PWT. The capital is measured using capital stock at current PPPs (in millions 2017 US$), i.e., the variable cn in PWT. The GDP is measured by output-side real GDP at current PPPs (in millions 2017 US$), i.e., the variable $cgdpo$ in PWT. The output price is the same for all countries. To illustrate the evolution of the aggregate inefficiency over the latest 30 years (from 1990 to 2019), we will use the same sub-set of 84 countries/regions that was considered in Badunenko et al. (2008) and Pham et al. (2023).[19]

In previous DEA studies that utilized PWT data, output orientation is usually deployed (Färe et al., 1994; Ray & Desli, 1997; Kumar & Russell, 2002; and Badunenko et al., 2008) and so we focus on this case here. We use CRS-DEA estimators and $p = 2$, $q = 1$, thus $\kappa = n^{-2/3}$. When using data sharpening methods, we let $\gamma = \kappa$. We obtain the results using only 84 observations for each year separately from 1990 to 2019 as the technology might have changed over time. Further, we focus on aggregate inefficiency while also presenting the results for simple mean inefficiency to compare.

Fig. 3(a) presents the dynamic changes of the aggregate output-oriented inefficiency over 1990–2019.[20] As expected, the "Standard" (i.e., the estimated aggregate inefficiency without bias correction) is substantially below the "Bias-Corrected" (i.e., the estimated aggregate

[17] We also obtain the results using VRS-DEA estimators. Their similarities and differences are reported in the Appendix G.2 of the separate Appendices.

[18] The data are available at https://www.rug.nl/ggdc/productivity/pwt/?lang=en.

[19] The countries/regions included in our sample are listed in the Appendix C of the separate Appendices.

[20] Tables A.1 and A.2 in the Appendix G of the separate Appendices present more details on the results of our estimations for aggregate inefficiency and simple mean inefficiency, respectively.

Table 1
VRS-DEA estimates of aggregate input-oriented efficiency, standard deviations, and the 99% confidence intervals for rice producers in the Philippines.

Panel A: Without data sharpening

Year	$\hat{\varphi}_{n_x}$	$\hat{B}_{\varphi,n}$	$\hat{\varphi}_{n_x} - \hat{B}_{\varphi,n}$	$\hat{\sigma}_{\varphi}$			99% CI					
				Sol1	Sol2	Sol3	Sol1		Sol2		Sol3	
1990	0.8013	0.1708	0.6304	0.2124	0.2726	0.3476	0.5081	0.7528	0.4734	0.7874	0.4302	0.8307
1991	0.8157	0.1099	0.7058	0.2201	0.2460	0.3117	0.5791	0.8326	0.5641	0.8475	0.5263	0.8854
1992	0.8944	0.1034	0.7910	0.1317	0.1674	0.2586	0.7151	0.8668	0.6945	0.8874	0.6420	0.9399
1993	0.8586	0.0986	0.7600	0.2070	0.2293	0.2887	0.6408	0.8793	0.6280	0.8921	0.5938	0.9263
1994	0.8243	0.1751	0.6492	0.2663	0.3187	0.4500	0.4958	0.8026	0.4656	0.8327	0.3900	0.9083
1995	0.8665	0.1215	0.7451	0.2163	0.2481	0.3442	0.6204	0.8697	0.6021	0.8880	0.5468	0.9433
1996	0.8297	0.1273	0.7024	0.2466	0.2776	0.3540	0.5604	0.8445	0.5426	0.8623	0.4985	0.9063
1997	0.8166	0.2387	0.5779	0.2281	0.3302	0.4561	0.4465	0.7093	0.3877	0.7681	0.3152	0.8406
Pooled	0.6152	0.1867	0.4285	0.2654	0.3245	0.3445	0.3621	0.4949	0.3473	0.5097	0.3423	0.5147

Panel B: With data sharpening

Year	$\hat{\varphi}_{n_x}$	$\hat{B}_{\varphi,n}$	$\hat{\varphi}_{n_x} - \hat{B}_{\varphi,n}$	$\hat{\sigma}_{\varphi}$			99% CI					
				Sol4	Sol5	Sol6	Sol4		Sol5		Sol6	
1990	0.7747	0.1686	0.6061	0.1861	0.2511	0.3191	0.4989	0.7133	0.4615	0.7507	0.4223	0.7899
1991	0.7349	0.1076	0.6274	0.1696	0.2009	0.2456	0.5296	0.7251	0.5117	0.7430	0.4859	0.7688
1992	0.8057	0.0955	0.7102	0.1164	0.1505	0.2258	0.6432	0.7773	0.6235	0.7970	0.5802	0.8403
1993	0.7637	0.0926	0.6711	0.1607	0.1855	0.2327	0.5785	0.7636	0.5642	0.7779	0.5370	0.8051
1994	0.7400	0.1704	0.5695	0.1971	0.2606	0.3645	0.4560	0.6831	0.4195	0.7196	0.3596	0.7795
1995	0.7913	0.1173	0.6741	0.1616	0.1997	0.2856	0.5810	0.7672	0.5590	0.7891	0.5095	0.8386
1996	0.7521	0.1204	0.6317	0.2000	0.2335	0.3043	0.5165	0.7469	0.4972	0.7662	0.4564	0.8069
1997	0.7537	0.2339	0.5198	0.1583	0.2824	0.3757	0.4287	0.6110	0.3572	0.6825	0.3034	0.7362
Pooled	0.6135	0.1866	0.4270	0.2521	0.3136	0.3290	0.3639	0.4900	0.3485	0.5054	0.3447	0.5093

inefficiency with bias correction). This is also true for the case using the data sharpening method, where the "Standard+Sharpened" (i.e., the estimated aggregate inefficiency without bias correction and with the data sharpening method) is below the "Bias-Corrected+Sharpened" (i.e., the estimated aggregate inefficiency with bias correction and with the data sharpening method). Comparing the cases with and without the data sharpening method, it is also expected that the "Standard" is below the "Standard+Sharpened" and that the "Bias-Corrected" is below the "Bias-Corrected+Sharpened", as the data sharpening method "smoothes out" the observations that are spuriously equal to 1 and those within the neighborhood of 1.

Notice that the average values of annual aggregate inefficiency estimates from 1990 to 2019 (i.e., we first calculate the aggregate inefficiency estimate for each year and then calculate average value of aggregate inefficiency estimates over these 30 years) are 1.7990 for "Bias-Corrected" and 1.8029 for "Bias-Corrected+Sharpened", implying that large inefficiency exists in these countries/regions covered in our sample and that they have the potential to nearly double their GDP without increasing inputs.

Further, Fig. 3(a) shows that, for all these four cases, the aggregate inefficiency generally first decreases from 1990 to 1997, then increases until 2006, then it is very disruptive from 2007 to 2011 (during the global financial crisis), then again decreases until 2016 and increases again until 2019.

Fig. 3(b) presents the corresponding results for the simple mean output-oriented inefficiency. The relationships among the "Standard", "Bias-Corrected", "Standard+Sharpened" and "Bias-Corrected+Sharpened" still hold as those for aggregate inefficiency. The average values of the annual simple mean inefficiency estimates from 1990 to 2019 are 2.1345 for "Bias-Corrected" and 2.1362 for "Bias-Corrected+Sharpened", which are slightly larger than the corresponding case for aggregate inefficiency (1.7990 and 1.8029, respectively).

Further, the dynamic changes of the simple mean inefficiency are substantially different from the aggregate inefficiency, especially before 2005. This confirms the expectations that it is useful for empirical researchers to report both unweighted and weighted inefficiency to check whether there is any substantial difference, potentially presenting a different big picture of the performance of a group of interest.

Fig. 4(a) presents the estimates of the standard deviation σ_{φ} for the aggregate output-oriented inefficiency for Sol1–Sol6 over 1990–2019. It is clear that Sol3 has larger estimates for σ_{φ} than Sol1 and Sol2, and that Sol6 has larger estimates for σ_{φ} than Sol4 and Sol5, suggesting that the estimated confidence intervals using the full variance correction method (3.15) will be slightly wider than the previous methods, whether taken with or without the data sharpening method. Further, for each year the results for Sol3 and Sol6 are generally similar in terms of the estimates for σ_{φ}. The results for the standard deviations for the simple mean inefficiency presented in Fig. 4(b) are similar to those for aggregate inefficiency where Sol3 and Sol6 have slightly larger values than the remaining methods.

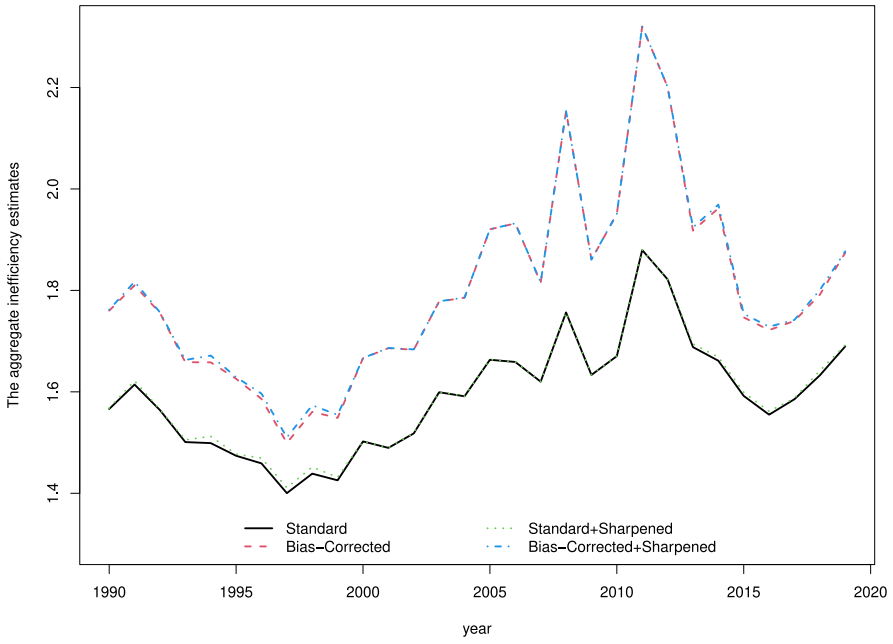
To compare Sol3 and Sol6 which could generate wider confidence intervals, we present the dynamics of the aggregate inefficiency estimates and their CIs in Fig. 5(a). The center of the CIs is the corresponding bias-corrected aggregate output-oriented inefficiency. The results for Sol3 and Sol6 in Fig. 5(a) are generally similar in terms of centers and boundaries of CIs, although the centers and lower boundaries for Sol6 are slightly higher than those for Sol3. Fig. 5(b) presents the corresponding results for simple mean inefficiency, which are generally similar to Fig. 5(a) for the aggregate inefficiency.

As a conclusion for this empirical illustration section, we use two different real data sets and different estimators (input-oriented VRS vs. output-oriented CRS) to show the practical differences of the full variance correction method for approximating the new CLT theorems for the aggregate efficiency.[21]

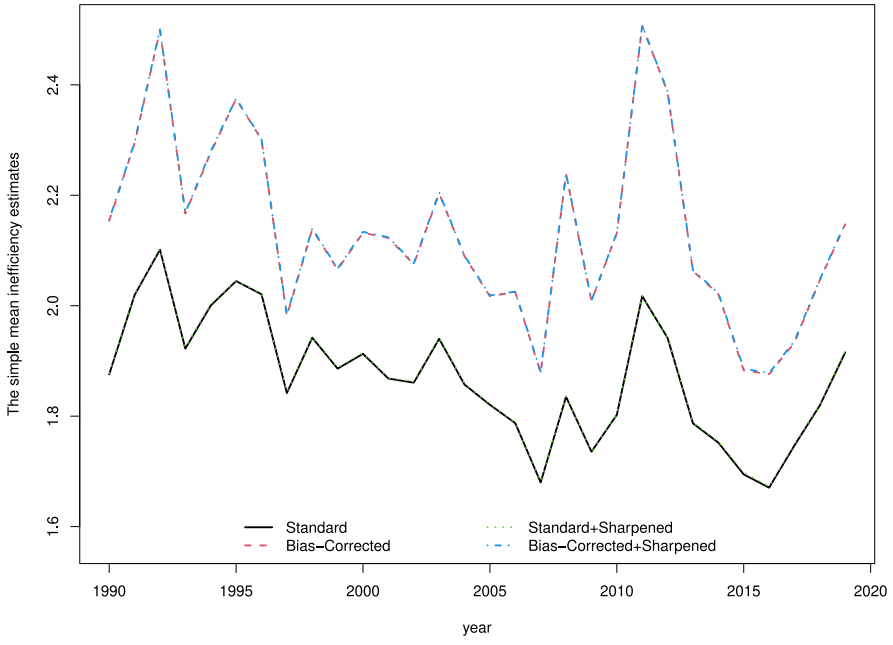
6. Conclusions

The CLT results for the aggregate output-oriented efficiency estimated via DEA and FDH developed by Simar and Zelenyuk (2018) are useful recent advancements of the statistical theory for efficiency and productivity analyses, complementing with the CLT results for the individual and simple mean efficiency developed by Kneip et al. (2015). Analogous to the output-oriented framework, we spell out the key CLT results for the aggregate input-oriented efficiency to better guide the applied researchers.

[21] The codes used in these two empirical illustrations are available in the GitHub. See https://github.com/srzhao89/szz-aggregate-efficiency for more details.



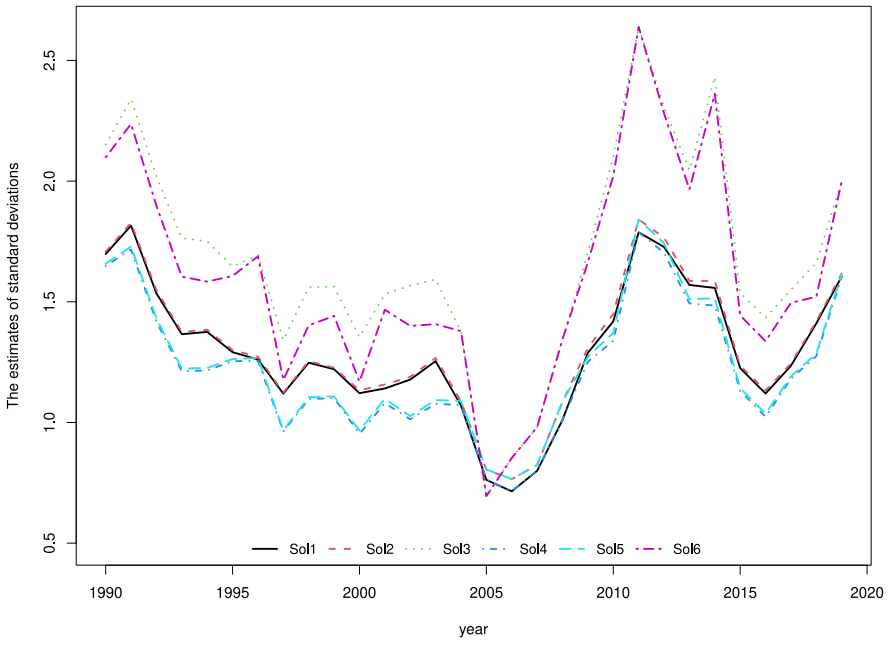
(a) For the Aggregate Output-Oriented Inefficiency



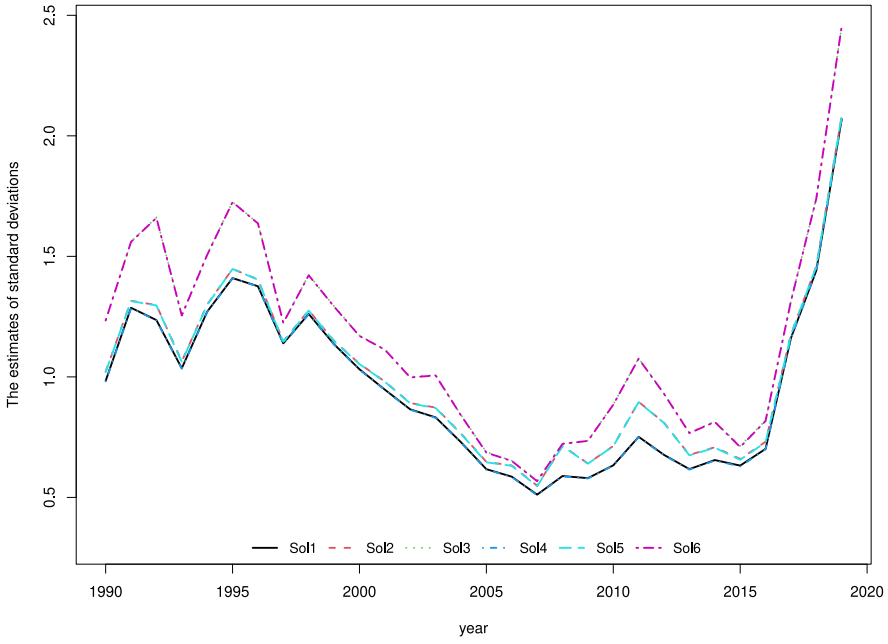
(b) For the Simple Mean Output-Oriented Inefficiency

Fig. 3. CRS-DEA Estimates of Aggregate/Simple Mean Output-Oriented Inefficiency for Countries/Regions from 1990 to 2019 when $\gamma = \kappa$.

Further, for relatively small sample sizes, the coverages of the estimated confidence intervals based on the CLT results for aggregate efficiency are well below the nominal coverage. This under-covering phenomenon in relatively small sample sizes was also observed for the simple mean efficiency. Some of the improvements were made through Simar and Zelenyuk (2020), Nguyen et al. (2022), and Simar et al. (2023) for the simple mean output-oriented efficiency, and through Simar and Zelenyuk (2020) for the aggregate output-oriented efficiency. However, whether the methods in Nguyen et al. (2022) and Simar et al. (2023) are effective to improve the finite sample performance of CLT results for aggregate input-oriented and output-oriented efficiency remains unknown. We fill the clear gap in the literature to thoroughly examine the performance of the partial variance correction method adapting from Simar and Zelenyuk (2020) and the full variance correction method adapting from Simar et al. (2023) for both cases—with and without the data sharpening method in Nguyen et al. (2022) (compared with the original CLT results) for the aggregate input-oriented and output-oriented efficiency through



(a) For the Aggregate Output-Oriented Inefficiency

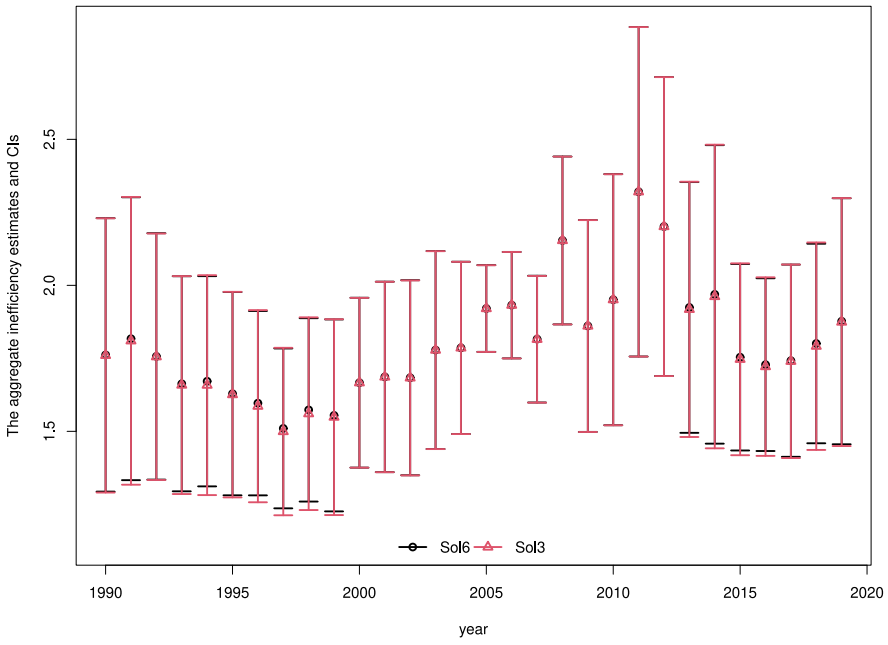


(b) For the Simple Mean Output-Oriented Inefficiency

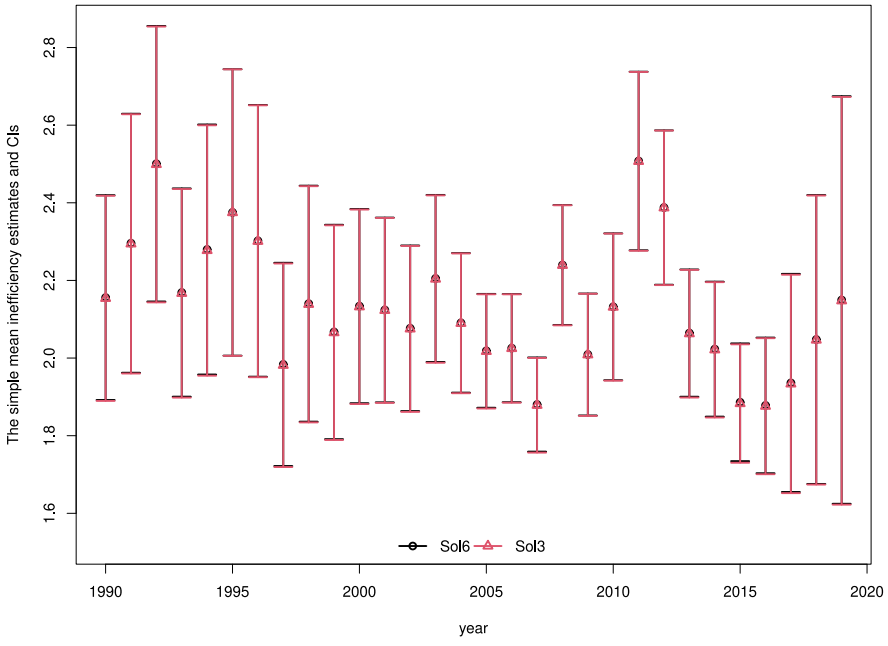
Fig. 4. CRS-DEA Estimates of Standard Deviations for Aggregate/Simple Mean Output-Oriented Inefficiency for Countries/Regions from 1990 to 2019 when $\gamma = \kappa$.

simulations. We find that: when the sample sizes are not large (≤ 200), the full variance correction method combined with the data sharpening method provides a better performance; and when the sample sizes are relatively large (> 200), the full variance correction method without the data sharpening method provides a better performance. In general, this holds for both small and large dimensions. Therefore, we are able to show that the methods in Nguyen et al. (2022) and Simar et al. (2023) are not only useful to improve the finite sample performance of CLT results for the simple mean efficiency, but also are useful for the aggregate efficiency, although of course this is not a panacea.

More importantly, we provide the guidance and the codes made available for practitioners to implement this framework for the aggregate efficiency. We also use the well-known Philippines rice data set and Penn World Table data set as two examples to illustrate the differences between all these methods in the estimated variance and the estimated confidence intervals for the aggregate input-oriented and output-oriented (respectively) efficiency.



(a) For the Aggregate Output-Oriented Inefficiency



(b) For the Simple Mean Output-Oriented Inefficiency

Fig. 5. CRS-DEA Estimates of Aggregate/Simple Mean Output-Oriented Inefficiency and their 95% Confidence Intervals for Countries/Regions from 1990 to 2019 when $\gamma = \kappa$.

Finally, this paper also shows that the selecting of the tuning parameter in the data sharpening methods plays an important role here, which was not discovered previously in Nguyen et al. (2022) and Simar et al. (2023), indicating that this might be a fruitful area for future research. We hope further research could hopefully bring more insights on this challenging topic. The full variance correction in this paper could be extended naturally to geometric means of the Malmquist index, technical efficiency change, scale efficiency change, technology change, and others.

Acknowledgments

Valentin Zelenyuk acknowledges the support from the Australian Research Council (FT170100401) and The University of Queensland. Shirong Zhao acknowledges the support from Liaoning Provincial Department of Education Research Fund (LJKMZ20221602). Feedback from Evelyn Smart and Zhichao Wang is appreciated. We also thank Paul Wilson as the programming codes used in this paper involve some earlier codes from Paul Wilson. These individuals and organizations are

not responsible for the views expressed here. These authors contributed equally: Léopold Simar, Valentin Zelenyuk, Shirong Zhao.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ejor.2024.01.028.

References

Badunenko, O., Henderson, D. J., & Zelenyuk, V. (2008). Technological change and transition: Relative contributions to worldwide growth during the 1990s. *Oxford Bulletin of Economics and Statistics*, *70*(4), 461–492.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*, 1078–1092.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*, 429–444.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. springer science & business media.

Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega*, *44*, 1–4.

Daraio, C., Simar, L., & Wilson, P. W. (2018). Central limit theorems for conditional efficiency measures and tests of the 'separability'condition in non-parametric, two-stage models of production. *The Econometrics Journal*, *21*(2), 170–191.

Deprins, D., Simar, L., & Tulkens, H. (1984). Measuring labor inefficiency in post offices. In M. M. P. Pestieau, & H. Tulkens (Eds.), *The performance of public enterprises: concepts and measurements* (pp. 243–267). Amsterdam: North-Holland.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1985). *The measurement of efficiency of production*. Boston: Kluwer-Nijhoff Publishing.

Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, *84*, 66–83.

Färe, R., Grosskopf, S., & Zelenyuk, V. (2004). Aggregation of cost efficiency: Indicators and indexes across firms. *Academia Economic Papers*, *32*, 395–411.

Färe, R., & Zelenyuk, V. (2003). On aggregate Farrell efficiencies. *European Journal of Operational Research*, *146*(3), 615–620.

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

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, *105*(10), 3150–3182.

Kneip, A., Park, B., & Simar, L. (1998). A note on the convergence of nonparametric DEA efficiency measures. *Econometric Theory*, *14*, 783–793.

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

Kneip, A., Simar, L., & Wilson, P. W. (2011). A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators. *Computational Economics*, *38*, 483–515.

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

Kneip, A., Simar, L., & Wilson, P. W. (2016). Testing hypotheses in nonparametric models of production. *Journal of Business & Economic Statistics*, *34*, 435–456.

Kneip, A., Simar, L., & Wilson, P. W. (2021). Inference in dynamic, nonparametric models of production: Central limit theorems for Malmquist indices. *Econometric Theory*, *37*(3), 537–572.

Koopmans, T. C. (1957). *Three essays on the state of economic science*. New York, NY: McGraw-Hill.

Kumar, S., & Russell, R. R. (2002). Technological change, technological catch-up, and capital deepening: relative contributions to growth and convergence. *American Economic Review*, *92*(3), 527–548.

Mayer, A., & Zelenyuk, V. (2019). Aggregation of individual efficiency measures and productivity indices. In *The palgrave handbook of economic performance analysis* (pp. 527–557). Springer.

Nguyen, H., Simar, L., & Zelenyuk, V. (2022). Data sharpening for improving CLT approximations for DEA-type efficiency estimators. *European Jouranl of Operational Research*, *303*(3), 1469–1480.

Park, B. U., Jeong, S.-O., & Simar, L. (2010). Asymptotic distribution of Conical-Hull estimators of directional edges. *The Annals of Statistics*, *38*, 1320–1340.

Park, B. U., Simar, L., & Weiner, C. (2000). FDH efficiency scores from a stochastic point of view. *Econometric Theory*, *16*, 855–877.

Pham, M. D., Simar, L., & Zelenyuk, V. (2023). Statistical inference for aggregation of Malmquist productivity indices. *Operations Research*, http://dx.doi.org/10.1287/opre.2022.2424.

Ray, S. C., & Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *American Economic Review*, *87*, 1033–1039.

Sickles, R. C., & Zelenyuk, V. (2019). *Measurement of productivity and efficiency*. Cambridge University Press.

Simar, L., & Wilson, P. W. (2015). Statistical approaches for non-parametric frontier models: A guided tour. *International Statistical Review*, *83*, 77–110.

Simar, L., & Wilson, P. W. (2019). Central limit theorems and inference for sources of productivity change measured by nonparametric Malmquist indices. *European Journal of Operational Research*, *277*(2), 756–769.

Simar, L., & Wilson, P. W. (2020). Technical, allocative and overall efficiency: Estimation and inference. *European Journal of Operational Research*, *282*(3), 1164–1176.

Simar, L., & Zelenyuk, V. (2006). On testing equality of two distribution functions of technical efficiency scores estimated via DEA. *Econometric Reviews*, *24*, 497–522.

Simar, L., & Zelenyuk, V. (2018). Central limit theorems for aggregate efficiency. *Operations Research*, *66*(1), 137–149.

Simar, L., & Zelenyuk, V. (2020). Improving finite sample approximation by central limit theorems for estimates from data envelopment analysis. *European Journal of Operational Research*, *284*(3), 1002–1015.

Simar, L., Zelenyuk, V., & Zhao, S. (2023). Further improvements of finite sample approximation of central limit theorems for envelopment estimators. *Journal of Productivity Analysis*, *59*(2), 189–194.

Walheer, B. (2019). Aggregating Farrell efficiencies with private and public inputs. *European Journal of Operational Research*, *276*(3), 1170–1177.

Wilson, P. W. (2018). Dimension reduction in nonparametric models of production. *European Journal of Operational Research*, *267*, 349–367.

Zelenyuk, V. (2020). Aggregation of efficiency and productivity: From firm to sector and higher levels. In S. C. Ray, R. Chambers, & S. Kumbhakar (Eds.), *Handbook of production economics* (pp. 1–40). Singapore: Springer Singapore.