STATISTICAL INFERENCE FOR AGGREGATION OF MALMQUIST PRODUCTIVITY INDICES

Manh D. Pham, Léopold Simar, Valentin Zelenyuk

LIDAM Discussion Paper ISBA
2022 / 05

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

Statistical Inference for Aggregation of Malmquist Productivity Indices*

Manh D. Pham[†1], Léopold Simar[2], and Valentin Zelenyuk[1,3]

[1] *School of Economics, The University of Queensland, Brisbane QLD 4072, Australia*
[2] *Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, B1348 Louvain-la-Neuve, Belgium*
[3] *Centre for Efficiency and Productivity Analysis, The University of Queensland, Brisbane QLD 4072, Australia*

January 2022

Abstract

The Malmquist Productivity Index (MPI) has gained popularity amongst studies on dynamic change of productivity of decision-making units (DMUs). In practice, this index is frequently reported at aggregate levels (e.g., public and private firms) in the form of simple equally-weighted arithmetic or geometric means of individual MPIs. A number of studies have emphasized that it is necessary to account for the relative importance of individual DMUs in the aggregations of indices in general and of MPI in particular. While more suitable aggregations of MPIs have been introduced in the literature, their statistical properties have not been revealed yet, preventing applied researchers from making essential statistical inferences such as confidence intervals and hypothesis testing. In this paper, we will fill this gap by developing a full asymptotic theory for an appealing aggregation of MPIs. On the basis of this, some meaningful statistical inferences are proposed and their finite-sample performances are verified via extensive Monte Carlo experiments.

Key words: aggregation, asymptotics, DEA, hypothesis test, inference, Malmquist index, productivity

JEL classification: C14, C44, C51, D24, M11

*Email addresses: ducmanh.pham@uq.net.au (M.D. Pham), leopold.simar@uclouvain.be (L. Simar), v.zelenyuk@uq.edu.au (V. Zelenyuk).

[†]Corresponding author at ducmanh.pham@uq.net.au.

1 Introduction

The Malmquist Productivity Index (MPI), since its first introduction by Caves et al. (1982), has become one of the most widely used tools for analyzing performance of decision making units (DMUs) in terms of productivity change over time. Färe et al. (1998) pointed out a large number of applications of MPI in a wide variety of areas, including "agriculture, airlines, banking, electric utilities, insurance companies, and public sectors."[1] Among the approaches to compute MPI, the nonparametric Data Envelopment Analysis (DEA) appears to be the most popular in the literature. According to Chen and Ali (2004), "this DEA-based Malmquist Productivity Index has proven itself to be a good tool for measuring the productivity change of DMUs" and "there is a substantial body of applications that uses the DEA-based Malmquist Productivity Index." The survey of Zhou et al. (2008) also pointed out that MPI is one of the three most common extensions of basic DEA models in energy and environment studies.

In practice, apart from measuring the productivity change of a single DMU, there is also an essential need for analyzing productivity change at an aggregate level (e.g., firms grouped by ownership status such as public and private).[2] For instance, while studying performances of European and US banking systems, Pastor et al. (1997) used the median, simple average, and weighted average by total assets of individual MPIs of banks in each country to make the cross-country comparison. Another example is Tortosa-Ausina et al. (2008) who examined productivity growth and the productive efficiency of Spanish savings banks over the period 1992-1998 and reported both the simple arithmetic and geometric means of the MPIs of individual banks.[3] However, when it comes to aggregation of indices, it has been pointed out in a number of studies that simple averages (i.e., arithmetic and geometric means) which assign the same weight to each individual regardless of their relative economic significance (e.g., market share) might lead to very different conclusions

[1]The Google Scholar web search found about 24,400 results for "Malmquist productivity index" (as of March 11, 2019).

[2]There is another research stream that studies decompositions of MPI (e.g., Färe et al., 1994b; Johnson and Ruggiero, 2014; Brennan et al., 2014; Simar and Wilson, 2019).

[3]A few more examples are Färe et al. (1990); Mukherjee et al. (2001); Ball et al. (2004); Murillo-Zamorano (2005); Abbott (2006).

in relation to the averages that account for economic weight (e.g., see Ylvinger, 2000; Ebert and Welsch, 2004).

Accounting for the economic weights of DMUs based on their relative importance is also consistent with reality where many industries are usually dominated by a minority of firms. For example, according to data from the Federal Reserve System (2019), about 40% of total domestic assets of 1,835 large commercial banks in the US were occupied by the four largest ones (J P Morgan Chase, Bank of America, Wells Fargo, Citibank) as of March 31, 2019. Consequently, the aggregate functional forms and weights should be considered carefully in order to achieve meaningful aggregations. This consideration also encompasses productivity and efficiency indices in general and MPI in particular (e.g., see Färe and Zelenyuk, 2003; Färe and Grosskopf, 2004; Zelenyuk, 2006; ten Raa, 2011; Wang et al., 2017; Walheer, 2018, 2019).

Zelenyuk (2006) proposed two aggregations of MPIs that resemble the weighted harmonic-type mean and the weighted geometric mean of efficiencies, accounting for the economic importance of individuals. Until now their statistical properties have not been unveiled, preventing applied researchers from making statistical inferences such as confidence intervals and hypothesis testing which are always in demand in practice. Following Kneip et al. (2015); Simar and Zelenyuk (2018a); Kneip et al. (2018) (hereafter KSW2015, SZ, KSW2018, respectively), in this paper we will fill this gap by developing a comprehensive asymptotic theory for the weighted harmonic-type mean aggregation of MPIs in two contexts: (i) individual efficiency scores are observable, and (ii) individual efficiency scores are non-observable and estimated via DEA relative to the conical hulls of the production technology sets. These new developments will enable applied researchers to obtain more meaningful statistical inferences on aggregate productivity change measured by MPI.[4]

It is important to note that the complexity of the used statistic will show that the traditional delta method on which KSW2015, SZ, KSW2018 were based is not sufficient, so we will need to refer to a uniform version of this method. This approach is a novel one compared to the previous works of KSW2015, SZ, KSW2018 and opens the path for

[4]A similar theory can be developed for other indices using the same methodology introduced in this paper.

deriving asymptotic properties of a variety of sophisticated indices such as the weighted geometric mean aggregation of MPIs and the Hicks-Moorsteen productivity index. We also conduct Monte Carlo experiments to verify the performance of the newly developed statistical inferences in finite samples.

2 Preliminaries

2.1 Foundations

We denote inputs and outputs by column vectors $x \in \mathbb{R}^p_+$ and $y \in \mathbb{R}^q_+$, respectively. The production technology set at time t is defined as

$$\Psi^t = \{(x, y) \in \mathbb{R}^p_+ \times \mathbb{R}^q_+ : x \text{ can produce } y \text{ at time } t\},\ t = 1, 2. \tag{1}$$

For each $t = 1, 2$, Ψ^t is assumed to satisfy several common regularity assumptions as indicated below (see, e.g., Shephard, 1970; Färe and Primont, 1995, for details).

Assumption 1. *Ψ^t is closed and strictly convex.*

Assumption 2. *[No free lunch]* $(0, y) \notin \Psi^t\ \forall y \geq 0$.[5]

Assumption 3. *[Strong disposability of inputs and outputs] If $(x, y) \in \Psi^t$ then $(x^*, y^*) \in \Psi^t$ for $x^* \geqq x$, $y^* \leqq y$.*

To date, a number of efficiency measures have been proposed to evaluate the performance of a particular DMU relative to a production technology set, and the Farrell-type efficiency measures appear to be the most popular in the literature (Farrell, 1957). They are also an important component in the decomposition of profit efficiency employed in studies on efficiency behavior of firms (Färe et al., 2019). Apart from this, the hyperbolic efficiency measure (Färe et al., 1985) is also appealing as it seeks to decrease input quantities and increase output quantities simultaneously and equiproportionally.

[5]For $a, b \in \mathbb{R}^m$, "$a \geqq b$" or "$b \leqq a$" means $a - b \in \mathbb{R}^m_+$, "$a \geq b$" or "$b \leq a$" means $a - b \in \mathbb{R}^m_+ \backslash \{0_m\}$, "$a > b$" or "$b < a$" means $a - b \in \mathbb{R}^m_{++}$.

Formally, efficiency of a DMU represented by an input-output combination $z = (x, y)$ can be evaluated relative to the production technology set Ψ^t via these measures as follows.

- Farrell-type output-oriented efficiency measure:

$$\lambda(z|\Psi^t) = \lambda(x, y|\Psi^t) = \sup_{\lambda}\{\lambda : (x, \lambda y) \in \Psi^t\}. \tag{2}$$

- Hyperbolic efficiency measure:

$$\gamma(z|\Psi^t) = \gamma(x, y|\Psi^t) = \inf_{\gamma}\{\gamma : (\gamma x, \gamma^{-1} y) \in \Psi^t\}. \tag{3}$$

By construction, for all $z = (x, y) \in \Psi^t$, we have $\lambda(z|\Psi^t) \geq 1$ and $0 \leq \gamma(z|\Psi^t) \leq 1$. Now we define the conical hull of the set Ψ^t as[6]

$$\mathcal{C}(\Psi^t) = \{(ax, ay) : (x, y) \in \Psi^t, a \in \mathbb{R}^1_+\}. \tag{4}$$

Obviously, $\Psi^t \subseteq \mathcal{C}(\Psi^t)$. Conventionally, Ψ^t is said to exhibit globally constant returns to scale (CRS) if $\Psi^t = \mathcal{C}(\Psi^t)$, and variable returns to scale (VRS) otherwise (which means Ψ^t might exhibit increasing, constant, or decreasing returns to scale in some local regions).

The conical Farrell-type output-oriented efficiency measure, denoted by λ_C, is defined as

$$\lambda_C(z|\Psi^t) = \lambda_C(x, y|\Psi^t) = \lambda(x, y|\mathcal{C}(\Psi^t)) = \sup_{\lambda}\{\lambda : (x, \lambda y) \in \mathcal{C}(\Psi^t)\}, \tag{5}$$

and γ_C can also be defined in a way similar to λ_C.

Now suppose that the input-output combinations of the interested DMU observed in periods 1 and 2 are $z^1, z^2 \in \mathbb{R}^p_+ \times \mathbb{R}^q_+$, respectively. Then the output-oriented conical

[6]For related discussion, see Zelenyuk (2014) who called it a conical closure of Ψ^t and considered its scale-homothetic decompositions.

MPIs which measure the productivity change of this DMU from period 1 to period 2 can be defined as follows.

$$M_O(z^1, z^2) = \left(\frac{\lambda_C(z^2|\Psi^1)}{\lambda_C(z^1|\Psi^1)} \times \frac{\lambda_C(z^2|\Psi^2)}{\lambda_C(z^1|\Psi^2)} \right)^{-1/2}. \tag{6}$$

It is clear that $M_O(z^1, z^2)$ take values in $(0, \infty)$. In particular, values in $(1, \infty)$, $\{1\}$, and $(0, 1)$ indicate that the productivity of firm i has improved, remained constant, and deteriorated from period 1 to period 2, respectively.

Similar to KSW2018, here we would like to emphasize the importance of measuring efficiency toward the conical hulls of the production technology sets rather than the sets themselves. On the one hand, Grifell-Tatjé and Lovell (1995) indicate that under VRS, the MPI does not account for productivity change accurately due to a systematic bias. In addition, Ray and Desli (1997), when analyzing productivity changes of 17 OECD countries, noted that computing MPI for some countries under VRS might be infeasible. On the other hand, it might be too restrictive to impose CRS on the production technology as discussed in, e.g., KSW2018. Interestingly, using the conical hull of the production technology set can solve this dilemma since measuring MPI relative to the conical hull is always feasible and furthermore, this approach allows the true production technology set to exhibit VRS rather than the more restricted CRS.

Therefore, we will focus on the conical Farrell-type output-oriented MPI in this paper while noting that similar results for the other orientations (e.g., input-oriented) can be developed analogously.

2.2 Aggregations of MPIs

Now we consider a sample $\mathcal{X}_n = \mathcal{X}_n^1 \cup \mathcal{X}_n^2$ consisting of n DMUs observed in period 1 (i.e., $\mathcal{X}_n^1$) and period 2 (i.e., $\mathcal{X}_n^2$). More precisely, for each $t = 1, 2$, $\mathcal{X}_n^t = \{Z_i^t\}_{i=1}^n$, where $Z_i^t = (X_i^t, Y_i^t)$, X_i^t and Y_i^t are column vectors of inputs and outputs of DMU i $(i = 1, \ldots, n)$, respectively.

A common approach to aggregate individual MPIs is to use the equally-weighted

arithmetic or geometric mean, where the latter seems to dominate the former due to the multiplicative essence of MPI. These types of aggregations treat individual DMUs equally, and hence ignore their relative economic importance (e.g., market share), which strongly motivates us to investigate the aggregate MPIs.

We assume that all DMUs in the same time period face common output prices (i.e., "law of one price"). The statistical theory developed here still applies to the context where DMUs face different prices and the assumption of the 'law of one price' (or common equilibrium) is required to maintain the Koopmans-type theorem of aggregation upon which the theory of aggregate efficiency and productivity are built.

Here we develop statistical theory for the following aggregate MPI:

$$\overline{M} = \left(\frac{\sum_{i=1}^{n} \beta_i^2 \lambda_C(Z_i^2|\Psi^1)}{\sum_{i=1}^{n} \beta_i^1 \lambda_C(Z_i^1|\Psi^1)} \times \frac{\sum_{i=1}^{n} \beta_i^2 \lambda_C(Z_i^2|\Psi^2)}{\sum_{i=1}^{n} \beta_i^1 \lambda_C(Z_i^1|\Psi^2)} \right)^{-1/2}, \tag{7}$$

where $\beta_i^s = \frac{w^s Y_i^s}{w^s \sum_{i=1}^{n} Y_i^s}$ $(i = 1, \ldots, n)$ are economic weights, $w^s \in \mathbb{R}_{++}^q$ are the row vector of output prices in the period s $(s = 1, 2)$ (for details of this measures, see Zelenyuk, 2006).

The above aggregation appears to be more appropriate than the conventional arithmetic and geometric means as the weight β_i^s represents the output share of DMU i in the period s and hence, reflects its relative economic importance in the aggregate indices. Zelenyuk (2006) argued that these weights are "not *ad hoc* but are derived from economic principles (agents' optimization behavior)" and using $\overline{M}$ enables the group revenue analog of the MPI to be decomposed in the same way as for the individual revenue analog of the MPI. Empirical applications of the system of weights β_i^s as well as the aggregation $\overline{M}$ can be found in a number of studies, e.g., Pilyavsky and Staat (2008); Gitto and Mancuso (2015).

3 Asymptotic theory when the true efficiency is known

In this section, we will develop central limit theorems in relation to $\overline{M}$ by assuming that the efficiency functions $\lambda_C(\cdot|\Psi^1)$ and $\lambda_C(\cdot|\Psi^2)$ are known. Deriving asymptotic

theories under this assumption is important since it enlightens the statistical essence of the aggregation form $\overline{M}$ and helps identify the underlying parameters these estimators gauge. Moreover, the results derived here also provide a statistical grounding for subsequent sections where we will develop asymptotic theories in the absence of the knowledge of $\lambda_C(\cdot|\Psi^1)$ and $\lambda_C(\cdot|\Psi^2)$ as usually happens in reality.

To begin with, we consider the log version of $\overline{M}$ as follows.

$$\log \overline{M} = -\frac{1}{2}\left[\log\left(\sum_{i=1}^{n}\beta_i^2\lambda_C(Z_i^2|\Psi^1)\right) + \log\left(\sum_{i=1}^{n}\beta_i^2\lambda_C(Z_i^2|\Psi^2)\right) - \log\left(\sum_{i=1}^{n}\beta_i^1\lambda_C(Z_i^1|\Psi^1)\right) - \log\left(\sum_{i=1}^{n}\beta_i^1\lambda_C(Z_i^1|\Psi^2)\right)\right]. \tag{8}$$

For $i = 1, \ldots, n$, let

$$U_{1,i} = \lambda_C(Z_i^2|\Psi^1)w^2Y_i^2,\ U_{2,i} = \lambda_C(Z_i^2|\Psi^2)w^2Y_i^2,\ U_{3,i} = \lambda_C(Z_i^1|\Psi^1)w^1Y_i^1,$$
$$U_{4,i} = \lambda_C(Z_i^1|\Psi^2)w^1Y_i^1, U_{5,i} = w^2Y_i^2,\ U_{6,i} = w^1Y_i^1. \tag{9}$$

Clearly, $U_{s,i}$ are scalar-valued random variables for all $s = 1, 2, \ldots, 6$ and $i = 1, \ldots, n$. Denote $\mu_s = E(U_{s,i})$ and $\widehat{\mu}_{s,n} = n^{-1}\sum_{i=1}^{n} U_{s,i}$ $(s = 1, \ldots, 6)$. Similar to SZ, we have

$$\sum_{i=1}^{n}\beta_i^2\lambda_C(Z_i^2|\Psi^1) = \sum_{i=1}^{n}\frac{w^2Y_i^2}{w^2\sum_{i=1}^{n}Y_i^2}\ \lambda_C(Z_i^2|\Psi^1) = \frac{\sum_{i=1}^{n}\lambda_C(Z_i^2|\Psi^1)w^2Y_i^2}{\sum_{i=1}^{n}w^2Y_i^2} = \frac{\sum_{i=1}^{n}U_{1,i}}{\sum_{i=1}^{n}U_{5,i}} = \frac{\widehat{\mu}_{1,n}}{\widehat{\mu}_{5,n}}.$$

Analogously,

$$\sum_{i=1}^{n}\beta_i^2\lambda_C(Z_i^2|\Psi^2) = \frac{\widehat{\mu}_{2,n}}{\widehat{\mu}_{5,n}}, \quad \sum_{i=1}^{n}\beta_i^1\lambda_C(Z_i^1|\Psi^1) = \frac{\widehat{\mu}_{3,n}}{\widehat{\mu}_{6,n}}, \quad \sum_{i=1}^{n}\beta_i^1\lambda_C(Z_i^1|\Psi^2) = \frac{\widehat{\mu}_{4,n}}{\widehat{\mu}_{6,n}}.$$

Consequently, the aforementioned aggregation of MPIs can be expressed as

$$\log \overline{M} = \frac{1}{2}\left(\log\widehat{\mu}_{3,n} + \log\widehat{\mu}_{4,n} - \log\widehat{\mu}_{1,n} - \log\widehat{\mu}_{2,n}\right) + \log\widehat{\mu}_{5,n} - \log\widehat{\mu}_{6,n}. \tag{10}$$

Therefore, it turns out that $\log \overline{M}$ is a point estimate of the following parameter:

$$\xi = \frac{1}{2}\left(\log \mu_3 + \log \mu_4 - \log \mu_1 - \log \mu_2\right) + \log \mu_5 - \log \mu_6, \tag{11}$$

As a consequence, $\overline{M}$ is a point estimate of $\exp(\xi)$. For this reason, hereafter we will write $\log \overline{M}$ as $\widehat{\xi}_n$, and develop statistical inferences for the parameters of interest ξ. Prior to doing so, we need an additional mild assumption similar to an assumption in SZ.

Assumption 4. *The first two moments of $w^1 Y_i^1$ and $w^2 Y_i^2$ are finite for all $i = 1, \ldots, n$.*

Now we consider i.i.d. random variables

$$T_i = [U_{1,i}, U_{2,i}, U_{3,i}, U_{4,i}, U_{5,i}, U_{6,i}]' \quad (i = 1, \ldots, n) \tag{12}$$

and denote their means and variances by

$$\mu = E(T_i) = [\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6]', \quad \Sigma = Var(T_i) = \left[\sigma_{jk}\right]_{j,k \in \{1,2,3,4,5,6\}}. \tag{13}$$

By standard central limit theorem (see, e.g., Van der Vaart, 2000, page 16), we have

$$\sqrt{n}(\widehat{\mu}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma). \tag{14}$$

where $\widehat{\mu}_n = [\widehat{\mu}_{1,n}, \widehat{\mu}_{2,n}, \widehat{\mu}_{3,n}, \widehat{\mu}_{4,n}, \widehat{\mu}_{5,n}, \widehat{\mu}_{6,n}]'$.

Define the function $\phi : (0, \infty)^6 \longrightarrow \mathbb{R}^1$ as

$$\phi(\eta_1, \eta_2, \eta_3, \eta_4, \eta_5, \eta_6) = \frac{1}{2}\left(\log \eta_3 + \log \eta_4 - \log \eta_1 - \log \eta_2\right) + \log \eta_5 - \log \eta_6. \tag{15}$$

Then under standard regularity conditions, we can apply the delta method (see, e.g., Theorem 3.1 of Van der Vaart, 2000) to (14) and obtain

$$\sqrt{n}(\phi(\widehat{\mu}_n) - \phi(\mu)) \xrightarrow{d} \mathcal{N}(0, [\nabla\phi(\mu)]'\Sigma[\nabla\phi(\mu)]), \tag{16}$$

or equivalently,

$$\sqrt{n}(\widehat{\xi}_n - \xi) \xrightarrow{d} \mathcal{N}(0, V_\xi), \tag{17}$$

where

$$V_\xi = nVAR(\widehat{\xi}_n) = [\nabla\phi(\mu)]'\Sigma[\nabla\phi(\mu)], \tag{18}$$

and $\nabla\phi(\cdot)$ is the column vector of the gradient of $\phi(\cdot)$. Formally, $\nabla\phi(\mu) = \left[\frac{\partial\phi}{\partial\eta_j}(\mu)\right]$, where

$$\begin{aligned} &\frac{\partial\phi}{\partial\eta_1}(\mu) = -\frac{1}{2\mu_1}, \quad \frac{\partial\phi}{\partial\eta_2}(\mu) = -\frac{1}{2\mu_2}, \quad \frac{\partial\phi}{\partial\eta_3}(\mu) = \frac{1}{2\mu_3}, \\ &\frac{\partial\phi}{\partial\eta_4}(\mu) = \frac{1}{2\mu_4}, \quad \frac{\partial\phi}{\partial\eta_5}(\mu) = \frac{1}{\mu_5}, \quad \frac{\partial\phi}{\partial\eta_6}(\mu) = -\frac{1}{\mu_6}. \end{aligned} \tag{19}$$

Evidently, $U_{5,i}$ and $U_{6,i}$ are observable for all $i = 1, \ldots, n$. Under the assumption that the functions $\lambda_C(\cdot|\Psi^1)$ and $\lambda_C(\cdot|\Psi^2)$ are known, $U_{s,i}$ $(s = 1, \ldots, 4; i = 1, \ldots, n)$ are also observable and so is $\widehat{\xi}_n$. Let $\widehat{V}_{\xi,n}$ denote the empirical version of V_ξ, where μ_s is replaced by $\widehat{\mu}_{s,n}$ and σ_{jk} is replaced by $\widehat{\sigma}_{jk,n}$ with

$$\widehat{\sigma}_{jk,n} = n^{-1}\sum_{i=1}^{n}(U_{j,i} - \widehat{\mu}_{j,n})(U_{k,i} - \widehat{\mu}_{k,n}), \quad j, k = 1, 2, \ldots, 6. \tag{20}$$

It is well-known that $\widehat{\mu}_{s,n} \xrightarrow{p} \mu_s$ and $\widehat{\sigma}_{jk,n} \xrightarrow{p} \sigma_{jk}$ $(s, j, k = 1, \ldots, 6)$ as $n \to \infty$. Hence, by Slutsky's theorem and continuous mapping theorem (see, e.g., Van der Vaart, 2000, page 7), we have $\widehat{V}_{\xi,n} \xrightarrow{p} V_\xi$. Combining these results with (17) and Slutsky's theorem, we can obtain

$$\sqrt{n}\left(\frac{\widehat{\xi}_n - \xi}{\sqrt{\widehat{V}_{\xi,n}}}\right) \xrightarrow{d} \mathcal{N}(0, 1), \tag{21}$$

which in turn can be used to make inferences about ξ. In particular, when the true efficiency is known, asymptotic $100(1-\alpha)\%$ symmetric confidence interval for ξ is given

by

$$\left[\widehat{\xi}_n \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\widehat{V}_{\xi,n}/n}\ \right], \tag{22}$$

where $\Phi^{-1}_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

4 DEA estimators from the statistical viewpoint

The technology sets Ψ^1, Ψ^2 as well as efficiency measures and indices are unobserved in reality and must be estimated from data, raising the need for respective statistical inferences. Among estimation methodologies to date, DEA has emerged as one of the most popular, attracting numerous theoretical and empirical works.[7] More specifically, DEA has also been the mostly used method to compute MPI since the inspirational works of Färe et al. (1990); Färe et al. (1992); Färe et al. (1994a) (see, e.g., Färe et al., 1998, for a survey on MPI). Prior to presenting the estimation details and setting up a statistical model for the DEA-based MPI, we need additional assumptions as in KSW2018. More precisely, they are Assumptions 2.4-2.7, 3.1 and 3.2 in KSW2018 and are listed in Appendix A of this article as Assumptions 5-10.

By virtue of these assumptions, the sample $\mathcal{X}_n$ can be viewed as a random set generated from Ψ^1 and Ψ^2. For $t = 1, 2$, Ψ^t and $\mathcal{C}(\Psi^t)$ can be estimated via DEA-type estimators as follows.

$$\widehat{\Psi}^t_n = \left\{(x,y) \in \mathbb{R}^p_+ \times \mathbb{R}^q_+ : x \geqq \sum_{i=1}^n X^t_i \zeta_i, y \leqq \sum_{i=1}^n Y^t_i \zeta_i, \sum_{i=1}^n \zeta_i = 1, \zeta_i \in \mathbb{R}^1_+,\ \forall i = 1, \ldots, n\right\}, \tag{23}$$

$$\mathcal{C}(\widehat{\Psi}^t_n) = \left\{(x,y) \in \mathbb{R}^p_+ \times \mathbb{R}^q_+ : x \geqq \sum_{i=1}^n X^t_i \zeta_i, y \leqq \sum_{i=1}^n Y^t_i \zeta_i, \zeta_i \in \mathbb{R}^1_+,\ \forall i = 1, \ldots, n\right\}. \tag{24}$$

Substituting Ψ^t by $\widehat{\Psi}^t_n$ in (2) and (5) gives the DEA-type estimators of the Farrell-type

[7] Another popular approach to measuring performance is Stochastic Frontier Analysis (SFA). Olesen and Petersen (2016) provided a thorough review of Stochastic DEA, an extension of deterministic DEA.

output-oriented efficiency measure and its conical hull version:

$$\lambda(z|\widehat{\Psi}_n^t) = \lambda(x,y|\widehat{\Psi}_n^t) = \sup_{\lambda}\left\{\lambda : x \geqq \sum_{i=1}^{n} X_i^t\zeta_i, \lambda y \leqq \sum_{i=1}^{n} Y_i^t\zeta_i, \sum_{i=1}^{n}\zeta_i = 1, \zeta_i \in \mathbb{R}_+^1,\ \forall i = 1,\ldots,n\right\},$$
$$\lambda_C(z|\widehat{\Psi}_n^t) = \lambda_C(x,y|\widehat{\Psi}_n^t) = \sup_{\lambda}\left\{\lambda : x \geqq \sum_{i=1}^{n} X_i^t\zeta_i, \lambda y \leqq \sum_{i=1}^{n} Y_i^t\zeta_i, \zeta_i \in \mathbb{R}_+^1,\ \forall i = 1,\ldots,n\right\}, \tag{25}$$

respectively. Note that since $\widehat{\Psi}_n^t$ is in turn determined by $\mathcal{X}_n^t$, henceforward in this paper we will use the notations $\lambda(z|\mathcal{X}_n^t)$ and $\lambda_C(z|\mathcal{X}_n^t)$ instead of $\lambda(z|\widehat{\Psi}_n^t)$ and $\lambda_C(z|\widehat{\Psi}_n^t)$, respectively, to emphasize the dataset from which the estimators are computed.

The aforementioned aggregate MPI can be estimated via DEA as follows.

$$\widehat{\overline{M}}(\mathcal{X}_n) = \left(\frac{\sum_{i=1}^{n}\beta_i^2\lambda_C(Z_i^2|\mathcal{X}_n^1)}{\sum_{i=1}^{n}\beta_i^1\lambda_C(Z_i^1|\mathcal{X}_n^1)} \times \frac{\sum_{i=1}^{n}\beta_i^2\lambda_C(Z_i^2|\mathcal{X}_n^2)}{\sum_{i=1}^{n}\beta_i^1\lambda_C(Z_i^1|\mathcal{X}_n^2)}\right)^{-1/2}. \tag{26}$$

KSW2015 were the first to uncover the asymptotic properties of the Farrell-type technical efficiency evaluated at random points. In addition, they proposed important central limit theorems for the arithmetic mean of the efficiency of random samples, enabling researchers to make inference about the efficiency of groups of firms. Lately, KSW2018 extended the work of KSW2015 to the dynamic context and provided the following result.

Lemma 1. *Under Assumptions 1-3 and 5-10, as* $n \to \infty$,

$$E\left(\log\gamma_C(Z_i^s|\mathcal{X}_n^t) - \log\gamma_C(Z_i^s|\Psi^t)\right) = \widetilde{C}_{st}n^{-\kappa} + R_{n,\kappa}, \tag{27}$$
$$E\left(\left[\log\gamma_C(Z_i^s|\mathcal{X}_n^t) - \log\gamma_C(Z_i^s|\Psi^t)\right]^2\right) = o(n^{-\kappa}), \tag{28}$$
$$\Big|E\left([\log\gamma_C(Z_i^s|\mathcal{X}_n^t) - E(\log\gamma_C(Z_i^s|\mathcal{X}_n^t))]\times\right.$$
$$\left.\times[\log\gamma_C(Z_j^{s^*}|\mathcal{X}_n^{t^*}) - E(\log\gamma_C(Z_j^{s^*}|\mathcal{X}_n^{t^*}))]\right)\Big| = o(n^{-1}), \tag{29}$$

for all $i,j \in \{1,\ldots,n\}, i \neq j$; $s,t,s^*,t^* \in \{1,2\}$, *where* $\kappa = 2/(p+q+1)$, $\widetilde{C}_{st}$ *is a constant,* $R_{n,\kappa}$ *is a remainder of order smaller than* $n^{-\kappa}$.

The above results pave the way for statistical inference in a dynamic context. Par-

ticularly, KSW2018 introduced asymptotic properties for DEA-based estimators of MPIs for individual DMUs as well as their geometric mean. Our goal is to generalize their results to the case where researchers want to account for the economic weights in the aggregation.

5 Asymptotic properties when the true efficiency is unknown

In reality the functions $\lambda_C(\cdot|\Psi^1)$ and $\lambda_C(\cdot|\Psi^2)$ are unknown and hence, $U_{s,i}$ $(s = 1, \ldots, 4; i = 1, \ldots, n)$ are not observable. As a result, $\widehat{\xi}_n$ is an infeasible estimator, and hence, it is impossible to obtain confidence interval (22) in practice. Therefore, we relax the assumption of knowledge of $\lambda_C(\cdot|\Psi^1)$ and $\lambda_C(\cdot|\Psi^2)$, and develop some new central limit theorems that allow researchers to make statistical inferences about ξ based on feasible DEA-type estimators.

For $i = 1, \ldots, n$, let

$$\widehat{U}_{1,i} = \lambda_C(Z_i^2|\mathcal{X}_n^1)w^2Y_i^2,\ \widehat{U}_{2,i} = \lambda_C(Z_i^2|\mathcal{X}_n^2)w^2Y_i^2,$$
$$\widehat{U}_{3,i} = \lambda_C(Z_i^1|\mathcal{X}_n^1)w^1Y_i^1,\ \widehat{U}_{4,i} = \lambda_C(Z_i^1|\mathcal{X}_n^2)w^1Y_i^1,$$

where we recall that $\lambda_C(.|.)$ was defined before in (25). Next, let

$$\widehat{\widehat{\mu}}_{s,n} = n^{-1}\sum_{i=1}^{n}\widehat{U}_{s,i}, \quad s = 1, \ldots, 4. \tag{30}$$

Since $\widehat{\widehat{\mu}}_{s,n}$ $(s = 1, \ldots, 4)$, $\widehat{\mu}_{5,n}$ and $\widehat{\mu}_{6,n}$ can be computed from the data, the estimators

$$\widehat{\widehat{\xi}}_n = \log\left(\widehat{\overline{M}}(\mathcal{X}_n)\right) = \frac{1}{2}\left(\log\widehat{\widehat{\mu}}_{3,n} + \log\widehat{\widehat{\mu}}_{4,n} - \log\widehat{\widehat{\mu}}_{1,n} - \log\widehat{\widehat{\mu}}_{2,n}\right) + \log\widehat{\mu}_{5,n} - \log\widehat{\mu}_{6,n}$$

are feasible. Hence, we will develop asymptotic theories for $\widehat{\widehat{\xi}}_n$ in order to make feasible statistical inferences for our parameters of interest that are ξ and $\exp(\xi)$.

Let $\kappa = 2/(p+q+1)$ and define a sequence of positive real numbers $\{v_{n,\kappa}\}_{n=1}^{\infty}$ by

$$v_{n,\kappa} = \left(\frac{\log n}{n}\right)^{\frac{3}{p+q+1}}. \tag{31}$$

Theorem 1 below establishes basic asymptotic properties of moments of $\widehat{U}_{s,i}$ for $s = 1, \ldots, 4$ and $i = 1, \ldots, n$.[8]

Theorem 1. *Under Assumptions 1-10, there exist constants $C_s \in (0, \infty)$ such that as $n \to \infty$,*

$$E(\widehat{U}_{s,i} - U_{s,i}) = C_s n^{-\kappa} + O(v_{n,\kappa}), \tag{32}$$

$$E\left([\widehat{U}_{s,i} - U_{s,i}]^2\right) = o(n^{-\kappa}), \tag{33}$$

$$\left|E\left([\widehat{U}_{s,i} - E(\widehat{U}_{s,i})][\widehat{U}_{t,j} - E(\widehat{U}_{t,j})]\right)\right| = o(n^{-1}), \tag{34}$$

for all $i, j \in \{1, \ldots, n\}, i \neq j$; $s, t \in \{1, 2, \ldots, 4\}$.

The following theorem develops upon Theorem 1 and provides essential tools for deriving asymptotic properties of $\widehat{\mu}_{s,n}$ $(s = 1, \ldots, 4)$ in the later stage.

Theorem 2. *Under Assumptions 1-10, as $n \to \infty$,*

$$\text{(i) } E(\widehat{U}_{s,i}) = \mu_s + C_s n^{-\kappa} + O(v_{n,\kappa}) \tag{35}$$

$$\text{(ii) } Cov(\widehat{U}_{t,i}, \widehat{U}_{t^*,i}) = \sigma_{tt^*} + o(n^{-\kappa/2}) \tag{36}$$

$$\text{(iii) } Cov(\widehat{U}_{s,i}, U_{r,i}) = \sigma_{sr} + o(n^{-\kappa/2}) \tag{37}$$

for all $i \in \{1, \ldots, n\}$, $s, t, t^* \in \{1, \ldots, 4\}$, $r \in \{5, 6\}$, C_s are the same constants as in Theorem 1.

It should be noted that Theorem 2 is more comprehensive than Lemma 1 of SZ since it encompasses the asymptotic property of the covariance of two estimators containing

[8] Theorem 1 in this paper is an analogue of Theorem 3.1 of KSW2015, Corollary 1 of SZ, and Theorem 3.4 of KSW2018.

efficiency scores (i.e., $Cov(\widehat{U}_{t,i}, \widehat{U}_{t^*,i})$ in Theorem 2(ii)), which we have solved by decomposing the covariance into four components and examining the asymptotic behaviors of each one separately.

Let $\widetilde{\mu}_{s,n} = E(\widehat{\widehat{\mu}}_{s,n})$ for $s = 1, \ldots, 4$. The following theorem presents important properties of $\widehat{\widehat{\mu}}_{s,n}$ and provides consistent estimators of σ_{st}.[9]

Theorem 3. *Under Assumptions 1-10, as* $n \to \infty$,

$$\text{(i) } \widetilde{\mu}_{s,n} = \mu_s + C_s n^{-\kappa} + O(v_{n,\kappa}) \tag{38}$$

$$\text{(ii) } \widehat{\widehat{\mu}}_{s,n} - \widetilde{\mu}_{s,n} = \widehat{\mu}_{s,n} - \mu_s + o_p(n^{-1/2}) \tag{39}$$

$$\text{(iii) } \sqrt{n}(\widehat{\widehat{\mu}}_{s,n} - \widetilde{\mu}_{s,n}) \xrightarrow{d} \mathcal{N}(0, \sigma_{ss}) \tag{40}$$

$$\text{(iv) } \widehat{\widehat{\sigma}}_{st,n} = n^{-1} \sum_{i=1}^{n} (\widehat{U}_{s,i} - \widehat{\widehat{\mu}}_{s,n})(\widehat{U}_{t,i} - \widehat{\widehat{\mu}}_{t,n}) \xrightarrow{p} \sigma_{st} \tag{41}$$

$$\text{(v) } \widehat{\widehat{\sigma}}_{sr,n} = n^{-1} \sum_{i=1}^{n} (\widehat{U}_{s,i} - \widehat{\widehat{\mu}}_{s,n})(U_{r,i} - \widehat{\mu}_{r,n}) \xrightarrow{p} \sigma_{sr} \tag{42}$$

$$\text{(vi) } \widehat{\sigma}_{rr^*,n} = n^{-1} \sum_{i=1}^{n} (U_{r,i} - \widehat{\mu}_{r,n})(U_{r^*,i} - \widehat{\mu}_{r^*,n}) \xrightarrow{p} \sigma_{rr^*} \tag{43}$$

for $s, t \in \{1, \ldots, 4\}$, $r, r^* \in \{5, 6\}$.

It should be noted that compared to Theorem 1 of SZ, Theorem 3 includes an additional result about the covariance of random variables containing the efficiency scores (i.e., part (iv)). Theorem 3 allows for the deriving of consistent estimators of V_ξ. Indeed, let $\widehat{\widehat{V}}_{\xi,n}$ be the empirical versions of V_ξ where μ_s is replaced by $\widehat{\widehat{\mu}}_{s,n}$ ($s = 1, \ldots, 4$) and by $\widehat{\mu}_{s,n}$ ($s = 5, 6$), σ_{st} is replaced by $\widehat{\widehat{\sigma}}_{st,n}$ ($s, t \in \{1, \ldots, 4\}$), σ_{sr} is replaced by $\widehat{\widehat{\sigma}}_{sr,n}$ ($s \in \{1, \ldots, 4\}$, $r \in \{5, 6\}$), and σ_{rr^*} is replaced by $\widehat{\sigma}_{rr^*,n}$ ($r, r^* \in \{5, 6\}$). Then $\widehat{\widehat{V}}_{\xi,n}$ is a feasible estimator and its consistency is revealed below.

Theorem 4. *Under Assumptions 1-10, as* $n \to \infty$,

$$\widehat{\widehat{V}}_{\xi,n} \xrightarrow{p} V_\xi. \tag{44}$$

[9] Theorem 3 here is an analogue of Theorem 4.1 of KSW2015 and Theorem 1 of SZ but encompasses an additional case (part (iv)), which corresponds to the covariance between random variables containing efficiency.

Theorem 3 provides a foundation to derive new central limit theorems for our target estimator that is $\widehat{\widehat{\xi}}_n$. However, this task cannot be done trivially as an adaptation of the previous works (KSW2015, SZ, KSW2018) due to the complicated functional form of this estimator. In particular, it involves the nonlinear operator $\log(\cdot)$ on DEA-based components whereas the target estimators in the other works (e.g., SZ) are linear with respect to the DEA-based components. To circumvent this difficulty, we employ a uniform delta method (Theorem 3.8 of Van der Vaart, 2000) which we mention here as Lemma 2 in Appendix T2.2. With Lemma 2, we can prove the asymptotic property of sequences $\sqrt{n}\left(\log\widehat{\widehat{\mu}}_{s,n} - \log\widetilde{\mu}_{s,n}\right)$ $(s = 1, 2, 3, 4)$ where the centering vectors $\widetilde{\mu}_{s,n}$ are dependant on the sample size n, as is the case in our complex statistic (see Remark 1 in Section 8 for more details). Using this approach, we can derive important central limit theorems below.

Theorem 5. *Under Assumptions 1-10, as* $n \to \infty$*,*

$$\sqrt{n}(\widehat{\widehat{\xi}}_n - \xi - C_\xi n^{-\kappa} + O(v_{n,\kappa})) \xrightarrow{d} \mathcal{N}(0, V_\xi), \tag{45}$$

where $C_\xi \in \mathbb{R}$ *is a constant,* $O(v_{n,\kappa}) = o(n^{-\kappa})$*.*

Theorem 5 has implications for the limiting distribution of the DEA-based estimator of ξ. In particular, $\widehat{\widehat{\xi}}_n$ is a consistent estimator of ξ with the leading bias term being $C_\xi n^{-\kappa}$. Moreover, the asymptotic behavior of this bias when multiplied with the norming rate $\sqrt{n}$ is revealed below.

- If $\kappa > 1/2$ $(p + q = 2)$, the bias term in (45) vanishes asymptotically and can be ignored.
- If $\kappa = 1/2$ $(p + q = 3)$, the bias term converges to an unknown constant, implying that Theorem 5 cannot be used immediately to make inferences about ξ.
- If $\kappa < 1/2$ $(p+q = 4, 5, 6, \ldots)$, the bias term explodes to infinity as n increases, and again, Theorem 5 cannot be used directly to make inferences about the parameter of interest.

As a consequence, there emerges a need to correct for the bias term in Theorem 5 in order to make inferences when $\kappa \leq 1/2$. In the spirit of KSW2015, SZ, KSW2018, we find another norming rate different from $\sqrt{n}$ for the case $\kappa \leq 1/2$. In particular, consider the factor $n_\kappa = \min\{\lfloor n^{2\kappa} \rfloor, n\} \leq n$ and the estimators $\widehat{\widehat{\xi}}_{n_\kappa}$ which is the subsample version of $\widehat{\widehat{\xi}}_n$ in the sense that the averages are taken over a random subsample $\mathcal{X}^*_{n_\kappa}$ of size n_κ where $\mathcal{X}^*_{n_\kappa} \subset \mathcal{X}_n$. Formally,

$$\begin{aligned}\widehat{\widehat{\xi}}_{n_\kappa} = \frac{1}{2}\Bigg[&\log\left(\frac{\sum_{\{i:Z_i^1 \in \mathcal{X}^*_{n_\kappa}\}} \lambda_C(Z_i^1|\mathcal{X}_n^1) w^1 Y_i^1}{n_\kappa}\right) + \log\left(\frac{\sum_{\{i:Z_i^1 \in \mathcal{X}^*_{n_\kappa}\}} \lambda_C(Z_i^1|\mathcal{X}_n^2) w^1 Y_i^1}{n_\kappa}\right) \\ &- \log\left(\frac{\sum_{\{i:Z_i^2 \in \mathcal{X}^*_{n_\kappa}\}} \lambda_C(Z_i^2|\mathcal{X}_n^1) w^2 Y_i^2}{n_\kappa}\right) - \log\left(\frac{\sum_{\{i:Z_i^2 \in \mathcal{X}^*_{n_\kappa}\}} \lambda_C(Z_i^2|\mathcal{X}_n^2) w^2 Y_i^2}{n_\kappa}\right)\Bigg] \\ &+ \log\left(\frac{\sum_{\{i:Z_i^2 \in \mathcal{X}^*_{n_\kappa}\}} w^2 Y_i^2}{n_\kappa}\right) - \log\left(\frac{\sum_{\{i:Z_i^1 \in \mathcal{X}^*_{n_\kappa}\}} w^1 Y_i^1}{n_\kappa}\right) \end{aligned} \tag{46}$$

Similar to KSW2015, SZ, KSW2018, we note that although the average in (46) is taken over the subsample $\mathcal{X}^*_{n_\kappa}$, the DEA efficiency scores are still estimated using all of the available observations in the original sample $\mathcal{X}_n$. The asymptotic property of this new estimator is revealed below.

Theorem 6. *Under Assumptions 1-10, when $\kappa \leq 1/2$, as $n \to \infty$,*

$$n^\kappa \left(\widehat{\widehat{\xi}}_{n_\kappa} - \xi - C_\xi n^{-\kappa} + O(v_{n,\kappa})\right) \xrightarrow{d} \mathcal{N}(0, V_\xi), \tag{47}$$

where C_ξ is the same constant as in Theorem 5 and $O(v_{n,\kappa}) = o(n^{-\kappa})$.

By virtue of Theorem 6, when $\kappa < 1/2$, the bias term converges to a constant instead of exploding to infinity as was demonstrated in Theorem 5.[10]

[10]Note that when $\kappa = 1/2$, Theorems 5 and 6 are equivalent.

6 Generalized jackknife estimators for the bias

In the spirit of KSW2015, SZ, KSW2018, we construct a jackknife-type bias estimator on the basis of splitting the original sample into two subsamples.[11] The details are presented below, where one may notice that the adaptation of the previous works in the related literature is non-trivial.

For each $l = 1, \ldots, L$ where $L \ll \binom{n}{\lfloor n/2 \rfloor}$, split $\mathcal{X}_n$ randomly into two subsamples $\mathcal{X}_{l,m_1}$ and $\mathcal{X}_{l,m_2}$ of sizes $m_1 = \lfloor n/2 \rfloor$ and $m_2 = n - m_1$, respectively. More precisely, $\mathcal{X}_n = \mathcal{X}_{l,m_1} \cup \mathcal{X}_{l,m_2}$, where for each $j = 1, 2$, $\mathcal{X}_{l,m_j} = \mathcal{X}^1_{l,m_j} \cup \mathcal{X}^2_{l,m_j}$ consists of sets of the same m_j DMUs observed in periods 1 and 2, i.e., $\mathcal{X}^1_{l,m_j}$ and $\mathcal{X}^2_{l,m_j}$, respectively. Then for each $j = 1, 2$, set

$$\widehat{\widehat{\mu}}_{l,1,m_j} = m_j^{-1} \sum_{\{i: Z_i^2 \in \mathcal{X}^2_{l,m_j}\}} \lambda_C(Z_i^2 | \mathcal{X}^1_{l,m_j}) w^2 Y_i^2, \tag{48}$$

and similarly, define $\widehat{\widehat{\mu}}_{l,s,m_j}$ $(s = 2, 3, 4)$ as the analogues of $\widehat{\widehat{\mu}}_{s,n}$ in the same way as $\widehat{\widehat{\mu}}_{l,1,m_j}$. Here it is important to highlight that unlike $\widehat{\widehat{\xi}}_{n_\kappa}$, the efficiency scores in $\widehat{\widehat{\mu}}_{l,s,m_j}$ are estimated by DEA using only observations in the corresponding subsample $\mathcal{X}_{l,m_j}$. For $j = 1, 2$, we also define

$$\widehat{\widehat{\xi}}_{l,m_j} = \frac{1}{2}\left(\log \widehat{\widehat{\mu}}_{l,3,m_j} + \log \widehat{\widehat{\mu}}_{l,4,m_j} - \log \widehat{\widehat{\mu}}_{l,1,m_j} - \log \widehat{\widehat{\mu}}_{l,2,m_j}\right) + \log \widehat{\mu}_{5,n} - \log \widehat{\mu}_{6,n}. \tag{49}$$

In essence, $\widehat{\widehat{\xi}}_{l,m_j}$ is an analogue of $\widehat{\widehat{\xi}}_n$ in the sense that components containing efficiency scores are evaluated over the subsample $\mathcal{X}_{l,m_j}$ whereas the other components (i.e., $\widehat{\mu}_{5,n}$ and $\widehat{\mu}_{6,n}$) are evaluated over the original full sample $\mathcal{X}_n$.

Now set

$$\widehat{\widehat{\xi}}^*_{l,n} = \frac{1}{2}(\widehat{\widehat{\xi}}_{l,m_1} + \widehat{\widehat{\xi}}_{l,m_2}).$$

[11] For earlier works on the jackknife bias estimators, see, for example, Quenouille (1949, 1956); Tukey (1958); Miller (1974); Efron (1979).

The generalized jackknife estimators of the bias associated with $\widehat{\widehat{\xi}}_n$ is given by

$$\widehat{A}_{\xi,n,\kappa,L} = L^{-1}\sum_{l=1}^{L}(2^{\kappa}-1)^{-1}(\widehat{\widehat{\xi}}^{*}_{l,n} - \widehat{\widehat{\xi}}_n).$$

The following theorem reveals an important asymptotic property of this bias estimator.

Theorem 7. *Under Assumptions 1-10, as $n \to \infty$,*

$$\widehat{A}_{\xi,n,\kappa,L} = C_{\xi}n^{-\kappa} + O(v_{n,\kappa}) + o_p(n^{-1/2}), \tag{50}$$

where C_{ξ} is the same constant as in Theorems 5 and 6, $O(v_{n,\kappa}) = o(n^{-\kappa})$.

It should be emphasized that under Assumptions 1-10, the variance of $\widehat{A}_{\xi,n,\kappa,L}$ is inversely proportional to L^2. Therefore, similar to Kneip et al. (2016), while Theorem 7 holds true even with $L = 1$, applied researchers might want to increase L (e.g., $L = 100$) to reduce the variance of this bias estimator and achieve more reliable bias corrections for a few particular samples in practice.[12]

7 Confidence intervals

As discussed in Section 5, statistical inferences for ξ can be obtained directly from Theorem 5 when $\kappa > 1/2$ (i.e., $p+q = 2$) by ignoring the bias since it disappears asymptotically when multiplied with the norming rate $\sqrt{n}$. However, it might not be ideal to do so in practice since the bias might still be significant in small samples. In light of Theorem 7, we can account for this issue by estimating the leading term of the bias. The following theorem presents important results which pave the way for making inferences about the parameter of interest ξ for all cases of κ.

[12]Following the literature, we set $L = 10$ in our Monte Carlo simulations. A sensitivity check for different choices of L confirmed the robustness of the results.

Theorem 8. *Under Assumptions 1-10, as $n \to \infty$, for $\kappa \geq 1/2$,*

$$\sqrt{n}\left(\widehat{\widehat{\xi}}_n - \xi - \widehat{A}_{\xi,n,\kappa,L} + O(v_{n,\kappa})\right) \xrightarrow{d} \mathcal{N}(0, V_\xi), \tag{51}$$

and for $0 < \kappa < 1/2$,

$$n^\kappa\left(\widehat{\widehat{\xi}}_{n_\kappa} - \xi - \widehat{A}_{\xi,n,\kappa,L} + O(v_{n,\kappa})\right) \xrightarrow{d} \mathcal{N}(0, V_\xi), \tag{52}$$

where $\kappa = 2/(p+q+1)$ and $O(v_{n,\kappa}) = o(n^{-\kappa})$.

In light of Theorems 4, 8 and Slutsky's theorem, feasible confidence intervals for ξ can now be derived straightforwardly with the note that $O(v_{n,\kappa})$ in Theorem 8 can be ignored since $\sqrt{n}O(v_{n,\kappa}) = \sqrt{n}o(n^{-\kappa}) = o(1)$ when $\kappa \geq 1/2$ and $n^\kappa O(v_{n,\kappa}) = n^\kappa o(n^{-\kappa}) = o(1)$ when $\kappa < 1/2$. In particular, asymptotically correct $100(1-\alpha)\%$ symmetric confidence intervals for ξ are given by

$$\left[\widehat{\widehat{\xi}}_n - \widehat{A}_{\xi,n,\kappa,L} \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\widehat{\widehat{V}}_{\xi,n}/n}\ \right], \tag{53}$$

$$\left[\widehat{\widehat{\xi}}_{n_\kappa} - \widehat{A}_{\xi,n,\kappa,L} \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\widehat{\widehat{V}}_{\xi,n}}/n^\kappa\ \right], \tag{54}$$

for $\kappa \geq 1/2$ (i.e., $p+q = 2, 3$) and $\kappa < 1/2$ (i.e., $p+q = 4, 5, 6, \ldots$), respectively. Here we recall $n_\kappa = \lfloor n^{2\kappa} \rfloor$ and $\Phi^{-1}_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

8 Remarks

The following remarks are in order.

Remark 1. The theoretical developments in this paper are substantial (and non-trivial) generalizations of the previous works, particularly SZ. The major difficulty here is that the functional form of the aggregate MPIs (i.e., weighted harmonic-type mean) is more complex than the form for the aggregate Farrell-type efficiency (e.g., the former include nonlinear operators on DEA-based components). As such, arguments for deriving limiting distributions and bias corrections in SZ cannot be adapted to the aggregate

MPIs. We overcome this problem by using the uniform delta method (i.e., Lemma 2 or equivalently, Theorem 3.8 of Van der Vaart (2000)) to express the DEA-based estimators of ξ as sums of: (i) the underlying parameter of interest, (ii) bias term, (iii) a stochastic term of order smaller than $n^{-1/2}$, and (iv) an expression which is linear with respect to DEA-based components (e.g., see equations (86) and (102) in Appendix T1). Based on this, the task of deriving the limiting distributions and the jackknife bias correction can be carried out smoothly. To the best of our knowledge, this approach is novel relative to the previous works.

Remark 2. Similar to KSW2015, SZ, KSW2018, although our theories suggest using two different confidence intervals corresponding to $\kappa \geq 1/2$ and $0 < \kappa < 1/2$, it is worth noting that the former is still valid for $\kappa = 2/5$ (i.e., $p + q = 4$). Indeed, when $\kappa = 2/5$, we have

$$\sqrt{n}O(v_{n,\kappa}) = \sqrt{n}O\left(\left(\frac{\log n}{n}\right)^{3/5}\right) = O\left(\frac{(\log n)^{3/5}}{n^{1/10}}\right) = o(1), \tag{55}$$

$$n^{\kappa}O(v_{n,\kappa}) = n^{2/5}O\left(\left(\frac{\log n}{n}\right)^{3/5}\right) = O\left(\frac{(\log n)^{3/5}}{n^{1/5}}\right) = o(1), \tag{56}$$

and hence, the bias term $O(v_{n,\kappa})$ disappears asymptotically with both norming rates $\sqrt{n}$ and n^{κ} in both confidence intervals. However, one may notice from the above equalities that $n^{\kappa}O(v_{n,\kappa})$ converges to zero faster than $\sqrt{n}O(v_{n,\kappa})$, which fortifies the use of the confidence interval with norming rate n^{κ} for $\kappa = 2/5 < 1/2$ as before. We also check and confirm this remark in our Monte Carlo experiments (Section 9.2).

Remark 3. As noted by KSW2015 and KSW2018, when $0 < \kappa < 1/2$, one can obtain more informative confidence intervals by employing a re-centering technique. To do so, one needs to replace the point estimator using only n_{κ} DMUs (i.e., $\widehat{\widehat{\xi}}_{n_{\kappa}}$) by its analogue evaluated over the full original sample (i.e., $\widehat{\widehat{\xi}}_{n}$). The re-centered version of confidence interval (54) is as follows.

$$\left[\widehat{\widehat{\xi}}_{n} - \widehat{A}_{\xi,n,\kappa,L} \pm \Phi^{-1}_{1-\alpha/2}\sqrt{\widehat{\widehat{V}}_{\xi,n}}/n^{\kappa}\right]. \tag{57}$$

Similar to KSW2015 and KSW2018, this re-centering technique helps average the center over all possible draws (without replacement) of subsamples of size n_κ, and hence, eliminates the randomness as well as any deviation due to calculation on only a random subset of n_κ DMUs. In fact, one can see, for example, that $\widehat{\widehat{\xi}}_n$ is a better estimator of ξ than $\widehat{\widehat{\xi}}_{n_\kappa}$ in terms of mean-square error. Moreover, the coverage of the re-centered confidence interval converges to 1 as $n \to \infty$, exhibiting superior performance while having the same width as the respective origin whose coverage converges to $1 - \alpha$ as $n \to \infty$.[13] This is confirmed by the Monte Carlo evidence in Section 9.2.

Remark 4. It is possible that the bias cancels out in some certain situations, e.g., $C_\xi = \frac{1}{2}\left(\frac{C_3}{\mu_3} + \frac{C_4}{\mu_4} - \frac{C_1}{\mu_1} - \frac{C_2}{\mu_2}\right) = 0$. As a result, statistical inferences using a naive application of the standard central limit theorem (i.e., using the confidence interval (22)) with unobserved elements being replaced by their respective DEA estimates) might still be valid under these circumstances.

KSW2018 mentioned the possibility of having no bias in the simple Malmquist indices, and show that under very peculiar (and may be unrealistic) assumptions on the data generating process, this bias may be equal to zero (see KSW2018, Theorem 3.6 and Remark 3.2). In particular, some necessary (but not sufficient) conditions for the bias to cancel out are: (i) two samples are generated from identical data generating processes, (ii) the joint density of input-output bundles in two time periods is symmetric.

Remark 5. We elaborate on **Remark 4** and note that even if the bias does not cancel out, in some peculiar cases it might still be so tiny that its theoretical explosion to infinity implied from Theorem 5 is not clearly observed in moderate sample sizes and low dimensions (i.e., the number of inputs and outputs). As such, the naive application of the standard central limit theorem might also perform fairly well in this context. For example, in Theorem 5, if $C_\xi = 0.01$ and $\kappa = 1/3$ (i.e., $p + q = 5$), then the leading bias term $C_\xi n^{-\kappa}$ multiplied with the norming rate $\sqrt{n}$ will be equal to $0.031, 0.046, 0.068$ for $n = 1000, 10000, 100000$, respectively.

Furthermore, it can be seen that the coefficient of the leading bias term in this paper

[13] In other words, the re-centered confidence intervals overcover when n is sufficiently large.

is obtained from subtracting symmetric terms. More precisely, in C_ξ, the constant C_s ($s = 1, 2, 3, 4$) is normalized by the respective population means μ_s (i.e., $\frac{C_s}{\mu_s}$). Thus, in practice the chance that the bias coefficient C_ξ is close to zero (or even cancels out), making the whole bias negligible in moderate sample sizes and low dimensions, might be relatively higher than that for KSW2015, SZ, KSW2018.

Since the magnitude of the bias is unknown in reality, the naive application of the standard central limit theorems in dimensions greater than two (i.e., $\kappa \leq 1/2$) is problematic as pointed out in Theorem 5. As such, the use of the theoretically justified theorems appears to be a safer choice.

9 Monte Carlo evidence

We conduct Monte Carlo experiments to investigate the finite-sample performance of the statistical inferences proposed in this paper, focusing on confidence intervals and hypothesis testing. Regarding the confidence intervals, we analyze their estimated coverages which are calculated as the percentage of times an estimated confidence interval covers the true value of the parameter of interest, given the pre-determined level of significance α. For hypothesis testing, we examine the size and power of tests of productivity change from period 1 to period 2 via rates of rejection of the null hypothesis. In particular, the null hypothesis of no productivity change corresponds to $\xi = 0$, depending on which estimator is used to conduct the test. Meanwhile, the alternative hypothesis that the productivity has changed from period 1 to period 2 corresponds to $\xi \neq 0$. The rejection rates are computed as the percentage of times an estimated confidence interval does not cover zero.

9.1 A data generating process

In this paper, we employ a new data generating process that possesses the essential properties needed for this type of experiment, typically: (i) the technologies in both periods exhibit VRS; (ii) the productivity change between two periods can be controlled

via a pre-determined parameter, say δ; (iii) the technical efficiency scores of any particular DMU in two time periods are correlated and can be controlled via a pre-determined parameter, namely ρ_λ; (iv) the sizes (i.e., inputs amounts) of any particular DMU in two time periods are correlated and can be controlled via a pre-determined parameter, say ρ.

In particular, we assume that the true technology in the period t is characterized by

$$\Psi^t = \{(x, y) \in \mathbb{R}^p_+ \times \mathbb{R}^1_+ : y \leq \psi^t(x)\}, \quad t = 1, 2, \tag{58}$$

where $\psi^t(x) = \Upsilon^t(x_1 - b_1^t)^{\beta_1^t} \dots (x_p - b_p^t)^{\beta_p^t}$, $\beta_j^t, b_j^t > 0$, $\sum_{j=1}^p \beta_j^t < 1$. Note that $\psi^t(x)$ resembles the Stone-Geary utility function (Geary, 1950; Stone, 1954).[14]

In order to control the degree to which the production frontier in period 2 is different from that in period 1, we set $\Upsilon^2 = \Upsilon^1 + \delta$ and $b_j^1 = b_j^2$, $\beta_j^2 = \beta_j^1 + \delta$ $(j = 1, \dots, p)$, where δ is the predetermined parameter mentioned above. In order to guarantee that the function $\psi^t(.)$ is well-defined, we also need $x_j \geq b_j^t$ for $j = 1, \dots, p$, $t = 1, 2$. This constraint is reasonable from the economic viewpoint since it illustrates that firms incur fixed costs or some threshold expenses (i.e., b_i^t) in order to start producing positive outputs. Our data generating process includes three steps, as described below.

1. For each $t = 1, 2$, generate n points of the form $(x_i^t, \psi^t(x_i^t))$ $(i = 1, \dots, n)$. Here $x_i^t = (x_{i_1}^t, \dots, x_{i_p}^t)$ is the realization of random vectors $X_i^t = (X_{i1}^t, \dots, X_{i_p}^t)$ where $X_{i_j}^t \sim$ Uniform$(1, 10)$ such that $corr(X_{i_j}^1, X_{i_j}^2) = \rho$ for all $i \in \{1, \dots, n\}$, $j \in \{1, \dots, p\}$. The points $(x_i^t, \psi^t(x_i^t))$ generated in this step serve as optimal points, i.e., hypothetically fully efficient DMUs. The parameter ρ predetermines the correlation of inputs of each DMU in two different time periods. The details on how to generate correlated random numbers are given in Appendix T2.1.

2. Generate $\lambda(x_i^t, y_i^t|\Psi^t)$ for DMUs such that the efficiencies of the same DMU in different time periods are correlated according to ρ_λ. Formally, for each $i \in \{1, \dots, n\}$,

$$\lambda(X_i^1, Y_i^1|\Psi^1),\ \lambda(X_i^2, Y_i^2|\Psi^2) \sim 1 + |\mathcal{N}(0, 0.25)^2|, \tag{59}$$

[14]For example, this function was used in a production context by Beattie and Aradhyula (2015).

where $corr\left(\lambda(X_i^1, Y_i^1|\Psi^t), \lambda(X_i^2, Y_i^2|\Psi^t)\right) \neq 0$ and depends on the pre-determined parameter ρ_λ mentioned above (see Appendix T2.1 for more details).

3. Use the generated efficiency scores to project the optimal points away from the corresponding frontiers to obtain the set $\{(x_i^t, y_i^t)\}_{i\in\{1,\ldots,n\},t\in\{1,2\}}$ where $y_i^t = \psi^t(x_i^t)/\lambda(x_i^t, y_i^t|\Psi^t)$, newly generated in each Monte Carlo trial. Note that we project the optimal points from the frontiers of Ψ^t, not the conical hull $\mathcal{C}(\Psi^t)$.

It is important to highlight that as $q = 1$, the vectors of output prices are scalars and they cancel out in calculations. As such, we set output prices to be unity in our Monte Carlo experiments, i.e., $w^1 = w^2 = 1$. The summary of parameters used in our Monte Carlo experiments is presented in Table 1 below. In addition, we provide graphical illustrations of production frontiers in the two and three dimensional spaces in Figure 1.

Table 1: Summary of experiment parameters.

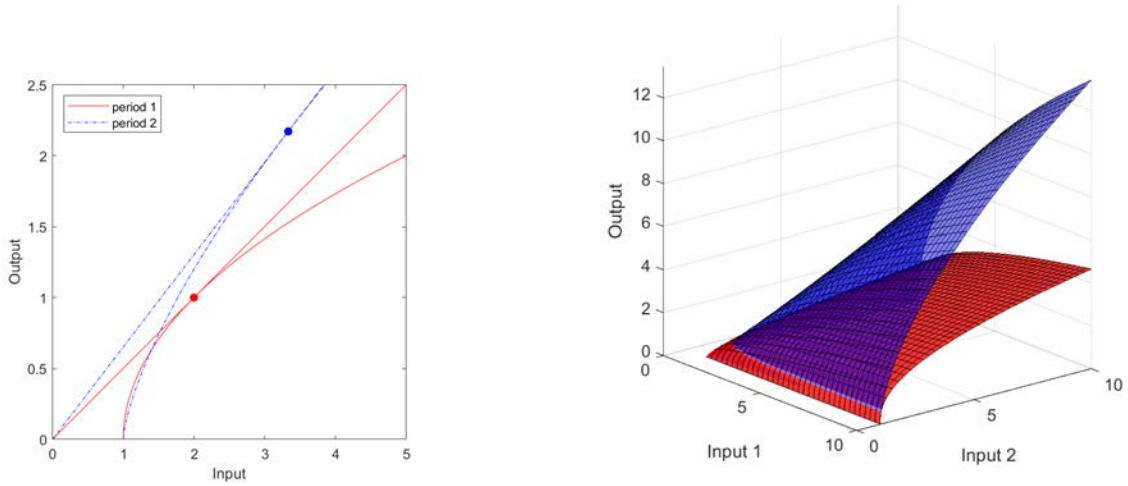
Parameters	Value
Number of Monte Carlo trials in each experiment (MC)	1000
Number of iterations in estimating the bias (L)	10
Level of significance (α)	$0.01, 0.05, 0.10$
Sample size (n)	10, 20, 50, 100, 500, 1000
Number of inputs (p)	$1, 2, 3, 4$
Parameter controlling deviation from the null hypothesis (δ)	$0.00, 0.01, 0.02, 0.03, 0.04$
Parameter controlling correlation of inputs in periods 1 and 2 (ρ)	0.5
Parameter controlling correlation of efficiencies in periods 1 and 2 (ρ_λ)	0.5
Distribution of the input variables	Uniform(1,10)
Distribution of efficiency	$1 + \lvert\mathcal{N}(0, 0.25^2)\rvert$
Exponents of the production frontier function in period 1 $(\beta_1^1, \ldots, \beta_p^1)$	
$p = 1$	0.5
$p = 2$	$(0.3, 0.4)$
$p = 3$	$(0.1, 0.2, 0.3)$
$p = 4$	$(0.1, 0.15, 0.2, 0.25)$

$\Upsilon^1 = 1$, $\Upsilon^2 = \Upsilon^1 + \delta$; $b_i^t = 1$ $(i = 1, \ldots, p;\ t = 1, 2)$; $\beta_i^2 = \beta_i^1 + \delta$ $(i = 1, \ldots, p)$; number of output $q = 1$.

9.2 Simulation results

As can be seen in Table 1, we conduct 120 Monte Carlo experiments, each including 1000 trials ($MC = 1000$) and corresponding to a combination of n, p and δ.[15] The rejection

[15] On a machine with 3.3GHz Intel Core i5-6600 CPU, 8GB RAM and MATLAB®R2018b, the runtime of one trial corresponding to the scenario where $p = 4$, $n = 1000$, $\delta = 0.04$ was about 7 minutes and 58 minutes for $L = 10$ and $L = 100$, respectively.



(a) One input one output

(b) Two inputs one output

Figure 1: Prodution frontiers used in Monte Carlo experiments. Periods 1 and 2 are illustrated by red and blue curves/surfaces, respectively. The parameter δ representing the difference between two time periods is set to be 0.2.

rates for testing for productivity change and the estimated coverage of confidence intervals corresponding to $p = 1, 2, 3, 4$ are reported in Appendices B.1, B.2, B.3, B.4, respectively. In order to obtain the true values of ξ for evaluating the coverages of confidence intervals, we conduct prior Monte Carlo simulations of $n = 10 \times 10^6$ observations without DEA estimation (see Appendix T2.3 for more details).

We focus on analyzing the performance of statistical inferences based on the following types of confidence intervals:

(i) Using the "naive" standard central limit theorems, i.e., using the confidence interval (22) with unobserved elements being replaced by their respective DEA estimates.

(ii) Using Theorem 8 for $\kappa \geq 1/2$ ($p + q = 2, 3$), i.e., confidence interval (53).

(iii) Using Theorem 8 for $\kappa < 1/2$ ($p + q = 4, 5, 6 \ldots$), i.e., confidence interval (54).

(iv) Using confidence interval (57), the re-centered version of (54) as discussed in Remark 4.

For convenience, we will refer to types (i)–(iv) as "ST", "CL1", "CL2", "RC2", respectively.

In general, the simulation results support our newly developed theory. Regarding the hypothesis testing, given a sample size, the rejection rates increase as the parameter δ, representing departure from the null hypothesis, increases. Moreover, for $\delta = 0$, the rejection rates (now representing the estimated size of the test) generally tend to approximate the nominal size (i.e., the respective value of α) as the sample size increases. Especially, the rejection rates for $\delta = 0$ corresponding to the re-centered confidence intervals RC2 converge to zero as n increases regardless of the values of the nominal size of the test α, confirming our Remark 3 in Section 8. Meanwhile, for $\delta \neq 0$, the rejection rates (now representing the estimated power of the test) increase toward unity as the sample size increases, and especially, the bigger values of n show faster convergence to unity.

With regard to the performance of confidence intervals, given a value of δ, when the sample size increases, the estimated coverage shows an upward trend to approximate the respective nominal coverage $1 - \alpha$. Similar to KSW2018, the re-centered confidence intervals RC2 show superior performance as their coverage converges to 1 rather than the nominal $1 - \alpha$, although they have the same width as the confidence intervals CL2. This superior performance verifies our Remark 3 in Section 8.

Furthermore, it can be seen that the performance of developed statistical inferences is quite impressive even with small sample sizes such as $n = 10$, e.g., the estimated coverage for ξ using confidence intervals CL1 when $p = 2, \alpha = 0.05, \delta = 0$ is 0.794.

The Monte Carlo evidence here also confirms our Remark 4 mentioned in Section 8. Indeed, when $\kappa = 2/5$ (i.e., $p = 3, q = 1$), the confidence intervals CL2 generally outperform the corresponding CL1 in terms of the size of the test and coverages (see Tables 7). Note that, on the contrary, the inferences CL1 perform better than the inferences CL2 in terms of the power of the test. A possible explanation might be due to the trade-off between the size and power of the test, similar to KSW2018.

Interestingly, our Monte Carlo evidence shows that the naive standard approach ST performs quite well in low dimensions (i.e., the number of inputs and outputs $p+q$) where $\kappa \geq 1/2$ (i.e., $p = 1, q = 1$ and $p = 2, q = 1$), illustrated by the fact that its performance

is quite similar to that of the CL1 there. Comparing the performance of ST to that of CL2 when $\kappa = 2/5$ and $\kappa = 1/3$ (i.e., $p = 3, q = 1$ and $p = 4, q = 1$), we recognize that they are relatively analogous for small values of δ such as 0 or 0.01, while for $\delta = 0.04$, the latter slightly outperforms the former in terms of estimated coverage of confidence intervals when $n = 1000$. These findings support the conjecture in Remarks 5 and 6 in Section 8 that in some peculiar cases, the bias might be tiny (or even cancel out in some more special symmetric cases such as $\delta = 0$) so that their explosion to infinity might not be observed clearly. Even so, the re-centered RC2 still outperforms ST in terms of estimated coverage.

10 Conclusion

While it is easy to see that accounting for economic weights of individuals in aggregations of indices makes sense and that it may yield very different conclusions compared to the equal-weight aggregations, its practical implementation still needs some statistical theories (for constructing confidence intervals and performing hypothesis tests), which have never been done before. In this paper we have filled this gap by developing a comprehensive set of asymptotic properties for the meaningful aggregation of MPIs, a weighted harmonic-type mean of individual efficiency scores. This provides operational researchers with the tools for constructing theoretically justified confidence intervals and statistical tests for the aggregate productivity indices. Our Monte Carlo evidence confirms that the newly developed statistical inferences perform well in finite samples similar to that in KSW2015, SZ, KSW2018.

In addition, we have contributed to the literature with an appropriate approach to derive asymptotic properties for complex indices that are nonlinear with respect to their DEA-based components. This approach paves the way for deriving a similar theory for other sophisticated indices such as the Hicks-Moorsteen productivity index. This paper also provides important statistical grounds for further theoretical developments. For instance, one may develop a test for equality of productivity change at the aggregate

level, which is very useful in practice (e.g., comparing productivity change of industries of an economy or groups of countries over time). Another potential is to improve the performance of the developed statistical inferences in small finite samples by correcting for the bias in the estimator of variance. This will be analogous to what Simar and Zelenyuk (2018b) did for the context of efficiency scores, yet it is likely to require considerable elaboration for the more complex context of MPIs, as our developments above have suggested, and so is a subject in itself.

Appendix A Regularity assumptions

This appendix includes the regularity assumptions needed to develop statistical properties for MPI, which are shown in KSW2018 as Assumptions 2.4-2.7 and 3.1-3.2. Note that these assumptions are presented for the Farrell-type input-oriented efficiency measure:

$$\theta(z|\Psi^t) = \theta(x,y|\Psi^t) = \inf_{\theta}\{\theta : (\theta x, y) \in \Psi^t\}, \tag{60}$$

and the output-oriented analogues can be expressed similarly.

Assumption 5. *(i) The random vector (X_i^t, Y_i^t) possesses a joint density f^t with support $\mathcal{D}^t \subset \Psi^t$; and (ii) f^t is continuously differentiable on $\mathcal{D}^t$.*

Assumption 6. *(i) $\mathcal{D}^{t*} \subset \mathcal{D}^t$ where $\mathcal{D}^{t*} = \{(\theta(x,y|\Psi^t)x, y) : (x,y) \in \mathcal{D}^t\}$; (ii) $\mathcal{D}^{t*}$ is compact; and (iii) $f^t(\theta(x,y|\Psi^t)x, y) > 0$ for all $(x,y) \in \mathcal{D}^t$.*

Assumption 7. *$\theta(x,y|\Psi^t)$ is three times continuously differentiable on $\mathcal{D}^t$.*

Assumption 8 (D^t is almost strictly convex)**.** *For any $(x,y), (\widetilde{x},\widetilde{y}) \in \mathcal{D}^t$ with $\left(\frac{x}{\|x\|}, y\right) \neq \left(\frac{\widetilde{x}}{\|x\|}, \widetilde{y}\right)$, the set $\{(x^*,y^*) : (x^*,y^*) = (1-\alpha)(x,y) + \alpha(\widetilde{x},\widetilde{y}), \alpha \in (0,1)\}$ is a subset of the interior of $\mathcal{D}^t$.*

Prior to mentioning Assumptions 9 and 10, we need some definitions as in KSW2018 below:

$$\mathcal{D}^t_{norm} = \left\{ \left(\frac{x}{\|x\|}, \frac{y}{\|y\|} \right) : (x,y) \in \mathcal{D}^t \right\} \tag{61}$$

and

$$\widetilde{g}_x\left(a\frac{y}{\|y\|}\right) = \min_{b>0}\left\{ b\frac{x}{\|x\|} : \left(b\frac{x}{\|x\|}, a\frac{y}{\|y\|} \right) \in \Psi^t \right\}. \tag{62}$$

Then there exists a unique $\alpha^{x,y}_{min} > 0$ such that

$$\frac{\widetilde{g}_x(\alpha^{x,y}_{min}\frac{y}{\|y\|})}{\alpha^{x,y}_{min}} = \min_{a>0}\left\{ \frac{\widetilde{g}_x(a\frac{y}{\|y\|})}{a} : \left(\widetilde{g}_x\left(a\frac{y}{\|y\|}\right)\frac{x}{\|x\|}, a\frac{y}{\|y\|} \right) \in \Psi^t \right\}. \tag{63}$$

Assumption 9. *(i) The support* $\mathcal{D}^t \subset \Psi^t$ *of* f^t *is such that for any* $\left(\frac{x}{\|x\|}, \frac{y}{\|y\|}\right) \in \mathcal{D}^t_{norm}$*, we have* $\left(\widetilde{g}_x(\alpha^{x,y}_{min}\frac{y}{\|y\|})\frac{x}{\|x\|}, \alpha^{x,y}_{min}\frac{y}{\|y\|}\right) \in \mathcal{D}^t$*; (ii) there exists a* $\delta > 0$ *such that for any* $\left(\frac{x}{\|x\|}, \frac{y}{\|y\|}\right) \in \mathcal{D}^t_{norm}$ *we also have* $\left(\widetilde{g}_x([\alpha^{x,y}_{min} - \delta]\frac{y}{\|y\|})\frac{x}{\|x\|}, [\alpha^{x,y}_{min} - \delta]\frac{y}{\|y\|}\right) \in \mathcal{D}^t$ *and* $\left(\widetilde{g}_x([\alpha^{x,y}_{min} + \delta]\frac{y}{\|y\|})\frac{x}{\|x\|}, [\alpha^{x,y}_{min} + \delta]\frac{y}{\|y\|}\right) \in \mathcal{D}^t$*; and (iii) there exists a constant* $0 < M < \infty$ *such that* $\|x\| \leq M \ \forall (x, y) \in \mathcal{D}^t$*.*

Assumption 10. *(i) For* $t \in \{1, 2\}$ *there are iid observations* (X^t_i, Y^t_i)*,* $i = 1, \ldots, n_t$*, such that Assumption 1-9 are satisfied with respect to the underlying densities* f^t *with supports* $\mathcal{D}^t$*; (ii)* $\mathcal{D}^1_{norm} = \mathcal{D}^2_{norm}$*; (iii) for some* $n \leq \min\{n_1, n_2\}$ *the observations* $((X^1_i, Y^1_i), (X^2_i, Y^2_i))$*,* $i = 1, \ldots, n$ *are iid and their joint distribution possesses a continuous density* f_{12} *with support* $\mathcal{D}^1 \times \mathcal{D}^2$*; (iv) for any* $i = 1, \ldots, n_1$*,* (X^1_i, Y^1_i) *is independent of* (X^2_j, Y^2_j) *for all* $j = 1, \ldots, n_2$ *with* $i \neq j$*; (v) for any* $i = 1, \ldots, n_2$*,* (X^2_i, Y^2_i) *is independent of* (X^1_j, Y^1_j) *for all* $j = 1, \ldots, n_1$ *with* $i \neq j$*.*

Appendix B Simulation results

B.1 Simulation results when $p = 1, q = 1$

Table 2: Rejection rates for test for aggregate productivity change using ξ when $p = 1, q = 1$.

n	δ	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
		ST	CL1	ST	CL1	ST	CL1
10	0.00	0.158	0.161	0.100	0.104	0.039	0.034
	0.01	0.169	0.182	0.110	0.112	0.043	0.050
	0.02	0.273	0.286	0.189	0.199	0.094	0.100
	0.03	0.366	0.371	0.276	0.281	0.144	0.142
	0.04	0.558	0.564	0.443	0.443	0.269	0.266
20	0.00	0.130	0.134	0.080	0.080	0.025	0.023
	0.01	0.192	0.192	0.117	0.119	0.046	0.046
	0.02	0.354	0.348	0.257	0.253	0.138	0.138
	0.03	0.558	0.558	0.450	0.453	0.256	0.253
	0.04	0.701	0.708	0.605	0.603	0.435	0.427
50	0.00	0.122	0.125	0.061	0.063	0.013	0.012
	0.01	0.255	0.254	0.163	0.160	0.065	0.063
	0.02	0.557	0.552	0.435	0.426	0.247	0.241
	0.03	0.833	0.834	0.754	0.752	0.558	0.556
	0.04	0.970	0.971	0.950	0.948	0.870	0.864
100	0.00	0.108	0.105	0.058	0.058	0.014	0.013
	0.01	0.356	0.351	0.254	0.254	0.116	0.112
	0.02	0.798	0.792	0.705	0.702	0.480	0.488
	0.03	0.978	0.977	0.963	0.961	0.888	0.888
	0.04	1.000	1.000	1.000	1.000	0.994	0.995
500	0.00	0.113	0.113	0.060	0.060	0.013	0.012
	0.01	0.878	0.878	0.807	0.808	0.589	0.590
	0.02	1.000	1.000	1.000	1.000	0.999	0.999
	0.03	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000
1000	0.00	0.130	0.127	0.069	0.069	0.008	0.008
	0.01	0.984	0.985	0.961	0.960	0.902	0.902
	0.02	1.000	1.000	1.000	1.000	1.000	1.000
	0.03	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000

$MC = 1000$, $L = 10$, $\rho = \rho_\lambda = 0.5$.

Table 3: Coverage of confidence intervals for ξ when $p = 1, q = 1$.

δ	n	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
		ST	CL1	ST	CL1	ST	CL1
0.00	10	0.842	0.839	0.900	0.896	0.961	0.966
	20	0.870	0.866	0.920	0.920	0.975	0.977
	50	0.878	0.875	0.939	0.937	0.987	0.988
	100	0.892	0.895	0.942	0.942	0.986	0.987
	500	0.887	0.887	0.940	0.940	0.987	0.988
	1000	0.870	0.873	0.931	0.931	0.992	0.992
0.01	10	0.851	0.848	0.912	0.905	0.968	0.967
	20	0.871	0.869	0.926	0.923	0.978	0.976
	50	0.890	0.896	0.936	0.938	0.983	0.985
	100	0.877	0.879	0.933	0.934	0.989	0.988
	500	0.897	0.898	0.943	0.943	0.993	0.993
	1000	0.887	0.887	0.936	0.936	0.991	0.991
0.02	10	0.834	0.828	0.896	0.887	0.960	0.956
	20	0.830	0.831	0.894	0.887	0.966	0.963
	50	0.883	0.885	0.950	0.945	0.992	0.992
	100	0.902	0.897	0.950	0.951	0.987	0.985
	500	0.918	0.921	0.961	0.961	0.993	0.993
	1000	0.889	0.889	0.949	0.949	0.990	0.991
0.03	10	0.847	0.841	0.904	0.903	0.962	0.961
	20	0.863	0.858	0.920	0.917	0.977	0.976
	50	0.883	0.883	0.942	0.938	0.981	0.981
	100	0.893	0.892	0.946	0.949	0.987	0.986
	500	0.912	0.909	0.954	0.954	0.988	0.989
	1000	0.919	0.922	0.953	0.953	0.991	0.991
0.04	10	0.861	0.851	0.894	0.893	0.960	0.950
	20	0.847	0.852	0.910	0.905	0.965	0.965
	50	0.893	0.891	0.938	0.937	0.984	0.985
	100	0.898	0.895	0.947	0.946	0.988	0.988
	500	0.897	0.895	0.939	0.940	0.986	0.987
	1000	0.894	0.893	0.942	0.942	0.991	0.991

$MC = 1000$, $L = 10$, $\rho = \rho_\lambda = 0.5$.

B.2 Simulation results when $p = 2, q = 1$

Table 4: Rejection rates for test for aggregate productivity change using ξ when $p = 2, q = 1$.

n	δ	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
		ST	CL1	ST	CL1	ST	CL1
10	0.00	0.215	0.290	0.136	0.206	0.050	0.114
	0.01	0.311	0.367	0.224	0.294	0.119	0.171
	0.02	0.584	0.566	0.471	0.482	0.322	0.347
	0.03	0.799	0.769	0.707	0.698	0.543	0.550
	0.04	0.928	0.878	0.889	0.849	0.775	0.749
20	0.00	0.142	0.186	0.087	0.115	0.030	0.036
	0.01	0.382	0.392	0.288	0.310	0.133	0.154
	0.02	0.749	0.735	0.672	0.642	0.451	0.464
	0.03	0.960	0.950	0.933	0.926	0.827	0.818
	0.04	0.996	0.992	0.989	0.981	0.956	0.957
50	0.00	0.100	0.109	0.051	0.056	0.009	0.012
	0.01	0.589	0.584	0.462	0.472	0.252	0.278
	0.02	0.975	0.976	0.953	0.952	0.861	0.865
	0.03	0.999	0.999	0.998	0.998	0.993	0.993
	0.04	1.000	1.000	1.000	1.000	1.000	1.000
100	0.00	0.106	0.112	0.053	0.062	0.016	0.018
	0.01	0.845	0.844	0.757	0.763	0.569	0.554
	0.02	1.000	0.999	0.999	0.999	0.994	0.994
	0.03	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000
500	0.00	0.106	0.106	0.052	0.052	0.009	0.008
	0.01	1.000	1.000	1.000	1.000	1.000	1.000
	0.02	1.000	1.000	1.000	1.000	1.000	1.000
	0.03	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000
1000	0.00	0.097	0.097	0.053	0.054	0.015	0.015
	0.01	1.000	1.000	1.000	1.000	1.000	1.000
	0.02	1.000	1.000	1.000	1.000	1.000	1.000
	0.03	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000

$MC = 1000$, $L = 10$, $\rho = \rho_\lambda = 0.5$.

Table 5: Coverage of confidence intervals for ξ when $p = 2, q = 1$.

δ	n	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
		ST	CL1	ST	CL1	ST	CL1
0.00	10	0.785	0.710	0.864	0.794	0.950	0.886
	20	0.858	0.814	0.913	0.885	0.970	0.964
	50	0.900	0.891	0.949	0.944	0.991	0.988
	100	0.894	0.888	0.947	0.938	0.984	0.982
	500	0.894	0.894	0.948	0.948	0.991	0.992
	1000	0.903	0.903	0.947	0.946	0.985	0.985
0.01	10	0.794	0.709	0.855	0.800	0.939	0.886
	20	0.849	0.813	0.914	0.883	0.971	0.955
	50	0.880	0.869	0.937	0.933	0.984	0.986
	100	0.878	0.867	0.934	0.938	0.990	0.988
	500	0.900	0.895	0.946	0.946	0.984	0.983
	1000	0.897	0.897	0.949	0.945	0.988	0.987
0.02	10	0.787	0.702	0.859	0.790	0.931	0.892
	20	0.852	0.825	0.917	0.896	0.969	0.960
	50	0.880	0.879	0.939	0.943	0.982	0.976
	100	0.895	0.888	0.950	0.949	0.991	0.990
	500	0.908	0.906	0.949	0.948	0.985	0.982
	1000	0.894	0.895	0.946	0.945	0.989	0.988
0.03	10	0.812	0.737	0.882	0.802	0.951	0.905
	20	0.848	0.813	0.905	0.877	0.969	0.956
	50	0.904	0.888	0.953	0.938	0.988	0.980
	100	0.906	0.897	0.953	0.950	0.988	0.984
	500	0.882	0.883	0.947	0.949	0.992	0.991
	1000	0.914	0.908	0.963	0.964	0.994	0.993
0.04	10	0.812	0.723	0.865	0.794	0.945	0.897
	20	0.845	0.835	0.916	0.895	0.974	0.966
	50	0.878	0.859	0.935	0.920	0.984	0.982
	100	0.897	0.890	0.952	0.948	0.992	0.990
	500	0.908	0.907	0.944	0.950	0.986	0.991
	1000	0.909	0.899	0.963	0.958	0.994	0.993

$MC = 1000$, $L = 10$, $\rho = \rho_\lambda = 0.5$.

B.3 Simulation results when $p = 3, q = 1$

Table 6: Rejection rates for test for aggregate productivity change using ξ when $p = 3, q = 1$.

n	δ	$\alpha = 0.10$				$\alpha = 0.05$				$\alpha = 0.01$			
		ST	CL1	CL2	RC2	ST	CL1	CL2	RC2	ST	CL1	CL2	RC2
10	0.00	0.235	0.364	0.333	0.254	0.171	0.278	0.240	0.189	0.078	0.172	0.132	0.099
	0.01	0.365	0.470	0.394	0.350	0.273	0.378	0.318	0.265	0.170	0.249	0.185	0.145
	0.02	0.648	0.648	0.551	0.546	0.567	0.569	0.472	0.471	0.394	0.455	0.348	0.312
	0.03	0.878	0.816	0.713	0.750	0.827	0.765	0.654	0.674	0.681	0.648	0.514	0.511
	0.04	0.967	0.914	0.852	0.880	0.946	0.892	0.799	0.822	0.860	0.811	0.688	0.701
20	0.00	0.196	0.256	0.237	0.143	0.126	0.185	0.160	0.087	0.046	0.097	0.054	0.031
	0.01	0.460	0.491	0.371	0.316	0.360	0.391	0.273	0.219	0.203	0.236	0.149	0.099
	0.02	0.830	0.784	0.621	0.662	0.755	0.727	0.539	0.562	0.575	0.584	0.380	0.359
	0.03	0.988	0.965	0.872	0.913	0.977	0.942	0.822	0.857	0.904	0.873	0.667	0.702
	0.04	1.000	1.000	0.976	0.992	1.000	0.995	0.961	0.977	0.993	0.980	0.900	0.929
50	0.00	0.120	0.152	0.142	0.045	0.078	0.090	0.078	0.018	0.019	0.032	0.021	0.002
	0.01	0.690	0.697	0.439	0.424	0.594	0.588	0.348	0.302	0.391	0.386	0.169	0.109
	0.02	0.997	0.995	0.902	0.965	0.994	0.990	0.863	0.925	0.953	0.955	0.716	0.764
	0.03	1.000	1.000	0.996	1.000	1.000	1.000	0.987	1.000	1.000	1.000	0.964	0.986
	0.04	1.000	1.000	0.999	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.998	1.000
100	0.00	0.123	0.135	0.120	0.022	0.068	0.069	0.056	0.006	0.013	0.022	0.011	0.001
	0.01	0.915	0.917	0.655	0.722	0.879	0.868	0.538	0.542	0.730	0.732	0.309	0.219
	0.02	1.000	1.000	0.989	0.999	1.000	1.000	0.978	0.998	0.999	0.999	0.919	0.985
	0.03	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	0.00	0.115	0.126	0.107	0.005	0.062	0.069	0.052	0.000	0.016	0.018	0.012	0.000
	0.01	1.000	1.000	0.994	1.000	1.000	1.000	0.981	1.000	1.000	1.000	0.915	0.987
	0.02	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.03	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	0.00	0.113	0.115	0.103	0.002	0.052	0.058	0.052	0.000	0.015	0.014	0.008	0.000
	0.01	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.995	1.000
	0.02	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.03	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$MC = 1000$, $L = 10$, $\rho = \rho_\lambda = 0.5$.

Table 7: Coverage of confidence intervals for ξ when $p = 3, q = 1$.

δ	n	$\alpha = 0.10$				$\alpha = 0.05$				$\alpha = 0.01$			
		ST	CL1	CL2	RC2	ST	CL1	CL2	RC2	ST	CL1	CL2	RC2
0.00	10	0.765	0.636	0.667	0.746	0.829	0.722	0.760	0.811	0.922	0.828	0.868	0.901
	20	0.804	0.744	0.763	0.857	0.874	0.815	0.840	0.913	0.954	0.903	0.946	0.969
	50	0.880	0.848	0.858	0.955	0.922	0.910	0.922	0.982	0.981	0.968	0.979	0.998
	100	0.877	0.865	0.880	0.978	0.932	0.931	0.944	0.994	0.987	0.978	0.989	0.999
	500	0.885	0.874	0.893	0.995	0.938	0.931	0.948	1.000	0.984	0.982	0.988	1.000
	1000	0.887	0.885	0.897	0.998	0.948	0.942	0.948	1.000	0.985	0.986	0.992	1.000
0.01	10	0.778	0.650	0.675	0.759	0.851	0.731	0.744	0.811	0.939	0.828	0.859	0.906
	20	0.810	0.732	0.781	0.850	0.875	0.799	0.851	0.915	0.959	0.908	0.944	0.968
	50	0.864	0.831	0.863	0.954	0.913	0.900	0.919	0.975	0.969	0.965	0.979	0.999
	100	0.882	0.868	0.886	0.980	0.932	0.917	0.939	0.991	0.988	0.976	0.990	1.000
	500	0.902	0.893	0.908	0.998	0.937	0.934	0.956	1.000	0.989	0.984	0.994	1.000
	1000	0.888	0.883	0.892	1.000	0.950	0.937	0.951	1.000	0.991	0.988	0.988	1.000
0.02	10	0.768	0.615	0.640	0.710	0.838	0.685	0.728	0.782	0.933	0.796	0.864	0.884
	20	0.805	0.736	0.785	0.866	0.889	0.816	0.857	0.922	0.965	0.917	0.953	0.975
	50	0.860	0.840	0.860	0.961	0.923	0.903	0.917	0.986	0.987	0.969	0.979	0.998
	100	0.886	0.881	0.880	0.985	0.940	0.929	0.941	0.993	0.983	0.984	0.983	1.000
	500	0.896	0.897	0.916	0.998	0.946	0.937	0.951	1.000	0.993	0.991	0.988	1.000
	1000	0.895	0.900	0.883	0.999	0.960	0.952	0.931	1.000	0.991	0.990	0.992	1.000
0.03	10	0.779	0.636	0.702	0.736	0.846	0.708	0.774	0.809	0.932	0.820	0.874	0.897
	20	0.820	0.742	0.786	0.849	0.881	0.807	0.867	0.916	0.969	0.904	0.943	0.970
	50	0.842	0.832	0.859	0.947	0.927	0.896	0.927	0.982	0.976	0.961	0.978	0.996
	100	0.860	0.842	0.885	0.979	0.921	0.914	0.949	0.994	0.980	0.979	0.985	0.999
	500	0.889	0.885	0.913	0.997	0.940	0.933	0.950	1.000	0.987	0.987	0.993	1.000
	1000	0.900	0.894	0.897	1.000	0.949	0.952	0.941	1.000	0.994	0.995	0.989	1.000
0.04	10	0.778	0.630	0.682	0.739	0.856	0.712	0.757	0.812	0.936	0.825	0.879	0.907
	20	0.819	0.738	0.779	0.871	0.884	0.816	0.852	0.935	0.967	0.931	0.950	0.978
	50	0.858	0.834	0.886	0.952	0.926	0.902	0.932	0.977	0.980	0.962	0.983	0.994
	100	0.873	0.862	0.866	0.979	0.937	0.929	0.931	0.988	0.977	0.979	0.981	0.999
	500	0.910	0.902	0.896	0.999	0.953	0.949	0.948	1.000	0.987	0.985	0.985	1.000
	1000	0.906	0.907	0.913	0.999	0.946	0.949	0.960	0.999	0.992	0.992	0.993	1.000

$MC = 1000$, $L = 10$, $\rho = \rho_\lambda = 0.5$.

B.4 Simulation results when $p = 4, q = 1$

Table 8: Rejection rates for test for aggregate productivity change using ξ when $p = 4, q = 1$.

n	δ	$\alpha = 0.10$				$\alpha = 0.05$				$\alpha = 0.01$			
		ST	CL1	CL2	RC2	ST	CL1	CL2	RC2	ST	CL1	CL2	RC2
10	0.00	0.239	0.465	0.359	0.293	0.170	0.387	0.285	0.213	0.086	0.262	0.178	0.118
	0.01	0.458	0.547	0.428	0.393	0.382	0.480	0.354	0.305	0.241	0.368	0.243	0.196
	0.02	0.810	0.727	0.610	0.605	0.755	0.676	0.530	0.519	0.617	0.568	0.411	0.390
	0.03	0.975	0.888	0.785	0.814	0.951	0.860	0.728	0.753	0.884	0.792	0.596	0.594
	0.04	0.993	0.955	0.890	0.915	0.992	0.941	0.856	0.886	0.976	0.905	0.771	0.803
20	0.00	0.197	0.351	0.236	0.140	0.120	0.274	0.156	0.078	0.049	0.155	0.083	0.019
	0.01	0.618	0.602	0.399	0.363	0.523	0.524	0.325	0.265	0.346	0.388	0.190	0.115
	0.02	0.964	0.903	0.720	0.780	0.941	0.874	0.637	0.689	0.868	0.802	0.477	0.474
	0.03	1.000	0.991	0.919	0.950	0.999	0.984	0.888	0.925	0.987	0.960	0.797	0.831
	0.04	1.000	1.000	0.988	0.997	1.000	1.000	0.973	0.994	1.000	0.999	0.939	0.973
50	0.00	0.179	0.249	0.154	0.028	0.123	0.177	0.088	0.011	0.041	0.071	0.030	0.001
	0.01	0.912	0.875	0.533	0.549	0.862	0.828	0.434	0.385	0.716	0.700	0.262	0.132
	0.02	1.000	1.000	0.926	0.990	1.000	0.998	0.877	0.964	0.999	0.994	0.731	0.841
	0.03	1.000	1.000	0.996	1.000	1.000	1.000	0.991	1.000	1.000	1.000	0.971	0.999
	0.04	1.000	1.000	0.999	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.996	1.000
100	0.00	0.122	0.176	0.132	0.005	0.074	0.110	0.079	0.002	0.019	0.036	0.022	0.000
	0.01	0.995	0.991	0.656	0.796	0.989	0.979	0.541	0.610	0.956	0.931	0.333	0.230
	0.02	1.000	1.000	0.987	1.000	1.000	1.000	0.979	1.000	1.000	1.000	0.912	0.997
	0.03	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
500	0.00	0.111	0.123	0.105	0.000	0.061	0.069	0.051	0.000	0.018	0.016	0.016	0.000
	0.01	1.000	1.000	0.980	1.000	1.000	1.000	0.952	1.000	1.000	1.000	0.832	0.997
	0.02	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.03	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1000	0.00	0.106	0.119	0.103	0.000	0.044	0.056	0.052	0.000	0.010	0.009	0.009	0.000
	0.01	1.000	1.000	0.999	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.984	1.000
	0.02	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.03	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

$MC = 1000$, $L = 10$, $\rho = \rho_\lambda = 0.5$.

Table 9: Coverage of confidence intervals for ξ when $p = 4, q = 1$.

δ	n	$\alpha = 0.10$				$\alpha = 0.05$				$\alpha = 0.01$			
		ST	CL1	CL2	RC2	ST	CL1	CL2	RC2	ST	CL1	CL2	RC2
0.00	10	0.761	0.535	0.641	0.707	0.830	0.613	0.715	0.787	0.914	0.738	0.822	0.882
	20	0.803	0.649	0.764	0.860	0.880	0.726	0.844	0.922	0.951	0.845	0.917	0.981
	50	0.821	0.751	0.846	0.972	0.877	0.823	0.912	0.989	0.959	0.929	0.970	0.999
	100	0.878	0.824	0.868	0.995	0.926	0.890	0.921	0.998	0.981	0.964	0.978	1.000
	500	0.889	0.877	0.895	1.000	0.939	0.931	0.949	1.000	0.982	0.984	0.984	1.000
	1000	0.894	0.881	0.897	1.000	0.956	0.944	0.948	1.000	0.990	0.991	0.991	1.000
0.01	10	0.738	0.513	0.606	0.680	0.816	0.601	0.687	0.760	0.920	0.704	0.821	0.867
	20	0.791	0.633	0.766	0.868	0.870	0.718	0.845	0.920	0.948	0.848	0.935	0.975
	50	0.840	0.752	0.837	0.967	0.893	0.843	0.898	0.987	0.967	0.923	0.975	0.999
	100	0.876	0.822	0.873	0.995	0.932	0.884	0.936	0.999	0.982	0.959	0.986	1.000
	500	0.895	0.874	0.903	1.000	0.945	0.942	0.955	1.000	0.989	0.984	0.988	1.000
	1000	0.901	0.897	0.905	1.000	0.948	0.942	0.955	1.000	0.993	0.991	0.995	1.000
0.02	10	0.754	0.509	0.625	0.686	0.814	0.573	0.687	0.770	0.906	0.710	0.820	0.882
	20	0.758	0.642	0.729	0.847	0.836	0.722	0.804	0.904	0.944	0.826	0.924	0.962
	50	0.827	0.732	0.848	0.968	0.906	0.818	0.910	0.991	0.974	0.933	0.969	0.999
	100	0.858	0.798	0.854	0.995	0.922	0.868	0.935	0.998	0.976	0.960	0.986	1.000
	500	0.897	0.871	0.889	1.000	0.946	0.935	0.946	1.000	0.990	0.990	0.990	1.000
	1000	0.912	0.901	0.892	1.000	0.957	0.948	0.947	1.000	0.990	0.991	0.991	1.000
0.03	10	0.759	0.519	0.633	0.684	0.827	0.598	0.702	0.759	0.920	0.709	0.822	0.877
	20	0.787	0.601	0.747	0.840	0.855	0.691	0.827	0.900	0.934	0.817	0.916	0.968
	50	0.833	0.739	0.843	0.969	0.892	0.822	0.904	0.985	0.962	0.918	0.968	1.000
	100	0.869	0.817	0.879	0.994	0.928	0.897	0.935	0.999	0.980	0.955	0.983	1.000
	500	0.891	0.861	0.899	1.000	0.941	0.929	0.947	1.000	0.990	0.989	0.991	1.000
	1000	0.875	0.880	0.892	1.000	0.939	0.943	0.943	1.000	0.989	0.983	0.989	1.000
0.04	10	0.736	0.495	0.606	0.659	0.804	0.560	0.684	0.733	0.898	0.684	0.804	0.863
	20	0.776	0.641	0.751	0.855	0.855	0.725	0.835	0.916	0.938	0.840	0.925	0.974
	50	0.821	0.735	0.859	0.963	0.896	0.813	0.902	0.985	0.973	0.918	0.969	0.998
	100	0.855	0.815	0.870	0.994	0.923	0.896	0.932	0.998	0.984	0.958	0.983	1.000
	500	0.891	0.877	0.898	1.000	0.938	0.923	0.951	1.000	0.989	0.982	0.995	1.000
	1000	0.900	0.889	0.902	1.000	0.943	0.938	0.950	1.000	0.986	0.981	0.986	1.000

$MC = 1000$, $L = 10$, $\rho = \rho_\lambda = 0.5$.

References

Abbott, M. (2006). The productivity and efficiency of the Australian electricity supply industry. *Energy Economics*, 28(4):444–454.

Ball, V. E., Lovell, C. K., Luu, H., and Nehring, R. (2004). Incorporating environmental impacts in the measurement of agricultural productivity growth. *Journal of Agricultural and Resource Economics*, pages 436–460.

Beattie, B. R. and Aradhyula, S. (2015). A note on threshold factor level(s) and Stone-Geary technology. *Journal of Agricultural and Applied Economics*, 47(4):482–493.

Brennan, S., Haelermans, C., and Ruggiero, J. (2014). Nonparametric estimation of education productivity incorporating nondiscretionary inputs with an application to dutch schools. *European Journal of Operational Research*, 234(3):809–818.

Caves, D. W., Christensen, L. R., and Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50(6):1393–1414.

Chen, Y. and Ali, A. I. (2004). DEA Malmquist productivity measure: New insights with an application to computer industry. *European Journal of Operational Research*, 159(1):239–249.

Ebert, U. and Welsch, H. (2004). Meaningful environmental indices: A social choice approach. *Journal of Environmental Economics and Management*, 47(2):270–283.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Färe, R. and Grosskopf, S. (2004). *New Directions: Efficiency and Productivity*. Springer.

Färe, R., Grosskopf, S., Lindgren, B., and Roos, P. (1994a). Productivity developments in Swedish hospitals: a Malmquist output index approach. In *Data envelopment analysis: Theory, methodology, and applications*, pages 253–272. Springer.

Färe, R., Grosskopf, S., and Lovell, C. K. (1985). *The Measurement of Efficiency of Production*, volume 6. Springer.

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

Färe, R., Grosskopf, S., and Roos, P. (1998). Malmquist productivity indexes: A survey of theory and practice. In *Index numbers: Essays in honour of Sten Malmquist*, pages 127–190. Springer.

Färe, R., Grosskopf, S., Yaisawarng, S., Li, S. K., and Wang, Z. (1990). Productivity growth in Illinois electric utilities. *Resources and Energy*, 12(4):383–398.

Färe, R., He, X., Li, S., and Zelenyuk, V. (2019). A unifying framework for farrell profit efficiency measurement. *Operations Research*, 67(1):183–197.

Färe, R. and Primont, D. (1995). *Multi-Output Production and Duality: Theory and Applications*. Springer.

Färe, R. and 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*, 120(3):253–290.

Federal Reserve System (2019). Large Commercial Bank releases for March 31, 2019 [Data file]. Retrieved June 21, 2019 from https://www.federalreserve.gov/releases/lbr/.

Färe, R., Grosskopf, S., and Roos, B. L. (1992). Productivity changes in Swedish pharamacies 1980–1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis*, 3(1-2):85–101.

Geary, R. C. (1950). A note on "a constant-utility index of the cost of living". *The Review of Economic Studies*, 18(1):65–66.

Gitto, S. and Mancuso, P. (2015). The contribution of physical and human capital accumulation to Italian regional growth: A nonparametric perspective. *Journal of Productivity Analysis*, 43(1):1–12.

Grifell-Tatjé, E. and Lovell, C. K. (1995). A note on the Malmquist Productivity Index. *Economics Letters*, 47(2):169–175.

Johnson, A. L. and Ruggiero, J. (2014). Nonparametric measurement of productivity and efficiency in education. *Annals of Operations Research*, 221(1):197–210.

Kneip, A., Simar, L., and Wilson, P. W. (2015). When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores. *Econometric Theory*, 31(2):394–422.

Kneip, A., Simar, L., and Wilson, P. W. (2016). Testing hypotheses in nonparametric models of production. *Journal of Business & Economic Statistics*, 34(3):435–456.

Kneip, A., Simar, L., and Wilson, P. W. (2018). Inference in Dynamic, Nonparametric Models of Production: Central Limit Theorems for Malmquist Indicies. *Working Paper*.

Miller, R. G. (1974). The jackknife – A review. *Biometrika*, 61(1):1–15.

Mukherjee, K., Ray, S. C., and Miller, S. M. (2001). Productivity growth in large US commercial banks: The initial post-deregulation experience. *Journal of Banking & Finance*, 25(5):913–939.

Murillo-Zamorano, L. R. (2005). The role of energy in productivity growth: A controversial issue? *The Energy Journal*, 26(2):69–88.

Olesen, O. B. and Petersen, N. C. (2016). Stochastic data envelopment analysis–A review. *European Journal of Operational Research*, 251(1):2–21.

Pastor, J. M., Perez, F., Quesada, J., et al. (1997). Efficiency analysis in banking firms: An international comparison. *European Journal of Operational Research*, 98(2):395–407.

Pilyavsky, A. and Staat, M. (2008). Efficiency and productivity change in Ukrainian health care. *Journal of Productivity Analysis*, 29(2):143–154.

Quenouille, M. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(1):68–84.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4):353–360.

Ray, S. C. and Desli, E. (1997). Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *The American Economic Review*, 87(5):1033–1039.

Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton University Press.

Simar, L. and Wilson, P. W. (2019). Central limit theorems and inference for sources of productivity change measured by nonparametric malmquist indices. *European Journal of Operational Research*.

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

Simar, L. and Zelenyuk, V. (2018b). Improving Finite Sample Approximation by Central Limit Theorems for DEA and FDH efficiency scores. *CEPA Working Paper WP07/2018*.

Stone, R. (1954). Linear expenditure systems and demand analysis: An application to the pattern of British demand. *The Economic Journal*, 64(255):511–527.

ten Raa, T. (2011). Benchmarking and industry performance. *Journal of Productivity Analysis*, 36(3):285–292.

Tortosa-Ausina, E., Grifell-Tatjé, E., Armero, C., and Conesa, D. (2008). Sensitivity analysis of efficiency and Malmquist productivity indices: An application to Spanish savings banks. *European Journal of Operational Research*, 184(3):1062–1084.

Tukey, J. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, 29:614.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge University Press.

Walheer, B. (2018). Aggregation of metafrontier technology gap ratios: The case of European sectors in 1995–2015. *European Journal of Operational Research*, 269(3):1013–1026.

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

Wang, H., Ang, B., Wang, Q., and Zhou, P. (2017). Measuring energy performance with sectoral heterogeneity: A non-parametric frontier approach. *Energy Economics*, 62:70–78.

Ylvinger, S. (2000). Industry performance and structural efficiency measures: Solutions to problems in firm models. *European Journal of Operational Research*, 121(1):164–174.

Zelenyuk, V. (2006). Aggregation of Malmquist productivity indexes. *European Journal of Operational Research*, 174(2):1076–1086.

Zelenyuk, V. (2014). Scale efficiency and homotheticity: equivalence of primal and dual measures. *Journal of Productivity Analysis*, 42(1):15–24.

Zhou, P., Ang, B., and Poh, K. (2008). A survey of data envelopment analysis in energy and environmental studies. *European Journal of Operational Research*, 189(1):1–18.

T1 Proofs of theorems

T1.1 Proof of Theorem 1

Consider the function $\Lambda : \mathbb{R}_{++} \to \mathbb{R}_{++}$ defined by $\Lambda(x) = x^{-1}$. Clearly, $\Lambda(\cdot)$ is monotonic, invertible, and differentiable with nonzero derivatives on $\mathbb{R}_{++}$. Hence, Theorem 3.2 of KSW2018 still holds true if $\Lambda(\cdot)$ substitutes the function $\Gamma : \mathbb{R}_{+} \to \mathbb{R}_{+}$ defined by $\Gamma(x) = x^{0.5}$ in the theorem. Similarly, the proof of Theorem 3.4 of KSW2018 is still valid if $\Lambda(\cdot)$ replaces the function $\Gamma : \mathbb{R}_{++} \to \mathbb{R}_{++}$ defined by $\Gamma(x) = \log x^{0.5}$ there. Combining these facts with Lemmas 3.1 and 3.2 of KSW2018, we can obtain asymptotic properties of the conical Farrell-type output-oriented efficiency estimators, which is analogous to Theorem 3.4 of KSW2018. Formally, there exist constants $\overline{C}_{st} \in (0, \infty)$ such that as $n \to \infty$,

$$E\left(\lambda_C(Z_i^s|\mathcal{X}_n^t) - \lambda_C(Z_i^s|\Psi^t)\right) = \overline{C}_{st} n^{-\kappa} + O(v_{n,\kappa}), \tag{64}$$

$$E\left(\left[\lambda_C(Z_i^s|\mathcal{X}_n^t) - \lambda_C(Z_i^s|\Psi^t)\right]^2\right) = o(n^{-\kappa}), \tag{65}$$

$$\left|E\left([\lambda_C(Z_i^s|\mathcal{X}_n^t) - E(\lambda_C(Z_i^s|\mathcal{X}_n^t))][\lambda_C(Z_j^{s^*}|\mathcal{X}_n^{t^*}) - E(\lambda_C(Z_j^{s^*}|\mathcal{X}_n^{t^*}))]\right)\right| = o(n^{-1}), \tag{66}$$

for all $i, j \in \{1, \ldots, n\}, i \neq j$; $s, t, s^*, t^* \in \{1, 2\}$.

For each $i = 1, \ldots, n$, since the first two moments of $w^2 Y_i^2$ are finite and

$$\widehat{U}_{1,i} - U_{1,i} = \left(\lambda_C(Z_i^2|\mathcal{X}_n^1) - \lambda_C(Z_i^2|\Psi^1)\right) w^2 Y_i^2, \tag{67}$$

the order of the first two moments of $\widehat{U}_{1,i} - U_{1,i}$ inherits those from $\lambda_C(Z_i^2|\mathcal{X}_n^1) - \lambda_C(Z_i^2|\Psi^1)$ and similarly, we have the same conclusions for $\widehat{U}_{s,i} - U_{s,i}$ $(s = 2, 3, 4)$. Also, the order of $Cov(\widehat{U}_{1,i}, \widehat{U}_{2,j})$ is the same as that of $Cov(\lambda_C(Z_i^2|\mathcal{X}_n^1), \lambda_C(Z_j^2|\mathcal{X}_n^2))$ $(j = 1, \ldots, n; j \neq i)$, and similar conclusions can be drawn for the other analogous covariances. These facts imply that Theorem 1 holds true for $s, t \in \{1, 2, 3, 4\}$.

T1.2 Proof of Theorem 2

(i) This part is a direct consequence of (32).

(ii) To begin with, we decompose $Cov(\widehat{U}_{t,i}, \widehat{U}_{t^*,i})$ as

$$\begin{aligned}
Cov(\widehat{U}_{t,i}, \widehat{U}_{t^*,i}) &= E\left(\left[\widehat{U}_{t,i} - E(\widehat{U}_{t,i})\right]\left[\widehat{U}_{t^*,i} - E(\widehat{U}_{t^*,i})\right]\right) \\
&= E\left(\left[\widehat{U}_{t,i} - U_{t,i} + U_{t,i} - E(\widehat{U}_{t,i})\right]\left[\widehat{U}_{t^*,i} - U_{t^*,i} + U_{t^*,i} - E(\widehat{U}_{t^*,i})\right]\right) \\
&= E\left([\widehat{U}_{t,i} - U_{t,i}][\widehat{U}_{t^*,i} - U_{t^*,i}]\right) + E\left([U_{t,i} - E(\widehat{U}_{t,i})][U_{t^*,i} - E(\widehat{U}_{t^*,i})]\right) \\
&\quad + E\left([\widehat{U}_{t,i} - U_{t,i}][U_{t^*,i} - E(\widehat{U}_{t^*,i})]\right) + E\left([U_{t,i} - E(\widehat{U}_{t,i})][\widehat{U}_{t^*,i} - U_{t^*,i}]\right).
\end{aligned} \tag{68}$$

Now we evaluate the rates of convergence of each term in the above decomposition of $Cov(\widehat{U}_{t,i}, \widehat{U}_{t^*,i})$.

First, by Cauchy-Schwartz inequality and (33), we have

$$\left(E([\widehat{U}_{t,i} - U_{t,i}][\widehat{U}_{t^*,i} - U_{t^*,i}])\right)^2 \leq E\left([\widehat{U}_{t,i} - U_{t,i}]^2\right) E\left([\widehat{U}_{t^*,i} - U_{t^*,i}]^2\right) = o(n^{-\kappa})o(n^{-\kappa}), \tag{69}$$

and hence,

$$E\left([\widehat{U}_{t,i} - U_{t,i}][\widehat{U}_{t^*,i} - U_{t^*,i}]\right) = o(n^{-\kappa}). \tag{70}$$

Next, the second term in (68) can be expressed as

$$\begin{aligned}
&E\left([U_{t,i} - E(\widehat{U}_{t,i})][U_{t^*,i} - E(\widehat{U}_{t^*,i})]\right) \\
&= E\left([U_{t,i} - E(U_{t,i}) + E(U_{t,i}) - E(\widehat{U}_{t,i})][U_{t^*,i} - E(U_{t^*,i}) + E(U_{t^*,i}) - E(\widehat{U}_{t^*,i})]\right) \\
&= E\left([U_{t,i} - E(U_{t,i})][U_{t^*,i} - E(U_{t^*,i})]\right) + E\left([U_{t,i} - E(U_{t,i})][E(U_{t^*,i}) - E(\widehat{U}_{t^*,i})]\right) \\
&\quad + E\left([E(U_{t,i}) - E(\widehat{U}_{t,i})][U_{t^*,i} - E(U_{t^*,i})]\right) + E\left([E(U_{t,i}) - E(\widehat{U}_{t,i})][E(U_{t^*,i}) - E(\widehat{U}_{t^*,i})]\right) \\
&= \sigma_{tt^*} + 0 + 0 + E(U_{t,i} - \widehat{U}_{t,i})E(U_{t^*,i} - \widehat{U}_{t^*,i}) \\
&= \sigma_{tt^*} + E(U_{t,i} - \widehat{U}_{t,i})E(U_{t^*,i} - \widehat{U}_{t^*,i}).
\end{aligned} \tag{71}$$

Thus, by (32), we have

$$E\left([U_{t,i} - E(\widehat{U}_{t,i})][U_{t^*,i} - E(\widehat{U}_{t^*,i})]\right) = \sigma_{tt^*} + (C_t n^{-\kappa} + o(n^{-\kappa}))(C_{t^*} n^{-\kappa} + o(n^{-\kappa}))$$
$$= \sigma_{tt^*} + o(n^{-\kappa}). \tag{72}$$

Third, by Cauchy-Schwartz inequality, (33) and applying (72) with $t = t^*$, we obtain

$$\left(E([\widehat{U}_{t,i} - U_{t,i}][U_{t^*,i} - E(\widehat{U}_{t^*,i})])\right)^2 \leq E\left([\widehat{U}_{t,i} - U_{t,i}]^2\right) E\left([U_{t^*,i} - E(\widehat{U}_{t^*,i})]^2\right)$$
$$= o(n^{-\kappa})\left(\sigma_{t^*t^*} + o(n^{-\kappa})\right) = o(n^{-\kappa}). \tag{73}$$

Consequently,

$$E\left([\widehat{U}_{t,i} - U_{t,i}][U_{t^*,i} - E(\widehat{U}_{t^*,i})]\right) = o(n^{-\kappa/2}). \tag{74}$$

By swapping the positions of t and t^* in (74), we obtain the result for the last term in (68) below.

$$E\left([U_{t,i} - E(\widehat{U}_{t,i})][\widehat{U}_{t^*,i} - U_{t^*,i}]\right) = o(n^{-\kappa/2}). \tag{75}$$

Combining (68), (70), (72), (74), and (75), we have part (ii) proved.

(iii) It can be seen that

$$\begin{aligned} Cov(\widehat{U}_{s,i}, U_{r,i}) &= E\left([\widehat{U}_{s,i} - E(\widehat{U}_{s,i})][U_{r,i} - \mu_r]\right) \\ &= E\left([\widehat{U}_{s,i} - U_{s,i} + U_{s,i} - \mu_s + \mu_s - E(\widehat{U}_{s,i})][U_{r,i} - \mu_r]\right) \\ &= E\left([\widehat{U}_{s,i} - U_{s,i}][U_{r,i} - \mu_r]\right) + \sigma_{sr} + E\left([\mu_s - E(\widehat{U}_{s,i})][U_{r,i} - \mu_r]\right) \\ &= E\left([\widehat{U}_{s,i} - U_{s,i}][U_{r,i} - \mu_r]\right) + \sigma_{sr}. \end{aligned} \tag{76}$$

By Cauchy-Schwartz inequality and (33),

$$\left(E([\widehat{U}_{s,i} - U_{s,i}][U_{r,i} - \mu_r])\right)^2 \leq E\left([\widehat{U}_{s,i} - U_{s,i}]^2\right) E\left([U_{r,i} - \mu_r]^2\right) = o(n^{-\kappa})\sigma_{rr} = o(n^{-\kappa}).$$

Thus, $E\left([\widehat{U}_{s,i} - U_{s,i}][U_{r,i} - \mu_r]\right) = o(n^{-\kappa/2})$ and part (iii) follows directly from (76).

T1.3 Proof of Theorem 3

Consider the random variables $\chi_{s,n} = n^{-1}\sum_{i=1}^{n}(\widehat{U}_{s,i} - U_{s,i}) = \widehat{\widetilde{\mu}}_{s,n} - \widehat{\mu}_{s,n}$ $(s = 1, \ldots, 4)$. Part (i) follows directly from Theorem 2(i) as follows.

$$\widetilde{\mu}_{s,n} - \mu_s = E(\chi_{s,n}) = n^{-1}\sum_{i=1}^{n} E(\widehat{U}_{s,i} - U_{s,i}) = C_s n^{-\kappa} + o(n^{-\kappa}). \tag{77}$$

Now from Theorem 1 we have

$$\begin{aligned} Var(\chi_{s,n}) &= n^{-2}\sum_{i=1}^{n} Var(\widehat{U}_{s,i} - U_{s,i}) = n^{-2}\sum_{i=1}^{n}\left(E\left([\widehat{U}_{s,i} - U_{s,i}]^2\right) - \left(E(\widehat{U}_{s,i} - U_{s,i})\right)^2\right) \\ &= n^{-1}\left(o(n^{-\kappa}) - (C_s n^{-\kappa} + o(n^{-\kappa}))^2\right) = n^{-1}o(n^{-\kappa}). \end{aligned} \tag{78}$$

By Markov's inequality (see, e.g., Van der Vaart, 2000, page 10), we have that for any $\epsilon > 0$,

$$Pr(\sqrt{n}|\chi_{s,n} - E(\chi_{s,n})| > \epsilon) \leq \frac{E\left(n(\chi_{s,n} - E(\chi_{s,n}))^2\right)}{\epsilon^2} = \frac{nVar(\chi_{s,n})}{\epsilon^2} = \frac{o(n^{-\kappa})}{\epsilon^2}. \tag{79}$$

Therefore, $\sqrt{n}(\chi_{s,n} - E(\chi_{s,n})) = o_p(1)$ or equivalently, $\chi_{s,n} - E(\chi_{s,n}) = o_p(n^{-1/2})$. Consequently,

$$(\widehat{\widetilde{\mu}}_{s,n} - \widetilde{\mu}_{s,n}) - (\widehat{\mu}_{s,n} - \mu_s) = (\widehat{\widetilde{\mu}}_{s,n} - \widehat{\mu}_{s,n}) - (\widetilde{\mu}_{s,n} - \mu_s) = \chi_{s,n} - E(\chi_{s,n}) = o_p(n^{-1/2}), \tag{80}$$

implying (ii).

To prove part (iii), we recall from (14) that

$$\sqrt{n}(\widehat{\mu}_{s,n} - \mu_s) \xrightarrow{d} \mathcal{N}(0, \sigma_{ss}), \quad s = 1, \ldots, 6. \tag{81}$$

Combining this with (ii), we have

$$\sqrt{n}(\widehat{\widehat{\mu}}_{s,n} - \widetilde{\mu}_{s,n}) = \sqrt{n}(\widehat{\mu}_{s,n} - \mu_s + o_p(n^{-1/2})) = \sqrt{n}(\widehat{\mu}_{s,n} - \mu_s) + o_p(1) \xrightarrow{d} \mathcal{N}(0, \sigma_{ss}), \tag{82}$$

and hence, (iii) is proved.

Now it follows from parts (i) and (iii) that $\widehat{\widehat{\mu}}_{s,n} \xrightarrow{p} \mu_s$ for $s = 1, \ldots, 4$. Combining this with Theorem 2 and using the standard central limit theorem, we have that as $n \to \infty$,

$$\begin{aligned}
\widehat{\widehat{\sigma}}_{st,n} &= n^{-1} \sum_{i=1}^{n} (\widehat{U}_{s,i} - \widehat{\widehat{\mu}}_{s,n})(\widehat{U}_{t,i} - \widehat{\widehat{\mu}}_{t,n}) = n^{-1} \sum_{i=1}^{n} \widehat{U}_{s,i}\widehat{U}_{t,i} - \widehat{\widehat{\mu}}_{s,n}\widehat{\widehat{\mu}}_{t,n} \\
&\xrightarrow{p} E(\widehat{U}_{s,1}\widehat{U}_{t,1}) - \mu_s\mu_t = Cov(\widehat{U}_{s,1}, \widehat{U}_{t,1}) + E(\widehat{U}_{s,1})E(\widehat{U}_{t,1}) - \mu_s\mu_t \\
&= (\sigma_{st} + o(n^{-\kappa/2})) + (\mu_s + C_s n^{-\kappa} + O(v_{n,\kappa}))(\mu_t + C_t n^{-\kappa} + O(v_{n,\kappa})) - \mu_s\mu_t \\
&\longrightarrow \sigma_{st},
\end{aligned} \tag{83}$$

which implies part (iv). Part (v) can be proved by using similar arguments as those used for part (iv); meanwhile part (vi) is a well-known result as mentioned in (20).

T1.4 Proof of Theorem 4

In light of Theorem 3, it is clear that $\widehat{\widehat{V}}_{\xi,n}$ is obtained by replacing unknown parameters in their formulas by the corresponding consistent estimates. Hence, by continuous mapping theorem and Slutsky's theorem, we have the desired results.

T1.5 Proof of Theorem 5

For each $s = 1, 2, 3, 4$, from Theorem 3(i), we have $\lim_{n\to\infty} \widetilde{\mu}_{s,n} = \mu_s > 0$. Hence, we can apply Lemma 2(i) to the result of Theorem 3(iii) and obtain

$$\begin{aligned}
\log \widehat{\widehat{\mu}}_{s,n} &= \log \widetilde{\mu}_{s,n} + \frac{1}{\mu_s}(\widehat{\widehat{\mu}}_{s,n} - \widetilde{\mu}_{s,n}) + o_p(n^{-1/2}) \\
&= \log(\mu_s + C_s n^{-\kappa} + O(v_{n,\kappa})) + \frac{\widehat{\widehat{\mu}}_{s,n} - \widetilde{\mu}_{s,n}}{\mu_s} + o_p(n^{-1/2})
\end{aligned}$$

$$= \log \mu_s + \frac{1}{\mu_s}[C_s n^{-\kappa} + O(v_{n,\kappa})] + O([C_s n^{-\kappa} + O(v_{n,\kappa})]^2) + \frac{\widehat{\mu}_{s,n} - \mu_s + o_p(n^{-1/2})}{\mu_s} + o_p(n^{-1/2})$$
$$= \log \mu_s + \frac{C_s}{\mu_s} n^{-\kappa} + O(v_{n,\kappa}) + \frac{\widehat{\mu}_{s,n} - \mu_s}{\mu_s} + o_p(n^{-1/2}). \quad (84)$$

In the above expression, the third equality follows from Taylor expansion and Theorem 3(ii), the last equality follows from $O([C_s n^{-\kappa} + O(v_{n,\kappa})]^2) = O(n^{-2\kappa}) = o(v_{n,\kappa})$. Hereafter, we will use this reasoning in subsequent proofs immediately to save space.

On the other hand, we can also apply Lemma 2(i) to $\sqrt{n}(\widehat{\mu}_{s,n} - \mu_s) \xrightarrow{d} \mathcal{N}(0, \sigma_{ss})$ and obtain

$$\log \widehat{\mu}_{s,n} = \log \mu_s + \frac{\widehat{\mu}_{s,n} - \mu_s}{\mu_s} + o_p(n^{-1/2}). \quad (85)$$

Note that (85) is also valid for $s = 5, 6$ since $\mu_5, \mu_6 > 0$. From (84) and (85), we have

$$\widehat{\widehat{\xi}}_n = \xi + \frac{1}{2}\left(\frac{C_3}{\mu_3} + \frac{C_4}{\mu_4} - \frac{C_1}{\mu_1} - \frac{C_2}{\mu_2}\right) n^{-\kappa} + O(v_{n,\kappa}) + o_p(n^{-1/2}) + \widehat{B}_{\xi,n}, \quad (86)$$
$$\widehat{\xi}_n = \xi + o_p(n^{-1/2}) + \widehat{B}_{\xi,n}, \quad (87)$$

where

$$\widehat{B}_{\xi,n} = \frac{\widehat{\mu}_{3,n} - \mu_3}{2\mu_3} + \frac{\widehat{\mu}_{4,n} - \mu_4}{2\mu_4} - \frac{\widehat{\mu}_{1,n} - \mu_1}{2\mu_1} - \frac{\widehat{\mu}_{2,n} - \mu_2}{2\mu_2} + \frac{\widehat{\mu}_{5,n} - \mu_5}{\mu_5} - \frac{\widehat{\mu}_{6,n} - \mu_6}{\mu_6}. \quad (88)$$

Subtracting (87) from (86) yields

$$\widehat{\widehat{\xi}}_n - \widehat{\xi}_n = C_\xi n^{-\kappa} + O(v_{n,\kappa}) + o_p(n^{-1/2}), \quad (89)$$

where $C_\xi = \frac{1}{2}(\frac{C_3}{\mu_3} + \frac{C_4}{\mu_4} - \frac{C_1}{\mu_1} - \frac{C_2}{\mu_2}) \in \mathbb{R}$ is a constant.

Now combining (17) with (89), we have the desired result.

T1.6 Proof of Theorem 6

For convenience, we can assume that the observations in $\mathcal{X}_n$ are randomly sorted and $\mathcal{X}_{n_\kappa}^*$ consists of the first n_κ elements of the sorted sample. Before going into details of the

proof, we establish some asymptotic properties similar to Theorem 3 as follows.

For $s = 1, \ldots, 4$, let $\chi_{s,n_\kappa} = \widehat{\widehat{\mu}}_{s,n_\kappa} - \widehat{\mu}_{s,n_\kappa}$ where $\widehat{\widehat{\mu}}_{s,n_\kappa} = n_\kappa^{-1} \sum_{i=1}^{n_\kappa} \widehat{U}_{s,i}$ and $\widehat{\mu}_{s,n_\kappa} = n_\kappa^{-1} \sum_{i=1}^{n_\kappa} U_{s,i}$. Remember that DEA estimation in $\widehat{U}_{s,i}$ here is the same as before, i.e., using all observations in the original sample $\mathcal{X}_n$. In addition, let $\widetilde{\mu}_{s,n_\kappa} = E(\widehat{\widehat{\mu}}_{s,n_\kappa})$. Then by Theorem 1, we have

$$E(\chi_{s,n_\kappa}) = \widetilde{\mu}_{s,n_\kappa} - \mu_s = n_\kappa^{-1} \sum_{i=1}^{n_\kappa} E(\widehat{U}_{s,i} - U_{s,i}) = C_s n^{-\kappa} + O(v_{n,\kappa}), \tag{90}$$

and

$$\begin{aligned} Var(\chi_{s,n_\kappa}) &= n_\kappa^{-2} \sum_{i=1}^{n_\kappa} Var(\widehat{U}_{s,i} - U_{s,i}) = n_\kappa^{-2} \sum_{i=1}^{n_\kappa} \left(E\left([\widehat{U}_{s,i} - U_{s,i}]^2 \right) - \left(E(\widehat{U}_{s,i} - U_{s,i}) \right)^2 \right) \\ &= n_\kappa^{-1} \left(o(n^{-\kappa}) - (C_s n^{-\kappa} + O(v_{n,\kappa}))^2 \right) = n_\kappa^{-1} o(n^{-\kappa}), \end{aligned} \tag{91}$$

where C_s is the same constant as in Theorem 3 $(s = 1, \ldots, 4)$.

Consequently, by Markov inequality, for any $\epsilon > 0$, we have

$$Pr\left(\sqrt{n_\kappa} |\chi_{s,n_\kappa} - E(\chi_{s,n_\kappa})| > \epsilon\right) \leq \frac{E(n_\kappa (\chi_{s,n_\kappa} - E(\chi_{s,n_\kappa}))^2)}{\epsilon^2} = \frac{n_\kappa Var(\chi_{s,n_\kappa})}{\epsilon^2} = \frac{o(n^{-\kappa})}{\epsilon^2},$$

which implies that $\chi_{s,n_\kappa} - E(\chi_{s,n_\kappa}) = o_p(n_\kappa^{-1/2})$, or equivalently,

$$\widehat{\widehat{\mu}}_{s,n_\kappa} - \widetilde{\mu}_{s,n_\kappa} = \widehat{\mu}_{s,n_\kappa} - \mu_s + o_p(n_\kappa^{-1/2}), \tag{92}$$

and as a consequence,

$$\sqrt{n_\kappa} \left(\widehat{\widehat{\mu}}_{s,n_\kappa} - \widetilde{\mu}_{s,n_\kappa} \right) \xrightarrow{d} \mathcal{N}(0, \sigma_{ss}). \tag{93}$$

By virtue of (90), the equality above can also be presented as

$$\widehat{\widehat{\mu}}_{s,n_\kappa} - \widehat{\mu}_{s,n_\kappa} = \widetilde{\mu}_{s,n_\kappa} - \mu_s + o_p(n_\kappa^{-1/2}) = C_s n^{-\kappa} + O(v_{n,\kappa}) + o_p(n_\kappa^{-1/2}), \quad s = 1, \ldots, 4. \tag{94}$$

Since $\kappa \leq 1/2$, $n_\kappa = \lfloor n^{2\kappa} \rfloor \leq n$ and $\lim_{n\to\infty} \frac{n^\kappa}{\sqrt{n_\kappa}} = \lim_{n\to\infty} \frac{n^\kappa}{\sqrt{\lfloor n^{2\kappa} \rfloor}} = 1$, it is sufficient to prove that

$$\sqrt{n_\kappa}\left(\widehat{\widehat{\xi}}_{n_\kappa} - \xi - C_\xi n^{-\kappa} + O(v_{n,\kappa})\right) \xrightarrow{d} \mathcal{N}(0, V_\xi), \text{ as } n \to \infty. \tag{95}$$

From (90) we have $\lim_{n_\kappa\to\infty} \widetilde{\mu}_{s,n_\kappa} = \mu_s > 0$ for $s = 1, 2, 3, 4$. Thus, similar to the proof of Theorem 5(i), we can apply Lemma 2(i) to (93) and then use Taylor expansion to get the following result:

$$\begin{aligned}
\log \widehat{\widehat{\mu}}_{s,n_\kappa} &= \log \widetilde{\mu}_{s,n_\kappa} + \frac{1}{\mu_s}(\widehat{\widehat{\mu}}_{s,n_\kappa} - \widetilde{\mu}_{s,n_\kappa}) + o_p(n_\kappa^{-1/2}) \\
&= \log(\mu_s + C_s n^{-\kappa} + O(v_{n,\kappa})) + \frac{\widehat{\mu}_{s,n_\kappa} - \mu_s + o_p(n_\kappa^{-1/2})}{\mu_s} + o_p(n_\kappa^{-1/2}) \\
&= \log \mu_s + \frac{1}{\mu_s}(C_s n^{-\kappa} + O(v_{n,\kappa})) + O([C_s n^{-\kappa} + O(v_{n,\kappa})]^2) + \frac{\widehat{\mu}_{s,n_\kappa} - \mu_s}{\mu_s} + o_p(n_\kappa^{-1/2}) \\
&= \log \mu_s + \frac{C_s}{\mu_s} n^{-\kappa} + O(v_{n,\kappa}) + \frac{\widehat{\mu}_{s,n_\kappa} - \mu_s}{\mu_s} + o_p(n_\kappa^{-1/2}).
\end{aligned} \tag{96}$$

On the other hand, it can be deduced from (85) that

$$\log \widehat{\mu}_{s,n_\kappa} = \log \mu_s + \frac{\widehat{\mu}_{s,n_\kappa} - \mu_s}{\mu_s} + o_p(n_\kappa^{-1/2}). \tag{97}$$

Subtracting (97) from (96) yields

$$\log \widehat{\widehat{\mu}}_{s,n_\kappa} - \log \widehat{\mu}_{s,n_\kappa} = \frac{C_s}{\mu_s} n^{-\kappa} + O(v_{n,\kappa}) + o_p(n_\kappa^{-1/2}) \quad (s = 1, 2, 3, 4). \tag{98}$$

Therefore, we can obtain

$$\widehat{\widehat{\xi}}_{n_\kappa} - \widehat{\xi}_{n_\kappa} = C_\xi n^{-\kappa} + O(v_{n,\kappa}) + o_p(n_\kappa^{-1/2}), \tag{99}$$

where $C_\xi = \frac{1}{2}(\frac{C_3}{\mu_3} + \frac{C_4}{\mu_4} - \frac{C_1}{\mu_1} - \frac{C_2}{\mu_2})$ is the same constant as in Theorem 5 and $\widehat{\xi}_{n_\kappa}$ is the analogue of $\widehat{\xi}_n$ but computed using the subsample $\mathcal{X}^*_{n_\kappa}$ (note that there is no DEA estimation in $\widehat{\xi}_{n_\kappa}$). Clearly, this result when combined with $\sqrt{n_\kappa}(\widehat{\xi}_{n_\kappa} - \xi) \xrightarrow{d} \mathcal{N}(0, V_\xi)$ as $n_\kappa \to \infty$ will lead to the desired result.

T1.7 Proof of Theorem 7

If n is odd, $\dfrac{m_1}{m_2} = \dfrac{\lfloor n/2 \rfloor}{\lfloor n/2 \rfloor + 1} \to 1$ as $n \to \infty$. Thus, for simplicity, we can assume without affecting the asymptotical result of the theorem that n is even and $m_1 = m_2 = n/2$.

For each $l = 1, \ldots, L$, $s = 1, 2, 3, 4$ and $j = 1, 2$, it follows by the same arguments in the proof of Theorem 5(i) that

$$\begin{aligned}\widehat{\widehat{\xi}}_{l,m_j} &= \xi + C_\xi (n/2)^{-\kappa} + O(v_{n/2,\kappa}) + \widehat{B}_{\xi,l,m_j} + o_p(n^{-1/2}) \\ &= \xi + 2^\kappa C_\xi n^{-\kappa} + O(v_{n,\kappa}) + \widehat{B}_{\xi,l,m_j} + o_p(n^{-1/2}),\end{aligned} \tag{100}$$

where $\widehat{B}_{\xi,l,m_j}$ is the analogue of $\widehat{B}_{\xi,n}$ in the sense that components involving efficiency scores are evaluated over the subsample $\mathcal{X}_{l,m_j}$ whereas the other components (i.e., $\widehat{\mu}_{5,n}$ and $\widehat{\mu}_{6,n}$) are evaluated over the full sample $\mathcal{X}_n$. Formally,

$$\widehat{B}_{\xi,l,m_j} = \frac{\widehat{\mu}_{l,3,m_j} - \mu_3}{2\mu_3} + \frac{\widehat{\mu}_{l,4,m_j} - \mu_4}{2\mu_4} - \frac{\widehat{\mu}_{l,1,m_j} - \mu_1}{2\mu_1} - \frac{\widehat{\mu}_{l,2,m_j} - \mu_2}{2\mu_2} + \frac{\widehat{\mu}_{5,n} - \mu_5}{\mu_5} - \frac{\widehat{\mu}_{6,n} - \mu_6}{\mu_6},$$

where $\widehat{\mu}_{l,s,m_j}$ is the analogue of $\widehat{\mu}_{s,n}$ but evaluated over the subsample $\mathcal{X}_{l,m_j}$. Note also that in (100) and in some places below, we use the fact that $o_p((n/2)^{-1/2}) = o_p(n^{-1/2})$ and $O(v_{n/2,\kappa}) = O(v_{n,\kappa})$, which is due to

$$\lim_{n\to\infty} \frac{v_{n/2,\kappa}}{v_{n,\kappa}} = \lim_{n\to\infty} \frac{\left(\frac{n}{2}\right)^{-\frac{3}{p+q+1}} \left(\log \frac{n}{2}\right)^{\frac{3}{p+q+1}}}{n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}} = \lim_{n\to\infty} 2^{\frac{3}{p+q+1}} \left(1 - \frac{\log 2}{\log n}\right)^{\frac{3}{p+q+1}} = 2^{\frac{3}{p+q+1}}. \tag{101}$$

Taking the average of (100) over $j = 1, 2$, we can come up with

$$\widehat{\widehat{\xi}}^*_{l,n} = \xi + 2^\kappa C_\xi n^{-\kappa} + O(v_{n,\kappa}) + \frac{1}{2}(\widehat{B}_{\xi,l,m_1} + \widehat{B}_{\xi,l,m_2}) + o_p(n^{-1/2}). \tag{102}$$

Subtracting (86) from the above equality yields

$$\begin{aligned}\widehat{\widehat{\xi}}^*_{l,n} - \widehat{\widehat{\xi}}_n &= (2^\kappa - 1) C_\xi n^{-\kappa} + O(v_{n,\kappa}) + \left[\frac{1}{2}(\widehat{B}_{\xi,l,m_1} + \widehat{B}_{\xi,l,m_2}) - \widehat{B}_{\xi,n}\right] + o_p(n^{-1/2}) \\ &= (2^\kappa - 1) C_\xi n^{-\kappa} + O(v_{n,\kappa}) + o_p(n^{-1/2}).\end{aligned} \tag{103}$$

In the above expression, the second equality follows from $\frac{1}{2}(\widehat{B}_{\xi,l,m_1}+\widehat{B}_{\xi,l,m_2})-\widehat{B}_{\xi,n}=0$, which is due to the fact that for $s=1,\ldots,6$,

$$\begin{aligned}\widehat{\mu}_{s,n}-\frac{1}{2}(\widehat{\mu}_{l,s,m_1}+\widehat{\mu}_{l,s,m_2}) &= \frac{1}{n}\sum_{i=1}^{n}U_{s,i}-\frac{1}{2}\left(\frac{1}{n/2}\sum_{\{i:Z_i^t\in\mathcal{X}_{l,m_1}\}}U_{s,i}+\frac{1}{n/2}\sum_{\{i:Z_i^t\in\mathcal{X}_{l,m_2}\}}U_{s,i}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}U_{s,i}-\frac{1}{n}\sum_{i=1}^{n}U_{s,i}=0. \end{aligned} \tag{104}$$

Finally, taking the average of (103) over $l=1,\ldots,L$ will lead to the desired result.

T1.8 Proof of Theorem 8

This theorem follows from substituting the results from Theorem 7 into the corresponding ones in Theorem 5 while noting that $\sqrt{n}o_p(n^{-1/2})=o_p(1)$ and $n^{\kappa}o_p(n^{-1/2})=o_p(1)$ when $\kappa<1/2$.

T2 Other technical details

T2.1 A procedure to generate correlated random numbers for Monte Carlo experiments

Here we describe how we generate correlated random numbers in our Monte Carlo simulations. Basically, the method is inspired by Kneip et al. (2018). Assume that we want to generate a dataset of n observations from random variables $V_j^{(t)}$ $(j=1,\ldots,p;t=1,2)$, where $V_j^{(t)}\sim\text{Uniform}(d_{j,min}^{(t)},d_{j,max}^{(t)})$ and $corr(V_j^{(1)},V_j^{(2)})=\rho$ for all j. The details are as follows.

1. Construct the matrix

$$U=\begin{bmatrix} I_p & \rho I_p \\ 0 & \sqrt{1-\rho^2}I_p \end{bmatrix} \tag{105}$$

where I_p is the identity matrix of order p. It can be seen that $U'U = C$ where

$$C = \begin{bmatrix} I_p & \rho I_p \\ \rho I_p & I_p \end{bmatrix}. \tag{106}$$

2. Generate an $n \times 2p$ matrix R of iid $\mathcal{N}(0,1)$.

3. Compute the matrix $D = \Phi(RU)$ where $\Phi(.)$ is the cdf of the standard normal distribution. Partition D into $D = [D^{(1)}, D^{(2)}]$, where $D^{(t)} = [D_1^{(t)}, \ldots, D_p^{(t)}]$ for $t = 1, 2$. Then it is clear that $D_j^{(t)}$ follows uniform distribution on $(0,1)$ and $corr(D_j^{(1)}, D_j^{(2)}) = \rho$ $(j = 1, \ldots, p; t = 1, 2)$.

4. For each $j = 1, \ldots, p$, transform $D_j^{(t)}$ to $V_j^{(t)} = (d_{j,max}^{(t)} - d_{j,min}^{(t)})D_j^{(t)} + d_{j,min}^{(t)}$. We then have numbers $V_j^{(t)}$ follow uniform distribution on $(d_{j,min}^{(t)}, d_{j,max}^{(t)})$ and $corr(V_j^{(1)}, V_j^{(2)}) = \rho$.

5. If we want to change the distribution of $V_j^{(t)}$ for some t and j to a new one other than uniform distribution (e.g., half normal distribution plus unity: $1 + |N(0, \sigma^2)|$), we just simply need to replace the transformation in step 4 by $V_j^{(t)} = F^{-1}(D_j^{(t)})$ where F^{-1} is the quantile function of the desired distribution. By this way, $corr(V_j^{(1)}, V_j^{(2)})$ might not be exactly ρ but is determined via a function of ρ which in turn depends on the new distribution. However, this does not raise any issue since our purpose in this paper is to ensure that the numbers generated are correlated and we can control the degree of correlation via the parameter ρ.

T2.2 A uniform delta method (Van der Vaart, 2000)

We reproduce Theorem 3.8 of Van der Vaart (2000) here with a small change of the notation and presentation for consistency throughout this paper and for convenience in the application in the proofs of our theorems.

Lemma 2. *Let $\Theta : \mathbb{R}^k \to \mathbb{R}^m$ be a continuously differentiable function in a neighborhood of $h \in \mathbb{R}^k$. Let H_n be random vectors taking their values in the domain of Θ. If $\vartheta_n(H_n -$*

$h_n) \xrightarrow{d} H$ *for vectors* $h_n \to h$ *and numbers* $\vartheta_n \to \infty$*, then we have*

$$\text{(i) } \Theta(H_n) - \Theta(h_n) = \nabla\Theta(h)(H_n - h_n) + o_p(\vartheta_n^{-1}), \tag{107}$$

$$\text{(ii) } \vartheta_n \left(\Theta(H_n) - \Theta(h_n)\right) \xrightarrow{d} \nabla\Theta(h)H, \tag{108}$$

where $\nabla\Theta(h)$ *denotes the* $(m \times k$ *matrix) gradients of* $\Theta(\cdot)$ *evaluated at* h*.*

Proof. See Theorem 3.8 of Van der Vaart (2000). □

T2.3 Computing the true parameters of interest

Similar to KSW2018, we only know the efficiency scores measured toward the true underlying production technology sets (i.e., $\lambda(.|\Psi^t)$). Meanwhile those measured toward the conical hull of the production technology sets (i.e., $\lambda_C(.|\Psi^t)$) are unknown (see step 3 of the data generating process mentioned in Section 9.1). We show how to obtain these figures below.

Assume that we have to measure the efficiency of a DMU represented by (x, y) toward $\mathcal{C}(\Psi^t)$ $(t = 1, 2)$. Taking advantage of the fact that $y \in \mathbb{R}^1_+$, we can transform the efficiency score as follows.

$$\begin{aligned}
\lambda_C(x, y|\Psi^t) &= \sup_\lambda\{\lambda : (x, \lambda y) \in \mathcal{C}(\Psi^t)\} = \sup_\lambda\{\lambda : (x, \lambda y) = (a\tilde{x}, a\tilde{y}), (\tilde{x}, \tilde{y}) \in \Psi^t, a \in \mathbb{R}^1_+\} \\
&= \sup_\lambda\{\lambda : (x, \lambda y) = (a\tilde{x}, a\tilde{y}), \tilde{y} \le \psi^t(\tilde{x}), \tilde{x} \in \mathbb{R}^p_+, a \in \mathbb{R}^1_+\} \\
&= \sup_\lambda\{\lambda : (x, \lambda y) = (a\tilde{x}, a\psi^t(\tilde{x})), \tilde{x} \in \mathbb{R}^p_+, a \in \mathbb{R}^1_+\} \\
&= \sup_{a,\tilde{x}}\left\{\frac{a\psi^t(\tilde{x})}{y} : a\tilde{x} = x, \tilde{x} \in \mathbb{R}^p_+, a \in \mathbb{R}^1_+\right\} = \sup_a\left\{\frac{a\psi^t(x/a)}{y} : a \in \mathbb{R}^1_+\right\} \\
&= \frac{1}{y}\max_{a>0}\left(a\psi^t\left(\frac{x}{a}\right)\right).
\end{aligned} \tag{109}$$

Now one can apply available optimization solvers (e.g., "fmincon" in MATLAB®) to solve for $\max_{a>0}\left(a\psi^t(\frac{x}{a})\right)$ and then obtain $\lambda_C(x, y|\Psi^t)$. It is noteworthy that there is a potential risk of receiving a local optimum instead of the global one. Our further analysis below will demonstrate that the global optimum will be achieved.

We fix (x, y) and consider the function

$$\varphi(a) \coloneqq a\psi^t(x/a) = a\Upsilon^t(x_1/a - b_1^t)^{\beta_1^t}(x_2/a - b_2^t)^{\beta_2^t} \dots (x_p/a - b_p^t)^{\beta_p^t} \tag{110}$$

with the domain $\mathbb{D} = \left(0, \min_{j=1,\dots,p} \frac{x_j}{b_j^t}\right)$. Its first derivative is presented below.

$$\begin{aligned}
\frac{d\varphi}{da} &= \Upsilon^t(x_1/a - b_1^t)^{\beta_1^t}(x_2/a - b_2^t)^{\beta_2^t} \dots (x_p/a - b_p^t)^{\beta_p^t} \\
&\quad - a\Upsilon^t \left[\beta_1^t(x_1/a - b_1^t)^{\beta_1^t - 1} x_1 a^{-2}\right] (x_2/a - b_2^t)^{\beta_2^t} \dots (x_p/a - b_p^t)^{\beta_p^t} \\
&\quad - a\Upsilon^t(x_1/a - b_1^t)^{\beta_1^t} \left[\beta_2^t(x_2/a - b_2^t)^{\beta_2^t - 1} x_2 a^{-2}\right] \dots (x_p/a - b_p^t)^{\beta_p^t} \dots \\
&\quad - a\Upsilon^t(x_1/a - b_1^t)^{\beta_1^t}(x_2/a - b_2^t)^{\beta_2^t} \dots \left[\beta_p^t(x_p/a - b_p^t)^{\beta_p^t - 1} x_p a^{-2}\right] \\
&= \Upsilon^t(x_1/a - b_1^t)^{\beta_1^t}(x_2/a - b_2^t)^{\beta_2^t} \dots (x_p/a - b_p^t)^{\beta_p^t} \left(1 - \frac{x_1\beta_1^t}{a(x_1/a - b_1^t)} - \dots - \frac{x_p\beta_p^t}{a(x_p/a - b_p^t)}\right) \\
&= \psi^t(x/a) \left(1 - \sum_{j=1}^{p} \frac{x_j\beta_j^t}{x_j - ab_j^t}\right) = \psi^t(x/a)\iota(a),
\end{aligned} \tag{111}$$

where

$$\iota(a) = 1 - \sum_{j=1}^{p} \frac{x_j\beta_j^t}{x_j - ab_j^t}. \tag{112}$$

Therefore, the sign of $d\varphi/da$ is the same as that of $\iota(a)$ and furthermore, $d\varphi/da = 0$ if and only if $\iota(a) = 0$. Since $\frac{d\iota}{da} = -\sum_{j=1}^{p} \frac{x_j\beta_j^t b_j^t}{(x_j - ab_j^t)^2} < 0 \; \forall a > 0$, $\iota(a)$ is strictly decreasing. In addition,

$$\lim_{a \to 0^+} \iota(a) = 1 - \sum_{j=1}^{p} \beta_j^t > 0, \quad \lim_{a \to \frac{x_j}{b_j^t}^-} \iota(a) = -\infty, \quad \lim_{a \to \frac{x_j}{b_j^t}^+} \iota(a) = +\infty, \quad \lim_{a \to +\infty} \iota(a) = 1 > 0.$$

Hence, $\iota(a) = 0$ has only one solution in the domain $\mathbb{D}$, say a^*. Moreover, the above findings show that $\iota(a)$ changes sign from positive to negative when passing a^* in the left-right direction, and hence, so is $d\varphi/da$. This implies that a^* is the global maximum of $\varphi(a)$ in the domain $\mathbb{D}$ as desired.

The true values of the parameter ξ are reported in Table 10. Note that we set $\xi = 0$

Table 10: True values of parameter ξ.

δ	$p=1,q=1$	$p=2,q=1$	$p=3,q=1$	$p=4,q=1$
0.00	0.00000	0.00000	0.00000	0.00000
0.01	0.02528	0.03916	0.05088	0.06385
0.02	0.05050	0.07820	0.10189	0.12791
0.03	0.07567	0.11745	0.15292	0.19215
0.04	0.10079	0.15663	0.20415	0.25657

$n = 10 \times 10^6$, $\rho = \rho_\lambda = 0.5$.

for $\delta = 0$ since the production frontier functions $\psi^1(x)$ and $\psi^2(x)$ are identical here and hence there is no productivity change. Prior Monte Carlo simulations (with $n = 10 \times 10^6$) also justify this point with some acceptable noises due to finite n, e.g., $\xi = 7 \times 10^{-8}$ for $p = 1$.