UCL Université catholique de Louvain

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

ISBA

DISCUSSION PAPER

2018/10

Inference in Dynamic, Nonparametric Models of Production: Central Limit Theorems for Malmquist Indices

Kneip A., Simar, L. and P. Wilson

Inference in Dynamic, Nonparametric Models of Production: Central Limit Theorems for Malmquist Indices

Alois Kneip Léopold Simar Paul W. Wilson*

April 2018

Abstract

The Malmquist index gives a measure of productivity in dynamic settings and has been widely applied in empirical work. The index is typically estimated using envelopment estimators, particularly data envelopment analysis (DEA) estimators. Until now, inference about productivity change measured by Malmquist indices has been problematic, including both inference regarding productivity change experienced by particular firms as well as mean productivity change. This paper establishes properties of a DEA-type estimator of distance to the conical hull of a variable-returns-to-scale production frontier. In addition, properties of DEA estimators of Malmquist indices for individual producers are derived as well properties of geometric means of these estimators. The latter requires new CLT results, extending the work of Kneip et al. (2015, *Econometric Theory*). Simulation results are provided to give applied researchers an idea of how well inference may work in practice in finite samples.

Keywords: asymptotic, DEA, hypothesis test, inference, Malmquist index, productivity change.

*Kneip: Institut für Finanzmarktökonomie und Statistik, Fachbereich Wirtschaftswissenschaften, Universität Bonn, Bonn, Germany; email: akneip@uni-bonn.de.

Simar: Institut de Statistique, Biostatistique, et Sciences Actuarielles, Université Catholique de Louvain-la-Neuve, Louvain-la-Neuve, Belgium; email: leopold.simar@uclouvain.be.

Wilson: Department of Economics and School of Computing, Clemson University, Clemson, South Carolina, USA; email: pww@clemson.edu.

Technical support from the Cyber Infrastructure Technology Integration group at Clemson University are gratefully acknowledged. Any remaining errors are solely our responsibility.

1 Introduction

Malmquist indices for measurement of productivity change in dynamic contexts are based on the quantity index of Malmquist (1953), with adaptation to the production framework by Caves et al. (1982) and Nishimizu and Page (1982). Malmquist indices are widely used to measure productivity change over time, and are often estimated using nonparametric, data envelopment analysis (DEA) estimators due to the work of Färe et al. (1992).[1] Applied researchers typically report geometric means of estimates of Malmquist indices, and sometimes report estimates of Malmquist indices for individual producers.

Most papers appearing in the literature presenting empirical estimates of Malmquist indices make no attempt at inference. The few that attempt inference either rely on standard central limit theorem (CLT) results (e.g., the Lindeberg-Feller CLT) or the bootstrap method proposed by Simar and Wilson (1999). Inference relying on standard CLT results is invalid for cases with more than one input and one output as shown below due to reasons similar to those given by Kneip et al. (2015) in the context of mean efficiency in a cross-sectional setting. Alternatively, Simar and Wilson (1999) provide only heuristic arguments to develop their bootstrap method without any theoretical results. Although the simulation evidence provided by Simar and Wilson seems to indicate that the smooth bootstrap on which the method of Simar and Wilson (1999) is based works well, the approach cannot be justified from a theoretical viewpoint due to the results obtained below.

This paper provides the theoretical developments needed to make inference about productivity change measured by Malmquist indices, both in the case of individual firms and the case of geometric means over a sample of firms. In particular, the properties of a Farrell (1957)-type estimator of distance to the boundary of the convex cone of the production set are developed. Additional results are then developed to permit the sub-sampling approach of Simar and Wilson (2011) to be used to make inference about the Malmquist index for a single firm. We also provide new CLTs that permit inference about geometric means of Malmquist indices across multiple firms. These in turn can be used to make statistical tests regarding differences in mean productivity change across groups of firms, along the lines of Kneip et al. (2016).

[1] A search on 5 February 2018 using Google Scholar and the keywords "Malmquist," "index" and "productivity" and excluding patents or citations yields approximately 15,400 papers and books.

Färe et al. (1992) explicitly assume constant returns to scale, as do others, and estimate their Malmquist index while imposing constant returns to scale on their nonparametric estimators. Grifell-Tatjé and Lovell (1995) note that others estimate Malmquist indices while allowing for variable returns to scale, and discuss the consequences of doing so. In fact, whether the underlying technology exhibits constant or variable returns to scale is a red herring; what is required is that the Malmquist index must be defined in terms of the convex cone of the production set in order to be properly interpreted as a measure of productivity change. At a given point in time, the convexity test developed by Kneip et al. (2016) can be used to test constant returns to scale versus variable returns to scale.

Malmquist indices present several problems for statistical inference. First, Malmquist indices for individual firms observed at times $t \in \{1,\ 2\}$ are defined in terms of a geometric mean of two *ratios* of Farrell (1957) efficiency measures. Even if the underlying technology exhibits global constant returns to scale, one cannot simply replace the *true* efficiency measures with corresponding DEA estimates (imposing constant returns to scale) and then rely on the results of Park et al. (2010) and Kneip et al. (2015) to make inference due to the definition in terms of ratios and a geometric mean. Second, if the underlying technology exhibits variable returns to scale, there are no results (to date) on estimates of measures of distance from an observed point in input-output space to the boundary of the convex cone of the variable-returns-to-scale production set. Third, if the applied researcher wants to report the geometric mean of estimated estimated Malmquist indices over the firms in his sample, neither standard CLTs nor the CLTs for sample means of nonparametric efficiency estimates developed by Kneip et al. (2015) can be used for inference.

These issues are addressed below as follows. Section 2 defines notation, lays out a statistical model and introduces the relevant estimators. Section 3 provides the theoretical developments, first for the static case and then for the dynamic case. These results are then used in Section 4 to provide methods for making inferences about productivity change and for testing hypotheses about productivity change. Simulation results showing how well the methods for inference can be expected to perform are given in Section 5, while Section 6 gives a summary and conclusions. Proofs and technical details appear in Appendix A.

2 A Dynamic, Nonparametric Production Process

2.1 A Statistical Model

In order to establish notation, let $x \in \mathbb{R}^p_+$ and $y \in \mathbb{R}^q_+$ be vectors of fixed input and output quantities. Throughout, vectors are assumed to be column-vectors, as opposed to row-vectors. At time t, the set of feasible combinations of inputs and outputs is given by

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

The *technology*, or *efficient frontier* of Ψ^t, is given by

$$\Psi^{t\partial} := \left\{(x, y) \mid (x, y) \in \Psi^t,\ (\gamma x, \gamma^{-1} y) \notin \Psi^t\ \forall\ \gamma \in (0, 1)\right\}. \tag{2.2}$$

Various economic assumptions regarding Ψ^t can be made; the assumptions of Shephard (1970) and Färe (1988) are typical and are used here.

Assumption 2.1. *Ψ^t is closed and strictly convex.*

Assumption 2.2. *$(x, y) \notin \Psi^t$ if $x = 0,\ y \geq 0,\ y \neq 0$; i.e., all production requires use of some inputs.*

Assumption 2.3. *For $\widetilde{x} \geq x,\ \widetilde{y} \leq y$, if $(x, y) \in \Psi$ then $(\widetilde{x}, y) \in \Psi^t$ and $(x, \widetilde{y}) \in \Psi^t$; i.e., both inputs and outputs are strongly disposable.*

Here and throughout, inequalities involving vectors are defined on an element-by-element basis, as is standard. Assumption 2.3 imposes weak monotonicity on the frontier, and is standard in microeconomic theory of the firm.

The Farrell (1957) output efficiency measure at time t measures the feasible proportionate expansion of output quantities and is defined by

$$\lambda(x, y \mid \Psi^t) := \sup\left\{\lambda \mid (x, \lambda y) \in \Psi^t\right\}. \tag{2.3}$$

This gives a *radial* measure of efficiency since all output quantities are scaled by the same factor λ. The Farrell (1957) input efficiency measure at time t is given by

$$\theta(x, y \mid \Psi^t) := \inf\left\{\theta \mid (\theta x, y) \in \Psi^t\right\} \tag{2.4}$$

and measures efficiency in terms of the amount by which input levels can be scaled downward by the same factor without reducing output levels. Clearly, $\lambda(x, y \mid \Psi^t) \geq 1$ and $\theta(x, y \mid \Psi^t) \leq 1$ for all $(x, y) \in \Psi^t$.

An alternative measure of efficiency is the hyperbolic graph measure of efficiency at time t introduced by Färe et al. (1985), i.e.,

$$\gamma(x, y \mid \Psi^t) := \inf \left\{ \gamma > 0 \mid (\gamma x, \gamma^{-1} y) \in \Psi^t \right\}. \tag{2.5}$$

By construction, $\gamma(x, y \mid \Psi^t) \leq 1$ for $(x, y) \in \Psi^t$. Just as the measures $\theta(x, y \mid \Psi^t)$ and $\lambda(x, y \mid \Psi^t)$ provide measures of the *technical efficiency* of a firm operating at a point $(x, y) \in \Psi^t$, so does $\gamma(x, y \mid \Psi^t)$, but along a hyperbolic path to the frontier of Ψ^t. The measure in (2.5) gives the amount by which input levels can be feasibly, proportionately scaled downward while simultaneously scaling output levels upward by the same proportion.

Now, for $(x, y) \in \mathbb{R}^{p+q}$, define

$$\mathcal{L}(x, y) := \{(\widetilde{x}, \widetilde{y}) \mid \widetilde{x} = ax,\ \widetilde{y} = ay \ \forall\ a \in \mathbb{R}^1_+)\}. \tag{2.6}$$

Then $\mathcal{L}(x, y)$ is the set of points in the ray emanating from the origin and passing through the point $(x, y) \in \mathbb{R}^{p+q}_+$.

Next, define the operator $\mathcal{C}(\cdot)$ so that

$$\mathcal{C}(\Psi^t) := \{\mathcal{L}(x, y) \mid (x, y) \in \Psi^t\} \tag{2.7}$$

is the convex cone of the set Ψ^t. Note that this is a pointed cone (i.e., $\mathcal{C}(\Psi^t)$ includes $\{(0, 0)\}$). Analogous to (2.2), the frontier of this set is given by

$$\mathcal{C}^{\partial}(\Psi^t) := \left\{ \mathcal{L}(x, y) \mid \mathcal{L}(x, y) \in \mathcal{C}(\Psi^t),\ \mathcal{L}(x, \lambda y) \notin \mathcal{C}(\Psi^t) \text{ for any } \lambda \in (1, \infty) \right\}. \tag{2.8}$$

If $\mathcal{C}(\Psi^t) = \Psi^t$, then the frontier $\Psi^{t\partial}$ at time t exhibits globally constant returns to scale (CRS), although this is ruled out by strict convexity of Ψ^t in Assumption 2.1. Otherwise, $\Psi^t \subset \mathcal{C}(\Psi^t)$ and $\Psi^{t\partial}$ is said to exhibit variable returns to scale (VRS), with returns to scale either increasing, constant, or decreasing depending on the particular region of the frontier.

It is well known that the choice of orientation (either input or output) can have a large impact on measured efficiency. As discussed by Wilson (2011), under VRS, a large firm could conceivably lie close to the frontier in the output direction, but far from the frontier

the input direction. Similarly, a small firm might lie close to the frontier in the input direction, but far from the frontier in the output direction. Such differences are related to the slope and curvature of the frontier. Moreover, there seems to be no criteria telling the applied researcher whether to use the input- or output-orientation. The hyperbolic measure in (2.5) can be viewed as a compromise between the two extremes (i.e., input or output orientations).

Now consider a sample $\mathcal{X}_n = \{(X_i^1, Y_i^1),\ (X_i^2, Y_i^2)\}_{i=1}^n$ of input-output combinations for n firms observed in periods $t = 1$ and 2. Firm i's change in productivity between periods 1 and 2 is measured by the hyperbolic Malmquist index

$$\mathcal{M}_i := \left(\frac{\gamma\left(X_i^2, Y_i^2 \mid \mathcal{C}(\Psi^1)\right)}{\gamma\left(X_i^1, Y_i^1 \mid \mathcal{C}(\Psi^1)\right)} \times \frac{\gamma\left(X_i^2, Y_i^2 \mid \mathcal{C}(\Psi^2)\right)}{\gamma\left(X_i^1, Y_i^1 \mid \mathcal{C}(\Psi^2)\right)} \right)^{1/2}. \tag{2.9}$$

This is the geometric mean of two ratios, each providing a measure of productivity change, in the first case using the boundary of $\mathcal{C}(\Psi^1)$ as a benchmark, and in the second case using the boundary of $\mathcal{C}(\Psi^2)$ as a benchmark. For firm i, $\mathcal{M}_i$ ($>$, $=$ or $<$) 1 if productivity (increases, remains unchanged or decreases) between periods 1 and 2.

In addition to estimating $\mathcal{M}_i$ for individual firms, applied researchers (e.g., Ball et al., 2004) often report estimates of *geometric* means

$$M_n = \left(\prod_{i=1}^{n} \mathcal{M}_i \right)^{1/n} \tag{2.10}$$

of Malmquist indices for each of n firms. Interest lies in whether such means are significantly greater than or less than 1. Similarly, geometric means of estimates of components of Malmquist indices are are also often reported. In the literature, geometric means are used to define Malmquist indices and to measure average productivity change over firms due to the multiplicative nature of radial efficiency measures such as those defined in (2.4), (2.3), and (2.5).

Of course, the production sets Ψ^t, $t \in \{1,\ 2\}$ and their frontiers are unobserved, and must be estimated from data. Due to this, the efficiency measures defined in (2.4), (2.3), and (2.5) are also unobserved, as are $\mathcal{M}_i$ and M_n defined in (2.9) and (2.10). All of these must be estimated, and inference is needed in order to have any idea of what is learned from the available data. In turn, the results of Bahadur and Savage (1956) make clear the necessity of a statistical model in order to make inference.

The following assumptions are analogous to Assumptions 3.1–3.4 of Kneip et al. (2015). Here, the assumptions below are understood to hold at time t, or when considering two points in time, they must hold at both points in time. In order to draw upon previous results, we state the assumptions below in terms of the input-oriented measure of efficiency. Of course one could also state the assumptions in terms of the output-oriented or hyperbolic efficiency measures introduced above as will be made clear by Lemma 3.1 which appears below in Section 3.

Assumption 2.4. *(i) The random variables (X^t, Y^t) possess 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 2.5. *(i) $\mathcal{D}^{t*} := \{\theta(x, y \mid \Psi^t)x,\ y) \mid (x, y) \in \mathcal{D}^t\} \subset \mathcal{D}^t$; (ii) $\mathcal{D}^{t*}$ is compact; and (iii) $f(\theta(x, y)x, y) > 0$ for all $(x, y) \in \mathcal{D}$.*

Recalling that the strong (i.e., free) disposability assumed in Assumption 2.3 implies that the frontier is weakly monotone, the next assumption strengthens this by requiring the frontier to be sufficiently smooth to obtain results on moments of DEA estimators.

Assumption 2.6. *$\theta(x, y \mid \Psi^t)$ is three times continuously differentiable on $\mathcal{D}^t$.*

In addition, the VRS-DEA estimator described below in Section 2.2 requires the following assumption.

Assumption 2.7. *$\mathcal{D}^t$ is* almost strictly convex*; i.e., for any $(x, y), (\widetilde{x}, \widetilde{y}) \in \mathcal{D}^t$ with $(\frac{x}{\|x\|}, y) \neq (\frac{\widetilde{x}}{\|\widetilde{x}\|}, \widetilde{y})$, the set $\{(x^*, y^*) \mid (x^*, y^*) = (x, y)+\alpha((\widetilde{x}, \widetilde{y})-(x, y))$ for some $0 < \alpha < 1\}$ is a subset of the* interior *of $\mathcal{D}^t$.*

Assumptions 2.1–2.7 comprise a statistical model similar to the one defined in Kneip et al. (2015). Some additional assumptions will be necessary for establishing properties of estimators of $\mathcal{M}_i$ and M_n, but these are introduced later as needed.

2.2 DEA Estimators

Given the sample $\mathcal{X}_n$, the production set Ψ^t can be estimated for $t \in \{1,\ 2\}$ by

$$\widehat{\Psi}^t = \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq \boldsymbol{Y}^t\omega,\ x \geq \boldsymbol{X}^t\omega,\ i_n'\omega = 1,\ \omega \in \mathbb{R}_+^n\right\} \tag{2.11}$$

where $\boldsymbol{X}^t = \begin{bmatrix} X_1^t & \dots & X_n^t \end{bmatrix}$, where $\boldsymbol{Y}^t = \begin{bmatrix} Y_1^t & \dots & Y_n^t \end{bmatrix}$, and i_n denotes an n-vector of ones. Note that for $t = 1$, only the pairs (X_i^1, Y_i^1) in $\mathcal{X}_n$ are used; similarly, for $t = 2$, only the pairs (X_i^2, Y_i^2) are used.

Substituting for Ψ^t in (2.3)–(2.5) leads to the variable returns to scale DEA (VRS-DEA) estimators

$$\lambda(x, y \mid \widehat{\Psi}^t) = \max_{\lambda,\omega} \left\{ \lambda \mid \lambda y \leq \boldsymbol{Y}^t \omega,\ x \geq \boldsymbol{X}^t \omega,\ i_n' \omega = 1,\ \omega \in \mathbb{R}_+^n \right\}, \tag{2.12}$$

$$\theta(x, y \mid \widehat{\Psi}^t) = \min_{\theta,\omega} \left\{ \theta \mid y \leq \boldsymbol{Y}^t \omega,\ \theta x \geq \boldsymbol{X}^t \omega,\ i_n' \omega = 1,\ \omega \in \mathbb{R}_+^n \right\} \tag{2.13}$$

and

$$\gamma(x, y \mid \widehat{\Psi}^t) = \min_{\gamma,\omega} \left\{ \theta \mid \gamma^{-1} y \leq \boldsymbol{Y}^t \omega,\ \gamma x \geq \boldsymbol{X}^t \omega,\ i_n' \omega = 1,\ \omega \in \mathbb{R}_+^n \right\} \tag{2.14}$$

of $\lambda(x, y \mid \Psi^t)$, $\theta(x, y \mid \Psi^t)$ and $\gamma(x, y \mid \Psi^t)$ (respectively). The estimators $\lambda(x, y \mid \widehat{\Psi}^t)$ and $\theta(x, y \mid \widehat{\Psi}^t)$ can be computed using standard linear programming methods, while Wilson (2011) gives a numerical algorithm for computation of $\gamma(x, y \mid \widehat{\Psi}^t)$. Asymptotic properties of the input-oriented VRS-DEA estimator of $\theta(x, y \mid \Psi^t)$ are developed by Kneip et al. (1998), Jeong (2004), Jeong and Park (2006), and Kneip et al. (2008, 2015). These results extend to the output-oriented VRS-DEA estimator of $\lambda(x, y \mid \Psi^t)$ after straightforward changes in notation. Wilson (2011) extends the asymptotic results to the VRS-DEA estimator of $\gamma(x, y, \mid \Psi^t)$. In each case, under appropriate assumptions, the VRS-DEA estimators are consistent, have non-degenerate limiting distributions, and converge at rate $n^{2/(p+q+1)}$ under variable returns to scale. See Simar and Wilson (2013, 2015) for summary and discussion.

The convex cone $\mathcal{C}(\Psi^t)$ of Ψ^t is estimated by $\mathcal{C}(\widehat{\Psi}^t)$ obtained by dropping the constraint $i_n \omega = 1$ in (2.11). Substituting for Ψ^t in (2.3)–(2.5) leads to the conical-DEA (CDEA) estimators $\lambda(x, y \mid \mathcal{C}(\widehat{\Psi}^t))$, $\theta(x, y \mid \mathcal{C}(\widehat{\Psi}^t))$ and $\gamma(x, y \mid \mathcal{C}(\widehat{\Psi}^t)$ of $\lambda(x, y \mid \mathcal{C}(\Psi^t))$, $\theta(x, y \mid \mathcal{C}(\Psi^t))$ and $\gamma(x, y \mid \mathcal{C}(\Psi^t))$ corresponding to (2.12)–(2.14) after dropping the constraint $i_n' \omega = 1$ (respectively). As with the DEA estimators, CDEA estimators $\lambda(x, y \mid \mathcal{C}(\Psi^t))$ $\theta(x, y \mid \mathcal{C}(\Psi^t))$ can be computed by standard linear programming methods (again, see Simar and Wilson, 2013 for details), and $\gamma(x, y \mid \mathcal{C}(\Psi^t))$ can be computed from the other two estimators due to Lemma 3.2 below. In the input-oriented case under constant returns to scale (i.e., when $\Psi^t = \mathcal{C}(\Psi^t)$) and appropriate regularity conditions, properties of the CDEA estimator $\theta(x, y \mid \mathcal{C}(\widehat{\Psi}^t))$ are provided by Park et al. (2010). Due to Lemmas 3.1 and 3.2 that appear below

in Section 3, these results extend trivially to the hyperbolic and output-oriented estimators. Under constant returns to scale and other appropriate assumptions given by Park et al. (2010), $\mathcal{C}(\Psi^t) = \Psi^t$ and each of the efficiency estimators $\lambda(x, y \mid \mathcal{C}(\widehat{\Psi}^t))$, $\theta(x, y \mid \mathcal{C}(\widehat{\Psi}^t))$ and $\gamma(x, y \mid \mathcal{C}(\widehat{\Psi}^t)$ are consistent, with non-degenerate limiting distributions, and converge at rate $n^{2/(p+q)}$. But under variable returns to scale where $\mathcal{C}(\Psi^t) \neq \Psi^t$, the properties of these estimators are (until now) unknown when used to estimate distance to $\mathcal{C}^\partial(\Psi^t)$.

An estimator $\widehat{\mathcal{M}}_i$ of the Malmquist index defined in (2.9) is obtained by replacing Ψ_t with $\widehat{\Psi}_t$ in (2.9), i.e., by replacing the four unknown, true efficiency measures with their corresponding estimators. Note, however, that in applied work the question of whether $\Psi^{t\partial}$ exhibits globally constant returns to scale is an empirical question. Kneip et al. (2016) provide a test of constant versus variable returns to scale, but even if the null of constant returns is not rejected, this does not mean that $\Psi^{t\partial}$ is characterized by constant returns to scale in the sense that failure to reject the null does not provide evidence that the null is true. Properties of the input-oriented VRS-DEA estimator under constant returns to scale are established by Kneip et al. (2016), and these extend trivially to the output-oriented and hyperbolic cases. But the Malmquist index defined in (2.9) is defined in terms of distance functions $\gamma(X_i^s, Y_i^s \mid \mathcal{C}(\Psi^t))$ for $s,\ t \in \{1, 2\}$, but unless constant returns to scale prevails, the properties of CDEA estimators of such distances are (until now) unknown.

In order to estimate and make inference about the Malmquist index in (2.9) and geometric means of Malmquist indices such as (2.10), some additional results are needed. These are developed in the next section.

3 Theoretical Results

3.1 Static case

In this section, we omit the superscript t denoting time in order to simplify notation. The superscript is retained where ambiguity would result without it, e.g., when considering more than one point in time. The results derived below are understood to hold for all values of t for which the assumptions in Section 2 are satisfied.

The first two results, although not new (see, e.g., Färe et al., 1985) are useful for establishing properties of estimators of distance to the boundary $\mathcal{C}^\partial(\Psi)$ of the convex cone $\mathcal{C}(\Psi)$

of Ψ.

Lemma 3.1. *Under Assumptions 2.1–2.3, For any $(x,y) \in \Psi$, (i) $\theta(x,y \mid \mathcal{C}(\Psi))^{1/2} = \lambda(x,y \mid \mathcal{C}(\Psi))^{-1/2} = \gamma(x,y \mid \mathcal{C}(\Psi))$. In addition, for any $(\widetilde{x},\widetilde{y}) \in \mathcal{L}(x,y)$ with $(x,y) \in \mathbb{R}^{p+q}_+$, (ii) $\lambda(x,y \mid \mathcal{C}(\Psi)) = \lambda(\widetilde{x},\widetilde{y} \mid \mathcal{C}(\Psi))$; (iii) $\theta(x,y \mid \mathcal{C}(\Psi)) = \theta(\widetilde{x},\widetilde{y} \mid \mathcal{C}(\Psi))$; and (iv) $\gamma(x,y \mid \mathcal{C}(\Psi)) = \gamma(\widetilde{x},\widetilde{y} \mid \mathcal{C}(\Psi))$.*

Lemma 3.2. *Under Assumptions 2.1–2.3, For any $(x,y) \in \widehat{\Psi}$, (i) $\theta(x,y \mid \mathcal{C}(\widehat{\Psi}))^{1/2} = \lambda(x,y \mid \mathcal{C}(\widehat{\Psi}))^{-1/2} = \gamma(x,y \mid \mathcal{C}(\widehat{\Psi}))$. In addition, for any $(\widetilde{x},\widetilde{y}) \in \mathcal{L}(x,y)$, $(x,y) \in \mathbb{R}^{p+q}_+$, (ii) $\lambda(x,y \mid \mathcal{C}(\widehat{\Psi})) = \lambda(\widetilde{x},\widetilde{y} \mid \mathcal{C}(\widehat{\Psi}))$; (iii) $\theta(x,y \mid \mathcal{C}(\widehat{\Psi})) = \theta(\widetilde{x},\widetilde{y} \mid \mathcal{C}(\widehat{\Psi}))$; and (iv) $\gamma(x,y \mid \mathcal{C}(\widehat{\Psi})) = \gamma(\widetilde{x},\widetilde{y} \mid \mathcal{C}(\widehat{\Psi}))$.*

Lemmas 3.1 and 3.2 establish relationships between input, output, and hyperbolic measures of distance to frontiers of the convex cones $\mathcal{C}(\Psi)$ and $\mathcal{C}(\widehat{\Psi})$. However, no such relationships exist between measures of distance to frontiers of either Ψ and $\widehat{\Psi}$ when $\Psi \neq \mathcal{C}(\Psi)$ and $\widehat{\Psi} \neq \mathcal{C}(\widehat{\Psi})$.

To further simplify notation and derivations, we now focus on the input-orientation. For purposes of measuring distances to the boundary of the convex cone of Ψ, Lemmas 3.1 and 3.2 permit trivial extensions of results obtained for the input-oriented case to the hyperbolic and output-oriented cases. Furthermore, focusing on the input orientation allows us to use analytic methods similar to those used in Kneip et al. (2015) to establish moment properties for the input-oriented VRS-DEA estimator. In addition, define $\theta(x,y) := \theta(x,y \mid \Psi)$ and $\theta_C(x,y) := \theta_C(x,y \mid \mathcal{C}(\Psi))$ to reduce notational complexity where ambiguity does not result. Analogously, define $\widehat{\theta}(x,y) := \theta(x,y \mid \widehat{\Psi})$ and $\widehat{\theta}_C(x,y) := \theta_C(x,y \mid \mathcal{C}(\widehat{\Psi}))$.

Note that for a point $(x,y) \in \mathcal{D}$ the input-oriented efficiency $\theta_C(x,y)$ can be written as[2]

$$\theta_C(x,y) = \min_{a>0} \left\{ \frac{\theta(x,ay)}{a} \mid (\theta(x,ay)x,\ ay) \in \Psi \right\}. \tag{3.1}$$

In addition, let $a^{x,y}_{min} \in \mathbb{R}_+$ denote the smallest $a > 0$ such that

$$\theta_C(x,y) = \frac{\theta(x, a^{x,y}_{min} y)}{a^{x,y}_{min}} = \min_{a>0} \left\{ \frac{\theta(x,ay)}{a} \mid (\theta(x,ay)x, ay) \in \Psi \right\}. \tag{3.2}$$

[2] For any efficiency estimator $\theta(x,y)$ considered in this section we will use the following conventions: if $(x,y) \notin \Psi$ with $(bx,y) \in \Psi$ for some $b > 1$ we set $\theta(x,y) = b\theta(bx,y)$. Otherwise, $\theta(x,y) := 1$ (or $\widehat{\theta}(x,y) := 1$) whenever the set of all possible values satisfying the defining inequalities is the empty set. Asymptotically, this has negligible effect.

Necessarily, $a^{x,y}_{min} \in \mathbb{R}_+$ is uniquely defined if Ψ is strictly convex.

Recall that due to Assumptions 2.4–2.7, the support of any observable data in each period t is some subset $\mathcal{D} \equiv \mathcal{D}^t \subset \Psi \equiv \Psi^t$. In other words, $\mathcal{D}$ is the "observable part" of Ψ. The difference between D and Ψ does not play an important role in Kneip et al. (2008, 2015 and 2016) since Assumption 2.5 requires (i) $(\theta(x,y)x, y) \in \mathcal{D}$ for $(\theta(x,y)x, y) \in \mathcal{D}$ and (ii) $f(\theta(x,y)x, y) > 0$. Here, however, the difference between $\mathcal{D}$ and Ψ is problematic for dealing with $\theta_C(x,y)$. Furthermore, in order to ensure that Malmquist indices are well-defined, $\mathcal{D}^t$ and $\mathcal{D}^s$ must "fit together" for different periods t, s. Therefore, some additional conceptual work is necessary.

Let

$$\mathcal{D}_{norm} := \left\{ \left(\frac{x}{\|x\|}, \frac{y}{\|y\|} \right) \mid (x,y) \in \mathcal{D} \right\}. \tag{3.3}$$

If $p+q=2$ then trivially $\mathcal{D}_{norm} = \{(1,1)\}$. But when $p+q>2$, $\mathcal{D}_{norm}$ will quantify the set of all possible "directions" of vectors x and y where it is possible to define a frontier. Note that for any $(\widetilde{x}, \widetilde{y})$ with $\|\widetilde{x}\| = 1$ and $\|\widetilde{y}\| = 1$ and $(\widetilde{x}, \widetilde{y}) \notin \mathcal{D}_{norm}$, we necessarily have $\{a\widetilde{x}, b\widetilde{y} \mid a, b > 0\} \cap \mathcal{D} = \emptyset$. This means that "in the direction" of $(\widetilde{x}, \widetilde{y})$ it is not possible to define any type of identifiable efficiency measure, since there is no information about an efficient frontier in such directions.[3]

Introduction of $\mathcal{D}_{norm}$ is of particular importance in a dynamic context where efficiencies in two different time periods t and s are to be compared. Frontiers may change and we may have different supports $\mathcal{D}^t$ and $\mathcal{D}^s$ in the two periods. However, it is necessary that $\mathcal{D}^t_{norm} = \mathcal{D}^s_{norm}$. Otherwise, there will be observations in one period for which distance to the other-period frontier cannot be defined. In this case Malmquist indices will be undefined with non-zero, non-negligible probability.

On the other hand, for any $\left(\frac{x}{\|x\|}, \frac{y}{\|y\|}\right) \in \mathcal{D}_{norm}$ there exists a unique ray defining the corresponding part of the conical hull frontier $\mathcal{C}^{\partial}(\Psi)$. This can easily be seen by letting $\left(\frac{x}{\|x\|}, \frac{y}{\|y\|}\right) \in \mathcal{D}_{norm}$. In addition, for $a > 0$, define

$$\widetilde{g}_x \left(a \frac{y}{\|y\|} \right) := \min_{b>0} \left\{ b \frac{x}{\|x\|} \mid \left(b \frac{x}{\|x\|}, a \frac{y}{\|y\|} \right) \in \Psi \right\}. \tag{3.4}$$

[3] Under the strong disposability assumed in Assumption 2.3, the DEA and CDEA estimators of $\theta(x,y)$ and $\theta_C(x,y)$ described above in Section 2.2 are well-defined and can be computed, but they do not estimate anything that does not depend entirely upon Assumption 2.3 or that can be identified from data when $(x,y) \notin \mathcal{D}_{norm}$.

Then there exists some $\alpha_{min}^{x,y} > 0$ such that

$$\frac{\widetilde{g}_x(\alpha_{min}^{x,y}\frac{y}{\|y\|})}{\alpha_{min}^{x,y}} = \min_{a>0}\left\{\frac{\widetilde{g}_x(a\frac{y}{\|y\|})}{a} \mid \left(g_x\left(a\frac{y}{\|y\|}\right)\frac{x}{\|x\|},\ a\frac{y}{\|y\|}\right) \in \Psi\right\} \tag{3.5}$$

where $\alpha_{min}^{x,y} \in \mathbb{R}_+$ is necessarily uniquely defined if Ψ is strictly convex.[4]

Assuming that only values a leading to well-defined frontier points are taken into account, for any $(x,y) \in \mathcal{D}$ we now have

$$\min_{a>0}\frac{\theta(x,ay)}{a} = \min_{a>0}\frac{\widetilde{g}_x(\|y\|a\frac{y}{\|y\|})}{\|x\|a} = \frac{\|y\|}{\|x\|}\min_{a>0}\frac{\widetilde{g}_x(\|y\|a\frac{y}{\|y\|})}{\|y\|a} = \frac{\|y\|}{\|x\|}\frac{\widetilde{g}_x(\alpha_{min}^{x,y}\frac{y}{\|y\|})}{\alpha_{min}^{x,y}}, \tag{3.6}$$

and $a_{min}^{x,y}$ defined in (3.2) satisfies $a_{min}^{x,y} = \frac{\alpha_{min}^{x,y}}{\|y\|}$.

Obviously, all we can hope to estimate is the version of (3.5) where Ψ is replaced by the observable part $\mathcal{D} \subset \Psi$. If $\alpha_{min}^{x,y} \in \mathbb{R}_+$ is such that $(g_x(\alpha_{min}^{x,y}\frac{y}{\|y\|})\frac{x}{\|x\|}, \alpha_{min}^{x,y}\frac{y}{\|y\|}) \notin \mathcal{D}$, then it is impossible to estimate $\theta_C(x,y)$ consistently. Minimizing (3.5) with respect to $\mathcal{D}$ instead of Ψ will then lead to a "boundary solution" $\alpha^* \in \mathcal{D}$ which is "as close as possible" to $\alpha_{min}^{x,y} \in \mathbb{R}_+$. This can only be avoided by assuming that $\mathcal{D}$ is large enough such that (when minimizing (3.5) over $\mathcal{D}$ instead of Ψ) the solution $a_{min}^{x,y} \in \mathbb{R}_+$ is in the interior of $\mathcal{D}$ in the sense that $(g_x((\alpha_{min}^{x,y}-\delta)\frac{y}{\|y\|})\frac{x}{\|x\|}, (\alpha_{min}^{x,y}-\delta)\frac{y}{\|y\|}) \in \mathcal{D}$ as well as $(g_x((\alpha_{min}^{x,y}+\delta)\frac{y}{\|y\|})\frac{x}{\|x\|},\ (\alpha_{min}^{x,y}+\delta)\frac{y}{\|y\|}) \in \mathcal{D}$. Since $\mathcal{D}$ is almost strictly convex by Assumption 2.7, $\alpha_{min}^{x,y} \in \mathbb{R}_+$ is necessarily unique, and $\frac{\widetilde{g}_x((a_{min}^{x,y}-\delta)\frac{y}{\|y\|})}{(a_{min}^{x,y}-\delta)} > \frac{\widetilde{g}_x(\alpha_{min}^{x,y}\frac{y}{\|y\|})}{\alpha_{min}^{x,y}}$ as well as $\frac{\widetilde{g}_x((\alpha_{min}^{x,y}+\delta)\frac{y}{\|y\|})}{(\alpha_{min}^{x,y}+\delta)} > \frac{\widetilde{g}_x(\alpha_{min}^{x,y}\frac{y}{\|y\|})}{\alpha_{min}^{x,y}}$. Convexity of Ψ then necessarily implies implies that this value $\alpha_{min}^{x,y} \in \mathbb{R}_+$ also corresponds to the solution of the original minimization problem with respect to Ψ. In this sense the following assumption ensures well-defined estimators of $\theta_C(x,y)$.

Assumption 3.1. *(i) The support $\mathcal{D} \subset \Psi$ of f is such that for any $(\frac{x}{\|x\|}, \frac{y}{\|y\|}) \in \mathcal{D}_{norm}$ we have $(\widetilde{g}_x(\alpha_{min}^{x,y}\frac{y}{\|y\|})\frac{x}{\|x\|}, \alpha_{min}^{x,y}\frac{y}{\|y\|}) \in \mathcal{D}$; (ii) there exists a $\delta > 0$ such that for any $(\frac{x}{\|x\|}, \frac{y}{\|y\|}) \in \mathcal{D}_{norm}$ we also have $(\widetilde{g}_x([\alpha_{min}^{x,y}-\delta]\frac{y}{\|y\|})\frac{x}{\|x\|}, [\alpha_{min}^{x,y}-\delta]\frac{y}{\|y\|}) \in \mathcal{D}$ and $(\widetilde{g}_x([\alpha_{min}^{x,y}+\delta]\frac{y}{\|y\|})\frac{x}{\|x\|}, [\alpha_{min}^{x,y}+\delta]\frac{y}{\|y\|}) \in \mathcal{D}$; (iii) There exists a constant $0 < M < \infty$ such that $\|x\| \leq M$ for all $(x,y) \in \mathcal{D}$.*

Remark 3.1. *Part (iii) of Assumption 3.1 is necessary to guarantee existence of moments. Although moments necessarily exist for $\theta_C(X_i,Y_i) \in (0,1]$, $|\log\theta_C(X_i,Y_i)|$ is potentially un-*

[4] Note that $\widetilde{g}_x\left(a\frac{y}{\|y\|}\right)$ corresponds to the function $g_x\left(0, a\frac{y}{\|y\|}\right)$ defined in Kneip et al. (2008). The coordinate system introduced in Kneip et al. (2008) is not needed here, but is required in the proofs that follow in Appendix A.

bounded. Moreover, up to this point we have only assumed compactness of $\mathcal{D}^*$ and not necessarily of $\mathcal{D}$. In principle, (iii) could be replaced by a weaker version requiring only existence of all relevant moments, but boundedness of $\|x\|$ greatly simplifies asymptotic arguments that follow.

Since by Assumption 2.7 $\mathcal{D}$ is almost strictly convex (i.e., its frontier does not contain a straight segment), Assumption (3.1) ensures that $\alpha^{x,y}_{min} > 0$ is uniquely defined for any (x, y) with $\left(\frac{x}{\|x\|},\ \frac{y}{\|y\|}\right) \in \mathcal{D}_{norm}$. The above arguments then imply that the ray

$$\mathcal{L}\left(\widetilde{g}_x(\alpha^{x,y}_{min}\frac{y}{\|y\|})\frac{x}{\|x\|},\ \alpha^{x,y}_{min}\frac{y}{\|y\|}\right) \subset \mathcal{C}^{\partial}(\Psi) \tag{3.7}$$

defines the part of the frontier of $\mathcal{C}(\Psi)$ corresponding to the specific vector $\left(\frac{x}{\|x\|},\ \frac{y}{\|y\|}\right) \in \mathcal{D}_{norm}$.

Next, note that for any $(\widetilde{x}, \widetilde{y})$ with $\widetilde{x}\|\widetilde{x}\|^{-1} = x\|x\|^{-1}$ and $\widetilde{y}\|\widetilde{y}\|^{-1} = y\|y\|^{-1}$ we have

$$\begin{aligned}
(\widetilde{x}, \widetilde{y}) &= \left(\|\widetilde{x}\|\frac{x}{\|x\|}, \|\widetilde{y}\|\frac{y}{\|y\|}\right) \\
&= \left(\frac{\|\widetilde{x}\|}{\widetilde{g}_x(\alpha^{x,y}_{min}\frac{y}{\|y\|})}\widetilde{g}_x\left(\alpha^{x,y}_{min}\frac{y}{\|y\|}\right)\frac{x}{\|x\|},\ \frac{\|\widetilde{y}\|}{\alpha^{x,y}_{min}}\alpha^{x,y}_{min}\frac{y}{\|y\|}\right) \\
&= \left(\left(\frac{\|\widetilde{x}\|\alpha^{x,y}_{min}}{\widetilde{g}_x(\alpha^{x,y}_{min}\frac{y}{\|y\|})\|\widetilde{y}\|}\right)\cdot\left(\frac{\|\widetilde{y}\|}{\alpha^{x,y}_{min}}\right)\cdot\widetilde{g}_x\left(\alpha^{x,y}_{min}\frac{y}{\|y\|}\right)\frac{x}{\|x\|},\ \left(\frac{\|\widetilde{y}\|}{\alpha^{x,y}_{min}}\right)\cdot\alpha^{x,y}_{min}\frac{y}{\|y\|}\right)
\end{aligned} \tag{3.8}$$

which implies

$$\begin{aligned}
\theta_C(\widetilde{x}, \widetilde{y}) &= \frac{\|\widetilde{y}\|}{\|\widetilde{x}\|}\frac{\widetilde{g}_x\left(\alpha^{x,y}_{min}\frac{y}{\|y\|}\right)}{\alpha^{x,y}_{min}} \\
&= \operatorname{argmin}\left\{\theta \mid (\theta\widetilde{x}, \widetilde{y}) \in \mathcal{L}\left(\widetilde{g}_x\left(\alpha^{x,y}_{min}\frac{y}{\|y\|}\right)\frac{x}{\|x\|}, \alpha^{x,y}_{min}\frac{y}{\|y\|}\right) \subset \mathcal{C}^{\partial}(\Psi)\right\}.
\end{aligned} \tag{3.9}$$

This shows that we can define an input-oriented efficiency measure for any $(\widetilde{x}, \widetilde{y}) \in \{bx, ay \mid a, b > 0,\ \left(\frac{x}{\|x\|}, \frac{y}{\|y\|}\right) \in \mathcal{D}_{norm}\}$, i.e. $\theta_C(\widetilde{x}, \widetilde{y})$ is well-defined even if $(\widetilde{x}, \widetilde{y}) \notin \mathcal{D}$. The same then holds for the hyperbolic graph measure $\gamma(\widetilde{x}, \widetilde{y})$ which by Lemma 3.1 can be represented by

$$\gamma(\widetilde{x}, \widetilde{y}) = \theta_C(\widetilde{x}, \widetilde{y})^{1/2} = \left(\frac{\|\widetilde{y}\|}{\|\widetilde{x}\|}\right)^{1/2}\left(\frac{\widetilde{g}_x(\alpha^{x,y}_{min}\frac{y}{\|y\|})}{\alpha^{x,y}_{min}}\right)^{1/2}. \tag{3.10}$$

The result in (3.10) provides a basis for defining Malmquist indices for data from different periods t, s with $\mathcal{D}^t \neq \mathcal{D}^s$. Recall, however, that $\mathcal{D}^t_{norm} = \mathcal{D}^s_{norm}$ is a necessary condition to

ensure that (3.9) and (3.10) are well defined for all points $(x, y) := (X_i^t, Y_i^t)$ and $(x, y) := (X_i^s, Y_i^s)$.

Relying on the VRS-DEA estimator, the estimator $\widehat{\theta}_C(x, y)$ in a given period is given by

$$\widehat{\theta}_C(x, y \mid \mathcal{X}_n) = \min_{a>0} \left\{ \frac{\widehat{\theta}_{\text{VRS}}(x, ay \mid \mathcal{X}_n)}{a} \mid (\widehat{\theta}_{\text{VRS}}(x, ay \mid \mathcal{X}_n)x, ay) \in \widehat{\Psi} \right\}. \tag{3.11}$$

Assumptions 2.1–2.7 and 3.1 now provide the basis for inference about $\widehat{\theta}_C(\widetilde{x}, \widetilde{y})$.

Theorem 3.1. *Under Assumptions 2.1–2.7 as well as 3.1, for each* $(x, y) \in \mathcal{D}$

$$n^{\frac{2}{p+q+1}} \left(\frac{\widehat{\theta}_C(x, y \mid \mathcal{X}_n)}{\theta_C(x, y)} - 1 \right) \xrightarrow{d} F_{x,y} \tag{3.12}$$

as $n \to \infty$*, where* $F_{x,y}$ *is a continuous, non-degenerate distribution function with* $F_{x,y}(0) = 0$*. The analytical structure of* $F_{x,y}$ *is given by* (A.21) *in the proof of the theorem. Furthermore,* $\exists$ *a constant* $0 < C_0 < \infty$ *such that for all* $i, j \in \{1, \ldots, n\}$*,* $i \neq j$*,*

$$E\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i)\right) = C_0 n^{-\frac{2}{p+q+1}} + O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right), \tag{3.13}$$

$$VAR\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = O\left(n^{-\frac{4}{p+q+1}} (\log n)^{\frac{4}{p+q+1}}\right), \tag{3.14}$$

and

$$\left| COV\Big(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i), \widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n) - \theta_C(X_j, Y_j)\Big) \right| = O\left(n^{-\frac{p+q+2}{p+q+1}} (\log n)^{\frac{p+q+2}{p+q+1}}\right) = o\left(n^{-1}\right). \tag{3.15}$$

The value of the constant C_0 *depends on* f *and on the structure of the set* $\mathcal{D} \subset \Psi$*.*

Among other things, Theorem 3.1 establishes that the rate of $\widehat{\theta}_C(x, y \mid \mathcal{X}_n)$ is determined by the rate of the VRS-DEA estimator (as opposed to the faster $n^{2/(p+q)}$ rate of the CRS-DEA estimator). This is perhaps intuitive, but until now unproven.

The next result is a straightforward consequence of Theorem 3.1 and Lemmas 3.1 and 3.2, and extends Theorem 3.1 to the hyperbolic estimator and its logarithm.

Theorem 3.2. *Let either (i)* $\Gamma(\theta_C(x, y)) := [\theta_C(x, y)]^{1/2} = \gamma(x, y)$ *or (ii)* $\Gamma(\theta_C(x, y)) := \log [\theta_C(x, y)]^{1/2} = \log \gamma(x, y)$*. Under Assumptions 2.1–2.7 as well as Assumption 3.1, for each* $(x, y) \in \mathcal{D}$

$$n^{\frac{2}{p+q+1}} \left(\Gamma(\widehat{\theta}_C(x, y \mid \mathcal{X}_n)) - \Gamma(\theta_C(x, y)) \right) \xrightarrow{d} F_{x,y}^{\Gamma} \tag{3.16}$$

as $n \to \infty$*, where* $F^{\Gamma}_{x,y}$ *is a continuous, non-degenerate distribution function with* $F^{\Gamma}_{x,y}(0) = 0$*. The analytical structure of* $F^{\Gamma}_{x,y}$ *depends on* Γ *and on the distribution function* $F_{x,y}$ *defined in Theorem 3.1. Furthermore,* $\exists$ *a constant* $0 < C^{\Gamma}_0 < \infty$ *such that for all* $i, j \in \{1, \ldots, n\}$*,* $i \neq j$*,*

$$E\left(\Gamma(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_C(X_i, Y_i))\right) = C^{\Gamma}_0 n^{-\frac{2}{p+q+1}} + O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right), \quad (3.17)$$

$$VAR\left(\Gamma(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_C(X_i, Y_i))\right) = O\left(n^{-\frac{4}{p+q+1}} (\log n)^{\frac{4}{p+q+1}}\right), \quad (3.18)$$

and

$$\left|COV\Big(\Gamma\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)\right) - \Gamma\left(\theta_C(X_i, Y_i)\right), \Gamma\left(\widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n)\right) - \Gamma\left(\theta_C(X_j, Y_j)\right)\Big)\right| = O\left(n^{-\frac{p+q+2}{p+q+1}} (\log n)^{\frac{p+q+2}{p+q+1}}\right) = o\left(n^{-1}\right). \quad (3.19)$$

The value of the constant C^{Γ}_0 *depends on* f*,* Γ *and the structure of the set* $\mathcal{D} \subset \Psi$*.*

Note that it is trivial to extend the results of Theorems 3.1 and 3.2 to the output-oriented estimator and its logarithm as well as the hyperbolic estimator and its logarithm due to Lemmas 3.1 and 3.2. The value of the constants C_0 and C^{Γ}_0 will differ of course depending on the direction in which efficiency is estimated.

3.2 Dynamic case

Now turn to the dynamic case. Suppose that for two different time periods $t \in \{1, 2\}$ we have the set $\mathcal{X}_n = \{(X^1_i, Y^1_i), (X^2_i, Y^2_i)\}^n_{i=1}$ defined earlier in Section 2.1 of independent, identically distributed (iid) pairs (of pairs) of input and output quantities for the two different periods. In each period there may exist additional observations which do not possess a counterpart in the other period. More precisely, there are $n_1 \geq n$ observations in period 1 which are used to estimate the hyperbolic distance $\gamma^1(x, y) := \gamma(x, y \mid \mathcal{C}(\Psi^1))$, while there are $n_2 \geq n$ observations in period 2 which are used to estimate the hyperbolic distance $\gamma^2(x, y) := \gamma(x, y \mid \mathcal{C}(\Psi^2))$.

Assumption 3.2. *(i) For* $t \in \{1, 2\}$ *there are iid observations* (X^t_i, Y^t_i)*,* $i = 1, \ldots, n_t$*, such that Assumptions 2.1–2.7 and 3.1 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 $i = 1, \ldots, n$, (X_i^1, Y_i^1) is independent of (X_j^2, Y_j^2) for all $j = 1, \ldots, n_2$ with $i \neq j$; (v) for any $i = 1, \ldots, n$, (X_i^2, Y_i^2) is independent of (X_j^1, Y_j^1) for all $j = 1, \ldots, n_1$ with $i \neq j$.

Note that condition (i) of this assumption only guarantees that all estimators $\widehat{\theta}_C^t(x, y \mid \mathcal{X}_{n_t}^t)$ and $\widehat{\gamma}^t(x, y \mid \mathcal{X}_{n_t}^t)$ follow the asymptotic distributions derived in Theorems 3.1 and 3.2. Condition (ii) together with (3.9) ensures that the cross-efficiency estimators $\widehat{\theta}_C^1(X_i^2, Y_i^2 \mid \mathcal{X}_{n_1}^1)$ and $\widehat{\theta}_C^2(X_i^1, Y_i^1 \mid \mathcal{X}_{n_2}^2)$ are asymptotically well-defined and possess the same rates of convergence as the contemporaneous efficiency estimators. Conditions (iv)–(v) permit dependence of a given firm's input-output quantities across periods 1 and 2, but require independence of the firm's input-output quantities from those of other firms in other periods.

Before analyzing "average" productivity change, first consider a firm operating at observed, fixed points (x^1, y^1) and (x^2, y^2) in periods 1 and 2. Then from (2.9) the Malmquist index for this firm is

$$\mathcal{M} = \left(\frac{\gamma\left(x^2, y^2 \mid \mathcal{C}(\Psi^1)\right)}{\gamma\left(x^1, y^1 \mid \mathcal{C}(\Psi^1)\right)} \times \frac{\gamma\left(x^2, y^2 \mid \mathcal{C}(\Psi^2)\right)}{\gamma\left(x^1, y^1 \mid \mathcal{C}(\Psi^2)\right)} \right)^{1/2}. \tag{3.20}$$

Using the data $\mathcal{X}_{n_1}^1 := \{(X_i^1, Y_i^1\}_{i=1,\ldots,n_1}$ and $\mathcal{X}_{n_2}^2 := \{(X_i^2, Y_i^2\}_{i=1,\ldots,n_2}$ $\mathcal{M}$ can be estimated by

$$\widehat{\mathcal{M}} = \left(\frac{\widehat{\gamma}\left(x^2, y^2 \mid \mathcal{X}_{n_1}^1\right)}{\widehat{\gamma}\left(x^1, y^1 \mid \mathcal{X}_{n_1}^1\right)} \times \frac{\widehat{\gamma}\left(x^2, y^2 \mid \mathcal{X}_{n_2}^2\right)}{\widehat{\gamma}\left(x^1, y^1 \mid \mathcal{X}_{n_2}^2\right)} \right)^{1/2}. \tag{3.21}$$

Theorem 3.3. *Under Assumptions 2.1–2.7, 3.1 and 3.2, for $(x^1, y^1) \in \mathcal{D}^1$ and $(x^2, y^2) \in \mathcal{D}^2$*

$$n^{\frac{2}{p+q+1}} \left(\widehat{\mathcal{M}} - \mathcal{M} \right) \xrightarrow{d} F^{\mathcal{M}} \tag{3.22}$$

as $n \to \infty$, where $F^{\mathcal{M}}$ is a continuous, non-degenerate distribution function with $F^{\mathcal{M}}(0) = 0$.

Theorem 3.3 establishes the existence of a non-degenerate limiting distribution as well as the convergence rate for the estimator in (3.21) of the Malmquist index for a given firm observed in periods 1 and 2. These results permit inference about the unobserved, true Malmquist index $\mathcal{M}$ using the subsampling methods described by Simar and Wilson (2011).

Now consider the log Malmquist index

$$\log \mathcal{M}_i = \frac{1}{2} \left[\log \gamma^1(X_i^2, Y_i^2) + \log \gamma^2(X_i^2, Y_i^2) - \log \gamma^1(X_i^1, Y_i^1) - \log \gamma^2(X_i^1, Y_i^1) \right] \tag{3.23}$$

with mean

$$\mu_{\mathcal{M}} := E\left(\log \mathcal{M}_i\right). \tag{3.24}$$

The log Malmquist index in (3.23) is estimated by

$$\begin{aligned}\log \widehat{\mathcal{M}}_i = \frac{1}{2}\Big(&\log \widehat{\gamma}^1(X_i^2, Y_i^2 \mid \mathcal{X}_{n_1}^1) + \log \widehat{\gamma}^2(X_i^2, Y_i^2 \mid \mathcal{X}_{n_2}^2) - \\ &\log \widehat{\gamma}^1(X_i^1, Y_i^1 \mid \mathcal{X}_{n_1}^1) - \log \widehat{\gamma}^2(X_i^1, Y_i^1 \mid \mathcal{X}_{n_2}^2)\Big)\end{aligned} \tag{3.25}$$

for firms indexed by i and observed in *both* periods 1 and 2, with sample average

$$\widehat{\mu}_{\mathcal{M},n} := \frac{1}{n}\sum_{i=1}^{n} \log \widehat{\mathcal{M}}_i. \tag{3.26}$$

In finite samples, it may be the case that $(X_i^2, Y_i^2) \notin \mathcal{C}(\widehat{\Psi}^1)$ or $(X_i^1, Y_i^1) \notin \mathcal{C}(\widehat{\Psi}^2)$ for some observations i. Here, we use the convention that $\widehat{\gamma}^1(x, y \mid \mathcal{X}_{n_1}^1) := 1$ if $(x, y) \notin \mathcal{C}(\widehat{\Psi}^1)$, as well as $\widehat{\gamma}^2(x, y \mid \mathcal{X}_{n_2}^2) := 1$ if $(x, y) \notin \mathcal{C}(\widehat{\Psi}^2)$, although this is seldom needed in practice. Asymptotically this does not impose a problem due to Assumption 3.2(ii).

The next theorem provides one of the main results of the paper by enabling inference on the difference between $\mu_{\mathcal{M}}$ and $\widehat{\mu}_{\mathcal{M},n}$. Theorem 3.2 describes the asymptotic behavior of $\log \widehat{\gamma}^2(X_i^2, Y_i^2 \mid \mathcal{X}_{n_2}^2)$ and $\log \widehat{\gamma}^1(X_i^1, Y_i^1 \mid \mathcal{X}_{n_1}^1)$. The following theorem provides rates of convergence for the moments of all efficiency estimators determining $\log \widehat{\mathcal{M}}_i$.

Theorem 3.4. *Under Assumptions 2.1–2.7, 3.1 and 3.2, and for all $t, s \in \{1, 2\}$, there exist constants $0 < C_0^{ts} < \infty$ such that for all $i \in \{1, \ldots, n\}$ and as $n \leq \min\{n_1, n_2\} \to \infty$,*

$$E\left(\log \widehat{\gamma}^s(X_i^t, Y_i^t \mid \mathcal{X}_{n_s}^s) - \log \gamma^s(X_i^t, Y_i^t)\right) = C_0^{ts} n_s^{-\frac{2}{p+q+1}} + O\left(n_s^{-\frac{3}{p+q+1}} (\log n_s)^{\frac{3}{p+q+1}}\right) \tag{3.27}$$

and

$$E\left(\left[\log \widehat{\gamma}^s(X_i^t, Y_i^t \mid \mathcal{X}_{n_s}^s) - \log \gamma^s(X_i^t, Y_i^t)\right]^2\right) = O\left(n_s^{-\frac{4}{p+q+1}} (\log n_s)^{\frac{4}{p+q+1}}\right), \tag{3.28}$$

and for $t^*, s^* \in \{1, 2\}$, $j \neq i$,

$$\begin{aligned}\Big| E\Big(&[\log \widehat{\gamma}^s(X_i^t, Y_i^t \mid \mathcal{X}_{n_s}^s) - E(\log \widehat{\gamma}^s(X_i^t, Y_i^t \mid \mathcal{X}_{n_s}^s))] \\ &[\log \widehat{\gamma}^{s^*}(X_j^{t^*}, Y_j^{t^*} \mid \mathcal{X}_{n_{s^*}}^{s^*}) - E(\log \widehat{\gamma}^{s^*}(X_j^{t^*}, Y_j^{t^*} \mid \mathcal{X}_{n_{s^*}}^{s^*}))]\Big)\Big| &= O\left(n^{-\frac{p+q+2}{p+q+1}} (\log n)^{\frac{p+q+2}{p+q+1}}\right) \\ &= o\left(n^{-1}\right).\end{aligned} \tag{3.29}$$

The value of the constant C_0^{ts} depends on f, g and the structure of the sets $\mathcal{D}^s \subset \Psi^s$ and $\mathcal{D}^t \subset \Psi^t$.

These results permit inference about the log Malmquist index. The following corollary is an immediate consequence of Theorem 3.4. A proof is omitted.

Theorem 3.5. *Under Assumptions 2.1–2.7, 3.1 and 3.2* $\exists$ *a constant* $C_{\mathcal{M}} < \infty$ *such that as* $n \leq \min\{n_1, n_2\} \to \infty$,

$$E\left(\widehat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}}\right) = C_{\mathcal{M}} n^{-\frac{2}{p+q+1}} + O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right) \tag{3.30}$$

and

$$VAR\left((\widehat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}}\right) = \frac{1}{n} VAR(\log \mathcal{M}_i) + o(n^{-1}). \tag{3.31}$$

The constant $C_{\mathcal{M}}$ is a sum of positive and negative biases, and it can be either positive, negative or equal to zero. It will be zero if the two distributions coincide, leading to the following result.

Theorem 3.6. *Assume Assumptions 2.1–2.7, 3.1 and 3.2 hold. In addition, assume that (i)* $f := f^1 = f^2$ *(and hence* $\mathcal{D} := \mathcal{D}^1 = \mathcal{D}^2$*); (ii)* $n = n_1 = n_2$*; and (iii)* $f_{12}(x, y; x^*, y^*) = f_{12}((x^*, y^*), (x, y))$ *for all* $(x, y), (x^*, y^*) \in \mathcal{D}$*. Then as* $n \leq \min\{n_1, n_2\} \to \infty$,

$$E\left(\widehat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}}\right) = 0 \tag{3.32}$$

and

$$VAR\left((\widehat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}}\right) = \frac{1}{n} VAR(\log \mathcal{M}_i) + o(n^{-1}). \tag{3.33}$$

Remark 3.2. *Note that* (3.32) *is a consequence of a somewhat trivial fact: If (a) there are two samples with* identical *data generating processes, and (b) for both samples the same type of estimator is applied, then all resulting biases are identical (and hence cancel out when subtracting). In our context there only exists the difficulty that the roles of* (X_i^1, Y_i^1) *and* (X_i^2, Y_i^2) *are different in* $\log \widehat{\gamma}^2(X_i^1, Y_i^1 \mid \mathcal{X}_{n_2}^2)$ *and* $\log \widehat{\gamma}^1(X_i^2, Y_i^2 \mid \mathcal{X}_{n_1}^1)$*, which is resolved by the additional assumption (iii) of "symmetry" on the joint density.*

Remark 3.3. *Section 3.1 of Kneip et al. (2016) overlooks the point raised in Remark 3.2. Indeed the results in Section 3.1 of Kneip et al. (2016) are incomplete (but not false; the results provide bad approximations in the case of identical distributions). In Kneip et al. (2016, Section 3.1) if* $n_1 = n_2$ *and if both samples possess identical distributions then the biases cancel out, and in the notation of Kneip et al. (2016)* $E(\widehat{\mu}_{1,n_1} - \widehat{\mu}_{2,n_2}) = 0$.

Remark 3.4. *It is possible to achieve* (3.32) *while only requiring* $f^1 = f^2$ *(i.e., without assuming* $n_1 = n_2$ *and symmetry of the joint density). This is possible by modifying the estimator and using*

$$\begin{aligned}\log \widetilde{\mathcal{M}}_i = \frac{1}{2}(&\log \widehat{\gamma}^1(X_i^2, Y_i^2 \mid \mathcal{X}^1_{n_1,-i}) + \log \widehat{\gamma}^2(X_i^2, Y_i^2 \mid \mathcal{X}^2_{n_2,-i}) \\ &- \log \widehat{\gamma}^1(X_i^1, Y_i^1 \mid \mathcal{X}^1_{n_1,-i}) - \log \widehat{\gamma}^2(X_i^1, Y_i^1 \mid \mathcal{X}^2_{n_2,-i})),\end{aligned}$$

where $\mathcal{X}^s_{n_s,-i}$ *is the reduced sample of size* $(n-1)$ *obtained by eliminating the* i*th observation* (X_i^s, Y_i^s)*,* $s = 0,\ 1$*. In other words, for any* $i = 1,\ \ldots,\ n$ *the estimates* $\widehat{\gamma}$ *are constructed without taking into account the i-th observation. In this case everything is symmetric, and for identical distributions arguments similar to those used above above lead to*

$$E\left(\log \widehat{\gamma}^1(X_i^1, Y_i^1 \mid \mathcal{X}^1_{n_1,-i})\right) = E\left(\log \widehat{\gamma}^1(X_i^2, Y_i^2 \mid \mathcal{X}^1_{n_1,-i})\right) \tag{3.34}$$

and

$$E\left(\log \widehat{\gamma}^2(X_i^1, Y_i^1 \mid \mathcal{X}^2_{n_2,-i})\right) = E\left(\log \widehat{\gamma}^2(X_i^2, Y_i^2 \mid \mathcal{X}^2_{n_2,-i})\right) \tag{3.35}$$

independent of n_1 *and* n_2*. Hence* $E\left(\log \widetilde{\mathcal{M}}_i\right) = 0$*.*

Remark 3.5. *Tests based on Theorem 3.6 are tests of* $f^1 = f^2$ *rather than of* $\mu_{\mathcal{M}} := E(\log \mathcal{M}_i) = 0$*. Note that the true mean* $\mu_{\mathcal{M}}$ *may be zero even if* $f^1 \neq f^2$*. But if* $f^1 \neq f^2$ *then biases do not cancel out in general, and one is back to* (3.30)*. Since for large* $(p+q)$ *bias dominates variance, the test will (asymptotically) reject the null hypotheses even if* $\mu_{\mathcal{M}} = 0$ *if bias is not accounted for.*

The results obtained so far lead to the following CLT.

Theorem 3.7. *Under Assumptions 2.1–2.7, 3.1 and 3.2,*

$$\sqrt{n}\left(\widehat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}} - \mathcal{R}_n\right) \xrightarrow{d} N(0,\, VAR(\log \mathcal{M}_i)) \tag{3.36}$$

as $n \leq \min\{n_1, n_2\} \to \infty$ *where*

$$\mathcal{R}_n := E\left(\widehat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}}\right) = C_{\mathcal{M}} n^{-\frac{2}{p+q+1}} + O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right). \tag{3.37}$$

The geometric mean of the estimated Malmquist indices $\widehat{\mathcal{M}}_i$ is given by

$$\widehat{M}_n = \exp(\widehat{\mu}_{\mathcal{M},n}) = \left(\prod_{i=1}^{n} \widehat{\mathcal{M}}_i\right)^{1/n}. \tag{3.38}$$

The following CLT may be used for inference about $\widehat{M}_n$.

Theorem 3.8. *Under Assumptions 2.1–2.7, 3.1 and 3.2,* $n \leq \min\{n_1, n_2\} \to \infty$

$$\sqrt{n}\left(\widehat{M}_n - \widetilde{M}\right) \xrightarrow{d} N(0, \exp(2\mu_{\mathcal{M}})\, VAR(\log \mathcal{M}_i)), \tag{3.39}$$

where $\mathcal{R}_n$ *is defined by* (3.37) *and*

$$\begin{aligned}\widetilde{M} &:= \exp(\mu_{\mathcal{M}} + \mathcal{R}_n) = \exp(\mu_{\mathcal{M}})(1 + \mathcal{R}_n)) + O\left(n^{-\frac{4}{p+q+1}}\right) \\ &= \exp(\mu_{\mathcal{M}}) + \exp(\mu_{\mathcal{M}}) C_{\mathcal{M}} n^{-\frac{2}{p+q+1}} + O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right).\end{aligned} \tag{3.40}$$

Remark 3.6. *Recall that under the additional conditions of Theorem 3.6 we even have* $\mathcal{R}_n = 0$.

Remark 3.7. *Consider the geometric mean* $M_n = \exp(\frac{1}{n}\sum_{i=1}^n \log \mathcal{M}_i) = (\prod_{i=1}^n \mathcal{M}_i)^{1/n}$ *of the true indices. We have* $E(n^{-1}\sum_{i=1}^n \log \mathcal{M}_i) = \mu_{\mathcal{M}}$*, but* $E(M_n)$ *depends on* n*, and by Jensen's inequality*

$$E(M_n) = E\left(\exp(\frac{1}{n}\sum_{i=1}^n \log \mathcal{M}_i)\right) > \exp\left(E(\frac{1}{n}\sum_{i=1}^n \log \mathcal{M}_i)\right) = \exp(\mu_{\mathcal{M}}). \tag{3.41}$$

Fortunately, Theorem 3.8 shows that this additional "bias" is asymptotically negligible, and inference for $\widehat{M}_n$ *follows from inference for* $\log \widehat{M}_n$.

4 Making Inference about Productivity Change

Applied researchers often report estimates of Malmquist indices for individual producers observed in periods 1 and 2. Confidence intervals for Malmquist indices of individual firms can be estimated using the sub-sampling methods described by Simar and Wilson (2011) while noting that the rate of convergence is n^κ where $\kappa = 2/(p+q+1)$. In addition, geometric means of estimated Malmquist indices are typically reported by applied researchers to give an idea of productivity change over all firms in a group or in the entire sample. Means are also useful for summarizing results when the sample size is large. Geometric means are used instead of arithmetic means in order to preserve the multiplicative nature of Malmquist indices (recall that the Malmquist index $\mathcal{M}_i$ defined in (2.9) is a geometric mean of two ratios). Values of such geometric means greater than one suggest improved overall productivity between periods 1 and 2, while values less than one suggest a decrease in overall

productivity between the two periods. In this section, we show how the results obtained so far can be used to make inference about overall changes in productivity working either with arithmetic means of logs of estimated Malmquist indices in Section 4.1 or with geometric means of estimated Malmquist indices in Section 4.2.

4.1 Inference Based on Arithmetic Means of Logs

Theorem 3.7 from Section 3.2 provides the basis for making inference about productivity change while working with arithmetic means of estimated Malmquist indices. To simplify notation, let $\sigma_{\mathcal{M}} = \text{VAR}(\log \mathcal{M}_i) = E\left((\log \mathcal{M}_i - E(\log \mathcal{M}_i))^2\right)$ where the expectations are over (X, Y) in both periods 1 and 2. Recall that in the definition of $\mu_{\mathcal{M}}$ in (3.24) the expectation is also with respect to (X, Y) in both periods 1 and 2. Assume both $\mu_{\mathcal{M}}$ and $\sigma_{\mathcal{M}}$ are finite.

Theorem 3.7 can now be written as

$$\sqrt{n}\left(\widehat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}} - C_{\mathcal{M}} n^{-\kappa} - \xi_{n,\kappa}\right) \xrightarrow{d} N\left(0, \sigma^2_{\mathcal{M}}\right) \tag{4.1}$$

where $\kappa = 2/(p+q+1)$ and $\xi_{n,\kappa} = O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right) = o\left(n^{-\kappa}\right)$.

The following lemma provides a consistent estimator of the variance term $\sigma^2_{\mathcal{M}}$ in (4.1).

Lemma 4.1. *Under the conditions of Theorem 3.7,*

$$\widehat{\sigma}^2_{\mathcal{M},n} = n^{-1} \sum_{i=1}^{n} \left(\log \widehat{\mathcal{M}}_i - \widehat{\mu}_{\mathcal{M},n}\right)^2 \xrightarrow{p} \sigma^2_{\mathcal{M}}. \tag{4.2}$$

From (4.1) it is clear that $\widehat{\mu}_{\mathcal{M},n}$ is a consistent estimator of $\mu_{\mathcal{M}}$, but with a bias of $C_{\mathcal{M}} n^{-\kappa}$ since $E\left(\mu_{\mathcal{M},n}\right) = \mu_{\mathcal{M}} + C_{\mathcal{M}} n^{-\kappa}$. If $\kappa > 1/2$, then the bias term as well as the remainder term $\xi_{n,\kappa}$ are dominated by the factor $\sqrt{n}$ and therefore can be ignored. Hence when $\kappa > 1/2$, a $(1-\alpha) \times 100$-percent confidence interval for $\widehat{\mu}_{\mathcal{M},n}$ is estimated by

$$\left[\widehat{\mu}_{\mathcal{M},n} \pm z_{1-\frac{\alpha}{2}} \frac{\widehat{\sigma}_{\mathcal{M},n}}{\sqrt{n}}\right], \tag{4.3}$$

where $z_{1-\frac{\alpha}{2}}$ is the corresponding quantile of the standard normal distribution function. Under the conditions of Theorem 3.7, provided $\kappa > 1/2$ (i.e., $p + q \leq 2$), the interval in (4.3) has asymptotically correct coverage.

However, if $\kappa = 1/2$, the bias in (4.1) is constant, and if $\kappa < 1/2$, the bias tends to infinity as $n \to \infty$. In cases where $\kappa \leq 1/2$, Replacing the scaling factor $\sqrt{n}$ with n^{ζ} with $\zeta \in (0, \kappa)$

does not solve the problem. Doing so would drive the bias to zero as $n \to \infty$, but would also drive the variance to zero, resulting in a degenerate limiting distribution and preventing inference from being made. Moreover, note that $\kappa > 1/2$ if and only if $(p+q) \leq 2$. Theorem 3.7 and (4.1) can be used for inference in situations where the dimension-reduction methods of Wilson (2018) can be reasonably used to reduce a $(p+q)$-dimensional problem to only two dimensions, but otherwise the bias term must be dealt with. An approach similar to the one of Kneip et al. (2015) is useful here.

Suppose $\kappa \leq 1/2$. Let $n_\kappa = \min(\lfloor n^{2\kappa} \rfloor,\ n)$, where $\lfloor a \rfloor$ denotes the largest integer less than or equal to a. Assume that the observations in $\mathcal{X}_n$ are randomly sorted (the algorithm described by Daraio et al., 2018, Appendix D can be used to randomly sort the observations while allowing results to be replicated by other researchers using the same data and the same sorting algorithm). Let

$$\widehat{\mu}_{\mathcal{M},n_\kappa} := n_\kappa^{-1} \sum_{i=1}^{n_\kappa} \log \widehat{\mathcal{M}}_i \tag{4.4}$$

where the estimates $\widehat{\mathcal{M}}_i$ are computed using n (not n_κ) observations; i.e., the 4 estimates comprising $\widehat{\mathcal{M}}_i$ are each computed using all of the available observations in each period. The next result establishes the properties of this estimator.

Theorem 4.1. *Under the conditions of Theorem 3.7, for cases where $\kappa \leq 1/2$,*

$$n^\kappa \left(\widehat{\mu}_{\mathcal{M},n_\kappa} - \mu_{\mathcal{M}} - C_{\mathcal{M}} n^{-\kappa} - \xi_{n,\kappa} \right) \xrightarrow{d} N\left(0, \sigma^2_{\mathcal{M}}\right) \tag{4.5}$$

as $n \to \infty$, where $\xi_{n,\kappa} = O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}} \right)$.

The bias term $C_{\mathcal{M}} n^{-\kappa}$ remains in (4.5), but it is now multiplied by the factor n^κ and hence is constant instead of exploding to infinity as before when $\kappa < 1/2$. In order to estimate the bias, a generalized jackknife estimator similar to the one proposed by Kneip et al. (2015) can be used, taking care to split the data into sub-samples appropriately for the two periods in which firms are observed.

In order to simplify notation, let $Z_i^t = (X_i^t, Y_i^t)$, $t \in \{1,\ 2\}$ so that the sample can be described by $\mathcal{X}_n = \{(Z_i^1, Z_i^2)\}_{i=1}^n$. Now split $\mathcal{X}_n$ randomly into two sub-samples $\mathcal{X}_{m_1}^{(1)}$ and $\mathcal{X}_{m_2}^{(2)}$ of sizes $m_1 = \lfloor n/2 \rfloor$ and $m_2 = n - \lfloor n/2 \rfloor$ (respectively). Note that if n is even, $m_1 = m_2$, but if n is odd then $m_1 = m_2 - 1$. Asymptotically, this makes no difference since $m_1/m_2 \to 1$

as $n \to \infty$. Define

$$\widehat{\mu}^{(j)}_{\mathcal{M},m_j} := m_j^{-1} \sum_{(Z_i^1, Z_i^2) \in \mathcal{X}^{(j)}_{m_j}} \log \widehat{\mathcal{M}}_i(\mathcal{X}^{(j)}_{m_j}) \tag{4.6}$$

for $j \in \{1,\ 2\}$, where the notation $\widehat{\mathcal{M}}_i(\mathcal{X}^{(j)}_{m_j})$ indicates that the four estimates comprising the estimated Malmquist index $\widehat{\mathcal{M}}_i$ are each computed for observation i in the jth sub-sample using only the observations in the jth sub-sample $\mathcal{X}^{(j)}_{m_j}$. Then set

$$\widehat{\mu}^{*}_{\mathcal{M},n/2} = \frac{1}{2}\left(\widehat{\mu}^{(1)}_{\mathcal{M},m_1} + \widehat{\mu}^{(2)}_{\mathcal{M},m_2}\right). \tag{4.7}$$

By following arguments similar to those in Kneip et al. (2015, Section 4) it is easy to show that

$$\widetilde{B}_{n,\kappa} = (2^{\kappa} - 1)^{-1}\left(\widehat{\mu}^{*}_{\mathcal{M},n/2} - \widehat{\mu}_{\mathcal{M},n}\right) = C_{\mathcal{M}} n^{-\kappa} + \xi^{*}_{n,\kappa} + o_p\left(n^{-1/2}\right), \tag{4.8}$$

where $\xi^{*}_{n,\kappa}$ is of the same order as $\xi_{n,\kappa}$ appearing in (4.1), provides an estimator of the bias $C_{\mathcal{M}} n^{-\kappa}$.

Note that there are $\binom{n}{n/2}$ possible splits of the original n observations. To reduce the variance of the bias estimate in (4.8), the sample can be split $K << \binom{n}{n/2}$ times while randomly shuffling the observations before each split, and computing $\widetilde{B}_{n,\kappa,k}$ using (4.8) for $k = 1,\ \ldots,\ K$. Then

$$\widehat{B}_{n,\kappa} = K^{-1} \sum_{k=1}^{K} \widetilde{B}_{n,\kappa} \tag{4.9}$$

gives a generalized jackknife estimate of the bias $C_{\mathcal{M}} n^{-\kappa}$ (Gray and Schucany, 1972, Definition 2.1). Averaging in (4.9) reduces the variance by a factor of K^{-1} relative to the bias in (4.8).

Combining Theorem 3.7 and (4.9) leads to the following result.

Theorem 4.2. *Under the conditions of Theorem 3.7, for cases where* $\kappa \geq 2/5$,

$$\sqrt{n}\left(\widehat{\mu}_{\mathcal{M},n} - \widehat{B}_{n,\kappa} - \mu_{\mathcal{M}} + \xi_{n,\kappa}\right) \xrightarrow{d} N\left(0, \sigma^2_{\mathcal{M}}\right) \tag{4.10}$$

as $n \to \infty$.

The interplay between the root-n scaling factor and the remainder term $\xi_{n,\kappa}$ ensures that the result in Theorem 4.10 holds for $\kappa \geq 2/5$, and hence for $(p + q) \leq 4$. However, it is important to note that Theorem 4.2 does not hold in cases where $\kappa < 2/5$. In such

cases, remainder term $\xi_{n,\kappa}$, when multiplied by $\sqrt{n}$, diverges toward infinity. Alternatively, combining Theorem 4.1 and (4.9) yields the following result.

Theorem 4.3. *Under the conditions of Theorem 3.7, for cases where $\kappa < 1/2$,*

$$n^{\kappa}\left(\widehat{\mu}_{\mathcal{M},n_{\kappa}} - \widehat{B}_{n,\kappa} - \mu_{\mathcal{M}} - \xi_{n,\kappa}\right) \xrightarrow{d} N\left(0, \sigma^2_{\mathcal{M}}\right) \tag{4.11}$$

as $n \to \infty$.

Note that in all cases (i.e., for all values of κ), $\xi_{n,\kappa} = o(n^{-\kappa})$ and hence $n^{\kappa}\xi_{n,\kappa} = o(1)$. Therefore the remainder term can be neglected.

Whenever $\kappa \geq 2/5$ and hence $(p+q) \leq 4$, Theorem 4.2 can be used to construct an asymptotically correct $(1-\alpha)$ confidence interval for $\mu_{\mathcal{M}}$ given by

$$\left[\widehat{\mu}_{\mathcal{M},n} - \widehat{B}_{n,\kappa} \pm z_{1-\frac{\alpha}{2}}\frac{\widehat{\sigma}_{\mathcal{M},n}}{\sqrt{n}}\right], \tag{4.12}$$

where as in (4.3) $z_{1-\frac{\alpha}{2}}$ represents the $\left(1-\frac{\alpha}{2}\right)$ quantile of the standard normal distribution function.

Alternatively, in cases where $\kappa < 1/2$ and hence $(p+q) \geq 4$, Theorem 4.3 permits construction of the asymptotically correct $(1-\alpha$ confidence interval

$$\left[\widehat{\mu}_{\mathcal{M},n_{\kappa}} - \widehat{B}_{n,\kappa} \pm z_{1-\frac{\alpha}{2}}\frac{\widehat{\sigma}_{\mathcal{M},n}}{n^{\kappa}}\right] \tag{4.13}$$

for $\mu_{\mathcal{M}}$. This interval is centered on $\widehat{\mu}_{\mathcal{M},n_{\kappa}} - \widehat{B}_{n,\kappa}$, and $\widehat{\mu}_{\mathcal{M},n_{\kappa}}$ computed from a random subset of estimates $\widehat{\mathcal{M}}_i$ (where each estimate $\widehat{\mathcal{M}}_i$ is computed using all of the sample observations in $\mathcal{X}_n$). While this may seem arbitrary, note that any confidence interval for $\mu_{\mathcal{M}}$ is arbitrary since any asymmetric confidence interval for $\mu_{\mathcal{M}}$ can be constructed simply by using different quantiles of the $N(0,1)$ distribution to establish the bounds. The main point is always to achieve a high level of coverage without making the confidence interval too wide to be informative.

In cases where $\kappa < 1/2$, the randomness of the interval in (4.13) due to centering on a mean over a subsample of size $n_{\kappa} < n$ can be eliminated by averaging the center of (4.13) over all possible draws (without replacement) of subsamples of size n_{κ}. This yields an interval

$$\left[\widehat{\mu}_{\mathcal{M},n} - \widehat{B}_{n,\kappa} \pm z_{1-\frac{\alpha}{2}}\frac{\widehat{\sigma}_{\mathcal{M},n}}{n^{\kappa}}\right] \tag{4.14}$$

centered on $\widehat{\mu}_{\mathcal{M},n} - \widehat{B}_{n,\kappa}$. The only difference between the intervals in (4.13) and (4.14) is the centering value. Both intervals have the same length and hence are equally informative. But the interval in (4.14) should be more accurate (i.e., should have higher coverage in finite samples) because the estimator $\widehat{\mu}_{\mathcal{M},n}$ uses more information than the estimator $\widehat{\mu}_{\mathcal{M},n_\kappa}$. Therefore, for $\kappa < 1/2$, $n_\kappa < n$ and hence the interval in (4.14) contains the true value $\mu_{\mathcal{M}}$ with probability greater than $(1-\alpha)$. Due to the results given above, it is clear that the coverage of the interval in (4.14) converges to 1 as $n \to \infty$.

Note that when $(p+q) = 4$, either Theorems 4.2 or 4.3 can be used to provide different but asymptotically correct confidence intervals for $\mu_{\mathcal{M}}$. The interval in (4.12) uses the scaling factor $\sqrt{n}$ and hence neglects the term $\sqrt{n}\xi_{n,\kappa} = O\left(n^{-1/10}\right)$ in Theorem 4.2. By contrast, the interval in (4.13) uses the scaling factor n^κ and hence neglects the term $n^\kappa \xi_{n,\kappa} = O\left(n^{-1/5}\right)$ in Theorem 4.3. Therefore one should expect (4.13) to provide a better approximation in finite samples than (4.12) when $(p+q) = 4$.

The null hypothesis of no change in productivity versus change in productivity between periods 1 and 2 can be tested by computing the appropriate interval for $\mu_{\mathcal{M}}$. Under the null, $\mu_{\mathcal{M}} = 0$, while under the alternative hypothesis, $\mu_{\mathcal{M}} \neq 0$. Hence the null is rejected whenever the estimated confidence interval does not include zero.

The intervals given so far in (4.3), (4.12) and (4.13) are for $\mu_{\mathcal{M}}$ defined in (3.24). Theorem 3.8 and Remark 3.7 ensure that these intervals can be used to make inference about the geometric mean $E(M_n)$ where M_n is defined by (2.10). In particular, asymptotically valid intervals for $E(M_n)$ are obtained by taking exponentials of the bounds of the appropriate interval for $\mu_{\mathcal{M}}$.

4.2 Inference Based on Geometric Means

Theorem 3.8 from Section 3.2 provides the starting point for making inference about productivity change while working with geometric means of estimated Malmquist indices. Combining (3.39) and (3.40), the result of Theorem 3.8 can be expressed as

$$\sqrt{n}\left(\widehat{M}_n - \exp(\mu_{\mathcal{M}})C_{\mathcal{M}}n^{-\kappa} - \exp(\mu_{\mathcal{M}}) + \eta_{n,\kappa}\right) \xrightarrow{d} N\left(0,\ \exp(2\mu_{\mathcal{M}})\sigma^2_{\mathcal{M}}\right) \tag{4.15}$$

where $\kappa = 2/(p+q+1)$ and $\eta_{n,\kappa+} = O\left(n^{-\frac{3}{p+q+1}}(\log n)^{\frac{3}{p+q+1}}\right) = o\left(n^{-\kappa}\right)$. Viewing the result this way confirms the observation in Remark 3.7; i.e., from (4.15) it is clear that $\widehat{M}_n$ is a

consistent estimator of its asymptotic expectation $\exp(\mu_{\mathcal{M}})$ with a bias of $\exp(\mu_{\mathcal{M}})C_{\mathcal{M}}n^{-\kappa}$.

The variance $\exp(2\mu_{\mathcal{M}})\sigma^2_{\mathcal{M}}$ in (4.15) is consistently estimated by $\exp\left(2\widehat{\mu}_{\mathcal{M},n}\right)\widehat{\sigma}^2_{\mathcal{M},n}$ due to the results in Section 4.1. However, similar to the situation in Section 4.1, the result in (4.15) can only be used to make inference when $\kappa > 1/2$. As before, if $\kappa = 1/2$, the bias is constant, whereas if $\kappa < 1/2$ the bias explodes as $n \to \infty$. When $\kappa > 1/2$, (4.15) can be used to construct the $(1-\alpha)\times 100$-percent confidence interval

$$\left[\widehat{M}_n \pm z_{1-\frac{\alpha}{2}} \frac{\exp\left(\widehat{\mu}_{\mathcal{M},n}\right)\widehat{\sigma}_{\mathcal{M},n}}{\sqrt{n}}\right] \tag{4.16}$$

for $\exp(\mu_{\mathcal{M}})$, but otherwise the bias must be reckoned with.

Suppose $\kappa \leq 1/2$, and let recall that $n_\kappa = \min\left(\lfloor n^{2\kappa} \rfloor,\ n\right)$. Assume the observations in $\mathcal{X}_n$ are randomly sorted. Define

$$\widehat{M}_{n_\kappa} := \left(\prod_{i=1}^{n_\kappa} \widehat{\mathcal{M}}_i\right)^{1/n_\kappa} \tag{4.17}$$

where the estimates $\widehat{\mathcal{M}}_i$ are computed using the full sample of observations, but the geometric mean is over only the first n_κ estimates of $\mathcal{M}_i$. The properties of this estimator are established by the following theorem.

Theorem 4.4. *Under the conditions of Theorem 3.8, for cases where $\kappa \leq 1/2$,*

$$n^\kappa\left(\widehat{M}_{n_\kappa} - \exp(\mu_{\mathcal{M}})C_{\mathcal{M}}n^{-\kappa} - \exp(\mu_{\mathcal{M}}) + \eta_{n,\kappa}\right) \xrightarrow{d} N\left(0,\ \exp(2\mu_{\mathcal{M}})\sigma^2_{\mathcal{M}}\right) \tag{4.18}$$

as $n \to \infty$.

The bias term is stabilized in (4.18) and no longer explodes as $n \to \infty$ when $\kappa < 1/2$. For any $\kappa \leq 1/2$, the bias is constant and can be estimated using a generalized jackknife similar to the one used earlier in Section 4.1. In particular, split the sample as described in Section 4.1 into two sub-samples $\mathcal{X}^{(j)}_{m_j}$, $j \in \{1,\ 2\}$. Define

$$\widehat{M}^{(j)}_{m_j} := \left(\prod_{(Z^1_i, Z^2_i)\in\mathcal{X}^{(j)}_{m_j}} \widehat{\mathcal{M}}_i\left(\mathcal{X}^{(j)}_{m_j}\right)\right)^{1/m_j} \tag{4.19}$$

for $j \in \{1,\ 2\}$. As in (4.6), the notation $\widehat{\mathcal{M}}_i\left(\mathcal{X}^{(j)}_{m_j}\right)$ serves to remind that the estimates under the product sign are computed using only the estimates in the jth subsample $\mathcal{X}^{(j)}_{m_j}$.

Next, define

$$\widehat{M}^*_{n/2} = \frac{1}{2}\left(\widehat{M}^{(1)}_{m_1} + \widehat{M}^{(2)}_{m_2}\right). \tag{4.20}$$

Using arguments similar to those of Kneip et al. (2015, Section 4) it is clear that

$$\widetilde{B}^*_{n,\kappa} = (2^\kappa - 1)^{-1}\left(\widehat{M}^*_{n/2} - \widehat{M}_n\right) = \exp(\mu_{\mathcal{M}})C_{\mathcal{M}}n^{-\kappa} + \eta^*_{n,\kappa} + o_p\left(n^{-1/2}\right), \tag{4.21}$$

where $\eta^*_{n,\kappa}$ is of the same order as $\eta_{n,\kappa}$ in (4.15), provides an estimator of the bias term $\exp(\mu_{\mathcal{M}})C_{\mathcal{M}}n^{-\kappa}$.

As discussed above in Section 4.1, the variance of the bias estimate $\widetilde{B}^*_{n,\kappa}$ can be reduced by randomly splitting the sample $K << \binom{n}{n/2}$ times and using (4.21) to compute $\widetilde{B}^*_{n,\kappa,k}$ for each split $k = 1, \ldots, K$ and then computing

$$\widehat{B}^*_{n,\kappa} = K^{-1}\sum_{k=1}^{K}\widetilde{B}^*_{n,\kappa,k}. \tag{4.22}$$

Combining (4.22) and Theorem 3.8 leads to the following result.

Theorem 4.5. *Under the conditions of Theorem 3.8, for cases where $\kappa \geq 2/5$,*

$$\sqrt{n}\left(\widehat{M}_n - \widehat{B}^*_{n,\kappa} - \exp(\mu_{\mathcal{M}}) + \eta_{n,\kappa}\right) \xrightarrow{d} N\left(0,\ \exp(2\mu_{\mathcal{M}})\sigma^2_{\mathcal{M}}\right) \tag{4.23}$$

as $n \to \infty$.

As with Theorem 4.2, the result here holds for $\kappa \geq 2/5$ for similar reasons. In addition, Theorem 4.5 does not hold for $\kappa < 2/5$. In such cases, the next result combining Theorem 4.4 and (4.22) is useful for inference.

Theorem 4.6. *Under the conditions of Theorem 3.8, for cases where $\kappa < 1/2$,*

$$n^\kappa\left(\widehat{M}_{n_\kappa} - \widehat{B}^*_{n,\kappa} - \exp(\mu_{\mathcal{M}}) + \eta_{n,\kappa}\right) \xrightarrow{d} N\left(0,\ \exp(2\mu_{\mathcal{M}})\sigma^2_{\mathcal{M}}\right) \tag{4.24}$$

as $n \to \infty$.

Recall that $\eta_{n,\kappa} = o\left(n^{-\kappa}\right)$. Therefore $n^\kappa\eta_{n,\kappa} = o(1)$, and hence the remainder term can be ignored.

For cases where $\kappa \geq 2/5$, Theorem 4.5 permits construction of an asymptotically correct $(1-\alpha)$ confidence interval for $\exp(\mu_{\mathcal{M}})$ given by

$$\left[\widehat{M}_n - \widehat{B}^*_{n,\kappa} \pm z_{1-\frac{\alpha}{2}}\frac{\exp\left(\widehat{\mu}_{\mathcal{M},n}\right)\widehat{\sigma}_{\mathcal{M},n}}{\sqrt{n}}\right]. \tag{4.25}$$

Alternatively, whenever $\kappa < 1/2$, Theorem 4.6 can be used to construct the asymptotically correct $(1-\alpha)$ confidence interval

$$\left[\widehat{M}_{n_\kappa} - \widehat{B}^*_{n,\kappa} \pm z_{1-\frac{\alpha}{2}} \frac{\exp\left(\widehat{\mu}_{\mathcal{M},n}\right)\widehat{\sigma}_{\mathcal{M},n}}{n^\kappa}\right]. \tag{4.26}$$

Analogous to the discussion in Section 4.1, one could also replace $\widehat{M}_{n_\kappa}$ with $\widehat{M}_n$ in (4.26), with the coverage of the resulting interval converging to 1 as $n \to \infty$.

Also as in Section 4.1, one can use either of the intervals in (4.25) and (4.26) when $(p+q) = 4$. The interval in (4.25) uses the scaling factor $\sqrt{n}$ and hence neglects the term $\sqrt{n}\eta_{n,\kappa} = O\left(n^{-1/10}\right)$ in Theorem 4.5, while the interval in (4.13) uses the scaling factor n^κ and hence neglects the term $n^\kappa\eta_{n,\kappa} = O\left(n^{-1/5}\right)$ in Theorem 4.6. Therefore one should expect (4.26) to provide a better approximation in finite samples than (4.25) when $(p+q) = 4$. For testing purposes, however, one cannot escape the tradeoff between size and power. This issue is further examined in the section reporting simulation results that follows below.

The null hypothesis of no productivity change corresponds to $\exp(\mu_{\mathcal{M}}) = 1$, while the alternative hypothesis of change in productivity between periods 1 and 2 corresponds to $\exp(\mu_{\mathcal{M}}) \neq 1$. Hence the null is rejected whenever the relevant estimated confidence interval in (4.25) or (4.26) does not include unity. The results of such tests are expected to be similar to the results of similar tests based on log values discussed near the end of Section 4.1, but some differences may arise due to the different asymptotic approximations involved. This issue is also examined in the next section.

5 Monte Carlo Evidence

5.1 Experimental Framework

In order to examine the size and power of tests of H_0: no productivity change versus H_1: productivity change between two periods using Malmquist indices, we need to simulate data from two different DGPs where (i) the technology in each period is characterized by VRS and (ii) the amount of productivity change between the two periods can be controlled, varied, and known.

First, consider the function $\mathbb{R}^1 \mapsto \mathbb{R}^1$: $t = \psi(s \mid \delta)$ such that

$$t = c\left(a^2 + \delta^2 s^2\right)^{1/2} - d \tag{5.1}$$

where $a = 0.5$, $c = 0.75$, $d = 0.375$ and $\delta \in \{0,\ 0.6\}$. Clearly, when $\delta = 0$ $t = 0$ for all values of s. When $\delta = 0.6$, $t \geq 0$ and the function is a convex (from below) parabola. The function $\psi(\cdot)$ is illustrated in panel (a) of Figure 1 for both values of δ. In the figure, the function is illustrated over the triangle with corners at $(-\sqrt{2}, 0)$, $(\sqrt{2}, 0)$ and $(0, \sqrt{2})$ in (s, t)-space.

Now consider the transformation from (s, t)-space to (v, y)-space obtained by rotating the curves shown in Figure 1(a) about the origin through a counter-clockwise angle of $5\pi/4$ radians, and then shifting by a distance 1 northeast along the ray from the origin at angle $\pi/4$. In general, the rotation matrix

$$R(\phi) = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \tag{5.2}$$

can be used to rotate a point around the origin by an angle ϕ.[5] Then for $R_1 = R(5\pi/4)$ we have

$$\begin{bmatrix} v & y \end{bmatrix}' = R_1 \begin{bmatrix} s & t \end{bmatrix}' + \begin{bmatrix} 1 & 1 \end{bmatrix}'. \tag{5.3}$$

The resulting curve (corresponding to $\delta = 0.6$) is shown in panel (b) of Figure 1, with the line from the origin (corresponding to $\delta = 0$) tangent at the point $(1, 1)$ in (v, y)-space.

Now consider rotating the curves shown in Figure 1(b) counter-clockwise around the origin by an angle $\omega \in [0, 0.1\pi]$. This amounts to transforming points (s, t) lying on the function $t = \psi(s \mid \delta)$ in (s, t)-space to points (v, y) where

$$\begin{bmatrix} v & y \end{bmatrix}' = R_2 \left(R_1 \begin{bmatrix} s & t \end{bmatrix}' + \begin{bmatrix} 1 & 1 \end{bmatrix}' \right) \tag{5.4}$$

with $R_2 = R(\omega)$. Panel (c) of Figure 1 shows the curves from panel (b) of the same figure as dashed curves—here, $\omega = 0$. Panel (c) also shows in solid curves the corresponding functions when $\omega = 0.05\pi$. It is easy to see that if $\omega = 0$ then R_2 is an identity matrix and (5.4) reduces to (5.3).

For $\delta = 0.6$ and a given value ω, simulating a sample of n input-output pairs $\{((X_i^1, Y_i^1),\ (X_i^2, Y_i^2))\}_{i=1}^n$ for periods 1 and 2 is now straightforward. In both periods, set $\delta = 0.6$. In period 1, set $\omega = 0$. In period 2 set $\omega = \beta\pi$ where $\beta \in [0,\ 0.1]$. First, consider the point in Figure 1(a) where the strictly convex (from below) curve $t = \psi(s \mid \delta = 0.6)$ intersects the line $t = \sqrt{2} - s$. Simple algebra reveals that he s-coordinate of this point is

$$s_0 = \frac{\left(\sqrt{2} + d\right) - \left[\left(\sqrt{2} + d\right)^2 - (1 - c^2\delta^2)\left(\left(\sqrt{2} + d\right)^2 - c^2a^2\right)\right]^{1/2}}{(1 - c^2\delta^2)}. \tag{5.5}$$

[5] See, for example, Noble and Daniel (1977, pp. 411–413).

Then set $s_{\max} = \min\left(s_0,\ 0.9\sqrt{(2)}\right)$, $s_{\min} = -s_{\max}$. In order to induce correlation across periods, as is typical in real production data, we use Gaussian copulas. Let C_v denote a (2×2) correlation matrix with off-diagonal elements ρ_v and with Cholesky decomposition $\mathcal{C}_v = \mathcal{U}_v'\mathcal{U}_v$ where $\mathcal{U}_v$ is upper-triangular. Generate an $(n \times 2)$ matrix $\mathcal{R}_v$ of iid $N(0,1)$ pseudo-random deviates and set $\mathcal{S} = s_{\min} + \Phi(\mathcal{R}_v\mathcal{U}_v)\,(s_{\max} - s_{\min})$ where $\Phi(\cdot)$ denotes the standard normal distribution function applied element-by-element to the $(n \times 2)$ matrix $\mathcal{R}_v\mathcal{U}_v$. Then $\mathcal{S}$ is an $(n \times 2)$ matrix of uniform deviates on $(s_{\min},\ s_{\max})$ and the columns of $\mathcal{S}$ are correlated with correlation ρ_v. Finally, for each element s_{ij} of $\mathcal{S}$ the corresponding value t_{ij} can be computed using (5.1), and then the pair (s,t) can be used in (5.3) and (5.4) to obtain pseudo-random draws $\{((\widetilde{V}_i^1, Y_i^1),\ (\widetilde{V}_i^2, Y_i^2))\}_{i=1}^n$. For the case where $p = q = 1$, simply set $\widetilde{X}_i^t = \widetilde{V}_i^t$ for $t \in \{1,\ 2\}$. Then $\{(\widetilde{X}_i^1, Y_i^1)\}_{i=1}^n$ and $\{(\widetilde{X}_i^2, Y_i^2)\}_{i=1}^n$ are sets of *technically efficient* input output pairs lying on the frontiers in periods 1 and 2, respectively.

For the case of $q = 1$ and multivariate inputs with $p > 1$, an additional step is needed. First, generate pseudo-random draws $\{((\widetilde{V}_i^1, Y_i^1),\ (\widetilde{V}_i^2, Y_i^2))\}_{i=1}^n$ exactly as before. In order to induce cross-period correlation among firms' mixes of outputs, we again use a Gaussian copula. Construct the $(2p \times 2p)$ correlation matrix

$$\mathcal{C}_u = \begin{bmatrix} I_p & \rho_u I_p \\ \rho_u I_p & I_p \end{bmatrix} \tag{5.6}$$

with Cholesky decomposition $\mathcal{U}_u'\mathcal{U}_u$ where I_p denotes a $(p \times p)$ identity matrix, $\rho_u \in (0,1)$ is a scalar correlation coefficient and $\mathcal{U}_u$ is upper-triangular. Next, generate an $(n \times 2p)$ matrix $\mathcal{R}_u$ of iid $N(0,1)$ deviates, and compute the $(n \times 2p)$ matrix $\mathcal{Z} = \Phi(\mathcal{R}_u\mathcal{U}_u)$. Then partition $\mathcal{Z}$ by writing $\mathcal{Z} = \begin{bmatrix}\mathcal{Z}^1 & \mathcal{Z}^2\end{bmatrix}$ where $\mathcal{Z}^1$ and $\mathcal{Z}^2$ are $(n \times p)$. Note that $\mathcal{Z}^1$ and $\mathcal{Z}^2$ each contain iid uniform (0,1) deviates, but corresponding elements of $\mathcal{Z}^1$ and $\mathcal{Z}^2$ have correlation ρ_u by construction. For $t \in \{1,\ 2\}$ let $\mathcal{Z}^{*t}$ denote the $(n \times 1)$ vector obtained by summing the p columns of $\mathcal{Z}^t$. Compute $\mathcal{W} = \begin{bmatrix}\mathcal{W}^1 & \mathcal{W}^2\end{bmatrix}$ where $\mathcal{W}^t = \mathcal{Z}^t/(\mathcal{Z}^{*t}\boldsymbol{i}_p')$ where $\boldsymbol{i}_p$ denotes a $(p \times 1)$ vector of ones and the division is understood to be element-by-element. The $(n \times p)$ sub-matrices $\mathcal{W}^t$ contain weights in each row that sum to one, and which are correlated across $t = 1$ and 2. The p weights in a given row of $\mathcal{W}^t$ amount to p independent uniform random deviates divided by their sum. Finally, for period $t \in \{1,\ 2\}$, set $\widetilde{X}_{ki}^t = (\widetilde{V}_i^t)^{p\mathcal{W}_{i,k}^2}$ to form the $(p \times n)$ matrices $\widetilde{X}^t - \left[\widetilde{X}_{ki}^t\right]$. The vector $\widetilde{X}_{\cdot i}^t$ is the efficient level of outputs for firm i in period j, and $\widetilde{X}_{\cdot i}^1$ and$\widetilde{X}_{\cdot i}^2$ are correlated (provided $\rho_u > 0$) by construction as is typical

in real production data.

Note that one can solve for s in the two equations represented by (5.4), equate the two resulting expressions, and then solve for y to obtain an explicit expression for the function $y = g(v \mid \delta, \omega)$. Then univariate input with $p = 1$, the DGP described above amounts a technology described by

$$Y_i^t = g(\widetilde{X}_i^t \mid \delta = 0.6, \omega). \tag{5.7}$$

In the case of multivariate inputs $\widetilde{X}_{ij}^t$, $j = 1, \ldots, p$, the technology corresponds to

$$Y_i^t = g\left(\prod_{j=1}^{p}\left(\widetilde{X}_{ij}^t\right)^{1/p} \mid \delta = 0.6, \omega\right). \tag{5.8}$$

It remains to add inefficiency in both periods. Again, we employ a Gaussian copula to maintain correlation between the efficiency of firm i in period 1 and in period 2. Let $\mathcal{C}_\theta$ denote the (2×2) correlation matrix with correlation coefficient ρ_θ, and via Cholesky decomposition let $\mathcal{C}_\theta = \mathcal{U}_\theta'\mathcal{U}_\theta$ where $\mathcal{U}_\theta$ is upper-triangular. Next, generate an $(n \times 2)$ matrix $\mathcal{R}_\theta$ of iid $N(0,1)$ deviates. Then compute the $(n \times 2)$ matrix $\mathcal{J} = \Phi(\mathcal{R}_\theta\mathcal{U}_\theta)$. The elements of $\mathcal{J}$ are thus uniformly distributed on $(0,1)$, and are independent within a given column but have correlation ρ_θ across the two columns. Now for the (i,t)th element $\mathcal{J}_{it}$ of $\mathcal{J}$, compute the $(n \times 2)$ matrix $\Theta = \left[\theta_{it}\right]$ where $\theta_{it} = \beta^{-1}(\mathcal{J}_{it} \mid a_\theta = 4, b_\theta = 1)$ and $\beta^{-1}(\cdot \mid a_\theta, b_\theta)$ is the quantile function of a beta distribution with shape parameters a_θ and b_θ.[6] Then the columns of Θ contain iid beta(4,1) random values with expected value $E(\theta_{it}) = a_\theta(a_\theta + b_\theta) = 0.8)$. The values in $\mathcal{T}$ are correlated across the two columns with correlation ρ_θ by construction. Finally, for $i = 1, \ldots, n$ and $t = 1, 2$ set $X_{\cdot,i}^t = \theta_{it}^{-1}\widetilde{X}_{\cdot,i}$ to form the simulated sample $\{((X_i^1, Y_i^1), (X_i^2, Y_i^2))\}_{i=1}^n$.

As noted above, it is easy to solve for $g(v \mid \delta, \omega)$. The function is monotonic, and hence after some additional algebra one obtains the inverse function $v = g^{-1}(y \mid \delta, \omega)$. The inverse function $g^{-1}(y \mid \delta, \omega)$ can be used to find the expected value of geometric means of Malmquist indices for various values of ω, with the expectation over X and Y.

Recall that in cases where $p > 1$, the $\widetilde{V}$s are replaced by a homogeneous (of degree 1) function of efficient input levels, which are then projected away from the frontier to reflect inefficiency. Clearly, however, one could obtain the same values of the inefficient levels of

[6] The beta quantile function has no closed-form expression, but can be computed numerically using the algorithm of Cran et al. (1977).

inputs by computing $V = \theta^{-1}\widetilde{V}$ and then replacing V with a homogeneous function of input levels simulated as before. Consequently, for purposes of evaluating the expected value of the geometric mean of Malmquist indices, we can work in the two dimensional space of (v, y).

Now consider a simulated observation $((\widetilde{V}_i^1, Y_i^1),\ (V_i^2, Y_i^2)))$ generated as described above for a particular value of the angle of rotation ω. For (V_i^1, Y_i^1) define $\zeta_i^{11} = g^{-1}(Y_i^1 \mid 0, 0)/V_i^1$, where $(g^{-1}(Y_i^1 \mid 0, 0), Y_i^1)$ gives the projection of (V_i^1, Y_i^1) onto the conical hull of the technology in period 1 in the input direction. We can also define $\zeta_i^{12} = g^{-1}(Y_i \mid 0, \omega)/V_i^1$, where $g^{-1}(Y_i^1 \mid 0, \omega)$ gives the projection of (V_i^1, Y_i^1) onto the conical hull of the technology in period 2 in the input direction. Similarly, for (V_i^2, Y_i^2) define $\zeta_i^{22} = g^{-1}(Y_i^2 \mid 0, \omega)/V_i^2$, where $(g^{-1}(Y_i^2 \mid 0, \omega), Y_i^2)$ gives the projection of (V_i^2, Y_i^2) onto the conical hull of the technology in period 2 in the input direction. In addition, define $\zeta_i^{21} = g^{-1}(Y_i^2 \mid 0, 0)/V_i^2$, where $g^{-1}(Y_i^2 \mid 0, 0)$ gives the projection of (V_i^2, Y_i^2) onto the conical hull of the technology in period 1 in the input direction. Then the "true" Malmquist index defined in terms of input-oriented efficiency measures for observation i is $\mathcal{M}_{\theta,i} = [(\zeta_i^{21}/\zeta_i^{11})(\zeta_i^{22}/\zeta_i^{12})]^{1/2}$. For a given value of ω, we generate a sample of size $n = 1.28 \times 10^{12}$ and then compute the sample mean $\widehat{\mu}_{\mathcal{M}_\theta,n}$ of the values $\{\log(\mathcal{M}_{\theta,i})\}_{i=1}^n$. We then take the exponential of this sample mean, yielding an estimate of $\exp(\mu_{\mathcal{M}_\theta}) = \exp(E(\widehat{\mu}_{\mathcal{M}_n}))$ accurate to 5–6 decimal places. For $\beta = 0$, 0.005, 0.010, 0.015, 0.020, 0.030, 0.040 and 0.050 we have $\exp(\widehat{\mu}_{\mathcal{M},n}) = 1.0000$, 1.0132, 1.0270, 1.0415, 1.0566, 1.0892, 1.1250 and 1.1645 (respectively). These values provide some perspective helpful for interpreting the results discussed below.

5.2 Simulation Results

Tables 1–2 report results of 200 Monte Carlo experiments for $q = 1$ and $p \in \{1,\ 2,\ 3,\ 4,\ 5\}$, $n \in \{25,\ 50,\ 100,\ 500,\ 1,000\}$ and $\beta \in \{0,\ 0.005,\ 0.010,\ 0.015,\ 0.020,\ 0.030,\ 0.040,\ 0.050\}$ (the values of beta yield values for the angle ω equal to 0.000, 0.016, 0.031, 0.047, 0.063, 0.094, 0.126 and 0.157 radians, or 0.000, 0.900, 1.800, 2.700, 3.600, 5.400, 7.200 and 9.000 degrees, respectively). For each experiment (i.e., for each combination of p, n and β) we draw a sample of size n and test the null hypothesis of no productivity change versus the alternative hypothesis of productivity change for each of 250,000 trials within a given experiment. In both tables we report the proportion of samples where we reject the null in tests of nominal size .10, .05 and .01. With 250,000 trials in each experiment, the standard deviation of

the estimated rejection rates for tests of size 0.1, 0.05 and 0.01 are 0.000600, 0.000436 and 0.000199.

The results in Table 1 for $p \in \{1,\ 2,\ 3\}$ are from tests based on Theorem 4.2 and intervals computed using (4.12), while results for $p \in \{4,\ 5\}$ are based on Theorem 4.3 and intervals computed from (4.13). Results in Table 2 are obtained with untransformed Malmquist indices as opposed to the logs of Malmquist indices as in Table 1. In Table 2, results for $p \in \{1,\ 2,\ 3\}$ are obtained using Theorem 4.5 and intervals computed from (4.25), whereas results for $p \in \{4,\ 5\}$ are obtained using Theorem 4.6 and intervals computed from (4.26).

Both tests show good size properties and good power. For each value of n, the rows of results where $\beta = 0$ reflect the realized sizes of the tests for each of the three levels 0.1, 0.05 and 0.01. Even with only 25 observations, there is very little deterioration in the realized sizes as dimensionality increases from 2 to 6 moving from left to right in the either table. For each value of n, increasing values of β reflect increasing departures from the null hypothesis of no productivity change. While size is hardly affected by dimensionality, but there is an effect on power. As dimensionality is increases, the power for a given value of β and a given nominal test size decreases. Even so, with only 100 observations and a nominal test size of 0.1, the probability of rejecting the null is about 0.5 when $p = 5$ and $q = 1$ when relying on Theorem 4.2 as seen in Table 1.

Comparing results overall between the two tables reveals similar qualitative results, but there are some differences in the qualitative results across the two tables. While both tests appear to have similar size properties, the test based on Theorem 4.2 (i.e., working in log-space) yields greater power than the test based on Theorem 4.5 (i.e., working with the untransformed indices) for moderate departures from the null (e.g., for β between 0.005 and 0.015 or 0.020). This is not surprising in view of Remark 3.7. While the additional bias accumulated in moving from Theorem 3.7 to Theorem 3.8 is asymptotically negligible, in finite samples it degrades the performance in terms of power of tests based on the untransformed Malmquist indices relative to the performance of tests based on logs of the indices. Further comparison of the results across Tables 1 and 2 shows that the differences in power become smaller as n increases for given values of $(p+q)$, as one should expect from Remark 3.7, but the effect persists even when $n = 1000$ in the larger dimensions.

Rejection rates were also estimated from the same Monte Carlo experiments but using the re-centered interval in (4.14) when working with the log-Malmquist index and the re-centered interval obtained by replacing $\widehat{M}_{n_\kappa}$ with $\widehat{M}_n$ in (4.26). Results from this exercise confirm the remarks in Sections 4.1 and 4.2; i.e., using the re-centered intervals, the rejection rate when the null is true tends to 0 as $n \to \infty$ for .90, .95, and .99 confidence intervals, even though the re-centered intervals are of the same width as those in (4.13) and (4.26) used in Tables 1 and 2. Remarkably, using the re-centered intervals gives (asymptotically) superior coverage without widening the confidence intervals.[7]

As a final remark, recall from the discussion in Section 4 that when $(p+q) = 4$, one can use either Theorem 4.2 or Theorem 4.3 while working in log-space, or either Theorem 4.5 or Theorem 4.6 while working with the original, untransformed indices. As noted above, the results in Tables 1–2 for $p = 3,\ q = 1$ are based on Theorems 4.2 and Theorem 4.5, respectively. Using instead Theorems 4.3 and 4.3 when $(p+q) = 4$ yields slightly better size for both tests than when using Theorems 4.2 and 4.5. However, the power of the tests is worse when relying on Theorems 4.3 and 4.3 rather than Theorems 4.2 and 4.5. For example, when $n = 100$, $\beta = 0.000$, and $p = 3,\ q = 1$, the realized sizes in Table 1 are 0.113, 0.058 and 0.011 for tests of size 0.1, 0.05 and 0.01, while using Theorem 4.3 results in realized sizes of 0.106, 0.054 and 0.011. But when $\beta = 0.010$, the rejection rates in Table 1 are 0.789, 0.688 and 0.445 compared to rejection rats of 0.451, 0.332 and 0.149 obtained using Theorem 4.3. As noted near the end of Section 4, there is not escape from the inherent tradeoff between size and power. In this case, the differences in sizes are much smaller than the differences in power.

6 Summary and Conclusions

Malmquist indices estimated by nonparametric DEA estimators are widely used to analyze changes in productivity in dynamic settings, and applied researchers typically report geometric means of estimated Malmquist indices to summarize results and to make statements about overall changes in productivity. Until now, however, valid inference in the context of geometric means has been impossible. Moreover, until now, even the statistical properties

[7] Tables showing the rejection rates obtained with the re-centered intervals are not shown here in order to conserve space, but are available from the authors on request.

of estimators of Malmquist indices for individual firms have been unknown.

The results of Kneip et al. (2015) strongly suggest that standard CLT results cannot be used to make inference about means of logs of Malmquist indices. However, the results of Kneip et al. (2015) cannot be used directly, as estimating Malmquist indices requires the conical-DEA estimator introduced in Section 3. Kneip et al. (2015) provide results on moments of a constant-returns-to scale version of the DEA estimator *under constant returns to scale,* but the situation here is very different. When Malmquist indices are estimated, researchers often assume that variable returns to scale prevail in both periods under consideration. Properties of the conical-DEA estimator in this setting have not been examined until now.

This paper provides the tools needed by researchers for making inference either about changes in productivity of an individual firm or overall change in productivity. Theoretical results are provided that allow researchers to work in terms of logs of Malmquist indices or in terms of the original, untransformed indices. The simulation results presented in Section 5 strongly suggest that for purposes of testing the null hypothesis of no productivity change on average versus a change in productivity, one should work in terms of logs of Malmquist indices.

A Proofs and Technical Details

A.1 Proof of Lemma 3.1

Consider rays $\mathcal{L}_1 = \mathcal{L}(x, y)$ and $\mathcal{L}_2 = \mathcal{L}(x, \lambda(x, y \mid \mathcal{C}(\Psi))y) \subset \mathcal{C}^{\partial}(\Psi)$.

Since $(\theta(x, y \mid \mathcal{C}(\Psi))x,\ y) \in \mathcal{L}_2$ and $(x,\ \lambda(x, y \mid \mathcal{C}(\Psi))y) \in \mathcal{L}_2$, $\frac{\lambda(x,y|\mathcal{C}(\Psi))\|y\|}{\|x\|} = \frac{\|y\|}{\theta(x,y|\mathcal{C}(\Psi))\|x\|\|}$ and hence $\lambda(x, y \mid \mathcal{C}(\Psi))^{-1} = \theta(x, y \mid \mathcal{C}(\Psi))$. In addition, $(\gamma(x, y \mid \mathcal{C}(\Psi))x,\ \gamma(x, y \mid \mathcal{C}(\Psi))^{-1}y) \in \mathcal{L}_2$. Therefore $\frac{\gamma(x,y|\mathcal{C}(\Psi))^{-1}\|y\|}{\gamma(x,y|\mathcal{C}(\Psi))\|x\|} = \frac{\|y\|}{\theta(x,y|\mathcal{C}(\Psi))\|x\|}$. Result (i) follows immediately.

To prove (ii), consider two points $(x, y) \in \mathcal{L}_1$ and $(\widetilde{x}, \widetilde{y}) \in \mathcal{L}_1$. Clearly, $(x,\ \lambda(x, y \mid \mathcal{C}(\Psi)y) \in \mathcal{L}_2$ and $(\widetilde{,}\ \lambda(\widetilde{x}, \widetilde{y} \mid \mathcal{C}(\Psi)\widetilde{y}) \in \mathcal{L}_2$. It follows that $\frac{\lambda(x,y|\mathcal{C}(\Psi)\|y\|}{\|x\|} = \frac{\lambda(\widetilde{x},\widetilde{y}|\mathcal{C}(\Psi)\|\widetilde{y}\|}{\|\widetilde{x}\|}$. Hence $\lambda(x, y \mid \mathcal{C}(\Psi) = \lambda(\widetilde{x}, \widetilde{y} \mid \mathcal{C}(\Psi)$ since $\frac{\|y\|}{\|x\|} = \frac{\|\widetilde{y}\|}{\|\widetilde{x}\|}$, establishing (ii). Results (iii) and (iv) follow from (i) and (ii). ■

A.2 Proof of Lemma 3.2

The results follow from the proof of Lemma 3.1 after replacing $\mathcal{C}(\Psi)$ with $\mathcal{C}(\widehat{\Psi})$). ■

A.3 Some Background Material used in Proof of Theorem 3.1

The proof of Theorem 3.1 that follows relies on the structural analysis used in the proof of Theorem 3.1 in Kneip et al. (2015). Let us first recall some of the notation used in there.

Consider an arbitrary point $(x, y) \in \mathcal{D}$. Let $\mathcal{V}(x)$ denote the $(p-1)$-dimensional linear space of all vectors $z \in \mathbb{R}^p$ such that $z^T x = 0$. Any input vector X_i adopts a unique decomposition of the form $X_i = \gamma_i \frac{x}{\|x\|} + Z_i$ for some $Z_i \in \mathcal{V}(x)$ and $\gamma_i = \frac{x^T X_i}{\|x\|}$, where $\|\cdot\|$ denotes the Euclidean norm. Let $\Psi^*(x)$ denote the set of all $(z, y) \in \mathcal{V}(x) \times \mathbb{R}^q$ with $(\gamma \frac{x}{\|x\|} + z, y) \in \mathcal{D}$ for some $\gamma > 0$. Note that the point of interest $(x, y) \in \Psi$ has coordinates $(0, y)$ in $\Psi^*(x)$.

The maintained assumptions imply that for any $(z, y) \in \Psi^*(x)$, there exists $\gamma > 0$ such that $(\gamma \frac{x}{\|x\|} + z, y) \in \Psi$. The efficient boundary of Ψ can therefore be described by the function $g_x(z, y) := \inf \left\{ \gamma \mid \left(\gamma \frac{x}{\|x\|} + z, y \right) \in \Psi \right\}$ defined for any $(z, y) \in \Psi^*(x)$. Furthermore, with only a small abuse of notation, one may extend the definition of g_x to all (v, y) with $\left(v - \frac{x^T v}{\|x\|^2} x, y \right) \in \Psi^*(x)$ by taking $g_x(v, y) = g_x \left(v - \frac{x^T v}{\|x\|^2} x, y \right)$.

Properties of g_x are discussed in Kneip et al. (2008). In particular, under the assumptions of the theorem, g_x is a three times continuously differentiable, strictly convex function, and there exists a constant $C_1 > 0$ such that $w^T g_x''(0, y) w \geq C_1$ for all $w \in \mathcal{V}(x) \times \mathbb{R}^q$ with $\|w\| = 1$ and all $x \in \mathbb{R}^q$ with $(x, y) \in \mathcal{D}$. Moreover, $g_x''(0, y)$ changes continuously in x. In the following we will additionally use $g''_{x;zz}(\widetilde{z}, \widetilde{y})$ to denote the $(p-1) \times (p-1)$-matrix of partial derivatives with respect to the z-coordinates at a point $(\widetilde{z}, \widetilde{y})$, while $g''_{x;yy}(\widetilde{z}, \widetilde{y})$ will denote the $q \times q$-matrix of partial derivatives with respect to the y-coordinates.

The decomposition described above establishes a new coordinate system in which each observation (X_i, Y_i) can be equivalently represented by the corresponding vector (θ_i, Z_i, Y_i), where $\theta_i := \theta(X_i, Y_i)$. Any point (x, ay) of interest has coordinates $(\theta(x, ay), 0, ay)$ in this system.

Different from Kneip et al. (2015) we will need an additional decomposition of the vari-

ables Y_i

$$Y_i = \alpha_i y + V_i \quad \text{for some } V_i \in \mathbb{R}^q,\ V_i^t y = 0,\ \text{and } \alpha_i = \frac{y^T Y_i}{\|y\|^2}. \tag{A.1}$$

This establishes another coordinate system with $(Z_i, V_i) \in \mathcal{V}(x, y)$, where $\mathcal{V}(x, y)$ denotes the $(p-1)\times(q-1)$-dimensional linear space of all vectors $z \in \mathbb{R}^p$ and $v \in \mathbb{R}^p$ such that $z^T x = 0$ and $v^T y = 0$. Instead of using (θ_i, Z_i, Y_i), each observation (X_i, Y_i) can be equivalently represented by the corresponding vector $(\theta_i, Z_i, \alpha_i, V_i)$, where $\theta_i := \theta(X_i, Y_i)$. Any point (x, ay) of interest has coordinates $(\theta(x, ay),\ 0,\ ay,\ 0)$ in this new system.

Let $z_x^{(1)}, \ldots, z_x^{(p-1)}$ and $v_y^{(1)}, \ldots, v_y^{(q-1)}$ be orthonormal bases of Z_i and V_i. Clearly, the $z_x^{(j)}$ and $v_y^{(j)}$ can be chosen as continuous functions of x and y, respectively. Every vector Z_i can be expressed in the form $Z_i = \mathbf{Z}_x \zeta_i$, where $\mathbf{Z}_x$ is the $p \times (p-1)$ matrix with columns $z_x^{(j)}$, $j = 1, \ldots, p-1$, and $\zeta_i \in \mathbb{R}^{p-1}$. Similarly, every vector V_i can be expressed in the form $V_i = \mathbf{V}_y v_i$, where $\mathbf{V}_y$ is the $q \times (q-1)$ matrix with columns $v_y^{(j)}$, $j = 1, \ldots, q-1$, and $v_i \in \mathbb{R}^{q-1}$.

Since $\theta_i = \theta(X_i, Y_i)$, $Z_i = X_i - \frac{x^T X_i}{\|x\|^2} x$, $\alpha_i = \frac{y^T Y_i}{\|y\|^2}$, and $V_i = Y_i - \frac{y^T Y_i}{\|y\|^2} y$ are smooth functions of (X_i, Y_i), the joint density f of (X_i, Y_i) translates into a density $\widetilde{f}_{x,y}$ on $(0, 1] \times \mathbb{R}^{p-1} \times \mathbb{R} \times \mathbb{R}^{q-1}$ of $(\theta_i, \zeta_i, \alpha_i, v_i)$. Let $\widetilde{\mathcal{D}}$ denote the support of this density. Since f is continuously differentiable, $\widetilde{f}_{x,y}(\theta, \zeta, \alpha, v)$ is also continuously differentiable on $(\theta, \zeta, \alpha, v) \in \widetilde{\mathcal{D}}$. Furthermore, compactness of $\mathcal{D}^*$, as well as $f(\theta(x, y)x, y) > 0$ for all $(x, y) \in \mathcal{D}$, imply that there exists a constant $c_{\inf} > 0$ such that

$$\widetilde{f}_{x,y}(\theta, \zeta, \alpha, v) \geq c_{\inf} \tag{A.2}$$

whenever $(\mathbf{Z}_x \zeta, \alpha y + \mathbf{V}_y v) \in \Psi^*(x)$ and $(x, y) \in \mathcal{D}$.

A.4 Proof of Theorem 3.1

Consider an arbitrary point $(x, y) \in \mathcal{D}$ and recall the notation introduced above. First note that $g_x(0, ay) = \|x\|\theta(x, ay)$ and hence

$$\theta_{\mathrm{C}}(x, y) = \frac{1}{\|x\|} \cdot \min_{a>0} \left\{ \frac{g_x(0, ay)}{a} \mid (\frac{g_x(0, ay)}{\|x\|} x, ay) \in \Psi \right\}. \tag{A.3}$$

Assumption 3.1 together with strict convexity of g_x therefore imply that $a_{min}^{x,y} \in \mathbb{R}_+$ is uniquely defined and $(\theta(x, a_{min}^{x,y}y)x, a_{min}^{x,y}y) \in \mathcal{D}$. Taking derivatives yields

$$\frac{\partial}{\partial a}\frac{g_x(0,ay)}{a}\bigg|_{a=a_{min}^{x,y}} = 0, \quad A_{x,y} := \frac{\partial^2}{\partial a^2}\frac{g_x(0,ay)}{a}\bigg|_{a=a_{min}^{x,y}} = \frac{y^T g''_{x;yy}(0, a_{min}^{x,y}y)y}{a_{min}^{x,y}} > 0. \tag{A.4}$$

Since by assumption g_x is at least three times continuously differentiable, Taylor expansions lead to

$$\left|\frac{g_x(0,ay)}{a} - \frac{g_x(0,a_{min}^{x,y}y)}{a_{min}^{x,y}} - \frac{A_{x,y}}{2}(a - a_{min}^{x,y})^2\right| \leq D|a - a_{min}^{x,y}|^3 \tag{A.5}$$

for some $D > 0$ and all a with $(\theta(x, ay)x, ay) \in \mathcal{D}$. Since $\mathcal{D}^* = \{\theta(\widetilde{x}, \widetilde{y})\widetilde{x}, \widetilde{y})|(\widetilde{x}, \widetilde{y}) \in \mathcal{D}\}$ is compact, D can be chosen independent of $a > 0$ and $(x, y) \in \mathcal{D}$.

Let $\kappa = \frac{2}{p+q+1}$. Since by Assumption 3.1 no relevant point lies in the "observable boundary" for sufficiently large n, it follows from (A.6) and (A.9) in the proof of Theorem 3.1 in Kneip et al. (2015) that for any $a > 0$ with $|a - a_{min}^{x,y}| < \delta$ there exists some $0 < D_1, D_2 < \infty$, which can be chosen independent of (x, y), such that

$$\Pr\left(\left|\|x\|\widehat{\theta}_{\mathrm{VRS}}(x, ay \mid \mathcal{X}_n) - \|x\|\theta(x, ay)\right| \geq D_1 n^{-\kappa}(\log n)^{\kappa}\right) \leq D_2 n^{-2} \tag{A.6}$$

On the other hand, by (A.5) there exists a $0 < d_1 < \infty$ such that

$$\begin{aligned}
\frac{g_x(0, (a_{min}^{x,y} - d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}})y)}{a_{min}^{x,y} - d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}}} - \frac{g_x(0, a_{min}^{x,y}y)}{a_{min}^{x,y}} &\geq 3D_1 n^{-\kappa}(\log n)^{\kappa}, \\
\frac{g_x(0, (a_{min}^{x,y} + d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}})y)}{a_{min}^{x,y} + d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}}} - \frac{g_x(0, a_{min}^{x,y}y)}{a_{min}^{x,y}} &\geq 3D_1 n^{-\kappa}(\log n)^{\kappa}.
\end{aligned} \tag{A.7}$$

Since necessarily $\inf_{(x,y)\in\mathcal{D}} A_{x,y} > 0$, d_1 can be chosen independent of $(x, y) \in \mathcal{D}$. Inequalities (A.5) and (A.6) now imply that with probability converging to 1 we obtain

$$\|x\|\frac{\widehat{\theta}_{\mathrm{VRS}}(x, ay \mid \mathcal{X}_n)}{a} > \|x\|\frac{\widehat{\theta}_{\mathrm{VRS}}(x, a_{min}^{x,y}y \mid \mathcal{X}_n)}{a_{min}^{x,y}} \tag{A.8}$$

for $a = a_{min}^{x,y} - d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}}$ as well as for $a = a_{min}^{x,y} + d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}}$. But convexity then additionally implies that (A.8) also holds for all $a \leq a_{min}^{x,y} - d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}}$ and $a \geq a_{min}^{x,y} + d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}}$. More precisely, there exists a constant $0 < D_3 < \infty$, which can be chosen independent of (x, y), such that

$$1 - \Pr\left(\widehat{\theta}_{\mathrm{C}}(x, y \mid \mathcal{X}_n) = \min_{a_{min}^{x,y} - d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}} \leq a \leq a_{min}^{x,y} + d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}}} \frac{\widehat{\theta}_{\mathrm{VRS}}(x, ay \mid \mathcal{X}_n)}{a}\right) \leq D_3 n^{-2}. \tag{A.9}$$

Recall that $Y_i = \alpha_i y + V_i$. Representation (A.15) of the VRS-DEA estimator in the proof of Theorem 3.1 in Kneip et al. (2015) tells us that

$$\begin{aligned}\frac{\widehat{\theta}_{\mathrm{VRS}}(x,y \mid \mathcal{X}_n)}{a} &= \min\Big\{ \sum_{i=1}^{n} \omega_i \frac{g_x(\theta_i Z_i, \alpha_i y + V_i)}{a\|x\|\theta_i} \mid \boldsymbol{i}_n^T\boldsymbol{\omega} = 1,\ \boldsymbol{Z}\boldsymbol{\omega} = 0, \\ &\qquad\qquad \boldsymbol{V}\boldsymbol{\omega} = 0, \boldsymbol{\alpha}^T\boldsymbol{\omega} = a, \boldsymbol{\omega} \in \mathbb{R}_+^n \Big\} \\ &= \theta_{\mathrm{C}}(x,y) \times \min\Big\{ \sum_{i=1}^{n} \omega_i \frac{a_{min}^{x,y} g_x(\theta_i Z_i, \alpha_i y + V_i)}{a g_x(0, a_{min}^{x,y} y)\theta_i} \mid \boldsymbol{i}_n^T\boldsymbol{\omega} = 1,\ \boldsymbol{Z}\boldsymbol{\omega} = 0, \\ &\qquad\qquad \boldsymbol{V}\boldsymbol{\omega} = 0, \boldsymbol{\alpha}\boldsymbol{\omega} = a, \boldsymbol{\omega} \in \mathbb{R}_+^n \Big\} \end{aligned} \tag{A.10}$$

where $\boldsymbol{i}_n = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$, ω_i represents the ith element of $\boldsymbol{\omega}$, $\theta_i = \theta(X_i, Y_i)$, $Z_i = X_i - \dfrac{\boldsymbol{x}^T X_i}{||\boldsymbol{x}||^2}\boldsymbol{x}$ is a $(p \times 1)$ vector and $\boldsymbol{Z} = (Z_1,\ \ldots,\ Z_n)$ is a $(p \times n)$ matrix, $\boldsymbol{V} = (V_1,\ \ldots,\ V_n)$ is a $((q-1) \times n)$ matrix, and $\boldsymbol{\alpha}^T = (\alpha_1, \ldots, \alpha_n)$.

An essential step of the proof now consists in the localization argument developed in Kneip et al. (2008) and reconsidered in Kneip et al. (2015) which states that VRS-DEA estimators are is asymptotically determined by local information. In Kneip et al. (2008, 2015) the argument relies on using the coordinates (θ_i, Z_i, Y_i), but a generalization to the coordinates $(\theta_i, Z_i, \alpha_i, V_i)$ is immediate. For any $h > 0$, define the set

$$\begin{aligned} C(x, a_{min}^{x,y} y; h) = &\Big\{ (\widetilde{\theta}, \widetilde{z}, \widetilde{\alpha}, \widetilde{v}) \in (0,1] \times \mathbb{R}^{p-1} \times \mathbb{R}_+ \times \mathbb{R}^{q-1} \mid 1 - \widetilde{\theta} \leq h^2, |\widetilde{\alpha} - a_{min}^{x,y}| \leq h, \\ &\widetilde{z} = \sum_j \zeta_j z_x^{(j)},\ |\zeta_j| \leq h\ \forall\ j = 1,\ \ldots,\ p-1,\ \widetilde{v} = \sum_r v_r v_y^{(r)},\ |v_r| \leq h\ \forall\ r = 1,\ \ldots,\ q \Big\}, \end{aligned} \tag{A.11}$$

and let $\mathcal{X}_n(x, a_{min}^{x,y} y; h) := \{(X_i, Y_i) \in \mathcal{X}_n \mid (\theta_i, Z_i, \alpha_i, V_i) \in C(x, a_{min}^{x,y} y; h)\}$.

In the following it will be necessary to distinguish between points (x, y) lying in the interior and on the observable boundary of $\mathcal{D}$. For $(x, y) \in \mathcal{D}$ let

$$\begin{aligned} \Psi^{*\partial}(x,y) = \Big\{ &(\widetilde{z}, \widetilde{v}) \in \mathcal{V}(x,y) \mid (g_x(\widetilde{z}, a_{min}^{x,y} y + \widetilde{v}) \frac{x}{||x||} + \widetilde{z}, a_{min}^{x,y} y + \widetilde{v}) \in \mathcal{D}^* \ \text{ while for any} \\ &\epsilon > 0 \text{ there is some } (z, v) \in \mathbb{R}^{p-1} \times \mathbb{R}^{q-1} \text{ with } \|\widetilde{z} - z\| < \epsilon \text{ and } \|\widetilde{v} - v\| < \epsilon \\ &\text{such that } \ (g_x(z, a_{min}^{x,y} y + v) \frac{x}{||x||} + z, a_{min}^{x,y} y + \ v) \notin \mathcal{D}^* \Big\} \end{aligned} \tag{A.12}$$

denote the boundary of possible vectors (z, v), where of course $\Psi^{*\partial}(x, y) = \emptyset$ if $\min p, q = 1$.

Then define the observable boundary as

$$\mathcal{W}(h) := \Bigg\{ (x,y) \in \mathcal{D} \,|\, \min\Big\{ \min_{j=1,\,...,\,p-1} |\zeta_j|, \min_{r=1,\,...,\,q-1} |v_r| \Big\} \le h \\ \text{for some } (\widetilde{z},\widetilde{v}) \in \Psi^{*\partial}(x,y) \text{ with } \widetilde{z} = \sum_j \zeta_j z_x^{(j)}, \widetilde{v} = \sum_r v_r v_y^{(r)} \Bigg\}. \tag{A.13}$$

If $p \le 1$ and $q \le 1$, then $\mathcal{W}(h) = \emptyset$; but for $p+q > 2$, compactness of $\mathcal{D}^*$ implies that for any $h > 0$ the observable boundary $\mathcal{W}(h)$ is nonempty.[8]

Recall the constant d_1 in (A.8) and choose some constant $b \ge 4(p+q)(1+d_1)$. Then set $\nu_n := b(\frac{\log n}{n\widetilde{f}_{x,y}(1,0,a_{min}^{x,y},0)})^{\frac{1}{p+q+1}}$ as well as $\nu_n^* := b(\frac{\log n}{c_{\inf} n})^{\frac{1}{p+q+1}}$.

Case(i): We first consider the case where (x,y) is in the interior of $\mathcal{D}$ in the sense that $(x,y) \notin \mathcal{W}(\nu_n^*)$. In this case, by Assumption 3.1 we have $C(x, a_{min}^{x,y} y;\ \nu_n) \subset \mathcal{D}$ for all sufficiently large n.

Following the arguments in Kneip et al. (2008, 2015) one can construct $k = 2(p+q-1)$ hypercubes $B_s \subset \mathbb{R}^{p-1} \times \mathbb{R}^p$, $s = 1, \ldots, k$, of side lengths $\frac{\nu_n}{2(p-1)+2q}$ and centered at values (z_j, y_j) determined in the following way: $z_j = \sum_s \zeta_s z_x^{(s)}$, $y_j = (\alpha + a_{min}^{x,y}) y + \sum_r v_r v_y^{(r)}$, where for each $j = 1,\ \ldots,\ 2(p+q-1)$ exactly one of the coordinates $(\zeta_1,\ \ldots,\ \zeta_{p-1}, \alpha, v_1, \ldots, v_{q-1})$ equals $\nu_n \cdot \frac{2(p-1)+2q-1}{2(p-1)+2q}$ or $-\nu_n \cdot \frac{2(p-1)+2q-1}{2(p-1)+2q}$, while all others are identically zero. By definition of ν_n, the probability that there exist at least k observations $(\theta_{i_1}, Z_{i_1}, Y_{i_1}), \ldots, (\theta_{i_k}, Z_{i_k}, Y_{i_k})$ with $\theta_{i_j} \ge 1 - \nu_n^2$ and $(Z_{i_j}, Y_{i_j}) \in B_j$, $j = 1, \ldots, k$, is of order $1 - n^{-2}$ as $n \to \infty$.

On the other hand, if such a set of k observations exists, then by construction for any $a \in [a_{min}^{x,y} - d_1 n^{-\frac{\kappa}{2}} (\log n)^{\frac{\kappa}{2}}, a_{min}^{x,y} + d_1 n^{-\frac{\kappa}{2}} (\log n)^{\frac{\kappa}{2}}]$ the point $(0, ay)$ is in the interior of the convex hull of (Z_{i_j}, Y_{i_j}), $j = 1,\ \ldots,\ k$. If n is sufficiently large, by the strict convexity of g_x the arguments in the proof of Theorem 1 of Kneip et al. (2008) can then be used to show then for any other observation (θ_i, Z_i, Y_i) with $(\theta_i, Z_i, Y_i)) \notin C(x, a_{min}^{x,y} y; \nu_n)$ and any vector $\boldsymbol{\omega} \in \mathbb{R}_+^n$ with $\omega_i > 0$, satisfying the constraints in (A.10) for $a \in [a_{min}^{x,y} - d_1 n^{-\frac{\kappa}{2}} (\log n)^{\frac{\kappa}{2}}, a_{min}^{x,y} + d_1 n^{-\frac{\kappa}{2}} (\log n)^{\frac{\kappa}{2}}]$, there exists another vector $\boldsymbol{\omega}^* \in \mathbb{R}_+^n$ with $\omega_i^* = 0$ and $\omega_{i_j}^* \ge 0$, $j = 1,\ \ldots,\ k$, such that $\sum_{i=1}^n \omega_i \frac{g_x(\theta_i Z_i, Y_i)}{\theta_i} > \sum_{i=1}^n \omega_i^* \frac{g_x(\theta_i Z_i, Y_i)}{\theta_i}$. This

[8] Note that there is an error in Appendix A of Kneip et al. (2015). The concept of the boundary $\Psi^{*\partial}(x)$ used here is correct (as well as the arguments relying on $\Psi^{*\partial}(x)$). But the definition in formula (A.4) of Kneip et al. (2015) does not provide the proper boundary, and it should be replaced by an analog of A.12. The proof of Theorem 3.1 in Kneip et al. (2015) still holds after this change.

implies that for arbitrary $a \in [a_{min}^{x,y} - d_1 n^{-\frac{\kappa}{2}} (\log n)^{\frac{\kappa}{2}}, a_{min}^{x,y} + d_1 n^{-\frac{\kappa}{2}} (\log n)^{\frac{\kappa}{2}}]$ the minimum in (A.10) is achieved by assigning zero weight $\omega_i = 0$ to each observation with $(\theta_i, Z_i, Y_i) \notin C(x, a_{min}^{x,y} y; \nu_n)$. This then leads to $\widehat{\theta}_{\mathrm{VRS}}(x, ay \mid \mathcal{X}_n) = \widehat{\theta}_{\mathrm{VRS}}(x, ay \mid \mathcal{X}_n(x, a_{min}^{x,y} y; \nu_n))$, where $\widehat{\theta}_{\mathrm{VRS}}(x, ay \mid \mathcal{X}_n(x, a_{min}^{x,y} y; \nu_n))$ denotes the VRS-DEA estimator only based on the subset of all observations in $\mathcal{X}_n(x, a_{min}^{x,y} y; \nu_n)$.

Therefore, there exists a $D_4 \in (0, \infty)$, which can be chosen independent of $(x, y) \in \mathcal{D}$ with $(x, y) \notin \mathcal{W}(\nu_n^*)$, such that for all sufficiently large n,

$$\Pr\left(\widehat{\theta}_{\mathrm{C}}(x, y \mid \mathcal{X}_n) = \min_{a_{min}^{x,y} - d_1 (\frac{\log n}{n})^{\frac{\kappa}{2}} \le a \le a_{min}^{x,y} + d_1 (\frac{\log n}{n})^{\frac{\kappa}{2}}} \frac{\widehat{\theta}_{\mathrm{VRS}}(x, ay \mid \mathcal{X}_n(x, a_{min}^{x,y} y; \nu_n))}{a} \right) \ge 1 - D_4 n^{-2} \tag{A.14}$$

Now consider the sums in (A.10) with respect to the (random) number $K_n \le \#\mathcal{X}_n(x, a_{min}^{x,y} y; \nu_n)$ of all observations with coordinates $(\theta_{i_j}, Z_{i_j}, \alpha_{i_j}, V_{i_j}) \in C(x, a_{min}^{x,y} y; \nu_n)$. Furthermore, for some $a \in [a_{min}^{x,y} - d_1 n^{-\frac{\kappa}{2}} (\log n)^{\frac{\kappa}{2}}, a_{min}^{x,y} + d_1 n^{-\frac{\kappa}{2}} (\log n)^{\frac{\kappa}{2}}]$ consider arbitrary weight vectors $\boldsymbol{\omega} = (\omega_1, \ \ldots, \ \omega_{K_n})^T \in \mathbb{R}_+^{K_n}$ such that $\sum_{j=1}^{K_n} \omega_j = 1$, $\sum_{j=1}^{K_n} \omega_j Z_{i_j} = 0$, $\sum_{j=1}^{K_n} \omega_j \alpha_{i_j} = a$, and $\sum_{j=1}^{K_n} \omega_j V_{i_j} = 0$. Let $\theta_{i_j}^* := 1 - \theta_{i_j}$, $G(ay) := g_x''(0, ay)$, and note that $\sum_{j=1}^{K_n} \omega_j (\alpha_{i_j} - a_{min}^{x,y})^2 = \sum_{j=1}^{K_n} \omega_j (\alpha_{i_j} - a)^2 + (a - a_{min}^{x,y})^2$. It then follows from Taylor expansions of g_x as well as from (A.5) that for some $0 \le R_n, R_n^* < \infty$

$$\begin{aligned}
&\sum_{j=1}^{K_n} \omega_j \frac{g_x(\theta_{i_j} Z_{i_j}, \alpha_{i_j} y + V_{i_j})}{a \theta_{i_j}} = \frac{g_x(0, ay)}{a} \\
&\quad + \frac{1}{a} \sum_{j=1}^{K_n} \omega_j \left[\begin{pmatrix} Z_{i_j} \\ V_{i_j} \end{pmatrix}^T \frac{G(ay)}{2} \begin{pmatrix} Z_{i_j} \\ V_{i_j} \end{pmatrix} + \begin{pmatrix} 0 \\ (a_{i_j} - a) y \end{pmatrix}^T G(ay) \begin{pmatrix} Z_{i_j} \\ V_{i_j} \end{pmatrix} + (\alpha_{i_j} - a)^2 \frac{y^T g_{x;yy}''(0, ay) y}{2} + \theta_{i_j}^* \right] \\
&\quad + R_n \nu_n^3 \\
&= \frac{g_x(0, a_{min}^{x,y} y)}{a_{min}^{x,y}} \\
&\quad + \frac{1}{a_{min}^{x,y}} \underbrace{\sum_{j=1}^{K_n} \omega_j \left[\begin{pmatrix} Z_{i_j} \\ (\alpha_{i_j} - a_{min}^{x,y}) y + V_{i_j} \end{pmatrix}^T \frac{G(a_{min}^{x,y})}{2} \begin{pmatrix} Z_{i_j} \\ (\alpha_{i_j} - a_{min}^{x,y}) y + V_{i_j} \end{pmatrix} + \theta_{i_j}^* \right]}_{=: \tau((\theta_{i_1}^*, Z_{i_1}, \alpha_{i_1}, V_{i_1}), \ldots, (\theta_{i_{K_n}}^*, Z_{i_{K_n}}, \alpha_{i_{K_n}}, V_{i_{K_n}}); \boldsymbol{\omega})} + R_n^* \nu_n^3
\end{aligned} \tag{A.15}$$

By our assumptions there exists a constant $D_5 < \infty$ such that $R_n^* < D_5$ for all possible K_n, all possible sets $\{(\theta_{i_j}, Z_{i_j}, \alpha_{i_j}, V_{i_j})\} \subset C(x, a_{min}^{x,y} y; \nu_n)$, all a and all $(x, y) \in \mathcal{D}$ with $(x, y) \notin \mathcal{W}(\nu_n^*)$.

The result in (A.15) shows that $\widehat{\theta}_{\mathrm{C}}(x,y \mid \mathcal{X}_n)$ is essentially determined by minimizing $\tau(\cdot)$ over all possible $\boldsymbol{\omega}$ with $\sum_{j=1}^{K_n}\omega_i Z_{i_j}=0$ and $\sum_{j=1}^{K_n}\omega_j V_{i_j}=0$, independent of the corresponding value of $\sum_{j=1}^{K_n}\omega_j\alpha_{i_j}=a$ (even cases with $a\notin[a_{min}^{x,y}-d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}}, a_{min}^{x,y}+d_1 n^{-\frac{\kappa}{2}}(\log n)^{\frac{\kappa}{2}}]$ need not to be excluded since due to (A.5) they cannot constitute an optimal solution with probability tending to 1). Recall that $\theta_{i_j}^*:=1-\theta_{i_j}$, and let

$$
\begin{aligned}
&T_{K_n}\left((\theta_{i_1}^*,Z_{i_1},\alpha_{i_1},V_{i_1}),\ldots,(\theta_{i_{K_n}}^*,Z_{i_{K_n}},\alpha_{i_{K_n}},V_{i_{K_n}})\right)\\
&\quad=\min\left\{\tau((\theta_{i_1}^*,Z_{i_1},\alpha_{i_1},V_{i_1}),\ldots,(\theta_{i_{K_n}}^*,Z_{i_{K_n}},\alpha_{i_{K_n}},V_{i_{K_n}});\boldsymbol{\omega})\mid \boldsymbol{i}_{K_n}^T\boldsymbol{\omega}=1,\ \sum_{j=1}^{K_n}\omega_j Z_{i_j}=\sum_{j=1}^{K_n}\omega_j V_{i_j}=0\right\}
\end{aligned}
\tag{A.16}
$$

When combining these arguments with (A.10) and (A.14) one can conclude that there are constants $0<D_6,D_7<\infty$ such that with probability at least $1-D_6 n^{-2}$

$$
\left|\widehat{\theta}_{\mathrm{C}}(x,y\mid\mathcal{X}_n)-\theta_{\mathrm{C}}(x,y)\left(1+\frac{T_{K_n}\left((\theta_{i_1}^*,Z_{i_1},\alpha_{i_1},V_{i_1}),\ldots,(\theta_{i_{K_n}}^*,Z_{i_{K_n}},\alpha_{i_{K_n}},V_{i_{K_n}})\right)}{g_x(0,a_{min}^{x,y}y)}\right)\right|\leq D_7\nu_n^3
\tag{A.17}
$$

Here, D_6 and D_7 can be chosen independent of $(x,y)\in\mathcal{D}$ with $(x,y)\notin\mathcal{W}(\nu_n^*)$. Since necessarily, $\tau((\theta_{i_1}*,Z_{i_1},\alpha_{i_1},V_{i_1}),\ldots,(\theta_{i_{K_n}}^*,Z_{i_{K_n}},\alpha_{i_{K_n}},V_{i_{K_n}});\boldsymbol{\omega})\leq D_8\nu_n^2$, (A.17) immediately implies that for some constant $D_8<\infty$ and all $\beta>0$

$$
E\left(\left|\widehat{\theta}_{\mathrm{C}}(x,y\mid\mathcal{X}_n)-\theta(x,y)\right|^{\beta}\right)\leq D_8\max\{n^{-\frac{2\beta}{p+q+1}}(\log n)^{\frac{2\beta}{p+q+1}},n^{-2}\}\ \forall\ (x,y)\in\mathcal{D}\backslash\mathcal{W}(\nu_n^*).
\tag{A.18}
$$

More precise results are to be obtained from the distribution of T_{K_n}. When translating the results of Kneip et al. (2008, (2015)) into the alternative (θ,ζ,α,v)-coordinate system it turns out that the asymptotic behavior the VRS-DEA estimator $\widehat{\theta}(x,a_{min}^{x,y}y\mid\mathcal{X}_n)$ of $\theta(x,a_{min}^{x,y}y)$ is determined by a similar random variable $T_{K_n}^{DEA}\left((\theta_{i_1}^*,Z_{i_1},\alpha_{i_1},V_{i_1}),\ldots,(\theta_{i_{K_n}}^*,Z_{i_{K_n}},\alpha_{i_{K_n}},V_{i_{K_n}})\right)$ defined by minimizing $\tau((\theta_{i_1}^*,Z_{i_1},\alpha_{i_1},V_{i_1}),\ldots,(\theta_{i_{K_n}}^*,Z_{i_{K_n}},\alpha_{i_{K_n}},V_{i_{K_n}});\boldsymbol{\omega})$ with respect to all weight sequences with $\boldsymbol{i}_{K_n}^T\boldsymbol{\omega}=1$, $\sum_{j=1}^{K_n}\omega_j Z_{i_j}=\sum_{j=1}^{K_n}\omega_j V_{i_j}=0$, **and** $\sum_{j=1}^{K_n}\omega_j\alpha_{i_j}=a_{min}^{x,y}$. Therefore, the only difference between T_{K_n} and $T_{K_n}^{DEA}$ consists in the fact that (A.16) does not incorporate the additional constraint $\sum_{j=1}^{K_n}\omega_j\alpha_{i_j}=a_{min}^{x,y}$. But all arguments developed for analyzing $T_{K_n}^{DEA}$ readily generalize to T_{K_n}.

Obviously, the observations $(\theta^*_{i_j}, \zeta_{i_j}, \alpha_{i_j}, v_{i_j})$ are independent. The conditional distribution of $(\theta^*_{i_j}, \zeta_{i_j}, \alpha_{i_j}, v_{i_j})$ given $(X_{i_j}, Y_{i_j}) \in \mathcal{X}_n(x, a^{x,y}_{min} y; \nu_n)$ converges to a uniform distribution. Also note that for all (x, y) in the interior of $\mathcal{D}$ we necessarily have $(x, y) \notin \mathcal{W}(\nu^*_n)$ for all sufficiently large n. For deriving the asymptotic distribution of T_{K_n} we rely on the construction presented in Kneip et al. (2008). Let $(\widetilde{\theta}_1,\ \widetilde{\zeta}_1,\ \widetilde{\alpha}_1,\ \widetilde{v}_1),\ \ldots,\ (\widetilde{\theta}_k,\ \widetilde{\zeta}_k,\ \widetilde{\alpha}_k,\ \widetilde{v}_k)$ denote iid random variables uniformly distributed on $[0,1] \times [-1,1]^{p-1} \times [a^{x,y}_{min} - 1, a^{x,y}_{min} + 1] \times [-1,1]^{q-1}$, and set $\widetilde{Z}_i = \sum_j \widetilde{\zeta}_{ij} z^{(j)}_x$, $\widetilde{V}_i = \sum_r \widetilde{v}_{ir} v^{(r)}_y$, $i = 1, \ldots, k$. Then for any integer k and $\gamma > 0$ define the following event $\mathcal{U}[\gamma, k]$: there exists a weight vector $\boldsymbol{\omega} \in \mathbb{R}^k_+$ with $\boldsymbol{i}^T_k \boldsymbol{\omega} = 1$ and $\sum_{j=1}^k \omega_j \widetilde{Z}_j = \sum_{j=1}^k \omega_j \widetilde{V}_j = 0$ such that

$$\frac{\tau((\widetilde{\theta}_1,\ \widetilde{Z}_1,\ \widetilde{\alpha}_1,\ \widetilde{V}_1),\ \ldots,\ (\widetilde{\theta}_k,\ \widetilde{Z}_k,\ \widetilde{\alpha}_k,\ \widetilde{V}_k); \boldsymbol{\omega})}{g_x(0, a^{x,y}_{min} y)} \leq \gamma. \tag{A.19}$$

Applying the same type of arguments as those used in the proof of Theorem 2 of Kneip et al. (2008) it can then be derived that for any $\gamma > 0$

$$\begin{aligned} &\lim_{n\to\infty} \Pr\left(n^\kappa \Big(\frac{\widehat{\theta}_{\mathrm{C}}(x, y \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(x,y)}{\theta_{\mathrm{C}}(x,y)} \Big) \leq \gamma \right) \\ &= \lim_{n\to\infty} \Pr\left(n^\kappa \frac{T_{K_n}((\theta^*_{i_1}, Z_{i_1}, \alpha_{i_1}, V_{i_1}), \ldots, (\theta^*_{i_{K_n}}, Z_{i_{K_n}}, \alpha_{i_{K_n}}, V_{i_{K_n}}))}{g_x(0, a^{x,y}_{min} y)} \leq \gamma \right) = F_{x,y}(\gamma) \end{aligned} \tag{A.20}$$

where $F_{x,y}$ is a continuous distribution function with $F_{x,y}(0) = 0$ and

$$F_{x,y}(\gamma) = \lim_{k\to\infty} \Pr\left(\mathcal{U}\left[\gamma \frac{\widetilde{f}_{x,y}(1, 0, a^{x,y}_{min}, 0)^{\frac{2}{p+q+1}}}{k^{\frac{2}{p+q+1}}}, k \right] \right) \tag{A.21}$$

This proves (3.12). Analysis of expectations now relies on the techniques developed in Kneip et al. (2015).

Let $\widetilde{\nu}_n := \left(\frac{n}{\widetilde{f}_{x,y}(1,0,a^{x,y}_{min},0)} \right)^{\frac{1}{p+q+1}}$, $\widetilde{Z}^{(n)}_j = \mathbf{Z}_x \widetilde{\zeta}^{(n)}_j$, $\widetilde{V}^{(n)}_j = \mathbf{V}_y \widetilde{v}^{(n)}_j$ and let $(\widetilde{\theta}^{(n)}_j, \widetilde{\zeta}^{(n)}_j, \widetilde{\alpha}^{(n)}_j, \widetilde{v}^{(n)}_j)$, $j = 1, \ldots, n$, denote iid random variables with a uniform distribution on $[0, \widetilde{\nu}^2_n] \times [-\widetilde{\nu}_n, \widetilde{\nu}_n]^{p-1} \times [a^{x,y}_{min} - \widetilde{\nu}_n, a^{x,y}_{min} + \widetilde{\nu}_n] \times [-\widetilde{\nu}_n, \widetilde{\nu}_n]^{p-1}$. Similar to T_{K_n} one can then define the r.v. $T_n\left((\widetilde{\theta}^{(n)}_1, \widetilde{Z}^{(n)}_1, \widetilde{\alpha}^{(n)}_1, \widetilde{V}^{(n)}_1),\ \ldots,\ (\widetilde{\theta}^{(n)}_n, \widetilde{Z}^{(n)}_n, \widetilde{\alpha}^{(n)}_n, \widetilde{V}^{(n)}_n) \right)$ by minimizing (A.16) with respect to the set of observations $\{(\widetilde{\theta}^{(n)}_j, \widetilde{\zeta}^{(n)}_j, \widetilde{\alpha}^{(n)}_j, \widetilde{v}^{(n)}_j)\}$ instead of $\{(\theta^*_{i_j}, Z_{i_j}, \alpha_{i_j}, V_{i_j})\}$. In a straightforward generalization of the arguments leading to relations (A.13)–(A.18) in the proof of Theorem 3.1 of Kneip et al. (2015) it can then be shown that the asymptotic distributions of $n^\kappa T_{K_n}\left((\theta^*_{i_1}, Z_{i_1}, \alpha_{i_1}, V_{i_1}), \ldots, (\theta^*_{i_{K_n}}, Z_{i_{K_n}}, \alpha_{i_{K_n}}, V_{i_{K_n}}) \right)$ and of

$T_n\left((\widetilde{\theta}_1^{(n)}, \widetilde{Z}_1^{(n)}, \widetilde{\alpha}_1^{(n)}, \widetilde{V}_1^{(n)}),\ \ldots,\ (\widetilde{\theta}_n^{(n)}, \widetilde{Z}_n^{(n)}, \widetilde{\alpha}_n^{(n)}, \widetilde{V}_n^{(n)})\right)$ coincide, and that all moments of $T_n\left((\widetilde{\theta}_1^{(n)}, \widetilde{Z}_1^{(n)}, \widetilde{\alpha}_1^{(n)}, \widetilde{V}_1^{(n)}),\ \ldots,\ (\widetilde{\theta}_n^{(n)}, \widetilde{Z}_n^{(n)}, \widetilde{\alpha}_n^{(n)}, \widetilde{V}_n^{(n)})\right)$ converge rapidly to finite, fixed values as $n \to \infty$. Additionally using (A.17), we obtain the following generalization of relations (A.16)–(A.18) in the proof of Theorem 3.1 of Kneip et al. (2015):

$$\left| E\left(\widehat{\theta}_{\mathrm{C}}(x,y \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(x,y)\right) - \theta_{\mathrm{C}}(x,y) n^{-\frac{2}{p+q+1}} \frac{\widetilde{C}_{g_x'', \widetilde{f}_{x,y}(1,0,a_{min}^{x,y},0)}}{g_x(0, a_{min}^{x,y} y)} \right| \leq D_9 n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}} \tag{A.22}$$

for all $(x,y) \in \mathcal{D}$ with $(x,y) \notin \mathcal{W}(\nu_n^*)$. and some $D_9 \in (0,\infty)$, where

$$\widetilde{C}_{g_x'', \widetilde{f}_{x,y}(1,0,a_{min}^{x,y},0)} := \lim_{n\to\infty} E\left[T_n\left((\widetilde{\theta}_1, \widetilde{Z}_1, \widetilde{\alpha}_1^{(n)}, \widetilde{V}_1^{(n)}),\ \ldots,\ (\widetilde{\theta}_n, \widetilde{Z}_n, \widetilde{\alpha}_n^{(n)}, \widetilde{V}_n^{(n)})\right)\right] \tag{A.23}$$

only depends upon g_x'' and $\widetilde{f}_{x,y}(1,0,a_{min}^{x,y},0)$ and changes continuously in $(x,y) \in \mathcal{D}$. Furthermore, there exists some $D_{10} \in (0,\infty)$ such that

$$E\left(\left|\widehat{\theta}_{\mathrm{C}}(x,y \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(x,y)\right|^2\right) \leq D_{10} n^{-\frac{4}{p+q+1}} \tag{A.24}$$

for all $(x,y) \in \mathcal{D}$ with $(x,y) \notin \mathcal{W}(\nu_n^*)$.

Case (ii): For a further analysis of expectations we additionally have to consider the alternative case where $(x,y) \in \mathcal{W}(\nu_n^*)$. We again rely on arguments similar to those used in the proof of Theorem 3.1 of Kneip et al. (2015).

In this case, the problem arises that some of the sets B_j used in the above construction surpass the boundary and are no longer in $\mathcal{D}$. As a consequence, one cannot exclude that $\widehat{\theta}_{\mathrm{C}}(x,y \mid \mathcal{X}_n)$ is influenced by an observation with $\theta_i \leq 1 - \nu_n^2$. But let

$$\mathcal{H}_n(x,y;\nu_n) := \{(X_i,Y_i) \in \mathcal{X}_n \mid (1, Z_i, a_{min}^{x,y}, V_i) \in C(x, a_{min}^{x,y} y; , \nu_n)\} \tag{A.25}$$

By a straightforward generalization of the arguments in the proof of Theorem 3.1 of Kneip et al. (2015) it it follows that

$$\left|1 - \Pr\left(\widehat{\theta}_{\mathrm{C}}(x,y \mid \mathcal{X}_n) = \widehat{\theta}_{\mathrm{C}}(x,y \mid \mathcal{H}_n(x,y;\nu_n))\right)\right| \leq D_{11} n^{-2} \tag{A.26}$$

for all $(x,y) \in \mathcal{D}$, some $D_{11} \in (0,\infty)$, and all sufficiently large n.

Recall that boundary problems arise only if $p+q > 2$. In such cases, for $r = 1,\ \ldots,\ p-1$, define

$$v_{r;x,y} := \min_{\substack{(\zeta,v) \in \mathbb{R}^{p-1} \times \mathbb{R}^{q-1}, \\ (\sum_{j=1}^{p-1} \zeta_j z_x^{(j)}, v) \in \Psi^{*\partial}(x,y)}} \{\nu_n,\ |\zeta_r|\}\,. \tag{A.27}$$

Similarly, for $r = 1, \ldots, q-1$, define $v_{p-1+r;x,y}$ by replacing $|\zeta_r|$ with $|v_r|$ in (A.27). These $v_{r;x,y}$ can be viewed as measuring a "distance" from (x, y) to the boundary, with $v_{r;x,y} \leq \nu_n$.

If $\prod_{r=1}^{p+q-2} v_{r;x,y} \geq \nu_n^{p+q+1}$, i.e. (x, y) is not too near the boundary, an upper bound for $\widehat{\theta}_{\mathrm{C}}(x, y \mid \mathcal{X}_n)$ can then be obtained by relying on the observations with $1 - \theta_i \leq \left(\frac{\nu_n^{p+q+1}}{\prod_{r=1}^{p+q-2} v_{r;x,y}}\right)^{2/3}$ and $|\alpha_i - \alpha_{min}^{x,y}| \leq \left(\frac{\nu_n^{p+q+1}}{\prod_{r=1}^{p+q-2} v_{r;x,y}}\right)^{1/3}$. Arguments similar to those used above then show that for all $(x, y) \in \mathcal{W}(\nu_n^*)$ with $\prod_{r=1}^{p+q-2} v_{r;x,y} \geq \nu_n^{p+q+1}$, we have for $\alpha \in \{1, 2\}$

$$E\left(\left|\widehat{\theta}_{\mathrm{C}}(x, y \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(x, y)\right|^{\alpha}\right) \leq D_{12}^{\alpha} \left(\frac{\nu_n^{p+q+1}}{\prod_{r=1}^{p+q-2} v_{r;x,y}}\right)^{2\alpha/3}, \tag{A.28}$$

for some constant $D_{12} \in (0, \infty)$, and for all sufficiently large n.

Now the moments of $\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n)$ can be analyzed in a way similar to Kneip et al. (2015). Let $\mathcal{X}_{n,-i}$ denote the sample of size $n-1$ obtained by eliminating the i-th observation (X_i, Y_i). When relying on $\mathcal{X}_{n,-i}$, it is clear that all constants in the above inequalities can be chosen independently of (x, y) and thus also apply for the (random) coordinate system induced by the specific choice $(x, y) = (X_i, Y_i)$. Obviously,

$$\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n) = \min\left\{\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_{n,-i}),\ 1\right\}. \tag{A.29}$$

Since (X_i, Y_i) is independent of $\mathcal{X}_{n,-i}$, (A.18) and (A.29) imply that

$$E\left(\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(X_i, Y_i) \mid (X_i, Y_i) \notin \mathcal{W}(\nu_n^*)\right) = C_0 n^{-\frac{2}{p+q+1}} + O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right) \tag{A.30}$$

for some $C_0 \in (0, \infty)$. If $p = 1$ and $q \leq 1$, then assertion (3.13) follows directly from (A.30), since in this case there is no boundary problem due to $\mathcal{W}(\nu_n^*) = \emptyset$.

In order to quantify the influence of boundary effects for $p + q > 2$, let $\mathcal{W}_{n,1} := \left\{(x, y) \in \mathcal{D} \mid \nu_n^{p+q-2} > \prod_{r=1}^{p+q-1} v_{r;x,y} \geq \nu_n^{p+q+1}\right\}$ contain points in $\mathcal{W}(\nu_n^*)$ but not too near the boundary, and let $\mathcal{W}_{n,2} := \left\{(x, y) \in \mathcal{D} \mid \prod_{r=1}^{p+q-2} v_{r;x,y} < \nu_n^{p+q+1}\right\}$ contain the other points of $\mathcal{W}(\nu_n^*)$ very near the boundary where only the trivial upper bound $|\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)| \leq 1$ can be used. For points in $\mathcal{W}_{n,1}$, note that for all $r = 1, \ldots, p+q-2$, $\nu_n^4 \leq v_{r;x,y} \leq \nu_n$. Fortunately, the boundary is "smaller" than in the DEA-case, and its

influence is less pronounced. Note that

$$E\left(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)-\theta(X_i,Y_i)\right) = E\left(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)-\theta_{\mathrm{C}}(X_i,Y_i) \mid (X_i,Y_i)\notin \mathcal{W}(\nu_n^*)\right) \times \quad \Pr((X_i,Y_i))\notin \mathcal{W}(\nu_n^*)) + \sum_{s=1}^{2} E\left(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)-\theta(X_i,Y_i) \mid (X_i,Y_i)\in \mathcal{W}_{n,s}\right)\times \Pr((X_i,Y_i)\in \mathcal{W}_{n,s}). \tag{A.31}$$

When relying on (A.28), straightforward calculations similar to those in Kneip et al. (2015) yield that with for some constants $D_{13}, D_{14} < \infty$,

$$\begin{aligned} &E\left(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)-\theta(X_i,Y_i) \mid (X_i,Y_i)\in \mathcal{W}_{n,1}\right)\cdot \Pr((X_i,Y_i)\in \mathcal{W}_{n,1}) \\ &\quad \leq D_{13}\int_{\mathcal{W}_{n,1}} \left(\frac{\nu_n^{p+q+1}}{\prod_{r=1}^{p+q-2} v_{r;x,y}}\right)^{2/3} f(x,y)dxdy \\ &\quad \leq D_{14}\sum_{r=1}^{p+q-2}\int_{\mathcal{B}} \frac{\nu_n^{8/3}}{v_{r;x,y}^{2/3}}\,dx\,dy + O_p\left(n^{-\frac{4}{p+q+1}}(\log n)^{\frac{4}{p+q+1}}\right) \\ &\quad = O_p\left(n^{-\frac{4}{p+q+1}}(\log n)^{\frac{4}{p+q+1}}\right), \end{aligned} \tag{A.32}$$

where $\mathcal{B} := \{(x,y)\in\mathcal{D} \mid$
$\nu_n > v_{r,x,y} \geq \nu_n^4\}$ In addition, $\Pr((X_i,Y_i)\in\mathcal{W}_{n,2}) = O\left(n^{-\frac{4}{p+q+1}}(\log n)^{\frac{4}{p+q+1}}\right)$. Together with (A.30), this leads to (3.13).[9]

Recall (A.22) and (A.24). Assertion (3.14) follows from the fact that (A.28) implies the existence of constants $D_{15}, D_{16} < \infty$ such that

$$\begin{aligned} \mathrm{VAR}\big(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)-\theta_{\mathrm{C}}(X_i,Y_i)\big) &\leq D_{15}n^{-\frac{4}{p+q+1}} \times \Pr\left((X_i,Y_i)\notin \mathcal{W}(\nu_n^*)\right) \\ &+ D_{16}^2\int_{\mathcal{W}_{n,1}}\left(\frac{\nu_n^{p+q+1}}{\prod_{r=1}^{p+q-2} v_{r;x,y}}\right)^{4/3} f(x,y)\,dx\,dy + \Pr\left((X_i,Y_i)\in\mathcal{W}_{n,2}\right) \\ &= O\left(n^{-\frac{4}{p+q+1}} + n^{-\frac{4}{p+q+1}}(\log n)^{\frac{4}{p+q+1}}\right). \end{aligned} \tag{A.33}$$

It remains to prove (3.15). Estimators $\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)$ exhibit stronger correlations than the original VRS-DEA estimators $\widehat{\theta}_{\mathrm{VRS}}(X_i,Y_i \mid \mathcal{X}_n)$. The reason is that by (A.14), for any

[9] Note that a typographical error appears in Appendix A of Kneip et al. (2015). The quantity $E\left(\widehat{\theta}_{\mathrm{VRS}}(X_i,Y_i \mid \mathcal{X}_n)-\theta(X_i,Y_i) \mid (X_i,Y_i)\in\mathcal{W}_{n,1}\right)$ in formula (A.24) should be replaced by $E\left(\widehat{\theta}_{\mathrm{VRS}}(X_i,Y_i \mid \mathcal{X}_n)-\theta(X_i,Y_i) \mid (X_i,Y_i)\in\mathcal{W}_{n,1}\right)\cdot\Pr((X_i,Y_i)\in\mathcal{W}_{n,1})$.

$b > 0$ the estimators $\widehat{\theta}_{\mathrm{C}}(x, y \mid \mathcal{X}_n)$ and $\widehat{\theta}_{\mathrm{C}}(x, by \mid \mathcal{X}_n)$ depend on the same local observations in $\mathcal{X}_n(x, a_{min}^{x,y} y; \nu_n)$, while for sufficiently large b, DEA estimators $\widehat{\theta}_{\mathrm{DEA}}(x, y \mid \mathcal{X}_n)$ and $\widehat{\theta}_{\mathrm{DEA}}(x, by \mid \mathcal{X}_n)$ will be asymptotically uncorrelated.

However, for all $i, j \in 1, \ldots, n$, $i \neq j$, it follows from (A.26) that $\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)$ and $\widehat{\theta}_{\mathrm{C}}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j)$ are asymptotically uncorrelated if $\mathcal{H}_n(X_i, Y_i; \nu_n) \cap \mathcal{H}_n(X_j, Y_j; \nu_n) = \emptyset$. Since all observations are iid, the Cauchy-Schwarz inequality yields

$$\begin{aligned} &\left|\mathrm{COV}\left(\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(X_i, Y_i),\ \widehat{\theta}_{\mathrm{C}}(X_j, Y_j \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(X_j, Y_j)\right)\right| \\ &\quad \leq \Pr\left(\mathcal{H}_n(X_i, Y_i; \nu_n) \cap \mathcal{H}_n(X_j, Y_j; \nu_n) \neq \emptyset\right) \\ &\qquad \times \mathrm{VAR}\left(\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(X_i, Y_i)\right) + O\left(n^{-2}\right). \end{aligned} \tag{A.34}$$

Relation (3.14) as well as

$$\Pr\left(\mathcal{H}_n(X_i, Y_i; \nu_n) \cap \mathcal{H}_n(X_j, Y_j; \nu_n) \neq \emptyset\right) = O\left(n^{-\frac{p+q-2}{p+q+1}} (\log n)^{\frac{p+q-2}{p+q+1}}\right) \tag{A.35}$$

now lead to assertion (3.15), completing the proof of the theorem. ■

A.5 Proof of Theorem 3.2

The transformation defined by the respective function Γ is monotonic and differentiable with nonzero derivatives on $\mathbb{R}_+$. Therefore, (3.16) follows via the delta method.

By Assumption 3.1 (iii) $\Gamma(\theta_{\mathrm{C}}(X_i, Y_i))$ as well as its derivatives $\Gamma'(\theta_{\mathrm{C}}(X_i, Y_i))$ and $\Gamma''(\theta_{\mathrm{C}}(X_i, Y_i))$ are uniformly bounded for all $(X_i, Y_i) \in \mathcal{D}$. It thus follows from a Taylor expansion and (3.14) that

$$\begin{aligned} E\left(\Gamma(\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_{\mathrm{C}}(X_i, Y_i))\right) &= E\left(\Gamma'(\theta_{\mathrm{C}}(X_i, Y_i))[\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(X_i, Y_i)]\right) \\ &\quad + O\left(n^{-\frac{4}{p+q+1}} (\log n)^{\frac{4}{p+q+1}}\right). \end{aligned} \tag{A.36}$$

Recall that (A.29) states that $\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n) = \min\left\{\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_{n,-i}), 1\right\}$. Moreover, the arguments developed in the proof of Theorem 3.1 imply that

$$\Pr\left(\{\widehat{\theta}_{\mathrm{C}}(X_i, Y_i \mid \mathcal{X}_n) = 1\} \cap \{(X_i, Y_i) \notin \mathcal{W}(\nu_n^*)\}\right) = O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right), \tag{A.37}$$

where the boundary $\mathcal{W}(\nu_n^*)$ is defined as in the proof of Theorem 3.1. Since $\Gamma'(\theta_{\mathrm{C}}(X_i, Y_i) > 0$,

it follows from (A.36), (A.22), and (A.23) that similar to (A.30) we have

$$E\left(\Gamma(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_{\mathrm{C}}(X_i,Y_i)) \mid (X_i,Y_i) \notin \mathcal{W}(\nu_n^*)\right) = C_0^{\Gamma} n^{-\frac{2}{p+q+1}} + O\left(n^{-\frac{3}{p+q+1}} (\log n)^{\frac{3}{p+q+1}}\right) \tag{A.38}$$

for some $0 < C_0^{\Gamma} < \infty$. An immediate generalization of (A.31) yields

$$\begin{aligned} &E\left(\Gamma(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_{\mathrm{C}}(X_i,Y_i))\right) = \\ &\quad E\left(\Gamma(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_{\mathrm{C}}(X_i,Y_i)) \mid (X_i,Y_i) \notin \mathcal{W}(\nu_n^*)\right) \cdot \Pr((X_i,Y_i) \notin \mathcal{W}(\nu_n^*)) \\ &\quad + \sum_{s=1}^{2} E\left(\Gamma(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_{\mathrm{C}}(X_i,Y_i)) \mid (X_i,Y_i) \in \mathcal{W}_{n,s}\right) \cdot \Pr((X_i,Y_i) \in \mathcal{W}_{n,s}). \end{aligned} \tag{A.39}$$

With $0 < M_1 := \sup_{(x,y)\in\mathcal{D}} \Gamma'(\theta_{\mathrm{C}}(x,y)) < \infty$ a Taylor expansion leads to

$$\begin{aligned} &E\left(\Gamma(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_{\mathrm{C}}(X_i,Y_i)) \mid (X_i,Y_i) \in \mathcal{W}_{n,1}\right) \cdot \Pr((X_i,Y_i) \in \mathcal{W}_{n,1}) \\ &\qquad \leq M_1 E\left(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(X_i,Y_i) \mid (X_i,Y_i) \in \mathcal{W}_{n,1}\right) \cdot \Pr((X_i,Y_i) \in \mathcal{W}_{n,1}), \end{aligned} \tag{A.40}$$

and Assertion (3.17) then is an immediate consequence of (A.38), (A.32), and $\Pr((X_i,Y_i) \in \mathcal{W}_{n,2}) = O\left(n^{-\frac{4}{p+q+1}} (\log n)^{\frac{4}{p+q+1}}\right)$. Similarly, (A.33) implies

$$\begin{aligned} E\left([\Gamma(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_{\mathrm{C}}(X_i,Y_i))]^2\right) &\leq M_1^2 E\left([\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n) - \theta_{\mathrm{C}}(X_i,Y_i)]^2\right) \\ &= O\left(n^{-\frac{4}{p+q+1}} (\log n)^{\frac{4}{p+q+1}}\right), \end{aligned} \tag{A.41}$$

which proves Assertion (3.18). Analogous to (A.34) and (A.35) Assertion (3.19) finally follows from the fact that $\Gamma(\widehat{\theta}_{\mathrm{C}}(X_i,Y_i \mid \mathcal{X}_n)) - \Gamma(\theta_{\mathrm{C}}(X_i,Y_i))$ and $\Gamma(\widehat{\theta}_{\mathrm{C}}(X_j,Y_j \mid \mathcal{X}_n)) - \Gamma(\theta_{\mathrm{C}}(X_j,Y_j))$ are asymptotically uncorrelated if $\mathcal{H}_n(X_i,Y_i;\nu_n) \cap \mathcal{H}_n(X_j,Y_j;\nu_n) = \emptyset$. ■

A.6 Proof of Theorem 3.3

Note that Theorem 3.2 holds for both (x^1,y^1) and (x^2,y^2) due to Assumption 3.2. The log transformation in Theorem 3.2 is monotonic, differentiable, and invertible. Hence the result follows via the delta method. ■

A.7 Proof of Theorem 3.4

For $t = s$ Assertion (3.27) follows from (3.17). Now consider the case $t \neq s$. Following the notation introduced in (3.4) let

$$(\breve{X}_i^t, \breve{Y}_i^t) := (\tilde{g}_x(\alpha_{min}^{X_i^t, Y_i^t} \frac{Y_i^t}{\|Y_i^t\|}) \frac{X_i^t}{\|X_i^t\|}, \alpha_{min}^{X_i^t, Y_i^t} \frac{Y_i^t}{\|Y_i^t\|})$$

Since $\mathcal{D}_{norm}^1 = \mathcal{D}_{norm}^2$ we have $(\breve{X}_i^t, \breve{Y}_i^t) \in \mathcal{D}^s$. Then (3.10) implies that

$$\begin{aligned} \log \widehat{\gamma}^s(X_i^t, Y_i^t \mid \mathcal{X}_{n_s}^s) - \log \gamma^s(X_i^t, Y_i^t) &= \log \widehat{\gamma}^s(\breve{X}_i^t, \breve{Y}_i^t \mid \mathcal{X}_{n_s}^s) - \log \gamma^s(\breve{X}_i^t, \breve{Y}_i^t) \\ &= \Gamma(\widehat{\theta}_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t \mid \mathcal{X}_{n_s}^s)) - \Gamma(\theta_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t)), \end{aligned} \tag{A.42}$$

where $\Gamma(\theta) = \log \theta^{1/2}$ for all $\theta > 0$. Recall the arguments developed in the proofs of Theorems 3.1 and 3.2 and the definitions of the boundaries $\mathcal{W}(\nu_{n_s}^*) \equiv \mathcal{W}^s(\nu_{n_s}^*)$, $\mathcal{W}_{n_s,1} \equiv \mathcal{W}_{n_s,1}^s$ as well as $\mathcal{W}_{n_2,2} \equiv \mathcal{W}_{n_s,2}^s$. If $(\widehat{\theta}_{\mathrm{C}}^s(\breve{X}_i^s, \breve{Y}_i^s \mid \mathcal{X}_{n_s}^s) \neq 1$, then obviously $\widehat{\theta}_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t \mid \mathcal{X}_{n_s}^s) = \widehat{\theta}_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t \mid \mathcal{X}_{n_s,-i}^s)$, where again $\mathcal{X}_{n_s,-i}$ denote the sample of size $n-1$ obtained by eliminating the i-th observation (X_i, Y_i). Moreover, the arguments developed in the proof of Theorem 3.1 imply that $\Pr\left(\{\widehat{\theta}_{\mathrm{C}}^s(\breve{X}_i^s, \breve{Y}_i^s \mid \mathcal{X}_{n_s}^s) = 1\}\right) = O\left(n_s^{-\frac{3}{p+q+1}} (\log n_s)^{\frac{3}{p+q+1}}\right)$. Hence,

$$\begin{aligned} &E\left(\Gamma(\widehat{\theta}_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t \mid \mathcal{X}_{n_s}^s)) - \Gamma(\theta_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t))\right) \\ &= E\left(\Gamma(\widehat{\theta}_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t \mid \mathcal{X}_{n_s,-i}^s)) - \Gamma(\theta_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t)) \mid (\breve{X}_i^t, \breve{Y}_i^t) \notin \mathcal{W}^s(\nu_{n_s}^*)\right) \cdot \Pr((\breve{X}_i^t, \breve{Y}_i^t) \notin \mathcal{W}^s(\nu_{n_s}^*)) \\ &\quad + \sum_{l=1}^{2} E\left(\Gamma(\widehat{\theta}_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t \mid \mathcal{X}_{n_s,-i}^s)) - \Gamma(\theta_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t)) \mid (\breve{X}_i^t, \breve{Y}_i^t) \in \mathcal{W}_{n_s,l}^s\right) \cdot \Pr((\breve{X}_i^t, \breve{Y}_i^t) \in \mathcal{W}_{n_s,l}^s) \\ &\quad + O\left(n_s^{-\frac{3}{p+q+1}} (\log n_s)^{\frac{3}{p+q+1}}\right) \end{aligned} \tag{A.43}$$

Note that $(\breve{X}_i^t, \breve{Y}_i^t)$ is independent of $\mathcal{X}_{n_s,-i}^s$, and that by definition of our coordinate system $(\breve{X}_i^t, \breve{Y}_i^t) \notin \mathcal{W}^s(\nu_{n_s}^*)$ if and only if $(X_i^t, Y_i^t) \notin \mathcal{W}^s(\nu_{n_s}^*)$, as well as $(\breve{X}_i^t, \breve{Y}_i^t) \in \mathcal{W}_{n_s,l}^s$ if and only if $(X_i^t, Y_i^t) \in \mathcal{W}_{n_s,l}^s$ for $l = 1, 2$. As $n_s \to \infty$, our assumptions on the densities f^1 and f^2 the probabilities of these events are of the same order of magnitude as those obtained when analyzing (X_i^s, Y_i^s). Therefore, (3.27) follows from $\Pr((\breve{X}_i^t, \breve{Y}_i^t) \in \mathcal{W}_{n_s,2}^s) = O\left(n_s^{-\frac{4}{p+q+1}} (\log n_s)^{\frac{4}{p+q+1}}\right)$ and arguments similar to (A.38) and (A.40).

In an analogous manner a straightforward generalizations of the arguments in the proof of Theorem 3.1 lead to $E\left([\widehat{\theta}_{\mathrm{C}}^s(\breve{X}_i^t, \breve{Y}_i^t \mid \mathcal{X}_{n_s}^s) - \theta_{\mathrm{C}}^s(X_i^t, Y_i^t)]^2\right) = O\left(n_s^{-\frac{4}{p+q+1}} (\log n_s)^{\frac{4}{p+q+1}}\right)$,

and (3.28) is obtained by an argument similar to (A.41). Finally, (3.29) can be derived from straightforward generalizations of (A.34) and (A.35). ■

A.8 Proof of Theorem 3.6

We only have to show (3.32). First note that the additional assumptions a) - c) imply $\mu_{\mathcal{M}} = 0$. Since for both samples $s = 1, 2$ the same algorithm is employed to determine $\log \widehat{\gamma}^s(x, y \mid \mathcal{X}_n^s)$, there exist a measurable function G such that $\log \widehat{\gamma}^s(x, y \mid \mathcal{X}_n^s) = G((x, y); (X_1^s, Y_1^s), \ldots, (X_n^s, Y_n^s))$. Since by a) and b) distributions are identical, we necessarily have

$$\begin{aligned} E\left(\log \widehat{\gamma}^1(X_i^1, Y_i^1 \mid \mathcal{X}_n^1)\right) &= E\left(G((X_i^1, Y_i^1); (X_1^1, Y_1^1), \ldots, (X_n^1, Y_n^1))\right) \\ &= E\left(G((X_i^2, Y_i^2); (X_1^2, Y_1^2), \ldots, (X_n^2, Y_n^2))\right) \\ &= E\left(\log \widehat{\gamma}^2(X_i^2, Y_i^2 \mid \mathcal{X}_n^2)\right). \end{aligned}$$

for all $i = 1, \ldots, n$. When additionally using c) we furthermore obtain

$$\begin{aligned} &E\left(\log \widehat{\gamma}^1(X_i^2, Y_i^2 \mid \mathcal{X}_n^1)\right) = E\left(G((X_i^2, Y_i^2); (X_1^1, Y_1^1), \ldots, (X_n^1, Y_n^1))\right) \\ &= \int E\left(G((x^2, y^2); (X_1^1, Y_1^1), \ldots, (x^1, y^1), \ldots, (X_n^1, Y_n^1))\right) f_{12}(x^1, y^1, x^2, y^2) dx^1 \ldots dy^2 \\ &= \int E\left(G((x^2, y^2); (X_1^1, Y_1^1), \ldots, (x^1, y^1), \ldots, (X_n^1, Y_n^1))\right) f_{12}(x^2, y^2, x^1, y^1) dx^1 \ldots dy^2 \\ &= \int E\left(G((x^1, y^1); (X_1^2, Y_1^2), \ldots, (x^2, y^2), \ldots, (X_n^2, Y_n^2))\right) f_{12}(x^1, y^1, x^2, y^2) dx^1 \ldots dy^2 \\ &= E\left(G((X_i^1, Y_i^1); (X_1^2, Y_1^2), \ldots, (X_n^2, Y_n^2))\right) = E\left(\log \widehat{\gamma}^2(X_i^1, Y_i^1 \mid \mathcal{X}_n^2)\right). \end{aligned}$$

By definition of $\log \widehat{\mathcal{M}}_i$ this implies that $E\left(\widehat{\mu}_{\mathcal{M},n}\right) = E\left(\log \widehat{\mathcal{M}}_i\right) = 0$. ■

A.9 Proof of Theorem 3.7

By theorem 3.7 we obtain

$$\begin{aligned} &\sqrt{n}(\widehat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}} - \mathcal{R}_n) = \\ &\qquad \frac{\sqrt{n}}{n} \sum_{i=1}^{n} (\log \widehat{\mathcal{M}}_i - \log \mathcal{M}_i - E(\log \widehat{\mathcal{M}}_i) + \mu_{\mathcal{M}}) + \frac{\sqrt{n}}{n} \sum_{i=1}^{n} (\log \mathcal{M}_i - \mu_{\mathcal{M}}) \end{aligned} \tag{A.44}$$

Since (3.28) and (3.29) imply $\frac{\sqrt{n}}{n} \sum_{i=1}^{n} (\log \widehat{\mathcal{M}}_i - \log \mathcal{M}_i - E(\log \widehat{\mathcal{M}}_i) + \mu_{\mathcal{M}}) \to_P 0$, the assertion is now an immediate consequence of standard CLTs. ■

A.10 Proof of Theorem 3.8

The result follows from straightforward arguments based on the delta method: Indeed, a Taylor expansion yields

$$\sqrt{n}\left(\exp(\widehat{\mu}_{\mathcal{M},n}) - \exp(\mu_{\mathcal{M}} + \mathcal{R}_n)\right) = \exp(\mu_{\mathcal{M}} + \mathcal{R}_n) \cdot \sqrt{n}\left(\widehat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}} - \mathcal{R}_n\right) + O_P(\frac{1}{\sqrt{n}}). \tag{A.45}$$

Since $R_n = O\left(n^{-\frac{2}{p+q+1}}\right)$, the desired result follows from a further Taylor expansion of $\exp(\mu_{\mathcal{M}} + \mathcal{R}_n)$ and Theorem 3.7. ■

A.11 Proof of Lemma 4.1

The proof is straightforward:

$$\begin{aligned}
\widehat{\sigma}^2_{\mathcal{M},n} &= n^{-1}\sum_{i=1}^{n}\left(\log\widehat{\mathcal{M}}_i - \widehat{\mu}_{\mathcal{M},n}\right)^2 \\
&\xrightarrow{p} E\left[\left(\log\widehat{\mathcal{M}}_i\right)^2\right] - \mu^2_{\mathcal{M}} \\
&= \mathrm{VAR}\left(\log\mathcal{M}_i\right) + \left[E(\log\mathcal{M}_i)\right]^2 - \mu^2_{\mathcal{M}} \\
&= \sigma^2_{\mathcal{M}}
\end{aligned}$$

since $\left[E(\log\mathcal{M}_i)\right]^2 - \mu^2_{\mathcal{M}} = 0$. ■

A.12 Proof of Theorem 4.1

The result follows directly from Theorem 3.7 after noting that the big-O remainder term in (3.37) is $o\left(n^{-\kappa}\right)$ and noting that $n^{\kappa}o\left(n^{-\kappa}\right) = o(1)$. Since $\widehat{\mu}_{\mathcal{M},n}$ in (3.36) has been replaced with $\widehat{\mu}_{\mathcal{M},n_\kappa}$ in (4.5), the scale factor needed to stabilize variance is n^{κ}. ■

A.13 Proof of Theorem 4.2

The result follows after substituting $\widehat{B}_{n,\kappa}$ for the bias term in (3.37). For $(p+q) = 4$ we have $\kappa = 2/5$. The remainder term is $O\left(n^{-3\kappa/2}\right)$ ignoring the $(\log n)$ term which does not affect the rate. Then $\sqrt{n}O\left(n^{-3\kappa/2}\right) = O\left(n^{-1/10}\right)$. ■

References

Bahadur, R. R. and L. J. Savage (1956), The nonexistence of certain statistical procedures in nonparametric problems, *The Annals of Mathematical Statistics* 27, 1115–1122.

Ball, V. E., C. A. K. Lovell, H. Luu, and R. Nehring (2004), Incorporating environmental impacts in the measurement of agricultural productivity growth, *Journal of Agricultural and Resource Economics* 29, 436–460.

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

Cran, G. W., K. J. Martin, and G. E. Thomas (1977), Remark AS R19 and algorithm AS 109: A remark on algorithms: AS 63: The incomplete beta integral AS 64: Inverse of the incomplete beta function ratio, *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 26, 111–114.

Daraio, C., L. Simar, and P. W. Wilson (2018), Nonparametric estimation of efficiency in the presence of environmental variables, *Econometrics Journal* In press, doi: 10.1111/ectj.12103.

Färe, R. (1988), *Fundamentals of Production Theory*, Berlin: Springer-Verlag.

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

Färe, R., S. Grosskopf, and C. A. K. Lovell (1985), *The Measurement of Efficiency of Production*, Boston: Kluwer-Nijhoff Publishing.

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

Gray, H. L. and R. R. Schucany (1972), *The Generalized Jackknife Statistic*, New York: Marcel Decker, Inc.

Grifell-Tatjé, E. and C. A. K. Lovell (1995), A note on the Malmquist productivity index, *Economic Letters* 47, 169–175.

Jeong, S. O. (2004), Asymptotic distribution of DEA efficiency scores, *Journal of the Korean Statistical Society* 33, 449–458.

Jeong, S. O. and B. U. Park (2006), Large sample approximation of the distribution for convex-hull estimators of boundaries, *Scandinavian Journal of Statistics* 33, 139–151.

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

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

— (2015), When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores, *Econometric Theory* 31, 394–422.

— (2016), Testing hypotheses in nonparametric models of production, *Journal of Business and Economic Statistics* 34, 435–456.

Malmquist, S. (1953), Index numbers and indifference surfaces, *Trabajos de Estatística y de Investigación Operativa* 4, 209–242.

Nishimizu, M. and J. Page (1982), Total factor productivity growth, technological progress and technical efficiency change: Dimensions of productivity change in Yugoslavia, *Economic Journal* 92, 920–936.

Noble, B. and J. W. Daniel (1977), *Applied Linear Algebra*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 2nd edition.

Park, B. U., S.-O. Jeong, and L. Simar (2010), Asymptotic distribution of conical-hull estimators of directional edges, *Annals of Statistics* 38, 1320–1340.

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

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

— (1999), Estimating and bootstrapping Malmquist indices, *European Journal of Operational Research* 115, 459–471.

— (2011), Inference by the m out of n bootstrap in nonparametric frontier models, *Journal of Productivity Analysis* 36, 33–53.

— (2013), Estimation and inference in nonparametric frontier models: Recent developments and perspectives, *Foundations and Trends in Econometrics* 5, 183–337.

— (2015), Statistical approaches for non-parametric frontier models: A guided tour, *International Statistical Review* 83, 77–110.

Wilson, P. W. (2011), Asymptotic properties of some non-parametric hyperbolic efficiency estimators, in I. Van Keilegom and P. W. Wilson, eds., *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, Berlin: Springer-Verlag, pp. 115–150.

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

Table 1: Rejection Rates for Test for Productivity Change using Logs (Two-sided Test)

n	β	— $p=1,\ q=1$ —			— $p=2,\ q=1$ —			— $p=3,\ q=1$ —			— $p=4,\ q=1$ —			— $p=5,\ q=1$ —		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
25	0.000	0.116	0.058	0.012	0.124	0.063	0.014	0.130	0.068	0.016	0.123	0.064	0.014	0.126	0.069	0.016
	0.005	0.183	0.108	0.030	0.194	0.115	0.034	0.204	0.125	0.039	0.148	0.084	0.021	0.143	0.082	0.021
	0.010	0.354	0.245	0.099	0.373	0.264	0.111	0.389	0.278	0.122	0.214	0.137	0.045	0.192	0.123	0.040
	0.015	0.566	0.445	0.233	0.594	0.473	0.257	0.609	0.490	0.274	0.313	0.219	0.090	0.268	0.185	0.074
	0.020	0.756	0.650	0.419	0.782	0.681	0.453	0.795	0.698	0.474	0.429	0.325	0.158	0.362	0.266	0.124
	0.030	0.954	0.914	0.774	0.965	0.931	0.806	0.970	0.939	0.821	0.658	0.554	0.350	0.565	0.458	0.271
	0.040	0.996	0.989	0.946	0.997	0.992	0.959	0.998	0.994	0.963	0.822	0.743	0.561	0.736	0.641	0.448
	0.050	1.000	0.999	0.990	1.000	1.000	0.993	1.000	1.000	0.994	0.915	0.864	0.729	0.853	0.781	0.614
50	0.000	0.109	0.053	0.010	0.114	0.057	0.011	0.117	0.059	0.012	0.112	0.059	0.013	0.112	0.062	0.015
	0.005	0.231	0.143	0.043	0.241	0.151	0.048	0.250	0.158	0.051	0.147	0.085	0.023	0.135	0.079	0.023
	0.010	0.513	0.387	0.182	0.539	0.413	0.200	0.555	0.429	0.213	0.244	0.159	0.056	0.200	0.129	0.046
	0.015	0.789	0.684	0.442	0.814	0.717	0.480	0.827	0.734	0.502	0.379	0.274	0.120	0.298	0.208	0.088
	0.020	0.939	0.890	0.722	0.952	0.911	0.760	0.958	0.920	0.780	0.531	0.415	0.219	0.417	0.310	0.151
	0.030	0.999	0.996	0.976	0.999	0.998	0.983	1.000	0.998	0.986	0.791	0.694	0.483	0.660	0.547	0.338
	0.040	1.000	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	0.929	0.876	0.726	0.840	0.753	0.554
	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.979	0.956	0.874	0.935	0.883	0.737
100	0.000	0.104	0.051	0.010	0.109	0.055	0.010	0.113	0.058	0.011	0.105	0.055	0.012	0.106	0.058	0.015
	0.005	0.331	0.222	0.080	0.351	0.239	0.090	0.362	0.251	0.095	0.159	0.094	0.028	0.138	0.082	0.025
	0.010	0.751	0.638	0.389	0.776	0.672	0.424	0.789	0.688	0.445	0.299	0.202	0.079	0.224	0.147	0.056
	0.015	0.961	0.925	0.788	0.970	0.942	0.823	0.974	0.948	0.839	0.488	0.369	0.179	0.352	0.249	0.111
	0.020	0.998	0.994	0.968	0.999	0.996	0.978	0.999	0.997	0.982	0.678	0.560	0.330	0.502	0.381	0.195
	0.030	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.921	0.858	0.679	0.779	0.669	0.438
	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.971	0.896	0.932	0.871	0.696
	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.974	0.984	0.962	0.870

Table 1: Rejection Rates for Test for Productivity Change using Logs (continued)

n	β	— $p=1,\ q=1$ —			— $p=2,\ q=1$ —			— $p=3,\ q=1$ —			— $p=4,\ q=1$ —			— $p=5,\ q=1$ —		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
250	0.000	0.102	0.051	0.010	0.107	0.053	0.010	0.108	0.055	0.011	0.104	0.054	0.012	0.102	0.055	0.014
	0.005	0.592	0.464	0.234	0.617	0.493	0.258	0.633	0.507	0.273	0.201	0.125	0.041	0.159	0.096	0.031
	0.010	0.977	0.953	0.851	0.983	0.963	0.878	0.985	0.968	0.892	0.437	0.320	0.144	0.306	0.208	0.085
	0.015	1.000	1.000	0.997	1.000	1.000	0.998	1.000	1.000	0.999	0.703	0.582	0.344	0.505	0.380	0.187
	0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.889	0.811	0.597	0.705	0.581	0.342
	0.030	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.985	0.932	0.943	0.889	0.709
	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.995	0.986	0.930
	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.990
500	0.000	0.102	0.051	0.010	0.103	0.051	0.010	0.107	0.055	0.011	0.101	0.052	0.012	0.101	0.053	0.013
	0.005	0.842	0.753	0.522	0.862	0.779	0.558	0.871	0.793	0.580	0.256	0.167	0.060	0.185	0.114	0.039
	0.010	1.000	0.999	0.995	1.000	1.000	0.997	1.000	1.000	0.998	0.591	0.464	0.242	0.393	0.279	0.122
	0.015	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.868	0.782	0.556	0.647	0.519	0.286
	0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.977	0.950	0.836	0.850	0.753	0.514
	0.030	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.990	0.975	0.893
	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.993
	0.050	1.000	1.000	1.000	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	0.000	0.102	0.051	0.010	0.102	0.051	0.010	0.106	0.054	0.011	0.101	0.051	0.011	0.100	0.051	0.012
	0.005	0.982	0.962	0.875	0.986	0.971	0.899	0.988	0.975	0.911	0.341	0.234	0.093	0.228	0.145	0.051
	0.010	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.767	0.657	0.411	0.521	0.394	0.192
	0.015	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.969	0.936	0.807	0.806	0.699	0.454
	0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.975	0.954	0.907	0.740
	0.030	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.986
	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

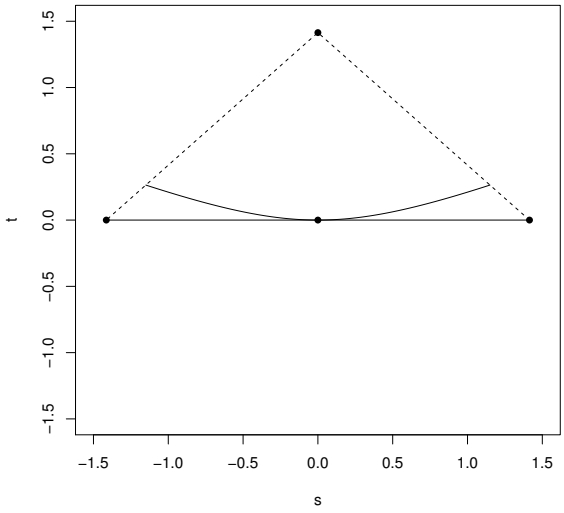
Table 2: Rejection Rates for Test for Productivity Change using Geometric Mean (Two-sided Test)

n	β	— $p=1,\ q=1$ —			— $p=2,\ q=1$ —			— $p=3,\ q=1$ —			— $p=4,\ q=1$ —			— $p=5,\ q=1$ —		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
25	0.000	0.118	0.061	0.014	0.126	0.066	0.016	0.134	0.073	0.018	0.125	0.066	0.014	0.127	0.070	0.017
	0.005	0.150	0.080	0.018	0.157	0.086	0.021	0.164	0.091	0.024	0.138	0.078	0.020	0.136	0.081	0.024
	0.010	0.295	0.188	0.062	0.307	0.200	0.070	0.314	0.209	0.078	0.191	0.121	0.040	0.175	0.114	0.040
	0.015	0.497	0.364	0.161	0.515	0.384	0.178	0.522	0.393	0.191	0.272	0.187	0.074	0.236	0.164	0.068
	0.020	0.692	0.562	0.314	0.712	0.586	0.341	0.717	0.595	0.355	0.369	0.271	0.125	0.312	0.226	0.105
	0.030	0.926	0.859	0.655	0.936	0.876	0.686	0.937	0.877	0.695	0.567	0.458	0.265	0.477	0.376	0.209
	0.040	0.988	0.968	0.873	0.990	0.972	0.888	0.989	0.971	0.889	0.723	0.628	0.426	0.626	0.523	0.335
	0.050	0.998	0.992	0.956	0.997	0.991	0.958	0.996	0.990	0.956	0.826	0.748	0.571	0.738	0.646	0.458
50	0.000	0.110	0.055	0.011	0.115	0.059	0.012	0.119	0.062	0.014	0.113	0.060	0.013	0.112	0.062	0.016
	0.005	0.198	0.114	0.027	0.205	0.118	0.030	0.209	0.122	0.033	0.139	0.082	0.024	0.130	0.079	0.027
	0.010	0.465	0.330	0.130	0.484	0.350	0.143	0.493	0.360	0.152	0.221	0.144	0.052	0.186	0.122	0.048
	0.015	0.750	0.625	0.355	0.772	0.653	0.386	0.782	0.666	0.402	0.335	0.239	0.103	0.266	0.186	0.082
	0.020	0.921	0.852	0.631	0.934	0.873	0.670	0.939	0.882	0.686	0.467	0.353	0.177	0.363	0.266	0.129
	0.030	0.998	0.992	0.948	0.998	0.994	0.958	0.998	0.994	0.961	0.705	0.595	0.376	0.567	0.453	0.262
	0.040	1.000	1.000	0.994	1.000	1.000	0.995	1.000	1.000	0.994	0.858	0.777	0.582	0.735	0.628	0.419
	0.050	1.000	1.000	0.999	1.000	1.000	0.998	1.000	1.000	0.998	0.933	0.882	0.737	0.846	0.760	0.567
100	0.000	0.105	0.052	0.010	0.110	0.055	0.011	0.114	0.059	0.012	0.106	0.056	0.013	0.106	0.059	0.016
	0.005	0.302	0.192	0.059	0.316	0.205	0.066	0.325	0.213	0.069	0.152	0.091	0.030	0.135	0.083	0.030
	0.010	0.723	0.596	0.326	0.747	0.627	0.358	0.757	0.641	0.375	0.274	0.185	0.074	0.210	0.140	0.058
	0.015	0.953	0.908	0.735	0.963	0.925	0.771	0.966	0.931	0.786	0.440	0.325	0.154	0.317	0.224	0.103
	0.020	0.997	0.992	0.952	0.998	0.994	0.964	0.998	0.995	0.968	0.612	0.488	0.270	0.442	0.329	0.166
	0.030	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.864	0.774	0.557	0.688	0.564	0.340
	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.963	0.922	0.784	0.856	0.762	0.541
	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.975	0.904	0.941	0.883	0.710

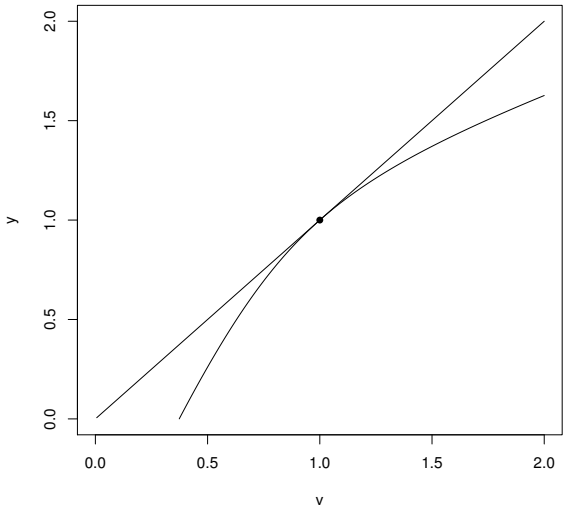
Table 2: Rejection Rates for Test for Productivity Change using Geometric Mean (continued)

n	β	— $p=1,\ q=1$ —			— $p=2,\ q=1$ —			— $p=3,\ q=1$ —			— $p=4,\ q=1$ —			— $p=5,\ q=1$ —		
		.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01	.10	.05	.01
250	0.000	0.103	0.051	0.010	0.107	0.054	0.011	0.109	0.055	0.012	0.104	0.054	0.013	0.102	0.055	0.014
	0.005	0.572	0.438	0.204	0.594	0.463	0.225	0.607	0.476	0.237	0.194	0.122	0.043	0.156	0.097	0.035
	0.010	0.974	0.946	0.826	0.980	0.957	0.854	0.983	0.962	0.870	0.407	0.294	0.131	0.286	0.196	0.083
	0.015	1.000	1.000	0.996	1.000	1.000	0.998	1.000	1.000	0.998	0.655	0.527	0.296	0.459	0.340	0.165
	0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.845	0.747	0.508	0.640	0.510	0.285
	0.030	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.985	0.963	0.859	0.895	0.808	0.580
	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.976	0.981	0.952	0.822
	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.998	0.991	0.942
500	0.000	0.102	0.051	0.010	0.103	0.052	0.010	0.108	0.055	0.012	0.101	0.052	0.012	0.100	0.054	0.014
	0.005	0.833	0.738	0.494	0.852	0.764	0.528	0.862	0.778	0.550	0.247	0.162	0.060	0.181	0.114	0.042
	0.010	1.000	0.999	0.994	1.000	1.000	0.996	1.000	1.000	0.997	0.558	0.429	0.218	0.368	0.261	0.117
	0.015	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.833	0.732	0.489	0.599	0.468	0.249
	0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.962	0.920	0.762	0.798	0.683	0.433
	0.030	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.983	0.976	0.942	0.797
	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996	0.962
	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996
1000	0.000	0.102	0.051	0.010	0.102	0.051	0.010	0.106	0.054	0.011	0.101	0.051	0.011	0.100	0.052	0.012
	0.005	0.981	0.960	0.865	0.985	0.969	0.889	0.987	0.973	0.902	0.329	0.226	0.091	0.223	0.144	0.054
	0.010	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.738	0.620	0.373	0.490	0.366	0.177
	0.015	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.956	0.911	0.749	0.763	0.645	0.396
	0.020	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.992	0.950	0.928	0.861	0.653
	0.030	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.994	0.956
	0.040	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
	0.050	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

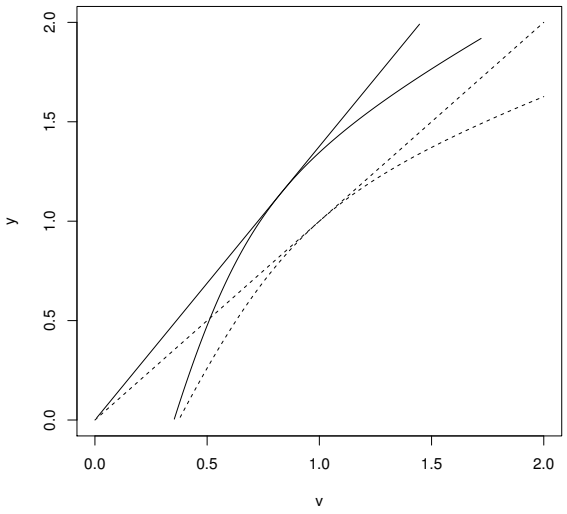
Figure 1: Technology in Two Periods



(a)



(b)



(c)