

CONICAL FDH ESTIMATORS FOR TESTING RETURNS TO SCALE AND MAKING INFERENCE ABOUT CHANGES IN PRODUCTIVITY

Alois Kneip, Léopold Simar, Paul W. Wilson

REPRINT | 2025 / 22

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

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

Conical FDH Estimators of General Technologies, with Applications to Returns to Scale and Malmquist Productivity Indices

ALOIS KNEIP

LÉOPOLD SIMAR

PAUL W. WILSON*

August 2022

Abstract

Nonparametric envelopment estimators are often used to estimate the attainable sets and its efficient boundary and to assess efficiency and changes in productivity. Kneip et al. (2015, 2016) provide asymptotic theory enabling inference about expected efficiency and testing constant versus variable returns to scale using these estimators, and Kneip et al. (2021) provide asymptotic results that can be used to make inference about expected changes in productivity measured by Malmquist indices. All of these results require convexity of the attainable set, but in a number of situations this assumption is questionable. Kneip et al. (2016) also provide a test of convexity versus non-convexity of a production set, and convexity is rejected in several recent studies. This paper extends the results mentioned above to allow for possibly non-convex technologies. Properties of a nonparametric envelopment estimator of distance to the boundary of the *cone* spanned by a possibly non-convex attainable set are derived. These new results are then extended to make inference about geometric means of Malmquist indices and to test constant versus non-constant returns to scale when the production set is not convex.

Keywords: Non-convex production sets, FDH, hypothesis test, returns to scale, Malmquist index, productivity change.

JEL classification: C12, C14, C18

*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 (ISBA), LIDAM, Université Catholique de Louvain-la-Neuve, Voie du Roman Pays 20, B 1348 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.

1 Introduction

Nonparametric envelopment estimators are widely used to evaluate the performance of firms, schools, utilities, local governments and other decision-making units in terms of efficiency and productivity. These estimators are based on economic theory of production provided by Koopmans (1951) and Debreu (1951). Farrell (1957) was the first to estimate efficiency from observations on production units, and his method—termed "data envelopment analysis" (DEA) by Rhodes (1978)—was later made popular by Charnes et al. (1978). DEA estimators impose convexity on the set of feasible combinations of inputs and outputs (i.e., the production set). Deprins et al. (1984) proposed alternative estimators based on the free-disposal hull (FDH) of the sample observations, thereby relaxing the assumption of convexity required by DEA estimators.

DEA and FDH estimators are used to estimate the technical efficiency of producers in terms of their observed distance in input-output space to the boundary, or frontier, of the unobserved production set. Various versions of DEA estimators exist; e.g., the version introduced by Farrell (1957) assumes constant returns to scale (CRS), while the version of Banker et al. (1984) allows for variable returns to scale (VRS), i.e., increasing, constant, or decreasing returns to scale. The FDH estimator introduced by Deprins et al. (1984) makes no restrictions on returns to scale, while Kerstens and Vanden Eeckaut (1997) describes a restricted version of the FDH estimator that imposes an assumption of CRS but does not impose convexity on the production set.

The statistical properties of DEA estimators are by now fully developed. Under appropriate assumptions, DEA estimators are consistent with known rates of convergence and properties of their moments (see Kneip et al., 1998; Kneip et al., 2008, 2015, 2016, 2021); and Park et al., 2010 for details). Similar properties of ordinary FDH estimators, with no restrictions on returns to scale, are also known (see Park et al., 2000; Daouia et al., 2017; and Kneip et al., 2015 for details). However, until now, properties of FDH estimators under CRS, as well as estimators of distance to the boundary of the cone spanned by the possibly non-convex production set (needed for inference about Malmquist productivity indices when production sets are not convex) have been elusive. This paper fills the missing gap, providing results that enable researchers to test constant versus non-constant returns to scale and to

make inference about changes in productivity using Malmquist indices when production sets are perhaps not convex.

An applied researcher confronted with data and wanting to use nonparametric envelopment estimators to examine the performance of observed firms must decide which estimator to use. If firms from more than one group are observed, one should first test whether firms in different groups face different frontiers; this amounts to testing “separability” with respect to a discrete variable describing group membership along the lines of Daraio et al. (2018). The FDH estimator should be used for this test since it is unknown whether production sets are convex. If the null hypothesis of a single frontier is rejected, then the groups must be treated separately; otherwise, the data can be pooled for further analysis. Next, one should test convexity versus non-convexity of the production sets for each group (if separability is rejected) or perhaps for the pooled data (if separability is not rejected) using the convexity test developed by Kneip et al. (2016). For a given group, if convexity is not rejected, one can test CRS versus non-CRS using DEA estimators to implement the returns-to-scale test described by Kneip et al. (2016). But if convexity is rejected, until now no test of CRS versus non-CRS has been available. In addition, in cases where convexity does not hold, until now there has been no theory to make inferences about productivity change using Malmquist indices. This is the strategy described by Apon et al. (2015) in their Figure 1, but their description is incomplete due to the lack (at the time of their study) of results for cases where production sets are not convex.

We provide new theory enabling tests of CRS versus non-CRS as well as inference about changes in productivity when production sets are not convex. In addition, we provide Monte Carlo evidence on the performance of our returns-to-scale test in terms of size and power as well as coverages of confidence intervals for expected efficiency under CRS and non-convexity. We use our new results to revisit the study of U.S. municipal governments by O’Loughlin and Wilson (2021) who found evidence against separability with respect to time and census regions and against convexity of production sets for region-years. We use the O’Loughlin and Wilson (2021) data to test CRS versus non-CRS without imposing convexity using our new test, and we provide inference on productivity change over time across census regions using our new results. We find statistically significant evidence against CRS, and we find significant evidence of increases in productivity in 10 out of 12 cases examined.

A number of recent studies have tested convexity versus non-convexity of production sets using nonparametric envelopment estimators and the test developed by Kneip et al. (2016) and have rejected convexity. Examples include Apon et al. (2015) and Daraio et al. (2021) (universities); O’Loughlin and Wilson (2021) (municipal governments’ production of local public goods); Wilson (2021) and Wilson and Zhao (2022) (banks); Qiao and Wang (2021) (semi-conductors); and Simar and Wilson (2022) (hospitals). Until now, researchers have lacked statistical tools to investigate either returns to scale or productivity change measured by Malmquist indices when convexity of the production set is rejected. Given the wide range of industries where convexity has been rejected, it seems that (i) convexity should be routinely tested in nonparametric efficiency studies, and (ii) the results provided in this paper are important for applied researchers.

The paper proceeds as follows. The next section establishes notation and a general non-parametric model of production where the production set is not necessarily convex. Section 3 discusses FDH estimators of distance to the cone of the production in the general model, and Section 4 extends the results to cases where CRS holds. Section 5 demonstrates that standard Central Limit Theorems (CLTs) (e.g., the Lindeberg-Feller CLT) do not hold for means of the new FDH estimators under either CRS or non-CRS, and establishes new CLT results that enable inference about expectations of efficiency under CRS as well as expected distance to the cone of the production set under non-CRS without requiring convexity of the production set. Section 6 extends the new CLT results to test CRS versus non-CRS under either convexity or non-convexity (Section 6.1) and to make inference about Malmquist indices when production sets are possibly not convex (Section 6.2). Monte Carlo evidence on the size and power of our returns-to-scale test is provided in Section 7. In Section 8, we use the new results to revisit and extend the study by O’Loughlin and Wilson (2021) of municipal governments’ provision of local public goods. We test constant versus non-constant returns to scale, and estimate Malmquist indices and make inference about changes in productivity over time.

2 A General Nonparametric Production Model

In order to establish notation and describe the basic model, 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. The set of feasible combinations of inputs and outputs is given by

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

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

$$\Psi^\partial := \{(x, y) \mid (x, y) \in \Psi, (\vartheta x, \vartheta^{-1}y) \notin \Psi \ \forall \ \vartheta \in (0, 1)\}. \quad (2.2)$$

The following assumptions are typical in micro-economic theory of the firm.

Assumption 2.1. Ψ is closed.

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

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

Here and throughout, inequalities involving vectors are defined on an element-by-element basis. Assumption 2.2 requires some input to be used to produce any non-zero output, while Assumption 2.3 imposes weak monotonicity on the frontier. Note that Ψ is *not* assumed to be convex.

The efficiency of a production plan (x, y) is measured by its distance to the efficient boundary, and of course this distance depends on the chosen direction to the frontier. The Farrell (1957) output efficiency measure

$$\zeta(x, y \mid \Psi) := \sup \{\zeta \mid (x, \zeta y) \in \Psi\} \quad (2.3)$$

gives the feasible proportionate expansion of output quantities to reach the frontier holding inputs constant. The Farrell (1957) input efficiency measure is given by

$$\theta(x, y \mid \Psi) := \inf \{\theta \mid (\theta x, y) \in \Psi\} \quad (2.4)$$

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, $\zeta(x, y \mid \Psi) \geq 1$ and $\theta(x, y \mid \Psi) \leq 1$ for all $(x, y) \in \Psi$.

An alternative measure of efficiency is the hyperbolic graph measure of efficiency

$$\gamma(x, y \mid \Psi) := \inf \{ \gamma > 0 \mid (\gamma x, \gamma^{-1} y) \in \Psi \} \quad (2.5)$$

introduced by Färe et al. (1985). By construction, $\gamma(x, y \mid \Psi) \leq 1$ for $(x, y) \in \Psi$. 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

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

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

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

is the cone spanned by the set Ψ . Note that this is a pointed cone (i.e., $\mathcal{C}(\Psi)$ includes $\{(0, 0)\}$), but since Ψ is not necessarily convex, neither is $\mathcal{C}(\Psi)$ necessarily convex. Analogous to (2.2), the frontier of this set is given by

$$\mathcal{C}^\partial(\Psi) := \{\mathcal{L}(x, y) \mid \mathcal{L}(x, y) \in \mathcal{C}(\Psi), \mathcal{L}(x, \zeta y) \notin \mathcal{C}(\Psi) \text{ for any } \zeta \in (1, \infty)\}. \quad (2.8)$$

If $\mathcal{C}(\Psi) = \Psi$, then the frontier Ψ^∂ exhibits globally constant returns to scale (CRS). In applications, whether $\mathcal{C}(\Psi) = \Psi$ is ultimately an empirical question. This is addressed below in Section 6.1.

In order to define Malmquist indices to measure changes in productivity, we must define a measure of distance from a point $(x, y) \in \mathbb{R}_+^{p+q}$ to the frontier $\mathcal{C}^\partial(\Psi)$ of the cone $\mathcal{C}(\Psi)$. Substituting $\mathcal{C}(\Psi)$ for Ψ in (2.3)–(2.5) yields measures $\zeta(x, y \mid \mathcal{C}(\Psi))$, $\theta(x, y \mid \mathcal{C}(\Psi))$ and $\gamma(x, y \mid \mathcal{C}(\Psi))$ of distances from (x, y) to $\mathcal{C}^\partial(\Psi)$ in the input, output and hyperbolic directions, respectively.

Note that Lemmas 3.1 and 3.2 of Kneip et al. (2021) remain valid when Ψ is not convex, as shown by the proofs of the Lemmas. We repeat here Lemma 3.1 of Kneip et al. (2021) because it will be useful for the subsequent analysis.

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

$\zeta(x, y \mid \mathcal{C}(\Psi)) = \zeta(\tilde{x}, \tilde{y} \mid \mathcal{C}(\Psi))$; (iii) $\theta(x, y \mid \mathcal{C}(\Psi)) = \theta(\tilde{x}, \tilde{y} \mid \mathcal{C}(\Psi))$; and (iv) $\gamma(x, y \mid \mathcal{C}(\Psi)) = \gamma(\tilde{x}, \tilde{y} \mid \mathcal{C}(\Psi))$.

This lemma ensures that results established for the input, output or hyperbolic direction extend trivially to the other two directions. Consequently, we simplify by focusing on the input direction in the analysis that follows. We further simplify the notation by defining $\theta(x, y) := \theta(x, y \mid \Psi)$ and $\theta_C(x, y) := \theta(x, y \mid \mathcal{C}(\Psi))$. We assume also that $p \geq 1$ and $q \geq 1$.

The assumptions in Kneip et al. (2021) have to be modified to account for the different framework here, where Ψ is not necessarily convex. First, we adopt the following assumption.

Assumption 2.4. (i) The random variables (X, Y) possess a joint density f with support $\mathcal{D} \subset \Psi$; (ii) The interior of $\mathcal{D}$ is a connected, open subset of Ψ , and $(\theta(x, y)x, y) \in \mathcal{D}$ for all $(x, y) \in \mathcal{D}$; (iii) $\exists$ some $\lambda \geq 1$ and some $d > 0$ such that for all $(x_0, y_0) \in \mathcal{D}$ and any $0 < \delta \leq d$, there is some $(x, y) \in \mathcal{D}$ with $\max_{j=1, \dots, p} |x_j - x_{0j}| \leq \lambda\delta$ and $\max_{j=1, \dots, q} |y_j - y_{0j}| \leq \lambda\delta$ such that $\{(x^*, y^*) \in \mathbb{R}_+^p \times \mathbb{R}_+^q \mid \max_{j=1, \dots, p} |x_j^* - x_j| < \delta \text{ and } \max_{j=1, \dots, q} |y_j^* - y_j| < \delta\} \subset \mathcal{D}$; and (iv) f is twice continuously differentiable on $\mathcal{D}$.

Note that $\mathcal{D}$ is the “observable” part of Ψ . Assumption 2.4 (iii) may look somewhat unusual, but it merely represents a regularity condition on the boundary $\mathcal{D}^\partial$ of $\mathcal{D}$. Note that we do not observe all possible values $(x, y) \in \Psi$, and hence $\mathcal{D}^\partial$ will also contain points (x, y) with $\theta(x, y) < 1$. For any point (x, y) in the interior of $\mathcal{D}$ there necessarily exists some δ such that $\mathcal{D}$ contains the corresponding δ -neighborhood of $\{(x^*, y^*) \in \mathbb{R}_+^p \times \mathbb{R}_+^q \mid \max_{j=1, \dots, p} |x_j^* - x_j| < \delta \text{ and } \max_{j=1, \dots, q} |y_j^* - y_j| < \delta\} \subset \mathcal{D}$. If Assumption 2.4 (iii) were violated, there would exist some boundary point $(x_0, y_0) \in \mathcal{D}^\partial$ such that for any δ , the closest point (x, y) with a δ -neighborhood in the interior of $\mathcal{D}$ is very far away. Then $\mathcal{D}$ would contain a very long, arbitrarily narrow fragment (with diameter less than δ) between (x, y) and the “edge” (x_0, y_0) . The assumption excludes the existence of such weird edges. The assumption necessarily holds if $\mathcal{D}$ is convex, as assumed in Kneip et al. (2021).

Now we can provide some insight into the distance from a given point (x_0, y_0) to the boundary of Ψ and to the boundary of $\mathcal{C}(\Psi)$, under the free disposability assumption but

without assuming convexity of Ψ .¹ We have

$$\theta(x_0, y_0) := \min\{\theta \mid (\theta x_0, y_0) \in \Psi\} \quad (2.9)$$

$$= \min_{(x,y) \in \Psi; y_0 \leq y} \max_{j=1, \dots, p} \frac{x_j}{x_{0j}}. \quad (2.10)$$

Noting that for any $a > 0$, $\theta(ax_0, ay_0) = \theta(x_0, ay_0)/a$, the input-oriented efficiency $\theta_C(x, y)$ with respect to $\mathcal{C}(\Psi)$ is given by

$$\begin{aligned} \theta_C(x_0, y_0) &:= \inf_{a>0} \left\{ \frac{\theta(x_0, ay_0)}{a} \mid (\theta(x_0, ay_0)x_0, ay_0) \in \Psi \right\} \\ &= \inf_{a>0} \inf_{(x,y) \in \Psi; ay_0 \leq y} \max_{j=1, \dots, p} \frac{x_j}{ax_{0j}} \\ &= \inf_{(x,y) \in \Psi} \inf_{a>0; ay_0 \leq y} \max_{j=1, \dots, p} \frac{x_j}{ax_{0j}} \\ &= \inf_{(x,y) \in \Psi} \frac{\max_{j=1, \dots, p} \frac{x_j}{x_{0j}}}{\min_{j=1, \dots, q} \frac{y_j}{y_{0j}}} \end{aligned} \quad (2.11)$$

The last equality follows from the fact that with $a_{x,y}^{x_0,y_0} := \min_{j=1, \dots, q} \frac{y_j}{y_{0j}}$ we have $a_{x,y}^{x_0,y_0} y_0 \leq y$, while $ay_0 \not\leq y$ for any $a > a_{x,y}^{x_0,y_0}$, and $\max_{j=1, \dots, p} \frac{x_j}{ax_{0j}} > \max_{j=1, \dots, p} \frac{x_j}{a_{x,y}^{x_0,y_0} x_{0j}}$ for any $a < a_{x,y}^{x_0,y_0}$.²

3 FDH Estimators and Their Basic Properties

Consider the random sample $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$. Deprins et al. (1984) proposed the FDH estimator

$$\widehat{\Psi}_{\text{FDH},n} := \bigcup_{(X_i, Y_i) \in \mathcal{S}_n} \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq Y_i, x \geq X_i\} \quad (3.1)$$

of Ψ . Substituting $\widehat{\Psi}_{\text{FDH},n}$ for Ψ in (2.10) leads to the FDH estimator of $\theta(x_0, y_0)$ which can be computed as

$$\widehat{\theta}(x_0, y_0) = \min_{i \in D(x_0, y_0)} \left[\max_{j=1, \dots, p} \left(\frac{X_{ij}}{x_{0j}} \right) \right] \quad (3.2)$$

¹For any efficiency measure $\theta(x, y)$ considered, we 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, $\theta(x, y) := 1$ whenever the set of all possible values satisfying the defining inequalities is the empty set. We use the same convention for the estimators $\widehat{\theta}(x, y)$ described below in Section 3.

²In the definition of $\theta_C(x_0, y_0)$ we use “inf” instead of “min” because the structural Assumptions 2.1–2.3 do not imply the existence of a minimum. For example, if the production set Ψ is strictly concave in some directions, then only an infimum is attained, and $\theta_C(x_0, y_0) = 0$. We assume that Ψ is such that for any (x_0, y_0) , a minimum $\theta_C(x_0, y_0) > 0$ is attained; this excludes strictly concave cases.

where $D(x_0, y_0) := \{i \mid X_i \leq x_0, Y_i \geq y_0\}$.

The conical FDH estimator of $\theta_C(x_0, y_0)$ is obtained by substituting $\widehat{\Psi}_{\text{FDH},n}$ for Ψ in (2.11) to obtain³

$$\widehat{\theta}_C(x_0, y_0) = \min_{i=1, \dots, n} \frac{\max_{j=1, \dots, p} \frac{X_{ij}}{x_{0j}}}{\min_{j=1, \dots, q} \frac{Y_{ij}}{y_{0j}}} \quad (3.3)$$

Note that the expression in (3.3) is much easier to compute than the expression suggested by Kerstens and Vanden Eeckaut (1997). Lemma 3.2 in Kneip et al. (2021) for convex sets extends trivially to the following.

Lemma 3.1. *Under Assumptions 2.1–2.3, for any $(x, y) \in \widehat{\Psi}_{\text{FDH},n}$, (i) $\widehat{\theta}_C(x, y)^{1/2} = \widehat{\zeta}_C(x, y)^{-1/2} = \widehat{\gamma}_C(x, y)$. In addition, for any $(\tilde{x}, \tilde{y}) \in \mathcal{L}(x, y)$ with $(x, y) \in \mathbb{R}_+^{p+q}$, (ii) $\widehat{\zeta}_C(x, y) = \widehat{\zeta}_C(\tilde{x}, \tilde{y})$; (iii) $\widehat{\theta}_C(x, y) = \widehat{\theta}_C(\tilde{x}, \tilde{y})$; and (iv) $\widehat{\gamma}_C(x, y) = \widehat{\gamma}_C(\tilde{x}, \tilde{y})$.*

Properties of the FDH estimator $\widehat{\theta}(\cdot, \cdot)$ for a fixed point (x_0, y_0) are developed in Park et al. (2000) and Daouia et al. (2017). Kneip et al. (2015) give results on the moments of FDH estimators as well as central limit theorem (CLT) results for making inferences about expected efficiency. In order to extend these results to the conical FDH estimator $\widehat{\theta}_C(x_0, y_0)$, some additional notation is needed.

For any p -vector $x = (x_1, \dots, x_p)^T \in \mathbb{R}_+^p$ and $j = 1, \dots, p$, define the p -vector $\tilde{x}^{(j)} = (x_1/x_j, \dots, x_{j-1}/x_j, 1, x_{j+1}/x_j, \dots, x_p/x_j)^T$. Then $x = x_j \cdot \tilde{x}^{(j)}$. Similarly, for any q -vector $y = (y_1, \dots, y_q)^T \in \mathbb{R}_+^q$ and any $j = 1, \dots, q$, define the q -vector $\tilde{y}^{(j)} = (y_1/y_j, \dots, y_{j-1}/y_j, 1, y_{j+1}/y_j, \dots, y_q/y_j)^T$. Again, $y = y_j \cdot \tilde{y}^{(j)}$.

In order to formulate necessary technical assumptions, we characterize the frontier by the function defined by

$$g_j(\tilde{x}^{(j)}, y) := \min_{\tau > 0} \{(\tau \tilde{x}^{(j)}, y) \in \Psi\} \quad (3.4)$$

for any $j = 1, \dots, p$. Note that for any $x^* = a \cdot x$, for some $a > 0$ we have $\tilde{x}^{*(j)} = \tilde{x}^{(j)}$, and for any $(x, y) \in \mathcal{D}$ we have $g_j(\tilde{x}^{(j)}, y) = \theta(x, y)x_j = \theta(ax, y)ax_j \forall a > 0$ with $(ax, y) \in \Psi$.

³If $(x_0, y_0) \notin \widehat{\Psi}_{\text{FDH},n}$, $\widehat{\theta}_C(x_0, y_0)$ may be much larger than 1. In the following we will assume in Assumption 3.2 that $\mathcal{D}$ is bounded, hence we have, for some $\theta_{\max} \geq 1$ and for any $(x_0, y_0) \in \mathcal{D}$, $\widehat{\theta}_C(x_0, y_0) \leq \theta_{\max}$.

If $\theta_C(x_0, y_0) > 0$ there necessarily exists some value $a_{min}^{x_0, y_0} \in \mathbb{R}_+$ such that

$$\begin{aligned}\theta_C(x_0, y_0) &= \frac{\theta(x_0, a_{min}^{x_0, y_0} y_0)}{a_{min}^{x_0, y_0}} = \min_{a>0} \left\{ \frac{\theta(x_0, ay_0)}{a} \mid (\theta(x_0, ay_0)x_0, ay_0) \in \Psi \right\} \\ &= \min_{a>0} \left\{ \frac{g_j(\tilde{x}_0^{(j)}, ay_0)}{ax_{0j}} \mid (\theta(x_0, ay_0)x_0, ay_0) \in \Psi \right\}\end{aligned}\quad (3.5)$$

for any $j = 1, \dots, p$. Note that the point $a_{min}^{x_0, y_0} \in \mathbb{R}_+$ is **not** necessarily unique. For example, under CRS, (3.5) is satisfied for any positive value $a_{min}^{x_0, y_0}$.

When analyzing properties of $\hat{\theta}_C(x_0, y_0)$, an important first result to be obtained is a modification of the Northwest Lemma established for the standard FDH estimator $\hat{\theta}(x_0, y_0)$, in Park et al. (2000, Lemma A.3). Note that the following proposition does not rely on uniqueness of $a_{min}^{x_0, y_0}$, and hence the proposition also holds under an additional CRS assumption.

Proposition 3.1. *Assume $\theta_C(x_0, y_0) > 0$ for some $(x_0, y_0) \in \mathcal{D}$, and let $a_{min}^{x_0, y_0} \in \mathbb{R}_+$ satisfy (3.5). For any $\delta > 0$ and $j = 1, \dots, q$, define*

$$S_{\delta, j}^{x_0, y_0} := \left\{ (x, y) \in \Psi \mid x \leq (1 + \delta) \frac{y_j}{y_{0j} a_{min}^{x_0, y_0}} \theta(x_0, a_{min}^{x_0, y_0} y_0) x_0 \text{ and } \tilde{y}^{(j)} \geq \tilde{y}_0^{(j)} \right\}, \quad (3.6)$$

and let $S_\delta^{x_0, y_0} = \bigcup_{j=1}^q S_{\delta, j}^{x_0, y_0}$. Then

$$\frac{\hat{\theta}_C(x_0, y_0)}{\theta_C(x_0, y_0)} \leq 1 + \delta \quad \text{if and only if} \quad \mathcal{X}_n \cap S_\delta^{x_0, y_0} \neq \emptyset. \quad (3.7)$$

The strong (i.e., free) disposability assumed in Assumption 2.3 implies that the frontier is weakly monotone, but similar to Kneip et al. (2015) for the usual FDH estimator, the next assumption strengthens this by requiring the frontier to be strictly monotone and sufficiently smooth in order to obtain results on moments of the conical FDH estimators.

Assumption 3.1. *For any $j = 1, \dots, p$, (i) the frontier function $g_j(\tilde{x}^{(j)}, y)$ is four-times continuously differentiable on $\mathcal{D}^* := \{(\theta(x_0, a_{min}^{x_0, y_0} y_0)x_0, a_{min}^{x_0, y_0} y_0) \mid (x_0, y_0) \in \mathcal{D}\}$, and (ii) the frontier is strictly monotone so that for all $(x, y) \in \mathcal{D}$, $\frac{\partial}{\partial x_k} g_j(\tilde{x}^{(j)}, y) < 0$, $k = 1, \dots, p$, $k \neq j$, and $\frac{\partial}{\partial y_k} g_j(\tilde{x}^{(j)}, y) > 0$, $k = 1, \dots, q$.*

From (3.5) and the definition of $a_{min}^{x_0, y_0}$,

$$\theta_C(x_0, y_0) = \frac{\theta(x_0, a_{min}^{x_0, y_0} y_0)}{a_{min}^{x_0, y_0}} \leq \frac{\theta(x_0, \alpha \cdot a_{min}^{x_0, y_0} y_0)}{\alpha \cdot a_{min}^{x_0, y_0}} \quad (3.8)$$

for all $\alpha > 0$. Therefore, $\alpha \cdot \theta(x_0, a_{\min}^{x_0, y_0} y_0) \leq \theta(x_0, \alpha \cdot a_{\min}^{x_0, y_0} y_0)$. If the equality in (3.8) holds only for $\alpha = 1$, then $a_{\min}^{x_0, y_0}$ is unique; this happens if and only if

$$(\theta(x_0, a_{\min}^{x_0, y_0} y_0)x_0, a_{\min}^{x_0, y_0} y_0) \in \Psi, \quad (3.9)$$

while

$$(\alpha \cdot \theta(x_0, a_{\min}^{x_0, y_0} y_0)x_0, \alpha \cdot a_{\min}^{x_0, y_0} y_0) \notin \Psi \quad \text{for all } \alpha > 0 \text{ with } \alpha \neq 1. \quad (3.10)$$

This obviously requires that “in the direction” of any possible ray $\mathcal{L}(x, y)$ the production set Ψ is locally strictly convex in neighborhoods of the frontier points $(\theta(x_0, a_{\min}^{x_0, y_0} y_0)x_0, a_{\min}^{x_0, y_0} y_0)$.

The case where the equality in (3.8) is valid for all $\alpha > 0$ corresponds to the case of global constant returns to scale (CRS) and is analyzed in Section 4 below. We can now formulate our final assumption.

Assumption 3.2. (i) The support $\mathcal{D} \subset \Psi$ of f is such that (3.9) and (3.10) are satisfied for all $(x_0, y_0) \in \mathcal{D}$; (ii) $\mathcal{D}^*$ is a compact subset of $\mathcal{D}$; (iii) $f(x, y) > 0$ for all $(x, y) \in \mathcal{D}^*$; (iv) There exists a constant $0 < M < \infty$ such that $\|x\| \leq M$ for all $(x, y) \in \mathcal{D}$.

We now obtain the following theorem, where $\Gamma(\cdot)$ denotes the well-known gamma function. This is the analog of Theorem 3.1 for the usual FDH estimator in Park et al. (2000). note that here a slightly better rate of convergence is achieved due to the geometric structure of the cone $\mathcal{C}(\Psi)$ near its boundary. Note also that we now denote our estimator by $\hat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n)$ to stress the fact that we use the sample $\mathcal{X}_n$ to define the frontier estimate, but we estimate distance to the frontier from a fixed point (x_0, y_0) .

Theorem 3.1. Under Assumptions 2.1–2.4, 3.1 and 3.2, for an arbitrary point (x_0, y_0) in the interior of $\mathcal{D}$ there exists a parameter $\lambda(x_0, y_0) > 0$ such that for all $z > 0$,

$$\Pr \left(n^{\frac{1}{p+q-1/2}} \left(\frac{\hat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n)}{\theta_C(x_0, y_0)} - 1 \right) \leq z \right) = 1 - \exp(-\lambda(x_0, y_0) z^{p+q-1/2}) + o(1), \quad (3.11)$$

i.e., $\frac{\hat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n)}{\theta_C(x_0, y_0)} - 1$ follows a Weibull distribution with parameters $\lambda(x_0, y_0)$ and $p + q - 1/2$, where $\lambda(x_0, y_0)$ depends on (x_0, y_0) , g and f . The analytical structure of $\lambda(x_0, y_0)$ may be inferred from the proof of the theorem.

Moreover,

$$E\left(\widehat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n)\right) = \theta_C(x_0, y_0) + \theta_C(x_0, y_0)\lambda(x_0, y_0)^{-1/\kappa}\Gamma(1 + \frac{1}{\kappa})n^{-1/\kappa} + O(n^{-2/\kappa}) \quad (3.12)$$

and

$$\text{VAR}\left(\widehat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n)\right) = \theta_C(x_0, y_0)^2\lambda(x_0, y_0)^{-2/\kappa}(\Gamma(1 + \frac{2}{\kappa}) - \Gamma(1 + \frac{1}{\kappa})^2)n^{-2/\kappa} + o(n^{-2/\kappa}) \quad (3.13)$$

where $\kappa = p + q - 1/2$.

A second result now concerns moments of $\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$, where the notation makes explicit that we analyze the moments of $\widehat{\theta}_C$ evaluated at the observation $(X_i, Y_i) \in \mathcal{X}_n$. The theorem is the analog of Theorem 3.3 in Kneip et al. (2015) for the usual FDH estimator.

Theorem 3.2. *Under the assumptions of Theorem 3.1 there exists a constant $0 < C_0 < \infty$ such that for all $i, j \in \{1, \dots, n\}$, $i \neq j$,*

$$E\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i)\right) = C_0 n^{-\frac{1}{p+q-1/2}} + O\left(n^{-\frac{2}{p+q-1/2}}(\log n)^{\frac{4}{p+q-1/2}}\right). \quad (3.14)$$

Furthermore,

$$\text{VAR}\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i)\right) = O\left(n^{-\frac{2}{p+q-1/2}}\right) \quad (3.15)$$

and

$$\begin{aligned} & \left| \text{COV}\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i), \widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n) - \theta_C(X_j, Y_j)\right) \right| \\ &= O\left(n^{-\frac{p+q+3/2}{p+q-1/2}}(\log n)^{\frac{2p+2q+3}{p+q-1/2}}\right) = o(n^{-1}). \end{aligned} \quad (3.16)$$

The value of the constant C_0 depends on f and on the structure of the set $\mathcal{D} \subset \Psi$, and may be inferred from the proof of the theorem.

Remark 3.1. *Due to Lemmas 2.1 and 3.1 it is straightforward to extend the results of Theorems 3.1 and 3.2 to the output oriented measures ζ_C , the hyperbolic version γ_C and its logarithm $\log \gamma_C$, under Assumptions 2.1–2.4, 3.1 and 3.2. The hyperbolic measure and its logarithm are useful for analyzing Malmquist indices of productivity change when technologies are possibly non-convex.*

The results obtained above are similar to those obtained by Park et al. (2000) and Kneip et al. (2015) for the usual FDH estimator, but the analysis is different here. The original estimator converges at rate $n^{1/\kappa}$ where $\kappa = p + q$, whereas under the assumptions of Theorem 3.1, the estimator $\theta_C(x, y \mid \mathcal{X}_n)$ converges at rate $n^{1/\kappa}$ with $\kappa = p + q - 1/2$, and this value appears in (3.11)–(3.16).

4 The CRS Case

As noted above, under (globally) CRS the equality in (3.8) holds for all $\alpha > 1$. In this case, instead of (3.9) and (3.10), we have that for any $(x, y) \in \Psi$ there exists some (non unique) $a_{\min}^{x_0, y_0} > 0$ such that

$$(\theta(x_0, a_{\min}^{x_0, y_0} y_0)x_0, a_{\min}^{x_0, y_0} y_0) \in \Psi \quad (4.1)$$

and

$$(\alpha \cdot \theta(x_0, a_{\min}^{x_0, y_0} y_0)x_0, \alpha \cdot a_{\min}^{x_0, y_0} y_0) = (\theta(x_0, \alpha \cdot a_{\min}^{x_0, y_0} y_0)x_0, \alpha \cdot a_{\min}^{x_0, y_0} y_0) \in \Psi \quad (4.2)$$

for all $\alpha > 0$. Hence, if the frontier exhibits CRS everywhere, Assumption 3.2 must be replaced by the following assumption.

Assumption 4.1. (i) For some $d > 0$ and all (x_0, y_0) in the interior of $\mathcal{D}$, $\exists$ some $a_{\min}^{x_0, y_0} > 0$ such that (4.2) is satisfied for all $\alpha > 0$, while $\{((1 + \delta) \cdot \theta(x_0, a_{\min}^{x_0, y_0} y_0)x_0, (1 + \delta) \cdot a_{\min}^{x_0, y_0} y_0) \mid \delta \in [-d, d]\}$ is a subset of the interior of $\mathcal{D}$; (ii) $\mathcal{D}^* := \{(\theta(x_0, y_0)x_0, y_0) \mid (x_0, y_0) \in \mathcal{D}\}$ is a compact subset of $\mathcal{D}$; (iii) $f(x, y) > 0$ for all $(x, y) \in \mathcal{D}^*$; and (iv) $\exists$ a constant $0 < M < \infty$ such that $\|x\| \leq M$ for all $(x, y) \in \mathcal{D}$.

We then obtain the following theorems, where the rates of convergence are improved again, here with $\kappa = p + q - 1$. We gain one dimension with respect to the usual FDH estimators, and 1/2 dimension compared with the estimator of the distance to the boundary of $\mathcal{C}(\Psi)$ when non-CRS prevails.

Theorem 4.1. Under Assumptions 2.1–2.4, 3.1 and 4.1, given an arbitrary point (x_0, y_0) in the interior of $\mathcal{D}$, there exists a parameter $\tilde{\lambda}(x_0, y_0) > 0$ such that for all $z > 0$

$$\Pr \left(n^{\frac{1}{p+q-1}} \left(\frac{\hat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n)}{\theta_C(x_0, y_0)} - 1 \right) \leq z \right) = 1 - \exp(-\tilde{\lambda}(x_0, y_0)z^{p+q-1}) + o(1), \quad (4.3)$$

i.e., $\frac{\widehat{\theta}_C(x, y | \mathcal{X}_n)}{\theta_C(x, y)} - 1$ follows a Weibull distribution with parameters $\widetilde{\lambda}(x_0, y_0)$ and $p+q-1$, where $\widetilde{\lambda}(x_0, y_0)$ depends on (x_0, y_0) , g and f . The analytical structure of $\widetilde{\lambda}(x_0, y_0)$ may be inferred from the proof of the theorem.

Moreover,

$$E\left(\widehat{\theta}_C(x_0, y_0 | \mathcal{X}_n)\right) = \theta_C(x_0, y_0) + \theta_C(x_0, y_0) \widetilde{\lambda}(x_0, y_0)^{-\frac{1}{\kappa}} \Gamma\left(1 + \frac{1}{\kappa}\right) n^{-1/\kappa} + O(n^{-2/\kappa}) \quad (4.4)$$

and

$$\text{VAR}\left(\widehat{\theta}_C(x_0, y_0 | \mathcal{X}_n)\right) = \theta_C(x_0, y_0)^2 \widetilde{\lambda}(x_0, y_0)^{-\frac{2}{\kappa}} \left(\Gamma\left(1 + \frac{2}{\kappa}\right) - \Gamma\left(1 + \frac{1}{\kappa}\right)^2\right) n^{-2/\kappa} + o(n^{-2/\kappa}) \quad (4.5)$$

where $\kappa = p + q - 1$.

A second result concerns moments of $\widehat{\theta}_C(X_i, Y_i | \mathcal{X}_n)$ under the CRS assumption.

Theorem 4.2. *Under the assumptions of Theorem 4.1 there exists a constant $0 < \widetilde{C}_0 < \infty$ such that for all $i, j \in \{1, \dots, n\}$, $i \neq j$,*

$$E\left(\widehat{\theta}_C(X_i, Y_i | \mathcal{X}_n) - \theta_C(X_i, Y_i)\right) = \widetilde{C}_0 n^{-\frac{1}{p+q-1}} + O\left(n^{-\frac{2}{p+q-1}} (\log n)^{\frac{4}{p+q-1}}\right). \quad (4.6)$$

Furthermore,

$$\text{VAR}\left(\widehat{\theta}_C(X_i, Y_i | \mathcal{X}_n) - \theta_C(X_i, Y_i)\right) = O\left(n^{-\frac{2}{p+q-1}}\right) \quad (4.7)$$

and

$$\begin{aligned} & \left| \text{COV}\left(\widehat{\theta}_C(X_i, Y_i | \mathcal{X}_n) - \theta_C(X_i, Y_i), \widehat{\theta}_C(X_j, Y_j | \mathcal{X}_n) - \theta_C(X_j, Y_j)\right) \right| \\ &= O\left(n^{-\frac{p+q+1}{p+q-1}} (\log n)^{\frac{2p+2q+2}{p+q-1}}\right) = o(n^{-1}). \end{aligned} \quad (4.8)$$

The value of the constant $\widetilde{C}_0$ depends on f and on the structure of the set $\mathcal{D} \subset \Psi$ and may be inferred from the proof of the theorem.

Remark 4.1. Remark 3.1 applies here too under the CRS assumption. So Theorems 4.1 and 4.2 extend to the output-oriented measures ζ_C , the hyperbolic version γ_C and its logarithm $\log \gamma_C$, under Assumptions 2.1–2.4, 3.1 and 4.1.

5 Central Limit Theorems

5.1 Basic Theorems

In order to make inference about returns to scale and about Malmquist indices of productivity for possibly non-convex technologies, new CLT results are needed for averages of $\hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$, both under CRS and non-CRS. The results established here are used below in Section 6.1 to develop a test of CRS versus non-CRS when technologies are not convex, and in Section 6.2 to make inference about Malmquist indices of productivity change when technologies are not convex. We use the results on moments obtained above in Theorems 3.2 and 4.2 to derive these new CLTs.

To summarize, the two theorems developed in this section are analogs of Theorem 3.2 in Kneip et al. (2015) for the usual FDH estimator of efficiency with respect to the boundary of Ψ . The only differences in the two theorems developed here involve the set of appropriate assumptions, the values of the constants appearing in the asymptotic moments and the rates of convergence. For the three cases we have the rate $n^{1/\kappa}$, where (i) $\kappa = p + q$ for the usual FDH estimator $\hat{\theta}(X, Y \mid \mathcal{X}_n)$ of the distances to the boundary of Ψ , (ii) $\kappa = p + q - 0.5$ for the CFDH estimator $\hat{\theta}_C(X, Y \mid \mathcal{X}_n)$ of the distance to the boundary of $\mathcal{C}(\Psi)$, the cone spanned by Ψ under non-CRS and (iii) $\kappa = p + q - 1$ for the CFDH estimators $\hat{\theta}_C(X, Y \mid \mathcal{X}_n)$ of the distance to the boundary of $\mathcal{C}(\Psi) \equiv \Psi$, under the global CRS assumption.

To begin, define

$$\mu_{\theta_C} = \mathbb{E}(\theta_C(X, Y)), \quad (5.1)$$

$$\sigma_{\theta_C}^2 = \text{VAR}(\theta_C(X, Y)) \quad (5.2)$$

and

$$\hat{\mu}_{n, \theta_C} = n^{-1} \sum_{i=1}^n \hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n). \quad (5.3)$$

Note that if the *true* efficiencies $\theta_C(X_i, Y_i)$ were observed, one could compute $\tilde{\mu}_{n, \theta_C} = n^{-1} \sum_{i=1}^n \theta_C(X_i, Y_i)$ and use standard CLTs to make inference about μ_{θ_C} . But, the $\theta_C(X_i, Y_i)$ are not observed; instead, only estimates $\hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$ are observed. Analogous to Theorem 4.1 of Kneip et al. (2015), we have

Theorem 5.1. *Under the assumptions of Theorem 3.1 and with $\kappa = p + q - 1/2$ if CRS does not hold everywhere, or under the assumptions of Theorem 4.1 and with $\kappa = p + q - 1$ if CRS holds everywhere, as $n \rightarrow \infty$, (i) $\hat{\mu}_{n,\theta_C}$ is a consistent estimator of μ_{θ_C} , with a bias term of order $Cn^{-1/\kappa}$; (ii) $\hat{\sigma}_{n,\theta_C}^2 = n^{-1} \sum_{i=1}^n \left[\hat{\theta}_C(X_i, Y_i | \mathcal{X}_n) - \hat{\mu}_{n,\theta_C} \right]^2$ is a consistent estimator of $\sigma_{\theta_C}^2$; and (iii)*

$$\sqrt{n} (\hat{\mu}_{n,\theta_C} - \mu_{\theta_C} - Cn^{-1/\kappa} - R_{n,\kappa}) \xrightarrow{\mathcal{L}} N(0, \sigma_{\theta_C}^2), \quad (5.4)$$

where $R_{n,\kappa} = O(n^{-2/\kappa}(\log n)^{4/\kappa})$ and where κ and the constant C depend on the particular case (i.e., CRS or non-CRS for Ψ^∂).

Note that $R_{n,\kappa} = o(n^{-1/\kappa})$. Hence (5.4) makes clear that problems arise when κ is too large. If $\kappa < 2$, a simple CLT can be derived from (5.4), because $\sqrt{n}(Cn^{-1/\kappa} + R_{n,\kappa}) = o(1)$. However this occurs if and only if $p + q \leq 2$ for the CFDH estimator, either with CRS ($\kappa = p + q - 1$) or non-CRS technologies ($\kappa = p + q - 0.5$). But if $\kappa = 2$, the bias is a constant, and if $\kappa > 2$ the bias explodes. Consequently, standard CLT results (e.g., the Lindeberg-Feller CLT) can be used with sample means of CFDH estimates to make inference about mean efficiency only if $1/\kappa > 1/2$. Hence if CRS does not hold, standard CLT results can be used to make inference about mean efficiency only if $p + q < 2.5$, while under CRS, standard CLTs can be used only if $p + q < 3$. Hence in either case, standard CLT results are valid only if $(p + q)$ equals 1 or 2. Replacing the scale factor $\sqrt{n}$ with n^ξ , with $\xi < 1/\kappa \leq 1/2$ is not a viable option. Doing so would cause the variance to converge to zero whenever $\kappa \geq 2$, rendering inference impossible.

For the most common cases where $\kappa \geq 2$, the solution suggested by Kneip et al. (2015) is based on two operations, (i) providing an estimate of the leading term $Cn^{-1/\kappa}$ in the inherent bias in (5.4), and (ii) computing sample mean from a subsample of efficiency estimates to control the variance of the resulting average.

The bias correction, when $\kappa \geq 2$, is based on a generalized jackknife method. Kneip et al. (2015) show that a consistent estimator of $Cn^{-1/\kappa}$ is provided by

$$\tilde{B}_{n,\kappa} = (2^{1/\kappa} - 1)^{-1} (\bar{\mu}_{n/2,\theta_C}^* - \hat{\mu}_{n,\theta_C}), \quad (5.5)$$

where

$$\bar{\mu}_{n/2,\theta_C}^* = (\bar{\mu}_{n/2,\theta_C}^{(1)} + \bar{\mu}_{n/2,\theta_C}^{(2)})/2, \quad (5.6)$$

is a jackknife analog of $\widehat{\mu}_{n,\theta_C}$, $\overline{\mu}_{n/2,\theta_C}^{(1)}$ is a version of $\widehat{\mu}_{n,\theta_C}$ based on a random subset of $\mathcal{X}_n$ of size $n/2$ and $\overline{\mu}_{n/2,\theta_C}^{(2)}$ is the analog based on the remaining part of the sample (to simplify notation, assume that n is even). Formally, for $j \in \{1, 2\}$,

$$\overline{\mu}_{n/2,\theta_C}^{(j)} = 2n^{-1} \sum_{\{i|(X_i, Y_i) \in \mathcal{X}_{n/2}^{(j)}\}} \widehat{\theta}_C(X_i, Y_i | \mathcal{X}_{n/2}^{(j)}), \quad (5.7)$$

where $\mathcal{X}_{n/2}^{(1)} \cap \mathcal{X}_{n/2}^{(2)} = \emptyset$ and $\mathcal{X}_{n/2}^{(1)} \cup \mathcal{X}_{n/2}^{(2)} = \mathcal{X}_n$. Note that the efficiency estimates are computed with the limited reference samples $\mathcal{X}_{n/2}^{(j)}$, as denoted explicitly by the notation. Kneip et al. (2015) prove that $\widetilde{B}_{n,\kappa} = Cn^{-1/\kappa} + o(n^{-1/\kappa}) + o_p(n^{-1/2})$. Kneip et al. (2016) suggest repeating the operations above a large number L times, shuffling the observations each time, to obtain L estimates $\widetilde{B}_{n,\kappa,\ell}$, $\ell = 1 \dots, L$ of the bias and then computing

$$\widehat{B}_{n,\kappa} = L^{-1} \sum_{\ell=1}^L \widetilde{B}_{n,\kappa,\ell} \quad (5.8)$$

to reduce the variance of the bias estimate (in practice, $L = 20$ is often sufficient).

This bias estimate in (5.8) provides leading part of the bias, but the remainder $R_{n,\kappa} = o(n^{-1/\kappa})$ must also be controlled when $\kappa > 2$. Define $n_\kappa = \lfloor n^{2/\kappa} \wedge n \rfloor \leq n$ where $\lfloor a \rfloor$ denotes the largest integer less than or equal to a and $a \wedge b$ denotes the minimum of a and b . Then let $\mathcal{X}_{n_\kappa}^* \subseteq \mathcal{X}_n$ denote a random subsample of $\mathcal{X}_n$ of size n_κ . Now define the subsample version of $\widehat{\mu}_n$, i.e.

$$\widehat{\mu}_{n_\kappa,\theta_C} = n_\kappa^{-1} \sum_{\{j|(X_j, Y_j) \in \mathcal{X}_{n_\kappa}^*\}} \widehat{\theta}_C(X_j, Y_j | \mathcal{X}_n), \quad (5.9)$$

where the average is taken over the random subsample $\mathcal{X}_{n_\kappa}^*$ but the estimates of efficiency for the observations in $\mathcal{X}_{n_\kappa}^*$ are computed using the full sample $\mathcal{X}_n$ as made explicit in the notation. The properties of $R_{n,\kappa}$ in Theorem 5.1 then lead to the following CLT for the conical version of the FDH estimator.

Theorem 5.2. *Under the assumptions of Theorem 4.1 for the CRS case, or of Theorem 3.1 for the non-CRS case, for the CFDH estimator, $p + q \leq 4$ and either CRS or non-CRS technology,*

$$\sqrt{n} \left(\widehat{\mu}_{n,\theta_C} - \mu_{\theta_C} - \widehat{B}_{n,\kappa} - R_{n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, \sigma_{\theta_C}^2) \quad (5.10)$$

as $n \rightarrow \infty$ with $R_{n,\kappa} = o(n^{-1/2})$. In addition, if $\kappa > 2$, then

$$\sqrt{n_\kappa} \left(\widehat{\mu}_{n_\kappa,\theta_C} - \mu_{\theta_C} - \widehat{B}_{n,\kappa} - R_{n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, \sigma_{\theta_C}^2) \quad (5.11)$$

as $n \rightarrow \infty$ with $R_{n,\kappa} = o(n^{-1/\kappa})$.

5.2 Alternative Bias Correction for the Conical FDH Estimators

It is possible to define a bias-corrected estimator (more precisely, the first order bias of the mean vanishes) for the estimator of $\theta_C(x_0, y_0)$ in the two cases analyzed in the preceding sections. We first consider the case of Section 3, where $\theta_C(x_0, y_0)$ is the distance of (x_0, y_0) to the boundary of $\mathcal{C}(\Psi)$.

Consider a point (x_0, y_0) in the interior of $\mathcal{D}$. There then exists an index $i_{max} \in \{1, \dots, n\}$ such that

$$i_{max} := \operatorname{argmin}_{i=1, \dots, n} \frac{\max_{j=1, \dots, p} \frac{X_{ij}}{x_{0j}}}{\min_{j=1, \dots, q} \frac{Y_{ij}}{y_{0j}}} \quad (5.12)$$

and $\hat{\theta}_C(x_0, y_0) = \left(\max_{j=1, \dots, p} \frac{X_{i_{max},j}}{x_{0j}} \right) \left(\min_{j=1, \dots, q} \frac{Y_{i_{max},j}}{y_{0j}} \right)^{-1}$. Then define the estimator

$$\hat{\theta}_C^*(x_0, y_0) := \min_{i=1, \dots, n; i \neq i_{max}} \frac{\max_{j=1, \dots, p} \frac{X_{ij}}{x_{0j}}}{\min_{j=1, \dots, q} \frac{Y_{ij}}{y_{0j}}}, \quad (5.13)$$

noting that $\hat{\theta}_C^*(x_0, y_0)$ uses the **second** smallest observation to estimate $\theta_C(x_0, y_0)$. Necessarily, $\hat{\theta}_C^*(x_0, y_0) > \hat{\theta}_C(x_0, y_0)$.

Proposition 3.1 generalizes to this alternative estimator in the following sense: $\frac{\hat{\theta}_C^*(x_0, y_0)}{\theta_C(x_0, y_0)} \leq 1 + \delta$ if and only if there are at least **two** observations in $S_\delta^{x_0, y_0}$. Following the arguments used to prove Theorem 3.1 we then have

$$\Pr \left(n^{\frac{1}{p+q-1/2}} \left(\frac{\hat{\theta}_C^*(x_0, y_0 \mid \mathcal{X}_n)}{\theta_C(x_0, y_0)} - 1 \right) \leq z \right) = 1 - (1 + \lambda(x_0, y_0) z^{p+q-1/2}) \exp(-\lambda(x_0, y_0) z^{p+q-1/2}) + o(1), \quad (5.14)$$

and with $\kappa = p + q - 1/2$, and

$$E \left(\hat{\theta}_C^*(x_0, y_0 \mid \mathcal{X}_n) \right) = \theta_C(x_0, y_0) + \theta_C(x_0, y_0) \lambda(x_0, y_0)^{-\frac{1}{\kappa}} \Gamma(2 + \frac{1}{\kappa}) n^{-\frac{1}{\kappa}} + O(n^{-\frac{2}{\kappa}}), \quad (5.15)$$

while of course the variance of the estimator is still of order $O(n^{-\frac{2}{\kappa}})$. Based on these results,

one may define the bias-corrected estimator

$$\begin{aligned}\tilde{\theta}_C(x_0, y_0 \mid \mathcal{X}_n) &:= \hat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n) - \frac{\Gamma(1 + \frac{1}{\kappa})}{\Gamma(2 + \frac{1}{\kappa}) - \Gamma(1 + \frac{1}{\kappa})} (\hat{\theta}_C^*(x_0, y_0 \mid \mathcal{X}_n) - \hat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n)) \\ &= \hat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n) - \kappa (\hat{\theta}_C^*(x_0, y_0 \mid \mathcal{X}_n) - \hat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n)).\end{aligned}\quad (5.16)$$

For the mean of the bias-corrected estimator we obtain

$$E \left(\tilde{\theta}_C(x_0, y_0 \mid \mathcal{X}_n) \right) = \theta_C(x_0, y_0) + O(n^{-\frac{2}{\kappa}}), \quad (5.17)$$

and the analog of Theorem 3.2 becomes

$$E \left(\tilde{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i) \right) = O \left(n^{-\frac{2}{p+q-1/2}} (\log n)^{\frac{4}{p+q-1/2}} \right), \quad (5.18)$$

$$\text{VAR} \left(\tilde{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i) \right) = O \left(n^{-\frac{2}{p+q-1/2}} \right), \quad (5.19)$$

and

$$\begin{aligned}\left| \text{COV} \left(\tilde{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i), \tilde{\theta}_C(X_j, Y_j \mid \mathcal{X}_n) - \theta_C(X_j, Y_j) \right) \right| \\ = O \left(n^{-\frac{p+q+3/2}{p+q-1/2}} (\log n)^{\frac{2p+2q+3}{p+q-1/2}} \right) = o(n^{-1}).\end{aligned}\quad (5.20)$$

Following the proofs of Theorems 3.1, and 3.2, it is straightforward to prove the properties of $\tilde{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$ listed above. The same modification can, of course, be done in the CRS case where only the value of κ has to be adjusted as in Theorems 4.1 and 4.2.

5.3 Confidence Intervals for μ_{θ_C}

Using Theorem 5.2 and the consistent estimator of $\sigma_{\theta_C}^2$ in (ii) of Theorem 5.1, an asymptotically correct $(1 - \alpha)$ confidence interval for μ_{θ_C} is given by

$$\left[\hat{\mu}_{n, \theta_C} - \hat{B}_{n, \kappa} \pm z_{1-\alpha/2} \hat{\sigma}_{n, \theta_C} / \sqrt{n} \right] \quad (5.21)$$

when $p + q \leq 4$, where $\hat{B}_{n, \kappa}$ is the generalized jackknife estimate of bias appearing in (5.8). Alternatively, when $\kappa > 2$,

$$\left[\hat{\mu}_{n_\kappa, \theta_C} - \hat{B}_{n, \kappa} \pm z_{1-\alpha/2} \hat{\sigma}_{n, \theta_C} / \sqrt{n_\kappa} \right], \quad (5.22)$$

gives an asymptotically valid $(1 - \alpha)$ confidence interval for μ_{θ_C} .

Note that either (5.10) or (5.11) in Theorem 5.2 can be used when $p + q = 4$ in the CRS case, or when $p + q = 3$ or 4 in the non-CRS case. However, as explained in Kneip et al. (2015, 2021), due to $R_{n,\kappa}$, (5.10) involves $n^{1/2}R_{n,\kappa}$ while (5.11) involves a term $n^{1/\kappa}R_{n,\kappa}$. These are neglected in (5.21) and (5.22), and since $n^{1/\kappa}R_{n,\kappa}$ is of smaller order than $n^{1/2}R_{n,\kappa}$, one might expect the interval in (5.22) to provide better coverage than the interval in (5.21) in cases where either can be used.

Below, in Section 7, we examine via Monte Carlo experiments the coverages of the confidence intervals in (5.21)–(5.22) with bias terms estimated by the generalized jackknife discussed in Section 5 as well as the alternative estimator used in (5.16). But first, we show in the next section how the results obtained so far can be used to test CRS versus non-CRS, and how inference can be made about changes in productivity using Malmquist indices, when production sets are possibly not convex.

6 Applications

6.1 Testing CRS in General Technologies

6.1.1 Test Statistics

The CLTs for sample means of efficiency measures have been used in Kneip et al. (2016) and Daraio et al. (2018) to develop statistical tests of various model features. These tests involve comparing a sample mean of efficiency estimates that impose the conditions of the null hypothesis against a sample mean of efficiency estimates where the null is not imposed. All of the tests require the two sample means under comparison to be independent of each other, and this requires randomly splitting the original sample into two independent parts. Simar and Wilson (2020) show how the ambiguity resulting from a single random split can be removed by splitting multiple times and then using bootstrap methods to implement tests based on sample means of the test statistics from each split, or based on the p -values obtained on each split.

We wish to test the null hypothesis $H_0: \Psi^\partial$ is CRS versus the alternative hypothesis $H_1: \Psi^\partial$ is not CRS without imposing convexity on Ψ as required in Kneip et al. (2016). The main idea is to compare sample means of FDH and CFDH estimates; under H_0 , the difference is expected to be “small,” whereas if the null is false, the difference is expected

to be “large.” In order to obtain a limiting distribution for the difference, we require the sample means to be independent, and this requires randomly splitting the sample $\mathcal{X}_n$ into two independent parts as in Kneip et al. (2016).

Under the null hypothesis of CRS, the FDH and CFDH estimators defined in (3.2) and (3.3) converge at rates $n^{1/(p+q)}$ and $n^{1/(p+q-1)}$, respectively. Let $\kappa_1 = p+q$ and $\kappa_2 = p+q-1$. To implement the test, randomly split the original, full sample $\mathcal{X}_n$ into two sub-samples $\mathcal{X}_{1,n_1}$, $\mathcal{X}_{2,n_2}$ of sizes n_1 and n_2 such that $\mathcal{X}_{1,n_1} \cup \mathcal{X}_{2,n_2} = \mathcal{X}_n$ and $\mathcal{X}_{1,n_1} \cap \mathcal{X}_{2,n_2} = \emptyset$, where $n_1 + n_2 = n$ and $n_1^{1/\kappa_1} = n_2^{1/\kappa_2}$. Next, compute the FDH estimates $\hat{\theta}(X_i, Y_i \mid \mathcal{X}_{1,n_1})$ at each observation $(X_i, Y_i) \in \mathcal{X}_{1,n_1}$, and compute the CFDH estimates $\hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_{2,n_2})$ at each observation $(X_i, Y_i) \in \mathcal{X}_{2,n_2}$. Then compute the sample means

$$\hat{\mu}_{\text{FDH},n_1} = n_1^{-1} \sum_{(X_i, Y_i) \in \mathcal{X}_{1,n_1}} \hat{\theta}(X_i, Y_i \mid \mathcal{X}_{1,n_1}) \quad (6.1)$$

and

$$\hat{\mu}_{\text{CFDH},n_2} = n_2^{-1} \sum_{(X_i, Y_i) \in \mathcal{X}_{2,n_2}} \hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_{2,n_2}). \quad (6.2)$$

By construction, $\hat{\mu}_{\text{FDH},n_1}$ and $\hat{\mu}_{\text{CFDH},n_2}$ are independent.

Now let $\hat{B}_{\text{FDH},n,\kappa_1}$ be the jackknife estimate of the bias $\hat{\mu}_{\text{FDH},n_1}$ (computed using L replications), and let $\hat{B}_{\text{CFDH},n,\kappa_2}$ be the estimate of the bias of $\hat{\mu}_{\text{CFDH},n_2}$ computed either by the jackknife method or by the new estimator in (5.13). Consistent estimates of the variance in both cases are given by

$$\hat{\sigma}_{\text{FDH},n_1}^2 = n_1^{-1} \sum_{(X_i, Y_i) \in \mathcal{X}_{1,n_1}} \left[\hat{\theta}(X_i, Y_i \mid \mathcal{X}_{1,n_1}) - \hat{\mu}_{\text{FDH},n_1} \right]^2 \quad (6.3)$$

and

$$\hat{\sigma}_{\text{CFDH},n_2}^2 = n_2^{-1} \sum_{(X_i, Y_i) \in \mathcal{X}_{2,n_2}} \left[\hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_{2,n_2}) - \hat{\mu}_{\text{CFDH},n_2} \right]^2. \quad (6.4)$$

Then under the null hypothesis of CRS and under assumptions of Theorem 4.1,

$$\hat{\tau}_{1,n} = \frac{(\hat{\mu}_{\text{FDH},n_1} - \hat{\mu}_{\text{CFDH},n_2}) - (\hat{B}_{\text{FDH},\kappa_1,n_1} - \hat{B}_{\text{CFDH},\kappa_2,n_2})}{\sqrt{\frac{\hat{\sigma}_{\text{FDH},n_1}^2}{n_1} + \frac{\hat{\sigma}_{\text{CFDH},n_2}^2}{n_2}}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (6.5)$$

as $n \rightarrow \infty$ provided $(p + q) \leq 3$, since the FDH convergence rate dominates that of the CFDH estimator.

Alternatively, if $\kappa_1 > 2$, the sample means must be computed using subsets of $\mathcal{X}_{1,n_1}$ and $\mathcal{X}_{2,n_2}$. For $\ell \in \{1, 2\}$, let $n_{\ell,\kappa_\ell} = \lfloor n_\ell^{2/\kappa_\ell} \rfloor$ and let $\mathcal{X}_{\ell,n_{\ell,\kappa_\ell}}^*$ be a random subset of size n_{ℓ,κ_ℓ} from $\mathcal{X}_{\ell,n_\ell}$. Compute the means

$$\hat{\mu}_{\text{FDH},n_1,\kappa_1} = n_{1,\kappa_1}^{-1} \sum_{(X_i,Y_i) \in \mathcal{X}_{1,n_1,\kappa_1}^*} \hat{\theta}(X_i, Y_i \mid \mathcal{X}_{1,n_1}) \quad (6.6)$$

and

$$\hat{\mu}_{\text{CFDH},n_2,\kappa_2} = n_{2,\kappa_2}^{-1} \sum_{(X_i,Y_i) \in \mathcal{X}_{2,n_2,\kappa_2}^*} \hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_{2,n_2}), \quad (6.7)$$

where it is clear from the notation that the full subsets $\mathcal{X}_{1,n_1}$ and $\mathcal{X}_{2,n_2}$ are used to compute the efficiency estimates in (6.6)–(6.7), but the summations are over subsets of the observations. Then under the null hypothesis of CRS and the assumptions of Theorem th:asymCHCRS,

$$\hat{\tau}_{2,n} = \frac{(\hat{\mu}_{\text{FDH},n_1,\kappa_1} - \hat{\mu}_{\text{CFDH},n_2,\kappa_2}) - (\hat{B}_{\text{FDH},\kappa_1,n_1} - \hat{B}_{\text{CFDH},\kappa_2,n_2})}{\sqrt{\frac{\hat{\sigma}_{\text{FDH},n_1}^2}{n_{1,\kappa_1}} + \frac{\hat{\sigma}_{\text{CFDH},n_2}^2}{n_{2,\kappa_2}}}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (6.8)$$

as $n \rightarrow \infty$. Depending on the value of $(p + q)$, either $\hat{\tau}_{1,n}$ or $\hat{\tau}_{2,n}$ (but not both, except when $p + q = 3$) can be used to test the null hypothesis of CRS versus non-CRS, with critical values obtained from the standard normal distribution. In particular, for $j \in \{1, 2\}$, the null hypothesis of CRS is rejected if $\hat{p} = 1 - \Phi(\hat{\tau}_{j,n})$ is less than α for test with asymptotic size α . If $(p + q) = 3$, either of the statistics $\hat{\tau}_{1,n}$ or $\hat{\tau}_{2,n}$ can be used. In this case, one should expect $\hat{\tau}_{2,n}$ to perform better in finite samples for reasons similar to those given above at the end of Section 5.

6.1.2 Consistency Issues

Unlike Kneip et al. (2016), here we are able to prove consistency of the proposed tests of CRS versus non-CRS since we can derive the distribution of the test statistics under the alternative.⁴ Define $\mu_\theta = \mathbb{E}(\theta(X, Y))$ and $\xi = \mu_\theta - \mu_{\theta_C} \geq 0$. Then the null hypothesis of

⁴Note that due to the more recent result in Theorem 3.1 in Kneip et al. (2021), it is easy to prove consistency of test of CRS versus non-CRS proposed in Kneip et al. (2016), for cases where Ψ is convex, by following the arguments developed here.

CRS is equivalent to $H_0: \xi = 0$, and the alternative hypothesis is equivalent to $H_1: \xi > 0$. Consider the case where $p + q \geq 3$ and use the statistic $\widehat{\tau}_{2,n}$ defined in (6.8) to test the null against the alternative.

Under the alternative hypothesis, the estimator $\widehat{\mu}_{\text{CFDH},n_2,\kappa_2}$ does not have the same properties as it does under the null; in particular, the rates of convergence and the bias are different under H_1 . However, from Theorem 5.1, it is easy to demonstrate that under H_1 ,

$$\sqrt{n_{2,\kappa_2}} \left(\widehat{\mu}_{n_2,\kappa_2,\theta_C} - \mu_{\theta_C} - Cn_2^{-1/\kappa_3} - R_{n_2,\kappa_3} \right) \xrightarrow{\mathcal{L}} N(0, \sigma_{\theta_C}^2), \quad (6.9)$$

as $n \rightarrow \infty$, where $R_{n_2,\kappa_3} = O\left(n_2^{-2/\kappa_3}(\log n_2)^{4/\kappa_3}\right)$ and $\kappa_3 = p + q - 0.5 > \kappa_2 = p + q - 1$. It is easy to see that $R_{n_2,\kappa_3} = o\left(n_2^{-1/\kappa_2}\right)$, and hence this reminder term can be neglected.

A problem may arise for the bias term, but this is easily solved. In defining $\widehat{\tau}_{2,n}$ in (6.8), the bias of the CFDH mean is estimated by $\widehat{B}_{\text{CFDH},\kappa_2,n_2}$, but this is not appropriate under the alternative. However, it is easy to prove that

$$\widehat{B}_{\text{CFDH},\kappa_2,n_2} = \widehat{B}_{\text{CFDH},\kappa_3,n_2} - an_2^{-1/\kappa_3}C, \quad (6.10)$$

where a is a positive constant, with $a = (\kappa_3 - \kappa_2)/\kappa_3$ if the new bias correction in equation (5.16) is used or $a = (2^{1/\kappa_2} - 2^{1/\kappa_3})/(2^{1/\kappa_2} - 1)$ if the jackknife bias correction is used. Here, $\widehat{B}_{\text{CFDH},\kappa_3,n_2}$ provides the appropriate bias correction.

Similar to the argument leading to the distribution of $\widehat{\tau}_{2,n}$ under the null, we obtain here when $\xi > 0$

$$\widehat{\tau}_{2,n} = \frac{\left(\widehat{\mu}_{\text{FDH},n_1,\kappa_1} - \widehat{\mu}_{\text{CFDH},n_2,\kappa_2} \right) - \left(\widehat{B}_{\text{FDH},\kappa_1,n_1} - \widehat{B}_{\text{CFDH},\kappa_3,n_2} \right) - \xi - an_2^{-1/\kappa_3}C}{\sqrt{\frac{\widehat{\sigma}_{\text{FDH},n_1}^2}{n_{1,\kappa_1}} + \frac{\widehat{\sigma}_{\text{CFDH},n_2}^2}{n_{2,\kappa_2}}}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (6.11)$$

as $n \rightarrow \infty$. From (6.11) it is easy to see that

$$\lim_{n \rightarrow \infty} \Pr(\widehat{\tau}_{2,n} > z_\alpha) = 1 - \lim_{n \rightarrow \infty} \Phi \left(z_\alpha - \frac{\xi + an_2^{-1/\kappa_3}C}{\sqrt{\frac{\widehat{\sigma}_{\text{FDH},n_1}^2}{n_{1,\kappa_1}} + \frac{\widehat{\sigma}_{\text{CFDH},n_2}^2}{n_{2,\kappa_2}}}} \right) \quad (6.12)$$

as $n \rightarrow \infty$. Clearly, the denominator of the fraction inside the parentheses on the right-hand side (RHS) of (6.12) converges to zero at the rate n_1^{-1/κ_1} as $n \rightarrow \infty$. Hence the limit of

the RHS of (6.12) is equal to zero, and therefore the test is consistent. A similar argument applies when $p + q = 2$ and the test statistics $\widehat{\tau}_{1,n}$ is used. It is also easy to show that the test is still consistent for asymptotic local alternatives $H_{1,n} : \xi = bn^{-\rho}$ for some constant b as soon as $0 < \rho < 1/\kappa_1$.

6.1.3 Multiple Random Sample Splits

All of the tests developed here, as well as those developed in Kneip et al. (2016), involve comparing two sample means of efficiency estimates. To avoid complicated covariances that cannot be estimated, the two sample means under comparison must be independent of each other, and this requires randomly splitting the original sample. Kneip et al. (2016) show that if one does not split the original sample and instead applies the two estimators to all of the n observations in $\mathcal{X}_n$, and then builds a statistic similar to (6.5) or (6.8) but using sample means over all n observations, the resulting covariances wipe out the variances, resulting in a test statistic with a degenerate distribution for any value of $(p + q)$. The random split into two parts avoids this problem, but the test results depend on an arbitrary random split of the original sample. Simar and Wilson (2020) have shown how this ambiguity can be removed. To conserve space we summarize the main ideas below and refer readers to Simar and Wilson (2020) for additional details.

Given the random sample $\mathcal{X}_n$, one can randomly split the sample once to obtain a value of, e.g. the statistic in (6.5) or (6.8), depending the value of $p + q$. Shuffling the observations and splitting again yields a second realization of the statistic. Suppose the researcher shuffles and then splits the sample S times, obtaining a set $\mathcal{T} = \{\widehat{\tau}_{\ell,n,s}\}_{s=1}^S$, of S different values of the test statistic, either for $\ell = 1$ or 2 as appropriate depending on the value of $(p + q)$. Corresponding to each $\widehat{\tau}_{\ell,n,s}$ is a realized p -value $\widehat{p}_s$, so that the researcher also obtains a set $\mathcal{P} = \{\widehat{p}_s\}_{s=1}^S$ of p -values after S sample-splits. If the statistics in $\mathcal{T}$ were independent, then by the probability integral transform, the p -values in $\mathcal{P}$ would be independent and uniformly distributed on $(0, 1)$. However, the realized values of the test statistic over S splits contained in $\mathcal{T}$ cannot be independent.

Simar and Wilson (2020) suggest two test statistics based on results from S sample-splits, as well as a bootstrap algorithm for evaluating their p -values. The test statistics are defined

as

$$\widehat{T}_n := S^{-1} \sum_{s=1}^S \widehat{\tau}_{\ell,n,s} \quad (6.13)$$

and

$$\widehat{K}_n := \sup_{u \in [0,1]} \left(\widehat{F}_{S,\widehat{p},n}(u) - u \right), \quad (6.14)$$

where $\widehat{\tau}_{\ell,n,s}$, $\ell \in \{1, 2\}$ denotes the appropriate test statistic given by (6.5) or (6.8) (depending on the value of $p + q$) obtained with the s th sample-split and $\widehat{F}_{S,\widehat{p},n}(u)$ is the empirical distribution function of the S values in $\mathcal{P}$. The statistic $\widehat{K}_n$ is a one-sided, one-sample Kolmogorov-Smirnov statistic, but its distribution is unknown due to the dependence induced by splitting repeatedly the same sample. The distribution of $\widehat{T}_n$ is also unknown, for the same reason.

Clearly, one should reject the null for “large” values of these statistics. Simar and Wilson (2020) describe an easy-to-implement bootstrap algorithm to approximate the p -values corresponding to the test statistics in (6.13) and (6.14), where the bootstrap samples are generated under the null hypothesis of CRS. Since many of the CFDH estimates are likely equal to 1, we modify the bootstrap described by Simar and Wilson (2020) to use the smooth bootstrap described by Simar and Wilson (1998). This involves computing n estimates $\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$ before splitting the sample, reflecting these estimates around 1, and drawing from a kernel density estimate obtained from the original and reflected estimates. See Simar and Wilson (1998) for details and discussion.

In practice, the number of splits should be as large as possible given the computational burden. Experience suggests that in practice, $S = 100$ splits is reasonable, with 1,000 bootstrap replications. The Monte-Carlo simulations for testing CRS with convex Ψ reported by Simar and Wilson (2020) indicate that tests based on either (6.13) or (6.14) perform well in finite samples, reaching the appropriate size with good power behavior as sample sizes increase. The test based on the Kolmogorov-Smirnov statistic often provides slightly better results.

6.2 Malmquist Indices

The Malmquist productivity index is widely used in empirical work to measure changes in productivity in dynamic settings (see Kneip et al., 2021, Appendix A, for a detailed discussion of the literature). Malmquist indices are based on the quantity index of Malmquist (1953) with adaptation to the production framework by Caves et al. (1982), and are often estimated using nonparametric DEA estimators which require the production sets at each time period to be convex. Kneip et al. (2021) develop statistical tools allowing asymptotic inference from DEA estimators of Malmquist indices for individual producers as well as for geometric means of these estimators.⁵ It is not surprising that the CLTs derived above for non-convex conical FDH estimators for potentially non-convex technologies can be used to generalize the results of Kneip et al. (2021) generalizes to the framework here where production sets are possibly non-convex. As seen above, the value of κ for the CFDH estimator depends on whether the technology is CRS or non-CRS. We summarize below the main results for the more general case of non-CRS technologies, where $\kappa = p + q - 1/2$.

In dynamic settings, the set of feasible combinations of input and output quantities at time t is given by

$$\Psi^t := \{(x, y) \mid x \text{ can produce } y \text{ at time } t\}. \quad (6.15)$$

Then the efficient frontier of Ψ^t is

$$\Psi^{t\partial} := \{(x, y) \mid (x, y) \in \Psi^t, (\gamma x, \gamma^{-1}y) \notin \Psi^t \forall \gamma \in (0, 1)\}. \quad (6.16)$$

Assume that Assumptions 2.1–2.4, 3.1 and 3.2 are satisfied for each time period t . As explained in Kneip et al. (2021), in order to ensure that Malmquist indices are well-defined, $\mathcal{D}^t$ and $\mathcal{D}^s$ must “fit together” for different periods t and s . Let

$$\mathcal{D}_{norm}^t := \left\{ \left(\frac{x}{\|x\|}, \frac{y}{\|y\|} \right) \mid (x, y) \in \mathcal{D}^t \right\}. \quad (6.17)$$

If $p + q = 2$ then trivially $\mathcal{D}_{norm}^t = \{(1, 1)\}$. But when $p + q > 2$, $\mathcal{D}_{norm}^t$ will quantify the set of all possible “directions” of vectors x and y where it is possible to define a frontier at time t .

⁵Geometric means of Malmquist index estimates, rather than arithmetic means, are typically reported in applied work due to the multiplicative nature of the Malmquist index.

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

$$\mathcal{M}_i := \left(\frac{\gamma(X_i^2, Y_i^2 \mid \mathcal{C}(\Psi^1))}{\gamma(X_i^1, Y_i^1 \mid \mathcal{C}(\Psi^1))} \times \frac{\gamma(X_i^2, Y_i^2 \mid \mathcal{C}(\Psi^2))}{\gamma(X_i^1, Y_i^1 \mid \mathcal{C}(\Psi^2))} \right)^{1/2}, \quad (6.18)$$

where we keep the same notation as in Lemma 2.1 to be explicit. 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 changes (increases, remains unchanged or decreases) between periods 1 and 2. Note that we define the Malmquist index in terms of hyperbolic distances to avoid issues of existence and numerical difficulties when the index is decomposed into sources of productivity change (see Simar and Wilson, 2019 for discussion and details).

Suppose that for two different time periods $t \in \{1, 2\}$ we have the set $\mathcal{X}_n^{1,2} = \{(X_i^1, Y_i^1), (X_i^2, Y_i^2)\}_{i=1}^n$ 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, let $\mathcal{X}_{n_1}^1 = \{(X_i^1, Y_i^1)\}_{i=1}^{n_1}$ be the set of $n_1 \geq n$ observations in period 1 available to estimate the hyperbolic distance $\gamma_C^1(x, y) := \gamma(x, y \mid \mathcal{C}(\Psi^1))$, and let $\mathcal{X}_{n_2}^2 = \{(X_i^2, Y_i^2)\}_{i=1}^{n_2}$ be the set of $n_2 \geq n$ observations in period 2 available to estimate the hyperbolic distance $\gamma_C^2(x, y) := \gamma(x, y \mid \mathcal{C}(\Psi^2))$.

Assumption 6.1. (i) For $t \in \{1, 2\}$ there are iid observations (X_i^t, Y_i^t) , $i = 1, \dots, n_t$, such that Assumptions 2.1–2.4, 3.1 and 3.2 are satisfied with respect to the underlying densities f^t with supports $\mathcal{D}^t$; (ii) $\mathcal{D}_{norm}^1 = \mathcal{D}_{norm}^2$; (iii) for some $n \leq \min\{n_1, n_2\}$ the observations $((X_i^1, Y_i^1), (X_i^2, Y_i^2))$, $i = 1, \dots, n$ are iid and their joint distribution possesses a continuous density f_{12} with support $\mathcal{D}^1 \times \mathcal{D}^2$; (iv) for any $i = 1, \dots, n$, (X_i^1, Y_i^1) is independent of (X_j^2, Y_j^2) for all $j = 1, \dots, n_2$ with $i \neq j$; (v) for any $i = 1, \dots, n$, (X_i^2, Y_i^2) is independent of (X_j^1, Y_j^1) for all $j = 1, \dots, n_1$ with $i \neq j$.

Note that condition (i) of this assumption only guarantees that all estimators $\hat{\theta}_C^t(x, y \mid \mathcal{X}_{n_t}^t)$ (and $\hat{\gamma}_C^t(x, y \mid \mathcal{X}_{n_t}^t)$, by Remark 3.1) follow the asymptotic distributions derived in Theorems 3.1 and 3.2. Condition (ii) ensures that the cross-efficiency estimators $\hat{\theta}_C^1(X_i^2, Y_i^2 \mid \mathcal{X}_{n_1}^1)$

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.

We first consider an individual firm operating at fixed points (x^1, y^1) and (x^2, y^2) in periods 1 and 2. From (6.18), the Malmquist index for this firm is

$$\mathcal{M} = \left(\frac{\gamma(x^2, y^2 \mid \mathcal{C}(\Psi^1))}{\gamma(x^1, y^1 \mid \mathcal{C}(\Psi^1))} \times \frac{\gamma(x^2, y^2 \mid \mathcal{C}(\Psi^2))}{\gamma(x^1, y^1 \mid \mathcal{C}(\Psi^2))} \right)^{1/2}. \quad (6.19)$$

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

$$\widehat{\mathcal{M}} = \left(\frac{\widehat{\gamma}_C(x^2, y^2 \mid \mathcal{X}_{n_1}^1)}{\widehat{\gamma}_C(x^1, y^1 \mid \mathcal{X}_{n_1}^1)} \times \frac{\widehat{\gamma}_C(x^2, y^2 \mid \mathcal{X}_{n_2}^2)}{\widehat{\gamma}_C(x^1, y^1 \mid \mathcal{X}_{n_2}^2)} \right)^{1/2}. \quad (6.20)$$

Theorem 3.3 of Kneip et al. (2021) for convex technologies extends to general technologies as follows.

Theorem 6.1. *Under Assumptions 2.1–2.4, 3.1, 3.2 and 6.1, for $(x^1, y^1) \in \mathcal{D}^1$ and $(x^2, y^2) \in \mathcal{D}^2$*

$$n^{\frac{1}{p+q-1/2}} \left(\widehat{\mathcal{M}} - \mathcal{M} \right) \xrightarrow{\mathcal{L}} F^{\mathcal{M}} \quad (6.21)$$

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

This result permits inference about the unobserved, true Malmquist index $\mathcal{M}$ in (6.19) using the subsampling methods described by Simar and Wilson (2011).

Now consider the log of the Malmquist index $\mathcal{M}_i$ defined in (6.18) for $((X_i^1, Y_i^1), (X_i^2, Y_i^2)) \in \mathcal{X}_n^{1,2}$, i.e.,

$$\log \mathcal{M}_i = \frac{1}{2} \left[\log \gamma(X_i^2, Y_i^2 \mid \mathcal{C}(\Psi^1)) - \log \gamma(X_i^1, Y_i^1 \mid \mathcal{C}(\Psi^1)) + \log \gamma(X_i^2, Y_i^2 \mid \mathcal{C}(\Psi^2)) - \log \gamma(X_i^1, Y_i^1 \mid \mathcal{C}(\Psi^2)) \right]. \quad (6.22)$$

Define

$$\mu_{\mathcal{M}} := E(\log \mathcal{M}_i), \quad (6.23)$$

where the expectation is with respect to (X_i^t, Y_i^t) in both periods $t \in \{1, 2\}$. An estimator of $\log \widehat{\mathcal{M}}_i$ is given by

$$\log \widehat{\mathcal{M}}_i = \frac{1}{2} \left[\log \widehat{\gamma}_C(X_i^2, Y_i^2 \mid \mathcal{X}_{n_1}^1) \log \widehat{\gamma}_C(X_i^1, Y_i^1 \mid \mathcal{X}_{n_1}^1) + \log \widehat{\gamma}_C(X_i^2, Y_i^2 \mid \mathcal{X}_{n_2}^2) - \log \widehat{\gamma}_C(X_i^1, Y_i^1 \mid \mathcal{X}_{n_2}^2) \right], \quad (6.24)$$

and hence $\mu_{\mathcal{M}}$ is estimated by

$$\widehat{\mu}_{\mathcal{M},n} := n^{-1} \sum_{i=1}^n \log \widehat{\mathcal{M}}_i. \quad (6.25)$$

The next theorem is useful for making inference about the difference between $\widehat{\mu}_{\mathcal{M},n}$ and $\mu_{\mathcal{M}}$. 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 . We use the convention that sets $\widehat{\gamma}_C^1(x, y \mid \mathcal{X}_{n_1}^1) := 1$ when $(x, y) \notin \mathcal{C}(\widehat{\Psi}^1)$, as well as $\widehat{\gamma}_C^2(x, y \mid \mathcal{X}_{n_2}^2) := 1$ when $(x, y) \notin \mathcal{C}(\widehat{\Psi}^2)$, although this is seldom needed in practice. Asymptotically this does not impose a problem due to Assumption 6.1 (ii).

The following result generalizes Theorem 3.4 of Kneip et al. (2021) to non-convex technologies, and establishes rates of convergence of the moments of logs of efficiency estimators contained in $\log \widehat{\mathcal{M}}_i$.

Theorem 6.2. *Under Assumptions 2.1–2.4, 3.1, 3.2 and 6.1, and for all $t, s \in \{1, 2\}$, there exist constants $0 < C_0^{ts} < \infty$ such that for all $i \in \{1, \dots, n\}$ and as $n \leq \min\{n_1, n_2\} \rightarrow \infty$,*

$$E \left(\log \widehat{\gamma}_C^s(X_i^t, Y_i^t \mid \mathcal{X}_{n_s}^s) - \log \gamma_C^s(X_i^t, Y_i^t) \right) = C_0^{ts} n_s^{-\frac{1}{p+q-1/2}} + O \left(n_s^{-\frac{2}{p+q-1/2}} (\log n_s)^{\frac{4}{p+q-1/2}} \right) \quad (6.26)$$

and

$$E \left(\left[\log \widehat{\gamma}_C^s(X_i^t, Y_i^t \mid \mathcal{X}_{n_s}^s) - \log \gamma_C^s(X_i^t, Y_i^t) \right]^2 \right) = O \left(n_s^{-\frac{2}{p+q-1/2}} \right), \quad (6.27)$$

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

$$\begin{aligned} & \left| E \left(\left[\log \widehat{\gamma}_C^s(X_i^t, Y_i^t \mid \mathcal{X}_{n_s}^s) - E(\log \widehat{\gamma}_C^s(X_i^t, Y_i^t \mid \mathcal{X}_{n_s}^s)) \right] \right. \right. \\ & \quad \left. \left. \left[\log \widehat{\gamma}_C^{s^*}(X_j^{t^*}, Y_j^{t^*} \mid \mathcal{X}_{n_{s^*}}^{s^*}) - E(\log \widehat{\gamma}_C^{s^*}(X_j^{t^*}, Y_j^{t^*} \mid \mathcal{X}_{n_{s^*}}^{s^*})) \right] \right) \right| = O \left(n^{-\frac{p+q+3/2}{p+q-1/2}} (\log n)^{\frac{2p+2q+3}{p+q-1/2}} \right) \\ & \quad = o(n^{-1}). \end{aligned} \quad (6.28)$$

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 $\mathcal{D}^t \subset \Psi^t$.

The analog of Theorem 3.5 in Kneip et al. (2021) for non-convex technologies follows immediately from the results in Theorem 6.2. This in turn leads to the following CLT, where $\sigma_{\mathcal{M}}^2 = E((\log \mathcal{M}_i - E(\log \mathcal{M}_i))^2)$ with the expectation over (X, Y) in both periods 1 and 2. The result is the analog of Theorem 3.6 and Lemma 3.3 in Kneip et al. (2021).

Theorem 6.3. *Under Assumptions 2.1–2.4, 3.1, 3.2 and 6.1,*

$$\sqrt{n}(\hat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}} - \mathcal{R}_n) \xrightarrow{\mathcal{L}} N(0, \sigma_{\mathcal{M}}^2) \quad (6.29)$$

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

$$\mathcal{R}_n := E(\hat{\mu}_{\mathcal{M},n} - \mu_{\mathcal{M}}) = C_{\mathcal{M}} n^{-\frac{1}{p+q-1/2}} + O\left(n^{-\frac{2}{p+q-1/2}} (\log n)^{\frac{4}{p+q-1/2}}\right). \quad (6.30)$$

In addition,

$$\hat{\sigma}_{\mathcal{M},n}^2 = n^{-1} \sum_{i=1}^n \left(\log \widehat{\mathcal{M}}_i - \hat{\mu}_{\mathcal{M},n}\right)^2 \xrightarrow{p} \sigma_{\mathcal{M}}^2. \quad (6.31)$$

Empirical papers (e.g. Färe et al., 1992; Färe et al., 1997; Wheelock and Wilson 1999 and 2013; Ball et al., 2004 and many others) typically report estimates of geometric means

$$M_n = \left(\prod_{i=1}^n \mathcal{M}_i\right)^{1/n} \quad (6.32)$$

of Malmquist indices for each of n firms. These estimates are computed from the estimator

$$\widehat{M}_n = \exp(\hat{\mu}_{\mathcal{M},n}) = \left(\prod_{i=1}^n \widehat{\mathcal{M}}_i\right)^{1/n} \quad (6.33)$$

of $\widehat{M}_n$ in (6.32). Using the results obtained so far, it is easy to prove the following CLT for making inference about M_n using $\widehat{M}_n$.

Theorem 6.4. *Under Assumptions 2.1–2.4, 3.1, 3.2 and 6.1,*

$$\sqrt{n}(\widehat{M}_n - \widetilde{M}) \xrightarrow{\mathcal{L}} N(0, \exp(2\mu_{\mathcal{M}}) \sigma_{\mathcal{M}}^2), \quad (6.34)$$

as $n \leq \min\{n_1, n_2\} \rightarrow \infty$, where $\mathcal{R}_n$ is defined by (6.30) and

$$\begin{aligned} \widetilde{M} &:= \exp(\mu_{\mathcal{M}} + \mathcal{R}_n) = \exp(\mu_{\mathcal{M}})(1 + \mathcal{R}_n) + O\left(n^{-\frac{2}{p+q-1/2}}\right) \\ &= \exp(\mu_{\mathcal{M}}) + \exp(\mu_{\mathcal{M}}) C_{\mathcal{M}} n^{-\frac{1}{p+q-1/2}} + O\left(n^{-\frac{2}{p+q-1/2}} (\log n)^{\frac{4}{p+q-1/2}}\right). \end{aligned} \quad (6.35)$$

Theorem 6.4 can be used for inference when $p + q = 2$, in which case the bias term $\exp(\mu_{\mathcal{M}})C_{\mathcal{M}}n^{-1/\kappa}$ in (6.35) can be neglected. But if $p + q \geq 3$, the bias explodes as $n \rightarrow \infty$. To avoid this problem, the bias can be estimated by a generalized jackknife method similar of the one described above (see Kneip et al., 2021 for details on how to implement the jackknife in dynamic settings), and the geometric mean must be computed over a subsample of size $n_{\kappa} = \lfloor n^{2/\kappa} \rfloor$. Assume the observations in $\mathcal{X}_n$ are randomly sorted, and define

$$\widehat{M}_{n_{\kappa}} := \left(\prod_{i=1}^{n_{\kappa}} \widehat{\mathcal{M}}_i \right)^{1/n_{\kappa}} \quad (6.36)$$

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 $\mathcal{M}_i$. Then it is straightforward to prove the following CLT.

Theorem 6.5. *Under the conditions of Theorem 6.4, for cases where $p + q \geq 3$,*

$$\sqrt{n_{\kappa}} \left(\widehat{M}_{n_{\kappa}} - \exp(\mu_{\mathcal{M}})C_{\mathcal{M}}n^{-1/\kappa} - \exp(\mu_{\mathcal{M}}) + \eta_{n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, \exp(2\mu_{\mathcal{M}})\sigma_{\mathcal{M}}^2) \quad (6.37)$$

as $n \rightarrow \infty$, where $\eta_{n,\kappa} = o(n^{1/\kappa})$.

7 Monte Carlo Evidence

7.1 Experimental Framework

We conduct extensive Monte Carlo experiments to provide evidence on coverages of estimated confidence intervals for μ_{θ_C} and the performance of our returns-to-scale test in terms of achieved size and power. Each experiment consists of 1,000 Monte Carlo trials. On each trial, we simulate n observations from a known data-generation process (DGP) with given values of number of inputs p and number of outputs q . To simulate data from non-convex technologies where constant returns to scale (CRS) may or may not hold, first consider the function

$$g(t \mid \delta) = [2t\mathcal{J}(2^{-1}t \mid \delta, \delta)]^{1/2} \quad (7.1)$$

where $t \in [0, 2]$, $\delta > 0$, and $\mathcal{J}(t \mid \delta, \delta)$ is the regularized incomplete beta function obtained by dividing the incomplete beta function by the beta function (e.g., see Abramowitz and

Stegun, 1964, formula 6.6.2 or 26.5.1).⁶ The parameter δ controls the shape the function $g(t \mid \delta)$. If $\delta = 1$, then $g(t \mid \delta)$ is linear over $t \in [0, 2]$, but for larger values $g(t \mid \delta)$ is non-linear and non-convex. Figure 1 shows plots of $g(t \mid \delta)$ for values of $\delta \in \mathfrak{D} := \{1.0, 1.4, 2.0, 2.8, 4.0, 5.5, 8.0, 12.0, 18.0\}$. As δ increases, $g(t \mid \delta)$ exhibits increasing curvature and non-linearity.

In our Monte Carlo experiments, we set $q = 1$ and choose particular values for $p \in \{1, 2, \dots, 5\}$ and $\delta \in \mathfrak{D}$. On each trial within an experiment we simulate efficient input values $\tilde{X}_{ij}$, $i = 1, \dots, n$ and $j = 1, \dots, p$, by drawing independently from the uniform distribution with support on $(0.25, 1.75)$. We simulate the inefficiency process by drawing values θ_i independently from the density

$$f(t \mid \lambda) = \begin{cases} \lambda t^{-2} e^{-\lambda(t^{-1}-1)} & \forall t \in (0, 1], \\ 0 & \text{otherwise} \end{cases} \quad (7.2)$$

with $\lambda = 2$. Then θ_i^{-1} has the exponential density $f_*(t \mid \lambda) = \lambda e^{-\lambda t} \forall t \in [1, \infty)$ and $E(\theta^{-1}) = 1.5$ and $E(\theta) \approx 0.722657$. We then compute “observed” input values $X_{ij} = \tilde{X}_{ij}\theta^{-1}$ and corresponding “observed” output values

$$Y_i = g\left(\prod_{j=1}^p (\tilde{X}_{ij})^{1/p} \mid \delta\right). \quad (7.3)$$

The parameter δ controls the shape of the technology. If $\delta = 1$, then $Y_i^{\text{eff}} = \prod_{j=1}^p (\theta_i X_{ij})^{1/p}$ and the technology is convex and exhibits CRS. But if $\delta > 1$, the technology is neither convex nor CRS; as δ increases above 1, the technology becomes increasingly non-convex and farther from CRS.

7.2 Coverages of Confidence Intervals for μ_{θ_C}

We conduct a series of experiments to estimate the coverages of the confidence intervals in (5.21) and (5.22) using either the generalized jackknife method or the bias estimate $\kappa(\hat{\theta}_C^*(x_0, y_0 \mid \mathcal{X}_n) - \hat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n))$ that appears in (5.16) to obtain bias estimates $\hat{B}_{n,\kappa}$. We set $q = 1$ and consider various values $p \in \{1, 2, 3, 4, 5\}$ and sample sizes $n \in \{100, 200, 500, 1000, 5000, 10000, 100000\}$. In addition, we set $\delta = 4.0$ to induce a mild degree of non-convexity in the frontier. For each value of p , we need to know

⁶The regularized incomplete beta function forms the distribution function corresponding to the beta distribution.

$\mu_{\theta_C} = E(\theta_C(X_i, Y_i))$, where the expectation is with respect to (X_i, Y_i) . This is difficult to calculate analytically for the DGP described above in Section 7.1. Instead, we approximate μ_{θ_C} for each value of p by generating a sample of $n = 50,000,000$ observations, computing $\hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$ for each of the n observations and then taking the sample mean of these estimates. This provides approximate values 0.56321, 0.60074, 0.62003, 0.62897 and 0.63924 corresponding to $p = 1, \dots, 5$ respectively.

On each trial within a given experiment, we generate a sample of size n , and compute the interval estimates given by (5.21) when $p + q = 2$ or by (5.22) when $p + q > 2$ with nominal sizes $\alpha \in \{0.1, 0.05, 0.01\}$. We then record the proportion of trials where the estimated confidence intervals include the true value μ_{θ_C} . The results of these experiments are shown in Table 1 where columns 3–5 give the estimated coverages corresponding to $(1 - \alpha) = 0.90, 0.95$ and 0.99 obtained using the estimator $\hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$ and the generalized jackknife estimates of bias. Similarly, columns 6–8 give estimated coverages obtained using the alternative bias-corrected estimator described in Section 5.2.

The results in Table 1 indicate that for either estimator, coverages improve as the sample size n increases, but worsen as the number of dimensions increases for a given sample size. This is as expected. For the ordinary estimator $\hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$ and the generalized jackknife estimates of bias, coverages are close to the nominal values $(1 - \alpha)$ even with $p + 1 = 6$ provided $n = 500$ or larger. Coverages obtained with the alternative bias-corrected estimator $\tilde{\theta}_C(X_i, Y_i)$ are less good, but still close to coverages obtained with the ordinary estimator when $p + q = 2$ or 3 provided n equals 500 or more. But coverages for intervals based on $\tilde{\theta}_C(X_i, Y_i)$ degrade as the number of dimensions increases beyond 3 unless the sample size is 5,000 or more with $(p + q) = 4$ or perhaps 100,000 or more when $(p + q) = 5$. When $(p + q) = 6$, the coverage at the .1 level is only 0.749 using $\tilde{\theta}_C(X_i, Y_i)$ with $n = 100,000$, compared with coverage of 0.892 obtained using the ordinary estimator $\hat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$ with the jackknife bias correction. While the alternative bias correction avoids the computational cost of the jackknife, the results in Table 1 indicate that there is a price to pay for avoiding the computational burden of the jackknife.

7.3 Performance of Returns-to-Scale Tests

In order to examine the performance of tests of CRS versus non-CRS in terms of achieved size and power in various situations, we conduct a series of Monte Carlo experiments using the DGP described above in Section 7.1. For $q = 1$ and fixed values of p and n , setting $\delta = 1$ yields a frontier that is CRS everywhere. As δ increases beyond 1, we have increasing departures from the null as illustrated in Figure 1.

In our experiments, we consider $p \in \{1, \dots, 5\}$, $n \in \{100, 200, 500, 1000\}$ and $\delta \in \mathfrak{D}$ where the set $\mathfrak{D}$ is defined above in Section 7.1. On each trial of each experiment, we split the sample 100 times and compute the statistics $\hat{T}_n$ and $\hat{K}_n$ given in (6.13) and (6.14) and use the smooth bootstrap procedure discussed in Section 6.1.3 with 1,000 bootstrap replications to compute corresponding p -values. In each case, biases are estimated using the generalized jackknife with 100 replications. For either test, we count the number of trials where the p -value is less than the nominal test size $\alpha = 0.1, 0.05$ or 0.01 , and we divide this number by 1,000, the number of Monte Carlo trials in each experiment. In addition, we similarly estimate rejection rates achieved with a single sample-split using the first split on each trial and computing a p -value $\hat{p} = 1 - \Phi^{-1}(\hat{\tau}_{\ell,n})$ corresponding to the statistics $\hat{\tau}_{\ell,n}$ defined in (6.5) and (6.8) for $\ell \in \{1, 2\}$, where $\Phi^{-1}(\cdot)$ denotes the standard normal quantile function. In all of our experiments, we use the statistic $\hat{\tau}_{1,n}$ when $(p + q) = 2$ or $\hat{\tau}_{2,n}$ when $(p + q) > 2$.

Estimated rejection rates for the Kolmogorov-Smirnov test based on $\hat{K}_n$ are shown in Table 2. Tests based on $\hat{T}_n$ yield similar rejection rates, but in many cases the size and power achieved with the Kolmogorov-Smirnov statistic are slightly better. The results in Table 2 indicate that the tests are conservative. For every case considered, the rejection rate (corresponding to $\delta = 1$) is less than the nominal size α . The results also show that rejection rates increase with increasing departures from the null (i.e., as δ becomes larger) for a given sample size. Table 2 also shows the effects of increasing dimensionality. As p increases, with modest departures from the null, rejection rates diminish for a given value of δ . Overall, the results shown in Table 2 demonstrate the usefulness of the test in applied settings, even when the sample size is 100, provided the number of dimensions is no more than 4 or 5 (corresponding to $p = 3$ or 4 in the table).

The rejection rates resulting from only one sample-split are problematic due to over-rejection as seen in Table 3. In all cases, with $\delta = 1$ so that the null hypothesis of CRS

is true, the tests based on just one split yield rejection rates larger than the nominal test sizes. With $n = 100$ and $\alpha = 0.1$, the rejection rates range from 0.199 when $p = 1$ to 0.302 when $p = 5$. With $n = 200$, rejection rates range from 0.164 to 0.228 when the null is true. With $n = 500$ or 1,000, the rejection rates range from 0.118 to 0.203, or 0.101 to 0.185 (respectively) when the null is true.

In addition, the results obtained with a single sample split in Table 3 reveal that the power of the tests with a single split is often smaller than the power of tests based on 100 splits. With $p = 1$ and $\alpha = 0.1$, the Kolmogorov-Smirnov test with 100 splits has greater power against the null than the test with only one split even with δ equal to only 1.4, and remains larger as δ increases. With $p = 2$, the Kolmogorov-Smirnov test has greater power than that of the one-split test when $p = 2$ and $\delta \geq 2.0$. As discussed by Simar and Wilson (2020), the tests based on a single sample-split can yield different results with different splits of the same sample, but with 1,000 Monte Carlo trials we obtain an estimate of the *expected* rejection rates. With only one split, the tests are too aggressive and over-reject, but with 100 splits, we obtain conservative tests that have good (i.e., useful) power, often with greater power than that achieved with a single split.

8 Empirical Illustration with Real Data

O’Loughlin and Wilson (2021) examine the efficiency and productivity of municipal governments in the U.S. using data from 1997, 2002, 2007 and 2012. O’Loughlin and Wilson specify a single input, total current operating expenditures (reported in the Annual Survey of State and Local Government Finances) so that $p = 1$. Their paper specifies $q = 6$ outputs, namely total population, total charges for sewerage and waste management, the reciprocal of the total crime rate, total land area (in square miles), total building permits and the employment rate. The authors note that their specification of outputs reflects the wide variety of goods and services provided by municipal governments, and that their input-output specification is similar to other studies that examine the efficiency of municipal governments.

Among their results, O’Loughlin and Wilson (2021) find compelling evidence from the test described by Daraio et al. (2018) that the frontiers faced by municipalities in the U.S. differ across years as well as four broad census regions (Northeast, Midwest South and West, represented as regions 1, 2, 3 and 4, respectively; see O’Loughlin and Wilson, Figure 1).

In addition, O’Loughlin and Wilson find evidence of non-convexity of the production sets in various region-years using the test of convexity versus non-convexity developed by Kneip et al. (2016) using multiple sample splits with the bootstrap method described by Simar and Wilson (2020). The authors note that the evidence is mixed, but among 96 different tests, convexity is rejected at the 10-percent level in more than 30-percent of the cases. Moreover, some of the region-years have small sample sizes, e.g., 72 to 95 observations in Region 1, and this may cause the tests to have low power. Given the evidence against convexity of production sets faced by U.S. municipalities, O’Loughlin and Wilson (2021) were unable to examine returns to scale since the test of constant versus non-constant returns to scale introduced by Kneip et al. (2016) is valid only if the production set is convex.

We revisit the O’Loughlin and Wilson (2021) study in order to test constant versus non-constant returns to scale and to estimate Malmquist indices in order to examine productivity changes over time in each of the four census regions. O’Loughlin and Wilson specify a single input corresponding to cost, measured by total current operating expenditures, and 6 outputs designed to measure services provided by municipal governments. Similar to O’Loughlin and Wilson, we exploit the high degree of collinearity among the outputs to reduce dimensionality of the problem using eigensystem methods discussed by Daraio and Simar (2007) and Wilson (2018). In particular, we use eigensystem methods to decompose the moment matrix for four outputs (total charges for sewerage and waste management, the reciprocal of the total crime rate, total building permits and the employment rate) and to replace these four outputs with their first principle component. These four outputs are substantially collinear; the ratio of the largest eigenvalue to the sum of the eigenvalues obtained from their moment matrix is 0.9974. Simulation results provided by Wilson (2018) suggest that in such situations, dimension-reduction is likely to results in an improvement in mean-square error when using FDH estimators to estimate efficiency. We retain two of the original output variables—population and land area—in addition to the reduced-dimensional output. After eliminating observations with missing values for one or more of the variables, 648, 730, 746 and 800 observations for 1997, 2002, 2007 and 2012 remain (respectively), for a total of 2,924 observations, and we have $p = 1$ input and $q = 3$ output quantities. See O’Loughlin and Wilson (2021) for further discussion of the variables and data sources.

Using the data and the test developed above in Section 6.1, we test constant versus non-constant returns to scale for each region-year while allowing for non-convex production sets. For a given region-year, we randomly split the sample 100 times. For each split, we compute the test statistic $\hat{\tau}_{1,n}$ in (6.5) and find its corresponding p -value. We then use these values to compute the statistics $\hat{T}_n$ and $\hat{K}_n$ in (6.13)–(6.14), and we use the bootstrap idea of Simar and Wilson (2020) with the smoothing modification discussed above in Section 6.1.3. We conduct the tests in each of three directions, i.e., input, output and hyperbolic directions.

The results of these tests using the hyperbolic direction are shown in Table 4. Table 4 shows that the null hypothesis of CRS is rejected in favor of non-CRS in each of the 32 tests, Working in the output direction yields similar results, with CRS rejected in each case. Using the input direction, CRS is rejected in favor of non-CRS in 25 (78.1 percent) of the 32 tests, but we fail to reject at .1 or better in the remaining 7 cases. In the input direction, both tests reject CRS in 11 of the 16 region-years, and in only two cases—region 1 in 2012 and region 2 in 1997—neither test rejects CRS. As noted by O’Loughlin and Wilson (2021), failure to reject the null should not be construed as evidence supporting the null. Overall, the results provide substantial evidence against CRS.

In order to examine changes in productivity, we estimate Malmquist indices using the hyperbolic CFDH estimator. The results of our returns-to-scale tests discussed above indicate that the slower, non-CRS convergence rate of $n^{1/(p+q-1/2)}$ should be used instead of the faster, CRS rate of $n^{1/(p+q-1)}$ for making inferences. Table 5 shows results for geometric means across municipalities for our estimated Malmquist indices for pairs of years in each of the four census regions. Note that the data used by O’Loughlin and Wilson (2021) constitute an unbalanced panel. The third and fourth columns of Table 5 give the numbers of observations in each region-year, while the fifth column gives the number of municipalities observed in *both* years for each pair of years considered. As discussed in Section 6.2, for a given region-year, we estimate the Malmquist index for each observation observed in both periods, and we use all of the available observations to do so.

The sixth column of Table 5 gives the geometric means of estimated Malmquist indices in each region-year, and the seventh column gives the jackknife estimate of bias. The last two columns in the table give lower and upper bounds of estimated 99-percent confidence intervals. Recall that the Malmquist index equals 1 when there is not productivity change,

and is greater than (less than) 1 when there is an increase (decrease) in productivity. To interpret the values shown in Table 5, note that a value $a > 1$ of a geometric mean of Malmquist indices indicates that productivity increased by $100 \times (a - 1)$ -percent between the relevant pair of years, while a value $a < 1$ indicates that productivity declined by $100 \times (1 - a)$ -percent.

Only two of the geometric means in Table 5 are less than 1, while all the others are greater than 1. For the two means that are less than 1, the upper bound on the corresponding 99-percent confidence interval estimates is also less than 1, suggesting that the declines in productivity for regions 1 and 2 during 2007–2012 are statistically significant. The period 2007–2012 corresponds to the Great Recession in the U.S., and so it is perhaps not surprising that some regions suffered declines in productivity among municipalities due to the recession’s effect on demand for services and on revenue). For all of the means that are greater than 1, the lower bound on the corresponding confidence interval estimates are also greater than 1, again suggesting that the increases in productivity are statistically significant. For the entire period examined by O’Loughlin and Wilson (2021), 1997 to 2012, the geometric means reported in Table 5 indicate that productivity increased by 7.79 to 18.35 percent on average among municipalities in each region. However, these estimates are biased downward as indicated by the bias estimates in column 7 of the table; the lower bounds on the estimates for the 15-year period 1997–2012 indicate that productivity increased by much more than indicated by the geometric means.

9 Summary and Conclusions

As noted in the introduction, a number of papers have used the test developed by Kneip et al. (2016) to test convexity of production sets, leading to rejection in favor of non-convexity. Until now, empirical researchers whose data reject convexity have been unable to test CRS versus non-CRS nor to estimate and make inference about productivity change using Malmquist indices. The paper provides the tools to do so.

We derive the asymptotic properties of the conical FDH estimator of distance to the boundary of the cone spanned by a possibly non-convex production set. We use these results to establish CLTs for means of these estimators over sample observations. Monte-Carlo experiments confirm the asymptotic approximations are useful for estimating confidence

intervals in finite samples. We illustrate the use of our CLT results by analyzing the productivity of U.S. municipal governments, where convexity is rejected in most region-years. We find that the CRS assumption should be rejected, and this provides information needed for making inference about productivity change using Malmquist indices. We estimate confidence intervals for geometric means of Malmquist index estimates for various region-years, and conclude that on average, U.S. municipalities significantly increased their productivity in providing local public goods except in two regions during 2007–2012. Our analysis serves as a useful template for how applied researchers might proceed when confronted with data that reject convexity of production sets.

Acknowledgments

Technical support from the Cyber Infrastructure Technology Integration group at Clemson University are gratefully acknowledged. Part of the research in this paper was performed while Wilson was a consultant with the Federal Reserve Bank of St. Louis, St. Louis, Missouri, USA. The views expressed in this paper do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis nor the Federal Reserve System. Any remaining errors are solely our responsibility.

References

- Abramowitz, M. and I. A. Stegun (1964), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Washington, DC: U.S. Department of Commerce, National Bureau of Standards, tenth printing (with corrections), December 1972 edition.
- Apon, A. W., L. B. Ngo, M. E. Payne, and P. W. Wilson (2015), Assessing the effect of high performance computing capabilities on academic research output, *Empirical Economics* 48, 283–312.
- 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.
- Banker, R. D., A. Charnes, and W. W. Cooper (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* 30, 1078–1092.
- 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.
- Charnes, A., W. W. Cooper, and E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research* 2, 429–444.
- Daouia, A., L. Simar, and P. W. Wilson (2017), Measuring firm performance using nonparametric quantile-type distances, *Econometric Reviews* 36, 156–181.
- Daraio, C. and L. Simar (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, New York: Springer Science+Business Media, LLC.
- Daraio, C., L. Simar, and P. W. Wilson (2018), Central limit theorems for conditional efficiency measures and tests of the ‘separability condition’ in non-parametric, two-stage models of production, *The Econometrics Journal* 21, 170–191.
- (2021), Quality as a latent heterogeneity factor in the efficiency of universities, *Economic Modelling* 99, 105485.
URL: <https://www.sciencedirect.com/science/article/pii/S0264999321000687>
- Debreu, G. (1951), The coefficient of resource utilization, *Econometrica* 19, 273–292.
- Deprins, D., L. Simar, and H. Tulkens (1984), Measuring labor inefficiency in post offices, in M. M. P. Pestieau and H. Tulkens, eds., *The Performance of Public Enterprises: Concepts and Measurements*, Amsterdam: North-Holland, pp. 243–267.
- 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.
- Färe, R., S. Grosskopf, and M. Norris (1997), Productivity growth, technical progress, and efficiency change in industrialized countries: Reply., *American Economic Review* 87, 1040–1043.

- Farrell, M. J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society A* 120, 253–281.
- Kerstens, K. and P. Vanden Eeckaut (1997), Estimating returns to scale using nonparametric deterministic technologies: A new method based on goodness-of-fit. Discussion paper #9713, Center for Operations Research and Econometrics, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- 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.
- (2021), Inference in dynamic, nonparametric models of production: Central limit theorems for Malmquist indices, *Econometric Theory* 37, 537–572.
- Koopmans, T. C. (1951), An analysis of production as an efficient combination of activities, in T. C. Koopmans, ed., *Activity Analysis of Production and Allocation*, New York: John-Wiley and Sons, Inc., pp. 33–97. Cowles Commission for Research in Economics, Monograph 13.
- Malmquist, S. (1953), Index numbers and indifference surfaces, *Trabajos de Estadística y de Investigación Operativa* 4, 209–242.
- O’Loughlin, C. T. and P. W. Wilson (2021), Benchmarking the performance of U.S. municipalities, *Empirical Economics* 60, 2665–2700.
- 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.
- Park, B. U., L. Simar, and C. Weiner (2000), FDH efficiency scores from a stochastic point of view, *Econometric Theory* 16, 855–877.
- Qiao, G. and Z. Wang (2021), Vertical integration vs. specialization: A nonparametric conditional efficiency estimate for the global semiconductor industry, *Journal of Productivity Analysis* 56, 139–150.
- Rhodes, E. (1978), *Data Envelopment Analysis and RElated Approaches for Measuring the Efficiency of Decision Making Units with an Application to Program Follow Through in U.S. Education*, Ph.D. thesis, Carnegie-Mellon University, School of Urban and Public Affairs, Pittsburgh, PA.
- Simar, L. and P. W. Wilson (1998), Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Science* 44, 49–61.
- (2011), Inference by the m out of n bootstrap in nonparametric frontier models, *Journal of Productivity Analysis* 36, 33–53.

- (2019), Central limit theorems and inference for sources of productivity change measured by nonparametric Malmquist indices, *European Journal of Operational Research* 277, 756–769.
 - (2020), Hypothesis testing in nonparametric models of production using multiple sample splits, *Journal of Productivity Analysis* 53, 287–303.
 - (2022), Modern tools for evaluating the performance of health-care providers, in S. Grosskopf, V. Valdmanis, and V. Zelenyuk, eds., *The Cambridge Handbook of Productivity, Efficiency & Effectiveness in Health Care*, Cambridge, UK: Cambridge University Press. Forthcoming.
- Wheelock, D. C. and P. W. Wilson (1999), Technical progress, inefficiency, and productivity change in U. S. banking, 1984–1993, *Journal of Money, Credit, and Banking* 31, 212–234.
- (2013), The evolution of cost-productivity and efficiency among U.S. credit unions, *Journal of Banking and Finance* 37, 75–88.
- Wilson, P. W. (2018), Dimension reduction in nonparametric models of production, *European Journal of Operational Research* 267, 349–367.
- (2021), U.S. banking in the post-crisis era: New results from new methods, in C. Parmeter and R. Sickles, eds., *Advances in Efficiency and Productivity Analysis*, Cham, Switzerland: Springer Nature Switzerland AG, pp. 233–264.
- Wilson, P. W. and S. Zhao (2022), Investigating the performance of Chinese banks over 2007–2014, *Annals of Operations Research* Forthcoming.

Table 1: Coverages of Confidence Intervals for $E(\theta_C(X_i, Y_i))$

p	q	n	$\hat{\theta}_C(X_i, Y_i)$			$\tilde{\theta}_C(X_i, Y_i)$		
			0.90	0.95	0.99	0.90	0.95	0.99
1	1	100	0.857	0.914	0.977	0.823	0.884	0.963
		200	0.857	0.917	0.974	0.832	0.900	0.961
		500	0.875	0.934	0.986	0.865	0.920	0.981
		1000	0.884	0.944	0.990	0.866	0.934	0.982
		5000	0.895	0.945	0.991	0.888	0.942	0.986
		10000	0.900	0.953	0.994	0.897	0.944	0.993
		100000	0.895	0.961	0.993	0.894	0.956	0.993
2	1	100	0.844	0.919	0.980	0.792	0.872	0.957
		200	0.859	0.929	0.984	0.807	0.885	0.967
		500	0.877	0.930	0.979	0.854	0.911	0.972
		1000	0.871	0.928	0.987	0.849	0.916	0.980
		5000	0.890	0.944	0.985	0.890	0.933	0.983
		10000	0.882	0.934	0.986	0.880	0.928	0.987
		100000	0.889	0.938	0.993	0.887	0.939	0.993
3	1	100	0.862	0.926	0.981	0.676	0.753	0.894
		200	0.879	0.936	0.988	0.714	0.805	0.916
		500	0.882	0.939	0.988	0.796	0.869	0.955
		1000	0.887	0.936	0.989	0.825	0.915	0.974
		5000	0.903	0.965	0.992	0.893	0.945	0.991
		10000	0.885	0.940	0.984	0.874	0.929	0.980
		100000	0.919	0.963	0.994	0.919	0.962	0.994
4	1	100	0.835	0.917	0.986	0.570	0.683	0.849
		200	0.849	0.923	0.984	0.565	0.689	0.853
		500	0.904	0.955	0.990	0.616	0.721	0.901
		1000	0.891	0.945	0.988	0.643	0.768	0.909
		5000	0.896	0.951	0.994	0.771	0.866	0.959
		10000	0.908	0.953	0.991	0.826	0.902	0.968
		100000	0.868	0.922	0.986	0.847	0.903	0.976
5	1	100	0.841	0.910	0.987	0.589	0.679	0.835
		200	0.881	0.935	0.989	0.587	0.702	0.856
		500	0.896	0.944	0.992	0.514	0.654	0.837
		1000	0.908	0.960	0.995	0.551	0.656	0.837
		5000	0.888	0.945	0.990	0.581	0.695	0.860
		10000	0.890	0.942	0.989	0.608	0.728	0.894
		100000	0.892	0.936	0.988	0.749	0.849	0.951

Table 2: Rejection Rates for Test of CRS versus non-CRS, $\hat{K}_n$ with 100 Sample Splits

n	δ	$p = 1$			$p = 2$			$p = 3$			$p = 4$			$p = 5$		
		0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
100	1.0	0.045	0.020	0.003	0.040	0.022	0.004	0.019	0.007	0.001	0.009	0.005	0.000	0.005	0.001	0.000
	1.4	0.514	0.350	0.103	0.081	0.028	0.007	0.025	0.010	0.002	0.011	0.006	0.000	0.007	0.001	0.000
	2.0	0.986	0.963	0.835	0.416	0.260	0.084	0.081	0.044	0.007	0.034	0.008	0.000	0.010	0.002	0.000
	2.8	1.000	0.999	0.992	0.862	0.761	0.498	0.357	0.231	0.063	0.119	0.055	0.007	0.051	0.012	0.001
	4.0	1.000	1.000	1.000	0.996	0.986	0.917	0.830	0.731	0.443	0.533	0.336	0.082	0.264	0.127	0.015
	5.5	1.000	1.000	1.000	1.000	0.999	0.994	0.989	0.965	0.848	0.858	0.756	0.454	0.632	0.440	0.153
	8.0	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.991	0.990	0.973	0.867	0.930	0.854	0.596
	12.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000	0.994	0.994	0.987	0.925
	16.0	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	0.987
200	1.0	0.060	0.033	0.005	0.047	0.014	0.000	0.023	0.009	0.001	0.014	0.000	0.000	0.002	0.001	0.000
	1.4	0.950	0.878	0.622	0.199	0.101	0.026	0.037	0.013	0.004	0.017	0.006	0.000	0.005	0.001	0.000
	2.0	1.000	1.000	0.999	0.848	0.747	0.481	0.220	0.121	0.032	0.072	0.029	0.006	0.015	0.004	0.000
	2.8	1.000	1.000	1.000	1.000	0.999	0.976	0.777	0.631	0.346	0.333	0.206	0.064	0.095	0.034	0.004
	4.0	1.000	1.000	1.000	1.000	1.000	1.000	0.995	0.987	0.941	0.851	0.744	0.477	0.497	0.330	0.093
	5.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.995	0.988	0.931	0.940	0.859	0.564
	8.0	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	0.999	0.979
	12.0	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
	16.0	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 of CRS versus non-CRS, $\hat{K}_n$ with 100 Sample Splits
(continued)

n	δ	$p = 1$			$p = 2$			$p = 3$			$p = 4$			$p = 5$		
		0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
500	1.0	0.048	0.014	0.001	0.049	0.018	0.002	0.025	0.009	0.002	0.007	0.002	0.000	0.005	0.001	0.000
	1.4	1.000	1.000	1.000	0.636	0.461	0.214	0.067	0.028	0.005	0.017	0.007	0.002	0.010	0.002	0.000
	2.0	1.000	1.000	1.000	1.000	1.000	0.996	0.652	0.511	0.251	0.152	0.073	0.013	0.037	0.016	0.001
	2.8	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.993	0.961	0.779	0.653	0.375	0.293	0.181	0.058
	4.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.981	0.937	0.860	0.637
	5.5	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	0.994
	8.0	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
	12.0	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	16.0	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
	1.0	0.052	0.017	0.003	0.040	0.017	0.003	0.015	0.004	0.001	0.011	0.005	0.001	0.004	0.001	0.001
	1.4	1.000	1.000	1.000	0.959	0.907	0.751	0.143	0.070	0.018	0.028	0.015	0.002	0.007	0.001	0.001
	2.0	1.000	1.000	1.000	1.000	1.000	1.000	0.962	0.906	0.725	0.384	0.243	0.086	0.067	0.028	0.003
	2.8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.987	0.966	0.873	0.639	0.485	0.225
	4.0	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	0.997	0.978
	5.5	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
	8.0	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
	12.0	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
	16.0	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 3: Rejection Rates for Test of CRS versus non-CRS, One Sample Split, Asymptotic Normality

n	δ	$p = 1$			$p = 2$			$p = 3$			$p = 4$			$p = 5$		
		0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
100	1.0	0.199	0.125	0.044	0.249	0.163	0.071	0.255	0.170	0.067	0.285	0.208	0.095	0.302	0.210	0.090
	1.4	0.381	0.259	0.135	0.269	0.193	0.096	0.285	0.194	0.080	0.303	0.223	0.102	0.305	0.225	0.090
	2.0	0.651	0.531	0.318	0.387	0.295	0.156	0.368	0.264	0.117	0.366	0.269	0.118	0.352	0.254	0.117
	2.8	0.819	0.725	0.507	0.557	0.423	0.250	0.483	0.367	0.193	0.462	0.340	0.179	0.429	0.313	0.159
	4.0	0.925	0.858	0.649	0.733	0.601	0.376	0.608	0.503	0.305	0.586	0.460	0.269	0.555	0.423	0.236
	5.5	0.968	0.918	0.760	0.842	0.755	0.526	0.713	0.611	0.401	0.689	0.578	0.367	0.648	0.535	0.332
	8.0	0.984	0.961	0.843	0.923	0.867	0.687	0.790	0.710	0.518	0.776	0.683	0.464	0.759	0.645	0.446
	12.0	0.992	0.978	0.888	0.978	0.950	0.814	0.852	0.773	0.602	0.847	0.738	0.544	0.830	0.737	0.509
	16.0	0.996	0.986	0.915	0.987	0.973	0.869	0.874	0.810	0.640	0.866	0.784	0.581	0.866	0.779	0.543
	200	1.0	0.164	0.090	0.216	0.140	0.057	0.196	0.115	0.028	0.203	0.121	0.048	0.228	0.142	0.054
200	1.4	0.512	0.388	0.171	0.296	0.194	0.078	0.231	0.132	0.038	0.229	0.140	0.052	0.261	0.155	0.061
	2.0	0.860	0.773	0.568	0.485	0.368	0.175	0.329	0.222	0.068	0.288	0.196	0.071	0.306	0.211	0.082
	2.8	0.978	0.941	0.826	0.688	0.577	0.340	0.477	0.352	0.153	0.405	0.281	0.123	0.412	0.297	0.126
	4.0	0.999	0.991	0.941	0.858	0.781	0.571	0.647	0.525	0.306	0.553	0.432	0.223	0.537	0.415	0.224
	5.5	0.999	0.999	0.981	0.949	0.903	0.759	0.785	0.669	0.445	0.688	0.568	0.340	0.662	0.542	0.322
	8.0	0.999	0.999	0.997	0.988	0.975	0.896	0.875	0.798	0.586	0.805	0.687	0.464	0.779	0.671	0.446
	12.0	1.000	0.999	0.998	0.997	0.991	0.975	0.932	0.870	0.700	0.879	0.786	0.559	0.849	0.757	0.538
	16.0	1.000	1.000	0.997	0.998	0.997	0.987	0.952	0.913	0.748	0.908	0.826	0.595	0.876	0.787	0.572
	200	1.0	0.164	0.090	0.216	0.140	0.057	0.196	0.115	0.028	0.203	0.121	0.048	0.228	0.142	0.054
	1.4	0.512	0.388	0.171	0.296	0.194	0.078	0.231	0.132	0.038	0.229	0.140	0.052	0.261	0.155	0.061

Table 3: Rejection Rates for Test of CRS versus non-CRS, One Sample Split, Asymptotic Normality (continued)

n	δ	$p = 1$			$p = 2$			$p = 3$			$p = 4$			$p = 5$		
		0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
500	1.0	0.118	0.059	0.012	0.203	0.125	0.031	0.180	0.101	0.021	0.162	0.095	0.026	0.181	0.107	0.032
	1.4	0.733	0.602	0.343	0.363	0.254	0.096	0.222	0.137	0.032	0.184	0.107	0.032	0.199	0.119	0.043
	2.0	0.993	0.980	0.907	0.649	0.543	0.311	0.344	0.228	0.091	0.269	0.168	0.055	0.265	0.155	0.058
	2.8	1.000	1.000	0.997	0.883	0.817	0.618	0.539	0.409	0.189	0.408	0.280	0.105	0.368	0.257	0.096
	4.0	1.000	1.000	1.000	0.979	0.959	0.873	0.748	0.631	0.387	0.569	0.452	0.216	0.543	0.405	0.199
	5.5	1.000	1.000	1.000	0.999	0.997	0.967	0.877	0.803	0.593	0.718	0.594	0.368	0.691	0.561	0.328
	8.0	1.000	1.000	1.000	1.000	0.999	0.997	0.941	0.910	0.751	0.838	0.747	0.518	0.811	0.704	0.470
	12.0	1.000	1.000	1.000	1.000	1.000	1.000	0.975	0.949	0.880	0.913	0.838	0.625	0.876	0.804	0.590
	16.0	1.000	1.000	1.000	1.000	1.000	1.000	0.985	0.970	0.915	0.941	0.880	0.684	0.898	0.847	0.648
1000	1.0	0.101	0.052	0.012	0.185	0.095	0.023	0.163	0.096	0.023	0.147	0.090	0.021	0.155	0.080	0.017
	1.4	0.893	0.812	0.616	0.397	0.264	0.103	0.205	0.123	0.035	0.172	0.102	0.027	0.176	0.099	0.023
	2.0	0.999	0.999	0.998	0.815	0.688	0.436	0.367	0.247	0.097	0.274	0.170	0.054	0.250	0.143	0.037
	2.8	1.000	1.000	1.000	0.970	0.943	0.841	0.636	0.474	0.221	0.440	0.299	0.122	0.372	0.249	0.093
	4.0	1.000	1.000	1.000	0.998	0.997	0.980	0.836	0.748	0.512	0.664	0.517	0.262	0.572	0.422	0.187
	5.5	1.000	1.000	1.000	1.000	1.000	0.999	0.941	0.896	0.735	0.802	0.689	0.430	0.735	0.603	0.335
	8.0	1.000	1.000	1.000	1.000	1.000	1.000	0.990	0.970	0.890	0.903	0.845	0.617	0.859	0.771	0.516
	12.0	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.995	0.971	0.960	0.920	0.761	0.927	0.857	0.658
	16.0	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.988	0.978	0.953	0.827	0.938	0.896	0.715

Table 4: RTS Tests for U.S. Municipalities by Year and Region, 100 Sample Splits, Hyperbolic Orientation

Region	1997	2002	2007	2012
	Average Statistic			
1	1.2021 0.0000	7.2983 0.0000	1.5132 0.0000	6.2097 0.0000
2	6.8972 0.0000	9.6692 0.0000	9.1267 0.0000	0.3357 0.0000
3	9.8795 0.0000	0.2759 0.0000	8.8284 0.0000	5.1745 0.0000
4	1.1994 0.0000	0.3102 0.0000	0.9384 0.0000	0.5167 0.0000
	KS-test			
1	0.9997 0.0000	0.9781 0.0000	0.9976 0.0000	0.9792 0.0000
2	0.9932 0.0000	0.9992 0.0000	0.9998 0.0000	0.9959 0.0000
3	0.9986 0.0000	1.0000 0.0000	1.0000 0.0000	1.0000 0.0000
4	1.0000 0.0000	1.0000 0.0000	1.0000 0.0000	1.0000 0.0000

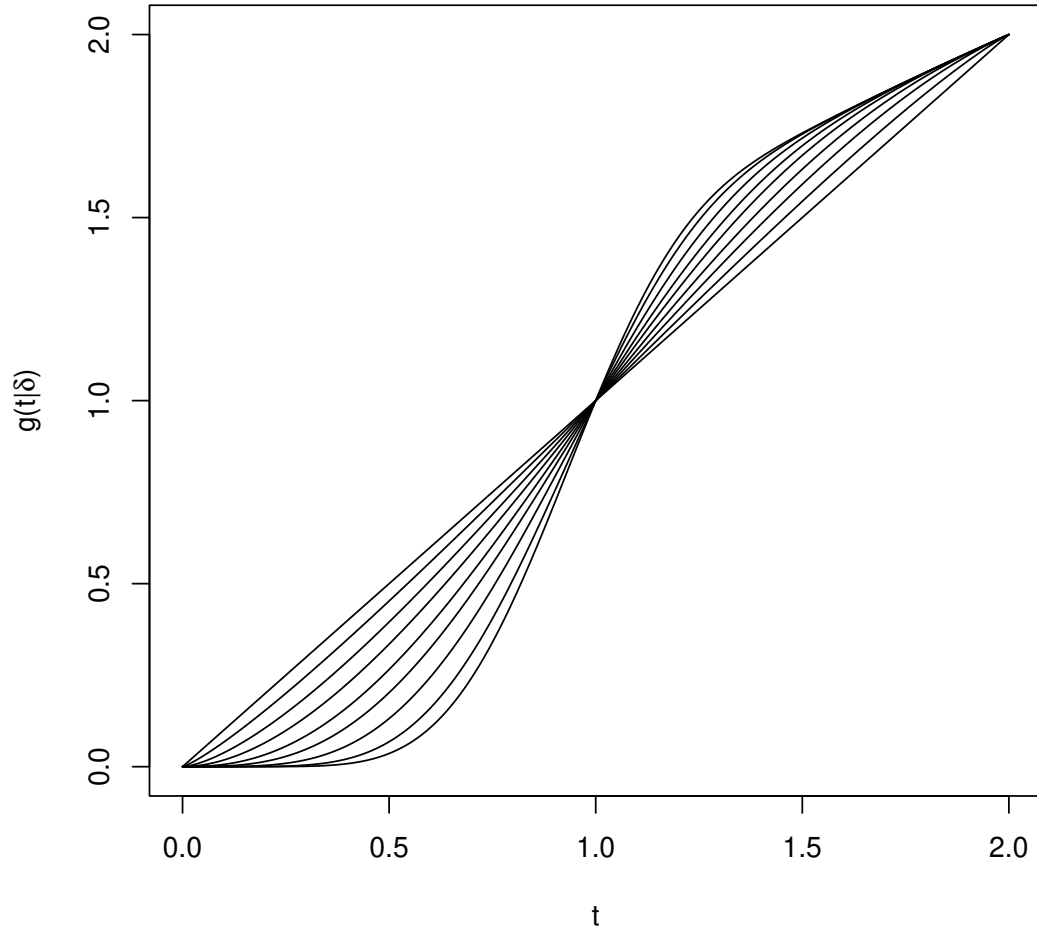
NOTE: Each cell gives the value of the test statistic (top) and the corresponding p -value (bottom).

Table 5: Geometric Means of Malmquist Index Estimates for U.S. Municipalities by Year and Region

Region	Years	n_1	n_2	n	Geometric Mean	Bias Estimate	Lower Bound	Upper Bound
1	1997–2002	90	75	64	1.0431	-0.3965	1.3606	1.5187
1	2002–2007	75	74	62	1.0492	-0.4266	1.4024	1.5492
1	2007–2012	74	72	63	0.9919	0.0511	0.8976	0.9839
1	1997–2012	90	72	52	1.0779	-0.7328	1.6911	1.9304
2	1997–2002	122	158	83	1.0843	-0.7667	1.7671	1.9349
2	2002–2007	158	160	114	1.0430	-0.3885	1.3657	1.4975
2	2007–2012	160	200	140	0.9874	0.1025	0.8030	0.9669
2	1997–2012	122	200	68	1.1197	-1.0469	2.0662	2.2671
3	1997–2002	200	289	179	1.0669	-0.6048	1.6150	1.7283
3	2002–2007	289	288	257	1.0638	-0.5741	1.5779	1.6980
3	2007–2012	288	299	273	1.0081	-0.0791	1.0421	1.1322
3	1997–2012	200	299	165	1.1286	-1.1343	2.1722	2.3536
4	1997–2002	236	208	119	1.0598	-0.4788	1.4612	1.6158
4	2002–2007	208	224	193	1.1035	-0.8730	1.8921	2.0610
4	2007–2012	224	229	213	1.0117	-0.0913	1.0560	1.1500
4	1997–2012	236	229	114	1.1835	-1.4466	2.4962	2.7639

NOTE: Lower and Upper Bounds are from estimates of 99-percent confidence intervals. Numbers of observations in periods 1 and 2 are given by n_1 and n_2 , and the number of municipalities observed in both periods 1 and 2 is given by n .

Figure 1: Function $g(t \mid \delta)$ used to Simulate Data



NOTE: The function $g(t \mid \delta)$ defined in (7.1) is plotted for $\delta = 1.0, 1.4, 2.0, 2.8, 4.0, 5.5, 8.0, 12.0$ and 18.0 . With $\delta = 1$, the function is linear, and becomes increasingly non-linear as δ increases.

Appendix A: Proofs

A.1 Proof of Proposition 3.1

First assume that $\mathcal{X}_n \cap S_\delta^{x_0, y_0} = \emptyset$, and choose some arbitrary $i \in \{1, \dots, n\}$. There necessarily exists some $j_i \in \{1, \dots, q\}$ such that

$$\min_{j=1, \dots, q} \frac{Y_{ij}}{y_{0j}} = \frac{Y_{ij_i}}{y_{0j_i}}, \quad \text{and hence } \tilde{Y}_i^{(j_i)} \geq \tilde{y}_0^{(j_i)},$$

since $\min_{j=1, \dots, q} \frac{Y_{ij} y_{0j_i}}{y_{0j} Y_{ij_i}} = 1$. Then $(X_i, Y_i) \notin S_{\delta, j_i}^{x_0, y_0}$ implies

$$\max_{l=1, \dots, p} \frac{X_{ij}}{x_{0j}} > (1 + \delta) \frac{Y_{ij_i}}{y_{0j_i} a_{\min}^{x_0, y_0}} \theta(x_0, a_{\min}^{x_0, y_0} y_0) = (1 + \delta) \frac{Y_{ij_i}}{y_{0j_i}} \theta_C(x_0, y_0),$$

and consequently

$$\frac{\max_{j=1, \dots, p} \frac{X_{ij}}{x_{0j}}}{\min_{j=1, \dots, q} \frac{Y_{ij}}{y_{0j}}} > \frac{(1 + \delta) \frac{Y_{ij_i}}{y_{0j_i}} \theta_C(x_0, y_0)}{\frac{Y_{ij_i}}{y_{0j_i}}} = (1 + \delta) \theta_C(x_0, y_0)$$

Since i is arbitrary, we then obtain $\hat{\theta}_C(x_0, y_0) = \min_{i=1, \dots, n} \frac{\max_{j=1, \dots, p} \frac{X_{ij}}{x_{0j}}}{\min_{j=1, \dots, q} \frac{Y_{ij}}{y_{0j}}} > (1 + \delta) \theta_C(x_0, y_0)$.

Now assume that there exists some $i_0 \in \{1, \dots, n\}$ as well as some $j \in \{1, \dots, q\}$ such that $(X_{i_0}, Y_{i_0}) \in S_{\delta, j}^{x_0, y_0}$. Since then $\tilde{Y}_{i_0}^{(j)} \geq \tilde{y}_0^{(j)}$ we have

$$\min_{l=1, \dots, q} \frac{Y_{i_0 l} y_{0j}}{y_{0l} Y_{i_0 j}} = 1, \quad \text{and hence } \min_{l=1, \dots, q} \frac{Y_{i_0 l}}{y_{0l}} = \frac{Y_{i_0 j}}{y_{0j}}.$$

Moreover, $X_{i_0} \leq (1 + \delta) \frac{Y_{i_0 j}}{y_{0j} a_{\min}^{x_0, y_0}} \theta(x_0, a_{\min}^{x_0, y_0} y_0) x_0$, and therefore

$$\hat{\theta}_C(x_0, y_0) \leq \frac{\max_{l=1, \dots, p} \frac{X_{i_0 l}}{x_{0l}}}{\min_{l=1, \dots, q} \frac{Y_{i_0 l}}{y_{0l}}} \leq \frac{(1 + \delta) \frac{Y_{i_0 j}}{y_{0j} a_{\min}^{x_0, y_0}} \theta(x_0, a_{\min}^{x_0, y_0} y_0)}{\frac{Y_{i_0 j}}{y_{0j}}} = (1 + \delta) \theta_C(x_0, y_0). \quad \blacksquare$$

A.2 Proof of Theorem 3.1

For the Proof of Theorem 3.1 we will need some additional notation. Recall that $\tilde{x}_k^{(k)} = 1$, and let $\tilde{x}_{-k}^{(k)}$ denote the $(p - 1)$ -vector $\tilde{x}_{-k}^{(k)} := (x_1/x_j, \dots, x_{j-1}/x_j, x_{j+1}/x_j, \dots, x_p/x_j)^T$. Similarly let $\tilde{y}_{-j}^{(j)} := (y_1/y_j, \dots, y_{j-1}/y_j, y_{j+1}/y_j, \dots, y_q/y_j)^T$.⁷

⁷If $p = 1$ one may set $\tilde{x}_{-k}^{(k)} = 0$, as well as $\tilde{y}_{-j}^{(j)} = 0$ if $q = 1$. In both cases all terms depending on the respective quantities simply cancel out in further analysis. This will go without saying.

Consider an arbitrary point (x_0, y_0) be a point in $\mathcal{D}$. We first analyze the probability $\Pr((X_i, Y_i) \in S_\delta^{x_0, y_0})$ for small $\delta > 0$.

Recall that $S_\delta^{x_0, y_0} = \bigcup_{j=1}^q S_{\delta, j}^{x_0, y_0}$. We first split the sets $S_{\delta, j}^{x_0, y_0}$ into p different subsets defined by the variables x_j and $\tilde{x}^{(j)}$. For any x there obviously exists some $k \in \{1, \dots, p\}$ such that $\max_{l=1, \dots, p} \frac{x_l}{x_{0l}} = \frac{x_k}{x_{0k}}$. Then $\tilde{x}^{(k)} \leq \tilde{x}_0^{(k)}$, and $(x, y) \in S_{\delta, j}^{x_0, y_0}$ if and only if $x_k \leq (1 + \delta) \frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} \theta(x_0, a_{\min}^{x_0, y_0} y_0) x_{0k}$. In other words, $(x, y) \in S_{\delta, j}^{x_0, y_0}$ if and only if there exists some $k \in \{1, \dots, p\}$ such that $(x, y) \in S_{\delta, j, k}^{x_0, y_0}$, where

$$S_{\delta, j, k}^{x_0, y_0} := \left\{ (x, y) \in \Psi \mid x_k \leq (1 + \delta) \frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} \theta(x_0, a_{\min}^{x_0, y_0} y_0) x_{0k} \text{ and } \tilde{x}^{(k)} \leq \tilde{x}_0^{(k)}, \tilde{y}^{(j)} \geq \tilde{y}_0^{(j)} \right\}. \quad (\text{A.1})$$

Since an observation is in $S_{\delta, j, k}^{x_0, y_0}$ if $\min_{l=1, \dots, q} \frac{Y_{il}}{y_{0l}} = \frac{Y_{ij}}{y_{0j}}$, $\max_{l=1, \dots, p} \frac{X_{il}}{x_{0l}} = \frac{X_{ik}}{x_{0k}}$, and $\Pr(\frac{Y_{ij}}{y_{0j}} = \frac{Y_{ij^*}}{y_{0j^*}}) = 0$, $\Pr(\frac{X_{ik}}{y_{0k}} = \frac{Y_{ik^*}}{y_{0k^*}}) = 0$ for $j \neq j^*$, $k \neq k^*$, we have

$$\Pr((X_i, Y_i) \in S_\delta^{x_0, y_0}) = \sum_{j=1}^q \sum_{k=1}^p \Pr((X_i, Y_i) \in S_{\delta, j, k}^{x_0, y_0}) \quad (\text{A.2})$$

Note that for $\delta = 0$ the set $S_\delta^{x_0, y_0}$ only contains the boundary point $(x_0^*, y_0^*) := (\theta(x_0, a_{\min}^{x_0, y_0} y_0) x_0, a_{\min}^{x_0, y_0} y_0) \in \mathcal{D}^*$ with $\tilde{x}_0^{(k)} = \tilde{x}_0^{(k)}$ and $\tilde{y}_0^{(j)} = \tilde{y}_0^{(j)}$ for all k, j .

Now choose some arbitrary j, k , and for recall the definition of $g_k(\tilde{x}^{(k)}, y)$, $k = 1, \dots, p$, in (3.4).

Now let $[z]_+ := z$ for $z > 0$ and $[z]_+ := 0$ for $z \leq 0$, and set

$$\begin{aligned} \eta_{jk}(\delta, y_j) &:= \left[(1 + \delta) \frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} \theta(x_0, a_{\min}^{x_0, y_0} y_0) x_{0k} - \theta(x_0, \frac{y_j}{y_{0j}} y_0) x_{0k} \right]_+ \\ &= \left[(1 + \delta) \frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} g_k(\tilde{x}_0^{(k)}, a_{\min}^{x_0, y_0} y_0 \tilde{y}_0^{(j)}) - g_k(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}) \right]_+ \end{aligned} \quad (\text{A.3})$$

By assumption, $\frac{\theta(x_0, a_{\min}^{x_0, y_0} y_0 \tilde{y}_0^{(j)}) x_{0k}}{a_{\min}^{x_0, y_0}} = \frac{\theta(x_0, a_{\min}^{x_0, y_0} y_0) x_{0k}}{a_{\min}^{x_0, y_0}} < \frac{\theta(x_0, \frac{y_j}{y_{0j}} y_0) x_{0k}}{y_j / y_{0j}} = \frac{\theta(x_0, y_j \tilde{y}_0^{(j)}) x_{0k}}{y_j / y_{0j}}$ and thus $\frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} g_k(\tilde{x}_0^{(k)}, a_{\min}^{x_0, y_0} y_0 \tilde{y}_0^{(j)}) < g_k(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)})$ if $y_j \neq a_{\min}^{x_0, y_0} y_{0j}$. Hence, for all y_j we obtain $\eta_{jk}(\delta, y_j) = 0$ if $\delta = 0$.

Furthermore, for (X_i, Y_i) with $Y_{ij} = y_j$, $\tilde{X}_i^{(k)} \leq \tilde{x}_0^{(k)}$, and $\tilde{Y}_i^{(j)} \geq \tilde{y}_0^{(j)}$ the asserted monotonicity of the boundary implies that

$$X_{ik} \geq \theta(X_i, Y_i) X_{ik} = g_k(\tilde{X}_i^{(k)}, Y_i) = g_k(\tilde{X}_i^{(k)}, y_j \tilde{Y}_i^{(j)}) \geq g(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}).$$

At this point we need some additional notation. Recall that $\tilde{x}_k^{(k)} = 1$, and let $\tilde{x}_{-k}^{(k)}$ denote the $(p-1)$ -vector $\tilde{x}_{-k}^{(k)} : (x_1/x_j, \dots, x_{j-1}/x_j, x_{j+1}/x_j, \dots, x_p/x_j)^T$. Similarly let $\tilde{y}_{-j}^{(j)} = (y_1/y_j, \dots, y_{j-1}/y_j, y_{j+1}/y_j, \dots, y_q/y_j)^T$.⁸

For given y_j (in some small neighborhood around $a_{min}^{x_0, y_0} y_{0j}$) consider the conditional distribution of $(X_i, Y_{i,-j}^{(j)})$ given y_j . This defines a reduced problem with p inputs and $q-1$ outputs. This distribution is characterized by a conditional density $f(x, y_{-j}^{(j)} | y_j)$, which by our assumption is also twice continuously differentiable. Since $\frac{y_j}{y_{0j}} y_0 = y_j \tilde{y}_0^{(j)}$, the above arguments lead to

$$\begin{aligned} \pi(\delta, y_j) &:= \Pr \left((X_i, Y_i) \in S_{\delta, j, k}^{x_0, y_0} \mid Y_{ij} = y_j \right) \\ &= \Pr \left(X_{ik} \leq (1 + \delta) \frac{y_j}{y_{0j} a_{min}^{x_0, y_0}} \theta(x_0, a_{min}^{x_0, y_0} y_0) x_{0k} \text{ and } \tilde{X}_i^{(k)} \leq \tilde{x}_0^{(k)}, \tilde{Y}_i^{(j)} \geq \tilde{y}_0^{(j)} \mid Y_{ij} = y_j \right) \\ &= \Pr \left(X_{ik} \leq g(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}) + \eta_{jk}(\delta, y_j) \text{ and } \tilde{X}_i^{(k)} \leq \tilde{x}_0^{(k)}, \tilde{Y}_i^{(j)} \geq \tilde{y}_0^{(j)} \mid Y_{ij} = y_j \right) \\ &= \int_{\mathcal{D}(\tilde{x}_0^{(k)}, \tilde{y}_0^{(j)})} I \left(x_k - g_k(\tilde{x}^{(k)}, y_j \tilde{y}^{(j)}) \leq g(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}) - g_k(\tilde{x}^{(k)}, y_j \tilde{y}^{(j)}) + \eta_{jk}(\delta, y_j) \right) f(x, y_{-j}^{(j)} | y_j) dx dy, \end{aligned} \quad (\text{A.4})$$

where $\mathcal{D}(\tilde{x}_0^{(k)}, \tilde{y}_0^{(j)}) := \{(x, y) \in \mathcal{D} \mid \tilde{x}^{(k)} \leq \tilde{x}_0^{(k)}, \tilde{y}^{(j)} \geq \tilde{y}_0^{(j)}\}$, and where $I(\cdot)$ denotes the indicator function. With $f_j(y_j)$ denoting the marginal distribution of the random variable Y_{ij} , the probabilities of interest are then given by

$$\Pr((X_i, Y_i) \in S_{\delta, j, k}^{x_0, y_0}) = \int \pi(\delta, y_j) f_j(y_j) dy_j \quad (\text{A.5})$$

We first analyze the conditional probabilities $\pi(\delta, y_j)$. Obviously,

$$\begin{aligned} I(x_k - g_k(\tilde{x}^{(k)}, y_j \tilde{y}^{(j)}) \leq g(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}) - g_k(\tilde{x}^{(k)}, y_j \tilde{y}^{(j)}) + \eta_{jk}(\delta, y_j)) &= 0 \\ \text{if } g(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}) - g_k(\tilde{x}^{(k)}, y_j \tilde{y}^{(j)}) &< -\eta_{jk}(\delta, y_j) \end{aligned} \quad (\text{A.6})$$

Now consider small values of δ resulting in a very small $\eta_{jk}(\delta, y_j) > 0$. Let $D_{0,k}^x$ denote the $(p-1)$ -vector of partial derivatives of $g_k(\tilde{x}^{(k)}, y_j \tilde{y}^{(j)})$ with respect to $\tilde{x}_{-k}^{(k)}$ evaluated at $(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)})$. Analogously, let $D_{0,k}^y$ denote the $(q-1)$ -vector of partial derivatives of $g_k(\tilde{x}^{(k)}, y_j \tilde{y}^{(j)})$ with respect to $\tilde{y}_{-j}^{(j)}$ evaluated at $(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)})$. By Assumption 3.1 all entries of $D_{0,k}^{(x)}$ are negative, while all entries of $D_{0,k}^{(y)}$ are positive. Furthermore, absolute values are uniformly bounded from below. A Taylor expansion of $g_k(\tilde{x}^{(k)}, y_j \tilde{y}^{(j)})$ at the point $(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)})$

⁸If $p = 1$ one may set $\tilde{x}_{-k}^{(k)} = 0$, as well as $\tilde{y}_{-j}^{(j)} = 0$ if $q = 1$. In both cases all terms depending on the respective quantities simply cancel out in further analysis. This will go without saying.

then leads to

$$\begin{aligned} & |g(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}) - g_k(\tilde{x}^{(k)}, y_j \tilde{y}^{(j)}) + (D_{0,k}^x)^T(\tilde{x}_{-k}^{(k)} - \tilde{x}_{0,-k}^{(k)}) + (D_{0,k}^y)^T(\tilde{y}_{-j}^{(j)} - \tilde{y}_{0,-j}^{(j)})| \\ & \leq C_1(\|\tilde{x}_{-k}^{(k)} - \tilde{x}_{0,-k}^{(k)}\|^2 + \|\tilde{y}_{-j}^{(j)} - \tilde{y}_{0,-j}^{(j)}\|^2), \end{aligned} \quad (\text{A.7})$$

where due to Assumption 3.1 as well as compactness of $\mathcal{D}^*$ the constant $0 < C_1 < \infty$ may be chosen independent of (x_0, y_0) . For $\tilde{x}_{-k}^{(k)} \neq \tilde{x}_{0,-k}^{(k)}$ with $\tilde{x}_{-k}^{(k)} \leq \tilde{x}_{0,-k}^{(k)}$ we necessarily have $(D_{0,k}^x)^T(\tilde{x}_{-k}^{(k)} - \tilde{x}_{0,-k}^{(k)}) > 0$, while $(D_{0,k}^y)^T(\tilde{y}_{-j}^{(j)} - \tilde{y}_{0,-j}^{(j)}) > 0$ for $\tilde{y}_{-j}^{(j)} \neq \tilde{y}_{0,-j}^{(j)}$ with $\tilde{y}_{-j}^{(j)} \geq \tilde{y}_{0,-j}^{(j)}$.

It is now useful to re-parameterize X_i by using the non-negative random variable $\xi_i := x_{ik} - g_k(\tilde{X}_i^{(k)}, Y_i)$ together with $\tilde{X}_{i,-k}^{(k)}$. Obviously, $X_i = (\xi_i + g_k(\tilde{X}_i^{(k)}, Y_i))\tilde{X}_i^{(k)}$. The density f implies a conditional density $\tilde{f}(\xi, \tilde{x}_{-k}^{(k)}, \tilde{y}_{-j}^{(j)} | y_j)$ of the random variables $(\xi_i, \tilde{X}_{i,-k}^{(k)}, Y_{i,-j}^{(j)})$, and by our assumptions also $\tilde{f}$ is twice continuously differentiable on its support. For $d > 0$ define

$$\begin{aligned} \tilde{\mathcal{D}}_{kj}(d) &:= \{(\xi, \tilde{x}^{(k)}, \tilde{y}^{(j)}) | \xi \leq d, 0 \leq \frac{\tilde{x}_{0l} - \tilde{x}_l}{|(D_{0,k}^x)_l|} \leq d, 0 \leq \frac{\tilde{y}_m - \tilde{y}_{0,m}}{(D_{0,k}^x)_m} \leq d \\ & \text{for all } l = 1, \dots, p, l \neq k, m = 1, \dots, q, m \neq j\} \end{aligned} \quad (\text{A.8})$$

The sets $\tilde{\mathcal{D}}_{kj}(d)$ have Lebesgue measure proportional to d^{p+q-1} . Furthermore, for sufficiently small $\eta_{jk}(\delta, y_j) > 0$ one can infer from (A.6) and (A.7) that in (A.4) the set of all (x, y) which a) contribute with a positive indicator function and b) satisfy $(x_k - g_k(\tilde{x}^{(k)}, \tilde{y}^{(j)}), \tilde{x}^{(k)}, \tilde{y}^{(j)}) \notin \tilde{\mathcal{D}}_{kj}(\eta_{jk}(\delta, y_j))$ have Lebesgue measure proportional to $\eta_{jk}(\delta, y_j)^{p+q}$.

Relying on additional Taylor expansions of $\tilde{f}$, for all sufficiently small $\eta_{jk}(\delta, y_j) > 0$ we thus arrive at

$$\begin{aligned} \pi(\delta, y_j) &= \int_{\tilde{\mathcal{D}}_{kj}(\eta_{jk}(\delta, y_j))} I\left(\xi \leq -(D_{0,k}^x)^T(\tilde{x}_{-k}^{(k)} - \tilde{x}_{0,-k}^{(k)}) - (D_{0,k}^y)^T(\tilde{y}_{-j}^{(j)} - \tilde{y}_{0,-j}^{(j)}) + \eta_{jk}(\delta, y_j)\right) \\ & \quad \cdot \tilde{f}(\xi, \tilde{x}_{-k}^{(k)}, \tilde{y}_{-j}^{(j)} | y_j) d\xi d\tilde{x}_{-k}^{(k)} d\tilde{y}_{-j}^{(j)} + O(\eta_{jk}(\delta, y_j)^{p+q}) \\ &= \int_{\tilde{\mathcal{D}}_{kj}(\eta_{jk}(\delta, y_j))} I\left(\xi \leq -(D_{0,k}^x)^T(\tilde{x}_{-k}^{(k)} - \tilde{x}_{0,-k}^{(k)}) - (D_{0,k}^y)^T(\tilde{y}_{-j}^{(j)} - \tilde{y}_{0,-j}^{(j)}) + \eta_{jk}(\delta, y_j)\right) \\ & \quad \cdot \tilde{f}(0, \tilde{x}_{0,-k}^{(k)}, \tilde{y}_{0,-j}^{(j)} | y_j) d\xi d\tilde{x}_{-k}^{(k)} d\tilde{y}_{-j}^{(j)} + O(\eta_{jk}(\delta, y_j)^{p+q}) \\ &= \lambda_{kj}(y_j) \eta_{jk}(\delta, y_j)^{p+q-1} + O(\eta_{jk}(\delta, y_j)^{p+q}), \end{aligned} \quad (\text{A.9})$$

where

$$\begin{aligned} \lambda_{kj}(y_j) &:= \int_{\tilde{\mathcal{D}}_{kj}(1)} I\left(\xi \leq -(D_{0,k}^x)^T(\tilde{x}_{-k}^{(k)} - \tilde{x}_{0,-k}^{(k)}) - (D_{0,k}^y)^T(\tilde{y}_{-j}^{(j)} - \tilde{y}_{0,-j}^{(j)}) + \eta_{jk}(\delta, y_j)\right) \\ & \quad \cdot \tilde{f}(0, \tilde{x}_{0,-k}^{(k)}, \tilde{y}_{0,-j}^{(j)} | y_j) d\xi d\tilde{x}_{-k}^{(k)} d\tilde{y}_{-j}^{(j)}. \end{aligned}$$

We want to note for further use that by our assumptions all second order terms bounded by $O(\eta_{jk}(\delta, y_j)^{p+q})$ converge uniformly for (x_0, y_0) in the interior of $\mathcal{D}$ and y_j in a neighborhood. Moreover $\lambda_{kj}(y_j)$ is a smooth, analytical function of $D_{0,k}^x, D_{0,k}^y$, and $\tilde{f}(0, \tilde{x}_{0,-k}^{(k)}, \tilde{y}_{0,-j}^{(j)}|y_j)$. Since $\mathcal{D}^*$ is compact, $D_{0,k}^x, D_{0,k}^y$, and $\tilde{f}(0, \tilde{x}_{0,-k}^{(k)}, \tilde{y}_{0,-j}^{(j)}|y_j)$ are bounded from below. Therefore, $\lambda_{kj}(y_j)$ is a twice continuously differentiable function of y_j which is uniformly bounded for all (x_0, y_0) in the interior of $\mathcal{D}$.

It remains to analyze (A.5). Therefore, let us first consider the structure of $\eta_{jk}(\delta, y_j)$ in more detail. Assumption 3.2 together $y_0 = y_{0j}\tilde{y}_0^{(j)}$ and smoothness of g_k imply that $\frac{\partial}{\partial a}(\frac{g_k(\tilde{x}_0^{(k)}, a\tilde{y}_0^{(j)})}{a})|_{a=a_{min}^{x_0,y_0}y_{0j}} = \frac{1}{y_{0j}}\frac{\partial}{\partial a}(\frac{g_k(\tilde{x}_0^{(k)}, a\tilde{y}_0^{(j)})}{a})|_{a=a_{min}^{x_0,y_0}y_0} = 0$, and hence

$$\frac{\partial}{\partial y}g_k(\tilde{x}_0^{(k)}, y\tilde{y}_0^{(j)})|_{y=a_{min}^{x_0,y_0}y_{0j}} = \frac{g_k(\tilde{x}_0^{(k)}, a_{min}^{x_0,y_0}y_{0j}\tilde{y}_0^{(j)})}{a_{min}^{x_0,y_0}y_{0j}} \quad (\text{A.10})$$

Abbreviate $D_{0,k}^2 := \frac{\partial^2}{\partial y^2}g_k(\tilde{x}_0^{(k)}, y\tilde{y}_0^{(j)})|_{y=a_{min}^{x_0,y_0}y_{0j}}$ and $D_{0,k}^3 := \frac{\partial^3}{\partial y^3}g_k(\tilde{x}_0^{(k)}, y\tilde{y}_0^{(j)})|_{y=a_{min}^{x_0,y_0}y_{0j}}$. By (A.10) and definition of $\eta_{jk}(\delta, y_j)$ in (A.3) a Taylor expansion of $g_k(\tilde{x}_0^{(k)}, y\tilde{y}_0^{(j)})$ around $y = a_{min}^{x_0,y_0}y_{0j}$ then leads to

$$\begin{aligned} \eta_{jk}(\delta, y_j) = & [\delta g_k(\tilde{x}_0^{(k)}, a_{min}^{x_0,y_0}y_{0j}\tilde{y}_0^{(j)}) + \delta(y_j - a_{min}^{x_0,y_0}y_{0j})\frac{g_k(\tilde{x}_0^{(k)}, a_{min}^{x_0,y_0}y_{0j}\tilde{y}_0^{(j)})}{a_{min}^{x_0,y_0}y_{0j}} \\ & - \frac{D_{0,k}^2}{2}(y_j - a_{min}^{x_0,y_0}y_{0j})^2 - \frac{D_{0,k}^3}{6}(y_j - a_{min}^{x_0,y_0}y_{0j})^3 + R(y_j - a_{min}^{x_0,y_0}y_{0j})]_+, \end{aligned} \quad (\text{A.11})$$

where the remainder term satisfies $|R(y_j - a_{min}^{x_0,y_0}y_{0j})| \leq C_2(y_j - a_{min}^{x_0,y_0}y_{0j})^4$ for some $0 < C_2 < \infty$ which may be chosen independent of (x_0, y_0) . From Assumption 3.2 and compactness of $\mathcal{D}^*$ it also follows that there are constants $0 < C_3 < C^4 < \infty$ such that $0 < C_3 < \frac{D_{0,k}^2}{2} < C^4$ for all (x_0, y_0) . Moreover $|D_{0,k}^3| \leq C_5$ for some $0 < C_2 < \infty$ which may be chosen independent of (x_0, y_0) . For all sufficiently small δ (A.11) implies that there exists some $0 < C_{6,\delta} < \infty$ such that

$$\eta_{jk}(\delta, y_j) = 0 \quad \text{if } y_j - a_{min}^{x_0,y_0}y_{0j} \notin [-C_{6,\delta}\delta^{1/2}, C_{6,\delta}\delta^{1/2}], \quad (\text{A.12})$$

As $\delta \rightarrow 0$, the smallest $C_{6,\delta}$ fulfilling (A.12) will satisfy $C_{6,\delta} = C_{k;(x_0,y_0)} + O(\delta)$, where $C_{k;(x_0,y_0)} := \sqrt{\frac{2g_k(\tilde{x}_0^{(k)}, a_{min}^{x_0,y_0}y_{0j}\tilde{y}_0^{(j)})}{D_{0,k}^2}}$, and $\eta_{jk}(\delta, y_j) = O(\delta^{3/2})$ for all $y_j - a_{min}^{x_0,y_0}y_{0j} \in [-C_{6,\delta}, -C_{k;(x_0,y_0)}]$ or $y_j - a_{min}^{x_0,y_0}y_{0j} \in [C_{k;(x_0,y_0)}, C_{6,\delta}]$

Obviously, $\eta_{jk}(\delta, y_j) \leq C_7 \delta$ for all y_j , where again the constant C_7 may be chosen independent of (x_0, y_0) . Hence, when combining (A.12) with (A.5) and (A.9) we obtain

$$\begin{aligned}
\Pr((X_i, Y_i) \in S_{\delta, j, k}^{x_0, y_0}) &= \int \pi(\delta, y_j) f_j(y_j) dy_j \\
&= \int_{[-C_{6, \delta} \delta^{1/2}, C_{6, \delta} \delta^{1/2}]} \lambda_{kj}(a_{\min}^{x_0, y_0} y_{0j} + u) \eta_{jk}(\delta, a_{\min}^{x_0, y_0} y_{0j} + u)^{p+q-1} f_j(a_{\min}^{x_0, y_0} y_{0j} + u) du + O(\delta^{p+q+1/2}) \\
&= \int_{[-C_{k; (x_0, y_0)} \delta^{1/2}, C_{k; (x_0, y_0)} \delta^{1/2}]} \lambda_{kj}(a_{\min}^{x_0, y_0} y_{0j} + u) \eta_{jk}(\delta, a_{\min}^{x_0, y_0} y_{0j} + u)^{p+q-1} f_j(a_{\min}^{x_0, y_0} y_{0j} + u) du + O(\delta^{p+q+1/2})
\end{aligned} \tag{A.13}$$

where all second order terms bounded by $O(\delta^{p+q+1/2})$ converge uniformly for (x_0, y_0) in the interior of $\mathcal{D}$. We now proceed by a Taylor expansion of $\eta_{jk}(\delta, a_{\min}^{x_0, y_0} y_{0j} + u)^{p+q-1}$ around $(\delta g_k(\tilde{x}_0^{(k)}, a_{\min}^{x_0, y_0} y_{0j} \tilde{y}_0^{(j)}) - \frac{D_{0,k}^2}{2} u^2)^{p+q-1}$. Note that for any constant $C > 0$

$$\begin{aligned}
&\int_{[-C\delta^{1/2}, C\delta^{1/2}]} \left(\delta g_k(\tilde{x}_0^{(k)}, a_{\min}^{x_0, y_0} y_{0j} \tilde{y}_0^{(j)}) - \frac{D_{0,k}^2}{2} u^2 \right)^{p+q-1} u^r du = 0 \text{ for } r = 1 \text{ and } r = 3, \\
&\int_{[-C\delta^{1/2}, C\delta^{1/2}]} \left(\delta g_k(\tilde{x}_0^{(k)}, a_{\min}^{x_0, y_0} y_{0j} \tilde{y}_0^{(j)}) - \frac{D_{0,k}^2}{2} u^2 \right)^{p+q-2} u^r du = 0 \text{ for } r = 1 \text{ and } r = 3.
\end{aligned}$$

Additionally, using Taylor expansions of $\lambda_{kj}(a_{\min}^{x_0, y_0} y_{0j} + u)$ and $f_j(a_{\min}^{x_0, y_0} y_{0j} + u)$ around $y_j = a_{\min}^{x_0, y_0} y_{0j}$ some straightforward computations lead to

$$\begin{aligned}
&\Pr((X_i, Y_i) \in S_{\delta, j, k}^{x_0, y_0}) \\
&= \int_{[-C_{k; (x_0, y_0)} \delta^{1/2}, C_{k; (x_0, y_0)} \delta^{1/2}]} \lambda_{kj}(a_{\min}^{x_0, y_0} y_{0j}) \left(\delta g_k(\tilde{x}_0^{(k)}, a_{\min}^{x_0, y_0} y_{0j} \tilde{y}_0^{(j)}) - \frac{D_{0,k}^2}{2} u^2 \right)^{p+q-1} f_j(a_{\min}^{x_0, y_0} y_{0j}) du \\
&\quad + O(\delta^{p+q+1/2}) \\
&= \underbrace{2f_j(a_{\min}^{x_0, y_0} y_{0j}) \lambda_{kj}(a_{\min}^{x_0, y_0} y_{0j}) g_k(\tilde{x}_0^{(k)}, a_{\min}^{x_0, y_0} y_{0j} \tilde{y}_0^{(j)})^{p+q-1} C_{k; (x_0, y_0)}}_{=: \lambda_{kj}(x_0, y_0)} \delta^{p+q-1/2} + O(\delta^{p+q+1/2}), \tag{A.14}
\end{aligned}$$

Since (A.14) holds for all k, j , and since convergence of the remainder terms is uniform in (x_0, y_0) , we finally arrive at

$$\begin{aligned}
\pi_\delta &:= \Pr((X_i, Y_i) \in S_\delta^{x_0, y_0}) = \lambda(x_0, y_0) \delta^{p+q-1/2} + R_\delta, \\
&\text{where } \lambda(x_0, y_0) = \sum_{k=1}^p \sum_{j=1}^q \lambda_{kj}(x_0, y_0) \text{ and } R_\delta \leq C_8 \delta^{p+q+1/2}
\end{aligned} \tag{A.15}$$

for some constant $C_8 < \infty$ which may be chosen independent of (x_0, y_0) . Also note that by our assumptions on g_j , f and compactness of $\mathcal{D}$ the values $\lambda(x_0, y_0)$ can be uniformly bounded from below.

We are now ready to consider the distribution of our estimator $\widehat{\theta}_C(x_0, y_0)$. For arbitrary $\delta > 0$ let $\delta_n := \delta n^{-\frac{1}{p+q-1/2}}$. We can then infer from Proposition 3.1 that

$$\Pr\left(\frac{\widehat{\theta}_C(x, y | \mathcal{X}_n)}{\theta_C(x_0, y_0)} \leq 1 + \delta_n\right) = \Pr(\mathcal{X}_n \cap S_{\delta_n}^{x_0, y_0} \neq \emptyset) \quad (\text{A.16})$$

Note that (A.15) implies that for $\delta_n = \delta n^{-\frac{1}{p+q-1/2}}$ we obtain $nR_{\delta_n} \leq C_8 \delta n^{-\frac{1}{p+q-1/2}}$. Since the observations (X_i, Y_i) are i.i.d. we can conclude from (A.15) that

$$\begin{aligned} \Pr(\mathcal{X}_n \cap S_{\delta_n}^{x_0, y_0} \neq \emptyset) &= 1 - (1 - \pi_{\delta_n})^n \\ &= 1 - \left(1 - \lambda(x_0, y_0) \frac{\delta^{p+q-1/2} + nR_{\delta_n}}{n}\right)^n \\ &= 1 - \exp(-\lambda(x_0, y_0) \delta^{p+q-1/2} + nR_{\delta_n}) + O\left(\frac{\lambda(x_0, y_0) \delta^{p+q-1/2} + nR_{\delta_n}}{n}\right) \\ &= 1 - \exp(-\lambda(x_0, y_0) \delta^{p+q-1/2})(1 + R_{n, \delta_n}^*) \quad \text{where } R_{n, \delta_n}^* \leq C_9 \delta n^{-\frac{1}{p+q-1/2}} \end{aligned} \quad (\text{A.17})$$

for some constant $0 < C_9 < \infty$. Setting $\lambda_{x_0, y_0} := \lambda(x_0, y_0)^{\frac{1}{p+q-1/2}}$, Assertion (3.11) is an immediate consequence.

Let us now consider the mean of $\widehat{\theta}_C(x_0, y_0)$ in more detail. First note that (A.16) and (A.17) can be rewritten in the form

$$\Pr\left(\frac{\widehat{\theta}_C(x_0, y_0 | \mathcal{X}_n) - \theta_C(x_0, y_0)}{\theta_C(x_0, y_0)} \leq \delta n^{-\frac{1}{p+q-1/2}}\right) = 1 - \exp(-\lambda(x_0, y_0) \delta^{p+q-1/2})(1 + R_{n, \delta}^*) \quad (\text{A.18})$$

Let $\Delta_n = (\frac{(\log n)^2}{\lambda(x_0, y_0)})^{\frac{1}{p+q-1/2}}$ and recall that $\widehat{\theta}_C(x_0, y_0) \leq \theta_{max}$ is bounded. Therefore, for any integer K

$$\begin{aligned} E\left(\widehat{\theta}_C(x_0, y_0 | \mathcal{X}_n)^K I(\widehat{\theta}_C(x_0, y_0)^K \geq \Delta_n n^{-\frac{K}{p+q-1/2}})\right) \\ \leq \theta_{max}^K \Pr\left(\widehat{\theta}_C(x_0, y_0 | \mathcal{X}_n) \geq C(x_0, y_0)(1 + \Delta_n n^{-\frac{1}{p+q-1/2}})\right) \\ = \theta_{max}^K (1 - \Pr\left(\widehat{\theta}_C(x_0, y_0 | \mathcal{X}_n) \leq C(x_0, y_0)(1 + \Delta_n n^{-\frac{1}{p+q-1/2}})\right)) \\ = \exp(-(\log n)^2)(1 + R_{n, \Delta_n}^{p+q-1/2}) = O(n^{-2}) \end{aligned} \quad (\text{A.19})$$

For an arbitrary integer M define $\delta_0 := 0$, as well as $\delta_j := \frac{j\Delta_n}{M}$, $\delta_{j,n} := \delta_j n^{-\frac{1}{p+q-1/2}}$ for $j = 1, \dots, M$, and let again $\kappa = p + q - 1/2$. $\widehat{\theta}_C(x_0, y_0)$ is a continuous random variable, and by (A.18) as well as (A.19) we obtain

$$\begin{aligned}
& \sum_{j=1}^M \delta_{j-1,n} \left(-\exp(-\lambda(x_0, y_0)\delta_j^\kappa)(1 + R_{n,\delta_{j,n}}^*) + \exp(-\lambda(x_0, y_0)\delta_{j-1}^\kappa)(1 + R_{n,\delta_{j-1,n}}^*) \right) \\
& \leq \frac{E \left(\widehat{\theta}_C(x_0, y_0 \mid \mathcal{X}_n) - \theta_C(x_0, y_0) \right)}{\theta_C(x_0, y_0)} \\
& \leq \sum_{j=1}^M \delta_{j,n} \left(-(\exp(-\lambda(x_0, y_0)\delta_j^\kappa)(1 + R_{n,\delta_{j,n}}^*) + \exp(-\lambda(x_0, y_0)\delta_{j-1}^\kappa)(1 + R_{n,\delta_{j-1,n}}^*)) \right) \\
& \quad + O(n^{-2})
\end{aligned} \tag{A.20}$$

When analyzing the terms in (A.20) note that

$$\begin{aligned}
& \sum_{j=1}^M \delta_{j,n} \left(-(\exp(-\lambda(x_0, y_0)\delta_j^\kappa)R_{n,\delta_{j,n}}^* + \exp(-\lambda(x_0, y_0)\delta_{j-1}^\kappa)R_{n,\delta_{j-1,n}}^*) \right) \\
& = \sum_{j=1}^{M-1} (\delta_{j+1,n} - \delta_{j,n}) \exp(-\lambda(x_0, y_0)\delta_j^\kappa)R_{n,\delta_j}^* + \Delta_n \exp(-\lambda(x_0, y_0)\Delta_n^\kappa)R_{n,\Delta_n}^* \\
& \leq C_9 \sum_{j=1}^{M-1} \frac{\Delta_n}{M} n^{-\frac{2}{\kappa}} \exp(-\lambda(x_0, y_0)(\frac{j\Delta_n}{M})^\kappa) (\frac{j\Delta_n}{M})^\kappa \\
& = C_9 n^{-\frac{2}{\kappa}} \int_0^1 \Delta_n \exp(-\lambda(x_0, y_0)(x\Delta_n)^\kappa) (x\Delta_n)^\kappa dx + O(\frac{\Delta_n}{M}) \\
& = C_9 n^{-\frac{2}{\kappa}} \int_0^{\Delta_n} \exp(-\lambda(x_0, y_0)x^\kappa) x^\kappa dx + O(\frac{\Delta_n}{M}) \\
& = O(n^{-\frac{2}{\kappa}}) + O(\frac{\Delta_n}{M}).
\end{aligned} \tag{A.21}$$

By similar arguments,

$$\begin{aligned}
& \sum_{j=1}^M \delta_{j-1,n} \left((\exp(-\lambda(x_0, y_0)\delta_j^\kappa)R_{n,\delta_{j,n}}^* - \exp(-\lambda(x_0, y_0)\delta_{j-1}^\kappa)R_{n,\delta_{j-1,n}}^*) \right) \\
& = O(n^{-\frac{2}{\kappa}}) + O(\frac{\Delta_n}{M}).
\end{aligned} \tag{A.22}$$

On the other hand,

$$\begin{aligned}
& \sum_{j=1}^M \delta_{j,n} (-\exp(-\lambda(x_0, y_0)\delta_j^\kappa) + \exp(-\lambda(x_0, y_0)\delta_{j-1}^\kappa)) \\
&= n^{-\frac{1}{\kappa}} \sum_{j=1}^M \delta_j \int_{\delta_{j-1}}^{\delta_j} \kappa \lambda(x_0, y_0) x^{\kappa-1} (\exp(-\lambda(x_0, y_0)x^\kappa)) dx \\
&= n^{-\frac{1}{\kappa}} \sum_{j=1}^M \int_{\delta_{j-1}}^{\delta_j} \kappa \lambda(x_0, y_0) x^\kappa (\exp(-\lambda(x_0, y_0)x^\kappa)) dx + O\left(\frac{\Delta_n}{M}\right) \\
&= \lambda(x_0, y_0)^{-\frac{1}{\kappa}} \Gamma\left(1 + \frac{1}{\kappa}\right) n^{-\frac{1}{\kappa}} + O\left(\frac{\Delta_n}{M}\right)
\end{aligned} \tag{A.23}$$

as well as

$$\begin{aligned}
& \sum_{j=1}^M \delta_{j-1,n} (-\exp(-\lambda(x_0, y_0)\delta_j^\kappa) + \exp(-\lambda(x_0, y_0)\delta_{j-1}^\kappa)) \\
&= \lambda(x_0, y_0)^{-\frac{1}{\kappa}} \Gamma\left(1 + \frac{1}{\kappa}\right) n^{-\frac{1}{\kappa}} + O\left(\frac{\Delta_n}{M}\right).
\end{aligned} \tag{A.24}$$

Since M is arbitrary, (3.12) follows from (A.20) - (A.24).

Additionally note that (A.19) also holds for $K = 2$. When multiplying probabilities with $\delta_{j,n}^2$ and $\delta_{j-1,n}^2$, respectively, $\frac{E(\widehat{\theta}_C(x_0, y_0 | \mathcal{X}_n) - \theta_C(x_0, y_0))^2}{\theta_C(x_0, y_0)^2}$ can then be analyzed by a straightforward generalization of the arguments in (A.20) - (A.24), leading to $\frac{E(\widehat{\theta}_C(x_0, y_0 | \mathcal{X}_n) - \theta_C(x_0, y_0))^2}{\theta_C(x_0, y_0)^2} = n^{-\frac{2}{\kappa}} \lambda(x_0, y_0)^{-\frac{2}{\kappa}} \Gamma\left(1 + \frac{2}{\kappa}\right) + o(n^{-\frac{2}{\kappa}})$. Assertion (3.13) is an immediate consequence. $\blacksquare$

A.3 Proof of Theorem 3.2

In the following it will be necessary to distinguish between points (x, y) lying in the interior and on the observable boundary of $\mathcal{D}$. Let $\partial\mathcal{D}$ denote the boundary of $\mathcal{D}$, and for $h > 0$ define

$$\mathcal{W}(h) := \left\{ (x_0, y_0) \in \mathcal{D} \left| \max_{k=1, \dots, p} \left| \frac{x_k}{x_{0k}} - 1 \right| \leq h \text{ or } \max_{j=1, \dots, q} \left| \frac{y_j}{y_{0j}} - 1 \right| \leq h \text{ for some } (x, y) \in \partial\mathcal{D} \right. \right\} \tag{A.25}$$

For any $h > 0$, all points in $\mathcal{D} \setminus \mathcal{W}(h)$ are obviously elements in the interior of $\mathcal{D}$. Recall that $\lambda(x, y)$ is uniformly bounded from below, and select some constant $b > 0$ with $b \geq \lambda(x, y)$ for all $(x, y) \in \mathcal{D} \setminus \mathcal{W}(h)$. With $h_n = n^{-\frac{1}{\kappa}} \left(\frac{(\log n)^2}{b} \right)^{\frac{1}{\kappa}}$ the results of Theorem 3.1 thus apply to all $(x, y) \in \mathcal{D} \setminus \mathcal{W}(h_n)$, since a distance of order h_n from the boundary guarantees that the approximations in (A.20) - (A.23) hold uniformly for all $(x, y) \in \mathcal{D} \setminus \mathcal{W}(h_n)$.

Since additionally $\theta_C(x, y)$ are bounded from above, we can conclude that there exists some constant $0 < C_{10} < \infty$ such that for all sufficiently large n and all $(x, y) \in \mathcal{D} \setminus \mathcal{W}(h_n)$

$$|E\left(\widehat{\theta}_C(x, y \mid \mathcal{X}_n)\right) - \theta_C(x, y) - \theta_C(x, y)\lambda(x, y)^{-\frac{1}{\kappa}}\Gamma(1 + \frac{1}{\kappa})n^{-\frac{1}{\kappa}}| \leq C_{10}n^{-\frac{2}{\kappa}} \quad (\text{A.26})$$

This immediately leads to

$$\begin{aligned} & E\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i) \mid (X_i, Y_i) \in \mathcal{D} \setminus \mathcal{W}(h_n)\right) \\ &= E\left(\theta_C(X_i, Y_i)\lambda(X_i, Y_i)^{-\frac{1}{\kappa}} \mid (X_i, Y_i) \in \mathcal{D} \setminus \mathcal{W}(h_n)\right) \Gamma(1 + \frac{1}{\kappa})n^{-\frac{1}{\kappa}} + O_P(n^{-\frac{2}{\kappa}}). \end{aligned} \quad (\text{A.27})$$

For any $(x, y) \in \mathcal{W}(h_n)$ let $(\check{x}_{h_n}, \check{y}_{h_n}) \in \mathcal{D} \setminus \mathcal{W}(h_n)$ denote the element in $\mathcal{D} \setminus \mathcal{W}(h_n)$ closest to (x, y) , i.e. possessing the minimal distance $\max_{k=1, \dots, p} |\frac{\check{x}_{h_n, k}}{x_k} - 1| + \max_{j=1, \dots, q} |\frac{\check{y}_{h_n, j}}{y_j} - 1|$. By Assumption 2.4 (iii) this minimal distance is of order h_n for all $(x, y) \in \mathcal{W}(h_n)$. Since $\mathcal{D}$ is compact and the frontier function is twice continuously differentiable, there exists a constant $0 < C_{11} < \infty$ such that

$$|\theta_C(x, y) - \theta_C(\check{x}_{h_n}, \check{y}_{h_n})| \leq C_{11}h_n \quad (\text{A.28})$$

Obviously,

$$\begin{aligned} \widehat{\theta}_C(\check{x}_{h_n}, \check{y}_{h_n} \mid \mathcal{X}_n) &= \frac{\min_{j=1, \dots, p} \frac{\check{x}_{h_n, j}}{x_j}}{\max_{j=1, \dots, q} \frac{\check{y}_{h_n, j}}{y_j}} \leq \min_{i=1, \dots, n} \frac{\max_{j=1, \dots, p} \frac{X_{ij}}{x_{h_n, j}} \frac{\check{x}_{h_n, j}}{x_j}}{\min_{j=1, \dots, q} \frac{Y_{ij}}{y_{h_n, j}} \frac{\check{y}_{h_n, j}}{y_j}} = \widehat{\theta}_C(x, y \mid \mathcal{X}_n) \\ &\leq \widehat{\theta}_C(\check{x}_{h_n}, \check{y}_{h_n} \mid \mathcal{X}_n) \frac{\max_{j=1, \dots, p} \frac{\check{x}_{h_n, j}}{x_j}}{\min_{j=1, \dots, q} \frac{\check{y}_{h_n, j}}{y_j}} \end{aligned} \quad (\text{A.29})$$

Since $(\check{x}_{h_n}, \check{y}_{h_n}) \in \mathcal{D} \setminus \mathcal{W}(h_n)$ we can infer from (A.26), (A.28) and (A.29) that there exists a further constant $0 < C_{12} < \infty$ such that

$$|E\left(\widehat{\theta}_C(x, y \mid \mathcal{X}_n)\right) - \theta_C(x, y)| \leq C_{12}h_n \quad \text{for all } (x, y) \in \mathcal{W}(h_n) \quad (\text{A.30})$$

Since necessarily $\Pr((X_i, Y_i) \in \mathcal{W}(h_n)) = O(h_n)$ with we can conclude that

$$\begin{aligned} E\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)\right) &= E\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) \mid (X_i, Y_i) \in \mathcal{W}(h_n)\right) \Pr((X_i, Y_i) \in \mathcal{W}(h_n)) \\ &\quad + E\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) \mid (X_i, Y_i) \in \mathcal{D} \setminus \mathcal{W}(h_n)\right) \Pr((X_i, Y_i) \in \mathcal{D} \setminus \mathcal{W}(h_n)) \end{aligned}$$

So that

$$E\left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i)\right) = C_0 n^{-\frac{1}{\kappa}} + O\left(n^{-\frac{2}{\kappa}} (\log n)^{\frac{4}{\kappa}}\right) \quad (\text{A.31})$$

with $C_0 = E \left(\theta_C(X_i, Y_i) \lambda(X_i, Y_i)^{-\frac{1}{\kappa}} \right)$. This proves (3.14). In a similar way, (3.15) follows from the results of Theorem 3.1 together with (A.28) and (A.29).

It remains to show (3.16). For $i = 1, \dots, n$ let $\mathcal{X}_{n,-i} := \{(X_k, Y_k) \mid k = 1, \dots, n, k \neq i\}$ denote the remaining sample of observations to be obtained when eliminating the observation (X_i, Y_i) . For some $0 < b^* < b$ let $h_n^* = n^{-\frac{1}{\kappa}(\frac{\log n}{b^*})^{\frac{1}{\kappa}}}$, and set $\mu_n := E \left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i) \right)$. Since observations are i.i.d., we can infer from (A.29), Proposition 3.1, and the arguments used in the proof Theorem 3.1 that if b^* is sufficiently small, then for all $i, j = 1, \dots, n$

$$\Pr \left(S_{h_n^*}^{X_i, Y_i} \cap \mathcal{X}_n = \emptyset \right) = O(n^{-2}), \quad \Pr \left((X_j, Y_j) \in S_{h_n^*}^{X_i, Y_i} \right) = O(n^{-1}(\log n)^2) \quad (\text{A.32})$$

$$\begin{aligned} & E \left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i) \mid (X_j, Y_j) \notin S_{h_n^*}^{X_i, Y_i} \right) \\ &= E \left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_{n,-j}) - \theta_C(X_i, Y_i) \mid (X_j, Y_j) \notin S_{h_n^*}^{X_i, Y_i} \right) + O_P(n^{-2}) = \mu_n(1 + O_P(n^{-1})) \end{aligned} \quad (\text{A.33})$$

Note that (A.32) and (A.33) also apply for $i = j$. For $i \neq j$, $\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n)$ and $\widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n)$ are thus asymptotically uncorrelated if $(X_i, Y_i) \notin S_{h_n^*}^{X_j, Y_j}$ and $(X_j, Y_j) \notin S_{h_n^*}^{X_i, Y_i}$. We have $\mu_n = O(n^{-\frac{1}{\kappa}})$, and the above arguments imply that for all $i, j = 1, \dots, n, i \neq j$,

$$\begin{aligned} &= E \left((\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i)) (\widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n) - \theta_C(X_j, Y_j)) \right) \\ &= E \left(E \left((\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i)) I((X_j, Y_j) \notin S_{h_n^*}^{X_i, Y_i}) \mid (X_j, Y_j) \right) (\widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n) - \theta_C(X_j, Y_j)) \right) \\ &+ E \left((\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i)) (\widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n) - \theta_C(X_j, Y_j)) \mid (X_j, Y_j) \in S_{h_n^*}^{X_i, Y_i} \right) \cdot \Pr((X_j, Y_j) \in S_{h_n^*}^{X_i, Y_i}) \\ &= \mu_n \cdot E \left(I((X_j, Y_j) \notin S_{h_n^*}^{X_i, Y_i}) (\widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n) - \theta_C(X_j, Y_j)) \right) + O((h_n^*)^2 n^{-1} (\log n)^2) \\ &= \mu_n^2 + O((h_n^*)^2 n^{-1} (\log n)^2) = \mu_n^2 + O(n^{-\frac{\kappa+2}{\kappa}} (\log n)^{\frac{2(\kappa+2)}{\kappa}}) \end{aligned} \quad (\text{A.34})$$

$$\begin{aligned} & (3.16) \text{ now follows from } \text{COV} \left(\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i), \widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n) - \theta_C(X_j, Y_j \mid \mathcal{X}_n) \right) = \\ & E \left((\widehat{\theta}_C(X_i, Y_i \mid \mathcal{X}_n) - \theta_C(X_i, Y_i)) (\widehat{\theta}_C(X_j, Y_j \mid \mathcal{X}_n) - \theta_C(X_j, Y_j \mid \mathcal{X}_n)) \right) - \mu_n^2. \quad \blacksquare \end{aligned}$$

A.4 Proof of Theorem 4.1

Let (x_0, y_0) be an arbitrary point in the interior of $\mathcal{D}$. The proof relies on a straightforward generalization of the arguments used to prove Theorem 3.1. Equations (A.1) and (A.2) remain valid. Assumption 4.1 implies that $\frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} g_k(\tilde{x}_0^{(k)}, a_{\min}^{x_0, y_0} y_{0j} \tilde{y}_0^{(j)}) = g_k(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)})$, and

thus (A.3) becomes

$$\begin{aligned}\eta_{jk}(\delta, y_j) &:= \left[(1 + \delta) \frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} \theta(x_0, a_{\min}^{x_0, y_0} y_0) x_{0k} - \theta(x_0, \frac{y_j}{y_{0j}} y_0) x_{0k} \right]_+ \\ &= \left[(1 + \delta) \frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} g_k(\tilde{x}_0^{(k)}, a_{\min}^{x_0, y_0} y_{0j} \tilde{y}_0^{(j)}) - g_k(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}) \right]_+ = \delta g_k(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)})\end{aligned}\quad (\text{A.35})$$

Probabilities $\pi(\delta, y_j) := \Pr(X_i, Y_i) \in S_{\delta, j, k}^{x_0, y_0} \mid Y_{ij} = y_j$ can now be analyzed similar to (A.4) - (A.9) resulting in

$$\begin{aligned}\pi(\delta, y_j) &= \lambda_{kj}(y_j) \eta_{jk}(\delta, y_j)^{p+q-1} + O(\eta_{jk}(\delta, y_j)^{p+q}) \\ &= \lambda_{kj}(y_j) (\delta g_k(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}))^{p+q-1} + O((\delta g_k(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}))^{p+q})\end{aligned}\quad (\text{A.36})$$

whenever $(\frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} x_0, y_j \tilde{y}_0^{(j)})$ is in the interior of $\mathcal{D}$. By our assumptions (compactness of $\mathcal{D}^*$) there has to exist some $0 < a_0^{x_0, y_0} < a_1^{x_0, y_0} \leq \infty$ such that $(\frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} x_0, y_j \tilde{y}_0^{(j)}) \in \mathcal{D}$ for $y_j \in [a_0^{x_0, y_0}, a_1^{x_0, y_0}]$, while $(\frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} x_0, y_j \tilde{y}_0^{(j)}) \notin \mathcal{D}$ for $y_j \notin [a_0^{x_0, y_0}, a_1^{x_0, y_0}]$. For any (x_0, y_0) in the interior of $\mathcal{D}$ we have $a_1^{x_0, y_0} - a_0^{x_0, y_0} \geq d$, where d corresponds to the constant defined in Assumption 4.1 (i). Then $\tilde{f}(\cdot, y_j) \equiv 0$ and $\lambda_{kj}(y_j) = 0$ for all $y_j \notin [a_0^{x_0, y_0}, a_1^{x_0, y_0}]$. For all $y_j \in (a_0^{x_0, y_0}, a_1^{x_0, y_0})$, $(\frac{y_j}{y_{0j} a_{\min}^{x_0, y_0}} x_0, y_j \tilde{y}_0^{(j)})$ is in the interior of $\mathcal{D}$ and the corresponding values $\lambda_{kj}(y_j)$ are well-defined and bounded from below. Different from the situation considered in Theorem 3.1 $\eta_{jk}(\delta, y_j)$ does not depend on y_j , and hence

$$\begin{aligned}\Pr((X_i, Y_i) \in S_{\delta, j, k}^{x_0, y_0}) &= \int_{a_0^{x_0, y_0}}^{a_1^{x_0, y_0}} \pi(\delta, y_j) f_j(y_j) dy_j \\ &= \int_{a_0^{x_0, y_0}}^{a_1^{x_0, y_0}} \lambda_{kj}(y_j) (\delta g_k(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}))^{p+q-1} f_j(y_j) dy_j + O(\delta^{p+q}) \\ &=: \tilde{\lambda}_{jk}(x_0, y_0) \delta^{p+q-1} + O(\delta^{p+q})\end{aligned}\quad (\text{A.37})$$

We want to note that due to boundary effects even for very small $\delta > 0$ we may have that $\pi(\delta, y_j)$ is considerably smaller than $\lambda_{kj}(y_j) (\delta g_k(\tilde{x}_0^{(k)}, y_j \tilde{y}_0^{(j)}))^{p+q-1}$ for values of y_j very close to $a_0^{x_0, y_0}$ or $a_1^{x_0, y_0}$. But due to our assumptions this can only happen in suitable δ -neighborhoods of $a_0^{x_0, y_0}$ or $a_1^{x_0, y_0}$ such that this effect is covered by the remainder term $O(\delta^{p+q})$. The remainder term converges uniformly for all (x_0, y_0) .

Similar to (A.15) we then arrive at

$$\begin{aligned}\pi_\delta &:= \Pr((X_i, Y_i) \in S_\delta^{x_0, y_0}) = \tilde{\lambda}(x_0, y_0) \delta^{p+q-1} + \tilde{R}_\delta, \\ \text{where } \tilde{\lambda}(x_0, y_0) &= \sum_{k=1}^p \sum_{j=1}^q \tilde{\lambda}_{kj}(x_0, y_0) \text{ and } \tilde{R}_\delta \leq C_{13} \delta^{p+q}\end{aligned}\quad (\text{A.38})$$

for some constant $C_{13} < \infty$ which may be chosen independent of (x_0, y_0) . Also note that by our assumptions on g_j , f and compactness of $\mathcal{D}$ the values $\tilde{\lambda}(x_0, y_0)$ can be uniformly bounded from below.

Based on (A.38) assertion (4.3) now follows from arguments similar to (A.16) and (A.17). The asserted bounds (4.4) and (4.5) for mean and variance can then be obtained by a straightforward generalization of the arguments in (A.18) - (A.24). ■

A.5 Proof of Theorem 4.2

Using $\kappa = p + q - 1$ instead of $\kappa = p + q - 1/2$, the proof follows from the same arguments as those used in the proof of Theorem 3.2.