Rønn-Nielsen, Anders; Kronborg, Dorte; Asmild, Mette

Working Paper

Permutation tests on returns to scale and common production frontiers in nonparametric models

IFRO Working Paper, No. 2022/05

Provided in Cooperation with:
Department of Food and Resource Economics (IFRO), University of Copenhagen

DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

Permutation tests on returns to scale and common production frontiers in nonparametric models

Anders Rønn-Nielsen
Dorte Kronborg
Mette Asmild

2022 / 05

IFRO Working Paper 2022 / 05

Permutation tests on returns to scale and common production frontiers in nonparametric models

Authors: Anders Rønn-Nielsen, Dorte Kronborg, Mette Asmild

JEL-classification: C12, C13, C15, C18, D24, C67

Published: August 2022

See the full IFRO Working Paper series here:
https://ifro.ku.dk/english/publications/ifro_series/working_papers/

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 23
DK 1958 Frederiksberg DENMARK
https://ifro.ku.dk/english/

Permutation tests on returns to scale and common production frontiers in nonparametric models

Anders Rønn-Nielsen & Dorte Kronborg
Center for Statistics, Department of Finance, Copenhagen Business School
and
Mette Asmild
Department of Food and Resource Economics, University of Copenhagen

August 11, 2022

Abstract

Permutation techniques, where one recompute the test statistic over permutations of data, have a long history in statistics and have become increasingly useful as the availability of computational power has increased. Until now, no permutation tests for examining returns to scale assumptions, nor for test of common production possibility sets, when analysing productivity have been available.

We develop three novel tests based on permutations of the observations. The first is a test for constant returns to scale. The other two are, respectively, tests for frontier differences and for whether the production possibility sets are nested. All tests are based on data envelopment analysis (DEA) estimates of efficiencies and are easily implementable. We show that our suggested permutations of the observations satisfy the necessary randomisation assumptions, and hereby that the sizes of the proposed tests are controlled. The advantages of permutation tests are that they are reliable even for relatively small samples and their size can generally be controlled upwards. We further add a lower bound showing that the proposed tests are very close to being exact. Finally, we show that our tests are consistent and illustrate the rate of convergence in simulation studies.

Keywords: Permutation tests, Returns to scale, Comparison of production frontiers, Data envelopment analysis (DEA), Size, Consistency

1 Introduction

When assessing the productivity of firms, or other production units, various assumptions are necessary. Typically, assumptions are made about the returns to scale characteristics of the underlying production technology. Furthermore, when comparing production possibility sets between, for example, privately and publicly owned enterprises, domestic and international companies, or organic and conventional farms, it is often relevant to determine whether the (frontiers for the) technologies coincide.

In this paper we will introduce and verify the use of permutation tests for inference in non-parametric production frontier models. The development of permutation methods in statistics began in the 1920's and the now well-known permutation test for exact inference in 2×2 contingency tables (Fisher (1935)) was among the early contributions. The historical development of permutation methods is described in Berry et al. (2014). Recently such methods have gained popularity due to improved computational possibilities and are mathematically formalized in e.g. Lehmann and Romano (2005).

We will describe and prove how a permutation test can be formulated for testing the hypothesis of constant returns to scale. Returns to scale is typically viewed as being technologically imposed and thus concerning the frontier (or production function) alone. Yet, we here explicitly acknowledge the distinction between the overall distribution of the units within the production possibility set and then the frontier estimated based on those units.

We will show that the suggested permutation test meets the randomization requirements and hereby that the size of the test is upward bounded. Additionally, we determine a lower bound showing that the test is very close to being exact. Also, the test is shown to be consistent. In the Supplementary Material (see Sections 8 and 9 below) the power of the permutation test is investigated through Monte-Carlo simulations.

Furthermore, we introduce and examine two new test statistics which we prove can be used for examining whether the frontier is independent of a discrete dichotomous environmental (or contextual) variable. The result of the first of these tests indicates whether the frontiers are likely to be different (including intersecting frontiers), and the second test supplements the first as it is designed to detect whether one group overall has better production possibilities than the other (nested frontiers). Like the test for returns to scale, these tests are also consistent and almost exact.

An often used method for estimation of productive efficiency of a set of production units is the non-parametric Data Envelopment Analysis (DEA) approach, which estimates the production possibility set as a convex envelopment of the observed set of input and output quantities (Farrell (1957), Charnes et al. (1978)). Our permutations tests utilise DEA's linear programming based estimation of the efficiencies and, indirectly, the production possibility set.

The statistical properties of the estimated efficiencies have been the subject of numerous studies, and Kneip et al. (2015) developed asymptotic results usable for inference about mean efficiencies. The asymptotics were later improved in Simar and Zelenyuk (2020) and Nguyen et al. (2022).

Methods for deciding appropriate technology assumptions have been the subject of several papers, lately Kneip et al. (2016), but Simar and Wilson (2002) and Banker (1996), among others, also address the importance of imposing the correct assumption and discuss

statistics for testing returns to scale. Kneip et al. (2016) and Simar and Wilson (2020) propose tests for returns to scale, with the hypothesis being constant returns to scale (CRS) within a model assuming variable returns to scale (VRS). The test is based on the asymptotic normal distribution of the difference of average sample means of efficiencies.

Most of the developed theory is concentrated on inference for individual or mean efficiencies, whereas the theory on another important issue - statistical tests concerning the production frontiers - is sparse. In Kneip et al. (2016), methods for comparisons of mean efficiencies *and* a common frontier for two independent samples, based on asymptotic normal approximations are developed. The proposed method focuses on testing equality of means for two independent groups when the efficiencies for the two samples are measured relative to a common frontier and when they are measured relative to different frontiers. The suggested test is formally a test for a composite hypothesis, namely whether the groups are facing the same frontier *and* whether the mean efficiencies are the same in the two groups. Consequently, it is not possible to determine whether rejection is due to different frontiers, different mean efficiencies, or both simultaneously.

The importance of the assumption of equal production possibility sets, known as the 'separability condition' is also addressed in e.g. Simar and Wilson (2007), Simar and Wilson (2015), and Daraio et al. (2018). Daraio et al. (2018) develop central limit theorems for means of conditional efficiencies and propose an asymptotic test for the separability condition when conditioning on a continuous environmental variable, later improved in Simar and Wilson (2020). When dealing with a discrete dichotomous environmental variable, the proposed test is basically the same as in the continuous case expect for the bias-correction method.

All the above tests rely on asymptotic results for the estimated efficiencies. The estimates are biased and therefore bootstrap methods for bias corrections are needed for improving the properties of the asymptotic tests. The permutation tests proposed here do not rely on asymptotic results for the estimated efficiencies and the size of the bias need not be estimated, as long as the sample sizes are equal in the tests comparing independent groups (otherwise a simple jackknife is implemented). Furthermore, this implies that the permutation tests also perform well for relatively small sample sizes.

The structure of the paper is as follows: In Section 2 we introduce our notation and the production frontier methodology. The method of inference on scalability (a concept that includes constant returns to scale, c.f. equation 9 below) and proof of the properties of the test are described in Section 3, and Section 4 similarly describes methods for comparisons of frontiers for independent groups. In Section 5 consistency of the proposed tests is proved and Section 6 concludes the paper. In the Appendix (see Section 7 below) we show the theoretical results on permutation test leading to the proposed permutation tests being close to exact. Finally, some additional technical results on consistency, together with a series of Monte Carlo experiments for evaluating the performance of the described tests, are presented in the Supplementary Material (see Sections 8 and 9 below).

2 The non-parametric frontier model

2.1 The production possibility set and returns to scale

Let the vector of input (x) and output (y) quantities be denoted by $(x, y) \in \mathbb{R}_+^{p+q}$. Using standard notation, the feasible set of input-output combinations, i.e. the production possibility set, is

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \,|\, x \text{ can produce } y\}.$$

The production possibility set is assumed to be closed, convex and satisfying strong disposablity in both inputs and outputs. The efficient frontier of Ψ is given by

$$\Psi^\delta = \{(x, y) \in \Psi \,|\, (\gamma^{-1}x, \gamma y) \notin \Psi, \ \forall \gamma > 1\}. \tag{1}$$

The efficiency of a given production unit is often measured by either the Farrell input index or the corresponding output index (Farrell (1957)), with the input index given by

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

and the output index given by

$$\vartheta(x, y \mid \Psi) = \sup\{\vartheta > 0 | \, (x, \vartheta y) \in \Psi\}.$$

If $\theta = 1$ the firm is said to be technically efficient in the input direction, while if $\vartheta = 1$ the firm is technically efficient in the output direction. Otherwise, the firm is referred to as technically inefficient in either the input and/or the output direction. Technical efficiency can alternatively be measured in hyperbolic distances, defined by Färe et al. (1985) as

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

Various assumptions about returns to scale are possible. Here we concentrate on the two most commonly used: Constant returns to scale (CRS) means that the production can be scaled arbitrarily up and down. Formally, that is

$$(x, y) \in \Psi \ \Rightarrow \lambda(x, y) \in \Psi \ \text{ for all } \lambda \geq 0, \tag{2}$$

and we will refer to this as Ψ satisfying CRS. On the other hand, the less restrictive assumption of variable returns to scale (VRS) means that rescaling of all points in the production possibility set is not necessarily possible, i.e. that (2) is not assumed to be satisfied. Constant returns to scale implies that the efficiency is invariant under simultaneous scaling of both inputs and outputs, contrary to the variable returns to scale scenario where the efficiency likely varies when scaling the units. We will denote the input efficiencies when assuming CRS or VRS as θ_{CRS} and θ_{VRS} respectively.

We define the operator $\mathcal{C}(\cdot)$ such that

$$\mathcal{C}(\Psi) = \{(x, y) \in \mathbb{R}_+^{p+q} \mid (x, y) = \eta(\tilde{x}, \tilde{y}) \text{ for some } (\tilde{x}, \tilde{y}) \in \Psi \text{ and } \eta \geq 0\}, \tag{3}$$

i.e. the smallest set containing Ψ that also exhibits CRS. Thus $\mathcal{C}(\Psi) = \Psi$ if (and only if) Ψ exhibits CRS.

For later reference, we collect all assumptions on the production possibility set in Assumption 1 below. We will distinguish between the VRS and the CRS cases.

Assumption 1. *A production possibility set Ψ is said to satisfy Assumption 1 (a) resp. (b) if*

(a) Ψ is closed, convex and satisfies strong disposability in both inputs and outputs, such that if $(x, y) \in \Psi$, and $\tilde{x} \geq x$ and $0 \leq \tilde{y} \leq y$, then $(\tilde{x}, y) \in \Psi$ and $(x, \tilde{y}) \in \Psi$. Furthermore, it satisfies that if $(x, y) \in \Psi$ and $x = 0$, then $y = 0$.

(b) In addition to (a), the production possibility set Ψ is assumed to satisfy CRS.

Under Assumption 1 (a) let $(x, y) \in \Psi \setminus \{(0, 0)\}$, and note that $\theta(x, y \mid \Psi) > 0$. Define

$$\delta(x, y \mid \Psi) = \theta(x, y \mid \Psi) \cdot x,$$

where clearly $(\delta(x, y \mid \Psi), y) \in \Psi^\delta$. Defining

$$z(y) = \|y\|, \qquad v(y) = \frac{y}{\|y\|} = \frac{y}{z(y)}, \qquad w(x) = \frac{x}{\|x\|},$$

we see from the definition of $\theta(x, y \mid \Psi)$ that $\delta(x, y \mid \Psi)$ has the form

$$\delta(x, y \mid \Psi) = \phi(z(y), v(y), w(x)) \tag{4}$$

for some measurable map ϕ. If furthermore Assumption 1 (b) is satisfied, this map in fact has the form

$$\delta(x, y \mid \Psi) = z(y) \cdot \tilde{\phi}(v(y), w(x)), \tag{5}$$

for some measurable map $\tilde{\phi}$.

2.2 Statistical model, estimation and notation

In applications the underlying production possibility set Ψ will typically be unknown and therefore has to be estimated from observed production units. We consider n such production units, represented as random vectors with a distribution that has support on Ψ. In the following Ψ is a production possibility set that satisfies either Assumption 1 (a) or (b). For convenience, we use the notation $\mathbf{X} = (X_1, \ldots, X_n)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$.

Assumption 2. *Let F be a continuous distribution on $\mathbb{R}^{p+q}_+$ with support Ψ, where Ψ is a production possibility set satisfying Assumption 1 (a) or (b). A family of random variables $(\mathbf{X}, \mathbf{Y})$ is said to satisfy Assumption 2 with distribution F, if $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent and identically distributed random vectors such that each (X_i, Y_i) has distribution F.*

A few additional pieces of notation will be convenient. In line with the notation introduced in the end of Section 2.1: Let $Z_i = \|Y_i\|$ and $V_i = \frac{Y_i}{Z_i}$. Furthermore, let $W_i = \frac{X_i}{\|X_i\|}$, $\Theta_i = \theta(X_i, Y_i \mid \Psi)$ and $X_{i,\delta} = \Theta_i \cdot X_i$. Note that $X_{i,\delta}$ is deterministically known from Z_i, V_i, and W_i in a way that is only determined by Ψ^δ since

$$X_{i,\delta} = \phi(Z_i, V_i, W_i). \tag{6}$$

Under Assumption 1 (a) and Assumption 2, the production possibility set Ψ can be estimated by

$$\hat{\Psi}_{VRS}(\mathbf{X}, \mathbf{Y}) = \left\{(x, y) \in \mathbb{R}^{p+q}_+ \,|\exists \omega \in \mathbb{R}^n_+ : x \geq \mathbf{X}\omega, y \leq \mathbf{Y}\omega, \sum \omega_i = 1\right\}.$$

The input efficiency index for (x, y) is estimated by

$$\hat{\theta}_{VRS}(x, y \mid \mathbf{X}, \mathbf{Y}) = \min_{\theta,\omega}\{\theta \,|\, \theta x \geq \mathbf{X}\omega, y \leq \mathbf{Y}\omega, \sum \omega_i = 1, \omega \in \mathbb{R}^n_+\}\,.$$

Analogously, under Under Assumption 1 (b) and Assumption 2, the production possibility set Ψ is estimated by

$$\hat{\Psi}_{CRS}(\mathbf{X}, \mathbf{Y}) = \{(x, y) \in \mathbb{R}^{p+q}_+ \,|\exists \omega \in \mathbb{R}^n_+ : x \geq \mathbf{X}\omega, y \leq \mathbf{Y}\omega\}\,,$$

whereas the input efficiency index for (x, y) is estimated by

$$\hat{\theta}_{CRS}(x, y \mid \mathbf{X}, \mathbf{Y}) = \min_{\theta,\omega}\{\theta \,|\, \theta x \geq \mathbf{X}\omega, y \leq \mathbf{Y}\omega, \omega \in \mathbb{R}^n_+\}\,. \tag{7}$$

Remark 1. *For any set of observations* $\mathbf{X}$ *and* $\mathbf{Y}$*, the additional requirement on* ω *in the definition of* $\hat{\theta}_{VRS}(x, y \mid \mathbf{X}, \mathbf{Y})$ *clearly implies that*

$$\hat{\theta}_{VRS}(x, y \mid \mathbf{X}, \mathbf{Y}) \geq \hat{\theta}_{CRS}(x, y \mid \mathbf{X}, \mathbf{Y})$$

, for all $(x, y) \in \mathbb{R}^{p+q}_+$*.*

Remark 2. *It should be noted that* $\hat{\theta}_{CRS}(x, y \mid \mathbf{X}, \mathbf{Y})$ *is well defined for all* $(x, y) \in \mathbb{R}^{p+q}_+$*. If* $(x, y) \notin \hat{\Psi}_{CRS}(\mathbf{X}, \mathbf{Y})$*, the value will be above 1. On the other hand* $\hat{\theta}_{VRS}(x, y \mid \mathbf{X}, \mathbf{Y})$ *is not necessarily well defined outside of* $\hat{\Psi}_{VRS}(\mathbf{X}, \mathbf{Y})$*.*

In Section 4 we consider observations from two different and independent groups of production units. There the notation will be extended appropriately.

The test procedures described in the following sections will rely on permutation methods. For this we will need to introduce a few additional pieces of notation. For any $d \in \mathbb{N}$ we let S_d denote the set of all permutations of $\{1, \ldots, d\}$, i.e. the set of all bijections of $\{1, \ldots, d\}$ into itself. Any $\sigma \in S_d$ can thus be represented as a reordering $\sigma = (\sigma(1), \ldots, \sigma(d))$ of the numbers in $\{1, \ldots, d\}$.

Furthermore, for any Borel space $\mathcal{X}$ and permutation $\sigma = (\sigma(1), \ldots, \sigma(d)) \in S_d$, we let $\sigma_{\mathcal{X}}$ denote the map $\mathcal{X}^d \to \mathcal{X}^d$ defined by

$$\sigma_{\mathcal{X}}(x_1, \ldots, x_d) = (x_{\sigma(1)}, \ldots, x_{\sigma(d)}),$$

for $x_1, \ldots, x_d \in \mathcal{X}$.

Finally, if $\mathbf{X} = (X_1, \ldots, X_n)$ and $S = \{s_1, \ldots, s_{n'}\} \subseteq \{1, \ldots, n\}$ is a set of indices, then we define

$$\mathbf{X}(S) = (X_{s_1}, \ldots, X_{s_{n'}}). \tag{8}$$

3 Inference on scalability of production

In this section we will work under Assumption 1 (a) and Assumption 2. We refer to the notation introduced after Assumption 2. Recalling, in particular, that $Z_i = \|Y_i\|$, the main interest in this section will be to test the hypothesis of distributional scalability, i.e. that production is independent of the scaling in the sense that

$$H \,:\, Z_i \text{ and } \frac{(X_i, Y_i)}{Z_i} \text{ are stochastically independent for each } i. \tag{9}$$

Remark 3. *In words, the hypothesis means that the output size factor out of the production distribution as an independent component.*

Remark 4. *If the hypothesis (9) is satisfied (in addition to Assumption 1 (a) and Assumption 2), then in particular the production possibility set Ψ satisfies CRS: The independence in (9) gives that the distribution of the variable $\left(Z_i, \frac{(X_i,Y_i)}{Z_i}\right)$ has support in a product set $A \times B$, where $A \subseteq [0,\infty)$ and $B \subseteq S_+^{p+q-1}$ with S_+^{p+q-1} denoting the positive unit sphere in $\mathbb{R}^{p+q}$. From the strong disposability in Assumption 1 (a) we have that in fact $A = [0,\infty)$. We then immediately see that for any possible value (x,y) of (X_i,Y_i), all rescalings $\lambda(x,y)$ can happen as well.*

In Proposition 1 below we give a condition under which the hypotheses (9) and CRS coincide.

Proposition 1. *Assume that Assumption 1 (a) and Assumption 2 are satisfied. Assume additionally for each $i = 1,\dots,n$ that Z_i is independent of (V_i, W_i, Θ_i), then the hypothesis in (9) is equivalent with*

$$H \ : \ \textit{The production possibility set } \Psi \textit{ satisfies CRS.} \tag{10}$$

Note that, the independence assumption in Proposition 1 can be checked graphically, using estimated efficiencies.

Proof. Due to Remark 4 it suffices to show that under the independence assumption of the proposition, hypothesis (10) implies hypothesis (9).

Thus we assume that Z_i is independent of (V_i, W_i, Θ_i) and that Ψ satisfies CRS. Then, recalling the notation Z_i, V_i, W_i from Section 2.2, we have from (5) that $X_{i,\delta}$ has the form

$$X_{i,\delta} = Z_i \cdot \tilde{\phi}(V_i, W_i),$$

such that

$$X_i = \frac{Z_i \cdot \tilde{\phi}(V_i, W_i)}{\Theta_i}.$$

Then,

$$\frac{(X_i, Y_i)}{Z_i} = \Big(\frac{\tilde{\phi}(V_i, W_i)}{\Theta_i}, V_i\Big),$$

from which the independence in (9) is directly obtained from the independence between Z_i and (V_i, W_i, Θ_i). □

A measure of the difference between CRS and VRS (also known as scale efficiency) for any (potential) observation $(x,y) \in \mathbb{R}_+^{p+q}$, is

$$\frac{\hat{\theta}_{VRS}(x,y \mid \mathbf{X}, \mathbf{Y})}{\hat{\theta}_{CRS}(x,y \mid \mathbf{X}, \mathbf{Y})},$$

and the overall difference can be measured by the geometric mean of the n ratios, when calculated in all (X_i, Y_i). That is, the test statistic for scalability $T_{rts}(\mathbf{X}, \mathbf{Y})$ is given by

$$T_{rts}(\mathbf{X}, \mathbf{Y}) = \left(\prod_{i=1}^{n} \frac{\hat{\theta}_{VRS}(X_i, Y_i \mid \mathbf{X}, \mathbf{Y})}{\hat{\theta}_{CRS}(X_i, Y_i \mid \mathbf{X}, \mathbf{Y})}\right)^{1/n}. \tag{11}$$

Remark 5. *Since, by construction, all* $(X_i, Y_i) \in \hat{\Psi}_{VRS}(\mathbf{X}, \mathbf{Y})$*, the test statistic* $T_{rts}(\mathbf{X}, \mathbf{Y})$ *is well defined due the considerations in Remark 2. Additionally, Remark 1 gives that* $T_{rts}(\mathbf{X}, \mathbf{Y}) \geq 1$*.*

The statistic will be used to test the hypothesis (9) but with special emphasis on the hypothesis (10) of CRS: A value of $T_{rts}(\mathbf{X}, \mathbf{Y})$ *significantly above one will be evidence against the hypothesis. To underline this interpretation, as we will see in Section 5.1,* $T_{rts}(\mathbf{X}, \mathbf{Y})$ *converges in probability to a value strictly larger than 1, when the sample size tends to infinity and CRS is not satisfied.*

Remark 6. *The information contained in* $\mathbf{X}$ *and* $\mathbf{Y}$ *enters the test statistic* $T_{rts}(\mathbf{X}, \mathbf{Y})$ *in two different ways: Via the DEA estimation procedure and through the points* $(x, y) \in \mathbb{R}_+^{p+q}$ *the estimators are calculated in. However, in the subsequent permutation procedures we perform the same alterations on both of the (identical) versions of* $(\mathbf{X}, \mathbf{Y})$*. Therefore, the notation for* $T_{rts}(\mathbf{X}, \mathbf{Y})$ *only expresses one unified dependence on* $(\mathbf{X}, \mathbf{Y})$*.*

The distribution of the test statistic $T_{rts}(\mathbf{X}, \mathbf{Y})$ is unknown under the hypothesis (9) of scalability of production. However, the significance of the hypothesis in (9) can be evaluated using permutation methods of the type described in Section 7. For this, Procedure 1 below can be applied to generate test statistics that under H have the same marginal distribution as $T_{rts}(\mathbf{X}, \mathbf{Y})$.

Procedure 1. *Based on the family of variables* $(\mathbf{X}, \mathbf{Y})$ *we follow 1–5 below to derive the variable* $\tilde{T}_{rts}$

1. *Recall the definitions* $Z_i = \|Y_i\|$ *and* $V_i = \frac{Y_i}{Z_i}$*. Furthermore define* $U_i = \frac{X_i}{Z_i}$*.*
2. *Let* $\sigma = (\sigma(1), \ldots, \sigma(n))$ *be randomly chosen according to the uniform distribution on* S_n*.*
3. *Define the new variables* $\tilde{Z}_i$*,* $i = 1, \ldots, n$*, by*
$$(\tilde{Z}_1, \ldots, \tilde{Z}_n) = (Z_{\sigma(1)}, \ldots, Z_{\sigma(n)}).$$
4. *Define* $\tilde{X}_i = \tilde{Z}_i \cdot U_i$ *and* $\tilde{Y}_i = \tilde{Z}_i \cdot V_i$ *for each* $i = 1, \ldots, n$*.*
5. *Define*
$$\tilde{T}_{rts} = T_{rts}(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}),$$
where $\tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n)$ *and* $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_n)$*.*

The following theorem gives that the comparison of $T_{rts}(\mathbf{X}, \mathbf{Y})$ with the empirical distribution of variables generated independently according to Procedure 1 can be used to construct a test of the hypothesis in (9), where the size is controlled. Note that the Theorem requires Assumption 1 (a) but then states the result under hypothesis H in (9). This also covers Assumption 1 (b), but it is usual practice to state the hypothesis in question explicitly in a theorem about the behaviour of a test statistic.

Theorem 1. *Let* $(\mathbf{X}, \mathbf{Y})$ *be a family of random variables satisfying Assumption 1 (a) and Assumption 2. Let* $N \in \mathbb{N}$ *be given, and suppose that* $\tilde{T}^1_{rts}, \ldots, \tilde{T}^N_{rts}$ *are generated independently according to Procedure 1. Define*

$$\tilde{p} = \frac{1}{N+1}\left(1 + \sum_{j=1}^{N} 1_{\{T_{rts}(\mathbf{X},\mathbf{Y}) \leq \tilde{T}^j_{rts}\}}\right).$$

Then under the hypothesis in (9) it holds that

$$\Big(\frac{n! - (N+1)}{n!}\Big)^N \Big(\alpha - \frac{1}{N+1}\Big) \leq P(\tilde{p} \leq \alpha) \leq \alpha, \tag{12}$$

for all $\alpha \in (0, 1)$.

Remark 7. *Theorem 1 provides a significance probability in terms of* $\tilde{p}$ *for testing the hypothesis in (9). The inequalities in (12) mean that the size of the test is controlled: If, for example,* $n = 50$ *and* $N = 1000$ *then the first factor* $\left(\frac{n!-(N+1)}{n!}\right)^N$ *is indistinguishable from* 1 *and the inequalities read*

$$\alpha - 0.001 \leq P(\tilde{p} \leq \alpha) \leq \alpha.$$

Thus the probability of making a type I error is very close to α.

Remark 8. *Obviously, it is a requirement in Theorem 1 that* $n! > N + 1$. *However, the theorem will only be useful if* $n! >> N + 1$. *In cases, where n is very small (however, still large enough to make DEA–estimates well defined), it is instead recommended to replace the randomly chosen permutations in Procedure 1 by systematically going through all permutations. In that case the inequalities in (12) can be simplified and sharpened by using (29) instead of (30) in Theorem 6 in the Appendix.*

Proof of Theorem 1. We use the framework and results from Section 7 in the Appendix. More precisely, we let $\mathbf{G}$ be all transformations of the dataset $(\mathbf{X}, \mathbf{Y})$ that are possible via Procedure 1. Then the number of elements in $\mathbf{G}$ is $M = n!$ since each transformation corresponds to a permutation of $(Z_1, \ldots, Z_n)$.

With this choice of $\mathbf{G}$ and the notation from Procedure 1 we easily find that Assumption 4 from Section 7 in the Appendix is satisfied: For any permutation $\sigma \in S_n$ and $(\tilde{Z}_1, \ldots, \tilde{Z}_n)$ defined accordingly we have, due to the independent and identical distribution of $(X_i, Y_i)_{i=1}^n$, that the variables $(\tilde{Z}_i)_{i=1}^n$ jointly have the same distribution as $(Z_i)_{i=1}^n$. Furthermore, the independence property given by (9) gives that both $(\tilde{Z}_i)_{i=1}^n$ and $(Z_i)_{i=1}^n$ are jointly independent of $(U_i, V_i)_{i=1}^n$ (recalling the definition $U_i = \frac{X_i}{Z_i}$ from Procedure 1). That leads to the conclusion that $(\tilde{X}_i, \tilde{Y}_i)_{i=1}^n$ are independent and distributed with the same distribution F as the original observations (X_i, Y_i).

Furthermore, we have that Assumption 5 from Section 7 in the Appendix is satisfied with $K = 1$: This is simply due to the fact that (X_i, Y_i) has an absolute continuous distribution, making the VRS–estimated frontier and efficiencies different across different permutations in Procedure 1.

Now the result of Theorem 1 follows from Theorem 6 in Section 7 in the Appendix. □

Corollary 1. *Assume that* $(\mathbf{X}, \mathbf{Y})$ *is a family of random variables satisfying Assumption 1 (a) and Assumption 2, such that additionally for each* $i = 1, \ldots, n$ *the variable* Z_i *is independent of* (V_i, W_i, Θ_i)*. Then the conclusion of Theorem 1 holds true under the hypothesis in (10).*

Proof. This is a direct consequence of Proposition 1. □

4 Inference on common frontiers for two groups of firms

In this section we will be concerned with two families of random variables $(\mathbf{X}_1, \mathbf{Y}_1)$ and $(\mathbf{X}_2, \mathbf{Y}_2)$, where $\mathbf{X}_1 = (X_1^1, \ldots, X_{n_1}^1)$, $\mathbf{Y}_1 = (Y_1^1, \ldots, Y_{n_1}^1)$, $\mathbf{X}_2 = (X_1^2, \ldots, X_{n_2}^2)$ and $\mathbf{Y}_2 = (Y_1^2, \ldots, Y_{n_2}^2)$ from two different groups of production units. Without loss of generality we assume that $n_1 \leq n_2$. For these two families we make the Assumption 3 below. In particular, we make the CRS assumption for both groups. Throughout the section we for simplicity, will use the notation $\hat{\theta}$ instead of $\hat{\theta}_{CRS}$.

Assumption 3. *The two families* $(\mathbf{X}_1, \mathbf{Y}_1)$ *and* $(\mathbf{X}_2, \mathbf{Y}_2)$ *are said to satisfy Assumption 3, if*

(i) for each $g = 1, 2$*, the continuous distribution* F_g *on* $\mathbb{R}_+^{p+q}$ *has support* Ψ_g*, where* Ψ_g *satisfies Assumption 1 (b) (CRS).*

(ii) for each $g = 1, 2$*, the variables* $(X_1^g, Y_1^g), \ldots, (X_{n_i}^g, Y_{n_g}^g)$ *are independent and identically distributed random vectors such that each* (X_i^g, Y_i^g) *has distribution* F_g*.*

(iii) $(\mathbf{X}_1, \mathbf{Y}_1)$ *and* $(\mathbf{X}_2, \mathbf{Y}_2)$ *are independent.*

Remark 9. *In Section 4.1 we discuss the implications of relaxing Assumption 1 (b) to 1 (a) in Assumption 3 (i),.*

Based on the notation from Assumption 3 we can now introduce the hypothesis that will be the main focus of this section

$$H \ : \ F_1 = F_2. \tag{13}$$

Remark 10. *In Proposition 2 below we demonstrate that (13) simplifies to a comparison of the two production possibility sets under an additional assumption about the distributions in the two groups. Note, that equality between production possibility sets is equivalent to equality between frontiers.*

Similar to the notation introduced in Section 2.2 we define for each group $g = 1, 2$ and production unit $i = 1, \ldots, n_g$: $Z_i^g = \|Y_i^g\|$, $V_i^g = \frac{Y_i^g}{Z_i^g}$, $W_i^g = \frac{X_i^g}{\|X_i^g\|}$, $\Theta_i^g = \theta(X_i^g, Y_i^g \mid \Psi_g)$, and $X_{i,\delta}^g = \Theta_i^g \cdot X_i^g$. Additionally, $X_{i,\delta}^g$ is known from (Z_i^g, V_i^g, W_i^g), since

$$X_{i,\delta}^g = \phi_g(Z_i^g, V_i^g, W_i^g), \tag{14}$$

where the deterministic function ϕ_g only depends on Ψ_g.

Proposition 2. *Assume that* $(\mathbf{X}_1, \mathbf{Y}_1)$ *and* $(\mathbf{X}_2, \mathbf{Y}_2)$ *satisfy Assumption 3. If, additionally,* $(V_i^1, W_i^1, Z_i^1, \Theta_i^1)$ *and* $(V_i^2, W_i^2, Z_i^2, \Theta_i^2)$ *have the same distribution, then the hypothesis in (13) is equivalent with*

$$H \ : \ \Psi_1 = \Psi_2. \tag{15}$$

Remark 11. *In the test procedure presented in Theorem 2 below it will, in fact, only be required that* $(V_i^1, W_i^1, \Theta_i^1)$ *and* $(V_i^2, W_i^2, \Theta_i^2)$ *have same distribution to ensure that the test concerns the hypothesis in (15), cf. Corollary 2.*

Proof of Proposition 2. Clearly, the requirement given by (13) is stronger than the requirement given by (15), so it suffices to demonstrate that the assumptions of Proposition 2 implies (13).

When assuming (15) we also have that $\phi_1 = \phi_2$ in (14). From (14) combined with the relations $X_i^g = \frac{X_{i,\delta}^g}{\Theta_i^g}$ and $Y_i^g = V_i^g \cdot Z_i^g$, we have a deterministic function H such that

$$(X_i^g, Y_i^g) = H(Z_i^g, V_i^g, W_i^g, \Theta_i^g),$$

for all $g = 1, 2$ and $i = 1, \ldots, n_g$. If also $(V_i^1, W_i^1, Z_i^1, \Theta_i^1)$ and $(V_i^2, W_i^2, Z_i^2, \Theta_i^2)$ have equal distributions, then (X_i^1, Y_i^1) and (X_i^2, Y_i^2) have equal distributions, which is exactly the assumption in (13). □

Below we introduce two methods to test the hypothesis (13). The tests are both designed to evaluate (13) but with a special emphasis on (15). However, the interpretation of the alternative hypothesis is different in the two cases. We will elaborate on the differences subsequently, but generally the test statistic T_{diff}^M defined in Procedure 2 below is designed to detect overall differences between the frontiers, whereas the second test statistic T_{nest}^M is designed to detect whether one of the groups has better production possibilities than the other, i.e. whether one of the technologies is nested within the other.

It should be emphasized that Assumption 3 allows for an unbalanced design, i.e. unequal sample sizes for the two groups of production units. Therefore, the comparison of the two groups of production units will be based on a jackknife procedure using several subsets of equal size.

Suppose that two pairs of random vectors $\mathbf{X} = (X_1, \ldots, X_{n_1})$, $\mathbf{Y} = (Y_1, \ldots, Y_{n_1})$ and $\tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_{n_1})$, $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_{n_1})$, respectively, represent observations from two production groups. Then we use the statistics

$$\begin{aligned} T_{\mathit{diff}}((\mathbf{X}, \mathbf{Y}), (\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})) = &\left(\prod_{i=1}^{n_1} \left(\frac{\max\{\hat{\theta}(X_i, Y_i \mid \mathbf{X}, \mathbf{Y}), \hat{\theta}(X_i, Y_i \mid \tilde{\mathbf{X}}, \tilde{\mathbf{Y}})\}}{\min\{\hat{\theta}(X_i, Y_i \mid \mathbf{X}, \mathbf{Y}), \hat{\theta}(X_i, Y_i \mid \tilde{\mathbf{X}}, \tilde{\mathbf{Y}})\}} \right)^{\frac{1}{2n_1}} \right) \\ &\cdot \left(\prod_{i=1}^{n_1} \left(\frac{\max\{\hat{\theta}(\tilde{X}_i, \tilde{Y}_i \mid \mathbf{X}, \mathbf{Y}), \hat{\theta}(\tilde{X}_i, \tilde{Y}_i \mid \tilde{\mathbf{X}}, \tilde{\mathbf{Y}})\}}{\min\{\hat{\theta}(\tilde{X}_i, \tilde{Y}_i \mid \mathbf{X}, \mathbf{Y}), \hat{\theta}(\tilde{X}_i, \tilde{Y}_i \mid \tilde{\mathbf{X}}, \tilde{\mathbf{Y}})\}} \right)^{\frac{1}{2n_1}} \right), \end{aligned}$$

and

$$T_{\mathit{nest}}((\mathbf{X}, \mathbf{Y}), (\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})) = \left(\prod_{i=1}^{n_1} \left(\frac{\hat{\theta}(X_i, Y_i \mid \mathbf{X}, \mathbf{Y})}{\hat{\theta}(X_i, Y_i \mid \tilde{\mathbf{X}}, \tilde{\mathbf{Y}})} \right)^{\frac{1}{2n_1}} \right) \cdot \left(\prod_{i=1}^{n_1} \left(\frac{\hat{\theta}(\tilde{X}_i, \tilde{Y}_i \mid \mathbf{X}, \mathbf{Y})}{\hat{\theta}(\tilde{X}_i, \tilde{Y}_i \mid \tilde{\mathbf{X}}, \tilde{\mathbf{Y}})} \right)^{\frac{1}{2n_1}} \right),$$

to obtain measures for the difference between the two production possibility sets.

Remark 12. *To explain the expressions, the statistic* $T_{diff}((\mathbf{X},\mathbf{Y}),(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}))$ *is a product of* $2n_1$ *factors that are each of the type*

$$\frac{\max\{\hat{\theta}(X,Y \mid \mathbf{X},\mathbf{Y}),\hat{\theta}(X,Y \mid \tilde{\mathbf{X}},\tilde{\mathbf{Y}})\}}{\min\{\hat{\theta}(X,Y \mid \mathbf{X},\mathbf{Y}),\hat{\theta}(X,Y \mid \tilde{\mathbf{X}},\tilde{\mathbf{Y}})\}},$$

where (X,Y) *is an input–output pair from either* $(\mathbf{X},\mathbf{Y})$ *or* $(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$*. This is a measure of the difference between the efficiencies for* (X,Y) *relative to each of the estimated production possibility sets* $\hat{\Psi}_{CRS}(\mathbf{X},\mathbf{Y})$ *and* $\hat{\Psi}_{CRS}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$*; a value substantially larger than 1 means a difference between the two estimated production possibility sets in that direction, and consequently indicates a difference between the true production possibility sets. By definition, it always hold that* $T_{diff}((\mathbf{X},\mathbf{Y}),(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})) \geq 1$*.*

Similarly, the test statistic $T_{nest}((\mathbf{X},\mathbf{Y}),(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}))$ *consists of* $2n_1$ *factors of the type*

$$\frac{\hat{\theta}(X,Y \mid \mathbf{X},\mathbf{Y})}{\hat{\theta}(X,Y \mid \tilde{\mathbf{X}},\tilde{\mathbf{Y}})},$$

that in the direction of (X,Y) *compares the production possibilities within* $\hat{\Psi}_{CRS}(\mathbf{X},\mathbf{Y})$ *and* $\hat{\Psi}_{CRS}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$*. A value above 1 means better production possibilities within* $\hat{\Psi}_{CRS}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$*. Thereby the value of* $T_{nest}((\mathbf{X},\mathbf{Y}),(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}))$ *makes an overall comparison of the two production possibility sets, such that a value above 1 indicates that the group represented by* $(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$*, on average over all observed directions, has better production possibilities. Conversely, a value below 1 means that the group represented by* $(\mathbf{X},\mathbf{Y})$ *has the best production possibilities.*

Remark 13. *It should be noted that if e.g. the input–output pair* (X,Y) *in Remark 12 are among the variables in* $(\mathbf{X},\mathbf{Y})$*, then due to Remark 2 using CRS–based DEA–estimates ensures that* $\hat{\theta}(X,Y \mid \tilde{\mathbf{X}},\tilde{\mathbf{Y}})$ *is always well defined - also when* $(X,Y) \notin \hat{\Psi}_{CRS}(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$*. In that case, the efficiency score will be greater than 1. Relaxing the CRS–assumption would require estimation of the hyperbolic efficiency measure instead. This is briefly discussed in Section 4.1.*

The test statistics $T_{diff}((\mathbf{X},\mathbf{Y}),(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}))$ and $T_{nest}((\mathbf{X},\mathbf{Y}),(\tilde{\mathbf{X}},\tilde{\mathbf{Y}}))$ are defined such that $\mathbf{X},\mathbf{Y},\tilde{\mathbf{X}},\tilde{\mathbf{Y}}$ all have length n_1 and is therefore not directly applicable when using $(\mathbf{X}_1,\mathbf{Y}_1)$ and $(\mathbf{X}_2,\mathbf{Y}_2)$. To make an evaluation of (13) based on all variables in $(\mathbf{X}_2,\mathbf{Y}_2)$ we use $T^M_{diff}((\mathbf{X}_1,\mathbf{Y}_1),(\mathbf{X}_2,\mathbf{Y}_2))$ and $T^M_{nest}((\mathbf{X}_1,\mathbf{Y}_1),(\mathbf{X}_2,\mathbf{Y}_2))$ defined in Procedure 2 below.

Procedure 2. *Let* $M \in \mathbb{N}$ *be given. Based on* $\mathbf{X}_1,\mathbf{Y}_1,\mathbf{X}_2,\mathbf{Y}_2$ *and for each* $m = 1,\ldots,M$ *we follow*

1. *Draw randomly without replacement, and independently for varying* m*, a vector* $s^m = (s^m_1,\ldots,s^m_{n_1})$ *of length* n_1 *from the set* $\{1,\ldots,n_2\}$*.*

2. *Define (recalling the notation introduced in (8))*

$$\tilde{\mathbf{X}}^m = \mathbf{X}_2(s^m) \quad \textit{and} \quad \tilde{\mathbf{Y}}^m = \mathbf{Y}_2(s^m).$$

Then define

$$T^M_{diff} = T^M_{diff}((\mathbf{X}_1,\mathbf{Y}_1),(\mathbf{X}_2,\mathbf{Y}_2)) = \left(\prod_{m=1}^{M} T_{diff}((\mathbf{X}_1,\mathbf{Y}_1),(\tilde{\mathbf{X}}^m,\tilde{\mathbf{Y}}^m))\right)^{\frac{1}{M}},$$

and

$$T^M_{nest} = T^M_{nest}((\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2)) = \left(\prod_{m=1}^{M} T_{nest}((\mathbf{X}_1, \mathbf{Y}_1), (\tilde{\mathbf{X}}^m, \tilde{\mathbf{Y}}^m)) \right)^{\frac{1}{M}}.$$

Remark 14. *We will assume that all repeated applications of Procedure 2 will rely on the same M randomly chosen subsets. In practice it does not make much difference, but it simplifies the subsequent proofs and it makes the comparison of T^M_{diff} and T^M_{nest} with $\tilde{T}_{diff}$ and $\tilde{T}_{nest}$, respectively, more direct.*

Remark 15. *Note that the purpose of the jackknife described in Procedure 2 above, as opposed to using e.g. $T_{diff}((\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2))$ when $(\mathbf{X}_1, \mathbf{Y}_1)$ and $(\mathbf{X}_2, \mathbf{Y}_2)$ have different lengths, is to make the bias that arises when estimating the frontier in each of the two groups of the same magnitude. This has no effect on the rejection rate under the hypothesis of equal distributions in the two groups – the permutation argument above works both with and without jackknifing. However, the use of jackknifing is of substantial importance for the rejection rate when the hypothesis of equal frontiers (or distributions) is false, i.e. for the power of the test. Without jackknifing, different group sizes will lead to unequal magnitudes of the biases for the two frontiers, and this may neutralize the real difference between them. This effect is illustrated as part of the simulation study in Section 9.2 in the Supplementary material.*

Remark 16. *The two test statistics T^M_{diff} and T^M_{nest} directly inherits the interpretation from Remark 12. Thus T^M_{diff} attaining a value substantially above 1, or T^M_{nest} having a value far from 1, will both be indications that the hypothesis (13) is incorrect.*

Remark 17. *It should be noted that T^M_{diff} and T^M_{nest} are random given $\mathbf{X}_1, \mathbf{Y}_1, \mathbf{X}_2, \mathbf{Y}_2$ due to the randomness in step 1 in Procedure 2. This randomness is easily controlled by adjusting M, and if n_1 and n_2 are close it is indeed possible to replace the M random vectors s^m by the non–random collection of all n_1–subsets of $\{1, \ldots, n_2\}$. If in fact $n_1 = n_2$, then Procedure 2 just means that $M = 1$ and*

$$T^1_{diff}((\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2)) = T_{diff}((\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2)),$$
$$T^1_{nest}((\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2)) = T_{nest}((\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2)).$$

From Remark 16 we have interpretations of the value of T^M_{diff} and T^M_{nest} in relation to the hypothesis (13). However, the distribution of both is unknown under (13). Instead we propose to apply Procedure 3 below to generate test statistics with the same marginal distribution as T^M_{diff} and T^M_{nest} under (13).

Procedure 3. *Let $M \in \mathbb{N}$ be given. Based on $\mathbf{X}_1, \mathbf{Y}_1, \mathbf{X}_2, \mathbf{Y}_2$ we follow*

1. *Define*

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) = (X^1_1, \ldots, X^1_{n_1}, X^2_1, \ldots, X^2_{n_2}),$$
$$\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2) = (Y^1_1, \ldots, Y^1_{n_1}, Y^2_1, \ldots, Y^2_{n_2}).$$

2. *Draw randomly, without replacement, a vector $s^1 = (s^1_1, \ldots, s^1_{n_1})$ of length n_1 from the set $\{1, \ldots, n_1 + n_2\}$.*

3. *Let* $s^2 = (s_1^2, \ldots, s_{n_2}^2\} = \{1, \ldots, n_1 + n_2\} \setminus s^1$.

4. *Define (recalling the notation defined in (8))*

$$\tilde{\mathbf{X}}_1 = \mathbf{X}(s^1), \quad \tilde{\mathbf{Y}}_1 = \mathbf{Y}(s^1), \quad \tilde{\mathbf{X}}_2 = \mathbf{X}(s^2), \quad \tilde{\mathbf{Y}}_2 = \mathbf{Y}(s^2).$$

5. *Based on* $\tilde{\mathbf{X}}_1, \tilde{\mathbf{Y}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{Y}}_2$ *use Procedure 2 (recalling Remark 14) to calculate*

$$\tilde{T}_{\mathit{diff}} = T_{\mathit{diff}}^M((\tilde{\mathbf{X}}_1, \tilde{\mathbf{Y}}_1), (\tilde{\mathbf{X}}_2, \tilde{\mathbf{Y}}_2)),$$

and

$$\tilde{T}_{\mathit{nest}} = T_{\mathit{nest}}^M((\tilde{\mathbf{X}}_1, \tilde{\mathbf{Y}}_1), (\tilde{\mathbf{X}}_2, \tilde{\mathbf{Y}}_2)).$$

Theorem 2. *Let* $(\mathbf{X}_1, \mathbf{Y}_1)$ *and* $(\mathbf{X}_2, \mathbf{Y}_2)$ *be two families of random variables that satisfy Assumption 3. Let* $N \in \mathbb{N}$ *and* $M \in \mathbb{N}$ *be given, and suppose that* $\tilde{T}_{\mathit{diff}}^1, \ldots, \tilde{T}_{\mathit{diff}}^N$ *and* $\tilde{T}_{\mathit{nest}}^1, \ldots, \tilde{T}_{\mathit{nest}}^N$ *are generated independently according to Procedure 3. Define*

$$\tilde{p}_{\mathit{diff}} = \frac{1}{N+1}\left(1 + \sum_{j=1}^{N} 1_{\{T_{\mathit{diff}}^M(\mathbf{X}_1, \mathbf{Y}_1, \mathbf{X}_2, \mathbf{Y}_2) \leq \tilde{T}_{\mathit{diff}}^j\}}\right),$$

and

$$\tilde{p}_{\mathit{nest}} = \frac{1}{N+1}\left(1 + \sum_{j=1}^{N} 1_{\{T_{\mathit{nest}}^M(\mathbf{X}_1, \mathbf{Y}_1, \mathbf{X}_2, \mathbf{Y}_2) \leq \tilde{T}_{\mathit{nest}}^j\}}\right).$$

Then under the hypothesis in (13) it holds that

$$\left(\frac{\binom{n_1+n_2}{n_1} - (N+1)}{\binom{n_1+n_2}{n_1}}\right)^N \left(\alpha - \frac{2}{N+1}\right) \leq P(\tilde{p}_{\mathit{diff}} \leq \alpha) \leq \alpha, \tag{16}$$

and

$$\left(\frac{\binom{n_1+n_2}{n_1} - (N+1)}{\binom{n_1+n_2}{n_1}}\right)^N \left(\alpha - \frac{1}{N+1}\right) \leq P(\tilde{p}_{\mathit{nest}} \leq \alpha) \leq \alpha, \tag{17}$$

for all $\alpha \in (0, 1)$.

Remark 18. *Theorem 2 provides significance probabilities for two tests of the hypothesis (13), which under the additional assumptions in Proposition 2 is equivalent with the hypothesis in (15). The inequalities in (16) and (17) display to what extend the sizes of the two tests are controlled.*

The difference between the tests lies in which part of the alternative hypotheses there is an emphasis on detecting. The test based on the significance probability $\tilde{p}_{\mathit{diff}}$ *is designed to detect overall differences between the two frontiers, while the test based on* $\tilde{p}_{\mathit{nest}}$ *will detect whether the hypothesis is violated in a way, where* $\Psi_1 \subset \Psi_2$*. If the emphasis is on the alternative* $\Psi_1 \supset \Psi_2$ *instead, the inequality in the definition of* $\tilde{p}_{\mathit{nest}}$ *should be turned around. That the two tests, in fact, has emphasis on these alternatives is underlined by the consistency results demonstrated in Section 5.2.*

Proof of Theorem 2. We use the framework and results of Section 7 in the Appendix. Let $\mathbf{G}$ be all transformations of the dataset $(\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2)$ that are possible via Procedure 3, that is, all unordered divisions into two sub–groups of n_1 and n_2 elements, respectively. Then there are $M = \binom{n_1+n_2}{n_1}$ elements in $\mathbf{G}$. That all new datasets $(\tilde{\mathbf{X}}_1, \tilde{\mathbf{Y}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{Y}}_2)$ created by Procedure 3 have the same distribution as $(\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2)$, when assuming the hypothesis in (13), is obvious.

All new divisions of the observations into two sub–groups will, with probability 1, give different values of $\tilde{T}_{nest}$. Thus, in that case and with the notation from Section 7 in the Appendix, we can choose $K = 1$. In the case $n_1 = n_2$ (or in the degenerate case, where all subsets in Procedure 2 are the same), two different applications of Procedure 3, i.e. two different divisions of the observations into sub–groups, may lead to the same value of $\tilde{T}_{diff}$ – namely the two divisions, where the two groups are interchanged. Thus, for $\tilde{T}_{diff}$ we choose $K = 2$.

The desired result now follows from Theorem 6 in Section 7 in the Appendix.

□

Up to now, the CRS assumption has played a role in making the test statistics well defined: this can be circumvented by changing the efficiency measure as described in Section 4.1 below. However, the assumption of CRS can furthermore provide a simplification of the assumptions required in Proposition 2, with the purpose of simplifying the hypothesis (13) under consideration. This is summarized in Corollary 2 below.

Corollary 2. *Assume that the two families* $(\mathbf{X}_1, \mathbf{Y}_1)$ *and* $(\mathbf{X}_2, \mathbf{Y}_2)$ *of random variables satisfy Assumption 3. If the data generating process furthermore is such that* $(V_i^1, W_i^1, \Theta_i^1)$ *and* $(V_i^2, W_i^2, \Theta_i^2)$ *have the same distribution, then the conclusion of Theorem 2 holds true under the hypothesis (15).*

Proof. Similarly to the proof of Proposition 2 we have for $g = 1, 2$ that X_i^g and Y_i^g are given by

$$X_i^g = Z_i^g \frac{H(V_i^g, W_i^g)}{\Theta_i^g} \qquad \text{and} \qquad Y_i^g = Z_i^g \cdot V_i^g, \tag{18}$$

where H is independent of $g = 1, 2$ due to the hypothesis assumption (15). The multiplicative structure of $\hat{\theta}(x, y \mid \mathbf{X}, \mathbf{Y})$ in $x, y, \mathbf{X}, \mathbf{Y}$ (cf. (7)) then gives that for all g, g' and all i the estimate $\hat{\theta}(X_i^g, Y_i^g \mid \mathbf{X}_{g'}, \mathbf{Y}_{g'})$ is independent of $Z_1^1, \ldots, Z_{n_1}^1, Z_1^2, \ldots, Z_{n_2}^2$. Consequently, the test statistics T_{diff}^M and T_{nest}^M will not depend on $Z_1^1, \ldots, Z_{n_1}^1, Z_1^2, \ldots, Z_{n_2}^2$ either.

For $g = 1, 2$, let the vectors $(\mathbf{X}_g^*, \mathbf{Y}_g^*)$ be the family of input–output pairs $(X_i^{g,*}, Y_i^{g,*})$ obtained by using (18) with $(V_i^g, W_i^g, \Theta_i^g, Z_i^g)$ replaced by $(V_i^g, W_i^g, \Theta_i^g, Z_i^{g,*})$, where $Z_i^{g,*}$ for all i, g are independent of each other and of everything else and follow the same (also for varying $g = 1, 2$) continuous distribution concentrated on $[0, \infty)$. Note that $(\mathbf{X}_g^*, \mathbf{Y}_g^*)$ has exactly the same production possibility set as $(\mathbf{X}_g, \mathbf{Y}_g)$ for $g = 1, 2$ and will therefore satisfy the requirements for Proposition 2. So Proposition 2 gives that when Theorem 2 is applied based on the variables $(\mathbf{X}_1^*, \mathbf{Y}_1^*)$ and $(\mathbf{X}_2^*, \mathbf{Y}_2^*)$, the result holds under the hypothesis (15).

On the other hand, due to the considerations above, using the two families $(\mathbf{X}_1^*, \mathbf{Y}_1^*)$ and $(\mathbf{X}_2^*, \mathbf{Y}_2^*)$ in Theorem 2 will lead to exactly the same test statistics and significance probabilities as when using $(\mathbf{X}_1, \mathbf{Y}_1)$ and $(\mathbf{X}_2, \mathbf{Y}_2)$. This concludes the proof. □

4.1 Other efficiency measures

In Section 4 we have until now made Assumption 1 (b) about the production possibility set satisfying CRS. Consequently, the test statistics T^M_{diff} and T^M_{nest} have been based on CRS DEA–estimates of the input efficiency. These could, alternatively, have been based on output-oriented efficiency measures, with trivial modifications (still assuming CRS).

Relaxing the assumptions to only cover Assumption 1 (a) and instead using the VRS input efficiency estimates, could be problematic, since $\hat{\theta}_{VRS}(x, y \mid \mathbf{X}, \mathbf{Y})$ is not necessarily well defined (not even asymptotically), when $(x, y) \notin \hat{\Psi}_{VRS}(\mathbf{X}, \mathbf{Y})$. On the other hand, the hyperbolic efficiency estimate is well defined for all (x, y). Färe et al. (2016) propose a linear programming approach for estimating the hyperbolic efficiency within the DEA framework. All results in Section 4, except Proposition 2 and Corollary 2, are still valid in the case, where input efficiency estimates in T^M_{diff} and T^M_{nest} are replaced by hyperbolic efficiency scores.

5 Consistency

The two subsections below will demonstrate consistency of the test methods introduced in Sections 3 and 4, i.e. that if the hypothesis under study is not satisfied, then the hypothesis will be rejected with increasing probability as the sample size increases.

To keep the presentation as simple as possible, we assume that necessary regularity conditions are satisfied, making various DEA–estimates point wise consistent, such that e.g. $\hat{\theta}_{VRS}(x, y \mid \mathbf{X}, \mathbf{Y}) \xrightarrow{P} \theta(x, y \mid \Psi)$ for all $(x, y) \in \Psi$ as the sample size n increases (see e.g. Kneip et al. (1998) for conditions ensuring such point wise consistency).

5.1 The test on scalability of frontiers

Recall Procedure 1 that constructs a new dataset $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ by a permutation of $(Z_1, \ldots, Z_n)$, where $Z_i = \|Y_i\|$. Theorem 3 below gives that an observation pair $(\tilde{X}, \tilde{Y})$ in $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ satisfies the hypothesis in (9), when the pair is constructed from two different original observations (X_i, Y_i) and (X_j, Y_j); also in the situation, where $(\mathbf{X}, \mathbf{Y})$ does not satisfy (9).

Theorem 3. *Suppose Assumptions 1 (a) and 2 are satisfied. Let $i, j \in \{1, \ldots, n\}$ with $i \neq j$ and define*

$$\overline{X} = Z_j \cdot U_i \qquad \text{and} \qquad \overline{Y} = Z_j \cdot V_i.$$

Defining $\overline{Z} = \|\overline{Y}\|$, and let $\overline{F}$ denote the distribution of $(\overline{X}, \overline{Y})$, and let $\overline{\Psi}$ denote the production possibility set of $\overline{F}$. Then

(i) $\overline{Z}$ and $\frac{(\overline{X},\overline{Y})}{\overline{Z}}$ are independent.

(ii) $\overline{F}$ is independent of n.

(iii) $\overline{\Psi}$ satisfies CRS.

Proof. Statement (i) follows directly from the independence between (X_i, Y_i) and (X_j, Y_j), and (ii) is a result of the fact that the two distributions are independent of n. Statement (iii) follows from combining (i) with Remark 4. □

For a permutation $\sigma \in S_n$ we say that $i \in \{1, \ldots, n\}$ is a fixed point, if $\sigma(i) = i$. Furthermore, σ is a derangement if it has no fixed points. The number of derangements in S_n is denoted $!n$, and it is well–known that

$$!n = \left\lfloor \frac{n!}{e} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ denotes the integer part. The following lemma gives the distribution of the number of fixed points, when a permutation is chosen randomly.

Lemma 1. *Suppose that σ is chosen randomly according to the uniform distribution on S_n. Let N_σ denote the number of fixed points of σ. Then the distribution of N_σ is given by*

$$P(N_\sigma = k) = \frac{!(n-k)}{k! \cdot (n-k)!}, \tag{19}$$

for $k \in \{0, \ldots, n\}$. Furthermore $N_\sigma \xrightarrow{\mathcal{D}} Pois(1)$ as $n \to \infty$. Here $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, and $Pois(1)$ denotes the Poisson distribution with mean 1.

Proof. For any choice of $k \in \{0, \ldots, n\}$ and distinct numbers $m_1, \ldots, m_k \in \{1, \ldots, n\}$, there are $!(n-k)$ permutations in S_n with $m_1, \ldots, m_k$ as fixed points: The permutation of the $n-k$ remaining numbers $\{1, \ldots, n\} \setminus \{m_1, \ldots, m_k\}$ needs to be a derangement. Furthermore for k given, there are $\binom{n}{k}$ possible choices of $m_1, \ldots, m_k$. Since there are $n!$ elements in S_n, we find

$$P(N_\sigma = k) = \frac{!(n-k) \cdot \binom{n}{k}}{n!} = \frac{!(n-k)}{k! \cdot (n-k)!},$$

which is (19).

The second statement of the lemma follows by letting $n \to \infty$ in (19), using that $!n \sim \frac{n!}{e}$. □

We now introduce the notation $(\mathbf{X}_n, \mathbf{Y}_n)$ which is the same as the notation $(\mathbf{X}, \mathbf{Y})$ defined previously, but making the n–dependence clear: I.e. $\mathbf{X}_n = (X_1, \ldots, X_n)$ and $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$, where $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent and identically distributed with common distribution F. Similarly, we assume that $(\overline{X}_1, \overline{Y}_1), \ldots, (\overline{X}_n, \overline{Y}_n)$ are independent and identically distributed with common distribution $\overline{F}$, defined in Theorem 3. We let $\overline{\mathbf{X}}_n = (\overline{X}_1, \ldots, \overline{X}_n)$ and $\overline{\mathbf{Y}}_n = (\overline{Y}_1, \ldots, \overline{Y}_n)$. Finally, we also replace the notation $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ defined in Procedure 1 by $(\tilde{\mathbf{X}}_n^\sigma, \tilde{\mathbf{Y}}_n^\sigma)$ to clarify the dependence on n and the permutation $\sigma \in S_n$.

We now consider the case, where the production possibility set Ψ of the distribution F does not satisfy CRS. In particular, hypothesis (9) is not satisfied, cf. Remark 4. We furthermore assume for both F and $\overline{F}$ that relevant regularity conditions are satisfied such that the DEA–estimates are point wise consistent and the logarithm of the efficiency score is integrable. Then from Lemma 4 in the supplementary material we have for F that

$$\frac{1}{n} \sum_{i=1}^{n} \log \hat{\theta}_{VRS}(X_i, Y_i \mid \mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{P} E \log(\theta(X_1, Y_1 \mid \Psi)), \tag{20}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\log\hat{\theta}_{CRS}(X_i,Y_i\mid\mathbf{X}_n,\mathbf{Y}_n)\xrightarrow{P}E\log(\theta(X_1,Y_1\mid\mathcal{C}(\Psi))). \tag{21}$$

And for $\overline{F}$ that

$$\frac{1}{n}\sum_{i=1}^{n}\log\hat{\theta}_{VRS}(\overline{X}_i,\overline{Y}_i\mid\overline{\mathbf{X}}_n,\overline{\mathbf{Y}}_n)\xrightarrow{P}E\log(\theta(\overline{X}_1,\overline{Y}_1\mid\overline{\Psi})), \tag{22}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\log\hat{\theta}_{CRS}(\overline{X}_i,\overline{Y}_i\mid\overline{\mathbf{X}}_n,\overline{\mathbf{Y}}_n)\xrightarrow{P}E\log(\theta(\overline{X}_1,\overline{Y}_1\mid\overline{\Psi})). \tag{23}$$

Note that $E\log(\theta(X_1,Y_1\mid\Psi))>E\log(\theta(X_1,Y_1\mid\mathcal{C}(\Psi)))$, since $\Psi\subset\mathcal{C}(\Psi)$ where the inclusion is strict due to the assumption that CRS is not satisfied. We also have that the common limit in (22) and (23) equals $E\log(\theta(\overline{X}_1,\overline{Y}_1\mid\mathcal{C}(\overline{\Psi})))$ since $\mathcal{C}(\overline{\Psi})=\overline{\Psi}$.

Theorem 4. *Under Assumption 1 (a), Assumption 2 and the additional assumptions stated above, we have for $\tilde{p}$ defined in Theorem 1 that*

$$\tilde{p}\xrightarrow{P}0,$$

as $n\to\infty$.

Proof. Recalling Procedure 1, we define for a permutation $\sigma\in S_n$ the n new variable pairs $(\tilde{X}_i,\tilde{Y}_i)$ for $i=1,\dots,n$ by

$$\tilde{X}_i=Z_{\sigma(i)}\cdot U_i\qquad\text{and}\qquad\tilde{Y}_i=Z_{\sigma(i)}\cdot V_i.$$

This can be represented by a directed graph $\mathcal{G}=(V,E)$ with vertices $V=\{1,\dots,n\}$ and with a directed edge from j to i if $(\tilde{X}_i,\tilde{Y}_i)$ has the form

$$(\tilde{X}_i,\tilde{Y}_i)=(Z_j\cdot U_i,Z_j\cdot V_i).$$

Note that for each vertex there is exactly one edge leaving and one edge arriving. For the fixed points $m_1,\dots,m_{N_\sigma}$ of σ it is the same edge that leaves and arrives at the vertex. Clearly, the $n-N_\sigma$ vertices $\{1,\dots,n\}\setminus\{m_1,\dots,m_{N_\sigma}\}$ that are not fixed points of σ will form a number of directed cycles.

It is now possible to divide the $n-N_\sigma$ non–fixed points into 3 groups such that no vertex has neighbours that are connected to vertices in the same group. Furthermore, the 3 groups can be chosen such that they each have either $\lfloor\frac{n-N_\sigma}{3}\rfloor$ or $\lfloor\frac{n-N_\sigma}{3}\rfloor+1$ elements respectively: All non–fixed points can, for example, be reordered such that vertices in the same cycle appear consecutively in the ordering, and using this new ordering, the first element is put into the first group, the second element is put into the second, the third into the third, the fourth into the first, and so on. Let $G_\sigma^1,G_\sigma^2,G_\sigma^3$ denote the three groups (recall that everything depends on n as well). Similarly, let $G_\sigma^4=\{m_1,\dots,m_{N_\sigma}\}$ denote the group of fixed points for σ.

Recalling the notation defined in (8), we always have

$$\theta(x,y\mid\overline{\Psi})\le\hat{\theta}_{VRS}(x,y\mid\tilde{\mathbf{X}}_n^\sigma,\tilde{\mathbf{Y}}_n^\sigma)\le\hat{\theta}_{VRS}(x,y\mid\tilde{\mathbf{X}}_n^\sigma(G_\sigma^\ell),\tilde{\mathbf{Y}}_n^\sigma(G_\sigma^\ell)), \tag{24}$$

for $\ell = 1, \ldots, 4$ (and similarly for the CRS estimator).

Now we find the following bounds for the statistic $T_{rts}(\tilde{\mathbf{X}}_n^\sigma, \tilde{\mathbf{Y}}_n^\sigma)$ defined in Procedure 1, using the inequalities in (24)

$$1 \leq T_{rts}(\tilde{\mathbf{X}}_n^\sigma, \tilde{\mathbf{Y}}_n^\sigma) \leq \prod_{\ell=1}^{4} \frac{\prod_{i \in G_\sigma^\ell} \hat{\theta}_{VRS}(\tilde{X}_i, \tilde{Y}_i \mid \tilde{\mathbf{X}}_n^\sigma(G_\sigma^\ell), \tilde{\mathbf{Y}}_n^\sigma(G_\sigma^\ell))^{1/n}}{\prod_{i \in G_\sigma^\ell} \theta(\tilde{X}_i, \tilde{Y}_i \mid \overline{\Psi})^{1/n}}. \tag{25}$$

By the choice of the groups G_σ^ℓ, the variable pairs $(\tilde{X}_i, \tilde{Y}_i)$ within a group are independent. Thus the collections $(\tilde{\mathbf{X}}_n^\sigma(G_\sigma^\ell), \tilde{\mathbf{Y}}_n^\sigma(G_\sigma^\ell))$ and $(\overline{\mathbf{X}}_n(G_\sigma^\ell), \overline{\mathbf{Y}}_n(G_\sigma^\ell))$ have the same distribution, where $(\overline{\mathbf{X}}_n, \overline{\mathbf{Y}}_n)$ and σ are independent. In particular, for $\ell = 1, \ldots, 4$ we find for the nominator on the right hand side of (25) that

$$\prod_{i \in G_\sigma^\ell} \hat{\theta}_{VRS}(\tilde{X}_i, \tilde{Y}_i \mid \tilde{\mathbf{X}}_n^\sigma(G_\sigma^\ell), \tilde{\mathbf{Y}}_n^\sigma(G_\sigma^\ell))^{1/n} \stackrel{\mathcal{D}}{=} \prod_{i \in G_\sigma^\ell} \hat{\theta}_{VRS}(\overline{X}_i, \overline{Y}_i \mid \overline{\mathbf{X}}_n(G_\sigma^\ell), \overline{\mathbf{Y}}_n(G_\sigma^\ell))^{1/n}. \tag{26}$$

For $\ell = 1, 2, 3$ the right hand side of this has the limit

$$\exp\left(\frac{1}{3} E \log(\theta(\overline{X}_1, \overline{Y}_1 \mid \overline{\Psi}))\right), \tag{27}$$

in probability by (22), using that $|G_\sigma^\ell| \sim \frac{n}{3}$ as $n \to \infty$, since $|G_\sigma^4| = N_\sigma = o_P(n)$ due to Lemma 1. Similarly for $\ell = 4$, the limit of the right hand side of (26) is 1 in probability, using again that $|G_\sigma^4| = o_P(n)$ as $n \to \infty$. Moreover, the denominator of (25) is equal in distribution to

$$\prod_{i \in G_\sigma^\ell} \theta(\overline{X}_i, \overline{Y}_i \mid \overline{\Psi})^{1/n},$$

which by a direct application of the law of large numbers is seen to converge in probability to (27) for $\ell = 1, 2, 3$ and to 1 for $\ell = 4$. In total, we have deduced that the upper bound in (25) converges to 1 in probability. Therefore, also

$$T_{rts}(\tilde{\mathbf{X}}_n^\sigma, \tilde{\mathbf{Y}}_n^\sigma) \xrightarrow{P} 1,$$

as $n \to \infty$.

By a similar, but simpler, reasoning without dividing into sub–groups, we obtain that

$$T_{rts}(\mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{P} \frac{\exp\left(E \log(\theta(X_1, Y_1 \mid \Psi)\right)}{\exp\left(E \log(\theta(X_1, Y_1 \mid \mathcal{C}(\Psi)))\right)},$$

which is strictly greater than 1. Therefore,

$$P(T_{rts}(\mathbf{X}_n, \mathbf{Y}_n) \leq T_{rts}(\tilde{\mathbf{X}}_n^\sigma, \tilde{\mathbf{Y}}_n^\sigma)) \to 0,$$

as $n \to \infty$, which in turn gives

$$E\tilde{p} \to 0.$$

From this the desired result is directly obtained. □

5.2 The test on common frontiers

The reasoning for the consistency of the tests proposed in Section 4 is similar to but somewhat simpler than the arguments in Section 5.1. The aim is again to demonstrate that $\tilde{p}_{diff} \xrightarrow{P} 0$ and $\tilde{p}_{nest} \xrightarrow{P} 0$ as $n_1 \to \infty$ (recall that $n_2 \geq n_1$ by assumption), when the hypothesis under investigation is false.

Suppose that the hypothesis in (13) is incorrect in such a way that

$$\begin{aligned} E_0 :=& E\log(\theta(X_1^1, Y_1^1 \mid \Psi_1)) - E\log(\theta(X_1^1, Y_1^1 \mid \Psi_2)) \\ &+ E\log(\theta(X_1^2, Y_1^2 \mid \Psi_1)) - E\log(\theta(X_1^2, Y_1^2 \mid \Psi_2)) \neq 0, \end{aligned}$$

and note that (13) would give that all four terms in E_0 are in fact equal, while a value larger than 0 means that, generally across the directions in the two groups, Ψ_2 is larger than Ψ_1, and similarly a negative value means that Ψ_1 is larger than Ψ_2. Also define

$$\begin{aligned} E_1 :=& E\log(\max\{\theta(X_1^1, Y_1^1 \mid \Psi_1), \theta(X_1^1, Y_1^1 \mid \Psi_2)\}) \\ &- E\log(\min\{\theta(X_1^1, Y_1^1 \mid \Psi_1), \theta(X_1^1, Y_1^1 \mid \Psi_2)\}) \\ &+ E\log(\max\{\theta(X_1^2, Y_1^2 \mid \Psi_1), \theta(X_1^2, Y_1^2 \mid \Psi_2)\}) \\ &- E\log(\min\{\theta(X_1^2, Y_1^2 \mid \Psi_1), \theta(X_1^2, Y_1^2 \mid \Psi_2)\}), \end{aligned}$$

and note that $E_1 > 0$ since $E_0 \neq 0$.

It can be seen that with $\tilde{\mathbf{X}}_1, \tilde{\mathbf{Y}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{Y}}_2$ constructed in Procedure 3 it holds that $(\tilde{X}_i^g, \tilde{Y}_i^g)$ are independent and have the same distribution across $g = 1, 2$ and varying observation indices i. Let the common distribution be denoted $\overline{F}$ and let $\overline{\Psi}$ denote the corresponding production possibility set.

Under suitable regularity conditions, we have point wise consistency and with another reference to Lemma 4 in the supplementary material, we can obtain convergences similar to (20)–(23). From such convergences we find that

$$\tilde{T}_{nest} \xrightarrow{P} 1 \qquad \text{and} \qquad \tilde{T}_{diff} \xrightarrow{P} 1,$$

while

$$T_{nest} \xrightarrow{P} \exp(E_0) \qquad \text{and} \qquad T_{diff} \xrightarrow{P} \exp(E_1),$$

as $n \to \infty$. This directly gives that $\tilde{p}_{diff} \xrightarrow{P} 0$ as $n \to \infty$.

If $E_0 > 0$, we similarly have that $\tilde{p}_{nest} \xrightarrow{P} 0$, while $\tilde{p}_{nest} \xrightarrow{P} 1$ if $E_0 > 0$. This demonstrates that $\tilde{p}_{nest}$ can be used to test the one–sided alternative corresponding to $E_0 > 0$, i.e. that group 2 has better production possibilities than group 1.

6 Concluding remarks

The test for returns to scale presented in Section 3 is, in its current form, testing the difference between CRS and VRS. This could, just as well, be (re)formulated as a test comparing non-decreasing returns to scale (NDRS) or non-increasing returns to scale (NIRS) to CRS. Furthermore, the theory is presented in the case of input-oriented productivity measures, but can, with appropriate modifications, be adapted to the case of output-orientation. Modifying the tests comparing frontiers for independent groups, from CRS to NDRS and NIRS, is left for future research.

APPENDIX

7 Permutation tests

This appendix is a both condensed and extended version of Lehmann and Romano (2005, Section 15.2) and displays definitions and relevant results concerning permutation tests, which are applied throughout the paper.

We consider a measurable space $(\Omega, \mathcal{F}, P)$ with $P \in \mathcal{P}$, where $\mathcal{P}$ is a family of probability measures on $(\Omega, \mathcal{F})$. We think of data as an observation of a random object X defined on $(\Omega, \mathcal{F}, P)$ with values in a sample space $\mathcal{X}$. Focus will be on the null hypothesis

$$H \ : \ P \in \mathcal{P}_0, \tag{28}$$

where $\mathcal{P}_0 \subset \mathcal{P}$ is a subset. Now let $\mathbf{G}$ be a group of transformations $g : \mathcal{X} \to \mathcal{X}$ with a finite number M of elements. The following, that is known as the *randomization hypothesis*, will be a crucial assumption when constructing permutation tests.

Assumption 4. *It holds that*

$$g(X)(P) = X(P) \quad \textit{for all } g \in \mathbf{G}$$

for all $P \in \mathcal{P}_0$.

In our applications, $\mathbf{G}$ will be all possible permutations of a specified type of the data set X, and Assumption 4 will be a requirement that such permutations cannot change the distribution of X under the null hypothesis H.

For each $x \in \mathcal{X}$ we define the $\mathbf{G}$-orbit set $\mathbf{G}^x$ as

$$\mathbf{G}^x = \{g(x) \ : \ g \in \mathbf{G}\}.$$

Across varying values of x, the sets $\mathbf{G}^x$ form a partition of $\mathcal{X}$. Let $\mathcal{G}$ be the σ-algebra generated by these sets.

Let furthermore $T : \mathcal{X} \to \mathbb{R}$ be a measurable function; making $T(X)$ a test statistic. In principle, T can be any measurable function, but to make the test procedure presented below useful, it should have the property that $T(X)$ has larger values under P, when $P \notin \mathcal{P}_0$ as compared to $P \in \mathcal{P}_0$. We have the following theorem describing the conditional distribution of $T(X)$ given the σ–algebra $\mathcal{G}$. The proof is found directly in Lehmann and Romano (2005, Section 15.2).

Theorem 5. *Assume, with the notation introduced above that Assumption 4 is satisfied. For any* $P \in \mathcal{P}_0$ *and Borel–measurable subset* A *of* $\mathbb{R}$ *it holds that*

$$P(T(X) \in A \mid \mathcal{G}) = \frac{1}{M} \sum_{g \in \mathbf{G}} 1_{\{T(g(X)) \in A\}}.$$

If the M values of $T(g(X))$ are distinct for varying $g \in \mathbf{G}$, then the conditional distribution of $T(X)$ given $\mathcal{G}$ is particularly simple; it is the uniform distribution on $T(\mathbf{G}^X) = \{T(g(X)) \ : \ g \in \mathbf{G}\}$.

In contrast to this, we will face situations, where $T(g(X)) = T(\tilde{g}(X))$ can naturally happen for some $g \neq \tilde{g}$. There will, however, be limits for how many different g that gives the same value of $T(g(X))$. More precisely, the following assumption will be satisfied in our applications, typically with $K = 1$ or $K = 2$.

Assumption 5. *There is a constant $K \in \mathbb{N}$ such that for all $g \in \mathbf{G}$*

$$P\Big(\sum_{g' \in \mathbf{G}} 1_{\{T(g'(X))=T(g(X))\}} \leq K \} \Big) = 1.$$

We now introduce a p–value for testing the null hypothesis H in (28). This is defined as

$$\hat{p} = \frac{1}{M} \sum_{g \in \mathbf{G}} 1_{\{T(g(X)) \geq T(X)\}}.$$

In practice, M will be very large, making this expression intractable. For a suitable and smaller number B, a stochastic approximation of $\hat{p}$ can be obtained by sampling the transformations $g_1, \ldots, g_{B-1}$ independently (also jointly independently of X) and identically distributed, such that each is chosen according to the uniform distribution on $\mathbf{G}$. Then the approximation is obtained as

$$\tilde{p} = \frac{1}{B} \left[1 + \sum_{i=1}^{B-1} 1_{\{T(g_i(X)) \geq T(X)\}} \right].$$

Now fix the nominal level $0 < \alpha < 1$. We have the following theorem giving that the two tests obtained by rejecting the null hypothesis, when $\hat{p} \leq \alpha$ and $\tilde{p} \leq \alpha$, respectively, are both very close to be of level α, i.e. being exact. The closeness will be controlled by M, K and B, where M in our applications will depend on the size of the data set and K is a result of the specific permutation method related to the type of the data X. The precision parameter B directly dictates the computational effort due to simulated permutations.

Theorem 6. *Assume, with the notation introduced above, that Assumptions 4 and 5 are satisfied. Then*

$$\alpha - \frac{K}{M} \leq P(\hat{p} \leq \alpha) \leq \alpha, \tag{29}$$

and

$$\left(\frac{M-B}{M} \right)^{B-1} \left(\alpha - \frac{K}{B} \right) \leq P(\tilde{p} \leq \alpha) \leq \alpha \tag{30}$$

for all $0 < \alpha < 1$.

The proof of the theorem is found below.

Remark 19. *It will, in fact, be clear from the proof below, that the upper bounds in (29) and (30) are also satisfied without Assumption 5. We will, however, only consider situations where Assumption 5 is satisfied for a relatively small number K.*

For the lower bound in (30) to be useful, B should be much smaller than M, but still large enough to make K/B small. In our applications, the lower bound will typically be very close to α, which makes a test procedure based on $\tilde{p}$ close to being exact. If, on the other hand, B and M are of the same magnitude, it would be easier to go through all permutations in $\mathbf{G}$ and use $\hat{p}$ directly instead of using the stochastic approximation.

Remark 20. *Suppose instead, in the definition of $\tilde{p}$, that the transformations $g_1, \ldots, g_{B-1}$ are chosen randomly without replacement from $\mathbf{G} \setminus \{\mathrm{id}\}$, where $\mathrm{id} : \mathcal{X} \to \mathcal{X}$ is the identity map. Then $\tilde{p}$ as defined in (30) satisfies*

$$\alpha - \frac{K}{B} \leq P(\tilde{p} \leq \alpha) \leq \alpha. \tag{31}$$

The argument for this follows from a simplification of the second part of the proof of Theorem 6. Though the lower bound in (31) is more precise than the lower bound in (30), it will, in most practical applications, be much simpler to implement sampling $g_1, \ldots, g_{B-1}$ with than without replacement.

Before we proceed to the proof of Theorem 6 we state and prove the following lemmas, that provide parts of the arguments for (30). For real random variables $Z_1, \ldots, Z_B$ we will use the notation $Z_{(i)}$ for the i'th ordered variable, i.e.

$$Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(B)}.$$

Furthermore, we write $Z = (Z_1, \ldots, Z_B)$ and $Z_{()} = (Z_{(1)}, \ldots, Z_{(B)})$. For a vector $z = (z_1, \ldots, z_B)$ we also introduce the i'th coordinate projection $\hat{Z}_i : \mathbb{R}^B \to \mathbb{R}$ as

$$\hat{Z}_i(z) = z_i,$$

and the i'th ordered coordinate projection $\hat{Z}_{(i)} : \mathbb{R}^B \to \mathbb{R}$ as

$$\hat{Z}_{(i)}(z) = z_{(i)}.$$

Lemma 2. *Assume that $Z_1, \ldots, Z_B$ are random variables with values in $\mathbb{R}$, and suppose that the distribution of $Z = (Z_1, \ldots, Z_B)$ is invariant to permutations, i.e.*

$$\big(\sigma_{\mathcal{X}}(Z)\big)(P) = Z(P),$$

for all $\sigma \in S^B$, with the notation from Section 2.2. Then, for any Borel–measurable set $A \subseteq \mathbb{R}$, it holds that

$$P(Z_B \in A \mid Z_{()}) = \frac{1}{B} \sum_{i=1}^{B} 1_{\{Z_{(i)} \in A\}}. \tag{32}$$

Proof. The right hand side of (32) is clearly $Z_{()}$–measurable, so it suffices to show that

$$\int_{(Z_{(1)} \in A_1, \ldots, Z_{(B)} \in A_B)} \frac{1}{B} \sum_{i=1}^{B} 1_{\{Z_{(i)} \in A\}} \, \mathrm{d}P = P(Z_B \in A, Z_{(1)} \in A_1, \ldots, Z_{(B)} \in A_B),$$

for all Borel measurable sets $A_1, \ldots, A_B$ in $\mathbb{R}$. First, we notice that

$$\frac{1}{B} \sum_{i=1}^{B} 1_{\{Z_{(i)} \in A\}} = \frac{1}{B} \sum_{i=1}^{B} 1_{\{Z_i \in A\}} = \frac{1}{B!} \sum_{\sigma \in S^B} 1_{\{Z_{\sigma(B)} \in A\}}.$$

Therefore,

$$\begin{aligned}
&\int_{(Z_{(1)}\in A_1,\ldots,Z_{(B)}\in A_B)} \frac{1}{B}\sum_{i=1}^{B} 1_{\{Z_{(i)}\in A\}}\,\mathrm{d}P \\
&= \frac{1}{B!}\sum_{\sigma\in S^B} P(Z_{\sigma(B)}\in A, Z_{(1)}\in A_1,\ldots,Z_{(B)}\in A_B) \\
&= \frac{1}{B!}\sum_{\sigma\in S^B} P(\hat{Z}_B(\sigma_{\mathcal{X}}(Z))\in A, \hat{Z}_{(1)}(Z)\in A_1,\ldots,\hat{Z}_{(B)}(Z)\in A_B) \\
&= \frac{1}{B!}\sum_{\sigma\in S^B} P(\hat{Z}_B(\sigma_{\mathcal{X}}(Z))\in A, \hat{Z}_{(1)}(\sigma_{\mathcal{X}}(Z))\in A_1,\ldots,\hat{Z}_{(B)}(\sigma_{\mathcal{X}}(Z))\in A_B) \\
&= \frac{1}{B!}\sum_{\sigma\in S^B} P(\hat{Z}_B(Z)\in A, \hat{Z}_{(1)}(Z)\in A_1,\ldots,\hat{Z}_{(B)}(Z)\in A_B) \\
&= \frac{1}{B!}\sum_{\sigma\in S^B} P(Z_B\in A, Z_{(1)}\in A_1,\ldots,Z_{(B)}\in A_B),
\end{aligned}$$

as desired, where the last but one equality follows from the fact that $Z(P)$ is invariant to permutations. □

Lemma 3. *Assume that $Z_1,\ldots,Z_B$ are random variables with values in $\mathbb{R}$, and suppose that the distribution of $(Z_1,\ldots,Z_B)$ is invariant to permutations. Define*

$$q = \frac{1}{B}\left[1+\sum_{i=1}^{B-1} 1_{\{Z_i\geq Z_B\}}\right].$$

Then, for all $0<\alpha<1$

$$P(q\leq\alpha)\leq\alpha. \tag{33}$$

Furthermore, suppose there is a constant $K\in\mathbb{N}$ such that

$$P\Big(\textit{for all } i\in\{1,\ldots,B\}\ :\ \sum_{j=1}^{B} 1_{\{Z_j=Z_i\}}\leq K\}\Big)=1. \tag{34}$$

Then for all $0<\alpha<1$

$$P(q\leq\alpha)\geq\alpha-\frac{K}{B}. \tag{35}$$

Proof. First, we focus on (33). From Lemma 2 we have that, conditioned on $Z_{()}$, the distribution of Z_B is the uniform distribution on the B values in $Z_{()}$ counted with multiplicity. Note also that clearly,

$$q=\frac{1}{B}\sum_{i=1}^{B} 1_{\{Z_{(i)}\geq Z_B\}}.$$

Given $Z_{()}$, the event that $q\leq\alpha$ corresponds to Z_B being among the αB largest elements in $\{Z_{(1)},\ldots,Z_{(B)}\}$. Now choose $Z_0\in\{Z_{(1)},\ldots,Z_{(B)}\}$ such that

$$\sum_{i=1}^{B} 1_{\{Z_{(i)}>Z_0\}}\leq\alpha B \qquad\text{and}\qquad \sum_{i=1}^{B} 1_{\{Z_{(i)}<Z_0\}}<(1-\alpha)B.$$

Then $q \leq \alpha$ corresponds to $Z_B > Z_0$. Since at most αB elements in $\{Z_{(1)}, \ldots, Z_{(B)}\}$ satisfy this, we immediately have that

$$P(q \leq \alpha \mid Z_{()}) \leq \alpha.$$

Furthermore, the assumption (34) gives that at most K elements in $\{Z_{(1)}, \ldots, Z_{(B)}\}$ are equal to Z_0. Thus under (34) we, in fact, have

$$\sum_{i=1}^{B} 1_{\{Z_{(i)} > Z_0\}} > \alpha B - K,$$

such that

$$P(q \leq \alpha \mid Z_{()}) > \alpha - \frac{K}{B}.$$

Now the results (33) and (35) follow from taking expectation. □

Proof of Theorem 6. From Theorem 5 we have the conditional distribution of $T(X)$ given $\mathcal{G}$ as the uniform distribution on $T(\mathbf{G}^X) = \{T(g(X)) : g \in \mathbf{G}\}$, counted with multiplicity. Following the same line of arguments as in the proof of Lemma 3, it is easily seen that under Assumption 5 it holds that

$$\alpha - \frac{K}{M} \leq P(\hat{p} \leq \alpha \mid \mathcal{G}) \leq \alpha,$$

from which (29) follows.

Next, we turn to the proof of (30). We recall that $g_1, \ldots, g_{B-1}$ are drawn independently from $\mathbf{G}$ according to the uniform distribution, and also independently of X. Clearly, from the invariance in Assumption 4, we have

$$\begin{aligned}&\big(T(g_1(X)), \ldots, T(g_{B-1}(X)), T(X)\big)(P)\\ &= \big(T(g_1(g_0(X))), \ldots, T(g_{B-1}(g_0(X))), T(g_0(X))\big)(P),\end{aligned}$$

where g_0 is drawn uniformly from $\mathbf{G}$, independently of everything else. Since $g_1 \circ g_0, \ldots, g_{B-1} \circ g_0, g_0$ are independent and uniform on $\mathbf{G}$, the second vector has the same distribution as

$$T(g_1(X)), \ldots, T(g_B(X)), \tag{36}$$

where $g_1, \ldots, g_B$ are independent and uniform on $\mathbf{G}$. Thus the vector (36) is invariant to permutations, so the original vector $\big(T(g_1(X)), \ldots, T(g_{B-1}(X)), T(X)\big)$ is also invariant. From Lemma 3 we therefore have

$$P(\tilde{p} \leq \alpha) \leq \alpha.$$

To use Lemma 3 to obtain a lower bound for $P(\tilde{p} \leq \alpha)$ we need that an assumption like (34) is satisfied. This is not the case since $g_1, \ldots, g_{B-1}$ are drawn with replacement. The probability that $g_1, \ldots, g_{B-1}$ are different, and different from the identity map on $\mathcal{X}$, denoted id is, however, very high when M is much larger than B. We define the event that this happens as

$$C = \{g_i \neq \text{id and } g_i \neq g_j \text{ for all } i, j, \text{ where } i \neq j\}.$$

Then with simple combinatorics,

$$P(C) = \frac{(M-1)(M-2)\cdots(M-1-(B-1))}{M^{B-1}} \geq \left(\frac{M-B}{M}\right)^{B-1}, \tag{37}$$

and furthermore, the conditional distribution of $g_1, \ldots, g_{B-1}$ given C is that the $B-1$ permutations are drawn from $\mathbf{G} \setminus \{\mathrm{id}\}$ without replacement. With arguments identical to the above it is seen that the conditional distribution of $\big(T(g_1(X)), \ldots, T(g_{B-1}(X)), T(X)\big)$ given C is invariant under permutations. If Assumption 5 is satisfied, we have that (34) is satisfied on C for the variables $T(g_1(X)), \ldots, T(g_{B-1}(X)), T(X)$. Therefore,

$$P(\tilde{p} \leq \alpha \mid C) > \alpha - \frac{K}{B}.$$

Writing

$$P(\tilde{p} \leq \alpha) = P(C)P(\tilde{p} \leq \alpha \mid C) + P(C^c)P(\tilde{p} \leq \alpha \mid C^c) \geq P(C)P(\tilde{p} \leq \alpha \mid C),$$

and combining with (37) gives the desired result. □

SUPPLEMENTARY MATERIAL

8 Technical result on consistency

The lemma below gives that if DEA–estimates are point wise consistent and the efficiency scores are integrable with respect to the distribution of an observation pair (X, Y), then the empirical mean of estimated efficiency scores is also consistent. Conditions ensuring point wise consistency can e.g. be found in Kneip et al. (1998).

Lemma 4. *Let $\mathbf{X}_n = (X_1, \ldots, X_n)$ and $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$ be random variables, such that Assumption 1 (a) and Assumption 2 are satisfied for $(\mathbf{X}_n, \mathbf{Y}_n)$. Furthermore, assume one of the following:*

(i) Assumption 1 (b) is satisfied, and it holds that

$$\hat{\theta}_{CRS}(x, y \mid \mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{P} \theta(x, y \mid \Psi)$$

for all $(x, y) \in \Psi$.

(ii) It holds that

$$\hat{\theta}_{VRS}(x, y \mid \mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{P} \theta(x, y \mid \Psi),$$

for all $(x, y) \in \Psi$.

(iii) It holds that

$$\hat{\theta}_{CRS}(x, y \mid \mathbf{X}_n, \mathbf{Y}_n) \xrightarrow{P} \theta(x, y \mid \mathcal{C}(\Psi))$$

for all $(x, y) \in \Psi$.

Let $\hat{\theta}(x, y \mid \mathbf{X}_n, \mathbf{Y}_n)$ denote the input oriented DEA efficiency estimate and $\theta(x, y)$ denote the limiting efficiency score from the relevant case. If furthermore,

$$E(-\log \theta(X_1, Y_1)) < \infty,$$

then

$$\frac{1}{n}\sum_{i=1}^{n} \log \hat{\theta}(X_i, Y_i) \xrightarrow{P} E \log \theta(X_1, Y_1)$$

as $n \to \infty$.

Proof. First we note that, for i fixed, it holds that

$$\hat{\theta}(X_i, Y_i \mid \mathbf{X}_n, \mathbf{Y}_n) \to \theta(X_i, Y_i),$$

since, eventually, $\hat{\theta}(X_i, Y_i \mid \mathbf{X}_n, \mathbf{Y}_n) = \hat{\theta}(X_i, Y_i \mid \mathbf{X}_n \setminus \{X_i\}, \mathbf{Y}_n \setminus \{Y_i\})$ as (X_i, Y_i) is an inner point (with probability 1) in Ψ.

From the law of large numbers we have

$$\frac{1}{n}\sum_{i=1}^{n} \log \theta(X_i, Y_i) \xrightarrow{a.s.} E \log \theta(X_1, Y_1),$$

so it suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n} Z_{n,i} \xrightarrow{P} 0,$$

where

$$Z_{n,i} = \log \hat{\theta}(X_i, Y_i \mid \mathbf{X}_n, \mathbf{Y}_n) - \log \theta(X_i, Y_i).$$

Note that $Z_{n,i} \xrightarrow{P} 0$ as $n \to \infty$ for every fixed i and that

$$0 \le Z_{n,i} \le -\log \theta(X_i, Y_i). \tag{38}$$

Now let $\epsilon > 0$ be given and use the integrability of $-\log \theta(X_1, Y_1)$ to choose a finite constant K such that

$$E\big[-\log \theta(X_1, Y_1) 1_{\{-\log \theta(X_1, Y_1) > K\}}\big] \le \frac{\epsilon}{3}.$$

Then choose $0 < \delta < \epsilon/3$ and $n_0 \in \mathbb{N}$ such that $KP(Z_{n,1} > \delta) \le \epsilon/3$ for $n \ge n_0$. Using (38) and that $Z_{n,1}, \ldots, Z_{n,n}$ have the same distribution, we find

$$\begin{aligned}
E(\frac{1}{n}\sum_{i=1}^{n} Z_{n,i}) &= E(Z_{n,1}) \\
&\le \delta P(Z_{n,1} \le \delta) + KP(Z_{n,1} > \delta, -\log \theta(X_1, Y_1) \le K) \\
&\quad + E\big[Z_{n,1} 1_{\{Z_{n,1} > \delta, -\log \theta(X_1, Y_1) \le K\}}\big] \\
&\le \delta P(Z_{n,1} \le \delta) + KP(Z_{n,1} > \delta) + E\big[-\log \theta(X_1, Y_1) 1_{\{-\log \theta(X_1, Y_1) > K\}}\big] \\
&\le \epsilon.
\end{aligned}$$

Thus, we have convergence in L_1 and, in particular, in probability. □

9 Monte Carlo experiments

9.1 Test for returns to scale

Simulation procedure 1

In the following simulations we let $p = 2$ and $q = 1$, and we will simulate such that the independence property stated in Proposition 1 is satisfied. Therefore, the hypothesis under study will be whether CRS is satisfied. First, the frontier is defined by a Cobb–Douglas function

$$f(x_1, x_2) = x_1^{\alpha} x_2^{\alpha} \,,$$

where $\alpha = \frac{\gamma}{2}$ and $0 < \gamma \leq 1$. A point $((x_1, x_2), y)$ is placed on the frontier, if

$$y = f(x_1, x_2) \,.$$

Note that the case $\gamma = 1$ corresponds to a CRS situation since then, with the frontier Ψ^{δ} defined by f, it always holds that $\|(x_1, x_2)\| = |y|$ when $((x_1, x_2), y) \in \Psi^{\delta}$. If, on the other hand, $\gamma < 1$ and $((x_1, x_2), y)$ satisfies $f(x_1, x_2) = y$, then for $a > 0$

$$f(a \cdot x_1, a \cdot x_2) = a^{\gamma} \cdot y \ ,$$

demonstrating that CRS cannot be assumed when $\gamma < 1$. Thus f is homogeneous of order 1 under the hypothesis of CRS, while $\gamma < 1$ corresponds to the alternative hypothesis of VRS. When γ decreases from one to zero then the 'distance' to the CRS hypothesis becomes 'larger'.

We generate each of the points $(X_i, Y_i)_{i=1,\ldots,n}$ in the following way, where we suppress i in the notation:

1. Generate U_1 and U_2 independently from a Beta$(3, 3)$–distribution.

2. Define the unit vector (W_1, W_2) by normalizing (U_1, U_2). That is

$$(W_1, W_2) = \frac{(U_1, U_2)}{\|(U_1, U_2)\|} \,.$$

3. Generate A from a $\Gamma(3, 3)$[1]–distribution, and calculate $(X_1^{\delta}, X_2^{\delta})$ as

$$(X_1^{\delta}, X_2^{\delta}) = A \cdot \frac{(W_1, W_2)}{f(W_1, W_2)^{1/\gamma}} \,.$$

4. Calculate Y as

$$Y = f(X_1^{\delta}, X_2^{\delta}).$$

5. Generate Θ from a Beta$(3, 1.5)$–distribution and calculate (X_1, X_2) as

$$(X_1, X_2) = \frac{(X_1^{\delta}, X_2^{\delta})}{\Theta}.$$

[1] We use the notation $\Gamma(shape, scale)$

We have chosen parameters that mimic what is often encountered in real data: In 1. we use a Beta–distribution with large shape parameters to make directions close to the axes less likely than directions "in the middle". In 5. the shape parameter 1.5 is chosen to limit the probability of observations very close to the frontier, but still ensuring that the major part of the observation are not too far away. In particular, this choice means that the joint density of (X, Y) is 0 at the frontier. Therefore based on this simulation study, the performance of the test procedure from Section 3 cannot be compared directly with the tests suggested in e.g. Section 3.2 in Kneip et al. (2016), where the density is required to be non–zero at the frontier.

Note that Y and (X_1, X_2) are related through A. This means that the simulation procedure satisfies the independence property stated in Proposition 1, requiring that the length of Y is independent of the joint distribution of V, W and Θ.

Results from simulation procedure 1

For different combinations of n, the number of observations, and γ, the degree of departure from the CRS hypothesis, we have simulated 1000 sets of observations. For each set we have used the permutation procedure proposed in Section 3 with the number of permutations $N = 1000$ to calculate a significance probability. From this we have derived the proportion of rejected hypotheses on a 5% significance level across the 1000 simulations. The upper part of Table 1 shows the resulting rejection rates.

The first column in Table 1 shows the rejection rates for $\gamma = 1$. This corresponds to the situation where the CRS hypothesis is actually true. As expected, the simulated rejection rate here is close to 5% for all the demonstrated values of n, meaning that the test has the correct size. In the next columns, the γ–parameter is decreased, which means that the departure from the CRS hypothesis increases. Here we see that rejection rates, i.e. the power of the test, increases rather fast, when γ decreases. It is also clear that the test procedure is more powerful for larger sets of observations.

9.2 Test for equality of frontiers

Simulation procedure 2

In this procedure we generate independent samples of observations from two production groups, each satisfying CRS, with $p = 2$, $q = 1$. Furthermore the production groups jointly satisfy the condition of Proposition 2, making the hypothesis under study to be a statement about equal frontiers. We can, without loss of generality, let $Y_i^g = 1$ for $g = 1, 2$ and all $i = 1, \ldots, n_g$ and focus on generating the points X_i^g. For each sample $g = 1, 2$, let the frontier be defined by the Cobb-Douglas function

$$f_g(x_1, x_2) = \beta_g x_1^{\alpha_g} x_2^{1-\alpha_g},$$

such that a point $((x_1, x_2), 1)$ is placed on the frontier, if

$$1 = f_g(x_1, x_2).$$

In each group $g = 1, 2$ we generate $X_i^g \in \mathbb{R}^2$ as follows, where i is suppressed in the notation.

	Permutation test (Section 3)										
	γ										
n	1.00	0.98	0.96	0.94	0.92	0.90	0.88	0.86	0.84	0.82	0.80
50	0.043	0.061	0.081	0.105	0.184	0.286	0.424	0.542	0.648	0.753	0.830
100	0.056	0.075	0.159	0.289	0.472	0.699	0.873	0.944	0.987	0.995	0.998
200	0.054	0.102	0.346	0.698	0.921	0.988	0.998	1.000	1.000	1.000	1.000
300	0.045	0.167	0.594	0.911	0.996	1.000	1.000	1.000	1.000	1.000	1.000

Table 1: Proportions of rejected hypotheses when testing returns to scale on a significance level of $\alpha = 5\%$ with observations simulated according to simulation procedure 1, and using varying values of n and γ.

$n_1 = 50$					$n_2 = 50$				
β_2	$n_2 = 100$		$n_2 = 200$		β_2	$n_1 = 100$		$n_1 = 200$	
	diff	*nest*	*diff*	*nest*		*diff*	*nest*	*diff*	*nest*
1.00	0.062	0.048	0.050	0.052	1.00	0.060	0.060	0.058	0.046
1.02	0.106	0.092	0.110	0.128	1.02	0.058	0.086	0.088	0.138
1.04	0.148	0.278	0.246	0.278	1.04	0.134	0.268	0.164	0.314
1.06	0.350	0.382	0.448	0.514	1.06	0.310	0.422	0.446	0.578
1.08	0.596	0.650	0.706	0.762	1.08	0.596	0.702	0.680	0.752
1.10	0.806	0.824	0.892	0.918	1.10	0.816	0.846	0.892	0.910

Table 2: Proportions of rejected hypotheses on a 5% significance level, when testing equality of frontiers using both the general difference (denoted *diff*) and the nested (denoted *nest*) test. Observations are generated by simulation procedure 2 with $\beta_1 = 1$ and varying values of β_2 such that the two production possibility sets are, in fact, nested.

1. Generate U_1 and U_2 independently from a Beta$(3, 3)$–distribution.
2. Define the unit vector (W_1, W_2) by normalizing (U_1, U_2). That is
$$(W_1, W_2) = \frac{(U_1, U_2)}{\|(U_1, U_2)\|}\,.$$
3. Generate Θ from a Beta$(3, 1.5)$–distribution and calculate (X_1, X_2) as
$$(X_1, X_2) = \frac{(W_1, W_2)}{f_g(W_1, W_2)\Theta}\,.$$

Results from simulation procedure 2

Now we investigate the performance of the test procedures proposed in Section 4, when the two independent groups of observations are simulated according to simulation procedure 2. In all simulation studies we generate 500 datasets each consisting of two groups of observations with varying group sizes n_1 and n_2. When calculating the test statistics, the number of jackknife replications, M, is chosen to be 50, and for the test procedures we have used $N = 1000$ permutations. For each combination of the parameters we have derived the proportion of rejected hypotheses on a 5% significance level across the 500 simulations.

First we have investigated the performance of the tests in a situation, where the two production possibility sets are nested. Thus we have chosen β_1 and β_2 to be unequal, while

$n_1 = 50$					$n_2 = 50$				
β_2	$n_2 = 100$		$n_2 = 200$		β_2	$n_1 = 100$		$n_1 = 200$	
	diff	*nest*	*diff*	*nest*		*diff*	*nest*	*diff*	*nest*
1.00	0.047	0.041	0.051	0.051	1.00	0.052	0.049	0.059	0.055
1.02	0.139	0.155	0.144	0.150	1.02	0.029	0.020	0.015	0.013
1.04	0.311	0.349	0.320	0.329	1.04	0.009	0.004	0.005	0.005
1.06	0.489	0.509	0.599	0.603	1.06	0.009	0.002	0.000	0.000
1.08	0.746	0.766	0.832	0.831	1.08	0.025	0.002	0.001	0.000
1.10	0.900	0.904	0.971	0.973	1.10	0.069	0.014	0.000	0.000

Table 3: Simulations similar to Table 2 but without jackknifing.

α_1 and α_2 are both chosen to be 0.5. We let $\beta_1 = 1$ and let β_2 vary from 1 to 1.1. Therefore, the production possibility set for group 1 is nested within the production possibility set for group 2. Furthermore, the sample sizes vary such that one of them is 50 and the other is either 100 or 200.

In the left part of Table 2 the rejection rates are seen for the cases, where sample 1 is smaller than sample 2, i.e. $n_1 = 50$ and $n_2 \in \{100, 200\}$, and in the right part group 2 is smaller than group 1. In the first row $\beta_2 = 1$, which means that the two frontiers are equal. Here the rejection rates are approximately 5% for both tests – as expected.

In the following rows β_2 is increased, which corresponds to the two frontiers becoming more and more different, such that the production possibility set for group 1 is nested within the production possibility set for group 2. Here the rejection rates are seen to increase substantially for both tests – no matter which of the groups is larger than the other. However, the rejection rate generally seems to be slightly higher for the nested test than for the general difference test. This is not surprising, since the nested test is, in fact, designed to detect exactly the kind of difference between the two frontiers that have been used to produce the two samples.

To illustrate the importance of the use of the jackknife method in the test procedures, we have included a simulation study similar to the one in Table 2, but without jackknifing, where the test statistics T^M_{diff} and T^M_{nest} are now replaced by $T_{diff}((\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2))$ and $T_{nest}((\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2))$ (allowing different lengths of $(\mathbf{X}_1, \mathbf{Y}_1)$ and $(\mathbf{X}_2, \mathbf{Y}_2)$). Here the findings are remarkably different from those of Table 2: When sample size in group 1 is smaller than in group 2, the rejection rate increases faster than before. On the other hand, when group 1 is larger than group 2, both of the tests seem to be unable to reject the false hypothesis of no difference for almost all of the simulated datasets. This is due to the different magnitude of bias when estimating the two frontiers. In the table to the right, the production possibility set for group 1 is nested within the production possibility set for group 2, but at the same time, the frontier of group 1 is estimated with a smaller bias.

	$n_1 = 50$			
α_2	$n_2 = 100$		$n_2 = 200$	
	diff	*nest*	*diff*	*nest*
0.1	0.988	0.022	1.000	0.016
0.2	0.948	0.020	0.992	0.030
0.3	0.646	0.024	0.782	0.036
0.4	0.184	0.034	0.262	0.054
0.5	0.062	0.048	0.050	0.052
0.6	0.160	0.038	0.230	0.054
0.7	0.654	0.020	0.778	0.040
0.8	0.952	0.018	0.998	0.012
0.9	0.952	0.014	0.996	0.028

Table 4: Proportions of rejected hypotheses on a 5% significance level, when testing equality of frontiers using both the general difference (denoted *diff*) and the nested (denoted *nest*) test. Observations are simulated according to simulation procedure 2 with $\beta_1 = \beta_2 = 1$, $\alpha_1 = 0.5$ and varying values of α_2.

These two effects counteract, such that the estimated frontiers are so close that the tests are unable to distinguish between them.

Another part of the evaluation of the two tests is to consider a situation, where the frontiers are different without one production possibility set being nested within the other. Here we have chosen $\beta_1 = \beta_2 = 1$, $\alpha_1 = 0.5$ and varying values of α_2. When α_1 and α_2 are different the two frontiers will be different and, since they intersect, the corresponding production possibility sets are not nested. The rejection rates from this simulation study are seen in Table 4. The middle row with $\alpha_2 = 0.5$ is identical to the first row in the left part of Table 2. Thus, the rejection rate is approximately 5%.

In the other rows the frontiers are different, with a larger difference as α_2 becomes more different from 0.5. Here we see that only the general difference test, in the table denoted as *diff*, detects the difference with an increasing rejection rate, as the two frontiers become more and more different. The nested test, on the other hand, is unable to distinguish between the two frontiers: This test keeps track of, which frontier corresponds to the best production possibilities for each observation, and when averaging over all observations these differences tend to cancel out when the frontiers intersect. Thus, in this situation,

the two tests jointly detect the difference between the two frontiers with high power and furthermore correctly concludes that the production possibility sets are not nested.

Simulation procedure 3

For completeness of the evaluation of our test for equality of frontiers, we have included a simulation study using the following simulation procedure that is inspired by a procedure described in Daraio et al. (2018). However, the two simulation procedures are still somewhat different: While the observations generated in our procedure are divided into two groups with a distinct frontier in each group, the datasets generated in Daraio et al. (2018) all have separate frontiers, determined parametrically by the value of a numerical covariate.

We let $p = q = 2$ and assume CRS. In each of the two groups $g = 1, 2$ we generate each of the observations (X_i, Y_i) for $i = 1, \ldots, n_g$ in the following way:

1. Generate U and V independently and each following the uniform distribution on the part of the unit circle, where both coordinates are positive.

2. Generate Z from a standard normal distribution.

3. Calculate $X, Y \in \mathbb{R}^2$ as

$$X = \left(1.01 - \frac{U}{\|U\|}\right) \cdot (1 + |Z|) \cdot (1 + \gamma_g) \qquad \text{and} \qquad Y = \frac{V}{\|V\|} + 0.01\,.$$

Choosing γ differently for the two groups corresponds to the two groups having different frontiers in such a way that one of the production possibility sets will be nested within the other.

Results from simulation procedure 3

The purpose of Table 5 is to investigate the performance of the two tests from Section 4, when the two groups of observations are simulated according to simulation procedure 3. For each combination of parameters we have generated 500 datasets. For the test procedures, we have again used $m = 50$ jackknife replications and $N = 1000$ permutations. Furthermore, the proportions of rejected hypotheses on a 5% significance level are derived across the 500 simulations.

For different combinations of group sizes, the first row of Table 5 shows the rejection rates when the two frontiers are equal, i.e. when $\gamma_1 = \gamma_2 = 0$. As expected theoretically, all rejection rates in this row are close to 5%. In the next rows the parameter $\gamma_1 = 0$ is fixed, while γ_2 increases. This corresponds to the two frontiers becoming more different. Here the two tests with a very high power correctly identifies both the difference and the fact that the production possibility sets are nested.

	$n_1 = 50$			
γ_2	$n_2 = 100$		$n_2 = 200$	
	F_{diff}	F_{nest}	F_{diff}	F_{nest}
0.0	0.040	0.062	0.050	0.050
0.2	0.634	0.902	0.756	0.938
0.4	0.998	1.000	1.000	1.000
0.6	1.000	1.000	1.000	1.000

Table 5: Simulation results for test for equality of frontiers, when data is simulated according to simulation procedure 3. Proportions of rejected hypotheses for varying values of γ_2.

References

Banker, R. D. (1996). Hypothesis tests using data envelopment analysis. *Journal of Productivity Analysis 7*, 139–159.

Berry, K. J., J. E. Johnston, and P. W. M. Jr (2014). *A Chronicle of Permutation Statistical Methods. 1920-2000, and beyond*. Springer.

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

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. *Econometrics Journal 21*, 170–191.

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

Färe, R., M. Margaritis, P. Rousea, and I. Roshdi (2016). Estimating the hyperbolic distance function: A directional distance function approach. *European Journal of Operational Research 251*, 312–319.

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

Fisher, R. A. (1935). The logic of inductive inference (with discussion). *Journal of the Royal Statistical Society 98*, 39–82.

Kneip, A., B. U. Park, , and L. Simar (1998). A note on the convergence of nonparametric dea estimators for production efficiency scores. *Econometric theory 14*, 783–793.

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

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

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses* (third edition ed.). Springer texts in statistics.

Nguyen, B. H., L. Simar, and V. Zelenyuk (2022). Data sharpening for improving central limit theorem approximations for data envelopment analysis - type efficiency estimators. *European Journal of Operational Research 303*, 1469–1480.

Simar, L. and P. W. Wilson (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research 139*, 115–132.

Simar, L. and P. W. Wilson (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics 136*, 31–64.

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

Simar, L. and P. W. Wilson (2020). Hypothesis testing in non-parametric models of production using multiple sample splits. *Journal of Productivity Analysis 53*(3), 287–303.

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