DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

Testing productivity change, frontier shift, and efficiency change

Mette Asmild
Dorte Kronborg
Anders Rønn-Nielsen

2018 / 07

IFRO Working Paper 2018 / 07

Testing productivity change, frontier shift, and efficiency change

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

JEL-classification: C12, C14, C44, C46, C61, D24

Published: June 2018

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

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk/english/

Testing productivity change, frontier shift, and efficiency change

Mette Asmild
Institute of Food and Resource Economics
University of Copenhagen,

Dorte Kronborg & Anders Rønn-Nielsen
Center for Statistics, Department of Finance
Copenhagen Business School.

Abstract

Inference about productivity change over time based on data envelopment (DEA) has focused primarily on the Malmquist index and is based on asymptotic properties of the index. In this paper we propose a novel set of significance tests for DEA based productivity change measures based on permutations and accounting for the inherent correlations when panel data are observed. The tests are easily implementable and give exact significance probabilities as they are not based on asymptotic properties. Tests are formulated both for the geometric means of the Malmquist index, and also of its components, i.e. the frontier shift index and the efficiency change index, which together enable analysis of not only the presence of differences, but also gives an indication of whether the productivity change is due to shifts in the frontiers and/or changes in the efficiency distributions. Simulation results show the power of, and suggest how to interpret the results of, the proposed tests. Finally, the tests are illustrated using a data set from the literature.

Keywords: Malmquist index, frontier shift, efficiency change, Data Envelopment Analysis (DEA), panel data, permutation tests, inference.

Correspondence: Mette Asmild, IFRO, University of Copenhagen, Rolighedsvej 25, 1958 Frederiksberg C., Denmark, email: meas@ifro.ku.dk

1 Introduction

Benchmarking production units with non-parametric data envelopment analysis (DEA) efficiency estimators, based on enveloping the observed set of input-output combinations with a convex set, is widely used. Farrell (1957) was the first to suggest the approach, which following the seminal paper by Charnes et al. (1978) became more widely used. Since then, the method has become very popular and many studies based on DEA efficiency estimators have been published.

The statistical properties of the DEA efficiency estimator have been subject to numerous studies. Simar and Wilson (2015) provide a review of the results on the asymptotics of its sampling distributions. Kneip et al. (2015) showed a central limit theorem for the means of the estimated inefficiencies. Based on these results, methods for inference about the mean efficiency has been developed in Kneip et al. (2016), who furthermore propose a test for equality of means in two *independent* samples (groups) of units and a test for equality of means *and* a common frontier for the two groups. Both these tests are based on the asymptotic distribution of the difference between the bias corrected sample means of inefficiencies, and the derived asymptotic normal distribution relies on the independence assumption. The bias correction terms are obtained using jackknife methods and are not easily calculated.

Measures of productivity change over time have been commonly used since Caves et al. (1982) introduced the Malmquist index of productivity change for a unit in different time periods. Färe et al. (1992) subsequently proposed to calculate the Malmquist index from DEA scores and furthermore suggested to decompose the index into an efficiency change and a frontier shift component respectively. Recently, Kneip et al. (2018) have derived central limit theorem results for the distribution of the geometric means of DEA-based Malmquist estimators. Their results can be used to construct asymptotic confidence intervals for the mean productivity changes, but can not directly be translated into corresponding results regarding the frontier shift and the efficiency change components.

Banker and Natarajan (2011) propose various tests for technical change, relative efficiency change and productivity change. These tests rely on supposed results on the asymptotic distributions of the empirical standard t-test statistic for e.g. the mean of the logarithm to the relative change in the estimated frontiers. Furthermore, the approach of Banker and Natarajan (2011) is at present limited to the single output case, and not immediately generalizable to multiple outputs.

In the present paper we propose exact, and easily implementable, tests for productivity change, frontier shift, and efficiency change, based on permutations. These tests do not rely on asymptotic distributional assumptions and are therefore also suitable for practical purposes with limited sample sizes. We conduct extensive simulation studies to assess the power of the proposed tests and conclude that they are extremely powerful. Moreover, we analyse the ability of the three tests to distinguish between frontier shift and efficiency change. Based on this, we give recommendations on how to interpret the test results. Finally, the approach is illustrated on a well-known data set from the literature.

The rest of this paper is structured as follows: In Section 2, the problem is formally introduced, the hypotheses specified and appropriate test statistics defined. Section 3 describes the design of the simulation studies, the results of which are presented in section 4, which showcases the power of the proposed tests and explains how to interpret the results. In Section 5 the tests are illustrated using an empirical example of U.S. electricity producing firms. The R-code used to implement the tests in the empirical example is provided in an appendix. Finally, Section 6 concludes the paper.

2 Test procedure

Following the notation of e.g. Kneip et al. (2016) let the vector of input quantities be denoted by $x \in \mathbb{R}_+^p$ and output vector denoted by $y \in \mathbb{R}_+^q$. The production possibility set, i.e the feasible set of input-output combinations, is given as

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

The efficient frontier of Ψ is given by

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

Farrell's index of technical input efficiency (Farrell 1957) is defined as

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

The smaller the $\theta(x, y) \leq 1$, the more inefficient is the firm and if $\theta(x, y) = 1$ the firm is said to be technically efficient. We could as well have chosen to measure the output oriented technical efficiency or any other directional efficiency measure.

As in e.g. Kneip et al. (2016) the production possibility set is assumed to be closed, convex and satisfying strong disposability in both inputs and outputs.

Since we are here concerned with frontiers from different time periods, let the efficient frontier be indexed by a subscript $t \in \{t_1, t_2\}$, such that Ψ_t^δ denotes the frontier for time period t. Assume that there exists a distribution F_t with density f_t on Ψ_t. Let (X_t, Y_t) denote random variables with distribution F_t and joint distribution F. Note that the random variables in the two time periods, (X_{t_1}, Y_{t_1}), and (X_{t_2}, Y_{t_2}), are allowed to be dependent.

In practice, the production possibility set Ψ_t is unobserved and estimated from a set of n observations, $i = 1, \dots, n$, from a balanced panel. The observations $(X_{t_1}^i, Y_{t_1}^i, X_{t_2}^i, Y_{t_2}^i)$ are assumed to be independent and identically distributed, such that $(X_{t_1}^i, Y_{t_1}^i)$ has distribution F_{t_1} and $(X_{t_2}^i, Y_{t_2}^i)$ has distribution F_{t_2} for all $i = 1, \dots, n$. For ease of notation, the vector $(X_t^1, \dots, X_t^n)$ will be denoted $\mathbf{X}_t$ and the vector $(Y_t^1, \dots, Y_t^n)$ will be denoted $\mathbf{Y}_t$ for $t \in \{t_1, t_2\}$.

The DEA estimator of the technical input efficiency is calculated by solving the usual linear programming problem, which e.g. in the case of constant return to scale (CRS) is formulated as

$$\hat{\theta}_t(x, y) = \min_{\theta, \omega}\{\theta \mid y \leq \mathbf{Y}_t\omega,\ \theta x \geq \mathbf{X}_t\omega,\ \omega \in \mathbb{R}_+^n\}.$$

for $x \in \mathbb{R}_+^p$, $y \in \mathbb{R}_+^q$ and $t \in \{t_1, t_2\}$. If another technology is appropriate the efficiencies are calculated accordingly.

2.1 Inference on changes

The traditional Malmquist index of productivity change from one period to another for a given firm (see e.g. Färe et al. 1992), is often decomposed into two effects: The technical change (frontier shift) and the efficiency change (catch-up). The technical change for an individual observation (x, y) is the ratio of the (here input) efficiencies for (x, y) relative to each of the two frontiers

$$FS(x, y) = \frac{\hat{\theta}_{t_1}(x, y)}{\hat{\theta}_{t_2}(x, y)}\,. \tag{1}$$

The efficiency change component for (x, y) observed in both time periods, t_1 and t_2, is

$$EC(x_{t_1}, y_{t_1}, x_{t_2}, y_{t_2}) = \frac{\hat{\theta}_{t_2}(x_{t_2}, y_{t_2})}{\hat{\theta}_{t_1}(x_{t_1}, y_{t_1})}. \tag{2}$$

Typically, the geometric means over all the observations of these components are reported and interpreted as the technical change viz. efficiency change for the whole technology, similar to the logic of Asmild and Tam (2007).

The geometric mean of the Malmquist index is calculated as

$$T_M = \prod_{t\in\{t_1,t_2\}} \prod_{i=1}^{n} \left(\frac{\hat{\theta}_t(X^i_{t_2}, Y^i_{t_2})}{\hat{\theta}_t(X^i_{t_1}, Y^i_{t_1})} \right)^{\frac{1}{2n}}. \tag{3}$$

Likewise, the geometric mean of the frontier shift component is

$$T_{FS} = \prod_{t\in\{t_1,t_2\}} \prod_{i=1}^{n} FS(X^i_t, Y^i_t)^{\frac{1}{2n}}, \tag{4}$$

and the geometric mean of the efficiency measure is

$$T_{EC} = \prod_{i=1}^{n} EC(X^i_{t_1}, Y^i_{t_1}, X^i_{t_2}, Y^i_{t_2})^{\frac{1}{n}}. \tag{5}$$

Note that $T_M = T_{FS} \times T_{EC}$ and all three statistics are positive. Also, we note that T_{FS} is close to one if the two frontiers are equal, and that a value far from 1 will be evidence against a hypothesis of the two frontiers being equal. Similarly, T_{EC} is close to one if the firms are equally close to the frontiers in the two periods and smaller than one if the firms overall are further away from the frontier in time period t_2 than in time period t_1.

The statistical properties of T_{FS} and T_{EC} are unknown. Therefore, we in the following propose a permutation test for the hypothesis $(F_{t_1}, F_{t_2}) = (F_{t_2}, F_{t_1})$, i.e. that the distribution in time period t_1, F_{t_1}, can be interchanged with the distribution in time period t_2, F_{t_2}. We use three different test statistics, cf. (3)-(5), to investigate this hypothesis, which provide different information about the nature of any difference between (F_{t_1}, F_{t_2}) and (F_{t_2}, F_{t_1}).

If the hypothesis is rejected because of an extreme value of the test statistic T_{FS}, this could be an indication of the frontiers in the two time periods being different, i.e. $\Psi^\delta_{t_1}$ is different from $\Psi^\delta_{t_2}$. However this could also be caused by other differences between the two distributions, for example the densities of observations close to the frontiers. The simulation studies in Section 3 provide insight into which types of differences are evident from the different test statistics.

Next, if the hypothesis is rejected because of an extreme value of the test statistic T_{EC} then there is an overall efficiency change and/or the frontiers are different.

2.2 A permutation test

We construct $3 \times N$ new test statistics T^j_{FS}, T^j_{EC}, and T^j_M for $j = 1, \ldots, N$, where N is the number of permutations, using the following procedure:

a. For each i in $1, \ldots, n$ construct a new observation pair $(\tilde{X}^i_{t_1}, \tilde{Y}^i_{t_1}, \tilde{X}^i_{t_2}, \tilde{Y}^i_{t_2})$

   With probability 0.5: Let $(\tilde{X}^i_{t_1}, \tilde{Y}^i_{t_1}, \tilde{X}^i_{t_2}, \tilde{Y}^i_{t_2})$ be the interchanged observation $(X^i_{t_2}, Y^i_{t_2}, X^i_{t_1}, Y^i_{t_1})$

   Otherwise: Let $(\tilde{X}^i_{t_1}, \tilde{Y}^i_{t_1}, \tilde{X}^i_{t_2}, \tilde{Y}^i_{t_2})$ be the unchanged observation $(X^i_{t_1}, Y^i_{t_1}, X^i_{t_2}, Y^i_{t_2})$

b. Calculate T^j_{FS}, T^j_{EC}, and T^j_M similar to the description in (3)-(5), but with the new set of observations $(\tilde{X}^i_{t_1}, \tilde{Y}^i_{t_1}, \tilde{X}^i_{t_2}, \tilde{Y}^i_{t_2})$ for $i = 1, \ldots, n$.

Under the hypothesis that $(F_{t_1}, F_{t_2}) = (F_{t_2}, F_{t_1})$ it holds that the observations $(\tilde{X}^i_{t_1}, \tilde{Y}^i_{t_1}, \tilde{X}^i_{t_2}, \tilde{Y}^i_{t_2})$ for $i = 1, \ldots, n$ are independent and identically distributed such that the joint distribution of these is the same as the distribution of the original observations. Consequently, the distributions of all T^j_{FS}'s are identical with the distribution of the original T_{FS}, and the significance probability for the hypothesis can be obtained by comparing the observed T_{FS} with the empirical distribution of the T^j_{FS}–variables: We count the number of T^j_{FS}'s that are further away from 1 than the observed T_{FS}. Similarly, significance probabilities can be obtained based on the empirical distributions of the T^j_{EC}'s and T^j_M's.

2.3 Frontier shift test for unbalanced panels

The test statistics introduced in Section 2.1 are formulated for balanced panels, but T_{FS}, as defined in (4), can easily be extended to cover the non-balanced case by calculating $FS(x, y)$ in (1) for all observations say n_{t_1} and n_{t_2} for each of the two time periods. The test statistic now becomes the geometric mean of these $n_{t_1}+n_{t_2}$ ratios and is calculated as

$$T_{FS} = \prod_{t \in \{t_1, t_2\}} \prod_{i=1}^{n_t} FS(X^i_t, Y^i_t)^{\frac{1}{n_{t_1}+n_{t_2}}}. \tag{6}$$

The permutation test described in Section 2.2 of interchanging the observations assume balanced panels. In the unbalanced case with a total of $n_{t_1}+n_{t_2}$ observations and where, say, n of these are complete observations (observations from both years) the method can be modified as follows, when constructing each of the new test statistics T^j_{FS} (for $j = 1, \ldots, N$):

I. Based on the set of n complete observation pairs derive $(\tilde{X}^i_{t_1}, \tilde{Y}^i_{t_1}, \tilde{X}^i_{t_2}, \tilde{Y}^i_{t_2})$ for $i = 1, \ldots, n$ as in (a) in Section 2.2.

II. Permute the remaining $(n_{t_1} - n)+(n_{t_2} - n)$ observations and divide them randomly into two groups of size $n_1 - n$ and $n_2 - n$ respectively. Let $(\tilde{X}^i_{t_1}, \tilde{Y}^i_{t_1})$ denote the observations that end up in the first group for $i = n+1, \ldots, n_{t_1}$, and similarly let $(\tilde{X}^i_{t_2}, \tilde{Y}^i_{t_2})$ denote the observations in the second group for $i = n+1, \ldots, n_{t_2}$.

III. Calculate T^j_{FS} based on the set of new observations $(\tilde{X}^i_{t_1}, \tilde{Y}^i_{t_1})$ and $(\tilde{X}^k_{t_2}, \tilde{Y}^k_{t_2})$ for $i = 1, \ldots, n_1$ and $k = 1, \ldots, n_2$ using (6).

3 Monte Carlo procedure

In the simulations we let $p = 2$ and $q = 1$ and assume CRS. The set of experiments is performed for varying "true frontiers" and varying distributions of the points both with and without frontier shift.

First we describe how we simulate each of the dependent pairs (X_{t_1}, X_{t_2}) under the hypothesis that $(F_{t_1}, F_{t_2}) = (F_{t_2}, F_{t_1})$.

Let the common frontier be defined by a Cobb–Douglas function, such that a point (x_1, x_2) is placed on the frontier, if

$$1 = f(x_1, x_2)\,, \quad \text{where} \quad f(x_1, x_2) = x_1^a x_2^{1-a},$$

for some parameter $a \in (0, 1)$. Each X_t will be determined by a direction U_t on the positive part of the unit sphere and a technical input efficiency $0 < \Theta_t \leq 1$. In the simulation procedure, we will let both pairs (U_{t_1}, U_{t_2}) and $(\Theta_{t_1}, \Theta_{t_2})$ be dependent. The direction vector (U_{t_1}, U_{t_2}) is simulated as follows:

1. Simulate two vectors (W^1_1, W^1_2) and (W^2_1, W^2_2) each with dependent uniform distributions on (0,1), using a Gumbel copula. The variables are positively correlated, and the strength of the dependence is determined by a non–negative parameter, we shall denote θ_0. A larger value means a stronger dependence.

2. Transform each of the variables $W^1_1, W^1_2, W^2_1, W^2_2$ by the quantile function for the Beta-distribution here chosen with parameters $(3, 3)$. Let $V^1_1, V^1_2, V^2_1, V^2_2$ denote the new variables, and note that they all follow a Beta$(3, 3)$–distribution.

3. Let U_{t_1} be the normalized version of the vector (V_1^1, V_1^2), and U_{t_2} the normalized version of the vector (V_2^1, V_2^2). That is

$$U_{t_1} = \frac{(V_1^1, V_1^2)}{\|(V_1^1, V_1^2)\|} \quad \text{and} \quad U_{t_2} = \frac{(V_2^1, V_2^2)}{\|(V_2^1, V_2^2)\|}.$$

   Consequently, the two directions U_{t_1} and U_{t_2} are dependent, since both their first and second coordinates are chosen to be dependent.

By this construction, the directions are not uniformly distributed on the positive part of the unit sphere. This is however intentional and has the purpose of mimicking typical datasets: The use of the Beta–distribution makes directions close to the axes less likely than directions in the middle of the positive part of the unit circle.

The input efficiency vector is simulated as follows

4. Simulate a vector (Z_1, Z_2) of two dependent uniformly distributed variables using a Gumbel copula with dependence parameter θ.

5. Transform each of the two variables Z_1, Z_2 by the quantile function for the Beta-distribution with parameters $(3, 1.5)$. Define Θ_{t_1} and Θ_{t_2} to be the new variables.

Note that the variables Θ_{t_1} and Θ_{t_2} both (marginally) follow a Beta(3,1.5)–distribution. The choice of the Beta–distribution is due to the flexibility of this class of probability distributions. The parameter 3 determines the tail behaviour of points far away from the frontier, and the parameter 1.5 determines the density of points very close to the frontier. Choosing this parameter greater than 1 makes points very close to the frontier less likely.

The final pair of points (X_{t_1}, X_{t_2}) is generated as

$$X_{t_1} = \frac{U_{t_1}}{f(U_{t_1})\Theta_{t_1}} \quad \text{and} \quad X_{t_2} = \frac{U_{t_2}}{f(U_{t_2})\Theta_{t_2}}.$$

3.1 Difference between frontiers

When simulating dependent pairs (X_{t_1}, X_{t_2}) in a situation, where there is a difference between the two distribution (F_{t_1}, F_{t_2}) and (F_{t_2}, F_{t_1}), we follow almost the same procedure as previously. The only difference will be that we apply two different Cobb–Douglas functions in the definitions of X_{t_1} and X_{t_2}, respectively. For X_{t_1} we apply the same function as previously

$$f(x_1, x_2) = x_1^a x_2^{1-a},$$

while for $\beta > 1$ we use

$$f_2(x_1, x_2) = \beta x_1^a x_2^{1-a} ,$$

to generate X_{t_2} with technological progress.

3.2 Difference between efficiency distributions

Another way of obtaining a difference between the two distributions (F_{t_1}, F_{t_2}) and (F_{t_2}, F_{t_1}) is by changing the efficiency distribution. Here we use the same frontier for the two time periods, and let the marginal distributions of Θ_{t_1} and Θ_{t_2}, respectively, be different. In this situation, we use two different modified versions of (5) when simulating the efficiency vector. For the first version, we only change the part of the efficiency distribution that is furthest away from 1:

5'. Transform each of the two variables Z_1, Z_2 by the quantile function for the Beta-distribution with parameters $(3, 1.5)$. Define Θ_{t_1} and Θ_{t_2} to be the new variables. Replace Θ_{t_2} by the following transformation of Θ_{t_2}

$$\begin{cases} \left(\frac{\Theta_{t_2}}{\delta}\right)^{1.5} \cdot \delta & \Theta_{t_2} \leq \delta \\ \Theta_{t_2} & \Theta_{t_2} > \delta \ . \end{cases}$$

The parameter δ is in the simulations chosen to be 0.7, 0.8, and 0.9. Using this procedure, Θ_{t_1} and Θ_{t_2} will have different distributions, but in a way such that the behaviour of efficiencies close to 1 is unchanged. That means that points close to the frontier will behave in the same way for the two time periods, while points further away from the frontier will behave differently: In the second time period these points are further away from the frontier than in the first period.

For the second version of modifying the efficiency distributions, we simply use another Beta-distribution for the observations in the second time period

5". Transform the variable Z_1 by the quantile function for the Beta-distribution with parameters $(3, 1.5)$ and the variable Z_2 by the quantile function for the Beta-distribution with parameters $(3, 1.3)$. Define Θ_{t_1} and Θ_{t_2} to be the new variables.

The densities for Θ_{t_1} and Θ_{t_2} (both when simulated according to 5', using $\delta = 0.8$, and 5") are shown in Figure 1.

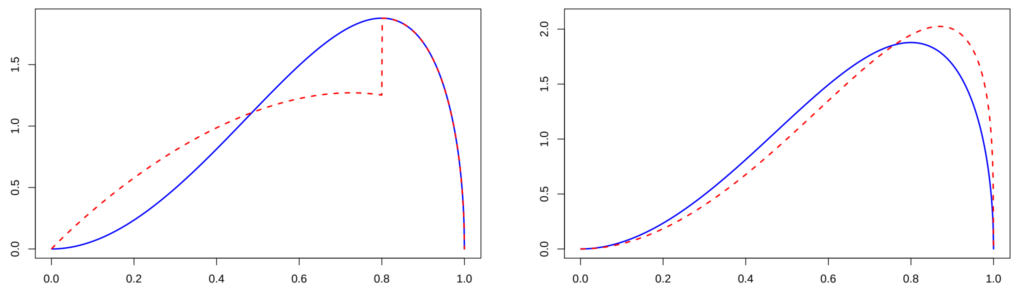

Figure 1: The left plot shows the density of the Beta(3,1.5)–distribution (blue solid line) together with the density described in 5" (red dashed line) with δ chosen to be 0.8. The right plot shows the density of the Beta(3,1.5)–distribution (blue solid line) together with the density of the Beta(3,1.3)–distribution (red dashed line).

4 Simulation results

In the following we have used the simulation procedure described above to investigate the power of the proposed tests. For each combination of parameters within each of the simulation scenarios from Section 3.1 and 3.2, we have generated 1000 sets of observations containing either $n = 50$ or $n = 100$ observations. For each set, we have used the permutation test procedure described in Section 2.2 with $N = 1000$ permutations, to derive the proportion of rejected hypotheses on a 5% significance level. The values shown in each of the cells in Tables 1-3 are the rejection rates across the 1000 simulations.

4.1 Results for differences between frontiers

Table 1 investigates the behaviour of the tests in the situations, where the frontiers are different but the efficiency distributions are identical, i.e. where there is frontier shift but no efficiency change. The results in the first 4 rows of Table 1 correspond to the situation with no frontier shift, i.e. $\beta = 1$, and thus identical distributions in the two groups, corresponding to the hypothesis $(F_{t_1}, F_{t_2}) = (F_{t_2}, F_{t_1})$. As expected, the proportions of rejected hypotheses, for all three tests and all parameter combinations, are close to 0.05 in these rows.

In the next rows we see that the larger the difference between the frontiers (i.e. the larger the β) the larger the proportions of rejections for the frontier shift (FS) and the Malmquist (M) index tests. This clearly illustrates that the tests have very high power for all parameter combinations. Note also that the higher the dependency between the observations in the two time periods, as expressed by the parameters θ_0 and θ, the higher the power of the test. However, the power of the efficiency change (EC) test remains low throughout the parameter combinations. This is as expected, since we in this set of simulations maintain the same efficiency distributions in the two groups.

To summarize, from Table 1 we see that in situations with only frontier shift, the M-test identifies differences between the two distributions, and the FS-test and the EC-test between them show that the differences are due to frontier shift alone.

Figure 2 illustrates the results from power calculations for FS-test for sample sizes of $n = 50$ and $n = 100$ respectively for the symmetric Cobb-Douglas function with $\theta_0 = 2.5$ and $\theta = 2.5$. The significance levels chosen are 0.05, 0.01, and 0.001. Again, we see that the power is high, and even for fairly small differences between the frontiers we are able to distinguish between them also for small samples ($n = 50$).

4.2 Results for differences between efficiency distributions

This section investigates the behaviour of the tests in the situations where the frontiers are identical but the efficiency distributions are different, i.e. where there is efficiency change but no frontier shift.

First consider the case where, even though the efficiency distributions are different, the distribution of points close to the frontier is the same in the two groups (using three different cut-off points), as illustrated in the left panel of Figure 1. This means that the estimated frontiers for the two groups are similar (apart from the natural variation). Therefore the FS-test should not identify any differences between the two frontiers, while both the M-test and the EC-test are expected to find a difference between the two group of observations.

Table 2 shows the power of the three tests with the cut-off parameter δ chosen to be 0.7, 0.8 and 0.9 and with a and the dependence parameters θ and θ_0 varying as before. As expected, both the M–test and the EC-test correctly detects that

			$a = 0.5$						$a = 0.8$					
			$n = 50$			$n = 100$			$n = 50$			$n = 100$		
β	θ_0	θ	FS	EC	M	FS	EC	M	FS	EC	M	FS	EC	M
1	2.5	2.5	0.047	0.046	0.049	0.049	0.051	0.045	0.065	0.051	0.052	0.061	0.045	0.043
		5	0.054	0.040	0.045	0.061	0.042	0.046	0.050	0.054	0.057	0.050	0.044	0.043
	5	2.5	0.049	0.046	0.048	0.054	0.046	0.050	0.036	0.055	0.051	0.040	0.058	0.0058
		5	0.057	0.050	0.061	0.047	0.045	0.056	0..55	0.060	0.060	0.0346	0.057	0.052
1.05	2.5	2.5	0.475	0.033	0.203	0.806	0.025	0.373	0.422	0.032	0.189	0.813	0.024	0.388
		5	0.532	0.018	0.495	0.888	0.019	0.788	0.519	0.012	0.487	0.845	0.008	0.783
	5	2.5	0.537	0.021	0.232	0.874	0.017	0.383	0.504	0.034	0.217	0.872	0.030	0.379
		5	0.713	0.015	0.523	0.944	0.009	0.833	0.652	0.012	0.539	0.930	0.011	0.822
1.1	2.5	2.5	0.885	0.014	0.578	0.999	0.010	0.896	0.884	0.015	0.584	0.998	0.010	0.864
		5	0.932	0.007	0.931	0.998	0.002	0.998	0.904	0.006	0.924	0.997	0.004	0.998
	5	2.5	0.920	0.016	0.571	0.999	0.016	0.861	0.912	0.018	0.591	1	0.006	0.894
		5	0.979	0.000	0.969	1	0.002	0.998	0.965	0.004	0.956	1	0.003	0.999
1.2	2.5	2.5	0.997	0.007	0.968	1	0.004	1	0.999	0.004	0.972	1	0.005	1
		5	1	0.002	1	1	0	1	1	0	1	1	0	1
	5	2.5	0.999	0.002	0.972	1	0.010	1	0.999	0.007	0.967	1	0.011	1
		5	1	0	1	1	0	1	0.999	0	1	1	0.011	1

Table 1: Simulation results: Proportions of rejected hypotheses for varying frontiers and parameter combinations.

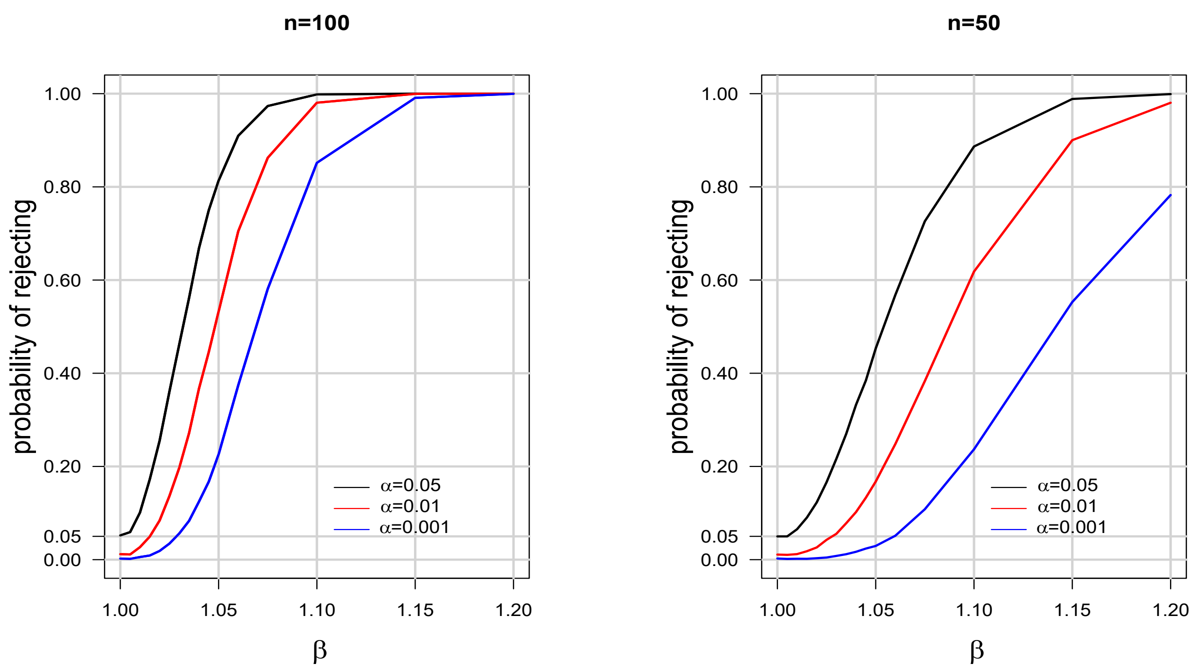

Figure 2: Power functions for varying sample sizes (n) and varying significance levels, α.

			$a = 0.5$						$a = 0.8$					
			$n = 50$			$n = 100$			$n = 50$			$n = 100$		
δ	θ_0	θ	FS	EC	M	FS	EC	M	FS	EC	M	FS	EC	M
0.7	2.5	2.5	0.052	0.436	0.442	0.046	0.750	0.772	0.053	0.436	0.443	0.044	0.751	0.774
		5	0.044	0.759	0.832	0.049	0.971	0.993	0.057	0.773	0.823	0.055	0.973	0.992
	5	2.5	0.057	0.414	0.448	0.048	0.737	0.759	0.054	0.422	0.449	0.051	0.739	0.761
		5	0.051	0.798	0.850	0.050	0.980	0.994	0.058	0.809	0.853	0.054	0.984	0.994
0.8	2.5	2.5	0.056	0.697	0.742	0.052	0.937	0.963	0.062	0.694	0.738	0.051	0.940	0.961
		5	0.051	0.951	0.974	0.065	0.999	1	0.057	0.952	0.980	0.061	0.999	1
	5	2.5	0.051	0.725	0.770	0.064	0.950	0.960	0.057	0.734	0.760	0.057	0.946	0.960
		5	0.050	0.977	0.985	0.060	1	1	0.047	0.972	0.986	0.063	1	1
0.9	2.5	2.5	0.066	0.886	0.926	0.054	0.995	0.998	0.080	0.889	0.921	0.054	0.994	0.998
		5	0.075	0.992	1	0.052	1	1	0.073	0.991	1	0.050	1	1
	5	2.5	0.059	0.932	0.953	0.074	0.998	1	0.056	0.932	0.954	0.069	0.998	1
		5	0.065	0.998	1	0.062	1	1	0.065	0.997	1	0.057	1	1

Table 2: Simulation results: Proportions of rejected hypotheses for varying efficiencies simulated according to 5'.

there is a difference between the two groups. The power of these two tests seem to be almost identical in the situations studied here. It is furthermore seen that the power increases drastically, when δ increases. This effect is due to the fact that the two efficiency distributions become more different with increasing values of δ.

Conversely, but not unexpectedly, the FS-test does not detect any differences between the two groups. In particular for δ being 0.7 and 0.8 the test behaves as if there was no difference at all. For $\delta = 0.9$ there seems to be a small increase in the power. This happens since with $\delta = 0.9$ only few of the points near the frontier in the second group follow the same distribution as the points in the first group. Thus, due to the bias when estimating frontiers, the two frontiers will be estimated differently.

The total collection of tests correctly concludes, that the efficiencies are different and that the frontier and the distribution close to the frontier is the same.

Next, consider the case where the entire efficiency distributions, i.e. also for points close to the frontiers, are in fact different in the two groups as illustrated in the right panel of Figure 1 and mathematically described in 5". This influences the (biases of the) estimations of the two frontiers and therefore the FS-test might identify frontier differences, even if these are caused by the different biases in the estimations rather than real frontier shifts.

Table 3 shows the simulated power of the three tests with the parameters a, θ and θ_0 varying as before. The M-test correctly identifies that there is a difference between the groups, and the EC-test furthermore establishes that (at least) part of the difference is due to a change in the efficiency distribution. However as expected, also the FS-test finds a difference between the groups. Thus, it can be concluded that the FS-test reacts on both real frontier shifts and on changes in the point distribution near the frontier.

4.3 Conclusion on simulation results

To sum up, our simulations show

- If the M-test shows non-significance, there is not statistical evidence for F_{t_1} and F_{t_2} being different and thereby $\Psi^{\delta}_{t_1}$ being unequal to $\Psi^{\delta}_{t_2}$.
- If the M-test shows significance, then F_{t_1} and F_{t_2} are unequal.
- If the FS-test shows non-significance, there is not statistical evidence for the frontiers being equal i.e. $\Psi^{\delta}_{t_1}$ is equal to $\Psi^{\delta}_{t_2}$ *and* the distributions F_{t_1} and

		$a = 0.5$						$a = 0.8$					
		$n = 50$			$n = 100$			$n = 50$			$n = 100$		
θ_0	θ	FS	EC	M	FS	EC	M	FS	EC	M	FS	EC	M
2.5	2.5	0.142	0.114	0.227	0.170	0.198	0.394	0.125	0.115	0.229	0.168	0.210	0.396
	5	0.169	0.234	0.543	0.258	0.460	0.832	0.159	0.253	0.550	0.244	0.478	0.836
5	2.5	0.178	0.123	0.240	0.268	0.225	0.423	0.149	0.113	0.239	0.251	0.238	0.424
	5	0.280	0.233	0.584	0.369	0.529	0.861	0.250	0.245	0.581	0.357	0.558	0.856

Table 3: Simulation results: Proportions of rejected hypotheses for varying efficiencies simulated according to 5".

F_{t_2} are equal *near* the frontier. However this does not exclude the possibility of the point distributions being substantially different further away from the frontier.

- If the FS-test shows significance, then the frontiers are different, i.e. $\Psi_{t_1}^{\delta}$ is different from $\Psi_{t_2}^{\delta}$ *or* the distributions F_{t_1} and F_{t_2} are different *near* the frontier.
- If the EC-test shows significance, then the distributions of efficiences are different in the two periods.

5 Empirical Example

To illustrate the proposed tests, we reconsider the data set on U.S. electric power generating firms, introduced by Christensen and Greene (1976) and also used by e.g. Pastor and Lovell (2005) and Pastor et al. (2011) to illustrate various Malmquist-type indexes. The data set comprises a balanced panel of 93 electricity producing firms observed in each of the years 1977, 1982, 1987, and 1992. The firms use three inputs (capital, fuel and labour) to produce one output (electricity) and the production model use an input-orientation and assume constant returns to scale. The justification for the choice of variables can be found in Christensen and Greene (1976) and average values for the variables in each year are provided by Pastor and Lovell (2005).

In Table 4 the test statistics (cf. equations 3-5) for the shifts 1977-1982, 1982-1987, and 1987-1992 in the electricity data set are shown.

	77-82	82-87	87-92
T_{FS}	0.370 (0.000)	0.822 (0.474)	1.118 (0.099)
T_{EC}	1.163 (0.678)	1.089 (0.986)	0.929 (0.314)
T_M	0.431 (0.000)	0.895 (0.000)	1.039 (0.084)

Table 4: Test statistics and significance probabilities for electricity data (based on 10000 permutations).

Regarding the frontier shifts shown in the first row of Table 4, we see that the frontier shift index, T_{FS} (cf. equation 4), can be found to be 0.370, 0.822, and 1.118 for the shifts 1977-1982, 1982-1987, and 1987-1992 respectively, which is traditionally interpreted as the frontier having worsened from 1977 to 1982 and from

1982 to 1987, but improved from 1987 to 1992.

Applying the permutation test (with 10000 permutations), the corresponding significance probabilities are 0.000, 0.474 and 0.099 respectively. Therefore, we can conclude that there is only a significant difference between the distributions of points, and thus likely a significant difference between the frontiers, for the shift from 1977 to 1982. For the other two shifts (1982-87 and 1987-92) we find no significant differences between the distributions, and thus we can not distinguish between these frontiers.

The observations from 1977 and 1982 are illustrated in the top-left panel of Figure 3 (and similarly for 1987 and 1992 in the top-right panel), where we have divided each of the three inputs by the output (allowable since we are assuming CRS). For the observations from 1977 and 1982 it is obvious that there are differences between the two frontiers (with the 1982 observations having much higher capital levels) whereas there are no clear distinctions between the observations from 1987 and 1992. The lower panels illustrate the unit specific frontier shifts, FS (cf. equation 1), and here we see that all observations in 1977 and 1987 have FS-measures lower than 1, meaning that the 1982 frontier is worse than that in 1977, whereas the individual FS measures for the shift from 1987 to 1992 are centered around 1. Thus there is an obvious difference between the 1977 and 1982 frontiers, but not between the 1987 and 1992 frontiers, in correspondence with the significance probabilities in Table 4.

Concerning the efficiency change index in the second row of Table 4, we note that there in none of the cases are evidence against the hypothesis $(F_{t_1}, F_{t_2}) = (F_{t_2}, F_{t_1})$. This means that there is no evidence of efficiency changes between any of the time periods.

Concerning the Malmquist index, the test statistics are extreme for two of the three shifts (1977-1982, and 1982-1987). This means that there in these two situations are significant differences between the distributions. For the first shift (1977-1982) it is likely that the difference is due to changes in the frontiers, since T_{FS} is significant and T_{EC} is not. That the productivity change (decline) is, in fact, likely due to the frontier shift in particular, is evident from Figure 3.

For the shift from 1982 to 1987 there is a significant productivity change, but since neither of the T_{FS} and T_{EC} components are significant we have no additional information about the nature of the difference.

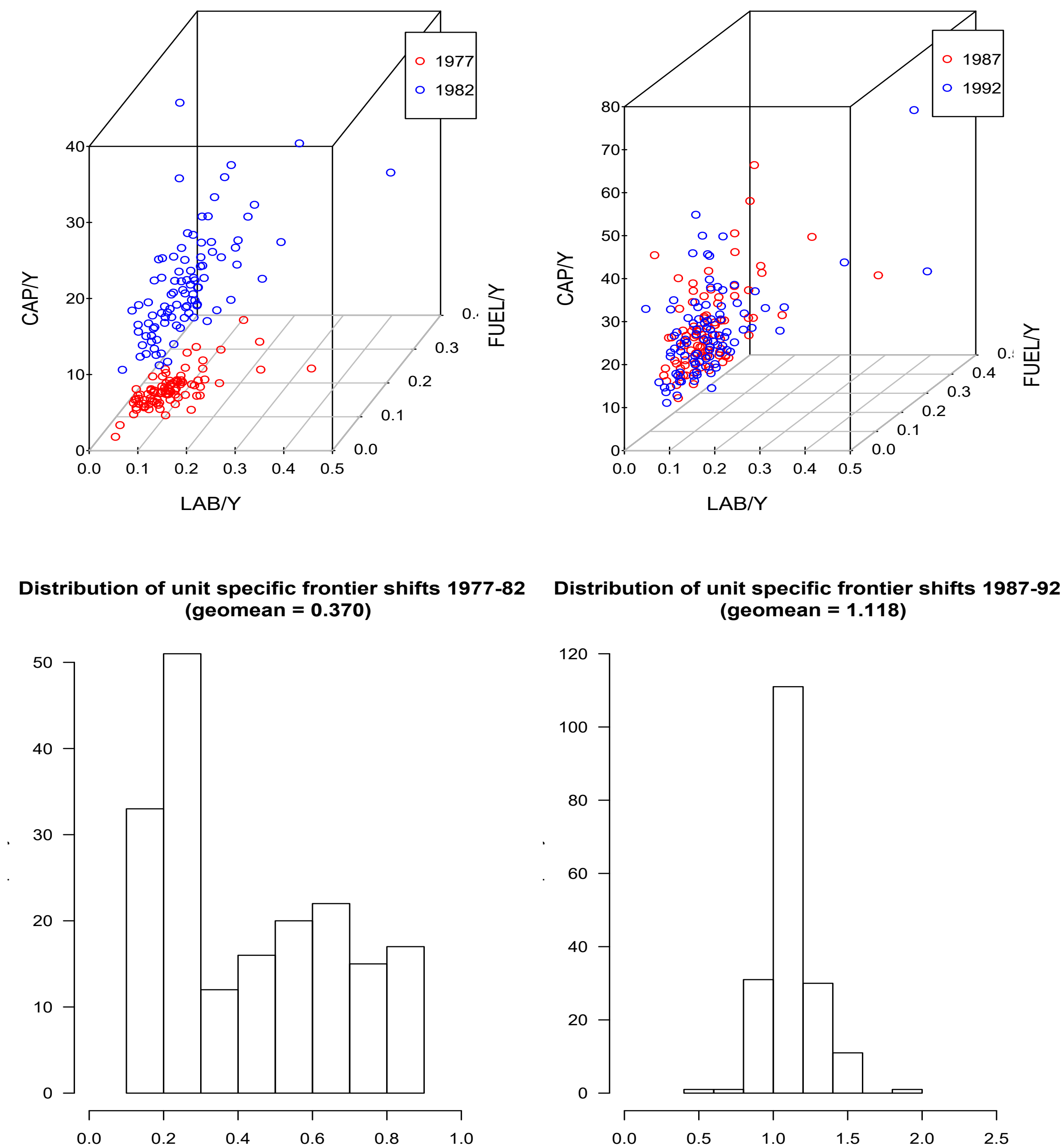

Figure 3: Top row shows the three inputs from Electric Power data for 1977 and 1982, resp. 1987 and 1992. The bottom row illustrate the unit specific frontier shifts, FS.

Finally, for the shift from 1987 to 1992 there is no evidence of any type of productivity change.

6 Conclusion

In this paper we have introduced exact tests (permutation tests) for the Malmquist index and its components (frontier shift and efficiency change) estimated using the non-parametric DEA approach. The tests do not depend on an assumed technology and a specific type of efficiency distribution, nor are they limited by the number of inputs and outputs. The tests are easily implementable and the R-code is shown in the Appendix. Further, it is described how to extend the the method to situations with unbalanced panel data.

Extensive simulation studies show that the suggested approach correctly detects differences between the two groups, as indicated by the individual tests relying on the Malmquist index, the frontier shift and (to some extent also) the efficiency change respectively. The proposed set of three tests is useful in the sense that when, for example, the Malmquist test identifies significant differences between time periods, then the other two tests generally are able to identify the nature of the difference, i.e. whether the difference is due to frontier shift or due to efficiency change.

The test procedure has furthermore shown to be useful when applied to an empirical case from the literature and we have provided the R-code for this example, showing that the approach is easily implementable in practice.

7 Appendix

The R-code used to implement the permutation test in the empirical illustration (balanced panel):

```
library(Benchmarking); library(EnvStats)
NN <- 10000 #number of permutations
N <- 93 #number of DMU's

#t1x and t1y are input and output matrices for time t1 and similary for t2
t1xy<-cbind(t1x,t1y)
t2xy<-cbind(t2x,t2y)
thetat1<-dea(t1x,t1y, RTS='crs',ORIENTATION="in")
```

```
thetat12<-dea(t1x,t1y,XREF=t2x,YREF=t2y, RTS='crs',
             ORIENTATION="in")
FSt1<-thetat1$eff/thetat12$eff
thetat21<-dea(t2x,t2y,XREF=t1x,YREF=t1y, RTS='crs',
              ORIENTATION="in")
thetat2<-dea(t2x,t2y, RTS='crs',ORIENTATION="in")
FSt2<-thetat21$eff/thetat2$eff
EC12<-thetat2$eff/thetat1$eff
FS<-geoMean(c(FSt1,FSt2) )
EC<-geoMean(EC12)
M<-FS*EC
perm_fun<-function(x,t){
  return(x*t)
}

FS_perm<-rep(0,NN)
EC_perm<-rep(0,NN)
M_perm<-rep(0,NN)
for (i in 1:NN){

  perm_index<-sample(0:1,size=N,replace=TRUE)

  t1xy_perm<-apply(t1xy,2,perm_fun, t=perm_index)+apply(t2xy,2,perm_fun, t=(1-perm_index))
  t2xy_perm<-apply(t2xy,2,perm_fun, t=perm_index)+apply(t1xy,2,perm_fun, t=(1-perm_index))

  t1x_perm<-t1xy_perm[,1:3] #three inputs
  t1y_perm<-t1xy_perm[,4] #one output

  t2x_perm<-t2xy_perm[,1:3] #three inputs
  t2y_perm<-t2xy_perm[,4] #one output

  efft1_perm<-dea(t1x_perm,t1y_perm, RTS='crs',ORIENTATION="in")
  efft12_perm<-dea(t1x_perm,t1y_perm,XREF=t2x_perm,YREF=t2y_perm, RTS='crs',
                  ORIENTATION="in")
  FSt1_perm<-efft1_perm$eff/efft12_perm$eff

  efft21_perm<-dea(t2x_perm,t2y_perm,XREF=t1x_perm,YREF=t1y_perm, RTS='crs',
                  ORIENTATION="in")
  efft2_perm<-dea(t2x_perm,t2y_perm, RTS='crs',ORIENTATION="in")
  FSt2_perm<-efft21_perm$eff/efft2_perm$eff
  EC12_perm<-efft2_perm$eff/efft1_perm$eff
  FS_perm[i]<-geoMean(c(FSt1_perm,FSt2_perm))
  EC_perm[i]<-geoMean(c(EC12_perm))
  M_perm[i]<-FS_perm[i]*EC_perm[i]
}

significance_probability_FS<-mean(abs(FS_perm-1)>abs(FS-1))
significance_probability_EC<-mean(abs(EC_perm-1)>abs(EC-1))
significance_probability_M<-mean(abs(M_perm-1)>abs(M-1))
```

8 Acknowledgement

The authors would like to thank C.A.K. Lovell for sharing the electricity data set used in the empirical illustration.

9 References

1. Asmild, M. and Tam, F. (2007). Estimating global frontier shifts and global Malmquist indices. *Journal of Productivity Analysis*, **27**, 137-148

2. Banker, R. D. and Natarajan, R. (2011). Statistical tests based on DEA efficiency scores. In *Handbook on data envelopment analysis* (Chap 11). Springer, Boston, MA.

3. Caves, D., Christensen, L. and Diewert, E. (1982). The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity. *Econometrica*, **50**, 1393-1414.

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

5. Christensen, L.R. and Greene, W.H. (1976). Economies of Scale in U.S. Electric Power Generation, *Journal of Political Economy*, **84**, 655-676.

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

7. Farrell, M.J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society*, **120**, 253-281.

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

9. Kneip, A., Simar, L. and Wilson, P.W. (2016). Testing Hyphoteses in Nonparametric Models of Production. *Journal of Business & Economic Statistics*, **34**, 435-456.

10. Kneip, A., Simar, L. and Wilson, P.W. (2015). When Bias Kills the Variance: Central Limit Theorems for DEA and FDH Efficiency Scores. *Econometric Theory*, **32**, 394-422.

11. Pastor, J.T., Asmild, M. and Lovell, C.A.K. (2011). The biennial Malmquist productivity change index, *Socio-Economic Planning Sciences*, **45**, 10-15.

12. Pastor, J.T. and Lovell, C.A.K. (2005). A global Malmquist productivity index, *Economics Letters*, **88**, 266-271.

13. Simar, L. and Wilson, P.W. (2015). Statistical Approaches for Nonparametric Frontier Models: A Guided Tour. *International Statistical Review*, **83**, 77-110.