Accepted Manuscript

Environmental Factors in Frontier Estimation - A Monte Carlo Analysis

Maria Nieswand, Stefan Seifert

PII: S0377-2217(17)30674-4
DOI: 10.1016/j.ejor.2017.07.047
Reference: EOR 14597

To appear in: *European Journal of Operational Research*

Received date: 3 June 2016
Revised date: 14 July 2017
Accepted date: 14 July 2017

Please cite this article as: Maria Nieswand, Stefan Seifert, Environmental Factors in Frontier Estimation - A Monte Carlo Analysis, *European Journal of Operational Research* (2017), doi: 10.1016/j.ejor.2017.07.047

Highlights

- Using Monte Carlo simulation we compare conditional DEA, latent class SFA and StoNEZD
- In 200 scenarios, we focus on estimators ability to account for environmental factors
- Latent class SFA outperforms cDEA and StoNEZD in most scenarios
- Noise-to-signal ratio is most important determinant of estimation accuracy

Environmental Factors in Frontier Estimation - A Monte Carlo Analysis

Maria Nieswand[a], Stefan Seifert[1b]

[a] *Loughborough University, Epinal Way, Loughborough, Leicestershire, UK, LE11 3TU*
[b] *Corresponding author, DIW Berlin – German Institute for Economic Research, Mohrenstrasse 58, D-10117 Berlin, Germany, sseifert@diw.de*

Abstract

We compare three recently developed frontier estimators, namely the conditional DEA (Daraio and Simar, 2005, 2007b), the latent class SFA (Orea and Kumbhakar, 2004; Greene, 2005), and the StoNEZD approach (Johnson and Kuosmanen, 2011) by means of Monte Carlo simulation. We focus on their ability to identify production frontiers and efficiency rankings in the presence of environmental factors. Our simulations match features of real life datasets and cover a wide range of scenarios with variations in sample size, distribution of noise and inefficiency, as well as in distributions, intensity, and number of environmental variables. Our results provide insight in the finite sample properties of the estimators, while also identifying estimator-specific characteristics. Overall, the latent class approach is found to perform best, although in many cases StoNEZD shows a similar performance. Performance of cDEA is most often inferior.

Keywords: OR in Energy, Monte Carlo Simulation, Environmental Factors, Conditional DEA, Latent Class SFA, StoNEZD
JEL: , C63, C14, D24

[1]This paper is partly produced as part of the KOMIED (Municipal infrastructure companies against the background of energy policy and demographic change) financed by Leibniz Association.

1. Introduction

1.1. Incorporating operational conditions in recent frontier approaches

Based on the seminal works of Koopmans (1951), Debreu (1951), and Farrell (1957), a wide range of frontier approaches for measuring the efficiency of firms and other decision making units (DMUs) evolved. Even in the very early stages of parametric frontier production models, Aigner et al. (1977, p. 25) mention "external events such as luck, climate, topography, and machine performance" as causes for deviations from the production frontier that are beyond the control of the DMUs and are considered different from productive inefficiency. Thus, frontier estimation needs to incorporate operational conditions when they affect the production possibilities; otherwise, efficiency estimates are meaningless. Consequently, numerous ways to account for operational conditions in frontier estimation are proposed. Among many others, influential parametric models include those by Kumbhakar and Hjalmarsson (1993), Battese and Coelli (1995), and Greene (2005). Non- and semi-parametric approaches include Cazals et al. (2002), Simar and Wilson (2007b), Bădin et al. (2012), Kumbhakar and Sun (2012), and Simar et al. (2016).

In this paper, we focus on three well established, yet newer, estimators and assess their ability to account for external factors[2] in a production setting by means of Monte Carlo Simulation (MCS). We compare the non-parametric conditional data envelopment analysis (cDEA, Daraio and Simar, 2005, 2007b), the parametric latent class stochastic frontier analysis (LC-SFA, Orea and Kumbhakar, 2004; Greene, 2005), and the stochastic semi-nonparametric envelopment of z variables data (StoNEZD, Johnson and Kuosmanen, 2011). We evaluate their performance in estimating the frontier and efficiency accurately. Respective conclusions are feasible because in controlled MCS environments all parameters and true values are known. Therefore, our analysis provides new insights into the finite sample properties of cDEA, LC-SFA, and StoNEZD. For this purpose, we measure their performance

[2]Throughout the paper, we use the terms operational conditions, environmental variables, and external factors interchangeably.

across a wide spectrum of scenarios with varying sample sizes and distributions of noise and inefficiency. We further consider different types of external factors (including the case with multiple environmental variables) and different intensities with which the operational conditions affect the frontier. Our analysis immediately confronts the estimators' performances against one another in the presence of heterogeneous operational conditions, which is of particular interest for most real-life applications of frontier analysis, such as benchmarking in electricity regulation. To address this, our data generating process (DGP) attempts to match features of real-life datasets, e.g., high correlations among inputs, non-normal distributions of environmental factors, and samples including a few large firms.

Although each of the three estimators is specifically designed to control for operational conditions, they differ considerably in the way they incorporate external factors. cDEA, for example, incorporates the environment by constructing for each unit a set of units of with a similar environment leading to a unit-specific effect of the environmental factors; LC-SFA accounts for a group-wise effect of the environment by grouping the units according to their environment, and StoNEZD assigns an average effect of the environment for all observations. As a result, the estimator also differ considerably in their statistical characteristics.

To the best of our knowledge, we are the first to provide a comparative MCS study covering non-parametric, parametric, and semi-parametric approaches while focusing on the estimators' ability to account for environmental factors. Closely related to our analysis is the work by Andor and Hesse (2014), who compare the stochastic non-smooth envelopment of data (StoNED), data envelopment analysis (DEA), and stochastic frontier analysis (SFA), i.e., the counterparts of the estimators we compare but without controlling for operational conditions. Our analysis complements their findings and, due to the methodological proximity, results of Andor and Hesse might give indications for the performance of cDEA, LC-SFA, and StoNEZD. The authors find SFA to perform best overall. However, StoNED performs well in noisy settings compared to DEA and SFA, but has a general tendency to underestimate the true frontier and is more sensitive to an increasing number of explanatory variables.

Further studies related to our analysis include Krüger (2012), who compares DEA, Free Disposable Hull (FDH), and SFA approaches with the more recently developed order-m and order-α estimators. Results indicate no advantage in using the more complex order-m and order-α approaches for well-behaved production settings and low levels of noise. Similar to our study, three other simulation studies focus on environmental factors in frontier estimation. Yu (1998) compares several SFA and DEA approaches, and finds a general advantage of SFA if the model is correctly specified. DEA approaches, such as two-stage Tobit DEA and the approach by Banker and Morey (1986), are found to perform reasonably well if the effect of the environmental variables is low, but performance deteriorates drastically otherwise. Cordero et al. (2009) compare different ways to account for environmental factors in DEA. The study finds that the four-stage model introduced by Fried et al. (2002) dominates other approaches. Finally, Cordero et al. (2016) compare conditional DEA and conditional FDH (cFDH) with DEA approaches. They find that cDEA performs best with respect to identifying efficient units, while other estimators can compete in terms of rank correlations.

The remainder of this paper is structured as follows. After a brief overview of regulatory benchmarking in Section 1.2, Section 2 outlines and compares the different frontier estimators. Section 3 explains the simulation design, the different scenarios considered in the simulation, and the implementation of the simulation. Section 4 presents the results, and Section 5 concludes.

1.2. Regulatory benchmarking in electricity distribution

To mitigate information asymmetries between regulated firms and regulators (see Laffont and Tirole, 1993), European regulators tend to combine price or revenue cap schemes with benchmarking techniques, especially in the regulation of electricity distribution networks (for overviews see Agrell et al., 2013b; Haney and Pollitt, 2009). Regulators approximate the unknown technologies of the firms by means of frontier estimation and use estimated inefficiency to set the firms' production or cost targets, which are ultimately translated into the respective regulatory outcome.

Given that the production process, costs, and observed data are likely to be influenced by external factors, the maximal (minimal) output (input) a firm is able to achieve, will vary due to the presence of external factors. Failing to control for the operational environment in this context is likely to transfer to the firms' inefficiency estimates and penalizes or favors firms and customers (see, e.g., Agrell and Brea-Solís, 2017). It remains, therefore, a challenge for regulatory authorities to implement appropriate benchmarking techniques.

Certainly, European regulators, who frequently implement DEA, SFA and variations of those, are well aware of the necessity of taking operational conditions into account. However, they also show interest in adapting their practice to methodological developments. In Finland, for example, StoNEZD replaced previously employed DEA and SFA approaches for reasonable pricing of electricity distribution network operators in the third regulatory period 2012-2015 (EMA, 2013). Further, the Norwegian regulatory model continuously advances, especially with respect to incorporating operational environments (Bjørndal et al., 2010). Monte Carlo studies, like ours, provide insights into how severe the deviations from the true frontier due to environmental factors are for a variety of scenarios and, therefore, contribute to improve actually implemented regulatory benchmarking procedures.

Apart from regulation, using information on external environments is a relevant issue in applications on energy companies (see, e.g., Anaya and Pollitt, 2017; Llorca et al., 2014). Emphasizing the broader relevance of our results for a variety of countries and industries, a large body of literature applies production frontier estimators to various other industries and network infrastructures worldwide (for overviews see, e.g., Estache et al., 2006; Coelli et al., 2003) as well as to banking (Aggelopoulos and Georgopoulos, 2017), health (Ancarani et al., 2009), and education services (Haelermans and De Witte, 2012; De Witte and Geys, 2013).

2. Methodology

We consider a production model with n $(i = 1, ..., n)$ DMUs, i.e., firms. Each firm employs M inputs $x_{i1}, ..., x_{iM}$ with the firm's input vector x_i to produce scalar output y_i. All firms have access to the same production technology with the production function $f(x)$ that gives maximum output for a given input level. A firm's actual output can deviate from the maximum due to random noise v_i, non-negative inefficiency u_i, and due to the impact $\delta = (\delta_1, ..., \delta_L)$ of L environmental factors $z_i = (z_{i1}, ..., z_{iL})'$. The multiplicative model is written as $y_i = f(x_i)exp(\varepsilon_i)$ with

$\varepsilon_i = \delta z_i + v_i - u_i$, and can be interpreted in two ways (Johnson and Kuosmanen, 2011): δz_i influences the location of the frontier and the attainable output for each i, or it influences i's distance to the production function.

2.1. Conditional DEA

Initially proposed by Cazals et al. (2002) and further developed by Daraio and Simar (2005, 2007a,b), cDEA is one extension of the standard DEA to incorporate environmental factors in performance evaluations. The approach aims to compare only units that operate under similar operational environments, i.e., the selection of the reference group for a particular observation is conditional on z. cDEA does not rely on the separability condition between the output space and the space of z-variables (Simar and Wilson, 2007a). Hence, z can influence the shape of the production set, and thus, the frontier.

cDEA estimates efficiency scores based on an attainable production set that is conditioned on a set of z-variables, denoted as $\hat{\Psi}^z_{DEA}$. The statistical properties of this estimator are derived in Kneip et al. (2008), and consistency is established in Jeong et al. (2010). Estimating $\hat{\Psi}^z_{DEA}$ implies the estimation of a nonstandard conditional distribution function, where the production process is conditional to a particular level of z (Daraio and Simar, 2007b; Bădin et al., 2010). Since the latter requires the application of a smoothing technique, we conduct kernel estimation. The kernel K is defined as

$$K_h = \frac{|z_i - z_k|}{h}, \tag{1}$$

where z_i is the vector of z-variables of the unit of interest i, z_k is the vector of all other observations, and h is the vector of selected bandwidths. The kernel function in eq. (1) provides firm-specific kernel probabilities that are used to define firm-specific reference sets. The firms closely located to firm i in terms of z thereby receive higher kernel probabilities, whereas small (or even zero) kernel probabilities are assigned to firms facing very different operating environments than firm i.

The cDEA efficiency measure $\hat{\theta}_{DEA}(x, y|z)$ for a single observation assuming variable returns to scale is defined as (Daraio and Simar, 2007b)

$$\hat{\theta}_{DEA}(x,y|z) = max\ \{\theta > 0 \mid \theta y \leq \textstyle\sum_{i|z-h\leq z_i\leq z+h}^{n} \lambda_i y_i, \quad x \geq \textstyle\sum_{i|z-h\leq z_i\leq z+h}^{n} \lambda_i x_i, \quad \textstyle\sum_{i|z-h\leq z_i\leq z+h}^{n} \lambda_i = 1, \quad i = 1,...,n\}. \tag{2}$$

For each observation, the bandwidths determine the range of z in which other observations are considered to be similar. Hence, only observations within this range are selected into the respective reference group, and are considered as potential references for the unit of interest. For implementation, the reference set of firm i is restricted to those firms with positive kernel probabilities. The frontier reference point of firm i is obtained by $\hat{y}_i = \hat{\theta}_{i,DEA}(x_i, y_i|z_i)y_i$.

2.2. Latent class SFA

The LC-SFA estimator, proposed by Greene (2002), Caudill (2003) and Orea and Kumbhakar (2004), belongs to the class of stochastic and parametric approaches. It accounts for heterogeneity among firms by endogenously sorting them into a prespecified number of groups. For each group, a separate frontier is estimated, and each firm gets a probability to belong to each group. The group-specific frontiers are allowed to differ in their parameters, thus, in their shapes. Therefore, contrary to standard SFA with environmental factors, LC-SFA accounts for a technology shift induced by the environmental factor and technological heterogeneity.

LC-SFA estimates the group-specific parameters for n observations with J ($j = 1,...,J$) groups of the form $\log y_i = f(x_i, \beta_j) - u_{i|j} + v_{i|j} = f(x_i, \beta_j) + \epsilon_{i|j}$. This can be estimated via maximum Likelihood (ML), and requires further assumptions: an a priori specified functional form for $f(x)$, e.g., Cobb-Douglas or Translog. Distributional assumptions are necessary for the noise and inefficiency components, which typically enter the likelihood function as normal noise $v \sim N(0, \sigma_v)$ and half-normal inefficiency $u \sim N^+(0, \sigma_u)$. Given these assumptions, a log-density for firm i for each group j can be imposed with the standard parameterization $\sigma_j^2 = \sigma_{uj}^2 + \sigma_{vj}^2$,

$\lambda_j = \sigma_{uj}^2/\sigma_{vj}^2$ and $\epsilon_{ij} = \epsilon_i(\beta_j) = \log y - \log f(x_i, \beta_j)$ such that

$$\log LF_{ij} = -\frac{1}{2}\log\frac{\pi}{2} - \frac{1}{2}\log\sigma_j^2 + \log\Phi\left(-\frac{\lambda_j\epsilon_{ij}}{\sqrt{\sigma_j^2}}\right) - \frac{\epsilon_{ij}^2}{\sigma_j^2}. \tag{3}$$

The contribution of firm i to the final likelihood function is the likelihood for each class j, LF_{ij}, weighted with the probability to belong to j, P_{ij}, such that

$$LF_i = \sum_{j=1}^{J} LF_{ij}P_{ij}. \tag{4}$$

The probability to belong to a certain group uses a multinomial logit with standard assumptions on probabilities ($0 \leq P_{ij} \leq 1; \sum_{j=1}^{J} P_{ij} = 1$). At this point, z_i enters the likelihood function and determines the probabilities of each firm to belong to each of the J groups, such that

$$P_{ij}(\zeta_j) = \frac{\exp(\zeta_j z_i)}{\sum_{j=1}^{J}\exp(\zeta_j z_i)}, \tag{5}$$

with ζ as logit parameters to estimate and $\zeta_J = 0$. Given this parameterization, the final log-likelihood function to be maximized is obtained as

$$LF = \sum_{i=1}^{n}\left\{\sum_{j=1}^{J} LF_{ij}P_{ij}(\zeta_j)\right\} \tag{6}$$

Note that each observation enters the likelihood function J times and can influence the shape of all group frontiers, depending on the weight $P_{ij}(\zeta_j)$. As a result, an observation has a reference point on each group frontier. To obtain one final reference point, one can either use the group frontier with the highest probability, or calculate a weighted reference point using the conditional posterior class probabilities $P(j|i)$ (Orea and Kumbhakar, 2004). Following Greene (2002), $P(j|i)$ can be calculated as $P(j|i) = \frac{LF_{ij}P_{ij}(\zeta_j)}{\sum_{j=1}^{J} LF_{ij}P_{ij}(\zeta_j)}$. We use the weighting approach to incorporate more information about the underlying data structure. Thus, a frontier reference point is calculated as probability weighted reference points from the J frontiers such that $\log \hat{y}_i = \sum_{j=1}^{J} P(j|i) \log \hat{f}(x_i, \beta_j)$.

2.3. StoNEZD

StoNEZD, proposed by Johnson and Kuosmanen (2011), is a semi-nonparametric approach that extends the standard StoNED estimator (Kuosmanen and Kortelainen, 2012). It incorporates an average effect of the operational environment common to all firms, and the production frontier is estimated in two stages: The first stage estimates an average production function $g(x)$ with concave non-parametric least squares (CNLS, Hildreth, 1954) accounting for z. In the second stage, $g(x)$ is shifted upwards to obtain a frontier estimate. This shift is based on the expected value of inefficiency derived from the residuals from the first stage based on parametric assumptions.
Johnson and Kuosmanen (2011) show that a multiplicative model $y_i = f(x_i)exp(\kappa_i)$ can be estimated with such a two-stage approach. For the first stage, a quadratic programming problem (QP) estimates the shape of $g(x)$ without any assumptions on a functional form but establishes concavity and monotonicity of the production function. Further, the QP directly takes into account that the firm's deviation from this average is influenced by the existence of z. No distributional assumptions for u and v are necessary in this stage, but u, v and z are assumed to be uncorrelated. To estimate $g(x)$, Johnson and Kuosmanen (2011) propose a minimization of the squared residual accounting for z using the following constrained QP:

$$\min_{\alpha,\beta,\delta,\phi} \sum_{i=1}^{n} (\log y_i - \log \hat{\phi}_i - \delta z_i)^2 \tag{7}$$

$$\begin{aligned} s.t. \quad & \hat{\phi}_i = \alpha_i + x_i\beta_i && \forall i = 1, ..., n, \\ & \alpha_i + x_i\beta_i \leq \alpha_h + x_i\beta_h && \forall h, i = 1, ..., n, \\ & \beta_i \geq 0 && \forall i = 1, ..., n. \end{aligned}$$

This QP estimates firm-specific coefficients α_i and β_i that can be interpreted as the marginal products of the inputs. They create linear hyperplanes $\hat{\phi}$, which are tangent to the average production function and deliver the fitted values on $\hat{g}(x)$. Microeconomic requirements on production functions are imposed as constraints in eq. (7): The first constraint establishes a linear form of the hyperplanes. The second constraint imposes concavity using Afriats inequalities (Afriat, 1967). These concavity constraints relate the piece-wise linear hyperplanes for all observations against each other leading to n^2 separate constraints. The third constraint ensures monotonicity of the estimated average production function. Further, the QP estimates the impact of the environmental factor, δ, which is identical for all firms. A residual for each observation containing noise and inefficiency is given by $\hat{\eta}_i = \log y_i - \log \hat{\phi}_i - \hat{\delta} z_i = \hat{\kappa}_i - \hat{\delta} z_i$. In the second stage, to shift the average production function $\hat{g}(x)$ to the frontier, the residuals $\hat{\eta}_i$ of the QP are used to estimate the expected value of inefficiency μ. Further distributional assumptions for noise and inefficiency are necessary. Following Kuosmanen and Kortelainen (2012), we assume a half-normal distribution for the inefficiency term, $u \sim N^+(0, \sigma_u)$, and a normal distribution for the noise term, $v \sim N(0, \sigma_v)$. Following Aigner et al. (1977), Kuosmanen and Kortelainen (2012) suggest using a method of moments (MM) estimator to derive the expected value of inefficiency.[3] This approach uses the property of the third central moment of a normal-half-normal residual to be a function of only one parameter, σ_u. Using the empirical third moment of the residuals, $\hat{M}_3$, an estimate $\hat{\sigma}_u$ can be recovered by calculating $\hat{\sigma}_u = \sqrt[3]{\frac{\hat{M}_3}{\sqrt{\frac{2}{\pi}}[1-\frac{4}{\pi}]}}$. Subsequently, the expected value of inefficiency is calculated as $\hat{\mu} = \hat{\sigma}_u \sqrt{2/\pi}$. Now, the frontier is derived as $\hat{f}(x) = \hat{g}(x) exp(\hat{\mu})$. Firm-specific frontier reference points can be estimated from the shifted average production function accounting for the impact of z, such that $\hat{y}_i = \hat{g}(x_i) exp(\hat{\delta} z_i) exp(\hat{\mu})$ for each observation.

2.4. Comparison of the estimators

The three estimators differ in their characteristics considerably. First, their a priori assumptions on the production process differ since cDEA is a non-parametric, StoNEZD a semi-parametric, and LC-SFA a parametric estimator. While cDEA

[3] Kuosmanen and Kortelainen (2012) also propose a pseudo-likelihood (PSL) approach following Fan et al. (1996) to estimate μ. We use the MM estimator due to its computational ease.

and StoNEZD need only few assumptions on the technology set, e.g., monotonicity, convexity of the production set and certain scaling assumptions, LC-SFA demands a functional form to be specified in advance (e.g., Cobb-Douglas or Translog). As a result, cDEA and StoNEZD are more flexible in the frontier estimation; however, LC-SFA can also estimate non-convex production sets. This difference in the nature of the estimators is also reflected in their asymptotic properties, and in particular in their rates of convergence: LC-SFA converges with the standard parametric rate of convergence, $n^{-1/2}$. cDEA converges with $n^{-2/(m+q+1)}$ with m and q being the number of inputs and outputs, respectively. Further, the conditional version is slower by factor $n^{-4/4+r}$ with r being the number of z-variables (Kneip et al., 1998; Jeong et al., 2010). For StoNEZD, Johnson and Kuosmanen (2011) show that δ converges with the standard parametric rate, $n^{-1/2}$, while convergence rate of the CNLS estimation is unknown (Kuosmanen et al., 2015). However, following Stone (1980, 1982), Johnson and Kuosmanen (2011) suggest an upper limit $n^{-2d/(2d+m)}$ with d being the degree of differentiability and m the number of inputs. Thus, both, cDEA and StoNEZD suffer from the "curse of dimensionality" with increasing data demand for additional dimensions on the input and output side.

Second, the estimators differ in the treatment of noise and inefficiency. cDEA is purely deterministic and the existence of noise is not considered, i.e., $\sigma_v^2 = 0$. Thus, the estimator is prone to outliers, which can lead to an overestimation of the frontier. On the contrary, LC-SFA and StoNEZD allow a differentiated treatment based on distributional assumptions with $\sigma_v^2 \geq 0$, which makes them less prone to noise. However, frontier and efficiency estimates depend on the distributional assumption for noise and inefficiency.

Third, the estimators vary considerably regarding the incorporation of environmental factors. cDEA constructs observation-specific reference sets depending on the realization of z and the estimated bandwidths. Thus, the estimated effect of z on the frontier is observation-specific. LC-SFA uses all observations in the frontier estimation, and the reference sets are weighted with the probabilities of group membership. As a result, with LC-SFA the effect of z-variables on the frontier is also weighted and varies from one firm to another. StoNEZD uses the whole sample as a reference set resulting in an effect of z on the frontier that is common to all observations.

Fourth, StoNEZD is a two-stage approach, cDEA is a one-stage approach estimated in two stages, and LC-SFA is a one-stage approach. Under cDEA, inefficiency and potential noise may influence the estimated bandwidths, which ultimately influence the reference sets. Similarly, the δ coefficient in the first stage of the StoNEZD might capture effects from the so-far neglected inefficiency term. These problems should not be present in the LC-SFA estimation, since LC-SFA considers environmental factors, noise, and inefficiency in one estimation.

And fifth, the approaches differ with respect to the identification of reference units. cDEA clearly identifies the reference set with similar units, and the units that span the frontier. StoNEZD identifies at least the observations that share a facet of the frontier. Such identification is not possible with LC-SFA. Further, the weighting approach in the LC-SFA can project all firms on different levels between multiple group frontiers.[4]

3. Simulation design and implementation

3.1. The data generating process

Using MCS, we can assess the performance of the estimators because the true position of the frontier and its characteristics are known. By varying the parameters of the DGP, the reaction of the estimator to these changes can be evaluated. Our DGP attempt to incorporate features of real-life datasets, as they could be also encountered by practitioners, such as regulatory authorities.[5]

To construct our datasets, we calculate a one-dimensional output y_i for each of the n observations. y_i is a function of the input vector x_i, which collects the M inputs $(x_{i1}, ..., x_{iM})$ that are transformed into outputs using a production function $f(x)$. L environmental factors $z_i = (z_{i1}, ..., z_{iL})'$ influenced the maximum output $f(x_i)$ with impact $\delta = (\delta_1, ..., \delta_L)$ that is common to all firms. The observed output of a firm

[4] The same applies for the counterparts of the estimators that do not control for the environment, DEA, SFA, and StoNED, respectively.

[5] To calibrate the parameters of the DGP, we use features of regulatory datasets. E.g., the sample composition, the number of and correlation among inputs, and the specification of environmental factors are inspired by analysis of regulatory data from Finland (Kuosmanen, 2012), Norway (Bjørndal et al., 2010), and Germany (Agrell et al., 2013a).

observed may deviate from this maximum due to noise v_i and inefficiency u_i. The output y_i is calculated as

$$y_i = f(x_i)exp(\delta z_i)exp(v_i - u_i) \tag{8}$$

Our defined sample sizes reflect that real datasets often contain heterogeneous firm sizes with few considerably larger firms. We assume that a sample contains n_{small} small and n_{large} large firms with $n_{small} = \{25, 50, 100, 150, 250\}$. n_{large} are additional 4% of the small firms, i.e., $n_{large} = 0.04n_{small}$. Thus, the total number of observations in the sample is $n = \{26, 52, 104, 156, 260\}$.
We assume noise to follow a normal and inefficiency a half-normal distribution, i.e., $v \sim N(0, \sigma_v^2)$ and $u \sim |N(0, \sigma_u^2)|$. We set $\sigma_v = \{0.01, 0.1\}$ and $\sigma_u = \{0.1, 0.3\}$, resulting in four potential noise/inefficiency set-ups with noise-to-signal ratios between $0.0\bar{3}$ and 1, and expected values of inefficiency, μ, of about 8% and 20% (see Table 1).

σ_u \ σ_v	0.01	0.1	μ
0.1	0.1	1	8.0%
0.3	$0.0\bar{3}$	$0.\bar{3}$	20.1%

Table 1: Noise-to-signal ratios and expected inefficiency

Each DMU employs $M = 4$ inputs, which are correlated uniform variables. We define the range of the inputs to vary between $[0, 10]$ for small firms, and $[20, 30]$ for large firms. Correlations among the inputs are given by the following matrix indicating moderate to high correlation between 0.55 and 0.85.

$$\rho_X = \begin{bmatrix} 1 & 0.55 & 0.65 & 0.75 \\ & 1 & 0.7 & 0.8 \\ & & 1 & 0.85 \\ & & & 1 \end{bmatrix} \tag{9}$$

For the production function $f(x)$, we assume a Translog specification, such that

$$f(x) = p_o \prod_{m=1}^{M} x_m^{(p_m)} \prod_{m=1}^{M} x_m^{(1/2)\sum_{l=1}^{M} p_{ml} \log x_l}. \quad (10)$$

Decreasing returns to scale are imposed with $p_m = 0.15$, $p_{11} = -0.10$, $p_{22} = -0.15$, $p_{33} = 0.10$, $p_{44} = 0.15$, and with symmetric cross-terms $p_{12} = 0.0$, $p_{13} = 0.01$, $p_{14} = 0.05$, and $p_{23} = p_{24} = p_{34} = -0.1$.[6]

We define $L = 3$ environmental factors, z_1, z_2, z_3. z_1 is drawn from $N(1, 0.15^2)$ and symmetrically distributed with mean one while values below zero are basically ruled out. z_2 is drawn from $TExp(5.5, 1)$, an exponential distribution with rate 5.5, truncated at 1, and with a mean of 0.178. Thus, only realizations in $(0, 1]$ are possible and small values of z_2 are more likely than values close to one. z_3 is drawn from $\Gamma(2, 1)$, a Gamma distribution with mean two. Again, values below zero cannot be observed, but the distribution is not truncated at the upper end. The distributions of the variables resemble operational conditions observed in real reuglatory data: Using a Kolmogorov-Smirnov (KS) test indicates that the percentage of underground cabling used in regulation of Finnish electricity distribution system operators (see Kuosmanen, 2012) stems, like z_2, from a truncated exponential distribution. Likewise, the KS test shows that the snow fall variable in the Norwegian DSO regulation follows a Gamma distribution.

3.2. Scenarios considered

Table 2 lists our ten scenarios. First, we construct three baseline scenarios $BL1$ to $BL3$ which include one environmental factor each, z_1 to z_3. For these baseline scenarios, the impact of the environmental variables on the frontier is small to moderate, with an average frontier shift of about 5% (cp. Table 2).

Second, in three high impact scenarios ($HI1, ..., HI3$) we triple the impact of the environmental variables leading to an average frontier shift between 14.7% and 15%.

[6] We thank an anonymous referee for the suggestion. As will be outlined in Section 3.3, we will further ensure monotonicity and concavity of the estimated production function in the implementation of the DGP.

	z_l	ρ_z	δ_l	1-exp($\delta_l \bar{z}_l$)
$BL1$	z_1	0	-0.05	4.9%
$BL2$	z_2	0	-0.3	5%
$BL3$	z_3	0	-0.025	5%
$HI1$	z_1	0	-0.15	14.7%
$HI2$	z_2	0	-0.9	15%
$HI3$	z_3	0	-0.075	15%
$MZ1$	z_1, z_2	0	-0.05, -0.3	9.8%
$MZ2$	z_1, z_2	0	-0.1, -0.6	19.6%
$MZ3$	z_1, z_2	0.5	-0.05, -0.3	9.8%
$MZ4$	z_1, z_2	0.5	-0.1, -0.6	19.6%

Table 2: Scenarios

Thus, in HI scenarios with low inefficiency ($\sigma_u = 0.01$), the average deviation of the firms from the frontier is driven more by the environment than by inefficiency.
Third, we construct four scenarios with multiple z-variables ($MZ1, ..., MZ4$), for which two environmental factors impact the frontier. For these scenarios, the average frontier shift varies between 9.8 and 19.6%. We use z_1 and z_2 to influence the firms' production potentials. In scenarios $MZ1$ and $MZ2$, the environmental variables are uncorrelated, whereas in $MZ3$ and $MZ4$, we set $\rho_{z_1,z_2} = 0.5$.

3.3. Implementation

To cover a large spectrum of potential datasets, we estimate each of the 10 scenarios for each sample size and for each noise-to-signal ratio in Table 1, resulting in 10·5·4=200 cases to simulate. We use 200 replications ($R = 200$) with fixed seeds, resulting in a total of 40,000 estimations for each estimator. This offers certain generality and a solvable number of cases to analyze.
For each of the 200 cases, we generate one set of inputs, environmental factors, and inefficiencies. We generate R random draws of noise and then calculate the observed output. Thus, variation among the R draws stems only from variation in noise. All estimations run with the correctly specified model, i.e., all inputs are considered. To ensure comparability of the datasets, we discard sets and redraw if (I) at least one

correlation deviates from the input correlation matrix by ± 0.05, (II) the resulting setting is not quasi-concave or monotonically increasing for more than 90% of the observations, (III) the composite error term $(v - u)$ has wrong (positive) skewness. For the StoNEZD estimator, the expected value of inefficiency is estimated using a method of moments estimator. Following Kuosmanen and Kortelainen (2012), if wrong skewness occurs in the estimation of the expected inefficiency, i.e., $\hat{M}_3 > 0$, we set $\hat{M}_3 = -0.0001$. For the LC-SFA estimator, we use a Cobb-Douglas specification for all sample sizes.[7] To choose the optimal number of groups, LC-SFA is estimated for $J = \{2, 3, 4\}$ and the estimation with the optimal Bayesian Information Criterion (BIC) is chosen and reported. Further, to reduce the risk of local optima in the ML procedure, optimization is carried out five times for each J with random starting values, and the solution with the best BIC is reported. For cDEA, we use an Epanechnikov kernel following Daraio and Simar (2005, 2007b). Bandwidths are computed using least squares cross validation proposed by Hall et al. (2004) and Li and Racine (2007, 2008).[8, 9]

Although our simulation aims at implementing a fair comparison of the estimators, some components may create estimator-specific (dis)advantages. Generally, the used production function for simulation (eq. 8) resembles the underlying model of the StoNEZD estimator, but does not contradict assumptions of the other approaches. The check for concavity and monotonicity of the production function enables StoNEZD and cDEA to compete with LC-SFA, which would be able to estimate non-concave settings. The specification of a normal noise and half-normal inefficiency matches the model assumptions for LC-SFA and StoNEZD, and a different specification might challenge these estimators additionally. This includes that we do not include fully efficient firms, which can disadvantage cDEA. However, although we model noise in every simulation, the small noise component in scenarios with

[7] This allows us to also estimate small samples, because the number of parameters to estimate increases with J, and, e.g., a LC-SFA with three groups and four inputs has 48 parameters in a Translog specification, but only 15 if a Cobb-Douglas model is estimated.

[8] The estimated efficiency scores depend on the kernel smoothing; other kernels and bandwidth selection procedures are available, see, e.g., Cazals et al. (2002); Daraio and Simar (2005, 2007b).

[9] We implement StoNEZD using GAMS, and LC-SFA and cDEA using R with the packages np, minqa, lpSolveAPI, and micEcon. Simulations run on a 32 CPU 2.8 MHz AMD with 512 GB RAM.

$\sigma_v = 0.01$ offers the deterministic cDEA settings close to its assumptions. Finally, for LC-SFA, the strong correlation of inputs might lead to problems of multicollinearity, while StoNEZD and cDEA may suffer from the curse of dimensionality due the fairly high number of inputs.

4. Results

4.1. Performance measures

We evaluate the performance of cDEA, StoNEZD and LC-SFA with respect to the frontier estimate and the rankings of the efficiency scores. Following the literature, we consider the bias and the mean squared error (MSE) as criteria for frontier estimation accuracy (e.g., Andor and Hesse, 2014; Kuosmanen et al., 2013). The bias delivers the average total deviation from the true frontier as a percentage. A positive (negative) sign indicates overestimation (underestimation) of the frontier, i.e., DMUs' output targets and inefficiencies are too high (low). The MSE is the average squared deviation from the frontier and is greater than or equal to zero. It penalizes larger deviations more strongly, and a higher value indicates a stronger deviation from the true frontier. For bias and MSE values close to zero are desirable. We calculate a firm's optimal output given its environmental conditions, $f^z(x_i)$, using the true value on the frontier $f(x_i)$ corrected by the effect of the environmental variable, i.e., $f^z(x_i) = f(x_i)exp(\sum_l \delta_l z_{li})$. Based on the estimated frontier reference point $\hat{f}^z(x_i)$ for n observations and R simulation replications, the performance measures are defined as:

$$BIAS = \frac{1}{nR}\sum_{r=1}^{R}\sum_{i=1}^{n}\frac{\hat{f}^z(x_i) - f^z(x_i)}{f^z(x_i)} \tag{11}$$

$$MSE = \frac{1}{nR}\sum_{r=1}^{R}\sum_{i=1}^{n}\left(\frac{\hat{f}^z(x_i) - f^z(x_i)}{f^z(x_i)}\right)^2 \tag{12}$$

To evaluate performance in terms of efficiency estimation, we calculate efficiency scores as the ratio of a DMU's observed and estimated optimal output, i.e., $\hat{TE}_i =$

$y_i/\hat{f}(x_i|z_i)$.[10] The true inefficiency is given by $TE_i = exp(-u_i)$. To analyze rank correlation coefficients, we use Kendall's τ, for which a value $\tau = 1$ ($\tau = -1$) indicates an identical ranking (a perfect inversion) of true and estimated efficiency scores, while a value of 0 indicates no association of the two measures. Defining n_c as the number of concordant pairs and n_d as discordant pairs, τ is calculated as

$$\tau = \frac{n_c - n_d}{n(n-1)/2}. \tag{13}$$

4.2. Precision of frontier estimates

4.2.1. Baseline scenarios

First, we analyze the precision of the frontier estimates for the BL scenarios in which the defined environmental variables only moderately influence the frontier. Figures 1 to 3 show BIAS and MSE for the BL scenarios on the y-axes of the upper and lower row of panels, respectively.[11]. Noise and inefficiency parameters are indicated on top of each column and sample sizes are on the x-axes. Thus, the panel in the upper left corner of Figure 1 (BIAS panel 1) depicts the BIAS for the low-noise-low-inefficiency setting (σ_v=0.01, σ_u=0.1) when z_1 has a moderate effect on the frontier. Accordingly, the panel in the bottom right corner of Figure 1 (MSE panel 4) shows the MSE for the high-noise-high-inefficiency setting (σ_v=0.1, σ_u=0.3). cDEA, LC-SFA and StoNEZD results are represented by filled rectangles, circles, and filled triangles, respectively.

In $BL1$ the environmental variable z_1 is drawn from a normal distribution. For cDEA, results indicate a positive BIAS between 0.1 and 0.4 in almost all cases (compare panels in top row of Figure 1). This indicates a general overestimation of the frontier and unreasonably high output requirements for firms to be considered as efficient. This upward bias is less pronounced in scenarios with high inefficiency ($\sigma_u = 0.3$, compare BIAS panel 1 with panel 2, and panel 3 with panel 4), but partly increases with additional noise (compare BIAS panel 2 with panel 4). Likewise, the

[10]Alternatively, for StoNEZD and LC-SFA stochastic efficiency estimators could be used (e.g., Jondrow et al., 1982), which are, however, inconsistent. As Kuosmanen (2012) points out, a consistently estimated frontier could be more suitable than an inconsistent point estimate of efficiency.

[11]Numerical results are provided in the supplementary material.

MSE for cDEA shows deviations of up to 0.35, indicating considerable error in the estimation. As for the bias, the lowest MSE values are found for cases with low noise and high inefficiency, although extreme MSEs occur sporadically, e.g., MSE panel 4 with n=156. Regarding sample sizes, results are not conclusive, but indicate a good performance with small samples that deteriorates in larger samples, especially with high noise-to-signal ratios (i.e., $\sigma_v/\sigma_u = 0.1/0.1$ and $\sigma_v/\sigma_u = 0.1/0.3$).

For the LC-SFA, the results show small biases (between -0.05 and +0.1) and MSE values (below 0.035) for all noise-to-signal ratios. However, while in low inefficiency cases ($\sigma_u = 0.1$) an upward bias can be observed, a slight underestimation is found for cases with high inefficiency (compare BIAS panels 1 and 3 with 2 and 4). Nonetheless, this has little effect in terms of MSE. Further, both BIAS and MSE indicate a good performance irrespective of the considered sample size.

The BIAS and MSE results obtained for StoNEZD indicate a moderate performance in low inefficiency cases with biases fluctuating around zero (mostly between -0.05 and +0.13) and low MSE values (up to 0.06). However, in high inefficiency cases, a considerable underestimation of the frontier occurs indicated by the negative BIAS and resulting in relatively large MSEs (compare MSE panels columns 2 and 4). This bias of about 0.2 corresponds to the expected value of inefficiency in these settings (20%). For all considered cases, StoNEZD performs well with small samples, but no systematic reduction of bias and MSE occurs with increasing sample size.

Summarizing the $BL1$ results, cDEA exhibits the largest differences between true and estimated frontiers among the three estimators and undercompensates for environmental conditions in most cases. The LC-SFA estimator is found to estimate the true frontier most precisely in nearly all simulations, but with tendencies to overestimate (underestimate) the frontier in low (high) inefficiency settings. StoNEZD shows a moderate performance but overcompensates in cases with high inefficiency.

Figures 2 and 3 show the results of $BL2$ and $BL3$, the baseline scenarios with truncated exponentially and Gamma distributed environmental factors. The patterns vary only slightly between the three baseline scenarios and the distribution of the z-variables seems to be of minor importance if the environmental factors' impact is low. All estimators show a fairly good performance with the smallest sample ($n = 26$), but no systematic reduction of bias and MSE occurs with increasing sam-

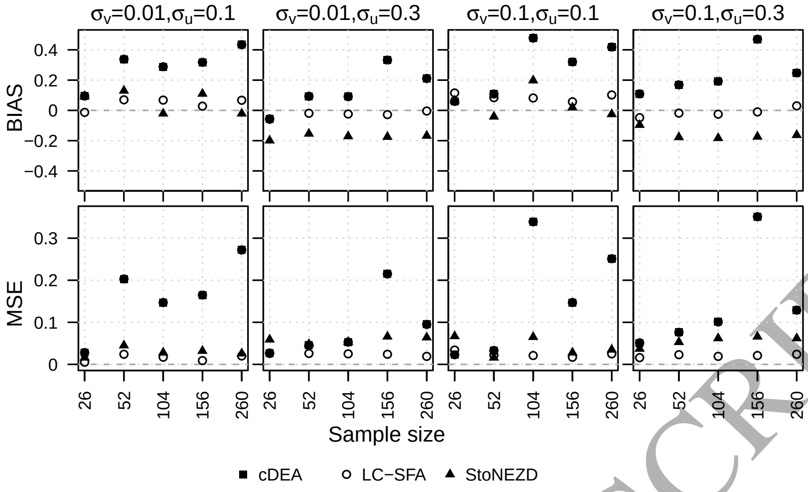

Figure 1: BL1 - Bias (top) and MSE (bottom)

ple size for any of the three. cDEA overestimates the frontier in nearly all cases with performance decreases with additional noise and increasing sample size. LC-SFA performs well in all cases in terms of MSE, but indicates an upward bias in low inefficiency cases. This upward bias is slightly more pronounced for the non-Gaussian z-variables in $BL2$ and $BL3$. On the contrary, performance of StoNEZD improves in these two scenarios, and the estimator outperforms the competitors in cases with low inefficiency. Further, an underestimation of the frontier in simulations with high inefficiency persists, but, compared to $BL1$, the absolute StoNEZD bias decreases (compare Figures 1 to 3, columns 2 and 4). This suggests that the estimator can better account for the non-Gaussian distributions of z-variables in $BL2$ and $BL3$.

4.2.2. High impact scenarios

In the next step, we triple the impact of the operational conditions on the firms' output potential by tripling the δ-coefficient. As a result, in cases with low inefficiency ($\sigma_u = 0.1$), the average effect of the environment is stronger than the average output loss due to inefficiency, i.e, $|\delta\bar{z}| > \mu$ and $exp(\delta\bar{z}) < exp(-\mu)$. Figures 4 to 6 show the results for these high impact scenarios $HI1$ to $HI3$.[12]

[12]Detailed numerical results are provided in the supplementary material.

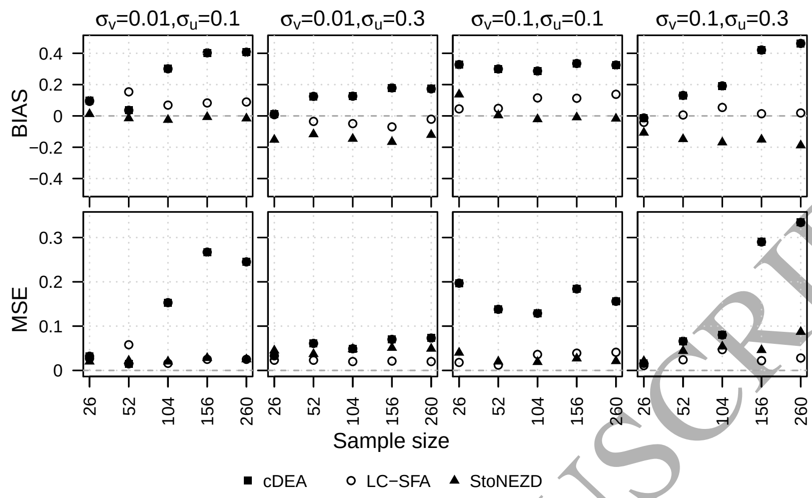

Figure 2: BL2 - Bias (top) and MSE (bottom)

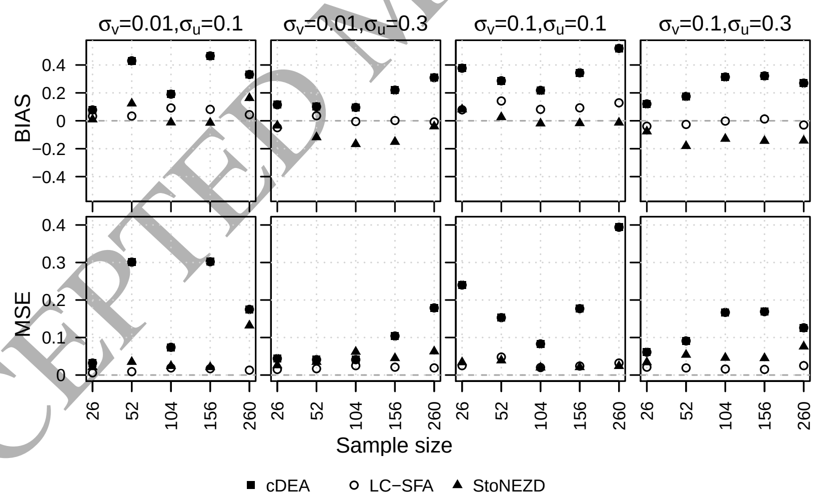

Figure 3: BL3 - Bias (top) and MSE (bottom)

$HI1$ results show generally similar patterns as the BL scenarios. For cDEA, results indicate in nearly all cases a considerable upward bias (compare Figure 4, upper row), i.e., an overestimation of the DMU's output potential. Bias and MSE increase with increasing noise levels (compare panel columns 1 with 3, and 2 with 4), while accuracy increases for higher levels of inefficiency (compare panel columns 1 with 2, and 3 with 4). As a result, cDEA, like in the BL scenarios, can compete with the other approaches, but only in cases with very low noise-to-signal ratios. For LC-SFA, results indicate again an overall good performance and the lowest MSE values among the three estimators in most settings. However, while the bias is most often negative for high inefficiency cases, an average 7% overestimation of the frontier occurs in the low inefficiency cases (compare BIAS panels 1 and 3). For StoNEZD, results indicate for the low inefficiency cases very small negative biases with an underestimation of the frontier ($\sim$ 1% to 3%). The corresponding low MSE values are competitive to LC-SFA results (compare panel columns 1 and 3). However, similar to the BL scenarios, StoNEZD underestimates the frontier in high inefficiency cases. These biases are again similar to the magnitude of the expected inefficiency (20%), and result in MSE values two to three times higher than the low inefficiency cases. Similar to the other estimators, StoNEZD shows no increase in performance with increasing sample sizes. In total, we find cDEA to generally overestimate the frontier, while LC-SFA outperforms the competitors in terms of MSE in nearly all settings. However, StoNEZD is found to perform best in terms of bias in simulations with low inefficiency, while a considerable underestimation of the frontier occurs in high inefficiency settings.

Comparing $HI1$ with $HI2$ and $HI3$ allows us to evaluate the impact of the z-variable distribution. Although the results of $HI2$ and $HI3$ show in general very similar patterns as the results for $HI1$, several points are worth emphasizing. In low inefficiency cases, the average bias for cDEA and LC-SFA increase to 31% and 11%, respectively. Further, for the LC-SFA, biases increase also in high inefficiency cases. As a result, in $HI2$ and $HI3$, LC-SFA overestimates the frontier in nearly all noise-to-signal ratios, although MSE values remain on a low level. StoNEZD, on the contrary, can furthermore improve its performance in these settings and outperforms the other estimators (compare MSE panels 1 and 3 in the Figures 4 to 6). Additionally, StoNEZD' downward biases reduce in cases with high inefficiency and the estimator achieves MSEs of similar magnitude as LC-SFA also in these settings (compare MSE panels 2 and 4 in the Figures 5 and 6). Given the small average bias of LC-SFA compared to StoNEZD in these cases, the similarity of the MSE values of both estimators suggests that LC-SFA estimates the frontier with considerable variation. Therefore, similar to the BL scenarios, StoNEZD seems to perform even better with the non-normally distributed z-variables, while performance of cDEA and LC-SFA deteriorates.

Comparing the HI to the corresponding BL scenarios (i.e., $BL1/HI1$, $BL2/HI2$, and $BL3/HI3$) shows the effect of the environmental intensity on estimators' precision. For cDEA, results show a mixed picture and indicate, on average, a positive but strongly varying impact on BIAS and MSE especially for the truncated exponential z_2 in $BL2$ and $HI2$. On the contrary, on average, no effects on BIAS and MSE from additional z-variable impact is found when comparing $BL1$ with $HI1$ and $BL3$ with $HI3$, although single deviations are of considerable magnitude. For LC-SFA, we observe a similar picture. While the additional impact of the z-variable in $HI1$ leads to very similar BIAS and MSE, performance deteriorates considerably in the scenarios with the non-Gaussian z_2 and z_3. These effects are more strongly pronounced in settings where the environmental impact exceeds the impact of inefficiency ($HI2$ and $HI3$ cases with $\sigma_u = 0.1$). This, again, suggests that LC-SFA is challenged by the distributions of z_2 and z_3. For StoNEZD, the comparison of BL and HI scenarios shows very stable results in all scenarios and settings. Especially for the low inefficiency scenarios, we find nearly constant MSE values for StoNEZD, which indicates that the estimator very well incorporates the additional z-variable impact.

In total, the results of the HI scenarios confirm the estimators patterns of frontier under- and overestimation found in the BL scenarios. Further, the results indicate that cDEA and LC-SFA are challenged by non-normal distributions of the environmental impact. In contrast, StoNEZD well accounts for these effects and competes with LC-SFA in terms of MSE in all cases.

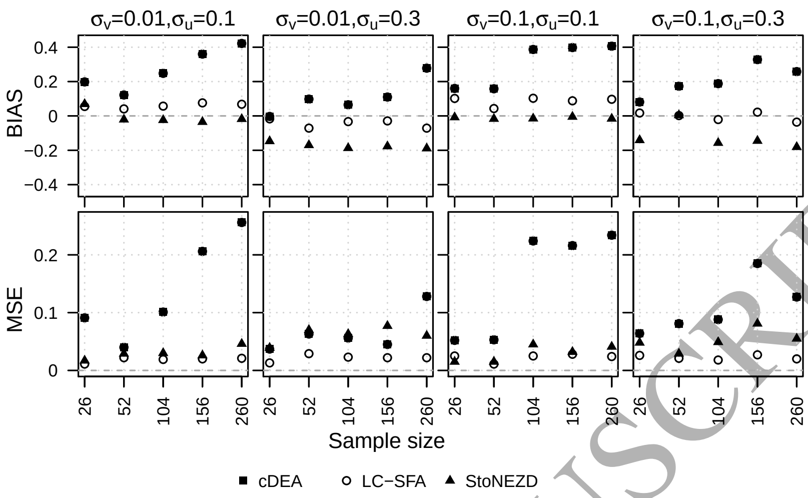

Figure 4: HI1 - Bias (top) and MSE (bottom)

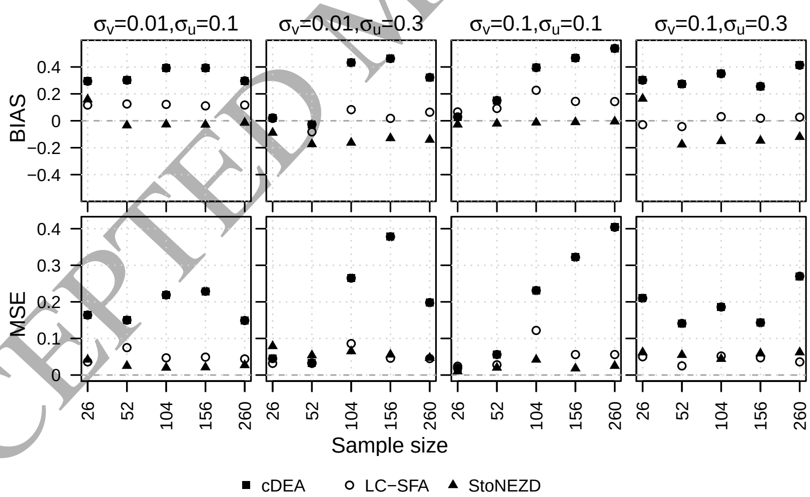

Figure 5: HI2 - Bias (top) and MSE (bottom)

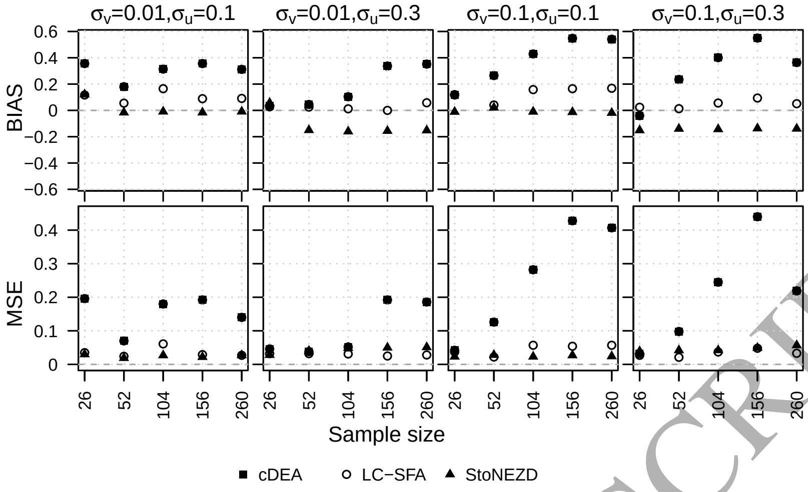

Figure 6: HI3 - Bias (top) and MSE (bottom)

4.2.3. Multiple-z scenarios

Next, we increase the number of z-variables that influence output potentials to two, which are drawn from a normal and a truncated exponential distribution. Figures 7 to 10 show the results for the four multiple-z scenarios $MZ1$ to $MZ4$.[13]

The results of the MZ scenarios show for all estimators similar patterns as the BL and HI scenarios. For $MZ1$, cDEA performs well in cases with high inefficiency, while performance deteriorates with increasing noise and increasing noise-to-signal ratios. Again, an upward bias is always present and is not mitigated in larger samples. LC-SFA achieves low biases in cases with high inefficiency, but an upward bias occurs in low inefficiency settings. LC-SFA MSE, however, is lowest of nearly all noise-to-signal ratios in $MZ1$. StoNEZD underestimates the frontier in high inefficiency settings, but no systematic bias occurs if low inefficiency is present. Often LC-SFA and StoNEZD perform similarly in terms of MSE, while cDEA estimates deviate from the true frontier considerably, especially with high noise-to-signal ratios.

Comparing $MZ1$ with $MZ2$ allows us to analyze again the influence of the inten-

[13]Detailed numerical results are shown in supplementary material.

sity of z-variables. All three estimators cope well with the additional signal and show similar trends as in $MZ1$. Moreover, for cDEA results indicate generally lower MSEs and lower upward biases in $MZ2$, which further decreases with additional inefficiency. For LC-SFA, the additional impact of the environment leads to a slight increase of the bias (with nearly constant MSE), confirming the results of the HI scenarios. Also for StoNEZD results from the HI scenarios are confirmed, and biases and MSEs are nearly identical in both $MZ1$ and $MZ2$.

Next, to analyze the impact of correlation among the z-variables on estimators performance, we compare $MZ1$ with $MZ3$, and $MZ2$ with $MZ4$. For cDEA, results are heterogeneous among the settings: in cases with high inefficiency, biases and MSE most often decrease for the correlated z-variables. On the contrary, in low inefficiency cases, we observe an increase of both performance measures. For LC-SFA and StoNEZD, results indicate generally similar performance in terms of MSE and bias independent of the z-variable correlation.

Finally, we compare the $MZ1$ with the BL and $MZ2$ with the HI scenarios, as they are, to a certain extent, similar. For cDEA, both MZ scenarios shows generally lower median bias and MSE values if inefficiency is low, while results for the high inefficiency scenarios are not conclusive. This, however, suggests that in DEA a second z-variable may allow to better disentangling the environmental effects. For LC-SFA, biases and MSEs increase in the MZ scenarios, which suggests that the additionally included non-normal z-variable has a detrimental effect on estimation accuracy of LC-SFA. These effects are more pronounced with the higher environmental impact in $MZ2$ and $HI2$. For StoNEZD, results indicate a negative impact of the additional z-variable on the median bias. However, overall results are fairly stable and no considerable changes in the MSE is found in the MZ cases.

4.3. Precision of efficiency estimates

The estimators considered in this paper are often also applied to estimate firm-specific efficiency scores. Therefore, we analyze the coherence of true (TE_i) and estimated inefficiency ($\hat{TE_i}$) in terms of rank correlations using Kendall's τ as defined in eq. (13). Figure 11 to 13 show the rank correlation coefficients for the BL, HI, and MZ scenarios. In these figures, a row shows the rank correlation coeffi-

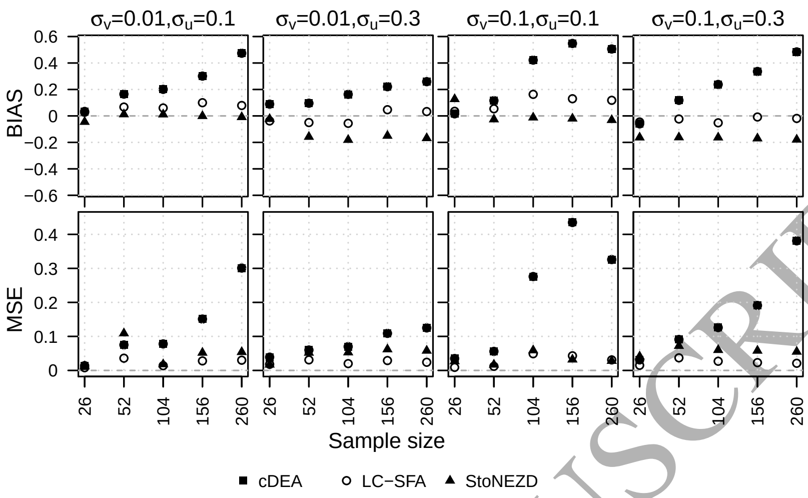

Figure 7: MZ1 - Bias (top) and MSE (bottom)

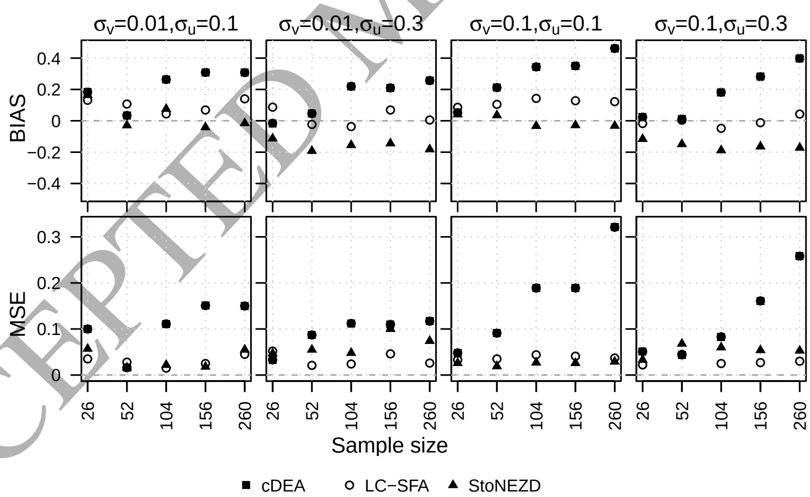

Figure 8: MZ2 - Bias (top) and MSE (bottom)

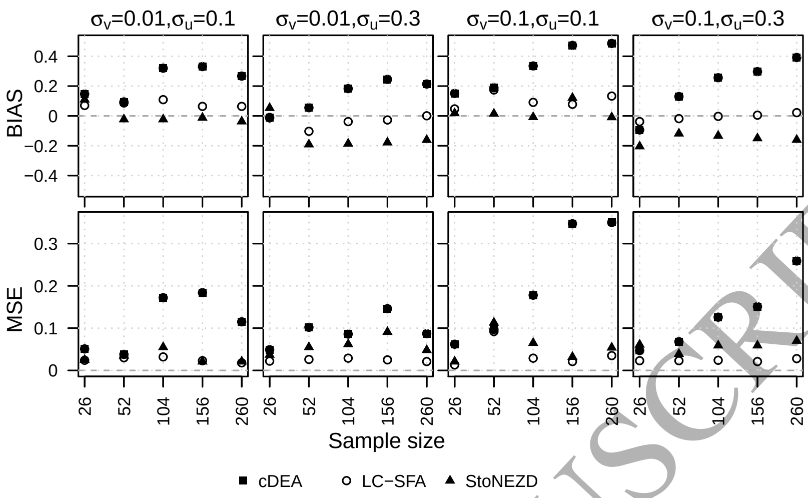

Figure 9: MZ3 - Bias (top) and MSE (bottom)

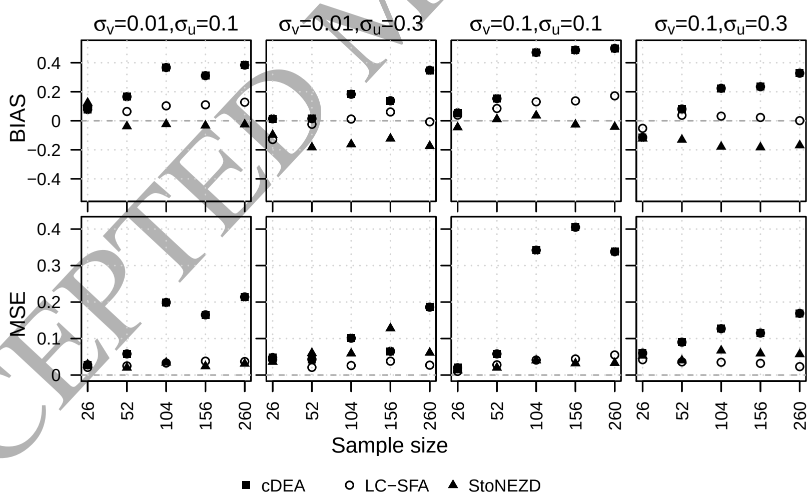

Figure 10: MZ4 - Bias (top) and MSE (bottom)

cients (y-axis) for the scenario indicated on the left, with the parameters for noise and inefficiency indicated on top, and the simulated sample sizes (x-axis).

For the baseline scenarios $BL1$, $BL2$, and $BL3$, rank correlation coefficients range most often between 0.1 and 0.5. Taking the median of all rank correlations per BL scenario reveals that, on average, estimated and true efficiency rank scores are equally correlated ($\tau_{median} = 0.3$). Further, for a given noise-to-signal ratio similar patterns of rank correlation coefficients are present across all the three scenarios (compare Figure 11 by column). Both findings suggest that z-variable distribution is of minor importance for rank correlations if z-variable impact is low. While rank correlations are found to be independent of the sample size, coherence of true efficiency and estimated efficiency varies strongly with the noise and inefficiency parameters. For simulations with high inefficiency ($\sigma_u = 0.3$, Figure 11 columns 2 and 4), rank correlations are of moderate ($0.3 \leq \tau \leq 0.5$) to high ($\tau \geq 0.7$) magnitude. In simulations with low inefficiency, τ decreases. In low-noise-low-inefficiency simulations, rank correlations reach moderate levels in several occasions, while rank correlations are the lowest ($\tau \leq 0.2$) if the noise-to-signal ratio is equal to one (i.e., $\sigma_u = \sigma_v = 0.1$). Regarding the single estimators, we find LC-SFA to outperform the competitors most often in the BL scenarios with noise-to-signal ratio below one, while no clear ranking is available if $\sigma_u = \sigma_v = 0.1$. However, although cDEA and StoNEZD rank behind, differences between the estimators are most often rather small. Additionally, although cDEA delivers few negative rank correlation coefficients, the gap to LC-SFA and StoNEZD in terms of rank correlations is less pronounced than in terms of frontier estimation.

In the HI scenarios (see Figure 12), median rank correlation is 0.26, thus slightly lower than in the BL cases. This suggests that the additional impact of the z-variable has a negative impact on the accuracy of efficiency rankings, but with only small magnitude. As in the BL scenarios, rank correlations vary in most cases between 0.1 and 0.5, but few extreme cases with negative rank correlations are found. However, although these extreme cases appear only with the smallest sample sizes, results do not indicate a clear relationship between sample size and rank correlations. Similar to the BL scenarios, rank correlations depend strongly on the noise and inefficiency parameters, and are highest in cases with high inefficiency ($0.2 \leq \tau \leq 0.7$).

Again, we find a considerable drop for the low inefficiency cases, and rank correlations are the lowest (mostly $0 \leq \tau \leq 0.2$) if the noise-to-signal ratio is equal to one ($\sigma_u = \sigma_v = 0.1$). Results indicate rather small differences between the gaps for the HI scenarios, although these differences vary the z-variable distribution. Overall, cDEA achieves the lowest average rank correlations, but gaps to StoNEZD and LC-SFA are rather small, especially when keeping in mind the rather large gaps in terms of bias and MSE in these settings. LC-SFA copes best with the Gaussian z_1 in $HI1$, where it outperforms the competitors in nearly all cases, but rank correlations in the $HI2$ and $HI3$ scenarios are lower than in the BL scenarios. This again confirms that LC-SFA is challenged by the non-normal z-variable distributions. StoNEZD, on the contrary, performs well with the skewed distributions of z, while performance in $HI1$ deteriorates relative to $BL1$, confirming also the results concerning frontier estimation accuracy.

In the MZ scenarios, we find a median rank correlation coefficient of 0.24, thus again slightly lower than the baseline and the high impact scenarios (compare Figure 13). For the different MZ scenarios, rank correlations are generally in a similar range between 0.1 and 0.4. As before, few negative rank correlations occur in settings with 26 and 52 observations, but again no clear relationship of sample size and rank correlation coefficients exists. As for BL and HI, MZ scenarios' rank correlations are highest in high efficiency cases, and the lowest for the $\sigma_u = \sigma_v = 0.1$ case. A comparison of $MZ1$ with $MZ2$ and $MZ3$ with $MZ4$ - with $MZ2$ and $MZ4$ as the cases in which the environmental impact exceeds the impact of noise and inefficiency - suggests that a strong signal from the environment in low inefficiency cases has detrimental effects on the rank correlations, while this effect is not found if inefficiency is high (compare, e.g., $MZ1$ with $MZ2$ in Figure 13, columns 1 and 3 with 2 and 4). Further, comparing the scenarios with uncorrelated and correlated z-variables, i.e., $MZ1$ with $MZ3$, and $MZ2$ with $MZ4$, indicates that correlation of environmental factors has no clear effect on estimators performance in terms of efficiency scores. For the single estimators, no clear patterns are available, although LC-SFA and StoNEZD most often outperform cDEA.

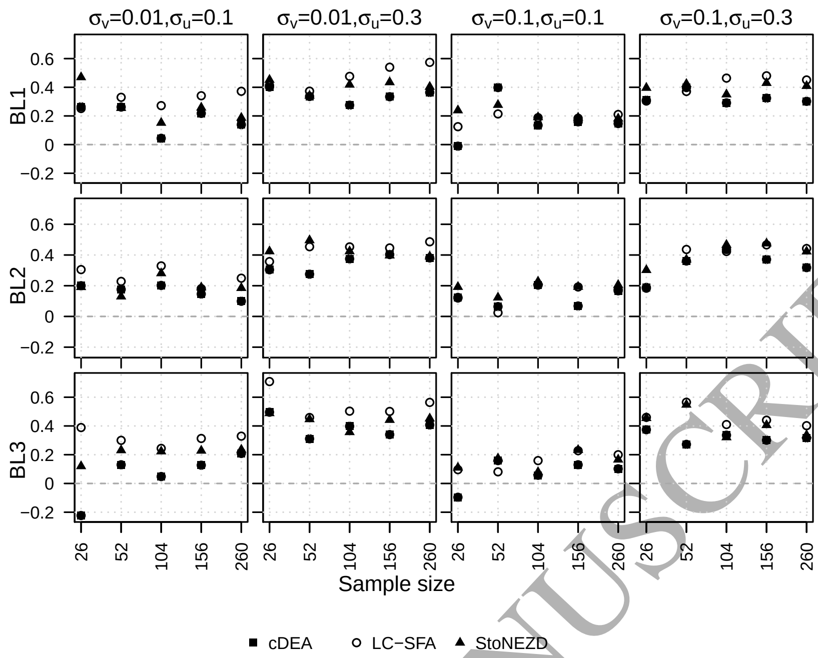

Figure 11: BL - Efficiency score rank correlations

4.4. Summary and discussion

Overall, we find that frontier and efficiency estimation accuracy strongly depends on the noise-to-signal ratio. Additionally, all estimators approximate the frontier accurately with small to medium sample sizes, but accuracy does not improve with increasing sample sizes. In summary, our results show that LC-SFA outperforms the competitors due its low MSE in nearly all scenarios. However, StoNEZD can compete and outperform LC-SFA in several settings, while results for cDEA indicate a clear gap in terms of frontier estimation accuracy. The results, however, reveal diverse and estimator-specific strengths and weaknesses, which are in line with the literature.

More specifically, cDEA generally overestimates the frontier and cannot compete with neither LC-SFA nor StoNEZD. Combining our findings with Andor and Hesse (2014) yields the conclusion that the nonparametric alternative is outperformed by its semi- and fully parametric competitors in noisy settings irrespective of the presence of operational conditions. It is, however, worth emphasizing that cDEA performs

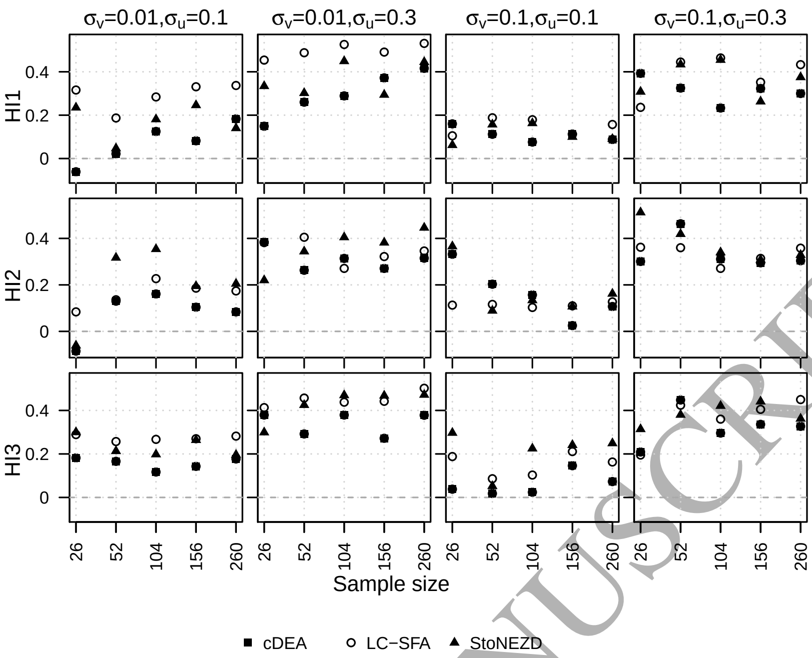

Figure 12: HI - Efficiency score rank correlations

comparably well in a substantial number of cases with very small samples, while performance deteriorates for larger sample sizes. Although the latter conflicts with Cordero et al. (2016), this divergence seems reasonable given that the authors model environmental factors to affect the inefficiency distribution, but not the frontier as in our case. The good performance of cDEA for small sample sizes is good news particularly for regulatory benchmarking which in most cases is subject to a limited number of observations due to the market structure of network-based industries. cDEA performs better if the signal from inefficiency is stronger than that from noise and the performance of cDEA is rather stable across different noise levels in scenarios with high inefficiency. In contrast, in scenarios with low inefficiency levels, cDEA estimates the frontier less accurately if noise is increased. Therefore, we partially confirm Andor and Hesse's findings and conclude that less noise is always beneficial for cDEA. Further, the intensity with which the environmental variable impacts production possibilities is relevant to a moderate extent: the smaller the impact, the more accurately cDEA approximates the frontier. This finding points in the same

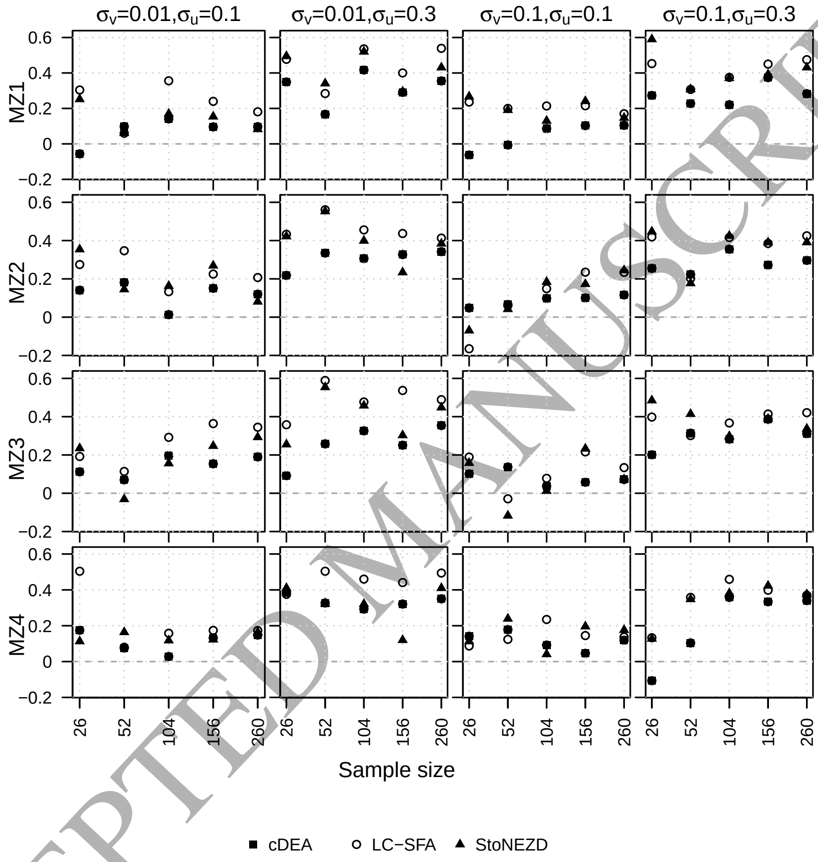

Figure 13: MZ - Efficiency score rank correlations

direction as Yu (1998), who finds that non-parametric estimators are an alternative to parametric approaches particularly in such scenarios. Neither the distribution nor the number of the environmental variables significantly affect how accurately cDEA estimates the frontier. This is an interesting result for real-life applications, where operational conditions are captured in various dimensions. Especially in regulatory applications where nonparametric estimators might be preferred, it is recommendable to focus on reducing noise in the data rather than, for example, limiting the number of external factors for the sake of a low dimensionality. With respect to rank correlation, cDEA's performance is, with few exceptions, inferior to both LC-SFA and StoNEZD. We find smaller rank correlations of true and estimated efficiency scores than Cordero et al. (2016), which is likely due to the different simulation designs. Further, rank correlations vary especially in high impact and multiple-z scenarios, and, in line with Badunenko et al. (2012), we find low rank correlations especially if the noise-signal ratio equals one.

LC-SFA performs generally well in terms of frontier estimation accuracy and outperforms StoNEZD and cDEA in most cases. Thus, in line with results by Andor and Hesse (2014), the parametric approach performs often better than the non- and semi-parametric counterparts. Overall, the results indicate that the methodology offers characteristics suitable for application in a regulatory context. However, several factors impede estimation accuracy. Our results show a bias of LC-SFA driven by the inefficiency component, with an overestimation of the frontier occurring if the data contains only little inefficiency. On the contrary, an underestimation takes place in high inefficiency settings, similar to the results for standard SFA provided by Kuosmanen et al. (2013). With respect to the noise component, we find, similarly to the DEA results, deteriorating estimation performance, but overall losses are of small magnitude.

With respect to the environmental factor, estimation performance depends on z-variable distribution. LC-SFA performs best with normally distributed z-variables, while estimation performance decreases for the skewed distributions considered. The latter effect becomes sizable if the impact of the environmental variable is stronger than the inefficiency component, and is present as well if Gaussian and non-Gaussian variables are simultaneously considered. Practitioners should therefore treat non-normal z-variables with care. However, as argued by Llorca et al. (2014), the LC-SFA is well able to control for heterogeneity if the environment is not explicitly modeled, i.e., if z-variables are omitted. Results in terms of rank correlations confirm these findings. LC-SFA performs overall best, especially if the simulated inefficiency in the model is high, and if the considered environmental factor has a normal distribution or overall low impact. However, as for the other estimators, and in line with Badunenko et al. (2012), rank correlations are low if the noise-signal ratio equals one.

For StoNEZD, results are mixed. Overall, the estimator performs well in terms of MSE of the frontier estimate, but the bias strongly depends on the simulated noise-to-signal ratio. Nonetheless, StoNEZD generally outperforms cDEA and can compete with LC-SFA in many cases, which is in line with Andor and Hesse (2014) for the equivalent models without environmental factors. Unlike Kuosmanen et al. (2013), we do not find consistency of StoNEZD, i.e., estimation accuracy does not improve for larger sample sizes. StoNEZD performs very well in settings with low inefficiency, and the frontier is most often estimated with small bias and MSE. In settings with high inefficiency, however, the estimator underestimates the frontier with a bias of similar magnitude as the expected inefficiency. On the contrary, StoNEZD results are stable with respect to the simulated noise. Andor and Hesse (2014) derive similar results for the StoNED, and argue that the resulting overestimation of efficiency can be useful from the perspective of practitioners, e.g., in a regulatory context. The results show that StoNEZD does not lead to unreasonable improvement potentials for the regulated firms if they are inefficient, while improvement potentials are very accurately estimated if they are efficient. StoNEZD performs better in simulations with non-Gaussian distributions of environmental variables than with the normal z-variable, independent of the intensity and the number of z-variables. This points toward an advantage for StoNEZD if many environmental factors with various distributions are applied, as it is often the case in applied work. Interestingly, our results therefore endorse the Finnish regulator which uses StoNEZD for such a setting in the electricity distribution sector (see Kuosmanen, 2012).

Regarding efficiency score rank correlations, StoNEZD performs overall well and often competes with LC-SFA. Our findings thereby contradict Andor and Hesse (2014), who find the lowest rank correlations for the semi-parametric approach. As for the other two estimators, efficiency rank correlations are generally higher if considerable inefficiency is simulated. Again, the estimator shows its strength with non-normal environmental factors, independent of the number and the intensity of environmental factors.

5. Conclusion

Based on Monte Carlo simulations for a wide range of scenarios, this paper compares the non-parametric conditional DEA, the parametric latent class SFA, and the semi-parametric StoNEZD with a focus on their ability to account for environmental factors when estimating production functions and efficiencies. Our results provide insights in the finite sample properties of the three estimators, and we identify estimator-specific strengths and weaknesses. Overall, our results suggest that the latent class approach performs best, although StoNEZD can compete in many cases. On the contrary, cDEA often fails to correctly account for the environment, leading to imprecise frontier estimates.

However, there are two caveats. First, like all simulation studies, our data generating process covers only a subset of potential parameter settings and model specifications. And second, our findings are based on correctly specified models. Therefore, additional research is necessary to further understand the estimators' finite sample properties. We suggest that future studies may consider more complex settings with additional environmental variables, correlations between inputs and environment, technological heterogeneity induced by environmental factors, model misspecification, and omitted variables.

Acknowledgements

We would like to thank three anonymous referees for their valuable comments and suggestions. Further, we thank Oleg Badunenko, Marc Andor, Astrid Cullmann, and the participants of the North American Productivity Workshop 2016 and INFRATrain 2016 for fruitful discussions and support. We also owe our gratitude to

Ann Stewart and Adam Mark Lederer for excellent editing. All remaining mistakes and errors are ours.

References

Afriat, S. N., 1967. The construction of utility functions from expenditure data. International Economic Review 8 (1), pp. 67–77.

Aggelopoulos, E., Georgopoulos, A., 2017. Bank branch efficiency under environmental change: A bootstrap DEA on monthly profit and loss accounting statements of Greek retail branches. European Journal of Operational Research 261 (3), 1170–1188.

Agrell, P. J., Bogetoft, P., Koller, M., Trinkner, U., 2013a. Effizienzvergleich für Verteilernetzbetreiber Strom 2013 - Ergebnisdokumentation und Schlussbericht. Tech. rep., Swiss Economics and Sumiscid.

Agrell, P. J., Bogetoft, P., et al., 2013b. Benchmarking and regulation. CORE DISCUSSION PAPER 2013/8.

Agrell, P. J., Brea-Solís, H., 2017. Capturing heterogeneity in electricity distribution operations: A critical review of latent class modelling. Energy Policy 104, 361–372.

Aigner, D., Lovell, C. A. K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6 (1), 21–37.

Anaya, K. L., Pollitt, M. G., 2017. Using SFA to measure the impact of weather on the efficiency of the distribution businesses in developing countries. European Journal of Operational Research (forthcoming).

Ancarani, A., Mauro, C. D., Giammanco, M., 2009. The impact of managerial and organizational aspects on hospital wards' efficiency: Evidence from a case study. European Journal of Operational Research 194 (1), 280–293.

Andor, M., Hesse, F., 2014. The StoNED age: the departure into a new era of efficiency analysis? A monte carlo comparison of StoNED and the oldies (SFA and DEA). Journal of Productivity Analysis 41 (1), 85–109.

Badunenko, O., Henderson, D. J., Kumbhakar, S. C., 2012. When, where and how to perform efficiency estimation. Journal of the Royal Statistical Society: Series A (Statistics in Society) 175 (4), 863–892.

Banker, R. D., Morey, R. C., 1986. Efficiency analysis for exogenously fixed inputs and outputs. Operations Research 34 (4), 513–521.

Battese, G. E., Coelli, T. J., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20, 325–332.

Bjørndal, E., Bjørndal, M. H., Fange, K.-A., 2010. Benchmarking in regulation of electricity networks in Norway: An overview. In: Bjørndal, E., Bjørndal, M. H., Pardalos, P., Rönnqvist, M. (Eds.), Energy, Natural Resources and Environmental Economics. Springer, pp. 317–342.

Bădin, L., Daraio, C., Simar, L., 2010. Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. European Journal of Operational Research 201 (2), 633–640.

Bădin, L., Daraio, C., Simar, L., 2012. How to measure the impact of environmental factors in a nonparametric production model. European Journal of Operational Reseach (223), 818–833.

Caudill, S. B., 2003. Estimating a mixture of stochastic frontier regression models via the EM algorithm: A multiproduct cost function application. Empirical Economics 28 (3), 581–598.

Cazals, C., Florens, J.-P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. Journal of Econometrics 106 (1), 1–25.

Coelli, T., Estache, A., Perelman, S., Trujillo, L., 2003. A primer on efficiency measurement for utilities and transport regulators. World bank institute development studies, The World Bank, Washington D.C.

Cordero, J. M., Pedraja, F., Santín, D., 2009. Alternative approaches to include exogenous variables in DEA measures: A comparison using monte carlo. Computers & Operations Research 36 (10), 2699–2706.

Cordero, J. M., Polo, C., Santin, D., Sicilia, G., 2016. Monte-carlo comparison of conditional nonparametric mathods and traditional appraoches to include exogeneous variables. Pacific Economic Review 21 (4), 483–497.

Daraio, C., Simar, L., 2005. Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach. Journal of Productivity Analysis 24 (1), 93–121.

Daraio, C., Simar, L., 2007a. Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications. Studies in productivity and efficiency. Springer, New York.

Daraio, C., Simar, L., 2007b. Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. Journal of Productivity Analysis 28 (1), 13–32.

De Witte, K., Geys, B., 2013. Citizen coproduction and efficient public good provision: Theory and evidence from local public libraries. European Journal of Operational Research 224 (3), 592–602.

Debreu, G., 1951. The coefficient of resource utilization. Econometrica 19 (3), 273–292.

EMA, 2013. Appendix 1 – regulation methods for the assessment of reasonableness in pricing of electricity distribution network operations and high-voltage distribution network operations in the third regulatory period starting on 1 january 2012 and ending on 31 december 2015. Tech. rep., Energy Market Authority.

Estache, A., Perelman, S., Trujillo, L., 2006. Infrastructure Reform in Developing Economies: Evidence from a survey of efficiency measures. Edward Elgar Publishing,Cheltenham, pp. 133–174.

Fan, Y., Li, Q., Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. Journal of Business & Economic Statistics 14, 460 – 468.

Farrell, M., 1957. The measurement of productive efficiency. Journal of the Royal Statistical Society 120 (3), 253–281.

Fried, H., Lovell, C. A. K., Schmidt, S., Yaisawarng, S., 2002. Accounting for environmental effects and statistical noise in data envelopment analysis. Journal of Productivity Analysis 17 (1-2), 157–174.

Greene, W. H., 2002. Alternative panel data estimators for stochastic frontier models, working Paper, Stern School of Business, New York University.

Greene, W. H., 2005. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126 (2), 269–303.

Haelermans, C., De Witte, K., 2012. The role of innovations in secondary school performance – evidence from a conditional efficiency model. European Journal of Operational Research 223 (2), 541–549.

Hall, P., Racine, J., Li, Q., 2004. Cross-validation and the estimation of conditional probability densities. Journal of the American Statistical Association 99 (468), 1015–1026.

Haney, A. B., Pollitt, M. G., 2009. Efficiency analysis of energy networks: An international survey of regulators. Energy Policy 37 (12), 5814–5830.

Hildreth, C., 1954. Point estimates of ordinates of concave functions. Journal of the American Statistical Association 49 (267), 598–619.

Jeong, S.-O., Park, B., Simar, L., 2010. Nonparametric conditional efficiency measures: asymptotic properties. Annals of Operations Research 173 (1), 105–122.

Johnson, A., Kuosmanen, T., 2011. One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent StoNEZD method. Journal of Productivity Analysis 36 (2), 219–230.

Jondrow, J., Lovell, C. A. K., Materov, I. S., Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19 (2-3), 233–238.

Kneip, A., Park, B. U., Simar, L., 1998. A note on the convergence of nonparametric DEA estimators for production efficiency scores. Econometric Theory 14 (6), 783–793.

Kneip, A., Simar, L., Wilson, P. W., 12 2008. Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. Econometric Theory 24, 1663–1697.

Koopmans, T. C., 1951. An analysis of production as an efficient combination of activities. In: Koopmans, T. C. (Ed.), Activity Analysis of Production and Allocation. Wiley, New York.

Krüger, J. J., 2012. A monte carlo study of old and new frontier methods for efficiency measurement. European Journal of Operational Research 222, 137 – 148.

Kumbhakar, S., Hjalmarsson, L., 1993. Technical efficiency and technical progress in Swedish dairy farms. Oxford University Press, pp. 256–270.

Kumbhakar, S., Sun, K., 2012. Estimation of TFP growth: a semiparametric smooth coefficient approach. Empirical Economics (43), 1–24.

Kuosmanen, T., 2012. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. Energy Economics 34 (6), 2189 – 2199.

Kuosmanen, T., Johnson, A., Saastamoinen, A., 2015. Stochastic nonparametric approach to efficiency analysis: a unified framework. In: Zhou, J. (Ed.), Data

Envelopment Analysis: A Handbook of Models and Methods. Springer, pp. 191–244.

Kuosmanen, T., Kortelainen, M., 2012. Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. Journal of Productivity Analysis 38 (1), 11–28.

Kuosmanen, T., Saastamoinen, A., Sipiläinen, T., 2013. What is the best practice for benchmark regulation of electricity distribution? Comparison of DEA, SFA and StoNED methods. Energy Policy 61, 740 – 750.

Laffont, J.-J., Tirole, J., 1993. A theory of incentives in procurement and regulation. MIT press.

Li, Q., Racine, J. S., 2007. Nonparametric econometrics: Theory and practice. Princeton University Press, Princeton.

Li, Q., Racine, J. S., 2008. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. Journal of Business and Economic Statistics 26 (4), 423–434.

Llorca, M., Orea, L., Pollitt, M. G., 2014. Using the latent class approach to cluster firms in benchmarking: An application to the US electricity transmission industry. Operations Research Perspectives 1 (1), 6 – 17.

Orea, L., Kumbhakar, S. C., 2004. Efficiency measurement using a latent class stochastic frontier model. Empirical Economics 29 (1), 169–183.

Simar, L., Vanhems, A., Van Keilegom, I., 2016. Unobserved hetergeneity and endogeneity in nonparametric frontier estimation. Journal of Econometrics (190), 360–373.

Simar, L., Wilson, P. W., 2007a. Estimation and inference in two-stage, semiparametric models of production processes. Journal of Econometrics 136 (1), 31–64.

Simar, L., Wilson, P. W., 2007b. Statistical inference in nonparametric frontier models: Recent developments and perspectives. In: Fried, H., Lovell, C. A. K., Schmidt, S. (Eds.), The Measurement of Productive Efficiency. Oxford University Press, Oxford.

Stone, C. J., 11 1980. Optimal rates of convergence for nonparametric estimators. The Annals of Statistics 8 (6), 1348–1360.

Stone, C. J., 12 1982. Optimal global rates of convergence for nonparametric regression. The Annals of Statistics 10 (4), 1040–1053.

Yu, C., 1998. The effects of exogenous variables in efficiency measurement - a monte carlo study. European Journal of Operational Research 105 (3), 569 – 580.