RUHR
ECONOMIC PAPERS

Mark Andor
Frederik Hesse

The StoNED age: The Departure Into a New Era of Efficiency Analysis?

A Monte Carlo Comparison of StoNED and the "Oldies" (SFA and DEA)

rwi #394

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #394

Responsible Editor: Christoph M. Schmidt

ISSN 1864-4872 (online) – ISBN 978-3-86788-449-5

Ruhr Economic Papers #394

Mark Andor and Frederik Hesse

The StoNED age: The Departure Into a New Era of Efficiency Analysis?

A Monte Carlo Comparison of StoNED and the “Oldies” (SFA and DEA)

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788449
ISSN 1864-4872 (online)
ISBN 978-3-86788-449-5

Mark Andor and Frederik Hesse[1]

The StoNED age: The Departure Into a New Era of Efficiency Analysis? – A Monte Carlo Comparison of StoNED and the "Oldies" (SFA and DEA)

Abstract

Based on the seminal paper of Farrell (1957), researchers have developed several methods for measuring efficiency. Nowadays, the most prominent representatives are nonparametric data envelopment analysis (DEA) and parametric stochastic frontier analysis (SFA), both introduced in the late 1970s. Researchers have been attempting to develop a method which combines the virtues – both nonparametric and stochastic – of these "oldies". The recently introduced Stochastic non-smooth envelopment of data (StoNED) by Kuosmanen and Kortelainen (2010) is such a promising method. This paper compares the StoNED method with the two "oldies" DEA and SFA and extends the initial Monte Carlo simulation of Kuosmanen and Kortelainen (2010) in several directions. We show, among others, that, in scenarios without noise, the rivalry is still between the "oldies", while in noisy scenarios, the nonparametric StoNED PL now constitutes a promising alternative to the SFA ML.

JEL Classification: C1, C5, D2, L5, Q4

Keywords: efficiency; stochastic non-smooth envelopment of data (StoNED); data envelopment analysis (DEA); stochastic frontier analysis (SFA); Monte Carlo simulation

January 2013

1 Mark Andor, RWI; Frederik Hesse, Finance Center Münster, University of Münster. *– We are deeply indebted to the participants of the 8th Asia-Pacic Productivity Conference (APPC) in Bangkok, Thailand, the 4th Workshop on Efficiency and Productivity Analysis (HAWEPA) in Halle, Germany, the 12th European Workshop on Efficiency and Productivity Analysis (EWEPA) in Verona, Italy, and the 11th IAEE European Conference in Vilnius, Lithuania, for providing valuable comments that have led to a considerable improvement of earlier versions of this paper. Furthermore, we would like to thank Brian Bloch, Finn Forsund, William Greene, Arne Henningsen, Uwe Jensen, Chonjoo Lee, Colin Vance, the editors and two anonymous referees for their helpful comments and suggestions. The authors are responsible for all errors and omissions.* *– All correspondence to Mark Andor, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany, E-Mail: mark.andor@rwi-essen.de.*

1 Introduction

In his classic paper, Farrell (1957) stated that measuring the efficiency of productivity is important to economic theorists and economic policy makers alike. Based on Farrell's work, researchers have developed several methods for measuring efficiency. Despite this progress, after more than five decades of efficiency analysis research, there is still no single superior method (see, among others, Resti (2000), Mortimer (2002) and Badunenko et al (2011)).

The efficiency analysis literature can be divided into two main branches of parametric and nonparametric methods. Data envelopment analysis (DEA) is the most popular representative of the nonparametric methods. It is a linear programming method which constructs a nonparametric envelopment frontier over the data points. Despite the fact that previous papers also proposed mathematical programming methods (see, for example, Afriat (1972)), DEA is generally attributed to Charnes et al (1978). DEA estimates efficiency without considering statistical noise and is thus a deterministic method. This is its main disadvantage. On the other hand, its main advantage is flexibility, due to its nonparametric nature.

In contrast, parametric methods require an assumption about the functional form of the production function. The corrected ordinary least squares method (COLS), originally proposed by Winsten (1957), estimates the efficient frontier by shifting the ordinary least squares regression towards the most efficient producer. Subsequently, it measures inefficiency as the distance to this frontier. COLS has the same disadvantage as DEA, since it is also deterministic. Aigner et al (1977) and Meeusen and van den Broeck (1977) developed a stochastic parametric model, called stochastic frontier analysis (SFA). Its main advantage is its ability to measure efficiency while simultaneously considering the presence of statistical noise.

The methodological differences and corresponding strengths and weaknesses lead to DEA and SFA being the two most popular economic approaches for measuring efficiency. However, in real-world applications, the problem arises that it is unknown which set of assumptions is closer to reality and the methods yield different efficiency scores. Hence, both in the literature as well as in practical application, it is desirable to find a way to combine the advantages of the two methods. Among others, Banker et al (1994) state that the "...use of more than one methodology can help to avoid the possible occurrence of 'methodological bias'...". In practical application, one common approach is to combine SFA and DEA by using, for example, the mean value of the estimates yielded by the two methods. For instance, Haney and Pollitt (2009) conclude that the combination approach is "best-practice" in energy regulation. Therefore, in Andor and Hesse (2011), we analyzed SFA and DEA, and applied combination approaches within a MC simulation, in order to evaluate the performance. Under our assumptions, the results confirm weakly that the mean performs better than the elementary results of DEA and SFA. Nevertheless, this approach is ad-hoc and lacks a theoretical foundation, raising the question of whether any theoretical method effectively combines the virtues of DEA and SFA.

In the efficiency analysis literature, there are ongoing attempts to develop this kind of method (cf., among others, Fan et al (1996), Kneip and Simar (1996), Kumbhakar et al (2007)). The Stochastic non-smooth envelopment of data (StoNED) method, recently introduced by Kuosmanen and Kortelainen (2010), is a promising candidate, as it is stochastic and semi-parametric, requiring no a priori explicit assumption about the functional form of the production function.

The aim of this present article is to evaluate the performance of StoNED in comparison to the "oldies" – DEA and SFA – within a Monte Carlo Simulation (MC). MC studies are widely used to evaluate efficiency estimation methods (see, for example, Gong and Sickles (1992), Banker et al (1993) and Resti (2000)). They enable researchers to reveal factors influencing the performance of the various methods and succeed in indicating a range of specific situations, in which a particular estimation method proves superior. An MC study considering StoNED can be found in the originating paper Kuosmanen and Kortelainen (2010). Our simulation study extends this initial one in three directions. Firstly, Kuosmanen and Kortelainen (2010) state that one of the most promising avenues for future research is to conduct further MC simulations under a wider range of conditions. We respond to this call by analyzing the influence of sample size, the production function (number of inputs, correlation between inputs, functional form, economies of scale and elasticity of substitution) and the error terms (distribution of the inefficiency term, ratio of inefficiency and noise, and heteroscedasticity of the inefficiency term). Secondly, Kuosmanen and Kortelainen (2010) restrict their study to the "simpler" method of moments estimator (MoM). Nevertheless, among others, Olson et al (1980) and Coelli (1995) demonstrate in MC experiments that the choice of estimation technique impacts on the performance of the method. Hence, in this paper, we also consider the maximum likelihood estimator (ML) and the pseudolikelihood estimator (PL) for SFA and StoNED, respectively. Thirdly, we particularly measure performance in terms of technical efficiency instead of the estimated production frontier and, thus, provide a complementary view to Kuosmanen and Kortelainen (2010). From our point of view, most of the practical applications of efficiency estimation methods as well as the empirical studies focus on technical efficiency. Nevertheless, to give a comprehensive comparison, we also examine the performance of the methods in terms of the estimated production frontier for a couple of settings. In total, we analyze the performance of the following five methods DEA, SFA MoM, SFA ML, StoNED MoM and StoNED PL within 200 different settings.

The remainder of this paper is organized as follows. In Section 2, we explain the methods used in this study, DEA, SFA and StoNED, and the estimation techniques MoM, ML and PL. Section 3 describes the general simulation design of the Monte Carlo experiment. In Section 4, we first show the aggregated results and highlights the strengths and weaknesses of the methods. Afterwards we present the detailed results and discuss the various influence factors. Finally in Section 5, we summarize the most important findings and provide some directions for further research.

2 Methods

In this section, we describe the efficiency estimation methods used in this study. Before describing the methods in detail, we first give an overview of the main differences and the general procedure. We assume that there is cross-sectional data of n decision making units (DMU), for example, firms or universities. Each $DMU_j (j = 1, ..., n)$ produces a single output q_j using

m inputs $z_{i,j}(i = 1, ..., m)$. The relationship between the inputs and the output, i.e. the deterministic production frontier, is expressed by $F(z_{i,j})$. The observable, factual output q_j can deviate from the optimal output, determined via $F(z_{i,j})$, by a factor ε_j:

$$q_j = \underbrace{F(z_{i,j})}_{Production\ Frontier} \cdot exp^{(\varepsilon_j)} \qquad j = 1, ..., n. \tag{1}$$

The efficiency estimation methods can be categorized into parametric vs. nonparametric, as well as deterministic vs. stochastic. The first component of efficiency estimation methods is to estimate the underlying production function. While the parametric SFA requires an assumption about the functional form of the production function, DEA is nonparametric and the construction of the frontier is only restricted via its axiomatic foundation.[1] This is the main disadvantage of SFA compared to DEA. The semi-parametric StoNED avoids this shortcoming by using convex nonparametric least squares (CNLS). CNLS does not need an assumption of a particular functional form, but chooses a function from the family of continuous, monotonically increasing, concave functions that can be non-differentiable (cf. Kuosmanen and Kortelainen (2010)). Therefore, these assumptions are comparable with those of DEA, but are less restrictive than those of SFA.

The second important difference between the efficiency estimation methods is the assumption about the composition of the factor ε_j. While the deterministic DEA assumes that the entire deviation ε_j attributes to inefficiency, stochastic methods – SFA and StoNED – estimate technical efficiency, while admitting that there could be random noise v_j in the data, for example, due to variation in weather conditions, measurement errors or just coincidence. Adding this stochastic term to equation (1) leads to:

$$q_j = \underbrace{F(z_{i,j})}_{Production\ Function} \cdot \underbrace{exp^{(\varepsilon_j)}}_{Composed\ error\ term} \qquad \text{with } \varepsilon_j = v_j - u_j \text{ and } j = 1, ..., n, \tag{2}$$

where the composed error term (ε_j) is the combination of inefficiency u_j and the noise term v_j. The challenge for stochastic models is the decomposition of the composed error term into a noise term and an inefficiency term. For this purpose, the skewness of the distribution of the error term ε_j is crucial. In general parlance: "Luck", expressed by the noise term v_j, can contribute positively or negatively and we expect by definition that, on average, it is balanced. Hence, it is plausible to assume a symmetric distribution with a zero mean. In contrast, inefficiency u_j only affects in one direction and therefore, its distribution is skewed. In the case of a production function, inefficiency can only impact negatively. Due to the fact that the distribution of the composed error term ε_j is the combination of these two distributions, it indicates the presence of inefficiency. The likelihood of inefficiency increases with the skewness of the distribution of ε_j. Using distributional assumptions for the noise term and the inefficiency term, SFA and StoNED estimate the error term ε_j as well as the ratio of noise and inefficiency, by means of the method of moments, maximum likelihood or pseudo-maximum likelihood technique.

[1] Axioms: Convexity, Inefficiency ("Free Disposability"), Ray Unboundedness ("Returns to Scale") and Minimum Extrapolation, see Banker et al (1984).

The second step is the determination of technical efficiency for each DMU. Independent of the stochastic method, using the estimates of step one – the error term ε_j and the ratio of noise and inefficiency – individual efficiency can be estimated. The deterministic DEA does not consider random noise and thus the technical efficiency is the entire deviation from the estimated frontier.

Figure 1 summarizes the main differences between the methods. In this respect, the recently introduced StoNED is arranged in the middle of the two oldies DEA and SFA, as it combines the flexibility of DEA with the stochastic nature of SFA, in a unified framework of frontier estimation.

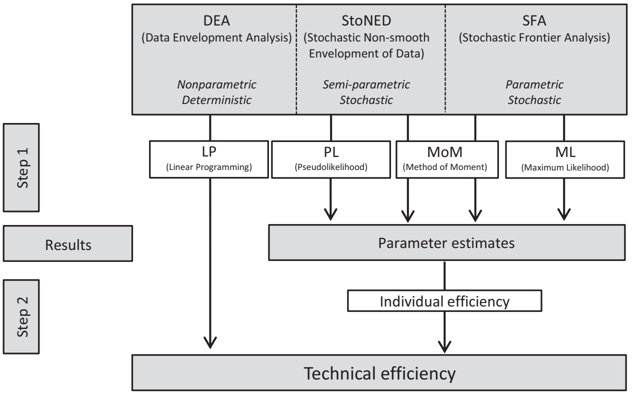

Figure 1: **Overview of the methods procedure.**

2.1 Data Envelopment Analysis (DEA)

DEA is generally attributed to Charnes et al (1978) who introduced the term *data envelopment analysis*. Their original model, also known as the CCR model, assumes constant returns to scale (CRS) and is input orientated. Nowadays, there is a wide range of different models which consider alternative sets of assumptions. An overview can be found, for example, in Cook and Seiford (2009).

In our study, we use the standard BBC model (Banker et al (1984)) which allows for variable returns to scale (VRS). In the multiple-input multiple-output context, each DMU_j produces s outputs $q_{r,j}(r = 1, ..., s)$ using m inputs $z_{i,j}(i = 1, ..., m)$. In order to determine the individual efficiency of the k-th DMU, the following output-oriented two-stage BBC model (cf. Banker et al (2004)) must be maximized

$$
\begin{aligned}
&\text{maximize}_{\phi,\lambda} && \phi_k - \theta\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) && && (3)\\
&\text{subject to} && z_{i,k} = \sum_{j=1}^{n} \lambda_j z_{i,j} + s_i^-, && i = 1, ..., m,\\
& && \phi_k q_{r,k} = \sum_{j=1}^{n} \lambda_j q_{r,j} - s_r^+, && r = 1, ..., s,\\
& && \sum_{j=1}^{n} \lambda_j = 1,\\
& && \lambda_j, s_i^-, s_r^+ \geq 0 && \forall \quad i, r, j,
\end{aligned}
$$

where θ is an infinitesimal non-Archimedean constant, λ_j are the weightings, ϕ_k is a scalar and $1 \leq \phi_k \leq \infty$.[2] The output and input slacks are s_r^+ and s_i^-, respectively. In order to obtain efficiency values for all DMUs, the linear programming model must be solved for each DMU, i.e. n times. The estimated technical efficiency (TE) is defined by

$$\hat{TE}_j = 1/\phi_k \qquad \text{with } 0 \leq \hat{TE} \leq 1. \tag{4}$$

A value of one indicates a point on the efficient frontier and thus a fully efficient DMU, according to Farrell (1957). Until now, only a StoNED model exists for the multiple-input single-output case (see Kuosmanen and Kortelainen (2010)). Hence, in order to compare the methods, we restrict our analysis to the simpler multiple-input single-output case, i.e. $s = 1$.

2.2 Stochastic Frontier Analysis (SFA)

2.2.1 SFA maximum likelihood (SFA ML)

Aigner et al (1977) and Meeusen and van den Broeck (1977) simultaneously developed a stochastic parametric model, the stochastic frontier analysis (SFA). A comprehensive treatment of SFA can be found in Kumbhakar and Lovell (2003). SFA is a parametric method and requires an assumption regarding the functional form of the production function. Assuming a log-linear Cobb-Douglas form, we can rewrite equation (2) as

$$y_j = \beta_0 + \sum_{i=1}^{m} \beta_i \cdot x_{i,j} + \varepsilon_j \qquad \text{with } \varepsilon_j = v_j - u_j \text{ and } j = 1, ..., n. \tag{5}$$

[2]This envelopment formulation is usually the preferred form, because it has fewer constraints than the multiplier form (see Coelli et al (2005)).

Note that $y_j = ln(q_j)$ and $x_j = ln(z_j)$. The noise term v_j and the inefficiency term u_j are assumed to be statistically independent of each other, as well as of the inputs x_j. The latter assumption implies that inefficiency and random noise are homoscedastic, i.e. independent of the scale of the DMU.

Throughout this paper, we use the standard normal-half normal model. That is, we assume a normally distributed noise term $v_j \sim N(0, \sigma_v^2)$ and a half normally distributed inefficiency term $u_j \sim |N(0, \sigma_u^2)|$.[3] Under these assumptions, the marginal density function of the composed error term is defined by:

$$f(\varepsilon) = \frac{2}{\sigma} \cdot \phi\left(\frac{\varepsilon}{\sigma}\right) \cdot \Phi\left(-\frac{\varepsilon\lambda}{\sigma}\right) \tag{6}$$

where $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$, $\lambda = \frac{\sigma_u}{\sigma_v}$, and ϕ and Φ are the standard normal cumulative distribution and the density function, respectively. The ratio of inefficiency and noise is represented by λ. If $\lambda \to 0$, the composed error term is dominated by the noise term. In contrast, if $\lambda \to \infty$, the inefficiency term dominates the composed error term.

Maximum likelihood estimation is an appropriate technique for estimating σ_u, σ_v and ε_j. The corresponding likelihood function must be maximized:

$$L(\alpha, \beta, \sigma, \lambda) = \text{constant} - n \cdot \ln(\sigma) + \sum_{j=1}^{n} \ln\, \Phi\left(-\frac{\varepsilon_j\lambda}{\sigma}\right) - \frac{1}{2\sigma^2}\sum_{j=1}^{n} \varepsilon_j^2, \tag{7}$$

where ε_j is defined by

$$\varepsilon_j = y_j - (\beta_0 + \sum_{i=1}^{m} \beta_i \cdot x_{i,j}). \tag{8}$$

After this first step, the individual technical efficiency can be obtained by decomposing the estimated error term $\hat{\varepsilon}_j$ into an estimated noise term $\hat{v}_j$ and an estimated inefficiency term $\hat{u}_j$. For the standard normal-half normal model, Jondrow et al (1982) (JMLS) showed that the conditional distribution of u, given the composed error term ε, is

$$f(u|\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_*} \cdot \frac{\exp\left[-\frac{(u-\mu_*)^2}{2\sigma_*^2}\right]}{\left[1 - \Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]}, \tag{9}$$

with $\mu_* = -\varepsilon\sigma_u^2/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$.

Based on the maximum likelihood estimates, individual technical efficiency can be estimated by several point estimators. In this study, we use the point estimator proposed by Battese and Coelli (1988):

[3]The normal-half normal model is the most common model. There are other models which mainly differ in the assumption with respect to the inefficiency distribution, e.g. the normal-exponential model. For a comprehensive treatment of the different models, see Kumbhakar and Lovell (2003).

$$\hat{TE}_j = \hat{E}(\exp(-u_j)|\hat{\varepsilon}_j) = \frac{\Phi(\hat{\mu}_{*j}/\hat{\sigma}_* - \hat{\sigma}_*)}{\Phi(\hat{\mu}_{*j}/\hat{\sigma}_*)} \cdot \exp\left(\frac{1}{2}\hat{\sigma}_*^2 - \hat{\mu}_{*j}\right). \tag{10}$$

This estimator is optimal in the sense of minimizing the mean square error and is mostly used in empirical and theoretical applications (cf. Bogetoft and Otto (2011)). It is worth emphasizing that the JMLS estimator is not a consistent estimator of u_j. Independent of the sample size n, there is only one observation of each DMU_j. Therefore, the JMLS estimator does not converge to u_j, even if the sample size reaches infinity, but converges to $E(u|\varepsilon)$, which is the mean of the distribution from which u_j is drawn (cf. Greene (2008)). However, note that we apply the Battese and Coelli (1988) point estimator, which estimates the technical efficiency more accurate than the JMLS estimator (see Battese and Coelli (1988)). Using cross-sectional data, this is the best that can be achieved.

In short, the SFA ML estimation consists of two steps. Firstly, the parameters are estimated using the maximum likelihood method (equation (7)). Based on the maximum likelihood estimates, the individual efficiency of each DMU is estimated using the Battese and Coelli (1988) point estimator, equation (10).

2.2.2 SFA method of moments (SFA MoM)

An alternative to the maximum likelihood estimation is the method of moments, which splits the first step into two parts. In the first part (A), an OLS regression is used to obtain estimates for the composed error term (see also Figure 2). Using OLS regression to estimate the production function, the estimates for all slope coefficients (β_i) are consistent. However, the intercept $\hat{\beta}_{0,OLS}$ is biased by $E(u_j)$ and therefore, the estimated OLS residuals $\hat{\varepsilon}_{j,OLS}$ does not equal the composed error term ε_j, which has to be estimated ($\hat{\varepsilon}_j \neq \hat{\varepsilon}_{j,OLS}$) (cf. Kumbhakar and Lovell (2003)).

Assuming the normal-half normal model, this bias can be corrected – as indicated in Part B of Figure 2 – by using the fact that $E(u_j)$ is a constant and the central moments of the composed error term ε_j to be estimated are the same as those of $\hat{\varepsilon}_{j,OLS}$. The second and third central moments of the distribution can be estimated from the OLS residuals $\hat{\varepsilon}_{j,OLS}$ in the following way (cf. Kuosmanen and Kortelainen (2010)):

$$\hat{M}_f = \frac{1}{n}\sum_{j=1}^{n}(\hat{\varepsilon}_{j,OLS} - \hat{E}(\varepsilon_{j,OLS}))^f \qquad f = 2, 3. \tag{11}$$

Consequently, we can estimate the standard deviation of the noise term σ_v and the inefficiency term σ_u by:

$$\hat{\sigma}_u = \sqrt[3]{\frac{\hat{M}_3}{\sqrt{\frac{2}{\pi}} \cdot \left(1 - \frac{4}{\pi}\right)}}, \tag{12}$$

$$\hat{\sigma}_v = \sqrt{\hat{M}_2 - \left(1 - \frac{2}{\pi}\right)\hat{\sigma}_u^2}. \quad (13)$$

Subsequently, a consistent estimate for the intercept of the production function is given by:

$$\hat{\beta}_0 = \hat{\beta}_{0,OLS} + \hat{E}(u_j) = \hat{\beta}_{0,OLS} + \sqrt{\frac{2}{\pi}}\hat{\sigma}_u. \quad (14)$$

After shifting the OLS frontier upwards by the expected value of the inefficiency term, all estimates are unbiased and consistent (see Aigner et al (1977), Kumbhakar and Lovell (2003) and Greene (2008)) and the estimate of the composed error term ε_j can be calculated by

$$\hat{\varepsilon}_j = \hat{\varepsilon}_{j,OLS} - \sqrt{\frac{2}{\pi}}\hat{\sigma}_u. \quad (15)$$

Analogously to the maximum likelihood technique, firm-specific efficiency is estimated by means of the Battese and Coelli (1988) point estimator, equation (10), in a second step.

2.3 Stochastic non-smooth envelopment of data (StoNED)

2.3.1 StoNED pseudolikelihood (StoNED PL)

Kuosmanen and Kortelainen (2010) recently introduced the *stochastic non-smooth envelopment of data* (StoNED). StoNED avoids the main disadvantage of SFA – its parametric nature – by using convex nonparametric least squares (CNLS) to estimate the production function. CNLS does not require an assumption about the functional form of the production function, but determines a frontier from the family of continuous, monotonically increasing, concave functions which best fits the data (see Kuosmanen (2008)).

Similar to the procedure for the SFA MoM, step one consists of two parts (see Figure 2). Instead of using OLS regression in Part A, the shape of the production function is estimated by CNLS regression. In order to obtain the CNLS residuals $\varepsilon_{j,CNLS}$, the following quadratic programming problem has to be solved (cf. Kuosmanen and Kortelainen (2010))

$$\begin{aligned}
&\text{minimize}_{\hat{q},\beta_0,\beta_i} && \sum_{j=1}^{n}(ln(q_j) - ln(\hat{q}_j))^2 && (16)\\
&\text{subject to} && \hat{q}_j = \beta_{0,j} + \sum_{i=1}^{m}\beta_{i,j}z_{i,j}, \\
& && \beta_{0,j} + \sum_{i=1}^{m}\beta_{i,j}z_{i,j} \leq \beta_{0,h} + \sum_{i=1}^{m}\beta_{i,h}z_{i,j} \quad \forall \quad h,j = 1,...,n \text{ and } i = 1,...,m, \\
& && \beta_{i,j} \geq 0 \quad \forall \quad j = 1,...,n \text{ and } i = 1,...,m. \\
&\text{with} && \varepsilon_{j,CNLS} = ln(q_j) - ln(\hat{q}_j).
\end{aligned}$$

Using CNLS, we obtain estimates $\hat{\varepsilon}_{j,CNLS}$ for the deviation from the estimated production function. However, these estimates are biased in a similar manner to the OLS residuals $\hat{\varepsilon}_{j,OLS}$. Therefore, in Part B, distributional assumptions on the inefficiency and noise term are required and an estimation technique – pseudolikelihood or method of moments – has to be applied.

Assuming the normal-half normal model, the pseudolikelihood (PL) approach, suggested by Fan et al (1996), can be applied. We set $\sigma = \sigma_u + \sigma_v$, $\lambda = \frac{\sigma_u}{\sigma_v}$ and maximize the following log-likelihood function:

$$\ln L(\lambda) = -n \ln \hat{\sigma} + \sum_{j=1}^{n} \ln \Phi\left[\frac{-\hat{\varepsilon}_j\lambda}{\hat{\sigma}}\right] - \frac{1}{2\hat{\sigma}^2}\sum_{j=1}^{n}\hat{\varepsilon}_j^2, \tag{17}$$

$$\hat{\varepsilon}_j = \hat{\varepsilon}_{j,CNLS} - \frac{\sqrt{2}\lambda\hat{\sigma}}{\sqrt{\pi\left(1+\lambda^2\right)}}, \tag{18}$$

$$\hat{\sigma} = \sqrt{\frac{\frac{1}{n}\sum\limits_{j=1}^{n}\hat{\varepsilon}_{j,CNLS}^2}{1 - \frac{2\lambda^2}{\pi(1+\lambda)}}}. \tag{19}$$

When the optimal solution for $\hat{\lambda}$ is found, the estimates for $\hat{\varepsilon}_j$ and $\hat{\sigma}$ can be calculated by equations (18) and (19).

In analogy to SFA, in the second step, the Battese and Coelli (1988) point estimator, equation (10), is used to calculate the technical efficiency for each DMU.

2.3.2 StoNED method of moments (StoNED MoM)

The method of moments can be used as an alternative estimation technique to pseudolikelihood. Accordingly, part A of step one is the same as described above. The shape of the production function is estimated by CNLS regression. In accordance with the SFA MoM, in Part B, the central moments of the CNLS residuals $\varepsilon_{j,CNLS}$ are calculated by using equation (11). The standard deviations of the inefficiency $\hat{\sigma}_u$ and noise $\hat{\sigma}_v$ term are then estimated using equations (12) and (13), respectively. To complete step one, $\hat{\varepsilon}_j$ is obtained by equation (15). Again, the technical efficiency is obtained by the Battese and Coelli (1988) point estimator, equation (10), in the second step. Figure 2 shows the procedure of the methods in detail. The numbers in

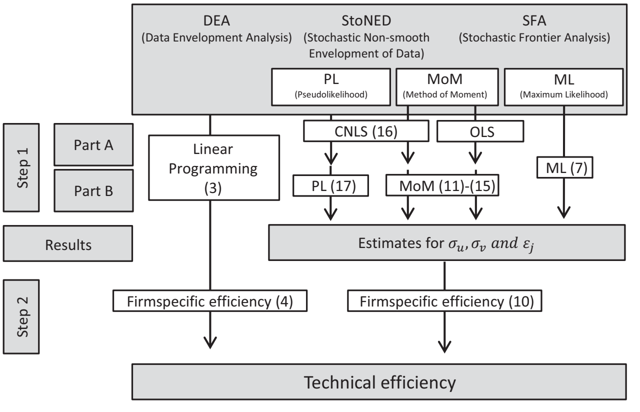

Figure 2: **Detailed overview of the methods procedure.**

brackets refer to the respective equations above.

3 Simulation Design

The aim of this paper is to evaluate the presented methods within the controlled environment of an MC simulation. Using empirical data, it is impossible to evaluate the performance of different methods, because the “true” efficiency is not known. Hence, MC simulations are used to avoid this problem. As stated by Perelman and Santin (2009), MC studies are the “statistical referee” most frequently used to verify the potential strengths and weaknesses of competing estimation methods. They enable researchers to generate their own artificial dataset under specific assumptions. For the data generating process (DGP), the underlying assumptions have to be defined. A certain set of assumptions is referred to as “setting”. Within a given setting, the DGP can be replicated several times in order to obtain reliable results. By analyzing different settings, for instance varying the number of DMUs, the influence of this specific factor can be measured. The difficulty is to decide how the settings should be varied, so as to derive a wide and meaningful spectrum.

The first best optimum would be to consider all possible specifications of influencing factors. As this approach becomes increasingly complex, an alternative has to be used. In Andor and Hesse (2011), we defined a standard set, i.e. one specification for all influencing factors. This standard set was used as the point of reference for the following sensitivity analysis. Accordingly, we varied the different influencing factors successively, while keeping the remaining factors unchanged. This kind of analysis facilitates the use of more specifications for a single factor. However, it is restricted in such a way that all the other parameters are kept unchanged. In this paper, we use a compromise to avoid this limitation. We create 12 standard sets which vary with respect to the number of decision making units, the production function and the composite error term. This seems to be an appropriate approach for our purpose, as our analysis is multidimensional and the conclusions are based on a wider basis. In total, we analyze the results of 200 settings and each is replicated 50 times (R=50), so that we consider 10.000 datasets. As especially the DEA and CNLS regression of the StoNED are time-consuming to replicate, this represents a reasonable compromise between accuracy and computational time.

Below, we define the DGP for the 12 standard sets. We follow Ruggiero (1999), Jensen (2005) and others, by using two inputs, z_1 and z_2, which are generated from a uniform distribution with the interval $(5, 15)$. Furthermore, we assume that there is no collinearity between z_1 and z_2 ($\rho = 0$) and that the inefficiency and the noise term are homoscedastic. The endogenous variable q_j, the output, is calculated by the following equation:

$$ln(q_j) = \underbrace{ln(F(z_{i,j}))}_{Production\ Function} + \underbrace{\varepsilon_j}_{Composed\ error\ term} \quad with\ \varepsilon_j = v_j - u_j, \tag{20}$$

where u_j and v_j represent the inefficiency term and the statistical noise term, respectively. We assume that the inefficiency term is exponentially distributed $u_j \sim \text{Exp}(\mu{=}1/6)$, with parameter μ representing the expected inefficiency. This leads to an expected (technical) efficiency of approximately 86%. The noise term is normally distributed $v_j \sim N(0, \sigma_v^2)$ with $\sigma_v = \rho_{nts} \cdot \mu$, where ρ_{nts} represents the noise-to-signal ratio, i.e. $\rho_{nts} = \frac{\sigma_v}{\sigma_u}$. This DGP calibration is similar to the procedure in Kuosmanen and Kortelainen (2010) and Simar and Zelenyuk (2011). Regarding the production function, we use three different specifications that are also used in other MC studies. They vary with respect to returns to scale and input substitution (see Table 1). The combination with a varying number of DMUs (50 and 100) and two specifications for the noise-to-signal ratio (0 and 1) results in the 12 standard settings. For the remaining 200 settings, we describe the variation of the DGP at the beginning of the specific analysis. The five

No	PF (F(x))	Description	Parametrization	Source
I	$\sum_{i=1}^{m} \beta_i \cdot ln(z_{i,j})$	Cobb-Douglas, IRS	$\beta_1 = \beta_2 = 0.6$	[a]
II	$\ln([\sum_{i=1}^{m} \alpha_i \cdot z_{i,j}^{-\rho_i}]^{-\delta/\rho})$	CRESH	δ=1, α_1=α_2=0.5, ρ=ρ_i=2	[b]
III	$\beta_0 + \sum_{i=1}^{m} \beta_i \cdot ln(z_{i,j}) + 0.5 \cdot \sum_{i=1}^{m} \sum_{f=1}^{m} \beta_{i,f} \cdot ln(z_{i,j}) \cdot ln(z_{i,j})$	Translog	β_0=1, β_1= β_2=0.3, $\beta_{11} = \beta_{22} = \beta_{12} = \beta_{21}$= 0.1	[c]

Table 1: **Standard sets: Production functions.** IRS: Increasing returns to scale. [a] Adler and Yazhemsky (2010) in modified form, [b] Yu (1998) in modified form, [c] Cordero et al (2009).

methods DEA, SFA MoM, SFA ML, StoNED MoM and StoNED PL are applied with the model specifications described in section 2 using the drawn inputs and the generated output.[4] Olson et al (1980) and Banker et al (1993) identify that there can be two problems with the method of moments approach. Type I failure occurs when the skewness of the error term ε is positive $\hat{M}_3 \geq 0$. We follow Kuosmanen and Kortelainen (2010) in these cases and set $\hat{M}_3 = -0.0001$. Type II failure occurs when the estimated standard deviation of the noise term ($\hat{\sigma}_v$) is negative. In accordance with Kuosmanen and Kortelainen (2010), we set $\hat{\sigma}_v = 0.0001$ in these cases.

Finally, the evaluation of the methods requires a performance criterion. Ruggiero (1999) and others focus on ranking accuracy, using the average rank correlation between "true" and estimated technical efficiency. However, from our perspective, ranking accuracy is an inferior performance criterion in real-world applications, because policy makers often have to set individual efficiency objectives. Hence, the ability to measure individual efficiency is the most important factor. Accordingly, we use the mean absolute deviation (MAD) between the estimated and the true technical efficiency value, as our main performance criterion.

$$MAD = \frac{1}{nR} \sum_{r=1}^{R} \sum_{j=1}^{n} \left| \hat{TE}_{r,j} - TE_{r,j} \right|, \tag{21}$$

where $\hat{TE}_j$ denotes the estimated and TE_j the true technical efficiency, and r is the index for the replications for a certain setting. In order to gain additional insight into the influence of a particular factor, we also calculate the following three additional information criteria: Mean deviation (MD), mean squared error (MSE) and mean rank correlation (MRC). We discuss them, whenever they yield additional information about performance variation. The information criteria are defined by:

$$MD = \frac{1}{nR} \sum_{r=1}^{R} \sum_{j=1}^{n} (\hat{TE}_{r,j} - TE_{r,j}), \tag{22}$$

$$MSE = \frac{1}{nR} \sum_{r=1}^{R} \sum_{j=1}^{n} (\hat{TE}_{r,j} - TE_{r,j})^2. \tag{23}$$

The Spearman rank correlation is defined as the Pearson linear correlation of the ranked technical efficiencies:

$$MRC = \frac{1}{R} \sum_{r=1}^{R} \frac{\sum_{j=1}^{n} (\hat{te}_{r,j} - \bar{\hat{te}}_r)(te_{r,j} - \bar{te}_r)}{\sqrt{\sum_{j=1}^{n} (\hat{te}_{r,j} - \bar{\hat{te}}_r)^2 \sum_{j=1}^{n} (te_{r,j} - \bar{te}_r)^2}}, \tag{24}$$

where the n technical efficiencies TE_j are converted to ranks te_j. The results for these three criteria are shown in the appendix. Furthermore, we briefly review the aggregated results in

[4]StataSE 11.2 is used for the implementation of the DGP and the estimation of DEA while General Algebraic Modeling System (GAMS) Version 23.3.2 is used to estimate the other four methods. The codes are available upon request from the authors.

the next section.

Note that we measure performance in terms of technical efficiency applying the Battese and Coelli (1988) point estimator introduced in equation (10), while the studies of Simar and Zelenyuk (2011) and Kuosmanen and Kortelainen (2010) rely on the estimated production function as the performance criterion. There are two reasons for doing this. First, this allows us to provide a complementary view on performance of the methods. Second, from our point of view, most of the practical applications of efficiency estimation methods as well as many of the empirical studies focus on the technical efficiency. Nevertheless, to provide a comprehensive comparison, we also examine the performance of the methods in terms of the estimated production frontier in Subsection 4.6.

4 Results

4.1 Overall results

In this section, we present and discuss the results of the simulation study. In total, we have results from 200 settings. In the interests of clarity, the analysis is carried out in two stages. Firstly, we focus on a comparison of the aggregated results of the 200 settings and discuss some important characteristics of the methods. In the second stage, we successively analyze the influence of specific factors on the performance of the various methods.

Table 2 shows the mean deviation and the average of the other three performance criteria for all 200 settings. We additionally order the methods from best to worst, for each setting under consideration, so as to calculate the mean rank for each performance criterion. A rank of one represents the "winner" and a rank of five the "loser".

	DEA	SFA MoM	SFA ML	StoNED MoM	StoNED PL
MD	-0.0724	-0.0519	-0.0242	-0.0412	0.0260
MAD	0.1105	0.0853	0.0649	0.0862	0.0700
Rank (MAD)	*3.63*	*3.36*	*2.13*	*3.52*	*2.38*
MSE	0.0268	0.0123	0.0104	0.0131	0.0102
Rank (MSE)	*3.58*	*3.27*	*2.16*	*3.58*	*2.44*
MRC	0.6627	0.6878	0.7052	0.6357	
Rank (MRC)	*3.18*	*2.01*	*1.38*	*3.44*	

Table 2: **Overview of the performance criteria for all 200 settings.**

The mean deviation (MD) is an important characteristic of the methods, as it shows the bias of the efficiency estimation. The results highlight an interesting difference between the StoNED PL and the other methods. While all other methods underestimate on average, the StoNED

PL overestimates the efficiency. The average MD also shows, that SFA is the method with the lowest bias, whereas DEA is the method with by far the greatest bias. The fact that DEA underestimates is not particularly surprising, because we consider noise in approximately 53% of the settings. A second general peculiarity is that the MoM methods, SFA MoM and StoNED MoM, achieve relatively similar results with regard to the MD, MAD and MSE. In contrast, the results of StoNED PL differ considerably from the SFA ML, as well as the StoNED MoM in general. This conclusion can be drawn for almost all 200 settings.

The aggregated results for the MAD suggest that SFA ML is the best method. Nevertheless, the recently introduced StoNED PL seems to be a serious competitor, as it has the second lowest MAD and its MSE is even slightly lower than that of SFA ML. The MoM estimation techniques, SFA MoM and StoNED MoM, achieve similar average MADs. DEA exhibits the highest MAD, but the rank(MAD) is similar to those of the MoM methods. Except for the comparison of StoNED PL and SFA ML, MAD and MSE come to the same conclusion.

The rank correlation demonstrates both a characteristic and a weakness of the StoNED methods. The former is that both methods, StoNED MoM and StoNED PL, have the same rank correlation. The weakness is that it has a lower average rank correlation than the other methods. As a result, if practitioners or researchers regard the rank correlation as the appropriate criterion for their purposes, our results advise against using StoNED.

As mentioned above, the chosen settings aim to cover a wide range of assumptions, and the aggregated results shed light on the overall performance. However, each method has its own strengths and weaknesses. Below, we consider two specific subsamples of the 200 settings in order to emphasize them.

	DEA	SFA MoM	SFA ML	StoNED MoM	StoNED PL
MD	0.0112	-0.0552	-0.0035	-0.0392	0.0200
MAD	0.0407	0.0694	0.0231	0.0704	0.0519
Rank (MAD)	*2.31*	*4.07*	*1.32*	*4.05*	*3.24*
MSE	0.0048	0.0075	0.0023	0.0084	0.0063
Rank (MSE)	*2.29*	*3.87*	*1.40*	*4.18*	*3.28*
MRC	0.8449	0.8665	0.9039	0.7973	
Rank (MRC)	*2.84*	*2.32*	*1.26*	*3.59*	

Table 3: **Overview of the performance criteria in the subsample without noise ($\rho_{nts} = 0$).**

The underlying assumption of DEA, that there is no noise in the data, is violated in every setting with $\rho_{nts} > 0$. Hence, we compare the deterministic DEA with the stochastic methods in a nondiscriminatory subsample, i.e. we restrict the analysis to the 94 settings without noise ($\rho_{nts} = 0$). Table 3 summarizes the results. In general, all performance criteria for all methods improve considerably in the subsample without noise. However, it is interesting to compare the relative performance of the methods. Even in this subsample, SFA ML is still the best method, but followed closely by DEA. For these methods, the MD yields what is to be expected. In the scenario without noise, the underestimation declines. This is particularly true for the DEA, as the MD changes from -0.0724 to 0.0112. DEA overestimates in settings without noise, whereas it underestimates in those with noise. The StoNED PL and SFA MoM underestimate more in the scenario without noise. Although this leads to a lower efficiency bias (MD) and MAD for the StoNED PL, its relative performance deteriorates in comparison to DEA and SFA. A further conclusion is that the relative performance of the MoM technique worsens considerably in the scenario without noise. This conclusion supports the recommendation of Olson et al (1980) and Coelli (1995) that the SFA ML method is preferable to the SFA MoM, when there is little noise in the data.

	DEA	SFA MoM	SFA ML	StoNED MoM	StoNED PL
MD	-0.1334	-0.0479	-0.0358	-0.0416	0.0330
MAD	0.1607	0.0959	0.0931	0.0968	0.0820
Rank (MAD)	*4.63*	*2.86*	*2.58*	*3.22*	*1.71*
MSE	0.0416	0.0154	0.0151	0.0160	0.0127
Rank (MSE)	*4.56*	*2.85*	*2.56*	*3.25*	*1.80*
MRC	0.5318	0.5622	0.5634	0.5226	
Rank (MRC)	*3.48*	*1.73*	*1.45*	*3.34*	

Table 4: **Overview of the performance criteria in the subsample with noise ($\rho_{nts} > 0$).**

For the contrary subsample, i.e. all 106 settings with noise ($\rho_{nts} > 0$), all performance criteria deteriorate. Here, it is particularly remarkable that StoNED PL outperforms the SFA ML (and all other methods) in terms of MAD and MSE. Consequently, we can conclude that a great virtue of StoNED PL is its ability to measure efficiency when there is (a lot of) noise in the data (see also section 4.3.1). In addition, the performance of the MoM methods, especially SFA MoM, are also relatively good and are similar to that of the SFA ML. Again, this supports the conclusion of Olson et al (1980) and Coelli (1995) that the MoM estimation technique has its comparative advantage vis-à-vis the maximum likelihood estimation technique, when the ratio of noise to inefficiency is high. However, this conclusion seems invalid for the StoNED PL, because it performs considerably better than the StoNED MoM. While the MAD and rank(MAD) are very similar for the StoNED MoM and the SFA methods, the rank(MAD) of StoNED PL is 1.71.

In the following analysis, we focus on analyzing the influence of factors on the particular method and the corresponding relative performance. We divide this analysis into three main categories of sample size, error term and production function. The MAD is our main performance criterion and the results are presented in tables in which the parameter values for the factor under inspection, as well as the five methods, are arranged vertically, while the remaining (control) variables are arranged horizontally. As mentioned earlier, the results for the other performance criteria can be found in the appendix.

4.2 Variation of sample size

In several MC studies, sample size has been identified as one important factor influencing the performance of efficiency estimation methods (see, for instance, Olson et al (1980), Banker et al (1993), Ruggiero (1999) and Badunenko et al (2011)). In addition to our standard sample size assumptions of 50 and 100 DMUs, we now consider two additional number of DMUs: 20 and 200 DMUs. In comparison to Olson et al (1980), these sample sizes are relatively small. However, problems with more than 300 observations can take several days for the StoNED method (see Kuosmanen (2012b)). Furthermore, from our perspective, these sample sizes are the most relevant for real-world applications. Table 5 contains the resulting MAD values for the variation of sample size.

	NTS	0			1		
Method	PF	PF I	PF II	PF III	PF I	PF II	PF III
DEA	DMU = 20	0.0539	0.0545	0.0551	0.1209	0.1055	0.1354
	DMU = 50	0.0314	0.0357	0.0373	0.1384	0.1335	0.1544
	DMU = 100	0.0206	0.0232	0.0326	0.1729	0.1648	0.1960
	DMU = 200	0.0127	0.0170	0.0319	0.1985	0.1814	0.2210
SFA MOM	DMU = 20	0.0569	0.0634	0.0621	0.0952	0.1004	0.0940
	DMU = 50	0.0688	0.0697	0.0755	0.0917	0.1032	0.0976
	DMU = 100	0.0791	0.0737	0.0811	0.0987	0.1034	0.0896
	DMU = 200	0.0838	0.0832	0.0771	0.1029	0.1086	0.0997
SFA ML	DMU = 20	0.0279	0.0422	0.0273	0.1135	0.1182	0.1080
	DMU = 50	0.0101	0.0344	0.0130	0.0949	0.1006	0.0960
	DMU = 100	0.0053	0.0332	0.0119	0.0907	0.0912	0.0894
	DMU = 200	0.0024	0.0315	0.0114	0.0866	0.0939	0.0868
STONED MOM	DMU = 20	0.0610	0.0665	0.0684	0.1018	0.0923	0.0951
	DMU = 50	0.0676	0.0634	0.0740	0.0906	0.1029	0.1011
	DMU = 100	0.0734	0.0674	0.0799	0.0980	0.1021	0.0900
	DMU = 200	0.0823	0.0798	0.0774	0.0965	0.1015	0.0994
STONED PL	DMU = 20	0.0658	0.0642	0.0710	0.0986	0.0930	0.0900
	DMU = 50	0.0466	0.0467	0.0509	0.0809	0.0814	0.0862
	DMU = 100	0.0378	0.0368	0.0414	0.0749	0.0765	0.0776
	DMU = 200	0.0376	0.0644	0.0520	0.0931	0.0851	0.0950

Table 5: **Variation of sample size. Performance criterion: Mean absolute deviation (MAD).** DGP: *Sample size*: DMU= 20, 50, 100, 200; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: NO; *Production function*: PF I (Cobb Douglas with increasing returns to scale), PF II (CRESH), PF III (Translog); Collinearity: 0; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z): m= 2.

DEA is affected by a variation in sample size, but the direction of the effect depends on the underlying scenario. In the scenario without noise (NTS=0), the performance of DEA improves with an increasing number of DMUs, while the performance deteriorates with a growing number of DMUs in the scenario with noise (NTS=1). This diametrical effect is not an exception, but we also find it when analyzing other influencing factors. The reason is that, in general, DEA overestimates in the scenario without noise and underestimates in the scenario with noise (see MD in Table 22 in the appendix). Furthermore, an increasing sample size leads to a decreasing MD, i.e. the more observations, the more DMUs are underestimated. This can be explained by the fact that the relative number of DMUs on the efficient frontier decrease with the sample size. As a result, the "sample size effect" leads to a "downward shift" of the average estimated efficiency and so partially counteracts the overestimation in the scenario without noise. Therefore, it has a positive impact on the average performance. In contrast, it enforces the "noise effect", so that the underestimation in the settings with NTS=1 and a sample size of 200 DMUs is glaringly obvious and the performance is considerably poorer. However, the rank correlation generally improves with a growing number of DMUs (see MRC in Table 24 in the appendix).

Regarding the variation of sample size, the MoM models are affected more in the scenario without noise than with noise. In the former scenario, an increasing sample size seems to worsen their performance. In contrast, the SFA ML performs better with increasing sample size. Interestingly, the StoNED PL performance also improves with an increasing number of DMUs, but for 200 DMUs, this relationship reverses, i.e. the performance worsens. While the effect on the performance of the stochastic methods in terms of the MAD is ambiguous, the effect on MD and MRC is generally unambiguous. The underestimation and the rank correlation increases with the sample size, i.e. the MD decreases and the MRC increases (see Table 22 and 24 in the appendix). The only exception is the setting with 200 DMUs for both StoNED models. For this setting the MD is higher and the MRC is lower as in comparison to the setting with 100 DMUs.

4.3 Variation of the error term

4.3.1 Noise-to-signal ratio (NTS)

The noise-to-signal ratio represents the relationship between noise and inefficiency and is expressed by $\rho_{nts} = \frac{\sigma_v}{\sigma_u}$. Several studies verify that this ratio has a crucial impact on efficiency estimation methods (see, Olson et al (1980), Banker et al (1993), Ruggiero (1999), Ondrich and Ruggiero (2001), Jensen (2005) and Badunenko et al (2011)). In order to analyze the influence, we generate data with $\rho_{nts} = 0, 0.5, 1$ and 2. Table 6 presents the results.

Method	DMU	50			100		
	PF	PF I	PF II	PF III	PF I	PF II	PF III
DEA	NTS = 0	0.0314	0.0357	0.0373	0.0206	0.0232	0.0326
	NTS = 0.5	0.0650	0.0607	0.0815	0.0728	0.0674	0.0952
	NTS = 1	0.1384	0.1335	0.1544	0.1729	0.1648	0.1960
	NTS = 2	0.2908	0.3050	0.3247	0.3480	0.3331	0.3586
SFA MoM	NTS = 0	0.0688	0.0697	0.0755	0.0791	0.0737	0.0811
	NTS = 0.5	0.0799	0.0832	0.0883	0.0900	0.0881	0.0813
	NTS = 1	0.0917	0.1032	0.0976	0.0987	0.1034	0.0896
	NTS = 2	0.1240	0.1255	0.1240	0.1153	0.1260	0.1311
SFA ML	NTS = 0	0.0101	0.0344	0.0130	0.0053	0.0332	0.0119
	NTS = 0.5	0.0623	0.0663	0.0617	0.0587	0.0642	0.0585
	NTS = 1	0.0949	0.1006	0.0960	0.0907	0.0912	0.0894
	NTS = 2	0.1427	0.1516	0.1501	0.1313	0.1444	0.1400
StoNED MoM	NTS = 0	0.0676	0.0634	0.0740	0.0734	0.0674	0.0799
	NTS = 0.5	0.0806	0.0791	0.0885	0.0894	0.0986	0.0826
	NTS = 1	0.0906	0.1029	0.1011	0.0980	0.1021	0.0900
	NTS = 2	0.1289	0.1279	0.1282	0.1208	0.1258	0.1348
StoNED PL	NTS = 0	0.0466	0.0467	0.0509	0.0378	0.0368	0.0414
	NTS = 0.5	0.0652	0.0636	0.0649	0.0580	0.0593	0.0587
	NTS = 1	0.0809	0.0814	0.0862	0.0749	0.0765	0.0776
	NTS = 2	0.1050	0.1100	0.1118	0.1036	0.1048	0.1048

Table 6: **Variation of noise-to-signal ratio. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0, 0.5, 1 and 2; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: NO; *Production function*: PF I (Cobb Douglas with increasing returns to scale), PF II (CRESH), PF III (Translog); Collinearity: 0; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z): m= 2.

Obviously, all methods perform worse with an increasing noise-to-signal ratio, with respect to both the MAD and the MRC. Hence, the relative comparison is of primary importance as to which methods are influenced most. As DEA is deterministic, it is the method which is most negatively affected by this variation. On average, the DEA MAD is eleven times higher when noise-to-signal ratio is 2 instead of 0. However, even in the scenario without noise, SFA ML performs better in most of the settings and the order of methods in almost all settings, from best to worst is as follows: SFA ML, DEA, StoNED PL, StoNED MoM and SFA MoM. In contrast, StoNED PL is the least affected method: Its MAD also increase with an increasing noise-to-signal ratio, but in comparison to the other methods, its "competitiveness" increases. The ability to handle a lot of noise seems to be a comparative advantage of StoNED PL. In the scenario with NTS=2, the order is generally the following: StoNED PL, SFA MoM, StoNED MoM, SFA ML and DEA. So we can conclude, that the higher the noise-to-signal ratio, the better the StoNED PL and the MoM methods perform.

In these opposing cases (NTS=0 and NTS=2), the order of methods is comparatively consistent and the conclusions are relatively unambiguous. However, an assumption somewhere between these extremes could be more realistic. Note that a noise-to-signal ratio of two assumes that the data has twice as much noise as inefficiency. Would an efficiency estimation make sense in this case? Unfortunately, the conclusions are more ambiguous for the settings between these extremes. Given a NTS=0.5, StoNED PL and SFA ML are the best methods and DEA also delivers comparable results in most settings. The MoM methods perform worse than the others.

4.3.2 Distribution of the inefficiency term

In order to measure the influence of the inefficiency distribution, we vary the DGP with respect to it (cf., among others, Jensen (2005)). Apart from our standard exponential distribution $u_j \sim \text{Exp}(\mu{=}1/6)$, we use a half normal N^+ (0,0.021) and a beta distribution B (0.068,4) to generate the ineffiency term. The parametrization is chosen in such a manner that they have the same expected inefficiency value (see Table 7), whereupon the distributions differ with regard to the expected standard deviation and the skewness. The skewness represents the asymmetry regarding the inefficiency of the DMUs. The greater the skewness, the more DMUs are relatively efficient, but some DMUs are indeed very inefficient. Note that we still assume a half normally distributed inefficiency term for the stochastic methods.

Distribution	Expected		
	Mean	Standard deviation	Skewness
$N^+(0, 0.021)$	0.167	0.127	1
$Exp(\mu = 1/6)$	0.167	0.168	2
$B(0.068, 4)$	0.167	0.057	5.57

Table 7: **Variations of the inefficiency distribution.**

In general, all methods are affected by a variation in the inefficiency distribution (see Table 8), but the direction of the effect on the MAD differs. However, we can see a homogeneous effect of the variation on the MD and this explains the diverging effects on the MAD. The more skewed the inefficiency distribution, the lower the MD, that is, the underestimation of DMUs increases. As a result, the methods which generally overestimates are positively affected. These are the StoNED PL and the DEA in the scenario without noise. Again, DEA is negatively affected in the scenario with noise. In this case, DEA performs very poorly when inefficiency is drawn from the (more skewed) beta distribution.

As expected, the MoM methods achieve the best results, if they are not misspecified, i.e. inefficiency is generated by a half normal distribution. Surprisingly, this conclusion does not apply for the performance of SFA ML and StoNED PL. In most settings, the results are worse, when the assumptions are in accordance with the real DGP. For the StoNED PL, we give the explanation above, while the effect on SFA ML is surprising. However, the results suggest that a misspecification does not affect the ML performance as much as the MoM performance. This finding is important as, in contrast to the SFA ML, the SFA MoM estimates the slope of the production function without an assumption about the error term distribution, which is why one might expect a misspecified inefficiency distribution to exert a stronger impact on the SFA ML performance. Except for a few settings, we can conclude that the best PL and ML results are obtained when the inefficiency is drawn from a beta distribution. Particularly in the noise scenarios, it seems that for these methods, the skewness of the inefficiency distribution is more decisive than the specific form of distribution.

Method	NTS	0						1					
	DMU	50			100			50			100		
	PF	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III
DEA	$u_i \sim HN(\mu = 1/6)$	0.0411	0.0450	0.0407	0.0262	0.0307	0.0344	0.1404	0.1306	0.1571	0.1617	0.1472	0.1800
	$u_i \sim Exp(\mu = 1/6)$	0.0314	0.0357	0.0373	0.0206	0.0232	0.0326	0.1384	0.1335	0.1544	0.1729	0.1648	0.1960
	$u_i \sim Beta(\mu = 1/6)$	0.0058	0.0016	0.0296	0.0068	0.0008	0.0360	0.2050	0.1905	0.2106	0.2338	0.2279	0.2624
SFA MoM	$u_i \sim HN(\mu = 1/6)$	0.0309	0.0422	0.0339	0.0248	0.0385	0.0259	0.0820	0.0824	0.0814	0.0742	0.0819	0.0793
	$u_i \sim Exp(\mu = 1/6)$	0.0688	0.0697	0.0755	0.0791	0.0737	0.0811	0.0917	0.1032	0.0976	0.0987	0.1034	0.0896
	$u_i \sim Beta(\mu = 1/6)$	0.0828	0.0872	0.0769	0.0901	0.0979	0.0899	0.0891	0.1006	0.0992	0.0893	0.0976	0.0876
SFA ML	$u_i \sim HN(\mu = 1/6)$	0.0170	0.0338	0.0150	0.0084	0.0326	0.0127	0.0981	0.1002	0.1001	0.0861	0.1001	0.0967
	$u_i \sim Exp(\mu = 1/6)$	0.0101	0.0344	0.0130	0.0053	0.0332	0.0119	0.0949	0.1006	0.0960	0.0907	0.0912	0.0894
	$u_i \sim Beta(\mu = 1/6)$	0.0000	0.0400	0.0217	0.0000	0.0399	0.0248	0.0771	0.0783	0.0873	0.0764	0.0902	0.0655
StoNED MoM	$u_i \sim HN(\mu = 1/6)$	0.0396	0.0409	0.0438	0.0315	0.0303	0.0354	0.0817	0.0803	0.0789	0.0754	0.0785	0.0767
	$u_i \sim Exp(\mu = 1/6)$	0.0676	0.0634	0.0740	0.0734	0.0674	0.0799	0.0906	0.1029	0.1011	0.0980	0.1021	0.0900
	$u_i \sim Beta(\mu = 1/6)$	0.0751	0.0644	0.0698	0.0821	0.0788	0.0818	0.0980	0.1057	0.1017	0.0929	0.0975	0.0945
StoNED PL	$u_i \sim HN(\mu = 1/6)$	0.0623	0.0618	0.0579	0.0463	0.0448	0.0554	0.0952	0.0929	0.0903	0.0935	0.0925	0.0950
	$u_i \sim Exp(\mu = 1/6)$	0.0466	0.0467	0.0509	0.0378	0.0368	0.0414	0.0809	0.0814	0.0862	0.0749	0.0765	0.0776
	$u_i \sim Beta(\mu = 1/6)$	0.0279	0.0238	0.0295	0.0269	0.0255	0.0320	0.0549	0.0548	0.0547	0.0489	0.0520	0.0465

Table 8: **Variation of the distribution of the inefficiency term. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6), N^+ (0,0.021) and B (0.068,4); Heteroscedasticity: NO; *Production function*: PF I (Cobb Douglas with increasing returns to scale), PF II (CRESH), PF III (Translog); Collinearity: 0; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z): m= 2.

4.3.3 Heteroscedasticity

In the efficiency analysis literature, the effect of heteroscedasticity has been investigated by Caudill and Ford (1993), Caudill et al (1995), Kumbhakar (1997), Hadri (1999), Hadri et al (2003) and others. Caudill and Ford (1993) and Caudill et al (1995) point out that the performance of the efficiency estimation methods are affected by a heteroscedastic inefficiency term. Additionally, Hadri et al (2003) showed that inefficiency measures are also sensitive to heteroscedasticity in the noise term. Analogously to Kuosmanen and Kortelainen (2010) and Simar and Zelenyuk (2011), we investigate the influence of a heteroscedastic inefficiency term and leave the influence of a heteroscedastic noise term for further research. In order to analyze the influence of a heteroscedastic inefficiency term, we have to change the DGP, so that inefficiency depends on the size of the DMU. Following Simar and Zelenyuk (2011), we draw the inefficiency term from the half normal distribution $u_j|z_j \sim |N(0, (\sigma_u(z_{1,j} + z_{2,j})/w)^2|$, where σ_u is 0.3. We set $w = 28.72$ to ensure that the expected inefficiency ($\mu = 1/6$) remains unchanged. Otherwise, we would be mixing the effect of heteroscedasticity with that of a change in expected inefficiency. The noise term remains normally distributed, $v_j \sim N(0, \sigma_v^2)$, with $\sigma_v = \rho_{nts} \cdot E(\sigma_u) \cdot \sqrt{(\pi - 2)/\pi}$. Because the inefficiency is size-related, the noise-to-signal ratio varies for each replication, so that the parameter ρ_{nts} should be interpreted here as the average noise-to-signal ratio. The results are shown in Table 9.

Method	NTS	0						1					
	DMU	50			100			50			100		
	PF	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III
DEA	Homoscedastic	0.0314	0.0357	0.0373	0.0206	0.0232	0.0326	0.1384	0.1335	0.1544	0.1729	0.1648	0.1960
	Heteroscedastic	0.0409	0.0453	0.0406	0.0265	0.0296	0.0325	0.0957	0.0930	0.1152	0.1115	0.1043	0.1370
SFA MoM	Homoscedastic	0.0688	0.0697	0.0755	0.0791	0.0737	0.0811	0.0917	0.1032	0.0976	0.0987	0.1034	0.0896
	Heteroscedastic	0.0400	0.0490	0.0390	0.0318	0.0424	0.0319	0.0700	0.0763	0.0780	0.0716	0.0680	0.0714
SFA ML	Homoscedastic	0.0101	0.0344	0.0130	0.0053	0.0332	0.0119	0.0949	0.1006	0.0960	0.0907	0.0912	0.0894
	Heteroscedastic	0.0170	0.0398	0.0177	0.0084	0.0329	0.0120	0.0807	0.0888	0.0948	0.0753	0.0680	0.0773
StoNED MoM	Homoscedastic	0.0676	0.0634	0.0740	0.0734	0.0674	0.0799	0.0906	0.1029	0.1011	0.0980	0.1021	0.0900
	Heteroscedastic	0.0486	0.0464	0.0502	0.0388	0.0357	0.0401	0.0696	0.0748	0.0798	0.0715	0.0673	0.0730
StoNED PL	Homoscedastic	0.0466	0.0467	0.0509	0.0378	0.0368	0.0414	0.0809	0.0814	0.0862	0.0749	0.0765	0.0776
	Heteroscedastic	0.0756	0.0704	0.0749	0.0589	0.0583	0.0640	0.0856	0.0913	0.0953	0.0897	0.0782	0.0904

Table 9: **Influence of a heteroscedastic inefficiency term. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: YES; *Production function*: PF I (Cobb Douglas with increasing returns to scale), PF II (CRESH), PF III (Translog); Collinearity: 0; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z): m= 2.

All methods are affected by the presence of heteroscedasticity in inefficiency, but the direction of the effect is (surprisingly) divergent. In the scenario without noise, the DEA performance is worse with heteroscedasticity, but in the scenario with a NTS=1 the performance is considerably better when the inefficiency term is heteroscedastic. Surprisingly, in all settings, SFA MoM and StoNED MoM are substantially positively affected by the heteroscedastic inefficiency term. SFA ML seems to be unaffected in the scenario without noise, but in the scenario with noise, the performance also improves. StoNED PL is the only method which is consistently negatively influenced.

Again, these performance variations can be explained by the effect on the MD. A heteroscedastic inefficiency term leads to an increasing MD for all methods. Consequently, the growing overestimation causes an upward shift of the average estimated efficiency. For methods which generally underestimate, especially the MoM methods, this precipitates a performance improvement, whereas StoNED PL and DEA in the noise scenario are negatively affected.

4.4 Production Function

4.4.1 Functional Form of the Production Function

The influence of the production function is frequently referred to as important in the literature, but the variation of production functions under consideration has been limited so far (see Perelman and Santin (2009)). For example, Gong and Sickles (1992) use three different production functions, while Banker et al (1993) use two very similar ones in their MC studies. We use three different production functions within our standard settings (PF I, II and III) and extend the analysis by four additional production functions. Accordingly, we generate the data with a total of seven different production functions, which vary with respect to returns-to-scale and flexibility, see Table 10. We first discuss the influence of returns to scale, then the influence of elasticity of substitution and finally, we compare the results of all settings.

PF	Description	Parametrization
I	Cobb-Douglas, Increasing Return to Scale	$\beta_1 = \beta_2 = 0.6$
I.B	Cobb-Douglas, Constant Return to Scale	$\beta_1 = \beta_2 = 0.5$
I.C	Cobb-Douglas, Decreasing Return to Scale	$\beta_1 = \beta_2 = 0.4$
II	CRESH (Inputsubstitution=0.33)	$\rho = \rho_i = 2$
II.B	CRESH (Inputsubstitution=1.33)	$\rho = \rho_i = -0.25$
II.C	CRESH (Inputsubstitution=3)	$\rho = \rho_i = -0.67$
III	Translog	

Table 10: **Parametrization of the additional production functions.**

Method	NTS	0		1	
	DMU	50	100	50	100
DEA	PF I Cobb-Douglas (IRS)	0.0314	0.0206	0.1384	0.1729
	PF I.B Cobb-Douglas (CRS)	0.0330	0.0220	0.1273	0.1642
	PF I.C Cobb-Douglas (DRS)	0.0354	0.0217	0.1326	0.1566
	PF II CRESH (Inputsub. = 0.33)	0.0357	0.0232	0.1335	0.1648
	PF II.B CRESH (Inputsub. = 1.33)	0.0349	0.0216	0.1317	0.1587
	PF II.C CRESH (Inputsub. = 3)	0.0317	0.0201	0.1406	0.1658
	PF III Translog	0.0373	0.0326	0.1544	0.1960
SFA MoM	PF I Cobb-Douglas (IRS)	0.0688	0.0791	0.0917	0.0987
	PF I.B Cobb-Douglas (CRS)	0.0743	0.0719	0.0990	0.0972
	PF I.C Cobb-Douglas (DRS)	0.0617	0.0787	0.0933	0.0967
	PF II CRESH (Inputsub. = 0.33)	0.0697	0.0737	0.1032	0.1034
	PF II.B CRESH (Inputsub. = 1.33)	0.0599	0.0692	0.0996	0.1035
	PF II.C CRESH (Inputsub. = 3)	0.0753	0.0761	0.1061	0.1020
	PF III Translog	0.0755	0.0811	0.0976	0.0896
SFA ML	PF I Cobb-Douglas (IRS)	0.0101	0.0053	0.0949	0.0907
	PF I.B Cobb-Douglas (CRS)	0.0094	0.0046	0.1018	0.0902
	PF I.C Cobb-Douglas (DRS)	0.0091	0.0046	0.0985	0.0905
	PF II CRESH (Inputsub. = 0.33)	0.0344	0.0332	0.1006	0.0912
	PF II.B CRESH (Inputsub. = 1.33)	0.0113	0.0078	0.0972	0.0950
	PF II.C CRESH (Inputsub. = 3)	0.0186	0.0176	0.1024	0.0910
	PF III Translog	0.0130	0.0119	0.0960	0.0894
StoNED MoM	PF I Cobb-Douglas (IRS)	0.0676	0.0734	0.0906	0.0980
	PF I.B Cobb-Douglas (CRS)	0.0730	0.0693	0.1139	0.0966
	PF I.C Cobb-Douglas (DRS)	0.0608	0.0756	0.0941	0.0968
	PF II CRESH (Inputsub. = 0.33)	0.0634	0.0674	0.1029	0.1021
	PF II.B CRESH (Inputsub. = 1.33)	0.0597	0.0678	0.1019	0.1036
	PF II.C CRESH (Inputsub. = 3)	0.0749	0.0735	0.1005	0.1009
	PF III Translog	0.0740	0.0799	0.1011	0.0900
StoNED PL	PF I Cobb-Douglas (IRS)	0.0466	0.0378	0.0809	0.0749
	PF I.B Cobb-Douglas (CRS)	0.0454	0.0361	0.0883	0.0778
	PF I.C Cobb-Douglas (DRS)	0.0443	0.0369	0.0788	0.0761
	PF II CRESH (Inputsub. = 0.33)	0.0467	0.0368	0.0814	0.0765
	PF II.B CRESH (Inputsub. = 1.33)	0.0471	0.0366	0.0831	0.0748
	PF II.C CRESH (Inputsub. = 3)	0.0442	0.0355	0.0819	0.0795
	PF III Translog	0.0509	0.0414	0.0862	0.0776

Table 11: **Variation of the functional form of the production function. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: NO; *Production function*: See Table 10; Collinearity: 0; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z): m= 2.

For the purpose of measuring the influence of returns to scale, we compare the results of PF I, I.B and I.C, where PF I has increasing returns to scale of 1.2, PF I.2 has constant returns to scale and PF I.3 has decreasing returns to scale of 0.8. The results in Table 11 suggest that there is no significant influence on the methods, but the performance can be affected in specific settings and the direction is ambiguous. For instance, StoNED MoM is affected in the NTS=1 and 50 DMUs scenario. This is one of the few settings in which the performance of SFA MoM and StoNED MoM diverge considerably. SFA ML and StoNED PL are not noticeably affected in any scenario.

In order to measure the influence of elasticity of substitution, we use three CRESH (II, II.B and II.C) production functions. The respective functions have an elasticity of substitution of 0.33 (PF II), 1.33 (PF II.B) and 3 (PF II.C). The results suggest that the elasticity of substitution only has an impact on SFA MoM, SFA ML and StoNED MoM in the scenario without noise. For the SFA, we assume a Cobb-Douglas production function which has an elasticity of substitution of one. Presumably, this is the reason why SFA performs considerably better, especially SFA ML, when the elasticity of substitution is close to one in the scenario without noise.

Finally, we compare all the results in Table 11 to analyze the effect of the functional form. Additionally to the six production functions described above, we consider our standard translog production function (PF III). It is surprising that DEA, as a nonparametric method, is affected, while the SFA, which is misspecified in some settings, is not affected in most of the settings. DEA performance deteriorates when the data are generated by the translog function. In contrast, our results confirm that the semi-parametric StoNED PL is more "successful", as the underlying production function has no influence on its performance.

However, the comparison is based on the simple two-input one-output case. The use of more than two inputs could affect the results on the impact of the functional form. Hence, we analyze the influence of the number of inputs in the following section.

4.4.2 Number of Inputs

The number of inputs could affect the performance of a given method, because the estimation of the production function is more challenging with an increasing number of inputs. Our first step is to vary the number of inputs of the Cobb-Douglas production function (PF I) and keep the scale elasticity constant, i.e. $\sum_i^m \beta_i = 1.2$.

The results in Table 12 show that the performance of DEA and StoNED PL are influenced particularly by variations in the number of inputs. The effect on DEA is once again diametrical. In the settings without noise, the performance deteriorates with an increasing number of inputs, because the overestimation of DEA increases, i.e. the MD increases (see Table 37 in the appendix). The opposite is true for the noisy scenarios. This can also be explained by the MD, because DEA substantially underestimates the efficiency in the scenario with noise and therefore the "upward shift" caused by the "dimensionality effect" has a positive impact on average performance.

The positive interaction between the number of inputs and MAD is also observable for the semi-parametric StoNED and is most pronounced for the change from three to four inputs. In order to understand the escalating performance deterioration of StoNED, it is helpful to take a look at MD. The MD indicates that the overestimation of StoNED PL increases constantly with an increasing number of inputs. Furthermore, the mean rank correlation of StoNED decreases dramatically (see Table 39). The analysis demonstrates that in particular, the consideration of four inputs exerts a crucial impact on the performance of nonparametric and semi-parametric methods, whereas the parametric methods are less affected.

Method	NTS	0		1	
	DMU	50	100	50	100
DEA	PF I.1 (1 Input)	0.0139	0.0108	0.1716	0.2087
	PF I.2 (2 Inputs)	0.0314	0.0206	0.1384	0.1729
	PF I.3 (3 Inputs)	0.0547	0.0406	0.1234	0.1409
	PF I.4 (4 Inputs)	0.0704	0.0588	0.1065	0.1231
SFA MoM	PF I.1 (1 Input)	0.0585	0.0679	0.1081	0.1103
	PF I.2 (2 Inputs)	0.0688	0.0791	0.0917	0.0987
	PF I.3 (3 Inputs)	0.0623	0.0674	0.0908	0.0943
	PF I.4 (4 Inputs)	0.0606	0.0598	0.0962	0.1035
SFA ML	PF I.1 (1 Input)	0.0066	0.0036	0.1012	0.0973
	PF I.2 (2 Inputs)	0.0101	0.0053	0.0949	0.0907
	PF I.3 (3 Inputs)	0.0138	0.0061	0.1050	0.0939
	PF I.4 (4 Inputs)	0.0150	0.0082	0.1052	0.0958
StoNED MoM	PF I.1 (1 Input)	0.0619	0.0686	0.1095	0.1107
	PF I.2 (2 Inputs)	0.0676	0.0734	0.0906	0.0980
	PF I.3 (3 Inputs)	0.0635	0.0674	0.0919	0.0940
	PF I.4 (4 Inputs)	0.0922	0.0854	0.1019	0.1046
StoNED PL	PF I.1 (1 Input)	0.0313	0.0272	0.0803	0.0755
	PF I.2 (2 Inputs)	0.0466	0.0378	0.0809	0.0749
	PF I.3 (3 Inputs)	0.0597	0.0517	0.0877	0.0796
	PF I.4 (4 Inputs)	0.1161	0.1053	0.1219	0.1140

Table 12: **Variation of the number of inputs. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: NO; *Production function*: PF I (Cobb Douglas with increasing returns to scale); Collinearity: 0; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z): m= 1, 2, 3 and 4.

Our second step is to focus on the four input case, but to consider more functional forms to evaluate if the functional form, in conjunction with a higher number of inputs, has an influence on the method performance. Therefore, we add a Cobb Douglas production function with decreasing returns to scale (PF I.C.4), as well as a CRESH (PF II.4) and a translog (PF III.4) production function. This is of particular interest for the parametric methods, because it can be expected that misspecification is more serious if a flexible functional form, such as translog, is used to generate the data in a multiple-input case. The parametrization for the production functions can be found in Table 13.

Nr	PF (F(x))	Description	Parametrization
I.4	$\sum_{i=1}^{m} \beta_i \cdot ln(z_{i,j})$	Cobb-Douglas, IRS	$\beta_i = 0.3$ for i=1,...,4.
I.C.4	$\sum_{i=1}^{m} \beta_i \cdot ln(z_{i,j})$	Cobb-Douglas, DRS	$\beta_i = 0.2$ for i=1,...,4.
II.4	$\ln([\sum_{i=1}^{m} \alpha_i \cdot z_{i,j}^{-\rho_i}]^{-\delta/\rho})$	CRESH	δ=1, $\alpha_i = 0.25$, ρ=ρ_i=2 for i=1,...,4
III.4	$\beta_0 + \sum_{i=1}^{m} \beta_i \cdot ln(z_{i,j}) + 0.5 \cdot \sum_{i=1}^{m} \sum_{f=1}^{m} \beta_{i,f} \cdot ln(z_{i,j}) \cdot ln(z_{i,j})$	Translog	β_0=1, $\beta_i = 0.15$, $\beta_{i,f} = 0.025$ for i,f=1,...4

Table 13: **Parametrization of the production functions (Four inputs).**

The results in Table 14 confirm that the misspecification of the functional form can exert a negative influence on the performance of SFA ML, for example, in the case of a CRESH production function and the scenario without noise. However, in the scenario without noise, SFA ML is considerably better than the semi-parametric methods, regardless of which production function is used. Considering the noise scenario, the performance of all methods becomes quite similar, but StoNED PL is still the weakest method especially when the number of DMU is small. In summary, as also stated by Kuosmanen (2008), the flexibility of the semi-parametric approach does have a price. The performance, in particular of StoNED PL, deteriorates when more explanatory variables are considered, keeping the number of DMUs constant.

Method	NTS	0		1	
	DMU	50	100	50	100
DEA	PF I.4 Cobb-Douglas (IRS)	0.0704	0.0588	0.1065	0.1231
	PF I.C.4 Cobb-Douglas (DRS)	0.0714	0.0523	0.1091	0.1225
	PF II.4 CRESH	0.0739	0.0618	0.1018	0.1215
	PF III.4 Translog	0.0668	0.0518	0.1112	0.1348
SFA MoM	PF I.4 Cobb-Douglas (IRS)	0.0606	0.0598	0.0962	0.1035
	PF I.C.4 Cobb-Douglas (DRS)	0.0747	0.0726	0.0906	0.1002
	PF II.4 CRESH	0.0618	0.0745	0.1000	0.1005
	PF III.4 Translog	0.0663	0.0800	0.0925	0.0951
SFA ML	PF I.4 Cobb-Douglas (IRS)	0.0150	0.0082	0.1052	0.0958
	PF I.C.4 Cobb-Douglas (DRS)	0.0171	0.0074	0.1015	0.0995
	PF II.4 CRESH	0.0368	0.0357	0.1105	0.0949
	PF III.4 Translog	0.0175	0.0094	0.1023	0.0889
StoNED MoM	PF I.4 Cobb-Douglas (IRS)	0.0922	0.0854	0.1019	0.1046
	PF I.C.4 Cobb-Douglas (DRS)	0.0944	0.0866	0.0951	0.1017
	PF II.4 CRESH	0.0867	0.0852	0.0967	0.1007
	PF III.4 Translog	0.0852	0.0872	0.0973	0.0969
StoNED PL	PF I.4 Cobb-Douglas (IRS)	0.1161	0.1053	0.1219	0.1140
	PF I.C.4 Cobb-Douglas (DRS)	0.1119	0.0934	0.1134	0.1095
	PF II.4 CRESH	0.1122	0.1010	0.1125	0.1072
	PF III.4 Translog	0.1130	0.1009	0.1119	0.1052

Table 14: **Variation of the functional form of the production function (Four inputs). Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: NO; *Production function*: See Table 13; Collinearity: 0; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z): m= 4.

4.4.3 Omitted Variables

In real-world applications, one of the most important problems is the consideration of the "true" production process. It is not only necessary to replicate the "true" functional form of the production function, but also to include all relevant inputs and outputs. The determination of the relevant inputs and outputs can be challenging. Good examples for this problem are the banking and the electricity sector. Jamasb and Pollitt (2001) present an overview of 20 empirical efficiency studies for electricity distribution utilities and the frequency of the different in- and outputs that are used within the studies. Interestingly, the variables differ considerably, although all of them in principle analyze the same context. Furthermore, some variables are even used as "inputs" as well as "outputs," depending on the study. For the banking sector, for instance, Bauer et al (1998) state that there is considerable variation within the literature and that many studies use other bank outputs, inputs, bank characteristics and environmental factors.

Method	NTS	0		1		*0*		*1*	
	DMU	50	100	50	100	*50*	*100*	*50*	*100*
DEA	PF I.4 Cobb-Douglas (IRS)	0.0644	0.0648	0.1290	0.1578	*0.0704*	*0.0588*	*0.1065*	*0.1231*
	PF I.C.4 Cobb-Douglas (DRS)	0.0537	0.0486	0.1325	0.1504	*0.0714*	*0.0523*	*0.1091*	*0.1225*
	PF II.4 CRESH	0.0662	0.0598	0.1279	0.1622	*0.0739*	*0.0618*	*0.1018*	*0.1215*
	PF III.4 Translog	0.0798	0.0796	0.1444	0.1731	*0.0668*	*0.0518*	*0.1112*	*0.1348*
SFA MoM	PF I.4 Cobb-Douglas (IRS)	0.0857	0.0865	0.0983	0.1033	*0.0606*	*0.0598*	*0.0962*	*0.1035*
	PF I.C.4 Cobb-Douglas (DRS)	0.0767	0.0766	0.0977	0.0966	*0.0747*	*0.0726*	*0.0906*	*0.1002*
	PF II.4 CRESH	0.0811	0.0876	0.1078	0.1016	*0.0618*	*0.0745*	*0.1000*	*0.1005*
	PF III.4 Translog	0.0896	0.0878	0.1101	0.1092	*0.0663*	*0.0800*	*0.0925*	*0.0951*
SFA ML	PF I.4 Cobb-Douglas (IRS)	0.0676	0.0698	0.1062	0.1023	*0.0150*	*0.0082*	*0.1052*	*0.0958*
	PF I.C.4 Cobb-Douglas (DRS)	0.0505	0.0487	0.1072	0.0933	*0.0171*	*0.0074*	*0.1015*	*0.0995*
	PF II.4 CRESH	0.0678	0.0683	0.1189	0.0955	*0.0368*	*0.0357*	*0.1105*	*0.0949*
	PF III.4 Translog	0.0888	0.0827	0.1231	0.1057	*0.0175*	*0.0094*	*0.1023*	*0.0889*
StoNED MoM	PF I.4 Cobb-Douglas (IRS)	0.0820	0.0841	0.0984	0.1036	*0.0922*	*0.0854*	*0.1019*	*0.1046*
	PF I.C.4 Cobb-Douglas (DRS)	0.0731	0.0798	0.0959	0.0975	*0.0944*	*0.0866*	*0.0951*	*0.1017*
	PF II.4 CRESH	0.0846	0.0855	0.1088	0.1038	*0.0867*	*0.0852*	*0.0967*	*0.1007*
	PF III.4 Translog	0.0904	0.0911	0.1093	0.1131	*0.0852*	*0.0872*	*0.0973*	*0.0969*
StoNED PL	PF I.4 Cobb-Douglas (IRS)	0.0681	0.0680	0.0865	0.0823	*0.1161*	*0.1053*	*0.1219*	*0.1140*
	PF I.C.4 Cobb-Douglas (DRS)	0.0632	0.0645	0.0851	0.0829	*0.1119*	*0.0934*	*0.1134*	*0.1095*
	PF II.4 CRESH	0.0704	0.0648	0.0857	0.0798	*0.1122*	*0.1010*	*0.1125*	*0.1072*
	PF III.4 Translog	0.0723	0.0765	0.0918	0.0876	*0.1130*	*0.1009*	*0.1119*	*0.1052*

Table 15: **Omitted Variables. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: NO; *Production function*: See Table 13; Collinearity: 0; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z) within the DGP (efficiency estimation): m= 4 (3).

In our analysis we account for this problem by deliberately omitting variables while estimating the efficiency. We assume a data generating process with four inputs as in the previous section, but consider only three inputs in the application of the methods. This enables us to identify how the methods performance is influenced by this misspecification error.

Table 15 shows the results. These results should be compared with the results in Table 14, which is the point of reference. To simplify the comparison, the results of Table 14 are denoted in italic numbers in Table 15. If only the MAD is considered as performance criterion, the effect of omitted variables seems to be complex and ambiguous. Here, it is again worthwhile to consider the MD because it shows that the general effect is for most methods unambiguous and as one might expect: The omission of an input tends to result in an underestimation of the "true" efficiency, i.e. the MD is lower than in the case when all relevant inputs are considered (see Table 43 in the appendix). This result is true for almost all settings and methods, except for the SFA MoM. DEA and StoNED seem to be particularly affected by this "omission effect". The underestimation caused by the "omission effect" leads to ambiguous effects for the MAD, our main performance criterion, due to the point of reference. The point of reference are the results of Table 14: settings with four inputs and all inputs are considered within the estimation. Therefore, it is important to consider the associated under- vs. overestimation constellation in these settings.

As explained in Section 4.4.2, the consideration of four inputs causes a "'dimensionality effect"' that induces an overestimation, whereas the "omission effect" induces an underestimation so that the two effects tend to cancel each other out. This is the reason why for some of the methods and settings, against expectations, the performance improves when omitting variables. This is especially of relevance for the StoNED PL; its MAD decreases considerably. However, the MD shows that it still overestimates the efficiency on average, but the bias is much lower than before. In contrast, the MAD of SFA MoM and SFA ML increases for most of the settings, especially in the settings without noise. For the DEA it is conspicuous that the MAD increases in the scenario with noise. Here, the underestimation of the efficiency increases considerably.

In brief, this analysis demonstrates once again that the inspection of only one performance criterion, in our case the MAD, is in certain circumstances insufficient. The examination of the MAD does not show the underlying effects because two effects counteract. By looking at the MD, it is obvious that the omission of a relevant variable leads to an underestimation of the efficiency for most of the methods. To analyze the "omission effect" in more detail, we conduct the same experiment with two inputs, but consider only one input for the estimation of the efficiency. Thereby, we focus on the "omission effect" and avoid the "dimensionality problem".

Method	NTS	0		1		*0*		*1*	
	DMU	50	100	50	100	*50*	*100*	*50*	*100*
DEA	PF I Cobb-Douglas (IRS)	0.1419	0.1527	0.2380	0.2675	*0.0314*	*0.0206*	*0.1384*	*0.1729*
	PF II CRESH	0.1156	0.1198	0.2244	0.2548	*0.0357*	*0.0232*	*0.1335*	*0.1648*
	PF III Translog	0.1855	0.2001	0.2647	0.2923	*0.0373*	*0.0326*	*0.1544*	*0.1960*
SFA MoM	PF I Cobb-Douglas (IRS)	0.1100	0.1046	0.1168	0.1123	*0.0688*	*0.0791*	*0.0917*	*0.0987*
	PF II CRESH	0.1007	0.1049	0.1214	0.1208	*0.0697*	*0.0737*	*0.1032*	*0.1034*
	PF III Translog	0.1142	0.1245	0.1252	0.1210	*0.0755*	*0.0811*	*0.0976*	*0.0896*
SFA ML	PF I Cobb-Douglas (IRS)	0.1237	0.1267	0.1260	0.1205	*0.0101*	*0.0053*	*0.0949*	*0.0907*
	PF II CRESH	0.1128	0.1163	0.1337	0.1266	*0.0344*	*0.0332*	*0.1006*	*0.0912*
	PF III Translog	0.1457	0.1621	0.1431	0.1393	*0.0130*	*0.0119*	*0.0960*	*0.0894*
StoNED MoM	PF I Cobb-Douglas (IRS)	0.1095	0.1038	0.1180	0.1118	*0.0676*	*0.0734*	*0.0906*	*0.0980*
	PF II CRESH	0.1002	0.1028	0.1199	0.1195	*0.0634*	*0.0674*	*0.1029*	*0.1021*
	PF III Translog	0.1143	0.1250	0.1264	0.1218	*0.0740*	*0.0799*	*0.1011*	*0.0900*
StoNED PL	PF I Cobb-Douglas (IRS)	0.0921	0.0830	0.0971	0.0871	*0.0466*	*0.0378*	*0.0809*	*0.0749*
	PF II CRESH	0.0867	0.0795	0.0971	0.0921	*0.0467*	*0.0368*	*0.0814*	*0.0765*
	PF III Translog	0.0962	0.1020	0.1036	0.0990	*0.0509*	*0.0414*	*0.0862*	*0.0776*

Table 16: **Omitted Variables. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: NO; *Production function*: PF I (Cobb Douglas with increasing returns to scale), PF II (CRESH), PF III (Translog); Collinearity: 0; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z) within the DGP (efficiency estimation): m= 2 (1).

The results in Table 16 show a clear effect on the MAD. All methods perform considerably worse when the second input is not considered in the estimation. We can see further that the relative performance of the methods changes. The DEA is the method that is most affected. The reason is that DEA does not consider random noise, which partly covers the "omission effect" for the stochastic methods. Furthermore, the technique to estimate the underlying production function seems to influence the performance of the stochastic methods. The SFA ML is more negatively affected than the MoM methods and the StoNED PL. For both scenarios (with and without noise) the order of the methods is from best to worst: StoNED PL, StoNED MoM, SFA MoM, SFA ML and DEA. The reason for the performance deterioration can be traced back to the underestimation caused by the "omission effect". The MD for all methods decreases, but for SFA and especially DEA the effect is stronger (see Table 46). In addition, the mean rank correlation (MRC) for all methods decreases dramatically, see Table 48.

4.4.4 Collinearity

A further factor considered in studies comparing efficiency methods is the collinearity between inputs (see, for example, Jensen (2005)). Andor and Hesse (2011) assumed that correlation is between 0 and 0.9. However, it might be more interesting to consider cases with an even higher correlation between the inputs. Therefore, we only use extreme values for the collinearity, namely $\rho_{coll}(z_1,z_2)$=0.0, 0.9 and 0.99.

The results suggest that DEA is the only method which is considerably influenced by collinearity (see Table 17). The reason is that increasing collinearity leads to a greater underestimation of DEA. As a result, it is, once again, diametrically affected. For the scenario without noise (except PF III), it is positively affected, while the opposite applies to the noise scenario. SFA MoM and StoNED MoM seem to be unaffected. Also, SFA ML is mainly unaffected, but in the scenario without noise, the performance improves with increasing collinearity, when the underlying production function is PF II. StoNED PL exhibits a performance improvement with increasing collinearity for the scenario without noise. Nevertheless, considering extreme values for the collinearity, we can conclude that the various methods – except DEA – are not influenced substantially. These findings concur with Jensen (2005), who concludes that collinearity has no influence on the performance of SFA ML.

Method	NTS	0						1					
	DMU	50			100			50			100		
	PF	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III
DEA	ρ=0.00	0.0314	0.0357	0.0373	0.0206	0.0232	0.0326	0.1384	0.1335	0.1544	0.1729	0.1648	0.1960
	ρ=0.90	0.0219	0.0255	0.0456	0.0138	0.0160	0.0491	0.1562	0.1469	0.1892	0.1890	0.1760	0.2262
	ρ=0.99	0.0174	0.0214	0.0511	0.0115	0.0126	0.0563	0.1732	0.1617	0.2170	0.1980	0.1835	0.2379
SFA MoM	ρ=0.00	0.0688	0.0697	0.0755	0.0791	0.0737	0.0811	0.0917	0.1032	0.0976	0.0987	0.1034	0.0896
	ρ=0.90	0.0679	0.0626	0.0704	0.0679	0.0693	0.0721	0.1048	0.0901	0.0965	0.0981	0.1050	0.0949
	ρ=0.99	0.0680	0.0643	0.0754	0.0709	0.0753	0.0799	0.0980	0.0972	0.0964	0.1022	0.0927	0.0990
SFA ML	ρ=0.00	0.0101	0.0344	0.0130	0.0053	0.0332	0.0119	0.0949	0.1006	0.0960	0.0907	0.0912	0.0894
	ρ=0.90	0.0084	0.0107	0.0141	0.0048	0.0060	0.0125	0.1086	0.0956	0.1097	0.0939	0.0960	0.0907
	ρ=0.99	0.0093	0.0093	0.0156	0.0050	0.0048	0.0142	0.0979	0.1013	0.0968	0.0931	0.0875	0.0897
StoNED MoM	ρ=0.00	0.0676	0.0634	0.0740	0.0734	0.0674	0.0799	0.0906	0.1029	0.1011	0.0980	0.1021	0.0900
	ρ=0.90	0.0701	0.0615	0.0741	0.0660	0.0688	0.0747	0.1055	0.0916	0.0966	0.0992	0.1053	0.0974
	ρ=0.99	0.0690	0.0641	0.0816	0.0676	0.0734	0.0826	0.0987	0.0982	0.1002	0.1024	0.0932	0.1005
StoNED PL	ρ=0.00	0.0466	0.0467	0.0509	0.0378	0.0368	0.0414	0.0809	0.0814	0.0862	0.0749	0.0765	0.0776
	ρ=0.90	0.0417	0.0382	0.0473	0.0313	0.0302	0.0403	0.0849	0.0798	0.0887	0.0791	0.0765	0.0752
	ρ=0.99	0.0360	0.0326	0.0461	0.0314	0.0299	0.0433	0.0795	0.0782	0.0835	0.0731	0.0778	0.0762

Table 17: **Variation of collinearity between the inputs. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: NO; *Production function*: PF I (Cobb Douglas with increasing returns to scale), PF II (CRESH), PF III (Translog); Collinearity: 0, 0.9, 0.99; Input distribution: $z_j \sim$ U(5,15); Number of inputs(z): m= 2.

4.4.5 Input distribution

Most simulation studies use uniform or normal distributions to generate the inputs. In fact, real-world input distributions are usually different with regard to the standard deviation and skewness of the distribution. For instance, Resti (2000) justifies his use of a skewed input distribution by the fact that there are usually more small and medium-sized companies than large ones and that an unrealistic assumption could influence the performance of the methods. However, in contrast to Resti (2000), we vary the input distribution and are therefore able to evaluate the influence. We use normal, gamma and uniform distributions, which differ regarding the standard deviation and the skewness (see Table 18).

Distribution	Mean	σ	Skewness
$z_{1,2} \sim N(10,1)$	10	1.00	0.00
$z_{1,2} \sim Gamma(100,0.1)$	10	1.00	0.20
$z_{1,2} \sim U(5,10)$	10	2.90	0.00
$z_{1,2} \sim Gamma(10,1)$	10	3.15	0.62

Table 18: **Variation of the input distribution and their respective moments.**

In general, the results suggest that the input distribution can exert an impact on the performance of all methods, but only in specific settings (see Table 19). For SFA MoM and StoNED MoM, it is difficult to identify a systematic pattern. For DEA, the performance deteriorates with an increasing standard deviation in the scenario without noise. This effect is notably significant for the translog function (PF III). For instance, the DEA MAD is more than twice as high than in comparable settings. The same effect, increasing MAD for an increasing standard deviation, is observable for the SFA ML in cases with a high standard deviation in combination with a misspecification of the production function (PF II, III). The analysis of the input distribution supports the supposition of Resti (2000) that the input distribution can have an influence on the performance of the methods, but it depends on the specifications of the other influencing factors and has only a minor impact in comparison to the other influencing factors.

Method	NTS	0						1					
	DMU	50			100			50			100		
	PF	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III
DEA	$z_{1,2} \sim N(10,1)$	0.0212	0.0241	0.0203	0.0142	0.0151	0.0146	0.1687	0.1629	0.1665	0.1952	0.1937	0.2029
	$z_{1,2} \sim G(100,0.1)$	0.0216	0.0229	0.0213	0.0149	0.0158	0.0139	0.1649	0.1783	0.1731	0.2100	0.1927	0.2005
	$z_{1,2} \sim U(1,15)$	0.0314	0.0357	0.0373	0.0206	0.0232	0.0326	0.1384	0.1335	0.1544	0.1729	0.1648	0.1960
	$z_{1,2} \sim G(10,1)$	0.0343	0.0410	0.0437	0.0229	0.0260	0.0504	0.1418	0.1290	0.1628	0.1681	0.1588	0.1987
SFA MoM	$z_{1,2} \sim N(10,1)$	0.0598	0.0635	0.0653	0.0746	0.0712	0.0805	0.1032	0.0936	0.0982	0.0934	0.0982	0.1018
	$z_{1,2} \sim G(100,0.1)$	0.0585	0.0841	0.0726	0.0689	0.0714	0.0770	0.1019	0.1050	0.0994	0.0896	0.0991	0.0998
	$z_{1,2} \sim U(1,15)$	0.0688	0.0697	0.0755	0.0791	0.0737	0.0811	0.0917	0.1032	0.0976	0.0987	0.1034	0.0896
	$z_{1,2} \sim G(10,1)$	0.0593	0.0707	0.0674	0.0725	0.0785	0.0749	0.1078	0.1030	0.0940	0.0961	0.1074	0.1008
SFA ML	$z_{1,2} \sim N(10,1)$	0.0098	0.0121	0.0091	0.0048	0.0069	0.0052	0.0998	0.0989	0.0993	0.0891	0.0911	0.0924
	$z_{1,2} \sim G(100,0.1)$	0.0092	0.0108	0.0094	0.0049	0.0066	0.0044	0.1026	0.1042	0.1005	0.0872	0.0967	0.0933
	$z_{1,2} \sim U(1,15)$	0.0101	0.0344	0.0130	0.0053	0.0332	0.0119	0.0949	0.1006	0.0960	0.0907	0.0912	0.0894
	$z_{1,2} \sim G(10,1)$	0.0099	0.0372	0.0155	0.0060	0.0360	0.0136	0.1001	0.1065	0.0964	0.0867	0.1013	0.0915
StoNED MoM	$z_{1,2} \sim N(10,1)$	0.0608	0.0628	0.0669	0.0743	0.0696	0.0789	0.1018	0.0936	0.0991	0.0948	0.0973	0.1019
	$z_{1,2} \sim G(100,0.1)$	0.0605	0.0806	0.0700	0.0674	0.0710	0.0765	0.1027	0.1047	0.1014	0.0909	0.1000	0.0988
	$z_{1,2} \sim U(1,15)$	0.0676	0.0634	0.0740	0.0734	0.0674	0.0799	0.0906	0.1029	0.1011	0.0980	0.1021	0.0900
	$z_{1,2} \sim G(10,1)$	0.0608	0.0629	0.0696	0.0718	0.0697	0.0756	0.1086	0.1033	0.0959	0.0989	0.1030	0.1033
StoNED PL	$z_{1,2} \sim N(10,1)$	0.0356	0.0332	0.0372	0.0295	0.0298	0.0328	0.0811	0.0806	0.0809	0.0779	0.0769	0.0768
	$z_{1,2} \sim G(100,0.1)$	0.0350	0.0385	0.0374	0.0290	0.0294	0.0300	0.0870	0.0872	0.0788	0.0767	0.0782	0.0784
	$z_{1,2} \sim U(1,15)$	0.0466	0.0467	0.0509	0.0378	0.0368	0.0414	0.0809	0.0814	0.0862	0.0749	0.0765	0.0776
	$z_{1,2} \sim G(10,1)$	0.0437	0.0466	0.0502	0.0348	0.0363	0.0450	0.0878	0.0812	0.0843	0.0725	0.0742	0.0759

Table 19: **Variation of the input distribution. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim \text{Exp}(\mu{=}1/6)$; Heteroscedasticity: NO; *Production function*: PF I (Cobb Douglas with increasing returns to scale), PF II (CRESH), PF III (Translog); Collinearity: 0; Input distribution: See Table 18; Number of inputs(z): m= 2.

4.5 Overview effects

Finally, Table 20 summarizes the main conclusions of the analysis of the influencing factors. To avoid a misleading interpretation of Table 20, it is important to reference the legend. For instance, a "+" for the MAD symbolizes a performance deterioration, whereas a "+" for the MRC signifies a performance improvement. The most important part of our study is without doubt the analysis of the recently introduced StoNED. Hence, we now focus on the influencing factors of StoNED. StoNED MoM generally performs in a very similar manner to SFA MoM. The noise-to-signal ratio, the sample size and the skewness of the inefficiency distribution have a negative impact on it, particularly in the scenario without noise. The comparative advantage of the MOM methods is the ability to handle a heteroscedastic inefficiency term. In short, our results suggest that the StoNED MoM does not seem to constitute a substantial advancement in efficiency estimation, as it behaves very similar to the SFA MoM, without offering any compelling advantages.

However, the StoNED PL seems to constitute progress in efficiency estimation, as it has an important unique comparative advantage. StoNED PL is the best method if a high noise-to-signal ratio is assumed. Furthermore, the performance of StoNED PL is less affected by omitting relevant inputs than the other methods. In contrast, the curse of dimensionality (a larger number of inputs) and scenarios without noise, are weaknesses of the StoNED PL, in comparison to the other methods.

	DEA			SFA MoM			SFA ML			StoNED MoM			StoNED PL		
Influencing factor	**MAD**	**MD**	**Rc**	**MAD**	**MD**	**Rc**	**MAD**	**MD**	**Rc**	**MAD**	**MD**	**Rc**	**MAD**	**MD**	**Rc**
Sample size *(Section 4.2)*	- /+	-	+	+ / o	-	+	-	-	+	+o/o	o	o	o	o	o
Noise-to-signal (NTS) *(Section 4.3.1)*	+	-	-	+	+o	-	+	o	-	+	o	-	+	+o	-
Distribution inefficiency *(Section 4.3.2)*	-/+	-	o	+/o	-	-/o	o	-	-/o	+/o+	-	-/o	-	-	-/o
Heteroscedasticity *(Section 4.3.3)*	+/-	+	=/+	-	+	- / +	=+/-	+	o/+	-	+	-/+	+	+	-/+
Number of Inputs *(Section 4.4.2)*	+/-	+	-/o	o	o	-/o	=+/=o	=+/o	=o/o	+o/o	+	-	+	+	-
Omitted Variables *(Section 4.4.3)*	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-
Collinearity *(Section 4.4.4)*	o/+	-	o	=/o=	o	o	o=/=	o	o	=/o=	o	+/o	-o/=	o	+/o
Distribtuion inputs *(Section 4.4.5)*	+/o	o	-/o	o/=o	o	o	o+/=o	=o/o	o	=o	=o/o	o	o/=	o=/o	o

Table 20: **Overview of influencing factors on methods performance.** Legend: The meaning of the symbols are the following: (+) increasing, (-) decreasing, (o) ambiguous effect and (=) no considerable effect. If the results depend on the noise-to-signal ratio the sign in front of a slash (/) refers to the without noise scenario (NTS=0), whereas the sign after the slash refers to the noise scenario (NTS=1). If there are two symbols, both are valid in specific settings.

4.6 Performance in terms of the estimated production frontier

So far, we have only considered the estimation of the technical efficiency value as a performance measure. All performance criteria – MAD, MSE, MD and MRC – are measured in terms of the ability of the method to estimate the efficiency value TE. From our point of view, this is the most relevant performance benchmark, because the technical efficiency is still used in most efficiency analysis studies and in real-world applications. However, for the stochastic SFA and StoNED, we apply the Battese and Coelli (1988) point estimator, see equations (9) and (10). As mentioned above, this estimator is known to be inconsistent, whereas the SFA and StoNED estimators for the frontier F are unbiased and consistent. Hence, it can make sense to estimate "only" the frontier F and use these estimates. For instance, the regulatory model in Finland, which was based on the JMLS estimator until 2011, changed so that the StoNED estimator used since 2012 is based on the frontier F. The regulatory model changed because the estimate of the frontier F should be a more reliable benchmark than the firm-specific TE estimates (cf. Kuosmanen (2012a)). However, the other European electricity regulation systems still rely on the firm-specific efficiency estimator. Nevertheless, we discuss in the following the performance of the methods in terms of the estimation of frontier F for our standard settings.

We thus redefine the mean absolute deviation (MAD) as follows:

$$MAD = \frac{1}{nR} \sum_{r=1}^{R} \sum_{j=1}^{n} \sum_{i=1}^{m} \left| \hat{F}_r(z_{i,j}) - F(z_{i,j}) \right|, \tag{25}$$

where $\hat{F}_r$ and F_r are the estimated production value and the true production value, respectively, and $r = 1, ..., R$ is the number of replications of a setting.

	NTS	0						1					
	DMU	50			100			50			100		
Method	PF	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III	PF I	PF II	PF III
DEA		0.5538	0.3656	1.3369	0.3648	0.2411	1.1595	2.5996	1.4510	5.7619	3.8680	2.1797	8.6428
SFA MOM		2.4523	1.4379	4.7802	2.3887	1.4489	4.6935	2.2916	1.5304	5.0062	2.3803	1.4544	4.7853
SFA ML		0.1815	0.3669	0.4625	0.0948	0.3539	0.4178	1.3312	0.9010	2.9745	1.3508	0.7931	2.3843
STONED		2.4505	1.4080	4.8189	2.3888	1.4269	4.7500	2.2885	1.4959	5.0467	2.3803	1.4290	4.8400

Table 21: **Variation of performance measure - $F(x_{i,j})$ instead of TE. Performance criterion: Mean absolute deviation (MAD).** DGP: DMU= 50, 100; *Error term*: Noise-to-signal ratio (NTS): 0 and 1; $u_j \sim$ Exp(μ=1/6); Heteroscedasticity: NO; *Production function*: PF I (Cobb Douglas with increasing returns to scale), PF II (CRESH), PF III (Translog); Collinearity: 0; $z_j \sim$ U(5,15); Number of inputs(z): m=2.

Table 21 shows the mean absolute deviation between the estimated production function value and the "true" value. The results show that SFA ML is still the best method, independent of the underlying setting. Again, DEA performs very well in the scenario without noise, whereas it is the worst method in the scenario with noise. StoNED and SFA MoM obtain relatively similar results. The results support our supposition that the strength of the StoNED PL relies more in the ability to distinguish between noise and inefficiency by means of the pseudolikelihood estimator than in the estimation of the production function.

5 Conclusions

In this simulation study, we compared the StoNED method, recently introduced by Kuosmanen and Kortelainen (2010), with the two most popular estimation methods, or rather the two "oldies" DEA and SFA. Our research objective was a systematic comparison of the three methods and the two different estimation techniques (method of moments and likelihood), using cross sectional data. Accordingly, we analyzed the performance of DEA, SFA MoM, SFA ML, StoNED MoM and StoNED PL in a Monte Carlo simulation. By using 200 different settings, we identified factors influencing the performance of the particular method and derive recommendations for practical applications.

The main findings can be summarized as follows. The likelihood estimation techniques, and especially the SFA ML, perform best in our study. The StoNED PL is a serious competitor for SFA ML and has its comparative advantage in an increasing noise-to-signal ratio. Additionally, the performance of StoNED PL is less affected by omitting relevant inputs than the other methods. Furthermore, our analysis reveals a specific characteristic of the StoNED PL. While all other methods underestimate efficiency, StoNED PL is the only method that overestimates on average. This finding can partly explain the performance of StoNED and could be useful to policy makers. For instance, in the German incentive regulation of electricity grid operators, the best-of-two-method is applied, meaning that the highest of the estimates of DEA and SFA is used as the efficiency value, so as to avoid underestimating the efficiency of grid operators. The relatively good performance of StoNED PL, in conjunction with a bias to overestimate the efficiency, seems a good argument for applying StoNED PL for this purpose. A disadvantage for the application in the real-world is the diminishing performance of StoNED for an increasing number of inputs. Nevertheless, an evaluation of the methods depends on the specific performance criterion. While StoNED PL and SFA ML achieve similar performance with regard to MAD and MSE, the consideration of rank correlation leads to a different conclusion. As StoNED is the poorest method under the latter performance criterion in our study, the ranking accuracy seems to constitute a weakness of StoNED. In general, our results indicate that the strength of the StoNED PL relies more in the ability to distinguish between noise and inefficiency by means of the pseudolikelihood estimator than in the estimation of the production function.

Using the method of moments as estimation technique, the performance of SFA and StoNED are generally similar. The switch between SFA MoM and StoNED MoM, namely the methodology change of the production function estimation from OLS to CNLS, does not seem to be promising. In particular, StoNED has the disadvantage of a lower rank correlation. However, the MoM estimation technique is advisable when a heteroscedastic inefficiency term has to be considered. To cope with the deterministic of DEA, we also considered a nondiscriminatory subsample of 94 settings without noise. Indeed, in this subsample, DEA and SFA perform best. Summarizing, while in scenarios without noise, the "battle" is still between the "oldies", in noisy scenarios, the nonparametric StoNED PL is a promising alternative to the SFA ML.

Our conclusions have, like every Monte Carlo simulation, some limitations, because they are only valid under the considered assumptions. The results show that the relative advantageousness of a method critically depends on the underlying assumptions. As a result, we would like to advice for practical applications to conduct a Monte Carlo simulation under the concrete real-world conditions, before deciding for an estimation method. For instance, the number of DMUs, the input distribution as well as the number of inputs are observable, whereas one has to define adequate assumptions about, for example, the distribution of the inefficiency as well as the noise term. Of course, the conduction of a Monte Carlo simulation with all methods is laborious. However, at least for regulator who derives financial objectives for regulated firms from efficiency benchmarks, the effort should be worthwhile. For practitioners who cannot conduct their own MC study, theoretical MC studies which consider a wide variety of assumptions can serve as a guideline. Accordingly, our study can be seen as a first step in indicating a range of specific situations in which one of the five considered estimation methods proves superior, but further research is needed.

This study focused on the single-input multiple-output case. An MC study considering the multiple-input multiple-output case could be of interest, as policy makers in the real world often face this problem (cf. Perelman and Santin (2009)). Furthermore, this is one of the main advantages of DEA. However, for this purpose, a multiple-output model for StoNED has to be developed. Further research objectives for StoNED can be found in Kuosmanen and Kortelainen (2010). Finally, future research should consider how StoNED performs in comparison to other approaches, which combine the advantages of parametric and nonparametric methods. For instance, Badunenko et al (2011) compare the nonparametric kernel SFA estimator of Fan et al (1996) to the nonparametric bias-corrected DEA estimator of Kneip et al (2008). A comparison of these methods with StoNED would surely be worth conducting.

6 Appendix

The appendix can be found online at: http://www.rwi-essen.de/media/content/pages/publikationen/ruhr-economic-papers/REP_12_394_appendix.pdf.

References

Adler N, Yazhemsky E (2010) Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction. European Journal of Operational Research 202(1):273–284

Afriat SN (1972) Efficiency estimation of production functions. International Economic Review 13(3):568–598

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production models. Journal of Econometrics 6:21–37

Andor M, Hesse F (2011) A Monte Carlo simulation comparing DEA, SFA and two simple approaches to combine efficiency estimates. CAWM Discussion Papers 51, Center of Applied Economic Research Münster (CAWM), University of Münster

Badunenko O, Henderson DJ, Kumbhakar SC (2011) When, where and how to perform efficiency estimation. MPRA Working Paper, forthcoming in Journal of the Royal Statistical Society, Series A

Banker RD, Charnes A, Cooper WW (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science 30(9):1078–1092

Banker RD, Gadh VM, Gorr WL (1993) A Monte Carlo comparison of two production frontier estimation methods: Corrected ordinary least squares and data envelopment analysis. European Journal of Operational Research 67(3):332–343

Banker RD, Cooper WW, Grifell-Tajte E, Pastor JT, Wilson PW, Ley E, Lovell CAK (1994) Validation and generalization of DEA and its uses. TOP 2(2):249–314

Banker RD, Cooper WW, Seiford LM, Thrall RM, Zhu J (2004) Returns to scale in different DEA models. European Journal of Operational Research 154(2):345–362

Battese GE, Coelli TJ (1988) Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics 38(3):387–399

Bauer P, Berger A, Ferrier G, Humphrey D (1998) Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods. Journal of Economics and Business 50(2):85–114

Bogetoft P, Otto L (2011) Benchmarking with DEA, SFA, and R. Springer, Berlin, Heidelberg

Caudill SB, Ford JM (1993) Biases in frontier estimation due to heteroscedasticity. Economics Letters 41(1):17–20

Caudill SB, Ford JM, Gropper DM (1995) Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity. Journal of Business & Economic Statistics 13:105–111

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

Coelli TJ (1995) Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis. Journal of Productivity Analysis 6(4):247–268

Coelli TJ, Rao DSP, O' Donnell CJ, Battese GE (2005) An Introduction to Efficiency and Productivity Analysis. Springer, Berlin, Heidelberg

Cook WD, Seiford LM (2009) Data envelopment analysis (DEA)-Thirty years on. European Journal of Operational Research 192(1):1–17

Cordero JM, Pedraja F, Santin D (2009) Alternative approaches to include exogenous variables in DEA measures: A comparison using Monte Carlo. Computers & Operations Research 36(10):2699–2706

Fan Y, Li Q, Weersink A (1996) Semiparametric Estimation of Stochastic Production Frontier Models. Journal of Business & Economic Statistics 14(4):460–468

Farrell MJ (1957) The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A (General) 120(3):253–290

Gong B, Sickles RC (1992) Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. Journal of Econometrics 51:259–284

Greene WH (2008) The Econometric Approach to Efficiency Analysis. In: Fried H, Lovell CAK, Schmidt S (eds) The Measurement of Productive Efficiency and Productivity Growth, Oxford University Press, New York, pp. 92-250

Hadri K (1999) Estimation of a doubly heteroscedastic stochastic frontier cost function. Journal of Business & Economic Statistics 17(3):359–363

Hadri K, Guermat C, Whittaker J (2003) Estimation of technical inefficiency effects using panel data and doubly heteroscedastic stochastic production frontiers. Empirical Economics 28(1):203–222

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

Jamasb T, Pollitt MG (2001) Benchmarking and regulation: international electricity experience. Utilities Policy 9:107–130

Jensen U (2005) Misspecification preferred: The sensitivity of inefficiency rankings. Journal of Productivity Analysis 23:223–244

Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics 19(3):233–238

Kneip A, Simar L (1996) A general framework for frontier estimation with panel data. Journal of Productivity Analysis 7:187–212

Kneip A, Simar L, Wilson PW (2008) Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models. Econometric Theory 24(6):1663–1697

Kumbhakar SC (1997) Efficiency estimation with heteroscedasticity in a panel data model. Applied Economics 29(3):379–386

Kumbhakar SC, Lovell CAK (2003) Stochastic frontier analysis. Cambridge University Press, Cambridge

Kumbhakar SC, Park BU, Simar L, Tsionas EG (2007) Nonparametric stochastic frontiers: A local maximum likelihood approach. Journal of Econometrics 137:1–27

Kuosmanen T (2008) Representation theorem for convex nonparametric least squares. The Econometrics Journal 11(2):308–325

Kuosmanen T (2012a) Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. Energy Economics p doi: 10.1016/j.eneco.2012.03.005

Kuosmanen T (2012b) Web site: StoNED Stochastic Nonparametric Envelopment of Data: http://www.nomepre.net/index.php/computations.

Kuosmanen T, Kortelainen M (2010) Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. Journal of Productivity Analysis :1–18

Meeusen W, van den Broeck J (1977) Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. International Economic Review 18(2):435–444

Mortimer D (2002) Competing Methods for Effciency Measurement: A Systematic Review of Direct DEA vs SFA/DFA Comparisons., Centre for Health Program Evaluation (CHPE), Working Paper 136

Olson JA, Schmidt P, Waldman DM (1980) A Monte Carlo Study of Estimators of Stochastic Frontier Production Functions. Journal of Econometrics 13:67–82

Ondrich J, Ruggiero J (2001) Efficiency measurement in the stochastic frontier model. European Journal of Operational Research 129(2):434–442

Perelman S, Santin D (2009) How to generate regularly behaved production data? A Monte Carlo experimentation on DEA scale efficiency meassurement. European Journal of Operational Research 199(1):303–310

Resti A (2000) Efficiency measurement for multi-product industries: A comparison of classic and recent techniques based on simulated data. European Journal of Operational Research 121(3):559–578

Ruggiero J (1999) Efficiency estimation and error decomposition in the stochastic frontier model: A Monte Carlo analysis. European Journal of Operational Research 115(3):555–563

Simar L, Zelenyuk V (2011) Stochastic FDH/DEA estimatiors for frontier analysis. Journal of Productivity Analysis 36(1):1–20

Winsten CB (1957) Discussion on Mr. Farrells paper. Journal of the Royal Statistical Society, Series A (General) 120(3):282–284

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