Benchmarking South African water utilities: A comparison of estimates from three methods

Genius Murwirapachena[1*], Richard Mulwa[2], Johane Dikgang[3] and Jugal Mahabir[4]

ABSTRACT

This paper compares the data envelopment analysis (DEA), stochastic frontier analysis (SFA), and stochastic non-parametric envelopment of data (StoNED) methods in the context of regulating water utilities. We estimate technical efficiency based on cross-sectional data from 102 South African water utilities in the period 2013/14. We compare the impact of methodological choices on the efficiency estimates. For StoNED, we compare two different estimation techniques (method of moments or MM and pseudolikelihood or PSL). Overall, we analyse the performance of StoNED MM, StoNED PSL, SFA and DEA. We also use the naïve method of averaging (NMA) and compare its results to those generated by the other methods. Results show that the choice of method has an impact on both efficiency estimates and monetary cost reduction targets. When water sector regulators in the developing countries intend to benchmark utilities based on each of these methods, heterogeneity and the operating environments of utilities should be considered before deciding on the estimation method.

Keywords: DEA, SFA, StoNED, water utilities

Classification-JEL: D24, H41, P28, Q25

---

[1] **Corresponding Author:** School of Economics and Development Studies, Nelson Mandela University, Port Elizabeth, South Africa. Email: murwiragenius@gmail.com

[*] Department of Public Management and Economics, Durban University of Technology, 41 ML Sultan Road, Berea, Durban, 4001, South Africa.

[2] Centre for Advanced Studies in Environmental Law and Policy (CASELAP), University of Nairobi, Nairobi, Kenya. Email: richard.mulwa@gmail.com

[3] Public and Environmental Economics Research Centre (PEERC), School of Economics, University of Johannesburg, Johannesburg, South Africa. Email: jdikgang@uj.ac.za / jmahabir@uj.ac.za

1. Introduction

Climate change has brought renewed and increasing attention to the productivity and efficiency of the water sector. This has stimulated interest, which has manifested itself in the increased application of statistical tools to measure productivity and efficiency in the water sector. Policymakers in developed countries are already making use of statistical analyses of water systems for determining productivity and efficiency. The best-known examples are Switzerland, the United Kingdom (UK), the United States of America (USA), Germany, the Netherlands and Italy (see Baranzini et al. 2008). One of the most commonly used statistical tools for determining productivity and efficiency is the frontier model. There are generally two types of techniques used in frontier analysis, namely, non-parametric approaches such as data envelopment analysis or DEA (see Charnes et al. 1978; Farrell 1957); and parametric approaches, such as the stochastic frontier analysis or SFA (see Aigner et al. 1977; Meeusen and Vandenbroeck 1977). DEA and SFA are commonly used by researchers, practitioners and policymakers when carrying out frontier efficiency analysis. DEA places less emphasis on the shape of the efficiency frontier and is credited for its axiomatic properties that can accommodate a multiplicity of inputs and outputs. On the other hand, SFA is able to decompose deviations from the frontier into random noise and inefficiency terms. These tools are not direct competitors, but complement each other, due to their respective advantages.

Since DEA and SFA use different assumptions, efficiency scores from the two methods may be inconsistent. Where these two techniques are used concurrently, it creates complexity as to which scores to adopt. In some cases, regulators use the arithmetic average of the firm-specific DEA and SFA efficiency estimates. Kuosmanen et al. (2013) calls this approach "the naïve method of averaging" (NMA) and has been used in the regulation of electricity distribution utilities in Finland. In Germany, for example regulators estimated DEA and SFA efficiency estimates then choose the highest out of the two (see Kuosmanen 2012; von Hirschhausen et al. 2006). However, both the arithmetic average and taking the highest score violate the assumptions of DEA and SFA. For this reason, Kuosmanen and Kortelainen (2012) developed the Stochastic Non-parametric Envelopment of Data (StoNED) method, which combines the axiomatic, non-parametric, piecewise linear DEA-style frontier with a stochastic SFA-style treatment of inefficiency and noise. StoNED is more robust to both model misspecification and noise, because its less restrictive assumptions imply a wider range of applicability.

Policymakers in developing and emerging economies are now beginning to collect data that can serve as a basis for performance comparison, which can assist decision-makers to identify under-performing water utilities (Corton and Berg 2009). An increasing number of countries are adopting performance appraisals to promote efficiency improvement in water provision. Performance appraisals almost invariably involve some form of benchmarking, or the comparison of actual performance versus some reference performance. Although there are various steps that can be used to undertake benchmarking, the process generally entails identifying relevant performance indicators; determining where performance should be, versus where it is at the time the evaluation is done (i.e. identifying performance gaps); determining the performance gap drivers; and designing an action plan to deal with the gaps. This process is repeated continuously, as the organisation continuously improves its products and services.

Regardless of some emerging interests in data collection, performance benchmarking of water distribution utilities is still lagging in most developing countries. This is mainly because there is generally no standardised data that can allow for robust efficiency analyses in most developing countries. In cases where some data exist, gaps and inconsistencies exist due to poor data management in most of these countries. As a result, very little is known about how water distribution utilities compare in supplying water services. In most cases the application of rigorous and more robust tools such as DEA and SFA in benchmarking the performance of water utilities in developing countries is often limited to academics.

In the management of utilities, there is usually debate on which method would be most preferred to guide in the choice of level of inputs or outputs to achieve the utilities' goals. Application of different approaches usually bring out different policy implications and have monetary implications on the utilities operations, if the methodological estimates require very significant reduction in inputs or huge increases in outputs. It is therefore important to present utilities with a set of different methodologies to aid them in decision making, and preferably advise them on the best amongst the many competing models. This paper attempts to accomplish this objective. The goal of this paper is therefore to determine the efficiency scores of water utilities in South Africa by comparing results obtained from parametric, non-parametric and semi-parametric approaches. This study is an extension of existing studies which predominantly use frontier analysis, data envelopment analysis, or other indicator based approaches in regulation of their operations.

This methodological cross-checking process provides more robust, reliable and useful information and diagnostics for regulatory analysis and policymakers. This is an innovative approach, and to the best of our knowledge this is one of few such studies – and, the first cross-checking process using three methods – to be applied to the water sector. Most studies that compare efficiency models in the regulation of water utilities usually compare DEA and SFA (see Dong et al. 2014; Lannier and Porcher 2014). A study similar to ours is Kuosmanen et al. (2013) which compares DEA, SFA and StoNED to establish the best practice for benchmark regulation of the Finnish electricity distribution sector. Another important study to note in the literature is the work of Andor and Hesse (2014) which use Monte Carlo simulations to evaluate the performance of StoNED relative to DEA and SFA. Furthermore, our study presents one of the first application of the StoNED approach to the water sector. The method has previously been applied to the in the electricity distribution sector, energy sector, agriculture sector, and the banking sector (see Dong et al. 2014; Eskelinen and Kuosmanen 2013; Kuosmanen 2012; Li et al. 2016; Vidoli and Ferrara 2015).

The rest of the paper is organised into seven sections. Section 2 discusses the benchmarking methods used in the study. Section 3 gives an overview of the South African water sector. Section 4 reviews some empirical literature, while Section 5 presents the empirical approach. Section 6 discusses the data used in the study. Section 7 presents the results. Section 8 concludes the study.

2. The frontier approaches for benchmarking

Parametric and non-parametric efficiency analysis techniques have become strategic tools for comparing the performance (benchmarking) of one decision-making unit (DMU) with best practice among all peer DMUs (Zhu 2014). Parametric techniques include corrected ordinary least squares or COLS (Winsten 1957), parametric programming (Timmer 1971), and the stochastic frontier analysis (SFA). Non-parametric methods include DEA and the free disposable hull (FDH). Each of the twin branches has its underlying assumptions which contribute to its advantages and disadvantages. DEA, and SFA methods are the commonly used approaches in benchmarking, but StoNED has been recently developed to exploit the advantages from the two approaches, and also overcome their weaknesses. In this section, we discuss these three efficiency analysis methods.

2.1 Data envelopment analysis (DEA)

DEA (Charnes et al. 1978; Farrell 1957) constructs a non-parametric envelopment frontier over given data points, such that all observed points are on or below the frontier. In the context of this study, the idea is to obtain a measure of the ratio of all outputs over all inputs, such as $u'y_i/v'x_i$, where u is an $Mx1$ vector of output weight and v is a $Kx1$ vector of input weights; given that M represents outputs and K represents inputs for each water utility. The process involves obtaining values for u and v such that the efficiency measure of the i^{th} water utility is maximised (subject to the constraint that all efficiency measures are equal to or less than one). To avoid the problem of an infinite number of solutions, the constraint $v'x_i = 1$ is imposed. When this happens, the multiplier form of the linear programming problem, where the change of notation from u and v to μ and v reflects the transformation emerges. DEA can be output orientated, which aims at expanding outputs while keeping the inputs constant, or input oriented, which aims at reducing the inputs while holding the outputs constant. In both orientations, the researcher can estimate constant returns to scale (CRS) or variable returns to scale (VRS) model. Depending on the depth of the analysis, one can obtain technical efficiency, allocative efficiency, or cost efficiency scores. One can also exploit the returns to scale and estimate the scale efficiency scores.

The DEA model proposed by Charnes et al. (1978) had an input orientation that assumed CRS. In this study, we adopt the input-oriented assumption proposed by Charnes et al. (1978) to estimate efficiency in South African water distribution utilities. However, since South African water utilities are quite diverse in terms of size, type and operating environment, we assume that they are at different stages of the production process. As such, we adopt the VRS assumption in our DEA estimation. More precisely, we estimate an input-oriented DEA that assumes VRS and produce technical efficiency scores only.

2.2 Stochastic frontier analysis (SFA)

SFA is a parametric efficiency analysis technique that assumes a Cobb-Douglas, a log-linear or a translog functional form (Aigner et al. 1977; Meeusen and van den Broeck 1977). The efficiency of DMUs is determined based on the specified functional form. The original formulation that is the foundation of SFA is:

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u, \tag{1}$$

where y is the observed outcome (goal attainment), $\boldsymbol{\beta}\mathbf{x} + v$ is the optimal frontier goal pursued by the DMU (e.g. minimum cost), $\boldsymbol{\beta}\mathbf{x}$ is the deterministic part of the frontier, and $v \sim \mathrm{N}[0, \sigma_v^2]$ is the stochastic part. The two parts together constitute the stochastic frontier. The amount by which the observed DMU fails to reach the optimum (i.e. the frontier) is u, where $u = |U|$ and $U \sim \mathrm{N}[0, \sigma_u^2]$. The stochastic cost frontier then changes to $v + u$, where u represents inefficiency.

Different specifications of the terms u and v distinguish stochastic frontier models. According to Aigner et al. (1977), the half normal model is the basic form of the stochastic frontier model. It assumes u to be independently half-normally $[\mathrm{N} + (0, \sigma_u{}^2)]$ distributed, with the idiosyncratic component v independently normally $[\mathrm{N}(0, \sigma_v)]$ distributed over the observation. Other SFA model specifications are the normal-exponential model (where u is independently exponentially distributed with variance $(\sigma_u{}^2)$), and the truncated-normal model (where u is independently $[\mathrm{N} + (\mu, \sigma_u{}^2)]$ distributed with truncation point at 0). Our study uses the commonly applied half normal SFA model specification to estimate technical efficiency in South African water utilities. One advantage of the half normal SFA model specification is that it can fit models with heteroskedastic error components, conditional on a set of covariates. Most water utilities are required to supply a specific population with a certain volume of water. The output level is therefore predetermined, and the specification that would fit this scenario would be one of cost reduction in achieving this set level of output.

2.3 Stochastic non-parametric envelopment of data (StoNED)

This method combines the two methodologies developed earlier into a unified model. Developed by Kuosmanen and Kortelainen (2012), the StoNED has two main stages. The first stage estimates the shape of the function using the convex non-parametric least squares (CNLS) regression, which belongs to the set of continuous, monotonic increasing and globally concave functions whose disturbances satisfy the Gauss-Markov assumptions. The second stage (which will be our focus in this study) estimates the expected inefficiency (μ), variance parameters (σ_u^2, σ_v^2), and DMU-specific inefficiencies. Kuosmanen (2012) suggests we introduce a composite error term $(\varepsilon_i = u_i + v_i)$, and linearise the cost frontier function by taking the natural logs of both sides, to obtain:

$$lnTC_i = lnf(\boldsymbol{y}_i) + \varepsilon_i = lnf(\boldsymbol{y}_i) - u_i + v_i \quad (2)$$

The main challenge in the least squares estimation of equation 2 is that the expected value of the composite error term is negative, due to the inefficiency term $u > 0$; that is, $E(u_i) = \mu > 0$. Kuosmanen (2008) reiterates that the composite error term in the model violates the Gauss-Markov properties, which can be restored by rewriting equation 2 as:

$$lnTC_i = (lnf(\boldsymbol{y}_i) - \mu)(\varepsilon_i + \mu) = lng(\boldsymbol{y}_i) + v_i \tag{3}$$
$$\hat{\varepsilon}_i = \hat{v}_i - \hat{\sigma}_u\sqrt{2/\pi}$$

where $lng(\boldsymbol{y}_i) = lnf(\boldsymbol{y}_i) - \mu$ is the average practice cost function which can be contrasted with the best practice cost frontier $lnf(\boldsymbol{y}_i)$, while $v_i = \varepsilon_i + \mu$ is the modified composite error term. Since μ is a constant, the average practice function $lng(\boldsymbol{y}_i)$ inherits concavity and monotonicity properties from the best practice function $lnf(\boldsymbol{y}_i)$. The modified error term v_i satisfies the Gauss-Markov assumptions. The average practice frontier function can be estimated by a non-parametric regression technique such as StoNED. In the StoNED model, the assumption is that the cost of providing water (TC) by utility depends on a vector of outputs $\boldsymbol{y}$. Therefore, for each water utility, the CNLS problem is to find $g \in F_2$ that minimises the sum of square deviations of the average practice function, given as:

$$\underset{f,v}{Min} \sum_{i=1}^{n} v_i^2 \left| lnTC_i = lng(\boldsymbol{y}_i) + v_i \qquad \forall_i = 1, \dots, n \right.$$
$$\text{s.t.} \quad g \in F_2 \tag{4}$$

The CNLS estimator for the water utilities cost function is obtained as the optimal solution to the following least squares problem, which can be solved by convex programming algorithms and solvers:

$$\underset{y,\boldsymbol{\beta},v}{Min} \sum_{i=1}^{n} \left(v_i^{CNLS}\right)^2$$
$$\begin{aligned} s.t. \quad & lnTC_i = \alpha_i + \boldsymbol{\beta}_i' ln\boldsymbol{y}_i + v_i^{CNLS} && \forall_i = 1, \dots, n \\ & \alpha_i + \boldsymbol{\beta}_i'\boldsymbol{y}_i \le \alpha_h + \boldsymbol{\beta}_h'\boldsymbol{y}_h && \forall_i, \forall_h = 1, \dots, n \\ & \boldsymbol{\beta}_i \ge 0 && \forall_i = 1, \dots, n \\ & && g \in F_2 \end{aligned} \tag{5}$$

where α_i is the intercept and $\boldsymbol{\beta_i}$ represents the coefficient of the tangent hyperplanes, which can also be interpreted as the marginal costs of output variables. These coefficients are analogous to the multiplier weights in DEA; and in contrast to the linear regression model, they are specific to each DMU (Kuosmanen 2012). Parameter v_i^{CNLS} is the CNLS residual, the CNLS estimator of g is monotonic increasing and concave, while ε_i^{CNLS} does not need to be identically and independently distributed but is uncorrelated with outputs $\boldsymbol{y}$ (see Kuosmanen and Johnson 2010).

After the estimation of the CNLS residuals ($\hat{v}_i^{CNLS}$), the next step – which is the basis of our study – disentangles inefficiency from noise by imposing more specific distributional assumptions. Following the basic SFA developed by Aigner et al. (1977), we assume the half-normal distribution for the inefficiency term, and a normally distributed noise term. Usually the noise term is symmetrically distributed, and any skewing in the CNLS residual estimates can be attributed to inefficiency. According to Kuosmanen and Fosgerau (2009) it is essential to test if the skewing is statistically significant, in which case one can use the method of moments (MM) or the pseudo-likelihood (PSL) functions to estimate the variance parameters of the inefficiency and noise terms (σ_u^2, σ_v^2). When MM is used, assuming a half-normal inefficiency term and a normally distributed noise term, the second and third central moments of the composite error are given by:

$$M_2 = \left[\frac{\pi-2}{\pi}\right]\sigma_u^2 + \sigma_v^2, \qquad M_3 = \left(\sqrt{\frac{2}{\pi}}\right)\left[1 - \frac{4}{\pi}\right]\sigma_u^3. \tag{6}$$

Based on the distribution of CNLS residuals, these moments can be expressed as;

$$\widehat{M}_2 = \sum_{i=1}^{n} \frac{(\hat{v}_i - \hat{E}(v_i))^2}{n}, \quad \widehat{M}_3 = \sum_{i=1}^{n} \frac{(\hat{v}_i - \hat{E}(v_i))^3}{n}. \tag{7}$$

The third moment M_3, which is the skewness of the distribution, depends on the standard deviation of the parameter σ_u. This implies that the estimated $\widehat{M}_3$ should be positive in the case of a cost frontier. The σ_u parameter can be estimated as:

$$\hat{\sigma}_u = \sqrt[3]{\frac{\widehat{M}_3}{\left(\sqrt{2/\pi}\right)\left[1-\frac{4}{\pi}\right]}}. \tag{8}$$

Additionally, the standard deviation of the error term σ_v can also be estimated as follows:

$$\hat{\sigma}_v = \sqrt[2]{\widehat{M}_2 - \left[\frac{\pi-2}{\pi}\right]\hat{\sigma}_u^2}. \tag{9}$$

Citing Aigner et al. (1977) and Greene (2008), Kuosmanen (2012) suggests that these MM estimators are unbiased and consistent, but not as efficient as maximum likelihood estimators. Using the estimator $\hat{\sigma}_u$ from MM, the best cost frontier function $lnf(\boldsymbol{y}_i)$ can be presented as:

$$ln\hat{f}(\boldsymbol{y}_i) = ln\hat{g}_{min}(\boldsymbol{y}_i) + \hat{\sigma}_u\sqrt{2/\pi} \tag{10}$$

According to Kuosmanen and Johnson (2010), this is like shifting the average practice frontier obtained from the CNLS by the expected value of the inefficiency term. The firm-specific inefficiency component u_i can be inferred indirectly from Jondrow et al. (1982) conditional distribution of inefficiency u_i given ε_i, irrespective of how the estimators of σ_u and σ_v are obtained. Under the assumption of a normally distributed error term and half-normally distributed inefficiency term, Jondrow et al. (1982) derive the conditional distribution of u_i given ε_i, and propose the conditional mean of the point estimate of u_i (i.e. $E(u_i|\varepsilon_i)$) as:

$$\hat{E}(u_i|\hat{\varepsilon}_i) = \mu_* + \sigma_*\left[\frac{f(-\mu_*/\sigma_*)}{1-F(-\mu_*/\sigma_*)}\right] \tag{11}$$

where f represents the standard normal density function $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$ and F is the cumulative density function. Note that $(-\mu_*/\sigma_* = \varepsilon\lambda/\sigma)$ where $(\lambda = \sigma_u/\sigma_v)$. Given the parameter estimates $\hat{\sigma}_u$ and $\hat{\sigma}_v$ of the conditional inefficiency obtained from the method of moments, the conditional mean of u (assuming a truncated normal distribution) is given as:

$$\hat{E}(u_i|\hat{\varepsilon}_i) = \frac{\hat{\sigma}_u\hat{\sigma}_v}{\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\left[\frac{f\left(\frac{\hat{\varepsilon}_i\hat{\sigma}_u}{\hat{\sigma}_v\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\right)}{1 - \mathrm{F}\left(\frac{\hat{\varepsilon}_i\hat{\sigma}_u}{\hat{\sigma}_v\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\right)} - \frac{\hat{\varepsilon}_i\hat{\sigma}_u}{\hat{\sigma}_v\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\right] \tag{12}$$

where $\hat{\varepsilon}_i = \hat{v}_i - \hat{\sigma}_u\sqrt{2/\pi}$ is the estimate of the composite error term, and not the CNLS residual. The obtained conditional expected value from equation 12 is an unbiased but inconsistent estimator of u_i; and irrespective of the sample size, each utility will have a unique value of u_i. Technical efficiency estimates are given by $TE_i = e^{\hat{E}(u_i|\hat{\varepsilon}_i)}$ (see Dios-Palomares et al. 2002). The TE estimates can also be estimated by defining the ratio $(y_i/\hat{g}(x_i)$, where $\hat{g}(x_i) = \hat{E}(y_i|x_i) + \hat{\mu}$ is the estimated non-parametric frontier, which can be expressed as $\hat{g}(x_i) = \hat{E}(y_i|x_i) + \hat{\sigma}_u\sqrt{2/\pi}$ for the method of moments. Jondrow et al. (1982) estimates for $\hat{u}_i$ can be converted to cost efficiency measures (CE), expressed in the percentage scale by using CE = 100% x exp ($\hat{u}_i$). The range of the cost efficiency scores CE is [0%, 100%], where CE=100% corresponds to the cost-efficient activity level (see Kuosmanen 2012).

3. Application of benchmarking approaches

Over the years, studies in the literature have relied on DEA to estimate the efficiency of water utilities (see Brettenny and Sharp 2016; Carvalho et al. 2015; Carvalho et al 2013; De Witte and Marques 2010; Guerrini et al. 2015). Although DEA is a useful efficiency-analysis tool, it is often criticised for not allowing random error by assuming that any deviation from the frontier is inefficiency; an assumption that exaggerates inefficiency if noise is present (see Coelli et al. 2005; Leleu 2006; Simar and Wilson 2008). Assuming away the noise term makes DEA biased in small samples, and sensitive to outliers. Because of these criticisms, several studies that estimate the efficiency of water utilities opt for SFA as a better tool (see Aubert and Reynaud 2005; Baranzini et al 2008; Filippini et al. 2007; Horn and Saito 2011; Souza et al. 2007; Vishwakarma and Kulshrestha 2010). SFA is credited for its ability to control for heterogeneity in the sample. However, it is often criticised for its functional form assumption, which is arbitrary and difficult to justify. Many commonly used functional forms fail to capture the economies of scope in joint production (Kuosmanen and Kortelainen 2012).

In order to benefit from the advantages of both DEA and SFA, there is a large body of studies in the literature that use both tools to estimate the efficiency of utilities (see Dong et al. 2014; Herwartz and Strumann 2012; Lannier and Porcher 2014; Zschill and Walter 2012). Apart from the scholarly literature, some regulators (mostly in Europe) use both DEA and SFA to measure efficiency and benchmark utilities for regulatory purposes. For example, from 2008 to 2011 the Finnish electricity regulator estimated both DEA and SFA and determined efficiency improvement targets using the arithmetic average of the firm-specific DEA and SFA scores (Kuosmanen et al. 2013). In Germany, the electricity regulator also estimated both DEA and SFA scores, but chose the maximum (von Hirschhausen et al. 2006). However, both taking the arithmetic average and taking the highest score violate the assumptions of DEA and SFA. Efficiency scores from DEA and SFA are estimated based on different assumptions; using the scores interchangeably violates the theories underpinning the models.

In a study that applied both DEA and SFA, Dong et al. (2014) used a panel data set of Chinese banks and found efficiency scores generated by SFA to be slightly higher than scores from DEA. The study also revealed that DEA and SFA were moderately consistent in identifying the best and worst quartile decision-making units regarding cost efficiency. In a different study, Herwartz and Strumann (2012) used DEA and SFA to examine whether hospital efficiency had emerged after the financial reform on spatial interdependence in Germany. Results showed that the SFA efficiency scores were higher than the DEA scores, reflecting that DEA identifies all deviations from the frontier as inefficiencies, while SFA separates inefficiency from noise.

The complexity associated with having to choose between DEA and SFA led to the development of StoNED. Introduced as a replacement to DEA and SFA in the regulation of electricity distribution utilities in Finland, StoNED is increasingly garnering attention in the literature. Kuosmanen et al. (2013) compared DEA, SFA and StoNED in the context of regulating electricity distribution, using data from Finland. The study compared the impacts of methodological choices on cost efficiency estimates and acceptable cost. In the results, the efficiency estimates were highly correlated, while the cost targets revealed major differences. StoNED yielded a root mean squared error of 4%, and its precision improved as the sample size increased. DEA yielded a root mean squared error of 10%, but performance deteriorated as the sample size increased. SFA had a root mean squared error of 144%, its poor performance explained to be due to the wrong functional form and multicollinearity. These comparisons demonstrate that the choice of method has significant effects on the regulatory outcomes.

Following the work of Kuosmanen (2012) and Kuosmanen et al. (2013), StoNED is gaining momentum in its use for efficiency analysis in the electricity sector. Cheng et al. (2015) examined the productivity development of Norwegian electricity distribution companies for the period 2004 to 2013. The study used DEA, SFA, and StoNED to examine productivity change, with the usual decompositions into efficiency change, technical change, and scale efficiency change. Based on the hypothesis that increasing investment and use of accounting-based capital costs leads to a negative bias in the productivity change estimates, analysis in the study was performed with and without capital costs, and results indicated a negative productivity development. In a different study, Li et al. (2016) applied SFA and StoNED to estimate efficiency for 23 Chinese power-grid companies, using data for the period 2005 to 2009. Among other findings, the study revealed that StoNED efficiency estimates were no different from those estimated by the various functional forms of SFA. StoNED has also been applied in various other studies in the energy literature (see Dai and Kuosmanen 2014; Johnson and Kuosmanen 2015; Mekaroonreung and Johnson 2012; Sabouhi Sabouni and Kenari 2014).

The application of StoNED to estimate efficiency is also found in the literature in areas such as banking, agriculture, and manufacturing. Eskelinen and Kuosmanen (2013) used StoNED to examine the efficiency and performance of sales teams over time in a bank branch network. The study estimates the intertemporal sales frontier from a panel of monthly data for the years 2007 to 2010. Using StoNED to assess the efficiency and performance development of the sales teams of a bank is one major contribution to the banking sector, where efficiency is central to sustainability. In a different study conducted in the agriculture sector, Vidoli and Ferrara (2015) use StoNED to estimate efficiency on Italian citrus firms. Using agricultural micro-data, the study maps out the overall level of efficiency, focusing on the evaluation of the differences observed due to the presence of contextual variables. Following a different method, Andor and Hesse (2014) used Monte Carlo simulations to evaluate the performance of StoNED relative to DEA and SFA, and found that in scenarios without noise, the rivalry is between DEA and SFA; while in noisy scenarios, StoNED pseudo-likelihood was a promising alternative to SFA.

Despite StoNED being applied across various fields, a gap exists in the water literature on studies that use the approach to estimate efficiency. The use of StoNED to benchmark water utilities is more appealing to utilities from developing countries where there is poor data quality. Developing countries are mostly associated with inconsistent and inaccurate data, making it difficult to use deterministic approaches like DEA which will show all measurement errors directly in the efficiency estimates. Nevertheless, DEA and SFA are commonly used to estimate the efficiency of water utilities in developing countries(see Brettenny and Sharp 2016; Carvalho et al. 2015; Souza et al. 2007; Vishwakarma and Kulshrestha 2010). The ability of StoNED to produce an unbiased estimator of efficiency scores for each utility regardless of size makes it an ideal technique for benchmarking utilities in developing countries. Utilities in such countries usually vary in size, operating environments, and do not have reliably consistent datasets. To the best of our knowledge, the performance of StoNED has not been tested in benchmarking water utilities. A study closer to ours is Kuosmanen et al. (2013), which compared DEA, SFA and StoNED in the Finnish electricity distribution sector, and found that StoNED yielded the most precise efficiency scores in both heterogeneous and noisy samples.

4. Benchmarking efforts in the South African water sector

Water sectors across the world are characterised by natural monopolies. In developing countries, the provision of water is usually the responsibility of public entities. This is because water provision is not a lucrative business in the developing world, where most citizens are poor and access to water is considered a basic human right; hence, water in these countries is a public good. In developed countries, the private sector takes part in water provision. Nevertheless, government regulations in most of these countries still make water service providers monopolies (Aubert and Reynaud 2005; De Witte and Marques 2010).

Water services provision is a process involving the movement of water from source to final user. The process is comprised of water treatment works, storage and distribution. Key players in the South African water sector value chain are the Department of Water and Sanitation (DWS), Water Services Providers (WSPs), Water Services Authorities (WSAs) and the final water users. The DWS is the custodian of the country's water resources. It is primarily responsible for the formulation and implementation of the policies that govern the water sector.

The legislative mandate of the DWS is to ensure that the country's water resources are protected, managed, used, developed, conserved and controlled in a sustainable manner that benefits all people and the environment. In pursuing this mandate, the DWS plays a regulatory role through setting national norms and standards for water services, monitoring the performance of WSAs, providing support to WSAs, and intervening in cases of water service delivery failure. As a regulator, the DWS develops a knowledge base and implements policies, procedures and integrated planning strategies. The DWS regulatory role also includes enforcing set standards by way of incentivising performance and sanctioning non-performance.

The authority to supply potable water is a competence of municipalities, which act as water utilities. A municipality accorded the responsibility of providing water services is called a WSA. Although South Africa has 278 municipalities, only 152 are WSAs. These WSAs encompass district municipalities that deliver within the jurisdiction of their local municipalities, and local municipalities that deliver within their own jurisdictions. In most cases, where a district municipality is authorised to provide water, the local municipalities in the area do not have such authority; and in instances where the local municipalities within a district are authorised, the related district municipality is not authorised. If the local municipality is deemed to have a large enough budget, then it is authorised, as opposed to the district municipality. The asymmetric delivery of water services across municipalities is due to the incapacity of many local municipalities to deliver water services, particularly those in the former homeland areas[5].

[5] Under the pre-1994 apartheid government, these were areas that were designated for specific black ethnic groups, with a high degree of political autonomy and even so-called independent from South Africa.

The need for standardised information, transparency and accountability has recently intensified, resulting in benchmarking efforts gaining momentum in South Africa. The primary goal of benchmarking is to provide key performance indicators (KPIs) that will enable utilities to compare their performance with the performance of other utilities and identify areas of improvement. However, in South Africa as in other developing countries, the conventional benchmarking approaches used in developed countries are not applicable in cases where water supply is intermittent, accessed by non-piped means, unmetered, and/or has a significant number of poorer customers on shared public connections (Mehta et al. 2013). Although water services provision is widespread in South Africa, there is a lack of data regarding the quality and level of service. Very little is known about how South African municipalities compare, in their capacity as WSAs. This is mainly due to a lack of standardised data, gaps in existing data, and lack of data verification.

Earlier performance benchmarking, initiated by the South African Local Government Association (SALGA) in 2001, was a failure. In 2006, government made further efforts; and since then, much has been achieved in the monitoring of municipal service performance through the Blue and Green Drop Certifications. The former proactively measures aspects contributing to sustainable safe drinking water, while the latter identifies and develops the core competencies required to sustainably improve the level of wastewater management. Although these programmes are plausible efforts towards performance benchmarking, water sectors across the world use scientific benchmarking tools such as DEA and SFA (see Baranzini et al. 2008; da Cruz et al. 2013; Guerrini et al. 2015). In South Africa, DEA and SFA are only mentioned in the academic literature and have not been implemented by authorities for benchmarking (see Brettenny and Sharp 2016; Tsegai et al. 2009).

5. Empirical approach

Water provision is a process involving several operational costs (Filippini et al. 2007). To characterise the process, it is essential to assume the existence of a mathematical relationship between water supply inputs and outputs. Water provision costs (in the case of South African water utilities) include bulk water purchases, labour, interest on capital, depreciation of fixed assets, and other general expenditure, such as fuel and oil, printing and stationery, and hiring of plant equipment. We aggregate these cost components into one total operation cost variable denoted by TC_i. We then use frontier models to relate TC_i to output variables that influence each utility's cost structure. Output variables in this study are the volume of water supplied (Q_i), length of water pipes ($MAINS_i$), the number of connections (CON_i), and the population (POP_i), which is an exogenous variable. Therefore, our study assumes the cost frontier model to be:

$$TC_i = f\left(Q_i^{\alpha} MAINS_i^{\beta} CON_i^{\gamma} POP_i^{\Omega}\right). exp(v_i + u_i) \tag{13}$$

where u_i is a random variable representing the technical inefficiency of the water utility i, v_i is a stochastic noise term that captures the effects of measurement errors, omitted variables and other random disturbances, and $\alpha, \beta, \gamma, \Omega$ are parameters to be estimated. When the vector $\boldsymbol{y}$ is used to represent the output variables $Q, MAINS$, CON and exogenous variable POP, equation 13 can be written as:

$$TC_i = f(\boldsymbol{y}_i). exp(v_i - u_i) \tag{14}$$

If a composite error term $\varepsilon_i = v_i - u_i$, which consists of an inefficiency term $u > 0$, and a random parameter term $E(v) = 0$ is introduced, and the cost function is linearized in logs, then equation 14 will be rewritten as:

$$lnTC_i = lnf(\boldsymbol{y}_i) + \varepsilon_i = lnf(\boldsymbol{y}_i) - u_i + v_i \tag{15}$$

The function presented in equation 15 is estimated using DEA, SFA and StoNED. For DEA, we estimate an input-oriented DEA that assumes VRS. We also estimate an SFA function where total operation cost of providing water services (TC_i) is expressed as a function of the volume of water supplied (Q_i), the length of the water pipes ($MAINS_i$), and the number of customers connected (CON_i)[6]. Therefore, the SFA functional form assumes:

$$lnTC_i = \alpha_i + lnQ_i + \ lnMAINS_i + lnCON_i - u_i + \ v_i \tag{16}$$

where v_i is the noise term assumed to be in normal distribution $v_i \sim \mathrm{N}[0, \sigma_v^2]$. The u_i notation is the non-negative inefficient term (which is the distance from the observed cost to the cost on the frontier). The assumptions on the error term require both u_i and v_i to be homoscedastic. However, South African water utilities are diverse in size and operating environment. Such differences are likely to be captured in v_i, resulting in heteroscedasticity. But heteroscedasticity in v_i could lead to biased estimates, while heteroscedasticity in u_i leads to deceptive efficiency

[6] Justifications for the selection of these variables are given in section 6 of the study.

scores (Kumbhakar and Lovell 2000). To address heteroscedasticity in an SFA function, one can account for the key drivers of the variation when estimating the efficiency term. This is done by estimating a simultaneous regression on the SFA function, the inefficiency term, and the random noise term.

In this study, the variation in the inefficiency term u_i driven by heteroscedasticity is controlled for by regressing u_i on the total number of water connections served by each utility (total connections in this regard is the sum of metered and unmetered water connections) i.e.

$$u_i = \alpha_1 + \alpha_2 lnCON_i + \delta_i \tag{17a}$$

This will ensure that the size of each utility is accounted for, minimising the impact of size on the efficiency estimates. Likewise, we also control for heteroscedasticity in the noise term v_i by regressing v_i on the total population served by each water utility (i.e)

$$v_i = \alpha_1 + \alpha_2 \ln\mathrm{POP}_i + \delta_i) \tag{17b}$$

Despite earlier studies raising concerns over modelling both u_i and v_i as being heteroskedastic (see Roehrig 1988), literature provides that if heteroskedasticity is utility specific, one can express the heteroskedasticity in u_i and v_i as functions of utility specific variables (Kumbhakar and Lovell 2000). In doing this, one controls for bias estimates that could have emanated from heteroscedasticity in the noise term, and biased efficiency scores because of heteroscedasticity in the inefficiency term. Although StoNED has two main stages, this study only reports results on the utility-specific efficiency scores, in line with the objective of the study. We present results from both the method of moments (MM) and pseudolikelihood (PSL) estimators of StoNED.

6. Data

The sample in our study comprises cross-sectional data for the 2013/4 period for 102 water utilities. We could not include all 152 water utilities, due to missing data; nor could we use panel data for other periods, due to too many gaps in the dataset. The period we are using for our analysis had the most complete data. The sample is representative of city, big-town, small-town and rural South African water utilities. As is the case in Dong et al. (2014), where input and output variables are defined, treating Chinese banks as multi-product firms that employ inputs X_i at given prices W_i that minimise total costs (TC_i) to produce outputs Q_i, this study treats South African water utilities in exactly the same manner. Our study uses a single input (TC_i) with three outputs $(Q_i, MAINS_i, CON_i)$ and an environmental variable (POP_i), used to control for heterogeneity in the operating environments of the utilities. Kuosmanen (2012) also uses the same variables in the context of electricity distribution utilities.

Total cost (TC) is the total water-related operating cost for each water utility. The total cost data is expressed in South African Rands[7], and comprises both direct and indirect costs resulting from bulk water purchases, labour, interest on capital, depreciation of fixed assets, and other general expenditure. In the context of this study, total cost is used as the only input variable, and water utilities are expected to minimise cost, given output variables.

Water output (Q) is the total quantity of water supplied by each water utility. The authorised consumption expressed in kilolitres (kl) per annum is used to account for water output. Authorised consumption is defined by the DWS as the total volume of metered and/or non-metered water taken by registered customers, the water supplier itself, or others who are implicitly or explicitly authorised to do so by the water supplier. Water output is used in this study as one of the three output variables.

Total connections (CON) is the total number of metred and non-metered water connections for each utility. The connections variable shows the number of water consumer units for a water utility and is used as one of the three output variables. More connections for a utility implies more consumer units, which may result in higher costs of providing water services, *ceteris paribus*.

Length of mains (MAINS) is the total length in kilometres of the water pipes owned by each water utility during the year. This shows the distance that the water moves, from point of extraction to the last consumer for each water utility. The variable is used as one of the three output variables. Utilities with longer pipe networks incur more costs from water losses and from transporting water over long distances.

Population (POP) is the number of people served by each utility. The hypothesis is that the total cost of providing water services is likely to be higher for utilities with higher population figures. Arbitrary higher total cost figures for utilities with lower population numbers may be attributed to inefficiency, ceteris paribus. The population figures express the size of each utility's distribution network. As such, this variable is used as an exogenous variable necessary

[7] The Rand is the South African currency. As at 31 March 2020, US$1 = ZAR17.80

to control for heterogeneity. In the literature, population is used extensively to control for heterogeneity (see Baranzini et al. 2008; Filippini et al. 2007; Tsegai et al. 2009). Summary statistics of all the variables used in the analysis are presented in Table 1.

Table 1 Descriptive statistics

Variable	Description	Mean	Std. Dev	Min	Max
TC	Total cost (thousand US$)	16 067	45 674	56	302 247
Q	Quantity of water (thousand kilolitres)	22 700	56 700	350	352 000
CON	Total number of connections (number)	73 543	130 269	2 306	713 143
MAINS	Length of mains (kilometres)	1 485	2 593	46	12 479
POP	Population served in thousands (number)	413	801	11	4 500

Notes: Sample size (N) =102 water utilities.

The table reveals that water distribution in South Africa is heterogeneous. The distinctions in size are evident from the statistics provided. Our sample contains all the categories of water utilities (i.e. city, big-town, small-town and rural water utilities). These categories vary in terms of size, operational environment, and resources. City and big-town water utilities serve huge populations, because they have urban cores that are highly populated due to urbanisation, which is prevalent in South Africa. On the other hand, utilities serving small towns and rural areas are relatively poor and have low densities in terms of population distribution. For such utilities, population levels may be relatively less and the number of water connections relatively few. However, the length of the mains and the total cost of providing water services may be relatively larger, because water is distributed across widely spaced household units.

There are statistical variations in the sample suggesting heterogeneity. For example, the total cost of providing water varies from $56 000 to $302 247 000, while the quantity of water supplied by utilities varies from 350 400 kilolitres to 352 million kilolitres. The number of water connections within the sample varies from 2 306 to 713 143, while the total length of water pipes varies from 46 kilometres to 12 479 kilometres. The population statistics also show evidence of heterogeneity, varying from 10 578 people to 4.5 million people. Total costs are higher in large water utilities, due to larger population sizes and many connections.

The implication of these variations is that they affect the selection of efficiency-analysis tools, and how the selected tools are used. The most applicable estimation tools in such heterogeneous samples are SFA and StoNED, because of their ability to separate and control for noise (Andor and Hesse 2014; Kuosmanen et al. 2013). If one decides to use DEA, as Brettenny and Sharp (2016) did, one needs to carefully separate water utilities according to their sizes and operating environments. This is because DEA is more susceptible to the influence of outliers (Banker 1993), and is likely to make utilities serving lower population numbers appear relatively more efficient, while utilities serving higher population numbers are deemed relatively less efficient. This is probably due to smaller utilities with smaller inputs being compared to larger utilities. Such results highlight the importance of accounting for utility size, given the heterogeneous nature of South African water utilities.

7. Results and discussion

This section presents the estimated results using StoNED, SFA, DEA and the naïve method of averaging (NMA) which averages scores from SFA and DEA. The use of scores from NMA is in line with other efficiency analysis studies which use the method when benchmarking utilities. NMA has also been used in the regulation of electricity distribution utilities in developed countries such as Finland (see Kuosmanen 2012; Kuosmanen et al. 2013). In terms of the StoNED method, results from both the MM and PSL estimators are presented. To analyse the efficiency scores reported by each method (and technique in terms of StoNED) discuss the results in five main steps. First, we present the correlation matrices of the efficiency scores generated by each method. Second, we present the descriptive statistics of efficiency scores based on method. Third, we compare the performance of each method based on the empirical estimates. Fourth, we group the utilities into six categories and compare the average efficiency scores for each category under each given method[8]. Finally, we analyse the impact of each method on the average monetary cost reduction targets.

Before various other analyses of efficiency scores are presented, it is important to understand the correlation matrices of the efficiency scores generated by each method. Correlation is performed using the Pearson's product moment correlation coefficients and the Spearman's rank correlation coefficients. Results for the correlation matrices are presented in Table 2, where the Pearson's correlation is shown at the top and Spearman's correlation at the bottom.

---

[8] South African municipalities are categorised into metropolitan, district and local municipalities. Metropolitans are municipalities with urban cores that are highly populated due to urbanisation, there are only eight metropolitans in South Africa. Districts are very large municipalities that are further divided into several local municipalities. Local municipalities vary in terms of size, population served, the existence of an urban core, among other characteristics and are further categorised into B1, B2, B3 and B4 (these will be discussed later in the study).

Table 2 Correlation analysis of efficiency scores

Pearson's correlation					
	StoNED (MM)	StoNED (PSL)	SFA	DEA	NMA
StoNED (MM)	1.0000				
StoNED (PSL)	0.9730	1.0000			
SFA	-0.2974	-0.2740	1.0000		
DEA	0.0375	0.0403	-0.3679	1.0000	
NMA	0.2174	0.2106	-0.0439	0.4203	1.0000
Spearman's rank-correlation					
StoNED (MM)	1.0000				
StoNED (PSL)	0.9996	1.0000			
SFA	-0.4234	-0.4149	1.0000		
DEA	0.5570	0.5501	-0.5232	1.0000	
NMA	0.2854	0.2826	-0.0478	0.8096	1.0000

The table shows a high correlation between StoNED MM and StoNED PSL in both the Pearson's correlation and the Spearman's rank-correlation. In terms of the latter, there is also a high correlation between DEA and NMA. Except for these, the correlation coefficients score are low between all estimation methods in both the Pearson's and Spearman's correlation matrices. The correlation matrices observed signify that although the estimator used under StoNED has little effect on the efficiency scores, the choice of method between StoNED, DEA and SFA has huge effect on the efficiency scores. This finding is not consistent with the correlation coefficients reported in Kuosmanen et al. (2013) where high and positive correlation was reported in every pair of efficiency scores. Therefore, it follows that further comparison of the efficiency scores is warranted. To gain more light in the efficiency scores, summary statistics for each estimation method (or technique) are presented in Table 3.

Table 3 Summary statistics of efficiency scores based on method

	Mean	Std. Dev	Median	Minimum	Maximum
StoNED (MM)	0.680	0.080	0.705	0.392	0.762
StoNED (PSL)	0.529	0.146	0.563	0.016	0.725
SFA	0.662	0.139	0.659	0.223	0.896
DEA	0.447	0.280	0.355	0.095	1.000
NMA	0.555	0.119	0.499	0.430	0.923

Table 3 shows that the average efficiency scores for the 102 water utilities range from 0.447 in DEA to 0.680 in StoNED MM. For StoNED MM, the average efficiency score means that on average, water utilities are 68% efficient (i.e. 32% inefficient). This implies that utilities could reduce their operating costs by 32% and still afford to supply the same quantity of water, serve both the same population, and number of connections. StoNED MM recorded considerably higher scores in terms of the mean, median and the minimum relative to the other methods. This could be because StoNED takes the noise term explicitly into account and captures heterogeneity of firms and their operating environments through the use of the contextual variable *z*, which is omitted in SFA and DEA (see Johnson and Kuosmanen 2015; Kuosmanen 2012; Kuosmanen et al. 2013). Interestingly, the same is not observed in StoNED PSL whose scores are all less than those reported in SFA, implying that StoNED PSL reports scores that are generally lower than those reported in StoNED MM and SFA. On the other hand, NMA had the largest minimum score of 0.430, suggesting that when this method is used, the least underperforming utility will have a score greater than those reported in the other methods.

To gain more insight into each estimation technique, we further analysed the frequency distribution of utilities with efficiency scores above the mean of each model. This implies that scores from each model are compared to the mean for the model. Results show that using StoNED MM, 66 of the 102 water utilities (i.e. 65% of the utilities) scored efficiency estimates above the method's mean of 0.681, implying that 65% of the utilities are more than 68.1% efficient. In terms of StoNED PSL, 63 of the 102 water utilities (i.e. 62% of the utilities) scored efficiency estimates above the method's mean of 0.529, implying that 62% of the utilities are more than 52.9% efficient. On the other hand, SFA results show that 51 utilities (i.e. 50% of the utilities) scored efficiency estimates above the method's mean of 0.662, implying that 66% of the utilities are more than 66.2% efficient. We acknowledge that the 50-50 split under SFA is rather strange and its drivers worth further exploration. In terms of DEA which reported a mean of 0.447, 38 utilities (i.e. 37% of the utilities) recorded efficiency scores above the model's mean, implying that only 37% of the utilities are more than 44.7% efficient under DEA. For NMA, 37 utilities (i.e. 36% of the utilities) scored efficiency estimates above the method's mean of 0.555.

Another key statistic Table 3 that can shed light on the performance of each model is the standard deviation. This expresses how much the utility-specific efficiency scores vary from the mean score. StoNED MM reported the lowest standard deviation, of 0.080; followed by NMA, with a standard deviation of 0.119. Interesting to note is that SFA recorded a standard deviation lower than the one recorded by StoNED PSL. If the standard deviation was to be used as an indicator for each model's ability to control for heterogeneity in the sample, it would be argued that StoNED MM and NMA control heterogeneity better than the other methods. To give detail on how the efficiency scores for all utilities in the sample were distributed around the mean, we illustrate utility-specific efficiency scores in Figure 1.

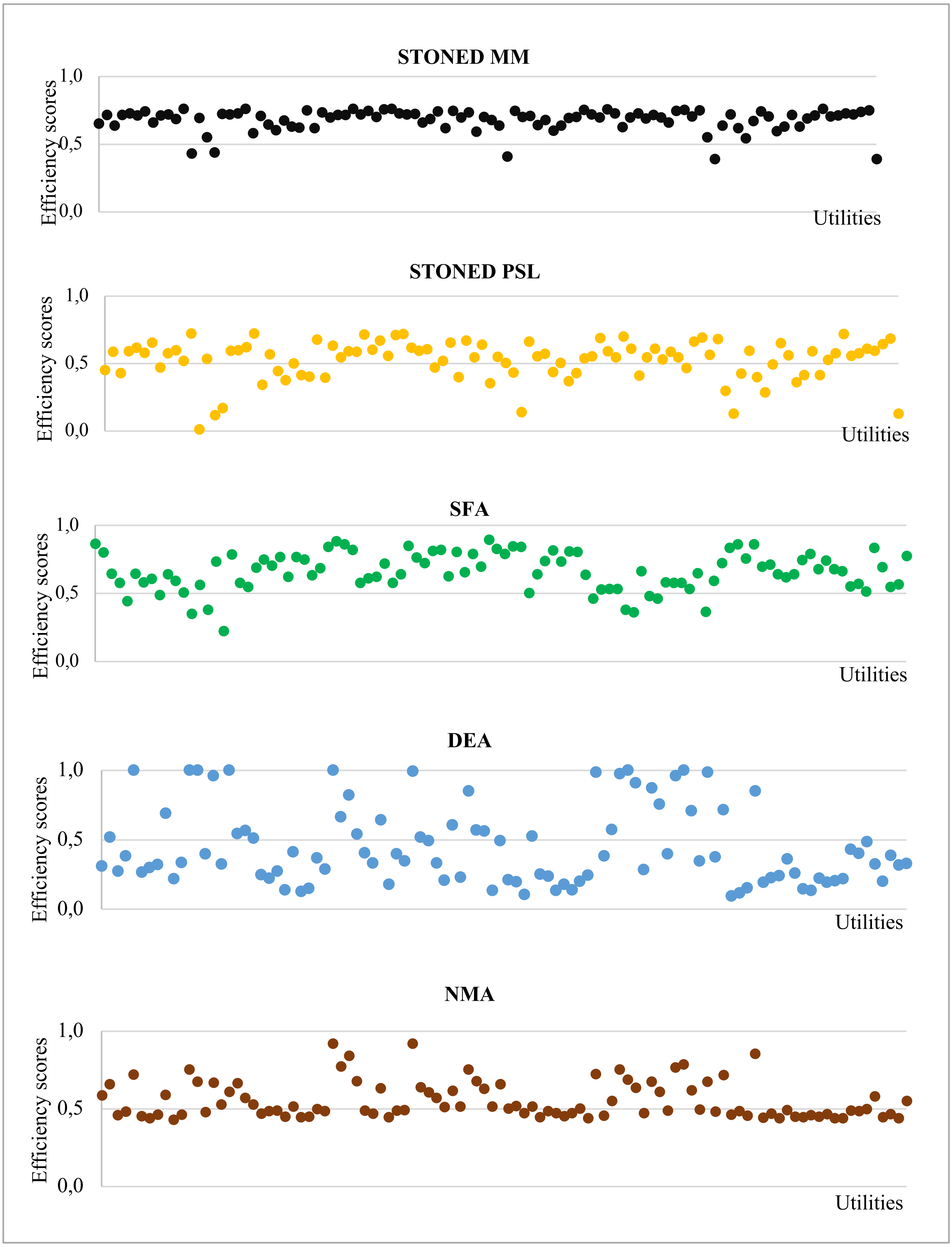


Fig. 1. Distribution of efficiency scores around the mean of each method

The scatter plots in Figure 1 show less variation around the mean in StoNED MM efficiency scores compared to scores from other models. StoNED MM had only five (5) utilities that recorded efficiency scores of less than 0.5, with the rest of utilities having almost uniformly distributed scores. The 5 utilities are smaller and poorer local municipalities with smaller towns as their urban cores. On the other hand, StoNED PSL had thirty two (32) utilities with scores below 0.5. These 32 utilities include both big and small municipalities, where among the big municipalities is one metropolitan and four district municipalities. In terms of SFA, eleven (11) utilities had efficiency scores below 0.5, and these were predominantly smaller municipalities and two district municipalities. DEA which reported the highest standard deviation of utilities compared to the other models had sixty seven (67) utilities with efficiency scores of less than 0.5. Municipalities that could not attain a 0.5 efficiency score under DEA include smaller and bigger municipalities. The existence of all municipal categories in the list of municipalities that performed poorly shows that the VRS property of DEA used in this study sufficiently controlled the differences in scale. In terms of NMA, although some fair distribution is observed, random spikes of highly performing utilities are observed.

Although the standard deviation can shed more light into the distribution of efficiency scores under each model, literature suggests that the statistic is not a proven scientific indicator of performance (see Andor and Hess 2014; Cheng et al. 2015; Kuosmanen et al. 2013). To understand the true performance of each method, studies various studies adopt other less subjective approaches. For example, Kuosmanen et al. (2013) graphically plotted StoNED efficiency scores against those obtained from NMA to evaluate the performance of each method in the Finnish electricity sector. NMA was used in that regard because it was used by the Finnish electricity regulator to benchmark the performance of electricity distribution utilities at the time. Since efficiency benchmarking using these methods is not yet adopted in the South African water sector, we follow Kuosmanen et al. (2013) and compare the performance of the two StoNED estimators against NMA[9]. This information is provided in Figure 2 where StoNED MM and NMA are compared in panel A while StoNED PSL and NMA are compared in panel B.

[9] Diagrams which compare StoNED efficiency scores to SFA and DEA, and SFA efficiency scores to DEA are provided in Appendix 1.

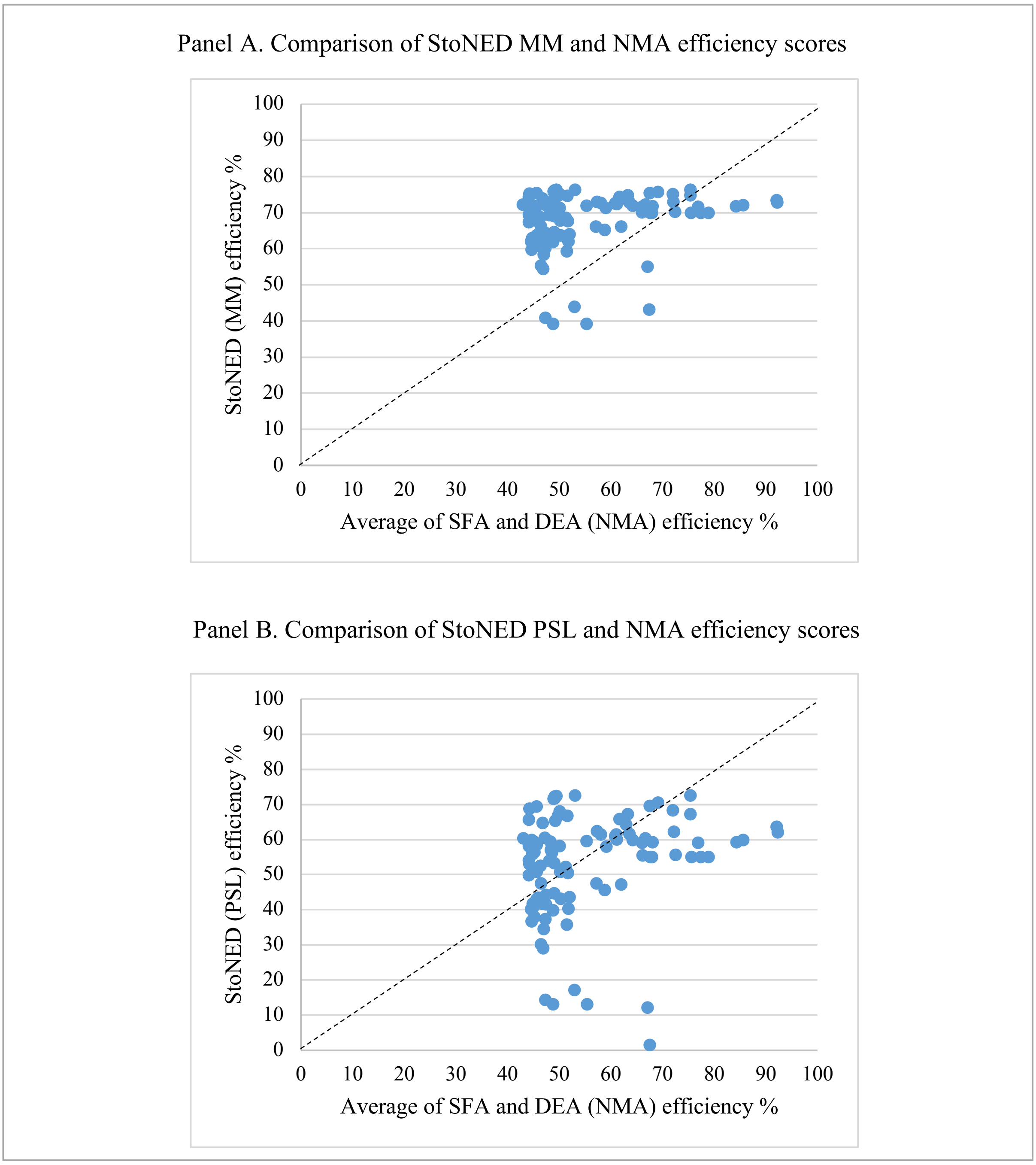


Fig. 2. Comparison of StoNED and NMA utility-specific efficiency scores

Panel A shows the pair of efficiency scores obtained by StoNED MM (on the vertical axis) and NMA (on the horizontal axis), while panel B shows the pair of efficiency scores obtained by StoNED PSL (on the vertical axis) and NMA (on the horizontal axis). The broken line in the middle of each diagram is a 45° line which in the context of panel A, all points above the line suggest that StoNED MM scores are greater than those of NMA. In terms of panel B, points above the 45° line suggest that StoNED PSL efficiency scores are greater than those of NMA. In panel A, there are more points above the 45° line, indicating that overall, StoNED MM produced efficiency scores that are greater than those reported in NMA. This suggests that when StoNED MM is used, cost reduction targets will be less than when NMA is used because the former produces higher efficiency scores than the latter. The same trend is not observed in panel B where some competition is noted between StoNED PSL and NMA, making it not clear which method is better than the other. However, it is noted in panel B that NMA reported some points where efficiency scores are very high (closer to 100%) than StoNED PSL. This was also noted in panel and could be due to the influence of DEA which had scores that were as high as 100%. If this is to be used as a measure of performance, utilities may prefer NMA to StoNED PSL because it entails less efficiency improvement targets.

Since South African municipalities are grouped into different categories, depending on the population dynamics served by each municipality, it is important to understand how each municipal category performed in terms of efficiency scores based on each estimation technique. South African municipalities are grouped into 6 categories (A, C, B1, B2, B3 and B4)[10]. After estimating efficiency scores for the whole sample of 102 utilities using the different techniques, we disentangle efficiency scores and group them according to the 6 municipal categories. This exercise was performed to understand how each group of utilities performed under each method. The average efficiency scores for each municipal category under each given method are presented in Table 4.

Table 4 Average efficiency scores for each municipal category in each estimation technique

	A	**C**	**B1**	**B2**	**B3**	**B4**
StoNED MM	0.712	0.676	0.703	0.707	0.662	0.604
StoNED PSL	0.580	0.498	0.565	0.587	0.497	0.373
SFA	0.845	0.698	0.728	0.660	0.596	0.810
DEA	0.714	0.517	0.359	0.335	0.464	0.137
NMA	0.779	0.607	0.543	0.407	0.530	0.473
N	*8*	*12*	*17*	*17*	*47*	*1*

[10] Category A is metropolitan municipalities, category C is district municipalities, category B1 is local municipalities with a large town or city as their urban core, category B2 is local municipalities with a medium town as their urban core, category B3 is local municipalities with a small town as their urban core, and category B4 is local municipalities without any urban core.

The table shows that under metropolitan municipalities (category A), all estimation techniques produced efficiency scores greater than 0.5. SFA produced the highest efficiency score of 0.845 while StoNED PSL had the least efficiency score of 0.580. SFA also recorded the highest scores for district municipalities (category C) and local municipalities (categories B1 and B4). However, unlike in categories A and C where StoNED PSL recorded the least efficiency scores, DEA had the least efficiency scores for all the local municipalities (categories B1 to B4). This implies that DEA reports metropolitan and district municipalities (i.e. bigger municipalities) as more efficient than local municipalities (relatively smaller municipalities). Although this result might be strange given that we adopted the VRS property which can control for scale, it can be argued that this could be because we did not model heteroskedasticity in the DEA model. Equally, it can be noted that StoNED MM and SFA maintained relatively higher efficiency scores across all municipal categories. On the other hand, StoNED PSL maintained relatively lower efficiency scores across all categories. For NMA, bigger municipalities generally recorded relatively higher efficiency scores than smaller municipalities.

The efficiency scores reported in this study are consistent with those reported in both the local and the international literature (see Brettenny and Sharp 2016; Estache and Rossi 2002; Horn and Saito 2011). Brettenny and Sharp (2016) estimated efficiencies for 88 South African water utilities and revealed average efficiency scores of 0.636 for urban water utilities and 0.526 for rural water utilities. As for scores from international water utilities, South African scores are comparable to those presented in Horn and Saito (2011), in which efficiencies were estimated for 831 Japanese utilities and the average scores were between 0.596 and 0.621. Similarly, Estache and Rossi (2002) estimated efficiencies for water utilities in Asia and the Pacific region. The study found the average efficiency to be within the range 0.72 to 0.78 in Bangkok, 0.66 to 0.69 in Beijing, 0.70 to 0.77 in Delhi, 0.66 to 0.77 in Hong Kong, 0.24 to 0.35 in Jakarta, 0.83 to 0.87 in Kuala Lumpur, and 0.74 to 0.75 in Singapore, among many others. These results indicate that South African water utilities compare well to international utilities.

Since one of the main reasons for efficiency analysis is to establish monetary cost reduction targets for each utility, we examine the impact of each model on monetary cost reduction targets. In this context, monetary cost reduction targets imply the amount by which utilities could reduce their costs to become more efficient. To do this, we adopt the approach used in Kuosmanen et al. (2013) where firm-specific efficiency scores were converted to monetary cost targets through multiplying the total cost for each utility by the inefficiency score, that is, $TC_i(1 - TE_i)$. In this context, TC_i is the total cost for each utility while TE_i is the technical efficiency score reported for each utility. Summary statistics on the monetary cost reduction targets are presented in Table 5.

Table 5 Monetary cost reduction targets (thousand US$)

	Mean	Std. Dev	Median	Minimum	Maximum
StoNED (MM)	4 605	12 560	693	17	80 697
StoNED (PSL)	6 597	17 715	1 056	25	110 427
SFA	3 198	6 871	830	13	47 780
DEA	4 183	7 127	1 634	0	46 910
NMA	3 690	5 932	1 309	12	31 701

A comparison of the average cost reduction targets generated by each method shows that each method (or estimation technique in the case of StoNED) yields different targets. StoNED PSL yielded the highest average cost reduction target of $6 597 000, implying that if StoNED PSL is used, utilities in the sample should reduce their total costs by about $6 597 000. SFA reported the least average cost reduction target of $3 198 000, which is almost half the cost reduction target reported in StoNED PSL. Revelations in Table 5 are consistent with the correlation coefficients estimates presented earlier which showed no correlation in estimates generated by the different methods. This implies that the choice of method and technique has an impact on the cost reduction targets. Such a result was also reported in Kuosmanen et al. (2013). In the context of our study, using SFA and NMA gives less average cost reduction target compared to using DEA and StoNED.

The empirical estimation performed in this study gives some indication on the efficiency scores generated by each method as well as the impact of model selection on both efficiency scores and cost reduction targets. However, literature suggests that the true performance of each method may not be sufficiently known from such estimations. To understand the true performance of each method, studies such as Andor and Hess (2014) and Kuosmanen et al. (2013) perform Monte Carlo simulations in addition to the empirical analysis and use performance measures such as the mean squared error (MSE), root mean squared error (RMSE), mean rank correlation (MRC), mean absolute deviation (MAD), among others. Although these are plausible activities on measuring the real performance of each model, this study only focused on a comparison efficiency scores based on empirical analysis.

8. Conclusion

In this paper, we employ parametric (SFA), non-parametric frontier (DEA) and semi-non-parametric (StoNED) approaches on a sample of South African water utilities, for methodological cross-checking purposes. In many developing countries there is a need to introduce rigorous benchmarking of the water sector, due to the low operational efficiency of existing public water utilities. As climate change intensifies, and competition increases between different needs for water, inefficiencies in the water sector in emerging economies such as South Africa are bound to rise significantly. Efficiency gains is a potential adaptation strategy that the water sector could use to address several emerging trends driven by climate change. Using the three efficiency analysis techniques and the naïve method of averaging, we report five key findings on the empirically generated efficiency scores.

First, low correlation coefficients were generally reported between the efficiency scores generated by each method. This finding suggests that the choice of method between StoNED, DEA and SFA has huge effect on the efficiency scores. Second, we find that StoNED MM recorded considerably higher scores in terms of the mean, median and the minimum relative to the other methods. The same was not observed in StoNED PSL whose mean, median and minimum scores were all less than those reported in SFA, implying that StoNED PSL reports scores that are generally lower than those reported in StoNED MM and SFA. On the other hand, NMA had the largest minimum score suggesting that when this method is used, the least underperforming utility will have a score greater than those reported in the other methods. In terms of the distribution of efficiency scores around the mean for each method, we found that StoNED MM reported the lowest standard deviation, followed by NMA. Interestingly, SFA recorded a standard deviation lower than that of StoNED PSL. If this statistic was be used indicate each model's ability to control for heterogeneity, it would be argued that StoNED MM and NMA control heterogeneity better than the other methods.

Third, a pairwise comparison of efficiency estimates from each method revealed that the choice of method used in estimation matters. For example, we find that StoNED MM produced efficiency scores that are greater than those reported in NMA, suggesting that if StoNED MM is adopted, cost reduction targets will be less than when NMA is used because the former produces higher efficiency scores. On the other hand the same relationship is not clear between StoNED PSL and NMA, even though it was observed that NMA reported points where efficiency scores were very high compared to StoNED PSL. Fourth, we examined the consistency of each method across the different types of water utilities and found that StoNED MM and SFA maintained relatively higher efficiency scores across all utility categories. On the other hand, StoNED PSL maintained relatively lower efficiency scores across all utility categories. However, DEA reported bigger utilities as more efficient than smaller utilities. This was also observed in NMA where bigger utilities generally recorded higher efficiency scores than smaller utilities. Finally, we found that the choice of method also has an impact on the average monetary cost reduction targets. StoNED PSL yielded the highest average cost reduction target, while SFA and NMA recorded almost half the cost reduction target reported in StoNED PSL.

Based on these results, we join other studies in the literature in arguing that the choice of estimation method and technique is important in efficiency analysis. The size of the water utility and its operating environment are equally important with factors to consider when deciding on the efficiency analysis method to use. This is hugely important when benchmarking water utilities in the developing countries which operate in diverse environments, with distinct budgets, and divergent supply of intermittent water. In consequence, benchmarking such utilities requires using methods that can control for heterogeneity and scale. One of the weaknesses of our study is that it focused only on the second stage of StoNED, which estimates utility-specific efficiency scores, future studies can also include the first stage of StoNED which estimates the shape of the function using the convex non-parametric least squares regression. Further, we recommend that future studies wishing to compare DEA, SFA and StoNED in the water sector should include other forms of distribution for SFA, e.g. exponential and truncated distributions. Our study only used the half-normal distribution. Regarding DEA, we recommend that future studies also test other forms of returns to scale. Our study focused only on an input-oriented VRS, because South African municipalities are different and are expected to be at different levels of production. Furthermore, we recommend future studies to also model heteroskedasticity in the DEA model. Although literature tends to use conventional DEA, its input/output data may contain random errors which may result in distorted efficiency scores due to statistical noise. Some studies may use the double-bootstrap DEA which adjusts efficiency estimates downwards because DEA estimates may be upwardly biased under the no noise assumption.

References

Aigner D, Lovell CK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6: 21-37.

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

Aubert C, Reynaud A (2005) The impact of regulation on cost efficiency: an empirical analysis of Wisconsin water utilities. Journal of Productivity Analysis 23: 383-409.

Banker RD (1993) Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. Management Science 39: 1265–1273.

Banker RD, Chang H (2006) The super-efficiency procedure for outlier identification, not for ranking efficient units. European Journal of Operational Research 175: 1311-1320.

Baranzini A, Faust AK, Maradan D (2008) Water supply: Costs and performance of water utilities: Evidence from Switzerland. 13th International Water Resources Association World Water Congress, 1-4 September 2008 Montpellier.

Brettenny W, Sharp G (2016) Efficiency evaluation of urban and rural municipal water service authorities in South Africa: A data envelopment analysis approach. Water SA 42: 11-19.

Carvalho P, Pedro I, Marques RC (2015) The most efficient clusters of Brazilian water companies. Water Policy 17: 902-917.

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

Cheng X, Bjorndal E, Lien G, Bjorndal MH (2015) Productivity development for Norwegian electricity distribution companies 2004-2013. Discussion Paper, Norwegian School of Economics.

Coelli T, Rao DSP, Battese GE (2005) An introduction to efficiency and productivity analysis. Boston/Dordrecht/London, Kluwer Academic Publishers.

Corton ML, Berg SV (2009) Benchmarking central American water utilities. Utilities Policy 17: 267-275.

Da Cruz NF, Carvalho P, Marques RC (2013) Disentangling the cost efficiency of jointly provided water and wastewater services. Utilities Policy 24: 70-77.

Dai X, Kuosmanen T (2014) Best-practice benchmarking using clustering methods: Application to energy regulation. Omega 42: 179-188.

De Witte K, Marques RC (2012) Gaming in a benchmarking environment. A non-parametric analysis of benchmarking in the water sector. Water Policy 14: 45-66.

Dios-Palomares R, Ramos A, Roldán J (2002) A Monte Carlo study on the technical efficiency estimation in the stochastic frontier model. Questiió 26: 443-459.

Dong Y, Hamilton R, Tippett M (2014) Cost efficiency of the Chinese banking sector: A comparison of stochastic frontier analysis and data envelopment analysis. Economic Modelling 36: 298-308.

Eskelinen J, Kuosmanen T (2013) Intertemporal efficiency analysis of sales teams of a bank: Stochastic semi-nonparametric approach. Journal of Banking and Finance 37: 5163-5175.

Estache A, Rossi MA (2002) How different is the efficiency of public and private water companies in Asia? The World Bank Economic Review 16: 139-148.

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

Filippini M, Hrovatin N, Zorić J, (2008) Cost efficiency of Slovenian water distribution utilities: An application of stochastic frontier methods. Journal of Productivity Analysis 29: 169-182.

Greenberg R, Nunamaker T (1987) A generalized multiple criteria model for control and evaluation of nonprofit organizations. Financial Accountability and Management 3: 331-342.

Greene WH (2008) The econometric approach to efficiency analysis. Oxford University Press.

Guerrini A, Romano G, Leardini C, Martini M (2015) Measuring the efficiency of wastewater services through data envelopment analysis. Water Science and Technology 71: 1845-1851.

Herwartz H, Strumann C (2012) On the effect of prospective payment on local hospital competition in Germany. Health Care Management Science 15: 48-62.

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

Johnson AL, Kuosmanen T (2015) An introduction to CNLS and StoNED methods for efficiency analysis: Economic insights and computational aspects, in Benchmarking for Performance Evaluation (117-186). Springer, New Delhi.

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

Kumbhakar SC, Lovell CAK (2000) Stochastic Frontier Analysis. Oxford University Press.

Kuosmanen T (2008) Representation theorem for convex nonparametric least squares. Economic Journal 11: 308-325.

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

Kuosmanen T, Fosgerau M (2009) Neoclassical versus frontier production models? Testing for the presence of inefficiencies in the regression residuals. Scandinavian Journal of Economics 111: 317–333.

Kuosmanen T, Johnson AL (2010) Data envelopment analysis as nonparametric least squares regression. Operations Research 58: 149-160.

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

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

Lannier AL, Porcher S (2014) Efficiency in the public and private French water utilities: prospects for benchmarking. Applied Economics 46: 556-572.

Leleu H (2006) A linear programming framework for free disposal hull technologies and cost functions: primal and dual models. European Journal of Operational Research 168: 340-344.

Li H-Z, Kopsakangas-Savolainen M, Xiao X-Z, Tian Z-Z, Yang X-Y, Wang J-L (2016) Cost efficiency of electric grid utilities in China: A comparison of estimates from SFA–MLE, SFA–Bayes and StoNED–CNLS. Energy Economics 55: 272-283.

Meeusen W, van Den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 435-444.

Mehta D, Meera M, Anitha I (2013) A Review of performance benchmarking urban water supply and sanitation. CEPT University.

Mekaroonreung M, Johnson AL (2012) Estimating the shadow prices of SO_2 and NO_x for US coal power plants: a convex nonparametric least squares approach. Energy Economics 34: 723-732.

Roehrig CS (1988) Conditions for identification in nonparametric and parametric models. Econometrica 56(2): 433-447.

Sabouhi Sabouni M, Kenari RE (2014) Measuring technological gap ratio of wheat production using StoNED approach to metafrontier. International Journal of Applied Operational Research 4.

Simar L, Wilson PW (2008) Statistical inference in nonparametric frontier models: recent developments and perspectives. The Measurement of Productive Efficiency and Productivity Growth 421-521.

Souza GS, Faria RC, Moreira TBS (2007) Estimating the relative efficiency of Brazilian publicly and privately-owned water utilities: a stochastic cost frontier approach. Journal of the American Water Resources Association 45: 1237-1244.

Timmer CP (1971) Using a probabilistic frontier production function to measure technical efficiency. Journal of Political Economy 79: 776-794.

Tsegai DW, Linz T, Kloos J (2009) Economic analysis of water supply cost structure in the Middle Olifants sub-basin of South Africa. Working Paper, Social Science Research Network (SSRN).

Vidoli F, Ferrara G (2015) Analyzing Italian citrus sector by semi-nonparametric frontier efficiency models. Empirical Economics 49: 641-658.

Vishwakarma A, Kulshrestha M, (2010) Stochastic production frontier analysis of water supply utility of urban cities in the State of Madhya Pradesh, India. International Journal of Environmental Sciences 1: 357-367.

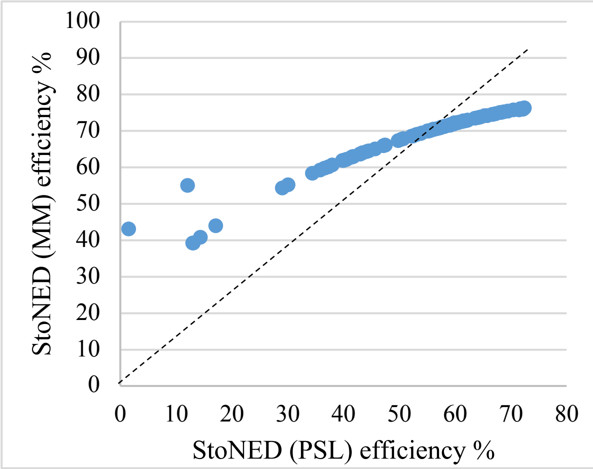
von Hirschhausen C, Cullmann A, Kappeler A (2006) Efficiency analysis of German electricity distribution utilities–non-parametric and parametric tests. Applied Economics 38: 2553-2566.

Winsten C (1957) Discussion on Mr. Farrells paper. Journal of the Royal Statistical Society 120: 282-284.

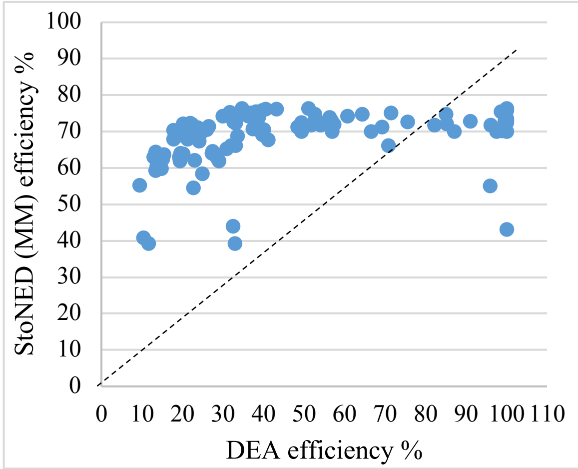
Zhu J (2014) Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets. Springer.

Zschill M, Walter M (2012) The performance of German water utilities: a (semi)-parametric analysis. Applied Economics 44: 3749-3764.

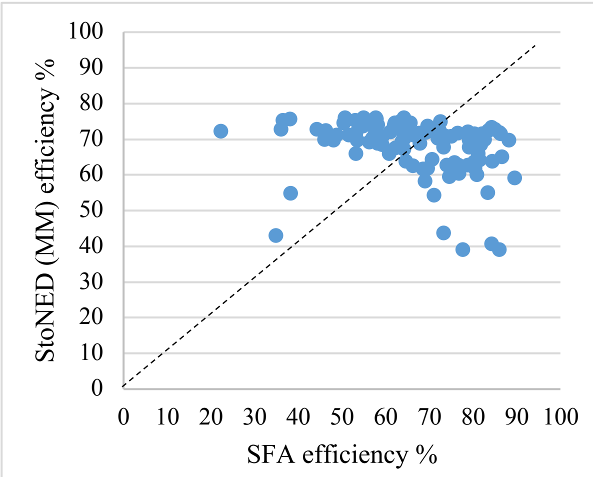
Appendix 1 Comparison of efficiency scores from each method



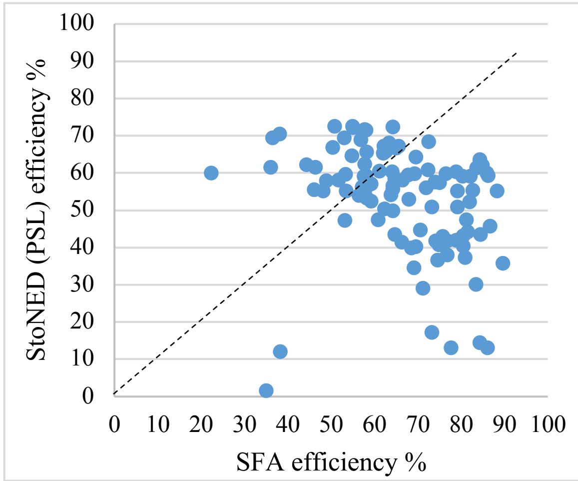
StoNED PSL versus StoNED MM



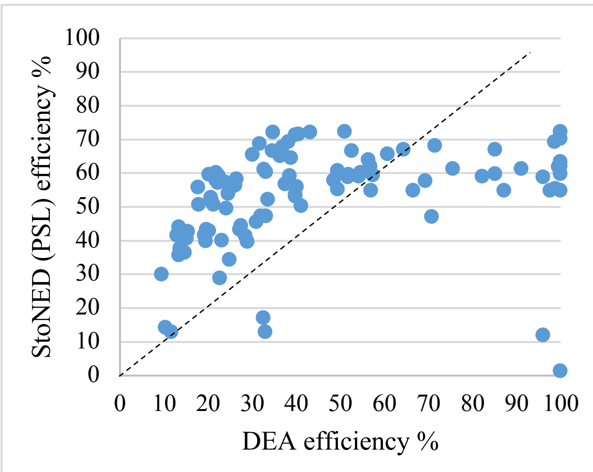
DEA versus StoNED MM



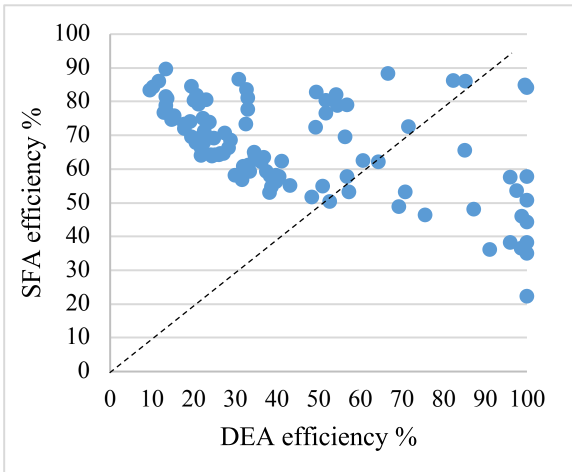
SFA versus StoNED MM



SFA versus StoNED PSL



DEA versus StoNED PSL



DEA versus SFA