Efficiency in South African water utilities: a comparison of estimates from DEA, SFA and StoNED

Genius Murwirapachena, Jugal Mahabir, Richard Mulwa and Johane Dikgang

ERSA working paper 780

April 2019

Economic Research Southern Africa (ERSA) is a research programme funded by the National Treasury of South Africa.

Efficiency in South African water utilities: a comparison of estimates from DEA, SFA and StoNED

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

ABSTRACT

For efficiency analysis to be useful to policymakers, the various approaches used should produce estimates that are consistent in identifying the best and worst firms, as well as overall rankings of firms in terms of their efficiency levels. This paper investigates the consistency of efficiency scores obtained from the data envelopment analysis (DEA), stochastic frontier analysis (SFA), and stochastic non-parametric envelopment of data (StoNED) methods. We estimate cost efficiency based on cross-sectional data from 102 South African water utilities in the period 2013/14. The results suggest that the StoNED method (based on the methods of moments estimator) outperforms SFA and DEA. However, based on the pseudo-likelihood estimator, SFA outperformed StoNED. Overall, the results suggest moderate consistency across the three methods. Based on the findings, we conclude that our results are robust.

Classification-JEL: D24, H41, P28, Q25

Keywords: water utilities, StoNED, DEA, SFA, frontier efficiency.

[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, Durban, South Africa.

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

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

[4] Public and Environmental Economics Research Centre (PEERC), School of Economics, University of Johannesburg, Johannesburg, South Africa. Email: jdikgang@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., 2010). One of the most commonly used statistical tools for determining productivity and efficiency is the efficiency frontier model.

There are generally two types of techniques used in frontier analysis, namely, non-parametric approaches (i.e. mathematical programming) such as data envelopment analysis or DEA (see Charnes et al., 1978; Farrell, 1957); and parametric (i.e. econometric) 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. In estimating efficiency, 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, as well as its ability to consider returns to scale. On the other hand, the strength of SFA is in its ability 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, whereas other regulators choose the highest out of the DEA and SFA estimates (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 economies such as Costa Rica, Panama and Honduras (in Central America) 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). In Asia and the Pacific region, examples include Thailand, India, Indonesia and Malaysia. 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 then repeated continuously, as the organisation continuously improves its products and services.

Although some developing countries have tried benchmarking, most are still lagging. Those that do perform benchmarking often use performance indicators, which provide the ratio of an input to an output and vice versa (for example, total debt to total assets, or workers to number of connections). Benchmarking using the ratio of one input to a single output lacks scientific rigour and does not accurately portray the overall performance of the utility (Greenberg and Nunamaker, 1987). But in developing countries, the application of rigorous and more robust tools such as DEA and SFA is often limited to academics.

The main objective of this paper is to compare efficiency results between parametric, non-parametric and semi-non-parametric approaches, using the South African water sector as a case study. This methodological cross-checking process using three methodological tools 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. Furthermore, to the best of our knowledge, this is the first application of the StoNED approach to the water sector. Given the recent drought in South Africa which resulted in water crises, the efficiency analyses for the water sector in South Africa is very important.

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 and discusses the results. Section 8 concludes the study.

2. The three frontier efficiency tools

Benchmarking is a process of comparing the performance of one decision-making unit (DMU) with best practice among all peer DMUs (Zhu, 2014). Predominantly, the rationale for benchmarking is to promote competition, encourage information-sharing and transparency, give performance trends, and provide accurate information to decision-makers. Parametric and non-parametric efficiency analysis techniques have become strategic tools for benchmarking activities. Parametric techniques include corrected ordinary least squares or COLS (Winsten, 1957), parametric programming (Timmer, 1971), and SFA. Non-parametric methods include convex non-parametric least squares or CNLS (Hildreth, 1954), corrected concave non-parametric least squares or C^2NLS (Kuosmanen and Johnson, 2010), DEA, and StoNED. Of these methods, DEA and SFA (and recently StoNED) are the most commonly used approaches for benchmarking utilities. In this section, we discuss these three efficiency analysis techniques.

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. If there are data on K inputs and M outputs on each of N decision-making units (DMUs), for the i^{th} DMU these variables are represented by the vectors x_i and y_i respectively. The KxN input matrix (x) and the MxN output matrix (y) represent the data for all N DMUs. For each DMU, 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. Optimal weights are selected by specifying the following mathematical problem:

$$\max_{u,v}(u'y_i/v'x_i),$$

$$\text{s.t.} \quad u'y_j/v'x_j \leq 1, j = 1,2,\dots,N,$$
$$u,v \geq 0 \qquad (1)$$

The process involves obtaining values for u and v such that the efficiency measure of the i^{th} DMU 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 $\mathrm{v}'\mathrm{x}_\mathrm{i} = 1$ is imposed. The imposed constraint provides that:

$$
\begin{aligned}
&\max_{\mu,\mathrm{v}}(\mu'\mathrm{y}_\mathrm{i}),\\
&\text{s.t.} \quad \mathrm{v}'\mathrm{x}_\mathrm{i} = 1,\\
&\qquad \mu'\mathrm{y}_\mathrm{j} - \mathrm{v}'\mathrm{x}_\mathrm{j} \le 0, \mathrm{j} = 1,2,\dots,\mathrm{N},\\
&\qquad\qquad \mu, \mathrm{v} \ge 0
\end{aligned}
\tag{2}
$$

This form is called the multiplier form of the linear programming problem, where the change of notation from u and v to μ and v reflects the transformation. When duality in linear programming is used, an equivalent envelopment form of the programming problem is derived. The equivalent envelopment form is presented as:

$$
\begin{aligned}
&\min_{\theta,\lambda} \theta,\\
&\text{s.t.} \quad -\mathrm{y}_\mathrm{i} + \mathrm{Y}\lambda \ge 0,\\
&\qquad \theta\mathrm{x}_\mathrm{i} - \mathrm{X}\lambda \ge 0,\\
&\qquad\qquad \lambda \ge 0
\end{aligned}
\tag{3}
$$

where θ is the scalar and λ is an $Nx1$ vector of constants. An envelopment form of this nature includes lesser constraints than the multiplier form $(K + M < N + 1)$. The obtained value of θ will be the efficiency score of the i^{th} DMU. This score satisfies $\theta \le 1$, with a value of 1 indicating a point on the frontier, that is, a technically efficient DMU (Farrell, 1957). Best-practice utilities are relatively efficient, shown by a DEA efficiency rating of $\theta = 1$, while inefficient utilities are shown by an efficiency rating of less than 1 (i.e. $\theta < 1$). DEA provides an efficiency rating that is generally between zero and 1, interchangeably referred to as an efficiency percentage between the range of 0 and 100%. The DEA model proposed by Charnes et al. (1978) had an input orientation that assumed constant returns to scale (CRS).

To estimate efficiency in South African water utilities, our study adopts the input-oriented assumption proposed by Charnes et al. (1978). However, since utilities in South Africa 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 variable returns to scale (VRS) assumption in our DEA estimation. More precisely, we estimate an input-oriented DEA that assumes VRS.

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:

$$y = \boldsymbol{\beta}'\mathbf{x} + v - u, \tag{4}$$

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 normal-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). For the sake of simplicity, our study uses the normal-half normal SFA model specification to estimate efficiency in South African water utilities. The study estimates an SFA cost function where total operation cost of providing water services (TC_i) is 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). Therefore, the SFA functional form assumes:

$$lnTC_i = \alpha_i + lnQ_i + lnMAINS_i + lnCON_i - u_i + v_i \tag{5}$$

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 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 cost function, the inefficiency term, and the random noise term. This is specified as follows:

$$u_i = \alpha_1 + \alpha_2 lnCON_i + \delta_i \tag{6}$$

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 metered and unmetered water connections. 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. $v_i = \alpha_1 + \alpha_2 \mathrm{lnPOP}_i + \delta_i$). In doing this, we control for bias estimates that could have emanated from heteroscedasticity in the noise term, and biased efficiency scores because of heteroscedasticity in the inefficiency term.

2.3. Stochastic non-parametric envelopment of data (StoNED)

Developed by Kuosmanen and Kortelainen (2012), StoNED combines the axiomatic, non-parametric, piecewise linear DEA-style frontier with a stochastic SFA-style treatment of inefficiency and noise. Combining a linear DEA-style frontier with a stochastic SFA-style treatment of inefficiency and noise makes StoNED more robust to both model misspecification and noise. The method has two main stages. The first stage estimates the shape of the total cost 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:

$$lnTC_i = lnf(\boldsymbol{y}_i) + \varepsilon_i = lnf(\boldsymbol{y}_i) - u_i + v_i \tag{7}$$

The main challenge in the least squares estimation of equation 7 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 7 as:

$$lnTC_i = (lnf(\boldsymbol{y}_i) - \mu)(\varepsilon_i + \mu) = lng(\boldsymbol{y}_i) + v_i \tag{8}$$

$$\hat{\varepsilon}_i = \hat{v}_i - \hat{\sigma}_u\sqrt{2/\pi}$$

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 WSPs 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:

$$\underset{f,v}{Min}\sum_{i=1}^{n} v_i^2 \left| lnTC_i = lng(\boldsymbol{y}_i) + v_i \qquad \forall_{\,i} = 1, \dots, n \right.$$

$$\text{s.t.} \qquad g \in F_2 \tag{9}$$

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:

$$\underset{\boldsymbol{y},\boldsymbol{\beta},v}{Min}\sum_{i=1}^{n} \left(v_i^{CNLS}\right)^2$$

$$\begin{aligned} s.t. \quad & lnTC_i = \alpha_i + \boldsymbol{\beta}_i' ln\boldsymbol{y}_i + v_i^{CNLS} && \forall_i = 1, \dots, n \\ & \alpha_i + \boldsymbol{\beta}_i'\boldsymbol{y}_i \leq \alpha_h + \boldsymbol{\beta}_h'\boldsymbol{y}_h && \forall_i, \forall_h = 1, \dots, n \\ & \boldsymbol{\beta}_i \geq 0 && \forall_i = 1, \dots, n \\ & && g \in F_2 \end{aligned} \tag{10}$$

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:

$$M_2 = \left[\frac{\pi-2}{\pi}\right]\sigma_u^2 + \sigma_v^2, \qquad M_3 = \left(\sqrt{\frac{2}{\pi}}\right)\left[1 - \frac{4}{\pi}\right]\sigma_u^3. \tag{11}$$

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

$$\widehat{M}_2 = \sum_{i=1}^{n}\frac{(\hat{v}_i - \hat{E}(v_i))^2}{n}, \quad \widehat{M}_3 = \sum_{i=1}^{n}\frac{(\hat{v}_i - \hat{E}(v_i))^3}{n}. \tag{12}$$

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:

$$\hat{\sigma}_u = \sqrt[3]{\frac{\widehat{M}_3}{\left(\sqrt{2/\pi}\right)\left[1-\frac{4}{\pi}\right]}}. \tag{13}$$

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

$$\hat{\sigma}_v = \sqrt[2]{\widehat{M}_2 - \left[\frac{\pi-2}{\pi}\right]\hat{\sigma}_u^2}. \tag{14}$$

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:

$$ln\hat{f}(\boldsymbol{y}_i) = ln\hat{g}_{min}(\boldsymbol{y}_i) + \hat{\sigma}_u\sqrt{2/\pi} \tag{15}$$

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's (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:

$$\hat{E}(u_i|\hat{\varepsilon}_i) = \mu_* + \sigma_* \left[\frac{f(-\mu_*/\sigma_*)}{1 - F(-\mu_*/\sigma_*)}\right] \tag{16}$$

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:

$$\hat{E}(u_i|\hat{\varepsilon}_i) = \frac{\hat{\sigma}_u\hat{\sigma}_v}{\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}} \left[\frac{f\left(\frac{\hat{\varepsilon}_i\hat{\sigma}_u}{\hat{\sigma}_v\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\right)}{1 - \mathrm{F}\left(\frac{\hat{\varepsilon}_i\hat{\sigma}_u}{\hat{\sigma}_v\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\right)} - \frac{\hat{\varepsilon}_i\hat{\sigma}_u}{\hat{\sigma}_v\sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}}\right] \tag{17}$$

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 17 is an unbiased but inconsistent estimator of u_i; and irrespective of the sample size, each DMU 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's (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. 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). Private participation in the water sector is also increasing in emerging economies (see Carvalho et al., 2015; Estache and Rossi, 2002; Souza et al., 2007).

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. The delivery of water services is dependent on a sequential process along a value chain. 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 sequential interrelation of key players in the South African water sector value chain is shown in Figure 1.

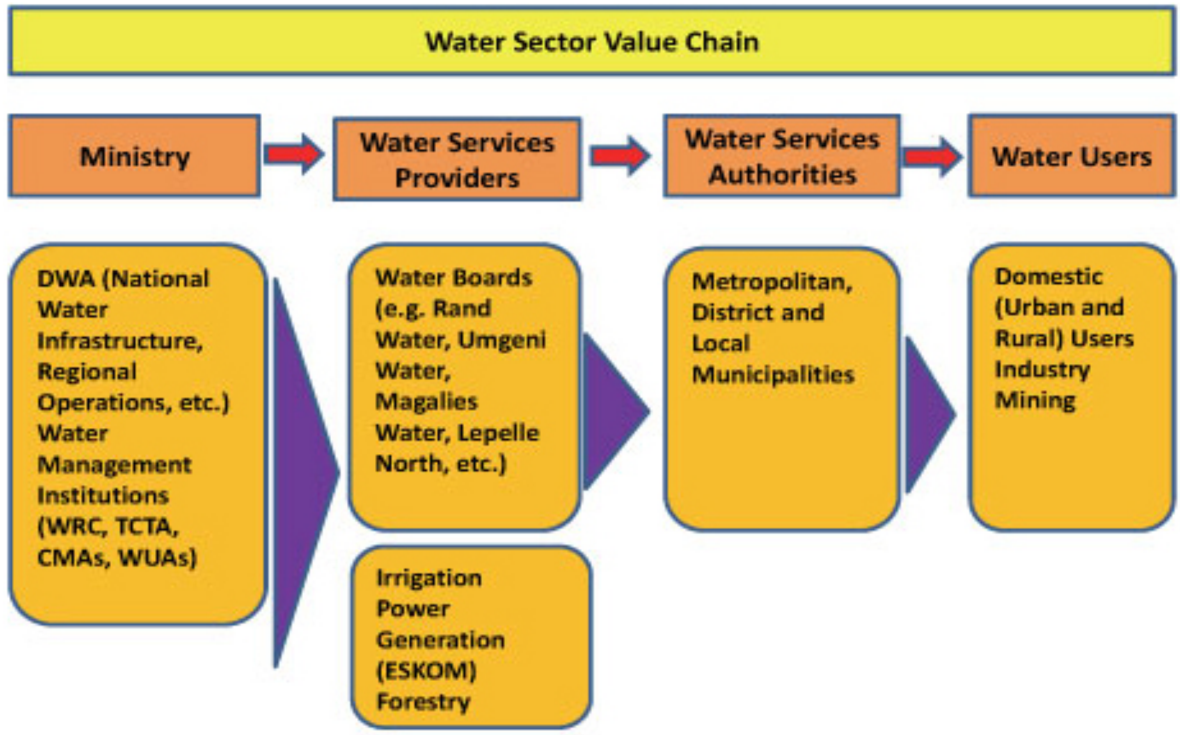


Figure 1: The water sector value chain in South Africa

Source: Ruiters (2013)

In the figure, Ministry refers to the DWS, which is the custodian of the country's water resources. The DWS 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 for both water resources and services. The DWS' regulatory role also includes supporting institutions in complying with existing regulations and institutional reforms, as well as 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. The Minister of

Cooperative Governance and Traditional Affairs (CoGTA) is responsible for determining which municipalities qualify to be WSAs. The 152 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. This usually occurs when the local municipality is considered a 'secondary city' (i.e. a local municipality with a large town or city as its urban core). 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].

Although these arrangements have some merit, the ultimate delivery of water services in the country can be complicated. This is apparent where WSAs have the legal option to appoint a third party to provide all or part of the water services on their behalf. Section 76 of the Municipal Systems Act (Act 32 of 2000) differentiates between internal and external service-delivery mechanisms. The former is the delivery of water services by a department, administrative unit or business unit within a municipality. The latter includes the partial or complete outsourcing or commercialising of the delivery of water services. External delivery of a service by an authorised WSP would include outsourcing the service to another municipality, municipal entity, organ of state or the private sector, through commercialising the delivery of the service or by public-private partnerships (PPPs).

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

[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 'independence' from South Africa.

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., 2010; da Cruz, et al., 2013; Guerrini et al., 2015). These tools estimate technical and cost efficiency scores that are essential to performance benchmarking. 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).

Although WSAs are the legal custodians of water services within their jurisdictions, the methods used to deliver water services vary. Using scientific efficiency-analysis tools to benchmark municipal performance can deal with the inconsistences in the way water is provided across municipalities in the country. Efficiency-analysis methods can cater for instances where there are differences in treatment methods, or instances in which both a district and a local municipality are providing water, even though only the former is authorised. These discrepancies are accounted for in the inputs used and outputs produced, thus making technical and cost efficiency analysis a more accurate benchmark than other one-dimensional or ordinal methods.

4. Literature review

Most water-efficiency studies use either DEA or SFA. 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; Cruz, 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, 2010; 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 literature on studies that use the approach to estimate efficiency in the water sector. To the best of our knowledge, the performance of StoNED has not been tested in benchmarking water utilities. The approaches common in the water literature are the conventional efficiency analysis techniques (i.e. DEA and SFA); they have also been extensively adopted by water regulators to benchmark utilities. A study closer to ours is Kuosmanen et al. (2013), which compared DEA, SFA and StoNED in the electricity distribution sector, and found that StoNED yielded the most precise efficiency scores in both heterogeneous and noisy samples.

By comparing efficiency estimates from DEA, SFA and StoNED in the South African water sector, our study becomes one of the first to use these three methods in the context of the water sector, especially in developing countries. Developing countries are mostly associated with inconsistent and inaccurate data, which make it difficult to use efficiency-analysis tools such as DEA, SFA and StoNED that require consistent data. However, some attempts have been made (see Brettenny and Sharp, 2016; Carvalho et al., 2015; Souza et al., 2007; Vishwakarma and Kulshrestha, 2010). These studies use either DEA or SFA, and report inefficiencies in the water utilities of developing countries. Considering that no study has used StoNED to compare the consistency of water efficiency levels, our study extends the current literature in that arena.

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 cost 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:

$$TC_i = f\left(Q_i^{\alpha} MAINS_i^{\beta} CON_i^{\gamma} POP_i^{\Omega}\right). exp(v_i + u_i) \tag{18}$$

where u_i is a random variable representing the cost 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 18 can be written as:

$$TC_i = f(\boldsymbol{y}_i). exp(v_i - u_i) \tag{19}$$

If a composite error term $\varepsilon_i = v_i - u_i$, which consists of an inefficiency term $u > 0$, and a random parameter term $v = 0$ is introduced, and the cost function is linearised in logs, then equation 19 will be rewritten as:

$$lnTC_i = lnf(\boldsymbol{y}_i) + \varepsilon_i = lnf(\boldsymbol{y}_i) - u_i + v_i \tag{20}$$

The cost function presented in equation 20 is estimated using DEA, SFA and StoNED. For DEA, we estimate an input-oriented DEA that assumes VRS. In doing this, we first estimate efficiency scores for the whole sample (consisting of all water utilities in the sample). Then we group utilities based on their sizes and estimate efficiency scores for each category. Grouping utilities is essential, since DEA is very susceptible to the influence of outliers (Banker, 1993). The South African water sector has variations in the sizes and operating environments of utilities; as such, categorisation is essential. Efficiency estimates from pooled data are compared to those derived from grouped utilities. Eventually, DEA efficiency scores will be compared to scores from the other two methods.

StoNED has two main process stages: the first stage estimates the shape of the cost function, while the second stage estimates inefficiencies (Kuosmanen et al., 2013). This study does not report on the shape of the cost frontier but presents results on the utility-specific efficiency scores, which are then compared to estimates from the other two models. This approach is in line with the objective of the study. We present results from both the MM and PSL estimators of StoNED; but for further analysis, we adopt the estimator with efficiency scores that have less variance around the mean. Kuosmanen (2012) suggests that the MM is unbiased and consistent, while Andor and Hesse (2014) suggest that PSL gives more robust estimates. Therefore, for the StoNED model we present scores from both MM and PSL, and subsequently compare their variance around the mean. A complete list of all DMU-specific efficiency scores will be presented for DEA, SFA and StoNED – that is, in addition to the separate comparison of efficiency scores for big and small water utilities.

6. Raw data

The sample in our study comprises cross-sectional data for the 2013/4 period for 102 water utilities. This implies that we have just one (i.e. annual) observation per utility. 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), which follows the intermediation approach (defining input and output variables), treating Chinese banks as multi-product firms that employ inputs X_i at given prices W_i that minimise total costs TC 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[6]. (Similar variables are also used in Kuosmanen (2012), in the context of electricity distribution utilities).

[6] In the context of this study, a water utility refers to a WSA that provides water services to final users.

Total cost (TC) is the total water-related operating cost[7] for each water utility. The total cost data is expressed in South African Rands[8], 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. The rationale for using operating cost is motivated by the reality that operating cost gives a true reflection of the actual costs of running a water services department each year.

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. Water utilities buy bulk raw water from water boards and are charged per kilolitre for the quantity bought. Hypothetically, the higher the quantity of water supplied by a utility, the higher will be the total costs, *ceteris paribus*. Therefore, utilities are expected to minimise the cost of providing water services, given a certain quantity supplied.

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. Hypothetically, more connections for a utility implies more consumer units, which may result in higher costs of providing water services, *ceteris paribus*. Therefore, the rationale for each water utility is to minimise its cost given a certain number of water connections.

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. Hypothetically, utilities with longer pipe networks incur more costs from water losses and from transporting water over long distances. As such, water utilities are

[7] A water utility typically distributes water and other related services. However, in the case of South Africa, this is the responsibility of a local municipality which also provides other services such as refuse removal, electricity, and street lighting. Considering local municipalities incur costs for services provided, we are only interested in "water-related" costs, hence the use of that term. Non-water-related operating costs therefore includes costs incurred towards provision of street lights, municipal roads and electricity.

[8] The Rand is the South African currency. As at 24 October 2018, US$1 = ZAR14.30.

expected to minimise the cost of providing water according to the length of the pipe network.

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 to control for heterogeneity. In the literature, population is used extensively to control for heterogeneity (see Baranzini et al., 2010; Filippini et al., 2007; Tsegai et al., 2009). The figure below shows the map of where the utilities (i.e. municipalities) are located.

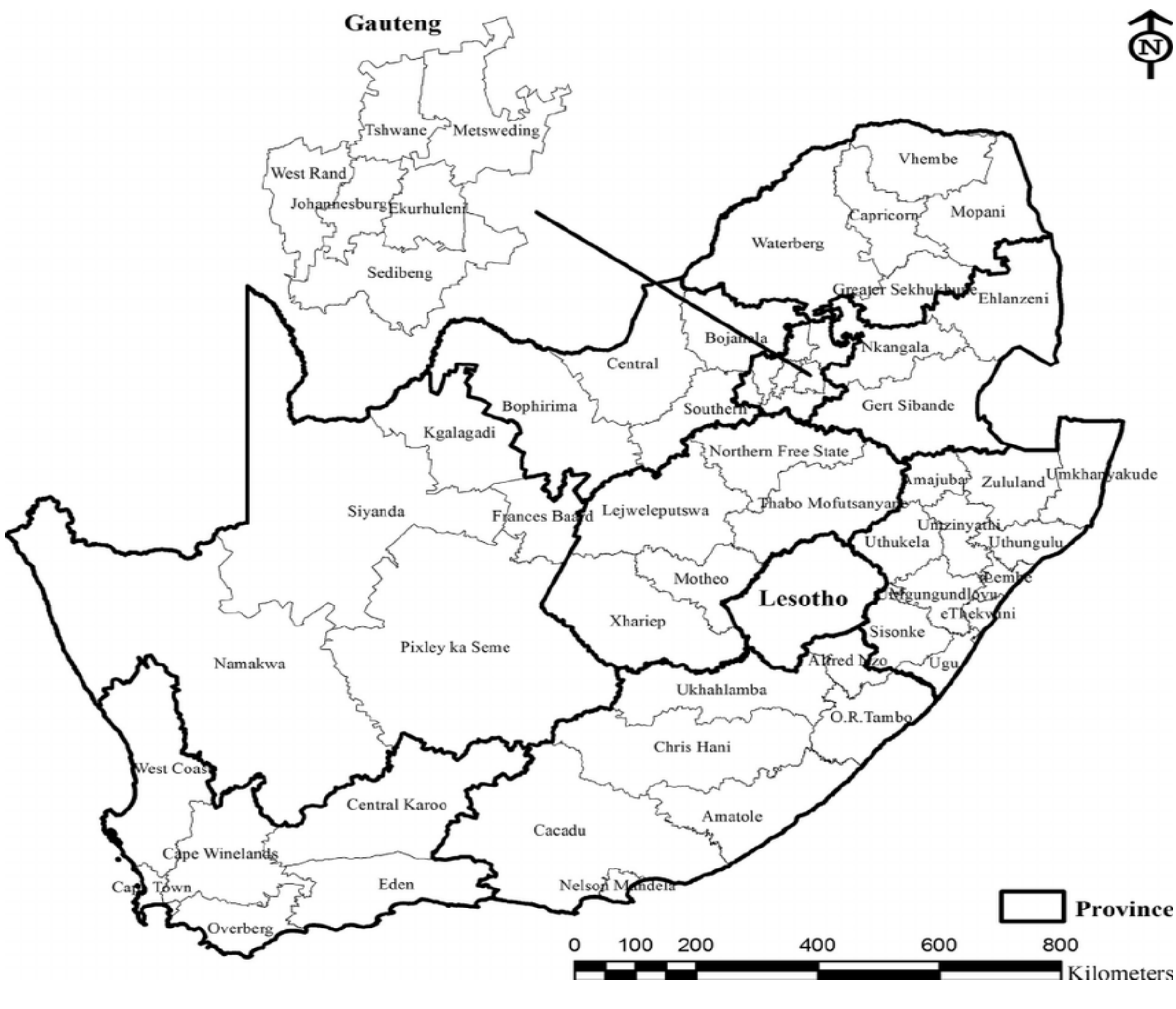


Figure 2: Map of where South African district municipalities are located
Source: Wabiri et al. (2016)

The analyses in the study are based on a sample of South African water utilities depicted in the figure above. The district municipalities depicted in the figure are further divided into various local municipalities which are also included in our analyses. 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 in thousands of Rands	286,000	813,000	990	5,380,000
Q	Quantity of water in thousands of kilolitres	22,700	56,700	350.4	352,000
CON	Total number of connections	73,543	130,269	2,306	713,143
MAINS	Length of mains in kilometres	1,485	2,593	46	12,479
POP	Population served in thousands	413.3	801.2	10.6	4,500

Notes: Observations (N) =102.

The table above 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 R990,000 to R5.38 billion, 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.

Sample size has been identified as one important factor influencing the performance of efficiency estimation methods (Andor and Hess, 2014). Where smaller samples are used, more variables should be included for each DMU, as a measure to increase the number of observations. One major weakness of few observations is that it becomes hard to get coefficient estimates with small standard errors in parametric modelling like SFA. On the other hand, DEA is affected by a variation in sample size, but the direction of the effect depends on the underlying scenario. Andor and Hess (2014) explain that in scenarios without noise, the performance of DEA improves with an increasing number of DMUs, while it deteriorates with a growing number of DMUs in noise scenarios. For StoNED, computations with more than 300 observations can take several days (see Kuosmanen, 2012; Andor and Hess, 2014). As such, the number of observations for this study may be deemed fewer for DEA but are relevant for real-world application of SFA and StoNED.

7. Results and discussion

This section presents the estimated results using the three models defined previously: the SFA, DEA and StoNED. The results are presented in four main steps. Firstly, we report on the summary statistics of the efficiency scores generated by each method. It is imperative to note that here, we present summary statistics for scores from an SFA model, two StoNED models (MM and PSL), and three DEA models (whole sample, big water utilities, and small water utilities). Secondly, we show the distribution of utility-specific scores around the mean for all methods.

After the first and second steps, we compare the standard deviations of the two StoNED methods (MM and PSL) and adopt the method with less variation for further analysis. The standard deviation is essential for showing how well each model controls for heterogeneity, where less variation around the mean efficiency score implies the model's ability to control heterogeneity. Thirdly, we present the frequency distribution of scores for each model by analysing the frequency of utilities below the mean efficiency score of each method, and the frequency of those above the mean. Finally, we group utilities into big and small water utilities and then compare the scores of each group for all models.

We extracted efficiency scores for each water utility from the SFA model and compared the SFA efficiency scores to those estimated using DEA and StoNED. Our study estimated an input-oriented DEA, which assumes VRS (because water utilities are at different levels of production). Since DEA is susceptible to the influence of outliers (Banker, 1993), we group utilities using their categories into big and small utilities and estimate efficiency scores. This allows for a comparison of water utilities of similar sizes. However, for consistency with the other estimation models used in this study, we also pooled all the utilities together and used DEA to estimate efficiency scores, which were then compared to those from big and small utilities as well as SFA and StoNED. For the StoNED analysis, we used both the MM and the PSL techniques. Summary statistics of the efficiency scores based on each method are given in Table 2.

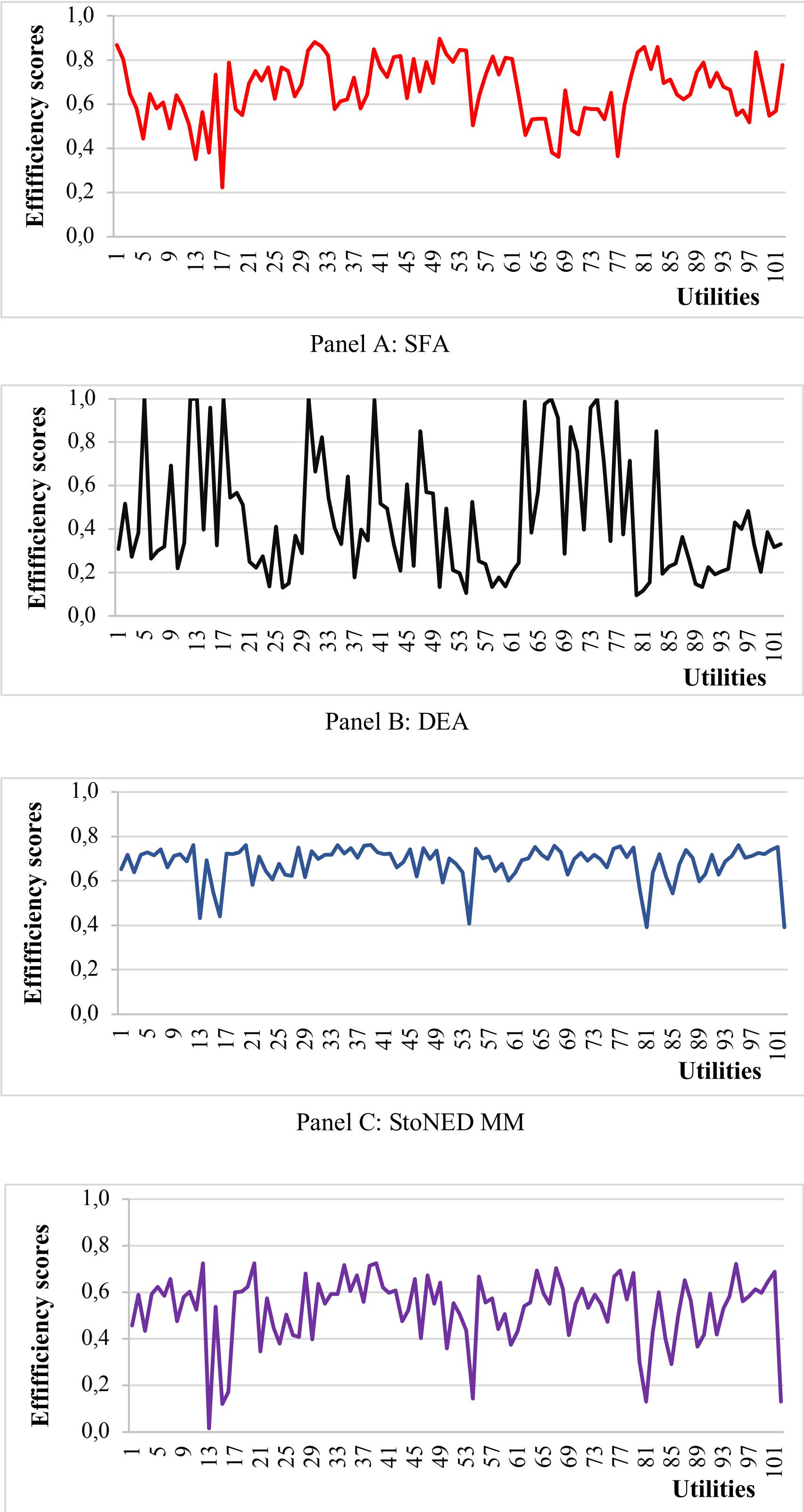
Table 2: Summary statistics of efficiency scores based on method

	StoNED (MM)	**StoNED (PSL)**	**SFA**	**DEA (All utilities)**	**DEA (Big utilities)**	**DEA (Small utilities)**
Mean	0.681	0.529	0.662	0.447	0.587	0.461
Minimum	0.396	0.018	0.223	0.095	0.154	0.104
Maximum	0.762	0.724	0.896	1.000	1.000	1.000
Std. Dev	0.079	0.146	0.139	0.280	0.256	0.289

The table shows average efficiency scores for the 102 water utilities ranging from 0.447 in DEA to 0.681 in StoNED MM. Using the StoNED MM estimate, the average efficiency score is interpreted to mean that on average, water utilities in the sample 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. When utilities are grouped together, the DEA average efficiency estimate is 45%, which is almost the same as the average estimate for small utilities (i.e. 46%), but less than the average estimate for big utilities, which is 59%.

In terms of standard deviation statistics, which express how much utility-specific efficiency scores vary from the mean score, the StoNED MM function reported the lowest standard deviation, of 0.079; followed by SFA, with a standard deviation of 0.139. In this regard, SFA performed better than the more sophisticated StoNED PSL. This confirms the findings in Andor and Hess (2014) that in noise samples, there is competition between StoNED PSL and SFA. Furthermore, it justifies why pioneering studies on StoNED, such as Kuosmanen et al. (2013) and Cheng et al. (2015), used the MM function. DEA reported the most variations which are similar across all three DEA categories. Based on the variations observed, the StoNED MM method performed better than the other models.

Although the summary statistics presented in Table 2 above give a clear snapshot of the performance of the three models, it is equally essential to examine how the efficiency scores for all the 102 water utilities in our sample were distributed around the mean. To do this, we graphically illustrate all utility-specific efficiency scores extrapolated from the three analysis models. For the DEA scores, we use estimates from the combined sample. The distribution of the efficiency scores for all water utilities around the mean is presented in Figure 3.



Panel A: SFA

Panel B: DEA

Panel C: StoNED MM

Panel D: StoNED PSL

Figure 1: Distribution of efficiency scores around the mean

The figure shows less variation around the mean in StoNED MM efficiency scores, followed by SFA. This is coherent with the summary statistics presented earlier in Table 2. The variation in scores generated by DEA was expected, because we took estimates for the whole sample; yet the method does not account for heterogeneity. Rather, DEA identifies the best-performing utilities in the sample, then compares the other utilities to the best-performing ones (see Charnes, et al., 1978; Farrell, 1957). This would then make smaller utilities with smaller populations and fewer connections appear more efficient compared to bigger utilities. Regarding StoNED, the theory assumes the PSL function to be more efficient than the MM function in estimating efficiency scores (see Fan et al., 1996). However, our estimates in both Figure 3 and Table 2 above show MM performing better than PSL in terms of variation from the mean efficiency score. Since the StoNED MM scores show less standard deviation relative to PSL scores, our subsequent analysis will be based on StoNED MM scores.

To gain more insight into each estimation technique, we present the frequency distribution of utilities with efficiency scores below the model's mean score, as well as the frequency of utilities with scores above the model's mean. This implies that scores from each model are compared to the mean for the model. The results are presented in Figure 4 below.

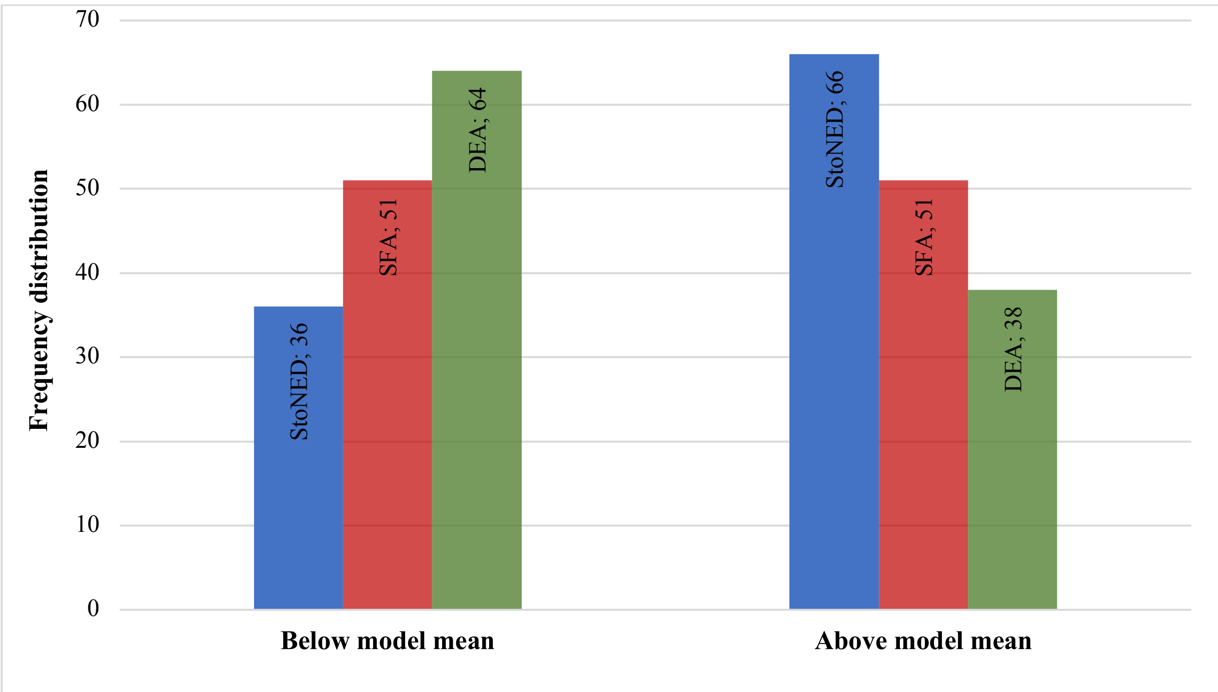


Figure 2: Frequency distribution of utilities below and above model average score

The results in Figure 4 show that using StoNED, 66 of the 102 water utilities (i.e. 65% of the utilities) scored efficiency estimates above the method's mean of 0.681. This implies that 65% of the utilities are more than 68.1% efficient under StoNED. On the other hand, SFA results show an even distribution of utilities across the two categories. Each of the two categories had 51 utilities. That is, 50% of the utilities reported efficiency scores below SFA's mean of 0.662. SFA results were somewhat coherent with those reported in StoNED. In fact, DEA reported the most inconsistent results relative to the other models. DEA scores show 64 utilities below the model's mean of 0.447, implying that 63% of the utilities reported efficiency estimates below 44.7% under DEA[9].

Since our main aim is to estimate the efficiency scores for each utility given the three methods, it is essential to present the estimated efficiency scores for each utility, using each method. Since we have a relatively large sample of 102 utilities, we therefore present efficiency scores for selected utilities from the sample. To give an intuitive snapshot of our diverse sample, we present a range of scores, from both big utilities (in cities and big towns) and small (in small towns and rural areas). The chosen utilities were randomly selected from the total of 102, and they cover the nine South African provinces. Table 3 presents efficiency scores for these selected utilities.

[9] Further analysis revealed that 5 utilities in StoNED, 3 utilities in SFA and 67 utilities in DEA recorded efficiency scores below 50%. Further analysis of the distribution of DEA scores reported for big and small utilities show that 56% of the former had scores below the mean, while 67% of the latter had scores below the mean.

Table 1: Efficiency scores for selected municipalities using the three estimation methods

Big water utilities				Small water utilities			
Utility	**StoNED**	**SFA**	**DEA**	**Utility**	**StoNED**	**SFA**	**DEA**
Nelson Mandela Bay	0.653	0.867	0.309	Camdeboo	0.644	0.647	0.272
Buffalo City	0.717	0.803	0.517	Ikwezi	0.737	0.443	1.000
Amathole	0.762	0.508	1.000	Sunday's River	0.663	0.609	0.320
Chris Hani	0.557	0.382	0.960	Baviaans	0.718	0.489	0.692
Joe Gqabi	0.724	0.223	1.000	Kouga	0.722	0.642	0.219
Mangaung	0.722	0.789	0.546	Kou-kamma	0.688	0.593	0.335
City of Johannesburg	0.735	0.843	1.000	Tsolwana	0.696	0.564	0.398
City of Tshwane	0.700	0.883	0.666	Tswelopele	0.588	0.691	0.249
Ekurhuleni Metro	0.718	0.863	0.823	Setsoto	0.649	0.706	0.275
eThekwini Metro	0.728	0.850	0.996	Mantsopa	0.680	0.623	0.411
Ugu	0.720	0.765	0.518	Richtersveld	0.712	0.461	0.988
Umgungundlovu	0.663	0.812	0.331	Karoo Hoogland	0.628	0.535	0.976
Uthukela	0.686	0.819	0.207	Umsobomvu	0.729	0.361	0.911
Amajuba	0.623	0.805	0.230	Thembelihle	0.700	0.482	0.872
Uthungulu	0.700	0.791	0.569	Siyathemba	0.729	0.464	0.756
iLembe	0.737	0.696	0.563	!Kai! Garib	0.693	0.582	0.398
Mopani	0.594	0.896	0.133	!Kheis	0.700	0.578	1.000
Vhembe	0.701	0.828	0.494	Tsantsabane	0.664	0.532	0.708
Capricorn	0.641	0.845	0.196	Dikgatlong	0.755	0.365	0.987
City of Cape Town	0.720	0.860	0.852	Gamagara	0.715	0.593	0.375
Average score	**0.690**	**0.756**	**0.596**	**Average score**	**0.691**	**0.548**	**0.607**

The table shows that the average efficiency scores for the selected big utilities lie in the range 0.596 to 0.756, while those for small utilities lie in the range 0.548 to 0.691. StoNED reported no major differences in the average scores for both big and small utilities. However, SFA reported big utilities to be more efficient (with an average efficiency score of 0.756, i.e. 75.6%) than small utilities (with an average efficiency score of 0.548, i.e. 54.8%). On the other hand, DEA reported small utilities to be relatively more efficient than big utilities. However, in the case of DEA the margin is not as large – small utilities are on average 60.7% efficient, while big utilities are on average 59.6% efficient. This revelation shows increasing returns to scale among water utilities.

Regarding the actual utility-specific efficiency scores, in many instances DEA reported very high scores (such as 100% efficiency) or very low scores, relative to the other methods. By contrast, StoNED estimates lay mostly in between the DEA and SFA scores. For example, for Nelson Mandela Bay, DEA reported an efficiency score of 0.309 while SFA reported a score of 0.867; StoNED was in between, with 0.653. The same trend is observed in utilities that were 100% efficient according to DEA. One example is Ikwezi, with a DEA score of 1, an SFA score of 0.443 and a moderating StoNED score of 0.737. In addition to the Nelson Mandela Bay and Ikwezi, the same trend is generally observed in other water utilities including Chris Hani, Joe Gqabi, Mopani, Capricorn, Richtersveld, and Tswelopele. StoNED proved to be more robust in estimating efficiency scores for our sample. This finding agrees with Kuosmanen et al. (2013), in which DEA, SFA and StoNED were compared and StoNED yielded the most precise results.

The efficiency scores reported in this study are consistent with those reported in the literature where DEA and SFA are used to estimate the efficiency of water utilities (see Brettenny and Sharp, 2016; Estache and Rossi, 2002; Horn and Saito, 2011). Brettenny and Sharp (2016) used an input-oriented DEA and 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 using SFA and the average scores were between 0.596 and 0.621. Similarly, Estache and Rossi (2002) SFA and 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.

After estimating efficiency scores, several efficiency analysis studies in the literature proceed to regress the estimated efficiency scores against the output variables. This practice is common in studies that use a two-stage DEA approach, as researchers try to ascertain the drivers of efficiency scores (see Dunghana et al., 2004; Lannier and Porcher, 2014; Ren, Li and Guo, 2017; Sharma et al., 1999; Zschill and Walter, 2012). According to Battese and Coelli (1995), one way of doing this is by including the variables in the stochastic frontier model, as presented earlier in the study. The advantages of this approach are explained in Kumbhakar and Lovell (2000).

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, our study reports four key findings.

Firstly, a comparison of standard deviations revealed that StoNED MM had the least standard deviation, followed by SFA, which outperformed StoNED PSL. DEA reported the most variations, which were similar across all three DEA categories employed in the study. Standard deviation is a key measure of robustness in the context of this study, as it expresses how much utility-specific efficiency scores within the given sample vary from the mean score. A technique that produces the least variation is deemed more robust, as it manages to control for heterogeneity in the sample. In this study DEA remained susceptible to outliers, even when we separated utilities and categorised them according to size.

Secondly, we observed that while SFA efficiency scores were distributed evenly, most utilities recorded scores above the model's mean under StoNED while most utilities reported scores below the model's mean under DEA (this was consistent even when utilities were grouped into big and small). Precisely, 65% of utilities had efficiency scores above the method's mean of 0.681 under StoNED while in SFA, 50% of the utilities reported efficiency scores above model's mean of 0.662. For DEA, only 37% of the utilities reported efficiency scores above the model's mean of 0.447. Another key observation regarding DEA is on its mean efficiency score which was below 50% efficiency. Further analysis revealed that 67 utilities under DEA had efficiency scores below 50%. This is consistent with the theory that DEA does not account for noise but treats any deviation from the frontier as inefficiency.

Thirdly, for most of the utilities, efficiency scores estimated using the StoNED model moderated those from DEA and SFA. Where DEA gave a higher efficiency score and SFA gave a lower efficiency score (and vice versa), StoNED usually gave a median score for the two. This trend was observed in most of the utilities in the sample. Fourthly, we observed some key empirical observations from the results. We noted that inefficiencies exist in the provision of water in South Africa. Average efficiency scores reported for the sample were 0.447 (DEA), 0.662 (SFA), 0.681 (StoNED MM) and 0.529 (StoNED PSL). After grouping utilities into big and small categories, we observed little variation between the average scores from big and small utilities using StoNED. However, SFA reported that big utilities were more efficient (75.6% efficient on average) than small utilities (54.8% efficient on average), whereas DEA reported that small utilities were more efficient than big utilities.

Based on the performance of StoNED MM relative to the other techniques, we join other studies in the literature in arguing that this method is more appropriate for heterogeneous samples. StoNED MM controls heterogeneity well and leads to efficiency estimates with low standard deviations even in noisy scenarios. Where StoNED MM cannot be used, we argue that SFA is the next-best efficiency-analysis tool for noisy samples. Our study shows that the use of DEA – even when water utilities are grouped according to size – is not ideal in heterogeneous samples. Utilities in developing countries operate in very distinct environments with hugely distinct budgets. In consequence, benchmarking such utilities requires using techniques that can control for heterogeneity.

One of the weaknesses of our study is that it focused only on the second stage of StoNED, which estimates utility-specific efficiency scores. 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 that future studies use the double-bootstrap DEA. 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. This shortcoming may be addressed by using the bias correcting double-bootstrap DEA.

List of references

Agrell, P., Bogetoft, P. and Tind, J. 2005. DEA and dynamic yardstick competition in Scandinavian electricity distribution. *Journal of Productivity Analysis* 23**,** 173-201.

Aigner, D., Lovell, C. K. and Schmidt, P. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics,* 6**,** 21-37.

Andor, M. and 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. and 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, R. D. 1993. Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science,* 39**,** 1265–1273.

Baranzini, A., Faust, A.-K. and Maradan, D. 2011. Water supply: Costs and performance of water utilities: Evidence from Switzerland. 13th International Water Resources Association World Water Congress, 1-4 September 2008 Montpellier.

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

Brettenny, W. and 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.

Byrnes, J., Crase, L., Dollery, B. and Villano, R. 2009. An analysis of the relative efficiency of wastewater utilities in non-metropolitan New South Wales and Victoria. *Australasian Journal of Regional Studies, The,* 15**,** 153.

Carvalho, P., Pedro, I. and Marques, R. C. 2015. The most efficient clusters of Brazilian water companies. *Water Policy,* 17**,** 902-917.

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

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

Coelli, T., Rao, D. S. P. and Battese, G. E. 2005. *An introduction to efficiency and productivity analysis,* Boston/Dordrecht/London, Kluwer Academic Publishers.

Coelli, T. J. 1995. Estimators and hypothesis tests for a stochastic frontier function: a Monte Carlo analysis. *Journal or Productivity Analysis,* 6**,** 247-268.

Corton, M. L. and Berg, S. V. 2009. Benchmarking central American water utilities. *Utilities Policy*, 17, 267-275.

Da Cruz, N. F., Carvalho, P. and Marques, R. C. 2013. Disentangling the cost efficiency of jointly provided water and wastewater services. *Utilities Policy,* 24**,** 70-77.

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

De Witte, K. and Marques, R. C. 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. and 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. and 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.

Dunghana, B. R., Nuthall, P. L. and Nartea, G. V. 2004. Measuring the Economic Efficiency of Nepalese Rice Farms Using Data Envelopment Analysis. *The Australian Journal of Agricultural and Resource Economics,* 48**,** 347-369.

Eskelinen, J. and 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. and Rossi, M. A. 2002. How different is the efficiency of public and private water companies in Asia? *The World Bank Economic Review,* 16**,** 139-148.

Estache, A., Rossi, M. A. and Ruzzier, C. A. 2004. The case for international coordination of electricity regulation: evidence from the measurement of efficiency in South America. *Journal of Regulatory Economics,* 25**,** 271-295.

Farrell, M. J. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General),* 120**,** 253-290.

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

Forsund, F. R. and Kittelsen, S. A. C. 1998. Productivity development of Norwegian electricity distribution utilities. *Resource Economics,* 20**,** 207-224.

Garcia-Valiñas, M. A. and Muñiz, M. A. 2007. Is DEA useful in the regulation of water utilities? A dynamic efficiency evaluation (a dynamic efficiency evaluation of water utilities). *Applied Economics,* 39**,** 245-252.

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

Greene, W. H. 2008. *The econometric approach to efficiency analysis*, Oxford University Press.

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

Herwartz, H. and 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.

Hjalmarsson, L. and Veiderpass, A. 1992. Productivity in Swedish electricity retail distribution. *Scandinavian Journal of Economics,* 94**,** 193-205.

Johnson, A. L. and Kuosmanen, T. 2012. One-stage and two-stage DEA estimation of the effects of contextual variables. *European Journal of Operational Research,* 220**,** 559–570.

Johnson, A. L. and Kuosmanen, T. 2015. An introduction to CNLS and StoNED methods for efficiency analysis: Economic insights and computational aspects. *Benchmarking for Performance Evaluation.* Springer.

Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Economics,* 19**,** 233–238.

Ke, W., Wei, H., Jie, W. and Ying-Nan, L. 2014. Efficiency measures of the Chinese commercial banking system using an additive two-stage DEA. *Omega,* 44**,** 5-20.

Kirkpatrick, C., Parker, D. and Zhang, Y.-F. 2006. An empirical analysis of state and private-sector provision of water services in Africa. *The World Bank Economic Review,* 20**,** 143-163.

Korhonen, P. and Syrjänen, M. 2003. Evaluation of cost efficiency in Finnish electricity distribution. *Annals of Operations Research,* 121**,** 105-122.

Kumbhakar, S. C. and Lovell, C. A. K. 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. and 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. and Johnson, A. L. 2010. Data envelopment analysis as nonparametric least squares regression. *Operations Research* 58**,** 149–160.

Kuosmanen, T. and 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. and 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, A. L. and 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.

Lewis, H. F., Mallikarjun, S. and Sexton, T. R. 2013. Unoriented two-stage DEA: The case of the oscillating intermediate products. *European Journal of Operational Research,* 559**,** 529–539.

Li, H.-Z., Kopsakangas-Savolainen, M., Xiao, X.-Z., Tian, Z.-Z., Yang, X.-Y. and 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. and Van Den Broeck, J. 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review***,** 435-444.

Mehta, D., Meera, M. and Anitha, I. 2013. A Review of Performance Benchmarking Urban Water Supply and Sanitation. CEPT University.

Mekaroonreung, M. and Johnson, A. L. 2012. Estimating the shadow prices of SO2 and NOx for US coal power plants: a convex nonparametric least squares approach. *Energy Economics,* 34**,** 723-732.

Olson, J. A., Schmidt, P. and Waldman, D. M. 1980. A Monte Carlo study of estimators of stochastic frontier production functions. *Journal of Economics,* 13**,** 67–82.

Ren, C., Li, R. and Guo, P. 2017. Two-stage DEA analysis of water resource use efficiency. *Sustainability* 9.

Ruiters, C. 2013. Funding models for financing water infrastructure in South Africa: framework and critical analysis of alternatives. *Water SA,* 39**,** 313-326.

Sabouhi Sabouni, M. and Kenari, R. E. 2014. Measuring technological gap ratio of wheat production using StoNED approach to metafrontier. *International Journal of Applied Operational Research,* 4.

Sharma, K. R., Leung, P. and Zaleski, H. M. 1999. Technical, allocative and economic efficiencies in Swine production in Hawaii: a comparison of parametric and nonparametric approaches. *Agricultural Economics* 20**,** 23-35.

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

Souza, G. S., Faria, R. C. and Moreira, T. B. S. 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, C. P. 1971. Using a probabilistic frontier production function to measure technical efficiency. *Journal of Political Economy*, 79, 776-794.

Tsegai, D. W., Linz, T. and 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. and Ferrara, G. 2015. Analyzing Italian citrus sector by semi-nonparametric frontier efficiency models. *Empirical Economics,* 49**,** 641-658.

Vishwakarma, A. and 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.

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

Wabiri, N., Shisana, O., Zuma, K. and Freeman, J. 2016. Assessing the spatial nonstationarity in relationship between local patterns of HIV infections and the covariates in South Africa: a geographically weighted regression analysis. *Spatial and Spatio-temporal Epidemiology*, 16**,** 88-99.

Westhuizen, G. and Dollery, B. 2009. Efficiency measurement of basic service delivery at South African district and local municipalities. *Journal for Transdisciplinary Research in Southern Africa,* 5**,** 162-174.

Winsten, C. 1957. Discussion on Mr. Farrell's paper. *Journal of the Royal Statistical Society,* 120, 282-284.

Yang, Z. J. 2006. A two-stage DEA model to evaluate the overall performance of Canadian life and health insurance companies. *Mathematical and Computer Modelling,* 43**,** 910-919.

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

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