UNIVERSIDADE FEDERAL DE MINAS GERAIS
Escola de Engenharia
Programa de Pós-graduação em Engenharia de Produção

Cláudio Vitor Maquiné Salvador

Modelo Estocástico de Análise Envoltória de Dados Aplicado ao Benchmarking das Empresas Brasileiras de Distribuição de Energia Elétrica no ano de 2015

Belo Horizonte
2021

Cláudio Vitor Maquiné Salvador

MODELO ESTOCÁSTICO DE ANÁLISE ENVOLTÓRIA DE DADOS APLICADO AO BENCHMARKING DAS EMPRESAS BRASILEIRAS DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA NO ANO DE 2015

Trabalho apresentado ao Programa de Pós-Graduação em Engenharia de Produção da UFMG como parte dos pré-requisitos para obtenção do título de Mestre em Engenharia de Produção

Orientador: Marcelo Azevedo Costa

Belo Horizonte

2021

S182m Salvador, Claudio Vitor Maquiné.
Modelo estocástico de análise envoltória de dados aplicado ao benchmarking das empresas brasileiras de distribuição de energia elétrica no ano de 2015 [recurso eletrônico] / Claudio Vitor Maquiné Salvador. - 2021.
1 recurso online (44 f. : il., color.) : pdf.

Orientador: Marcelo Azevedo Costa.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Bibliografia: f. 45.

1. Engenharia de produção - Teses. 2. Análise estocástica - Teses. 3. Benchmarking (Administração) - Teses. I. Costa, Marcelo Azevedo. II. Universidade Federal de Minas Gerais. Escola de Engenharia. III. Título.

CDU: 658.5(043)

Ficha catalográfica elaborada pela Bibliotecária Roseli Alves de Oliveira - CRB-6/2121
Biblioteca Prof. Mário Werneck - Escola de Engenharia da UFMG

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

UFMG

ATA DA DEFESA DA DISSERTAÇÃO DO ALUNO **CLÁUDIO VITOR MAQUINÉ SALVADOR**

Realizou-se, no dia 30 de março de 2021, às 09:00 horas, online em https://us02web.zoom.us/j/87237850102?pwd=b3l2NDd1QTlVS1hXRGlvTjh2Y09Gdz09, da Universidade Federal de Minas Gerais, a 355ª defesa de dissertação, intitulada *Modelo Estocástico de Análise Envoltória de Dados Aplicado ao Benchmarking das Empresas Brasileiras de Distribuição de Energia Elétrica no ano de 2015*, apresentada por CLÁUDIO VITOR MAQUINÉ SALVADOR, número de registro 2018702151, graduado no curso de ENGENHARIA DE PRODUÇÃO, como requisito parcial para a obtenção do grau de Mestre em ENGENHARIA DE PRODUÇÃO, à seguinte Comissão Examinadora: Prof(a). Marcelo Azevedo Costa - Orientador (DEP/UFMG), Prof(a). Magno Silvério Campos (UFOP), Prof(a). Aline Veronese da Silva (Arbor Consultoria).

A Comissão considerou a dissertação:

(X) Aprovada

( ) Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.
Belo Horizonte, 30 de março de 2021.

Marcelo Azevedo Costa

Prof(a). Marcelo Azevedo Costa (Doutor)

Prof(a). Magno Silvério Campos (Doutor)

ALINE VERONESE DA SILVA:00530656035

Assinado de forma digital por ALINE VERONESE DA SILVA:00530656035
Dados: 2021.04.19 11:11:57 -03'00'

Prof(a). Aline Veronese da Silva (Doutora)

Sou grato ao meu orientador, Professor
Marcelo Azevedo Costa pela atenção
dedicada ao longo desta pesquisa.

“Essencialmente, todos os modelos estão errados, mas alguns são úteis”

George E. P. Box

Resumo

Devido ao monopólio natural existente no setor elétrico brasileiro, são necessários mecanismos que garantem a proteção dos consumidores contra medidas que os desfavorecem, como por exemplo a alta de preços exorbitantes. Um destes mecanismos é a existência de uma agência reguladora, chamada Agência Nacional de Energia Elétrica, ANEEL. Entre suas várias funções, este órgão realiza a análise de parâmetros de produção de energia elétrica pelas 61 empresas ao longo do território nacional, chamadas de DSO (Distribution Service Operator). Atualmente, os parâmetros são utilizados para a realização de *benchmarking* das DSO's, de acordo com o modelo Data Envelopment Analysis (DEA), onde a eficiência é calculada durante o período de revisão tarifária O modelo atual conta ainda com um um procedimento *ad-hoc,* onde *bootstrap* é usado permitindo que algumas DSO's ultrapassem a fronteira de eficiência. Um novo método é proposto, utilizando o processo de *benchmarking* chamado Stochastic Data Envelopment Analysis, SDEA, que propõem uma equação paramétrica para a fronteira de eficiência, e que naturalmente permite que DSO's ultrapassem a fronteira, sem a necessidade de procedimentos adicionais. Na metodologia aplicada, foi proposto um novo modelo SDEA, e simulações demonstraram convergência entre o modelo atual utilizado pela agência reguladora e o modelo proposto, mostrando que o SDEA pode ser uma boa opção para o benchmarking no próximo ciclo tarifário.

***Keywords*:** Análise Envoltória de Dados Estocástica, Análise de Fronteira Estocástica, *benchmarking*.

Abstract

Due to the natural monopoly existing in the Brazilian electric sector, the elements that guarantee the protection of consumers against measures that disadvantage them are used, disadvantages such as the exorbitant high prices. One of these mechanisms is the existence of a regulatory agency, called the Agência Nacional de Energia Elétrica, ANEEL. Among its various functions, this agency carries out the analysis of parameters of electric energy production by the 61 companies throughout the national territory, called DSO (Distribution Service Operator). Currently, the parameters are used for benchmarking the DSO's, according to the Data Envelopment Analysis (DEA) model, where efficiency is obtained during the tariff review period. The current model also has an ad-hoc procedure, where bootstrap is used, allowing some DSOs to surpass the efficiency frontier. A new method is proposed, using the benchmarking process called Stochastic Data Envelopment Analysis, SDEA, which proposes a parametric equation for the efficiency frontier, and which naturally allows the DSO to cross the frontier, without the need for additional procedures. In the applied methodology, a new SDEA model was proposed, and simulations demonstrated convergence between the current model used by the regulatory agency and the proposed model, showing that SDEA can be a good option for benchmarking in the next tariff cycle.

Keywords**:** Análise Envoltória de Dados Estocástica, Análise de Fronteira Estocástica, *benchmarking*.

Súmario

1. INTRODUÇÃO

O órgão regulador do setor de energia elétrica brasileiro, denominado dea Agência Nacional de Energia Elétrica, ANEEL, realiza o *benchmarking* das empresas no Brasil desde 2003, onde utiliza Data Envelopment Analysis e alguns procedimentos adicionais, para garantir a proteção dos consumidores contra possíveis abusos tarifários e quedas de qualidade, devido ao monopólio de tais empresas.

1.1 ANÁLISE ENVOLTÓRIA DE DADOS ESTOCÁSTICA APLICADA AO MODELO DE BENCHMARKING DE DISTRIBUIÇÃO DE ENERGIA DE 2015

O processo de *benchmarking* analisa o mercado a alinha às melhores práticas entre empresas, visando otimizar os resultados. Atualmente, a Agência Nacional de Energia Elétrica regula o mercado e realiza o *benchmarking* em ciclos tarifários, que acontecem a cada 4 ou 5 anos, utilizando o método Data Envelopment Analysis (DEA) com retornos não decrescentes de escala, utilizando variáveis como número total de consumidores, mercado ponderado, rede de distribuição de alta tensão, rede de distribuição aérea, rede de distribuição subterrânea, perdas não técnicas, e tempo sem interrupção de serviço.

Uma alternativa a métodos de *benchmarking* conhecidos como métodos de fronteira estocástica (SFA - Stochastic Frontier Analysis) permite que os DSOs cruzem a fronteira de eficiência. Métodos de fronteira estocástica foram originalmente desenvolvidos por Aigner et al. (1977) usando uma equação paramétrica para a fronteira de eficiência e a soma de dois componentes aleatórios independentes. Um componente aleatório representa a ineficiência técnica e o segundo componente aleatório representa o componente de ruído. O componente de ruído permite que a fronteira de eficiência seja ultrapassada.

Banker (1986) propôs pela primeira vez um método estocástico não paramétrico usando DEA, conhecido como Stochastic Data Envelopment Analysis (SDEA), no qual o analista escolhe o número de pontos que cruzam a fronteira com antecedência, assim, a fronteira eficiente é estimada usando um modelo de programação linear.

Além da utilização do DEA, o regulador utiliza procedimentos adicionais que de maneira geral permitem que as eficiências ultrapassem o valor de 100%. Os procedimentos são explicados no Capítulo 2.2, e no Capítulo 2.7 é proposto um novo algoritmo, que utiliza o Stochastic Data Envelopment Analysis (SDEA) como opção de ferramenta de *benchmarking.* Os resultados do Capítulo 3 mostram que o modelo proposto possui alta relação com os modelos atuais pode ser uma boa opção para os próximos ciclos tarifarios.

Stochastic Data Envelopment Analysis applied to the 2015 Brazilian energy distribution benchmarking model

Abstract

The Brazilian energy regulator has been applying data envelopment analysis to set operating costs for distribution service operators since 2013. In addition to data envelopment analysis, further adjustments using Bootstrap and a reference efficiency, estimated as 79%, allow companies to have cost efficiencies above their observed costs. Thus, companies are allowed to cross the efficiency frontier. Similarly, stochastic frontier models also allow companies to cross the efficiency frontier. This work proposes the use of stochastic data envelopment analysis as an alternative for estimating efficient costs, thus providing a much simpler alternative. A new estimation algorithm is proposed in which the number of companies crossing the frontier comprises one important parameter. Simulation studies provide convergence evidence of the proposed model and results using the Brazilian database show that the stochastic data envelopment analysis is a promising model for upcoming tariff review cycles.

Keywords: stochastic data envelopment analysis, stochastic frontier analysis, benchmarking.

1. Introduction

The Brazilian regulator (ANEEL – Agência Nacional de Energia Elétrica) has been estimating regulatory revenues for energy distribution companies, such as regulatory operating costs, since 2003, during the first tariff review cycle (1TRC). Operating costs comprise a small part of the energy tariff and the regulatory operating cost, or efficient cost, represents the *price-cap* value that each company, hereafter named DSO (Distribution Service Operator), can charge consumers. In 2011, the regulator started to estimate regulatory operating costs using efficiency frontier methods such as corrected ordinary least squares (OLS) and data envelopment analysis (DEA). The proposed benchmarking methods estimate the efficient cost based on DSO characteristics such as number of consumers, distribution network length, energy market, non-technical losses, observed operating costs, among others. Further details about the Brazilian DSOs benchmarking models are found in Costa et al. (2015); da Silva et al. (2019) and Lopes et al. (2016).

In 2015, during the fourth tariff review cycle (4TRC), the regulator proposed only one frontier model, a DEA model, to estimate efficient costs. The current model is expected to be revised in the upcoming tariff review cycle, starting in 2020. The current DEA model uses non-decreasing returns of scale with operating costs as the input variable and number of consumers, weighted power consumption, high level network extension, low level network extension, underground network extension, non-technical losses and duration of interruption of energy as output variables. In addition, weight restrictions (Podinovski, 2004) are included in the DEA model. The total number of DSOs is 61 and the database comprises average yearly data observed for each DSO between tariff review cycles, every 4 to 5 years. Thus, the sample size is 61. Estimated cost efficiencies vary from 27% to 100%, meaning that some DSOs must reduce their observed cost by 73%. Cost efficiency is the ratio between efficient cost and observed cost. Costa et al. (2019); da Silva et al. (2019); Gil et al. (2017) and Lopes et al. (2016) have argued that the lower efficiencies are due to the lack of important output variables in the model and the lack of environmental adjustments.

In addition to the DEA model, the regulator is applying secondary *ad-hoc* adjustments. After estimating the cost efficiencies using DEA, a reference cost efficiency is estimated using the average of the cost efficiencies above 55%. In the last TRC, the reference cost was estimated as 79%. In addition, a Bootstrap simulation proposed by Simar and Wilson (1998) and Bogetoft and Otto (2010), is applied to generate confidence intervals for the cost efficiencies. Finally, the DEA cost efficiencies, the confidence intervals and the reference cost efficiency are combined generating final cost efficiencies varying from 37% to 119%. Further details are found in Technical Note 66/2015 (ANEEL, 2015). The regulator argues that fully efficient companies, i.e., DSOs with cost efficiencies of 100% estimated using the DEA model, must be rewarded for being fully efficient. Thus, their final efficiencies can be greater than 100%. This *ad-hoc* procedure also increases the minimum value of the cost efficiencies. The regulator also argues that the *ad-hoc* procedures adjusts for potential missing variables in the DEA model. One may argue that the *ad-hoc* procedure simply allows DSOs to cross the efficiency frontier.

An alternative class of benchmarking methods known as stochastic frontier methods (SFA - *Stochastic Frontier Analysis*) allows DSOs to cross the efficiency frontier. Stochastic frontier methods were originally developed by Aigner et al. (1977) using a parametric equation for the efficiency frontier and the sum of two independent random components. One random component represents the technical inefficiency and the second random component represents the noise component. The noise component allows the crossing of the efficiency frontier.

Banker (1986) first proposed a non-parametric stochastic method using DEA, known as Stochastic Data Envelopment Analysis (SDEA) in which the analyst chooses the number of points crossing the frontier in advance, thus the efficient frontier is estimated using a linear programming model. Banker and Maindiratta (1992) presented the use of maximum likelihood to estimate the SDEA model, and claims that *SDEA with multiple-outputs is somewhat harder to solve*. Therefore, *we leave the development of efficient solution method to future research*. Alternatively, Kuosmanen and Kortelainen (2012) proposed the Stochastic Non-Smooth Envelopment of Data (StoNED) which can be seen as an SDEA based on maximum likelihood, but with an estimation algorithm based on Modified Ordinary Least Squares (MOLS) (Greene, 1980). The StoNED was adopted to regulate electricity distribution companies in Finland in 2012 (Kuosmanen et al., 2013), achieving better performance than DEA and SFA.

Both SDEA and StoNED are semi-parametric frontier models that combine a piecewise linear efficiency frontier and a stochastic homoskedastic composite error. Recently, Jradi and Ruggiero (2019) compared the deterministic and stochastic DEA frontier using simulations, considering the error component as normally distributed and the inefficiency as half-normally distributed. The authors proposed an algorithm to estimate the efficiency frontier using maximum likelihood.

The present work proposes the use of SDEA to estimate the Brazilian DSO cost efficiencies. A new algorithm is presented based on the work of Jradi and Ruggiero (2019). Simulation studies provide convergence properties of the proposed algorithm under different returns of scale assumptions. Results show that the proposed SDEA model achieves similar cost efficiencies as compared to the ANEEL *ad-hoc* methodology. Therefore, we advocate the use of SDEA in upcoming tariff review cycles.

This paper is organized as follows. Section 2 presents the Brazilian electricity benchmarking model, the literature review and the proposed Stochastic Data Envelopment Analysis algorithm. Section 3 presents the simulation results and the case study. Section 4 presents the conclusion.

2. Materials and Methods

2.1. Historical background

The Brazilian electricity distribution sector comprises a monopoly market with 63 companies. Each company, or distribution service operator provides service to its own concession area within the 26 Brazilian states. To protect consumers from abusive tariffs, energy regulation is provided by the regulator agency (ANEEL). Until 1993, every consumer would paid the same energy price regardless of the state. Companies with negative revenues would get subsidies from the federal government. In 1994, different prices were set for each company, based on their own characteristics such as number of consumers, length of the distribution network and cost of the energy. In 2011, ANEEL started to apply benchmarking methodologies to regulate prices, so that the energy costs could be covered by the revenues, while protecting the consumers from abusive tariffs. The methodology for the estimate of energy tariffs is reviewed every 4 or 5 years in a process named tariff review cycle (TRC). In the beginning of the TRC, tariff prices are defined for each company. Thus, companies can optimize their costs and reach profitability. At the end of the TRC, tariff prices are revised and new values are defined. The fourth TRC was concluded in 2015 and the duration of the cycles are set individually for each company, making the process more effective.

2.2. The 2015 Brazilian electricity distribution regulation model

During the fourth tariff review cycle (2015-2018), the regulator applied a DEA-NDRS model to calculate cost efficiencies for each DSO. The Brazilian database is provided by ANEEL and comprises information about 61 electricity distribution companies. The input variable is operational cost. The outputs variables are number of consumers, weighted power consumption, high level network extension, low level network extension (network distribution), underground network extension, non-technical losses (energy loss) and duration of interruption of energy (amount of time without electricity service), as previously mentioned. The database comprises mean values from 2014 to 2016. A large number of companies has efficiency equal to one, thus weight restrictions are imposed on the DEA-NDRS model, limiting upper and lower values on the trade-offs between inputs and outputs.

The DEA-NDRS model, which is currently used by ANEEL (ANEEL, 2015), is shown in Equation 1.

$$
\begin{aligned}
&\max_{u,v,\varphi} h_0 = \textstyle\sum_{j=1}^{m1} v_j y_j^0 + \sum_{i=1}^{m2} v_i(-y_i^0) + \varphi \\
&\text{subject to:} \\
&u \cdot x^0 \leq 1, \\
&\textstyle\sum_{j=1}^{m1} v_j y_j^n + \sum_{i=1}^{m2} v_i(-y_i^n) + \varphi - u \cdot x^n \leq 0, \quad n = 1, 2, ..., N \\
&-v_r + \alpha_r u \leq 0, \quad r = 1, ..., R \\
&+v_t - \beta_t u \leq 0, \quad t = 1, ..., T \\
&u, v_j, \varphi \geq 0
\end{aligned}
\tag{1}
$$

where h_0 is the efficiency of the DSO under analysis, N is the total number of DSOs, $m1$: is the total number of positive outputs, $m2$ is the total number of negative outputs, y_j^n is the j-th output of DSO n, x_i^n is the i-th input of DSO n, u_i is the input parameter, v_j is the j-th output parameter, φ: is the scale parameter, α_r is the lower bound weight restriction between the parameters v_r and u, β_t is the upper bound weight restriction between the parameters v_t and u, R is the total number of lower bound weight restrictions and T is the total number of upper bound weight restrictions.

Table 1 shows the weight restrictions where u is the input parameter, related to the operational cost, and v's are the output parameters. Further details about the DEA-NDRS are available in Technical Note 162/2017-SRM/ANEEL (ANEEL, 2017).

Table 1: Trade-offs, i.e., weight restrictions between input and outputs variables imposed by ANEEL in the 2015 DEA-NDRS model.

Trade-offs	Lower and upper bounds (weight restrictions)
Input versus Network Distribution	$580 \leq \frac{v_{netdist}}{u} \leq 2200$
Underground Network versus Network Distribution	$1.00 \leq \frac{v_{undernet}}{v_{netdist}} \leq 2.00$
High Level Network versus Network Distribution	$0.40 \leq \frac{v_{highnet}}{v_{netdist}} \leq 1.00$
Input versus Total number of consumers	$30 \leq \frac{v_{cons}}{u} \leq 145$
Input versus Delivered MWh	$1 \leq \frac{v_{MWh}}{u} \leq 60$
Input versus Non-Technical Losses	$10 \leq \frac{v_{NonTechLoss}}{u} \leq 150$
Input versus Interrupted services	$\frac{v_{interrupt}}{u} \leq 2$

After calculating the efficiency scores, using the DEA-NDRS model with weight restrictions, as shown previously, the regulator applies additional steps to calculate the final cost effciencies (ANEEL, 2017). First, a confidence interval for the efficiency score is estimated for each DSO using a bootstrap method. Thus, lower (θ^i_{inf}) and upper (θ^i_{sup}) bounds for the efficiencies are estimated, i.e., $\theta^i_{inf} \leq \theta^i \leq \theta^i_{sup}$. Although the cost efficiency methodology is applied separately for each DSO at different years, under the hypothetical scenario in which all cost efficiencies (for all DSOs) are estimated in the first year of the TRC, it can be shown that the final efficiencies are calculated using Equation 2.

$$\theta^i_{final} = min(max(1, \theta^i_{inf}/\theta_{ref}), \theta^i_{sup}/\theta_{ref}) \quad (2)$$

where θ^i_{final} is the new adjusted efficiency score for the i-th DSO, θ^i_{inf} is lower bound of the cost efficiency, θ^i_{sup}: the upper bound of the cost efficiency and θ_{ref} is the reference score, estimated in the 4TRC as 0.79 (79%).

As mentioned, the value of θ_{ref} is calculated as the mean value of the efficiencies greater than 0.55 (55%), generated by the DEA-NDRS model with weight restrictions. The complete procedure (DEA-NDRS + Bootstrap + Equation 2) generates larger efficiencies as compared to the originals. In some cases, final efficiencies are greater than 1 (100%), which comprises DSOs crossing the efficiency frontier. Briefly, a DSO with bootstrap lower bound greater than the reference score has its efficiency score calculated as $\frac{\theta^i_{inf}}{\theta_{ref}}$. Consequently, the final efficiency is greater than 1. A DSO with efficiency within the interval $\theta^i_{inf} \leq \theta_{ref} \leq \theta^i_{sup}$ has its final efficiency equal to 1. Finally, a DSO with bootstrap upper bound lower than the reference score, $\theta^i_{sup} < \theta_{ref}$ has its efficiency calculated as $\frac{\theta^i_{sup}}{\theta_{ref}}$. Consequently, the final efficiency is greater or equal to the original.

In the last tariff review cycle (2014-2016), the DEA-NDRS resulted in 6 companies with efficiency scores equal to 1. After the recalculation, 11 companies (18.03%) achieved efficiencies greater than 1, and 15 companies (24.59%) achieved efficiencies equal to 1. Thus, 42.62% of the companies achieved efficiencies greater or equal to 1. This is illustrated in Figure 1 which compares the cost efficiencies calculated by the DEA-NDRS and using Equation 2, as proposed by ANEEL. The DSOs were sorted in increasing order of the DEA-NDRS cost efficiencies. The horizontal line represents the 100% cost efficiency. Thus, DSOs located below the horizontal line comprise companies with efficiencies below 100%. Points located on the horizontal line comprise companies with efficiencies of 100%, and points located above the horizontal line comprise companies with efficiencies greater than 100%. In this scenario, the Stochastic Data Envelopment Analysis (SDEA) can be a more suitable option, since the SDEA allows a portion of DSOs to cross the frontier. Furthermore, SDEA has the advantage of estimating linear equations to calculate efficiency cost for each DSO. Thus, the analyst can compare the linear coefficients among DSOs, evaluating the variables affecting their efficient costs.

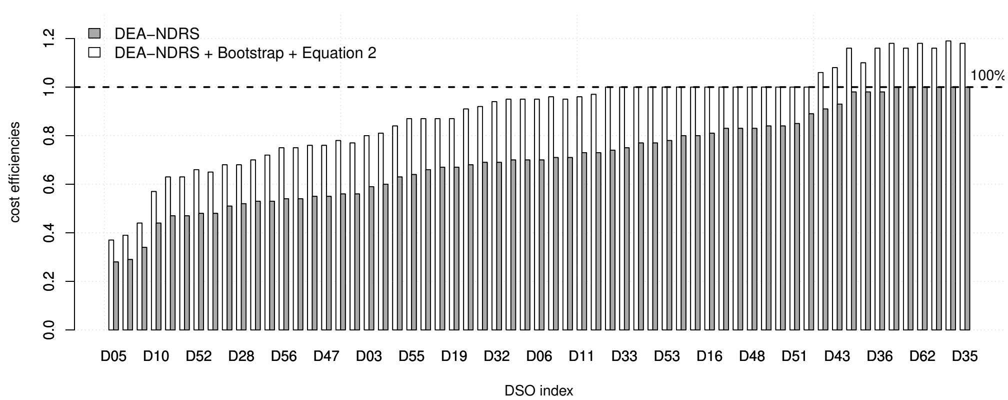


Figure 1: Comparison between cost efficiencies using the brazilian DEA-NDRS model with weight restrictions and the procedure using DEA-NDRS + Bootstrap + Equation 2.

2.3. Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a benchmarking tool proposed by Charnes et al. (1978) and extended by Banker et al. (1984), applied worldwide, which uses mathematical linear programming to measure the efficiency of DSOs using input and output variables. In general, DEA can be applied to minimize inputs or maximize outputs. Using an input-oriented approach, the DEA evaluates whether a DSO can reach the same outputs with fewer inputs. Using an output-oriented approach, DEA evaluates whether a DSO can produce more outputs with the same amount of inputs. In both cases, a DSO is fully efficient if there is no need either to minimize inputs or maximize outputs (Cook et al., 2014).

DEA models can assume different returns to scale properties. Constant returns to scale (CRS) or variable returns to scale (VRS) are the most common. The choice of the orientation must rely on the data and the objectives of the research (LaPlante and Paradi, 2015). Furthermore, DEA has a major advantage which is the non-parametric estimate of the frontier, i.e., without specifying the parametric equation of the production or cost function. Further details about DEA are found in Cook et al. (2014), Bogetoft and Otto (2010) and elsewhere.

2.4. Stochastic Frontier Analysis

DEA and SFA (Stochastic Frontier Analysis) have been used for both managerial and economic research, mainly in the last decade. SFA is more widely used in Economics. Lampe and Hilgers (2015) claim that DEA research activity is not as fast to adopt new concepts as SFA. SFA is a stochastic frontier model first proposed by Aigner et al. (1977) and Meeusen and van Den Broeck (1977), which has the advantage of distinguishing two type of errors: inefficiency and noise. The structure of the compound error for production frontier is given by $\epsilon_i = v_i - u_i$, where v_i and u_i are independent random variables. v_i is normally distributed $v_i \sim N(0, \sigma_v^2)$ and u_i follows a one-sided distribution, such as a half-normal distribution, $u_i \sim |N(0, \sigma_u^2)|$. For cost frontier, $\epsilon_i = v_i + u_i$.

In order to apply SFA, the parametric equation of the efficiency frontier must be specified. For production frontier the Cobb-Douglas function $C(y_j, ..., y_m) = \beta_0 y_1^{\beta} \times ... \times y_m^{\beta_m}$ (Jondrow et al., 1982) is widely applied, where y_j are the outputs and $\beta_0, \cdots \beta_m$ are the parameters of the production function. In the case of cost function, the Translog function (Christensen et al.,

1975) can be applied. However, even using the Translog function, the assumption of monotonicity and convexity of the cost function can be violated, creating perverse incentives to produce less outputs to improve the efficiency (Kuosmanen et al., 2013). Further details about cost frontier models are shown in section 2.6.

2.5. Stochastic Data Envelopment Analysis

Let $\mathbf{x}_j = (x_{1j}, ..., x_{Kj})$ be the vector of inputs, of dimension K, for the j-th decision making unit (DMU) and y_j the respective output, i.e., a single output. The SDEA model that estimates the production function is given by Equation 3, as originally proposed by Banker (1986):

$$
\begin{array}{ll}
min \sum_{i=1}^{n} \tau e_{1i} + (1-\tau) e_{2i} & \\
\text{subject to:} & \\
y_i = \alpha_i + \beta_{1i} x_{1i} + ... + \beta_{ki} x_{ki} & \\
+e_{1i} - e_{2i} & \forall_i = 1, ..., N \\
\alpha_i + \beta_{1i} x_{1i} + ... + \beta_{ki} x_{ki} \leq \alpha_j & \\
+\beta_{1j} x_{1i} + ... + \beta_{kj} x_{kj} & \forall_{i,j} = 1, ...N \\
\beta_{ki} \geq 0 & \forall_k = 1, ..., K; \\
 & i = 1, ..., N \\
e_{1i}, e_{2i} \geq 0, & \forall i = 1, .., N
\end{array}
\tag{3}
$$

where $\alpha_i + \beta_{1i} x_{1i} + ... + \beta_{ki} x_{ki}$ comprises the piecewise linear production frontier for each DMU i. Similar to SFA models, an SDEA compound error structure can be written as $e_i = e_{1i} - e_{2i}$. Thus, if $e_i > 0$ then the i-th DMU have an output above the production frontier. Likewise, if $e_i < 0$ then the i-th DMU have an output below the production frontier. If $e_i = 0$ the DMU output is located at the efficiency frontier. τ is the parameter, previously selected by the analyst, which controls the proportion of points crossing the production frontier.

Returns to scale properties in the SDEA model are implemented by restricting the value of α_i. Constant returns to scale are implemented assuming $\alpha_i = 0$. Non-decreasing returns to scale are implemented using $\alpha_i \geq 0$ and variable returns to scale are implemented using $\alpha_i \in \mathbb{R}$.

The SDEA model shown in Equation 3 is similar to a quantile multiple linear regression model where convexity and monotonicity are assumed (Jradi and Ruggiero, 2019). The solution of Equation 3 assumes that approximately 100τ percent of the data will be above the production frontier. If τ is equal to 1, the model becomes deterministic and the compound error expresses technical inefficiency only.

It can be shown that both mathematical representations of error, $e_i = e_{1i} - e_{2i}$ (SDEA) and $e_i = v_i - u_i$ (SFA), are equivalent. Therefore, the optimal value of τ can be chosen assuming an SFA stochastic compound error structure, as presented by Jradi and Ruggiero (2019). Equation 4 shows the probability density distribution (*pdf*) of the compound error for the production function.

$$f(\epsilon) = \frac{2}{\sigma}\phi\left(\frac{\epsilon}{\sigma}\right)\Phi\left(\frac{-\epsilon\lambda}{\sigma}\right) \tag{4}$$

where $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$, $\lambda = \frac{\sigma_u}{\sigma_v}$, $\phi(.)$ is the probability density function of a standard normal random variable and $\Phi(.)$ is the respective accumulated probability function. Jradi and Ruggiero (2019) proposes an algorithm to estimate the optimal value of τ by searching over a grid of values. For each value of $\tau = 0.5 + 0.001(k-1)$, $k = 1, ..., 491$, the frontier and the errors are estimated using Equation 3. Using the estimated residuals, $\hat{e}_i = \hat{e}_{1i} - \hat{e}_{2i}$, and the statistical properties of the first and second moments of the ϵ random variable, values for $\hat{\sigma}^2$ and $\hat{\lambda}$ are computed. The optimal estimate of $\hat{\tau}^*$ is the value of τ that achieves the maximum likelihood value based on Equation 4.

Finally, one may argue that the stochastic estimate of production SDEA model is the solution of Equation 5, shown below.

$$\max_{\alpha_i,\beta_i,\sigma,\lambda} \textstyle\sum_{i=1}^{n} \log\phi\left(\frac{e_i}{\sigma}\right) + \log\Phi\left(-\lambda\frac{e_i}{\sigma}\right) - \log\sigma$$

subject to:

$$\begin{aligned}
&y_i = \alpha_i + \beta_{1i}x_{1i} + ... + \beta_{ki}x_{ki} + e_i, && \forall i = 1, ..., N\\
&\alpha_i + \beta_{1i}x_{1i} + ... + \beta_{ki}x_{ki} \leq && \\
&\alpha_j + \beta_{1j}x_{1i} + ... + \beta_{kj}x_{kj}, && \forall i, j = 1, ...N\\
&\beta_{ki} \geq 0, && \forall k = 1, ..., K;\ \ \forall i = 1, ..., N.
\end{aligned} \tag{5}$$

Equation 5 was first presented by Banker and Maindiratta (1992). However, the solution of Equation 5, which comprises a non-linear maximization problem subject to linear constraints, is still an open topic for research.

2.6. Comparison of deterministic and stochastic cost frontier models

As mentioned, the present work proposes a new SDEA algorithm to estimate the cost efficiencies of the Brazilian DSOs. The main frontier models for cost regulation are based on Equation 6.

$$\ln x = \ln C(y_1, ..., y_k) + \delta z + u + v \tag{6}$$

where x is the observed cost (inputs), $C(.)$ is the function that characterizes the efficiency cost frontier, $y_1, ..., y_k$ are the outputs, z represents exogenous components, δ is the coefficient associated with the exogenous component, u is the random variable representing inefficiency and v is the random variable representing statistical noise. In deterministic models, deviations from the cost frontier are associated with inefficiency, i.e., $v = 0$ and $u \geq 0$. As mentioned, the DEA model assumes only inefficiency components and a non-parametric frontier equation.

Stochastic models (Kumbhakar and Lovell, 2003) assume both noise (v) and inefficiency (u) components. Consequently, the probability distributions for the v and u components must be specified. In general, the simplest model consists of assuming a truncated normal distribution (half-normal) for the inefficiency component, $u \sim |N(0, \sigma_u^2)|$ and a normal distribution for the noise component, $v \sim N(0, \sigma_v^2)$. The compound error is written as $\epsilon = u + v$. Using the probability distributions of u, $f_u(u)$ and v, $f_v(v)$, the probability distribution of the compound error ϵ is written as

$$f_\epsilon(\epsilon) = \int_0^\infty f_v(\epsilon - u) f_u(u) du \tag{7}$$

The solution is given by $f_\epsilon(\epsilon) = \frac{2}{\sigma}\phi(\frac{\epsilon}{\sigma})\Phi(\lambda\frac{\epsilon}{\sigma})$, where $\sigma^2 = \sigma_u^2 + \sigma_v^2$ and $\lambda = \frac{\sigma_u}{\sigma_v}$. Other probability distributions can also be defined (Meeusen and van Den Broeck, 1977; Aigner et al., 1977; Stevenson, 1980).

Given the density function of the compound error, the parameters of the cost function, $\ln x = \ln C(y_1, ..., y_k)$, as well as the parameters λ and σ^2 can be estimated using the maximization of the likelihood function (Casella and Berger, 2002). Nonetheless, according to Sartori (2006), the maximum likelihood estimate for parameter λ can be infinite with a non-zero probability for smaller samples. Consequently, the SFA model can wrongly identify all companies as fully efficient. The procedure above provides the estimates of the frontier or the cost function parameters. The estimates of the efficiency scores, i.e., the ratio between efficient cost and observed cost, imply the estimates of the inefficiency components u. In this case, the conditional density of $u|\epsilon$ can be calculated as shown in Equation 8.

$$f_{u|\epsilon}(u|\epsilon) = \frac{f(u,\epsilon)}{f_\epsilon(\epsilon)} = \frac{f_u(u)f_v(\epsilon - u)}{f_\epsilon(\epsilon)} \tag{8}$$

Given the conditional density distribution, Jondrow et al. (1982) and Bogetoft and Otto (2010) present three possible equations to estimate the efficiency score.

$$\hat{\theta}_i^{[1]} = e^{-E(u|\epsilon_i)} \tag{9}$$

$$\hat{\theta}_i^{[2]} = E(e^{-u}|\epsilon_i) \tag{10}$$

$$\hat{\theta}_i^{[3]} = e^{-M(u|\epsilon_i)} \tag{11}$$

where $E(u|\epsilon) = \int_0^\infty u \cdot f_{u|\epsilon}(u|\epsilon)du$, $E(e^{-u}|\epsilon_i) = \int_0^\infty e^{-u} \cdot f_{u|\epsilon}(u|\epsilon)$ and $M(u|\epsilon_i)$ is the conditional mode equation. In general, Equation 10 is the most commonly applied (Bogetoft and Otto, 2010). However, Equation 10 comprises an estimate for the mean value of the efficiency scores. Consequently, it is unlikely that one company will reach an efficiency score equal to one (100%). Thus, compound error models, in general, do not estimate fully efficient DSOs. Kuosmanen et al. (2013) mention that *"companies regulated by compound error models and the conditional mean estimator are not able to reach the efficiency frontier, even if its efficiency is adjusted according to the amount indicated by the model"*.

The StoNED model (Kuosmanen and Kortelainen, 2012) also applies a compound error structure, but the estimate of the cost function is acomplished in the first stage using a non-parametric convex least squares, as shown in Equation 12

$$\begin{array}{ll} \min_{\beta,\alpha,u} \sum_{i=1}^{N} (\ln x_i - \ln \varphi_i)^2 & \\ \text{subject to:} & \\ \varphi_i = \alpha_i + \beta_{1i}y_1 + ... + \beta_{ki}y_i & i = 1, ..., N \\ \alpha_i + \beta_{1i}y_1 + ... + \beta_{ki}y_i \geq \alpha_j & \\ +\beta_{1j}y_1 + ... + \beta_{kj}y_i & j, i = 1, ..., N \\ \beta_{ki} \geq 0 & i = 1, ..., N \end{array} \tag{12}$$

StoNED first apply a non-parametric minimum least squares to estimate the parameters of the frontier and the compound error. In the case of the StoNED, the compound error estimate is defined by the residual of the model, i.e., $\epsilon_i = \ln x_i - \ln \varphi_i$. In the original proposal, the normal and half-normal distributions are also used for the noise and the inefficiency components, respectively ($\epsilon_i = u_i + v_i$). The compound error parameters λ and σ^2 are estimated in a second stage using the method of moments (Kuosmanen and Kortelainen, 2012). The StoNED resembles a COLS estimate in which the parameters of the frontier are estimated using minimum least squares. However, the StoNED uses a non-parametric frontier equation. From the residuals of non-parametric OLS estimate, bias correction of the frontier and the compound error parameters are estimated. Final estimates of the efficiency scores, using StoNED is similar to the SFA model, in which the conditional distribution is required. The SDEA resembles the StoNED and DEA methods, by assuming a non-parametric form for the efficiency frontier.

It is worth noticing that SDEA, StoNED and SFA allow DSOs to cross the efficiency frontier. According to Tone (2017), other methods also allow points to cross the frontier, such as order-m and order-α frontier (Daraio and Simar, 2007) and chance-constrained programming (Land et al., 1993). Thus, these methodologies assume directly (as in the case of SFA and StoNED), or indirectly (in case of SDEA), a compound error structure.

As mentioned by Sartori (2006), da Silva et al. (2019) and Azzalini (2013) the composed error models, as SFA, StoNED and SDEA, presents serious problems of convergence, particularly related to the parameter λ . Those problems can be minimized using a large data base or a smaller number of variables, as compared to the number of DSOs. Nonetheless, this is not the case of Brazilian electricity distribution service operators (DSOs). da Silva et al. (2019) identified convergence problems in the compound error model using data from the 4TRC. Alternatives to adjust compound error models are shown in literature, such as the Bayesian approach described by Bayes and Branco (2007). However, the Bayesian approach requires the use of MCMC (Markov Chain Monte Carlo) methods which are sensitive to initial conditions of the algorithm.

In short, compound error models such as SFA and StoNED may present convergence problems when estimating their parameters, especially the efficiency scores. In addition, different stochastic assumptions for noise and inefficiency components may generate different results. Consequently, estimated operational cost efficiencies may be unreliable. Thus, international regulation agencies prefer DEA (de Barros Mesquita, 2017).

2.7. Proposed SDEA algorithm

The proposed SDEA estimation algorithm is based on the maximization of the compound error likelihood, subject to an SDEA piecewise linear cost frontier model, as shown in Equation 13.

$$\begin{aligned}
&\max_{\alpha_i,\beta_i,\sigma,\lambda} \textstyle\sum_{i=1}^{n} \log\phi\left(\frac{e_i}{\sigma}\right) + \log\Phi\left(\lambda\frac{e_i}{\sigma}\right) - \log\sigma \\
&\textit{subject to}: \\
&\begin{array}{ll}
C(y_{1i},...,y_{Ki}) = \alpha_i + \beta_{1i}y_{1i} + ... + \beta_{Ki}y_{Ki} & \\
e_i = \log x_i - \log C(y_{1i},...,y_{Ki}) & \forall i = 1,...,N \\
\alpha_i + \beta_{1i}y_{1i} + ... + \beta_{Ki}y_{Ki} \geq \alpha_j + \beta_{1j}y_{1i} + ... + \beta_{Kj}y_{Ki} & \forall i,j = 1,...,N \\
\beta_{ki}, \alpha_i \geq 0 & \forall i = 1,...,N \\
 & \forall k = 1,...,K \\
\sigma, \lambda > 0. &
\end{array}
\end{aligned} \tag{13}$$

Equation 13 comprises the original SDEA problem shown in Equation 5, but assuming a cost frontier formulation as described in Equation 12. Following Jradi and Ruggiero (2019), our proposed algorithm also applies the proportion of points crossing the frontier as a proxy for the estimation of λ. Nevertheless, a different representation is proposed as follows: from Equation 4, the probability of points crossing a production frontier can be written as a function of the λ parameter, as shown in Equation 14.

$$\begin{aligned}
P[\epsilon \leq 0] &= \textstyle\int_{-\infty}^{0} \frac{2}{\sigma}\phi\left(\frac{\epsilon}{\sigma}\right)\Phi\left(\frac{-\epsilon\lambda}{\sigma}\right)d\epsilon \\
&= 0.5 + \tfrac{1}{\pi}arctan(\lambda)
\end{aligned} \tag{14}$$

For the cost frontier, the compound error is written as $v + u$. Thus, the probability of points crossing the cost frontier can be written as a function of the λ parameter, as follows. Let $\epsilon_1 = v - u$ and $\epsilon_2 = v + u$, where ϵ_1 comprises the production function compound error and ϵ_2 comprises the cost frontier compound error. The pdfs of ϵ_1 and ϵ_2 can be written as follows.

$$f_{\epsilon_1}(x) = \frac{2}{\sigma}\phi\left(\frac{x}{\sigma}\right)\Phi\left(\frac{-x\lambda}{\sigma}\right) \tag{15}$$

$$f_{\epsilon_2}(x) = \frac{2}{\sigma}\phi\left(\frac{x}{\sigma}\right)\Phi\left(\frac{x\lambda}{\sigma}\right) \tag{16}$$

It can be shown that $\phi(\frac{x}{\sigma}) = \phi(\frac{-x}{\sigma})$. Consequently, $f_{\epsilon_1}(x) = f_{\epsilon_2}(-x)$ or $f_{\epsilon_1}(-x) = f_{\epsilon_2}(x)$. Thus, it can be shown that,

$$\begin{aligned} P[\epsilon_1 \leq 0] &= P[\epsilon_2 > 0] \\ P[\epsilon_1 \leq 0] &= 1 - P[\epsilon_2 \leq 0] \\ P[\epsilon_2 \leq 0] &= 1 - P[\epsilon_1 \leq 0] \end{aligned} \tag{17}$$

From Equation 14, the probability of points crossing the cost frontier can be written as follows.

$$\begin{aligned} P[\epsilon_2 \leq 0] &= 1 - (0.5 + \tfrac{1}{\pi}\arctan(\lambda)) \\ &= 0.5 - \tfrac{1}{\pi}\arctan(\lambda) \end{aligned} \tag{18}$$

Finally, from Equation 18, the λ parameter can be written as a function of $P[\epsilon_2 \leq 0]$, as shown in Equation 19.

$$\begin{aligned} P[\epsilon_2 \leq 0] &= 1 - (0.5 + \tfrac{1}{\pi}\arctan(\lambda) \\ \tfrac{1}{\pi}\arctan(\lambda) &= 0.5 - P[\epsilon_2 \leq 0] \\ \lambda &= \tan[\pi(0.5 - P[\epsilon_2 \leq 0])] \end{aligned} \tag{19}$$

Equation 19 shows that, using the SDEA cost frontier model, it is possible to estimate the λ parameter by defining the probability of points crossing the frontier, $P[\epsilon_2 \leq 0]$. Given the sample of size N, we propose to estimate the λ parameter by gradually allowing points to cross the efficiency frontier, thus approximating the value of $P[\epsilon_2 \leq 0]$ as the observed proportion of points crossing the frontier, $\hat{P}[\epsilon_2 \leq 0] = k/N$, as show in Equation 20

$$\hat{\lambda} = \tan[\pi(0.5 - k/N)] \tag{20}$$

where $k = 0, 1, ..., N/2$..

In order to estimate the cost efficiency frontier, i.e., the piecewise linear regression parameters, the SDEA linear model using $\tau = 1$ is applied, as shown in Equation 21.

$$
\begin{array}{ll}
min \sum_{i=1}^{n} e_i & \\
& \\
\text{subject to:} & \\
y_i = \alpha_i + \beta_{1i}x_{1i} + ... + \beta_{ki}x_{ki} + e_i & \forall i = 1, ..., N \\
\quad \alpha_i + \beta_{1i}x_{1i} + ... + \beta_{ki}x_{ki} \geq \alpha_j & \\
+\beta_{1j}x_{1i} + ... + \beta_{kj}x_{kj} & \forall i, j = 1, ...N \\
\beta_{ki} \geq 0 & \forall k = 1, ..., K; \quad \forall i = 1, ..., N \\
e_i \geq 0, & \forall i = 1, .., N
\end{array}
\tag{21}
$$

The proposed algorithm starts by assuming that no points are crossing the efficiency frontier, thus the frontier is estimated using equation 21 and $\hat{P}[\epsilon_2 \leq 0] = 0$. Next, the points located on the frontier are the candidates to cross the frontier. Each of these points is temporarily removed from the data and a new frontier is estimated, also using equation 21. Using the complete data, the residuals of the cost frontier model are calculated as $e_i = \log x_i - \log C(y_{1i}, ..., y_{Ki})$, as shown in Equation 13. Assuming that $\hat{\lambda} = \tan[\pi(0.5 - 1/N)]$, the σ^2 estimate is calculated using an univariate maximum likelihood search applying, for instance, the Golden-section search algorithm (Kiefer, 1953). Therefore, one maximum likelihood value is computed for each point on the frontier. The point with the maximum value is selected as the candidate to cross the frontier. The procedure is repeated until a maximum number of points, defined by the analyst, say $N/2$, crosses the frontier.

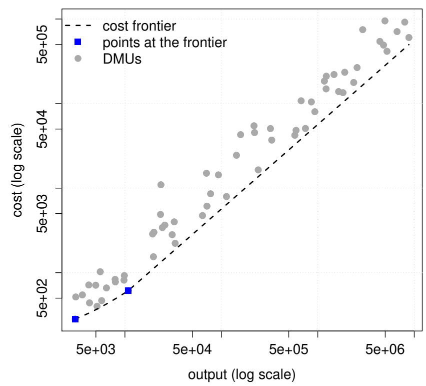
Briefly, the proposed algorithm comprises the following steps:

1. Set $k = 1$.
2. Compute $\hat{\lambda} = tan[\pi(0.5 - k/N)]$.
3. Solve the SDEA model (Equation 21).
4. Select the points located on the frontier, i.e., $e_i = 0$.
5. For each point located on the frontier,
   (a) Remove, temporarily, the point and solve the SDEA model (Equation 21).
   (b) Using the complete dataset calculate the residuals: $\hat{e}_i = \log x_i - \log C(y_{1i}, ..., y_{Ki})$.
   (c) Using the residuals, calculate $\hat{\sigma}^2 = arg \max_{\sigma} LL(\sigma)$, where $LL(\sigma) = \sum_{i=1}^{n} \left[\log \phi \left(\frac{\hat{e}_i}{\sigma} \right) + \log \Phi \left(\hat{\lambda} \frac{\hat{e}_i}{\sigma} \right) - \log \sigma \right]$ is the likelihood function.

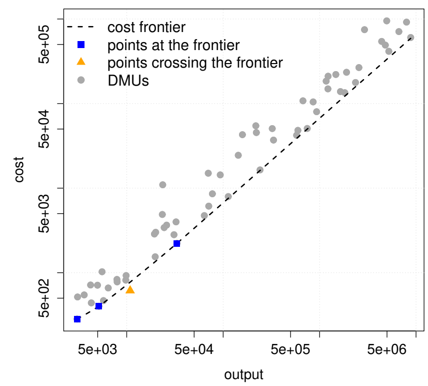
(d) Select the point located on the frontier that achieved the maximum likelihood value and remove it from the data set.
(e) Set $k = k + 1$ and return to Step 2.

6. Repeat steps 2 to 5 until a maximum of points, defined by the analyst, crosses the frontier.
7. The final solution comprises the value of k and the respective efficiency frontier which achieved the maximum likelihood function value.

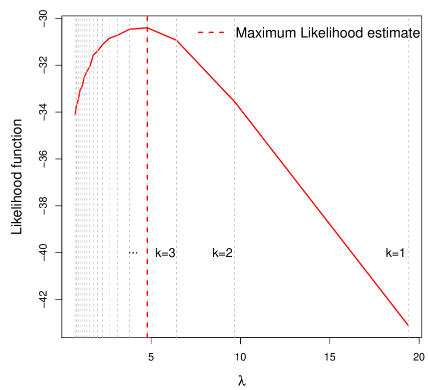
Figure 2 illustrates the proposed algorithm. Using Equation 21, the cost frontier and the points located on the frontier are estimated as shown in Figure 2(a). The points are gradually evaluated and the candidate with the maximum likelihood value crosses the frontier, as shown in Figure 2(b). The procedure is repeated until a maximum number of points crosses the frontier and the likelihood function is evaluated at each step, as shown in Figure 2(c). The final solution comprises the proportion of points crossing the frontier and the estimated cost frontier with the maximum likelihood value, as shown in Figure 2(d).



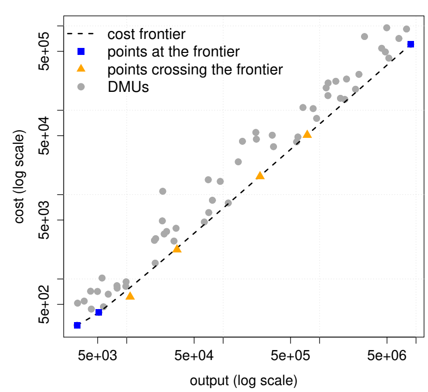
(a) Step 1: The cost frontier and the points located on the frontier are estimated.



(b) Step 2: The points crossing the frontier are selected gradually using likelihood maximization.



(c) Step 3: The optimal proportion of points is chosen based on the maximum likelihood value.



(d) Final cost frontier solution.

Figure 2: Proposed algorithm to estimate the optimal number of points crossing the cost frontier and the maximum likelihood parameters.

It is worth mentioning that by evaluating each point located on the frontier and gradually increasing the number of crossing points, the proposed algorithm avoids the estimate of overlapping frontiers or crossing quantiles (Wang et al., 2014).

2.8. Simulation study

In this section, a simulation study is presented to evaluate the statistical properties of the proposed SDEA algorithm. The simulation model is based on the statistical distribution and the statistical correlation between observed cost (input) and weighted market (ouput) of the Brazilian DSOs. Data is generated for $N = 61$ observations (DSOs). The output $\log(y_i) \sim Uniform(min = 8;\ max = 16)$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2 = 0.7$. Different values for λ are evaluated using different numbers of points crossing the frontier. The selected numbers of points crossing the frontier are $k = \{1, 3, 5, 8, 15\}$. Thus, based on Equation 20, the respective values of λ are $\lambda = \{19.4, 6.42, 3.8, 2.29, 1.02\}$. Consequently, $\sigma_u^2 = \sigma^2 \cdot \lambda^2/(1 + \lambda^2)$ and $\sigma_v^2 = \sigma^2/(1+\lambda^2)$. Therefore, $\epsilon = v + u$ with $v \sim N(0, \sigma_v)$ and $u \sim |N(0, \sigma_u)|$. The simulated costs are generated as $x_i = 0.07 \cdot y_i \times e^{\epsilon_i}$ denoting a constant returns to scale model.

2.9. The use of SDEA to approximate DEA results with weight restrictions

SDEA can be used to estimate DEA-NDRS cost frontier. To illustrate this approach, an SDEA model using one input (x) and one output (y) is presented in Equation 22

$$
\begin{array}{ll}
\min \sum_{i=1}^{n} e_i & \\
s.t. & \\
x_i = \alpha_i + \beta_i y_i + e_i & \forall i = 1, ..., N \\
\alpha_i + \beta_i y_i \geq \alpha_j + \beta_j y_i & \forall i, j = 1, ..., N \\
\beta_i, e_i, \alpha_i \geq 0 & \forall i = 1, ..., N
\end{array}
\tag{22}
$$

As mentioned, for non-decreasing returns to scale property $\alpha_i \geq 0$. For each DSO i, the efficient cost is given by $\hat{x}_i = \alpha_i + \beta_i y_i$ and the cost efficiency (input oriented) is given by $\frac{\hat{x}_i}{x_i}$. It is worth noticing that Equation 22 comprises a specific SDEA formulation in which no DSO crosses the frontier, i.e., $\tau = 0$. Consequently, both DEA and SDEA results are identical. If the DEA model applies weight restrictions, as proposed by the Brazilian regulator, one alternative to estimate a similar SDEA model is to use a data set of fully efficient DSOs as follows. First, the efficiency score (θ_i) is calculated using the DEA model with weight restrictions. Second, the DEA cost efficiency (θ_i) is included in the SDEA optimization model as shown in Equation 23.

$$
\begin{array}{ll}
\min \sum_{i=1}^{n} e_i & \\
s.t. & \\
\theta_i x_i = \alpha_i^* + \beta_i^* y_i + e_i & \forall i = 1, ..., N \\
\alpha_i^* + \beta_i^* y_i \geq \alpha_j^* + \beta_j^* y_i & \forall i, j = 1, ..., N \\
\beta_i^*, e_i, \alpha_i^* \geq 0 & \forall i = 1, ..., N
\end{array}
\tag{23}
$$

Briefly, by including DEA efficiencies, which were estimated using weight restrictions, in Equation 23, the SDEA model is estimated using all DSOs at the frontier. The SDEA cost frontier with weight restriction is given by $\hat{x}_i = \alpha_i^* + \beta_i^* y_i$. Thus, the estimated SDEA models using Equations 22 and 23 allows the analyst to compare the piecewise linear equations of the efficiency costs with and without weight restrictions for each DSO.

If p outputs are available, then the piecewise cost frontier is written as:

$$x_i = \alpha_i + \beta_{1i} y_{1i} + \beta_{pi} y_{pi} \tag{24}$$

Equation 24 can also be written as $x_i = \alpha_i + \mathbf{y}_i^T \beta$, where $\mathbf{y}_i$ is a vector of outputs $\mathbf{y}_i = [y_{1i}, y_{2i}, ..., y_{pi}]$.

3. Results

Table 2 presents the simulated results for the estimated number of points crossing the efficiency frontier. Three SDEA models using different returns to scale were evaluated. As mentioned, the simulated data sets were generated assuming constant returns do scale (CRS). Therefore, it is expected that the SDEA-CRS model achieves better performance. Using the SDEA-CRS model, the median values indicate that the model was able to detect the correct number of points crossing the frontier in most cases, except for simulations with the largest number of points crossing the frontier, $k = 15$. In these cases, the estimated number of points was slightly greater. Furthermore, the mean value indicates an estimation bias, i.e., a slightly larger number of points crossing the frontier. Differences between mean and median values indicate that the empirical distribution is asymmetric. In fact, the distribution of the estimated values is truncated at $k = 29$. It is worth noticing that the real value of the number of points crossing the frontier is within the quartile intervals, i.e., between the first and third quartiles, for all simulated values of k.

Using the SDEA-NDRS model, in general, the median values were slightly larger as compared do the SDEA-CRS model and the quartile intervals were also larger, as shown in Table 2. Nevertheless, simulated results using SDEA-CRS and SDEA-NDRS are closer.

Using the SDEA-VRS model, an interesting behavior is found. For $k = 1$, the SDEA-VRS presented similar results to SDEA-CRS. For $k = 3$, 5, 8, and 15, the SDEA-VRS model underestimated the number of points crossing the efficiency frontier. It is worth mentioning that the SDEA-VRS model is more flexible than SDEA-CRS and SDEA-NDRS. In general, the SDEA-VRS requires more piecewise linear functions. Consequently, more points are estimated on the frontier and therefore fewer points crosses the frontier.

Table 2: Summary statistics for the estimated number of points crossing the efficiency frontier using the simulated data.

		Number of points crossing the frontier					
RTS	$k(\lambda)$	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
CRS	1 (19.4)	1	1	1	1.4	1	22
	3 (6.42)	1	1	3	3.1	4	28
	5 (3.8)	1	3	5	5.5	6	29
	8 (2.29)	1	6	8	9.8	12	29
	15 (1.02)	1	11	17	17.5	25	29
NDRS	1 (19.4)	1	1	1	1.7	2	24
	3 (6.42)	1	1	2	3.4	4	29
	5 (3.8)	1	3	5	6.5	8	29
	8 (2.29)	1	6	9	11.4	15	29
	15 (1.02)	1	11	17	17.7	25	29
VRS	1 (19.4)	1	1	1	1.4	1	29
	3 (6.42)	1	1	1	2.8	3	29
	5 (3.8)	1	1	3	5	6	29
	8 (2.29)	1	2	7	9	13	29
	15 (1.02)	1	7	14	14.6	22	29

Table 3 presents the simulated results for the estimated λ parameter. As shown in Equation 20 there is a non-linear correlation between the number of points crossing the efficiency frontier (k) and the respective λ parameter. In general, the larger the number of points crossing the frontier, the lower the value of λ. In our proposed algorithm, the value of λ depends on the proportion of points crossing the efficiency frontier and, therefore, is sensitive to the sample size n. Consequently, a finite grid of λ values are evaluated. Using SDEA-CRS, for fewer points crossing the frontier, the median value of the estimated λ are closer to the real value, and the mean value is slightly larger. Using SDEA-NDRS and SDEA-VRS, the median values are closer to the real values but the mean values are overestimated. As mentioned, the λ estimate using maximum likelihood is problematic and usually requires large data sets. On the contrary, the estimate of the number of points crossing the efficiency frontier has shown promising results.

Table 3: Summary statistics for the λ parameter using the simulated data.

		$\hat{\lambda}$ summary statistics					
RTS	λ	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
CRS	19.4	0.47	19.40	19.40	16.86	19.40	19.40
	6.42	0.13	4.79	6.42	9.72	19.40	19.40
	3.8	0.08	3.13	3.80	5.63	6.42	19.40
	2.29	0.08	1.41	2.29	2.96	3.13	19.40
	1.02	0.08	0.29	0.83	1.32	1.57	19.40
NDRS	19.4	0.13	9.67	19.40	15.74	19.40	19.40
	6.42	0.08	4.79	9.67	10.39	19.40	19.40
	3.8	0.08	2.29	3.80	5.94	6.42	19.40
	2.29	0.08	1.03	2.00	3.07	3.13	19.40
	1.02	0.08	0.29	0.83	1.28	1.57	19.40
VRS	19.4	0.08	19.40	19.40	17.73	19.40	19.40
	6.42	0.08	6.42	19.40	13.16	19.40	19.40
	3.8	0.08	3.13	6.42	10.05	19.40	19.40
	2.29	0.08	1.26	2.65	6.16	9.67	19.40
	1.02	0.08	0.47	1.14	3.21	2.65	19.40

Table 4 shows the simulated results for the estimated σ^2 parameter. In general, both the median and mean values are slightly underestimated. Nevertheless, a small bias is expected for small samples. In general, results using SDEA-CRS, SDEA-NDRS and SDEA-VRS are very close.

To illustrate the effect of the sample size, Tables 5, 6 and 7 shows the simulation study results for the estimates of λ and σ^2 using a large data set with $n = 122$ observations. The simulated values of λ are identical to the simulated values using $n = 61$. Consequently, the largest the sample size the larger the number of points crossing the frontier using the same λ value. As expected, the quantiles using a larger sample size are narrower as compared to a smaller sample size. Furthermore, the mean and median values are closer to the simulated true parameters using a larger sample size. Table 4 shows that using a larger sample size, the mean and median values of the estimated σ^2 parameters are very close to the true value of $\sigma^2 = 0.7$.

Table 4: Summary statistics for the σ^2 parameter using the simulated data ($\sigma^2 = 0.7$).

	$\hat{\sigma}^2$ summary statistics					
RTS	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
CRS	0.33	0.63	0.68	0.72	0.72	0.89
	0.36	0.63	0.68	0.68	0.73	0.96
	0.32	0.62	0.69	0.68	0.75	0.96
	0.33	0.60	0.68	0.67	0.76	1.09
	0.43	0.58	0.66	0.69	0.79	1.22
NDRS	0.29	0.61	0.66	0.66	0.71	0.90
	0.36	0.62	0.68	0.67	0.74	0.95
	0.34	0.58	0.67	0.66	0.74	0.97
	0.33	0.55	0.65	0.65	0.75	1.13
	0.41	0.58	0.65	0.68	0.76	1.25
VRS	0.34	0.59	0.64	0.64	0.69	0.89
	0.34	0.61	0.67	0.66	0.72	0.93
	0.31	0.61	0.69	0.67	0.75	0.99
	0.34	0.56	0.69	0.68	0.79	1.12
	0.42	0.60	0.70	0.74	0.86	1.52

Table 5: Summary statistics for the estimated number of points crossing the efficiency frontier using the simulated data (n=122).

		Number of points crossing the frontier					
RTS	$k(\lambda)$	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
CRS	2 (19.4)	1	1	2	2.02	3	15
	6 (6.42)	1	4	5	5.60	7	18
	10 (3.8)	1	8	10	10.29	12	59
	16 (2.29)	2	13	16	17.68	20	60
	30 (1.02)	7	24	32	35.54	49	60
NDRS	2 (19.4)	1	1	1	2.1	3	21
	6 (6.42)	1	3	5	6.07	8	60
	10(3.8)	1	7	10	11.44	14	60
	16 (2.29)	1	13	17	19	23	60
	30 (1.02)	6	25	37	37.57	51	60

Table 6: Summary statistics for the λ parameter using the simulated data (n=122).

		$\hat{\lambda}$ summary statistics					
RTS	λ	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
CRS	19.4	2.45	12.91	19.40	26.03	38.82	38.82
	6.42	2.00	5.48	7.72	9.55	9.67	38.82
	3.8	0.05	3.13	3.79	4.48	4.78	38.82
	2.29	0.02	1.76	2.28	2.44	2.87	19.39
	1.02	0.02	0.31	0.92	0.97	1.40	5.48
NDRS	19.4	1.66	12.91	38.82	26.71	38.82	38.82
	6.42	0.02	4.78	7.72	10.66	12.91	38.82
	3.8	0.02	2.65	3.79	4.87	5.48	38.82
	2.29	0.02	1.48	2.13	2.53	2.87	38.82
	1.02	0.02	0.26	0.71	0.91	1.33	6.42

Table 7: Summary statistics for the σ^2 parameter using the simulated data (n=122, $\sigma^2 = 0.7$).

	$\hat{\sigma}^2$ summary statistics					
RTS	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
CRS	0.54	0.65	0.69	0.69	0.72	0.85
	0.52	0.65	0.69	0.69	0.73	0.86
	0.41	0.64	0.69	0.69	0.73	0.87
	0.42	0.62	0.68	0.68	0.74	0.93
	0.45	0.60	0.67	0.68	0.75	1.05
NDRS	0.50	0.65	0.68	0.68	0.71	0.83
	0.38	0.65	0.69	0.69	0.73	0.92
	0.37	0.62	0.68	0.68	0.73	0.97
	0.40	0.60	0.67	0.66	0.73	1.03
	0.48	0.58	0.64	0.66	0.74	1.13

Using the Brazilian dataset, Table 8 presents SDEA-NDRS models using different combinations of the outputs. Initially, models using one output are presented. Results show that using weighted power and underground network, the estimate of the number of points crossing the frontier is 29, i.e., close to half of the sample size. Similar results were described in the simulated study in which the estimate of the number of points crossing the frontier was 29 (see Table 2), even though the real number of points was lower. It is worth mentioning that most DSOs have underground network equal to zero. On the contrary, using number of consumers, high voltage network and aerial network the estimated number of points crossing the frontier is 4, 7 and

5, respectively. By combining two outputs, four models presented 4 points crossing the frontier, one model with 6 points crossing the frontier, one model with 14 points crossing the frontier and two models with 24 and 29 points crossing the frontier. The latter two models have underground network as one of the outputs. By combining three outputs: weighted power, aerial network and high voltage network, results show that the estimated number of points crossing the frontier is 5. By combining four output variables, two models were adjusted to indicate 5 points and 1 point crossing the frontier, respectively. By combining five outputs, the estimated number of points crossing the frontier is 1. This preliminary analysis indicates that the more outputs included in the model the lower the estimate of the number of points crossing the frontier. One may argue that the more outputs included in the model the more points are estimated on the frontier and the lower the estimate of the number of points crossing the frontier. A similar behavior was found in the SDEA-VRS simulation study.

In addition to the number of points crossing the frontier, Table 8 shows the number of piecewise linear functions and the maximum likelihood value estimated for each model. In general, the more outputs included in the model the larger the number of piecewise linear functions. One may argue that a large number of piecewise linear functions indicates a complex frontier, driven by outliers and specific outputs. Surprisingly, the greater the number of points crossing the frontier, the lower the estimated number of piecewise linear functions indicating that allowing points to cross the efficiency frontier reduces the complexity of the frontier. Thus, effects of outliers and the presence of additional outputs are reduced. Additionally, Table 8 shows the SDEA-NDRS results using all Brazilian output variables and different numbers of points crossing the frontier. If one point crosses the frontier then 57 piecewise linear functions are estimated; whereas, if 12 points cross the frontier then 41 piecewise linear functions are estimated. Finally, Table 8 also shows the number of piecewise linear functions using the DEA-NDRS Brazilian models, with and without weight restrictions, with no points crossing the frontier. In this case, the number of piecewise linear functions was estimated using an SDEA approximation, as described in section 2.9. Results show that the current DEA-NDRS (ANEEL) model has 61 piece-wise linear functions. This is the largest number of piece-wise linear functions among the evaluated models which indicates that the regulator model is highly complex. As mentioned, the ANEEL model includes output variables with a large proportion of null observations such as the underground network extension.

Interestingly, the SDEA estimate using the complete set of outputs and one DMU crossing the efficiency frontier has 57 piece-wise linear functions and the largest value of the likelihood function. Thus, a slightly simpler model, as compared to the current ANEEL model, can be achieved letting one DMU cross the efficiency frontier. On the contrary, the maximum likelihood SDEA solution may also indicate a saturated model, i.e., a model with too many output variables.

Table 8: Estimated SDEA-NDRS models using the Brazilian database with different output combinations.

Number of outputs	outputs	$\hat{\lambda}$	$\hat{\sigma}^2$	Number of crossing points	Number of piece-wise linear functions	Likelihood
1	Weighted power	0.08	0.36	29	3	-24.68
	Number of consumers	4.79	0.54	4	4	-16.17
	High voltage network	2.65	1.20	7	3	-74.53
	Underground network	0.08	1.48	29	3	-110.70
	Aerial network	3.80	0.79	5	3	-42.17
2	Number of consumers and Weighted power	4.79	0.52	4	5	-14.62
	Number of consumers and High voltage network	4.79	0.52	4	7	-14.50
	Number of consumers and Aerial network	4.79	0.45	4	10	-9.14
	High voltage network and Aerial network	4.79	0.81	4	9	-41.33
	Weighted power and Underground network	3.79	0.65	5	8	-28.65
	Weighted power and High voltage network	3.79	0.46	5	7	-10.50
	Underground network and High voltage network	1.76	0.87	10	9	-62.05
	Underground network and Aerial network	3.13	0.67	6	6	-34.93
3	Weighted power, Aerial network, and High voltage network	3.80	0.46	5	9	-10.10
4	Weighted power, Aerial network, High voltage network and Underground network	3.80	0.45	5	21	-10.37
	Weighted power, Aerial network, High voltage network, and Number of consumers	19.39	0.45	1	14	-4.05
5	Weighted power, Aerial network, Weighted power, Underground network, and High voltage network	19.40	0.45	1	25	-4.00
7	Full Model	19.40	0.36	1	57	3.87
		6.42	0.35	3	55	1.58
		3.13	0.34	6	48	0.08
		1.40	0.31	12	41	-2.70
7	DEA-NDRS (ANEEL) without weight restrictions	-	-	0	60	-
	DEA-NDRS (ANEEL) with weight restrictions	-	-	0	61	-

Figure 3 shows the Spearman correlation comparing the two models proposed by the regulator (DEA-NDRS and ad-hoc) and the proposed SDEA models with different number of outputs and different number of points crossing the efficiency frontier. For example, SDEA 2:1 comprises a SDEA model with two outputs and one crossing point. As expected, the DEA-NDRS and the ad-hoc (DEA-NDRS + Bootstrap + Equation 2) models share the largest correlation of 99%. In sequence, the SDEA model with 2 outputs and one crossing point (SDEA 2:1) has a correlations of 91% with the ad-hoc model. Thus, the SDEA 2:1 generates a very similar ranking of the DSOs as compared to the Brazilian regulator proposal, but using a fraction of the outputs. Nevertheless, the more points crossing the frontier, the lower the correlation as shown in models SDEA 2:12 and SDEA 2:6. Similarly, the greater the number of outputs in the SDEA model, the lower the Spearman correlation, as shown in SDEA 4:1 and SDEA 4:6. Finally, the correlations using the complete number of outputs (7) and using varying numbers of crossing points are the lowest. Using seven outputs, only the SDEA model with twelve points crossing the frontier (SDEA 7:12) shows the greatest correlation of 80 %. As mentioned, the ad-hoc benchmarking model proposed by the Brazilian regulator applies weight-restrictions in order to manage the large number of outputs. As shown in Table 8, the ad-hoc model has the greatest number of piecewise linear functions, therefore, it is the most complex model as compared to the proposed SDEA models.

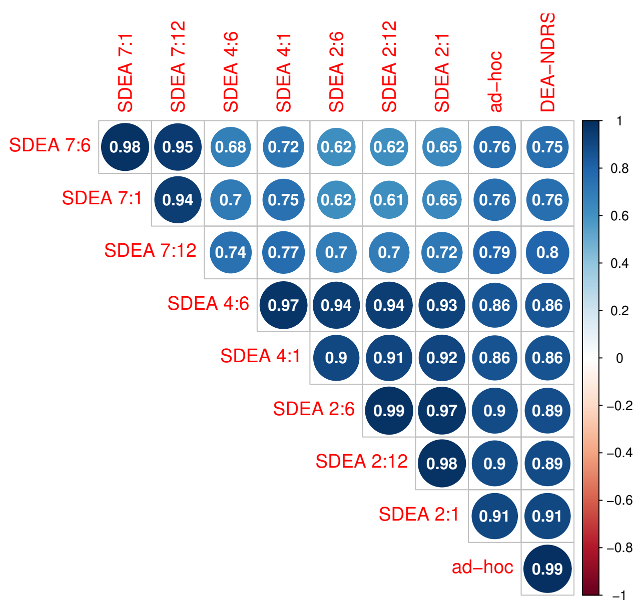


Figure 3: Pairwise Spearman correlation plot of the efficiency costs estimated by the Brazilian regulator and the proposed SDEA models.

Figure 4 compares the cost efficiencies of the DEA-NDRS and ad-hoc models, proposed by ANEEL, and the SDEA model with two outputs and one crossing point. The latter model achieved the largest Spearman correlation with the ad-hoc model. Results of DEA-NDRS (ANEEL) and SDEA 2:1 are very similar. A major difference is shown for the DSO with the largest DEA-NDRS cost efficiency. As shown in Figure 4, the ad-hoc model generates greater cost efficiencies with 11 DSOs having efficiencies greater than 100%.

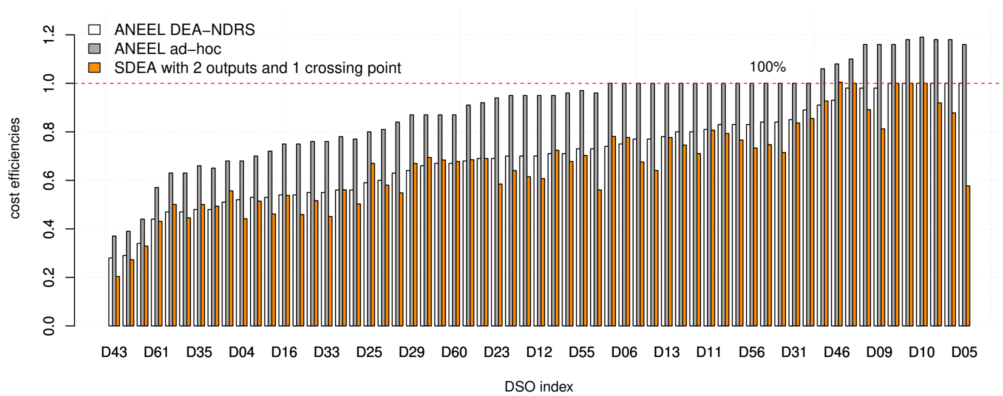


Figure 4: Comparison between cost efficiencies using the brazilian DEA-NDRS model with weight restrictions, the procedure using DEA-NDRS + Bootstrap + Equation 2 and the SDEA model using two outputs and one crossing point.

Figure 5 compares the final benchmarking model proposed by ANEEL (ad-hoc) and the proposed SDEA model using all the original seven outputs and with twelve points crossing the efficiency frontier. In general, the ANEEL model achieves greater efficiency scores for most DSOs, as compared to the SDEA 7:12. For two DSOs, the SDEA 7:12 generates cost efficiencies greater than using the ANEEL model. One may claim that the two largest efficiency costs represent outliers. It is worth mentioning that the proposed SDEA model does not include weight restrictions and most of the DSOs crossing the frontier do not achieve large cost efficiencies. Finally, Table 4 shows the original operational costs and the regulated costs estimated by the ad-hoc (ANEEL) and the proposed SDEA 7:12 models. Highlighted cells show regulated OPEX greater than observed OPEX and efficiency costs greater than 100%. Results show that, using the ad-hoc model, the total regulated costs comprise 93% of the total original costs. Whereas, using the proposed SDEA 7:12 model, the total regulated costs comprise 89% of the total original costs. The difference comprises a total value of R$ 769,114,635.17 or US$ 236,650,656.97, considering an exchange rate of R$3.25/US$1 in December 30, 2016, according to the Federal Reserve website. Therefore, the proposed SDEA 7:12 model rewards a few DSOs with a lower value of the total regulated costs as compared to the current Brazilian DEA (ad-hoc) model.

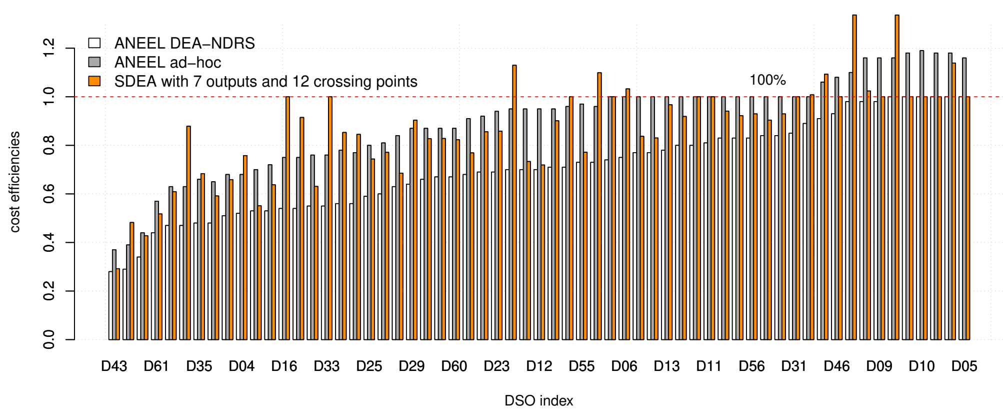


Figure 5: Comparison between cost efficiencies using the brazilian DEA-NDRS model with weight restrictions, the procedure using DEA-NDRS + Bootstrap + Equation 2 and the SDEA model using seven outputs and twelve crossing point.

4. Conclusion

The traditional DEA and SFA benchmarking models do not estimate an efficiency cost greater than 100%. The SFA model applies a compound error structure, in which the so-called noise component allows points to cross the efficiency frontier. Nevertheless, the original SFA cost efficiency estimates are below 100%. The noise component comprises a random variable with mean of zero and variance of σ_v^2. On the contrary, the Brazilian regulator has proposed an *ad-hoc* procedure to allow DSOs to achieve cost efficiencies greater than 100%. As shown in the present paper, the Brazilian *ad-hoc* procedure mimics an SDEA model in which the extent to which a DSO crosses the frontier is counted as an operating cost prize for over-efficiency.

In practice, the Brazilian regulator assumes that the noise component is meaningful and should not be counted as pure random error. On the contrary, it suggests that the compound error structure is the sum of inefficiency and a symmetric random variable component that allows DSOs to cross the frontier. As mentioned, for those DSOs crossing the frontier, a reward is given. Thus, one may suggest that the random noise component should be reevaluated as random reward component. Finally, the Brazilian cost efficiency model is in effect and has been applied since 2015 to estimate efficient operating costs for the Brazilian DSOs.

Finally, the Brazilian regulatory process of electricity distribution services must be technically consistent and reliable. Nonetheless, estimating an efficiency frontier using only a few set of inputs and outputs may generate biased results since the electricity distribution service is highly complex. One simple solution is to consider an underlying stochastic component which may capture misspecifications of inputs and outputs. The Brazilian regulator created a complex ad-hoc procedure that can be replaced by a simple stochastic data envelopment analysis model.

Acknowledgements

The authors thank CNPq [Grant number: PQ-303119/2019-5] and CAPES [PROBRAL Process number: 88881.370796/2019-01] for financial support.

References

M. A. Costa, A. L. M. Lopes, G. B. B. de Pinho Matos, Statistical evaluation of data envelopment analysis versus cols cobb–douglas benchmarking models for the 2011 brazilian tariff revision, Socio-Economic Planning Sciences 49 (2015) 47–60.

A. V. da Silva, M. A. Costa, A. L. M. Lopes, G. M. do Carmo, A close look at second stage data envelopment analysis using compound error models and the tobit model, Socio-Economic Planning Sciences 65 (2019) 111–126.

A. L. M. Lopes, B. de Almeida Vilela, M. A. Costa, E. A. Lanzer, Critical evaluation of the performance assessment model of brazilian electricity distribution companies, Revista Gestão & Tecnologia 16 (2016) 5–30.

V. V. Podinovski, Production trade-offs and weight restrictions in data envelopment analysis, Journal of the Operational Research Society 55 (2004) 1311–1322.

M. A. Costa, L. B. Mineti, V. D. Mayrink, A. L. M. Lopes, Bayesian detection of clusters in efficiency score maps: An application to brazilian energy regulation, Applied Mathematical Modelling 68 (2019) 66–81.

G. D. R. Gil, M. A. Costa, A. L. M. Lopes, V. D. Mayrink, Spatial statistical methods applied to the 2015 brazilian energy distribution benchmarking model: Accounting for unobserved determinants of inefficiencies, Energy Economics 64 (2017) 373–383.

L. Simar, P. W. Wilson, Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, Management science 44 (1998) 49–61.

P. Bogetoft, L. Otto, Benchmarking with dea, sfa, and r, volume 157, Springer Science & Business Media, 2010.

ANEEL, NT 66/2015-SRM/SGT/ANEEL: Retodologia de Custos Operacionais, 2015.

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

R. D. Banker, Stochastic data envelopment analysis, Pittsburgh: Carnegie-Mellon University (1986).

R. D. Banker, A. Maindiratta, Maximum likelihood estimation of monotone and concave production frontiers, Journal of Productivity Analysis 3 (1992) 401–415.

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

W. H. Greene, Maximum likelihood estimation of econometric frontier functions, Journal of econometrics 13 (1980) 27–56.

T. Kuosmanen, A. Saastamoinen, T. Sipiläinen, What is the best practice for benchmark regulation of electricity distribution? comparison of dea, sfa and stoned methods, Energy Policy 61 (2013) 740–750.

S. Jradi, J. Ruggiero, Stochastic data envelopment analysis: A quantile regression approach to estimate the production frontier, European Journal of Operational Research 278 (2019) 385–393.

ANEEL, NT 66/2017-SRM/ANEEL: Abertura de Audiência pública para atualização dos parâmetros relacionados à definição dos Custos Operacionais Regulatórios, 2017.

A. Charnes, W. W. Cooper, E. Rhodes, Measuring the efficiency of decision making units, European journal of operational research 2 (1978) 429–444.

R. D. Banker, A. Charnes, W. W. Cooper, Some models for estimating technical and scale inefficiencies in data envelopment analysis, Management science 30 (1984) 1078–1092.

W. D. Cook, K. Tone, J. Zhu, Data envelopment analysis: Prior to choosing a model, Omega 44 (2014) 1–4.

A. E. LaPlante, J. Paradi, Evaluation of bank branch growth potential using data envelopment analysis, Omega 52 (2015) 33–41.

W. D. Cook, K. Tone, J. Zhu, Data envelopment analysis: Prior to choosing a model, Omega 44 (2014) 1–4.

H. W. Lampe, D. Hilgers, Trajectories of efficiency measurement: A bibliometric analysis of dea and sfa, European Journal of Operational Research 240 (2015) 1–21.

W. Meeusen, J. van Den Broeck, Efficiency estimation from cobb-douglas production functions with composed error, International economic review (1977) 435–444.

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

L. R. Christensen, D. W. Jorgenson, L. J. Lau, Transcendental logarithmic utility functions, The American Economic Review 65 (1975) 367–383.

S. C. Kumbhakar, C. K. Lovell, Stochastic frontier analysis, Cambridge university press, 2003.

R. E. Stevenson, Likelihood functions for generalized stochastic frontier estimation, Journal of econometrics 13 (1980) 57–66.

G. Casella, R. L. Berger, Statistical inference, volume 2, Duxbury Pacific Grove, CA, 2002.

N. Sartori, Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions, Journal of Statistical Planning and Inference 136 (2006) 4259–4275.

K. Tone, Advances in DEA theory and applications: with extensions to forecasting models, John Wiley & Sons, 2017.

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

K. C. Land, C. K. Lovell, S. Thore, Chance-constrained data envelopment analysis, Managerial and decision economics 14 (1993) 541–554.

A. Azzalini, The skew-normal and related families, volume 3, Cambridge University Press, 2013.

C. L. Bayes, M. D. Branco, Bayesian inference for the skewness parameter of the scalar skew-normal distribution, Brazilian Journal of Probability and Statistics (2007) 141–163.

R. de Barros Mesquita, Regulação de custos de distribuição de energia elétrica: uma análise comparativa das abordagens de benchmarking utilizadas em países europeus e latino-americanos (2017).

J. Kiefer, Sequential minimax search for a maximum, Proceedings of the American mathematical society 4 (1953) 502–506.

Y. Wang, S. Wang, C. Dang, W. Ge, Nonparametric quantile frontier estimation under shape restriction, European Journal of Operational Research 232 (2014) 671–678.

Table 9: Comparison of observed operational expenditures (OPEX) and regulated OPEX using the ad-hoc DEA-NDRS model proposed by the Brazilian regulator and the proposed SDEA model with seven outputs and twelve points crossing the efficiency frontier.

DSO	OPEX	ad-hoc	SDEA 7:12	ad-hoc (OPEX)	SDEA 7:12 (OPEX)
NOVA PALMA	R$ 5,762,588.20	119%	100%	R$ 6,857,479.96	R$ 5,762,590.10
RGE	R$ 304,029,938.00	118%	114%	R$ 358,755,326.84	R$ 346,274,757.35
MUXFELDT	R$ 2,304,748.20	118%	100%	R$ 2,719,602.88	R$ 2,304,749.96
PIRATININGA	R$ 300,814,764.80	118%	100%	R$ 354,961,422.46	R$ 300,814,982.71
COELCE	R$ 571,992,827.50	116%	134%	R$ 663,511,679.90	R$ 764,280,253.18
CPFL PAULISTA	R$ 793,720,878.10	116%	102%	R$ 920,716,218.60	R$ 812,695,095.61
JAGUARI	R$ 12,324,818.70	116%	100%	R$ 14,296,789.69	R$ 12,324,828.53
ELEKTRO	R$ 522,991,305.90	116%	100%	R$ 606,669,914.84	R$ 522,991,611.85
ETO	R$ 250,567,173.30	110%	134%	R$ 275,623,890.63	R$ 334,911,539.74
JOAO CESA	R$ 2,294,867.00	108%	100%	R$ 2,478,456.36	R$ 2,294,867.52
EMT	R$ 532,233,923.40	106%	109%	R$ 564,167,958.80	R$ 581,588,013.58
COSERN	R$ 294,446,203.60	100%	103%	R$ 294,446,203.60	R$ 304,032,526.77
BANDEIRANTE	R$ 352,211,426.30	100%	101%	R$ 352,211,426.30	R$ 355,037,237.00
SANTA MARIA	R$ 34,789,487.60	100%	100%	R$ 34,789,487.60	R$ 34,789,505.29
EBO	R$ 44,780,721.70	100%	100%	R$ 44,780,721.70	R$ 44,780,741.40
EMS	R$ 345,073,715.70	100%	100%	R$ 345,073,715.70	R$ 345,073,844.27
CEMAR	R$ 497,342,577.90	100%	100%	R$ 497,342,577.90	R$ 497,342,725.16
EPB	R$ 304,633,122.60	100%	97%	R$ 304,633,122.60	R$ 294,600,010.23
ELETROPAULO	R$ 1,457,826,768.00	100%	94%	R$ 1,457,826,768.00	R$ 1,371,785,519.80
CPEE	R$ 17,364,001.60	100%	93%	R$ 17,364,001.60	R$ 16,141,696.20
RGE SUL	R$ 325,979,415.40	100%	93%	R$ 325,979,415.40	R$ 303,018,731.96
CSPE	R$ 21,269,422.30	100%	92%	R$ 21,269,422.30	R$ 19,609,508.12
ESCELSA	R$ 336,224,538.90	100%	92%	R$ 336,224,538.90	R$ 309,024,265.67
MOCOCA	R$ 11,970,286.20	100%	90%	R$ 11,970,286.20	R$ 10,821,569.86
SANTA CRUZ	R$ 52,034,961.60	100%	84%	R$ 52,034,961.60	R$ 43,561,323.45
VALE PARANAPANEMA	R$ 47,981,646.70	100%	83%	R$ 47,981,646.70	R$ 39,844,285.17
LIGHT	R$ 922,882,323.90	97%	77%	R$ 895,195,854.18	R$ 712,183,553.55
CELESC	R$ 837,578,560.50	96%	110%	R$ 804,075,418.08	R$ 920,405,774.50
EMG	R$ 126,325,482.90	96%	100%	R$ 121,272,463.58	R$ 126,325,496.90
CFLO	R$ 15,795,965.10	95%	113%	R$ 15,006,166.85	R$ 17,844,047.16
COELBA	R$ 1,341,275,532.00	95%	90%	R$ 1,274,211,755.40	R$ 1,208,668,438.24
CAIUÁ	R$ 66,782,170.50	95%	73%	R$ 63,443,061.98	R$ 48,982,546.32
NACIONAL	R$ 32,718,396.70	95%	72%	R$ 31,082,476.87	R$ 23,528,016.88
COPEL	R$ 1,326,796,300.00	94%	86%	R$ 1,247,188,522.00	R$ 1,138,725,514.49
CEMIG	R$ 2,260,483,577.00	92%	86%	R$ 2,079,644,890.84	R$ 1,934,428,329.91
CHESP	R$ 15,617,197.50	91%	77%	R$ 14,211,649.73	R$ 12,008,674.41
ESE	R$ 185,887,399.00	87%	90%	R$ 161,722,037.13	R$ 167,949,494.89
CEPISA	R$ 393,252,856.10	87%	83%	R$ 342,129,984.81	R$ 325,748,487.68
CELPE	R$ 861,117,202.40	87%	83%	R$ 749,171,966.09	R$ 712,486,061.39
CELG	R$ 948,403,942.60	87%	82%	R$ 825,111,430.06	R$ 780,551,510.86
BRAGANTINA	R$ 47,991,957.30	84%	69%	R$ 40,313,244.13	R$ 32,898,743.28
CELPA	R$ 678,124,176.50	81%	77%	R$ 549,280,582.97	R$ 522,941,361.18
AMPLA	R$ 658,257,317.10	80%	74%	R$ 526,605,853.68	R$ 489,546,050.76
SULGIPE	R$ 43,027,634.40	78%	85%	R$ 33,561,554.83	R$ 36,706,714.44
COOPERALIANÇA	R$ 13,182,982.90	77%	85%	R$ 10,150,896.83	R$ 11,140,716.51
CEB	R$ 381,490,986.00	76%	100%	R$ 289,933,149.36	R$ 381,491,276.67
ELETROCAR	R$ 16,210,774.90	76%	63%	R$ 12,320,188.92	R$ 10,226,617.49
FORCEL	R$ 4,285,406.80	75%	100%	R$ 3,214,055.10	R$ 4,285,406.80
HIDROPAN	R$ 7,568,700.80	75%	91%	R$ 5,676,525.60	R$ 6,924,650.25
IGUAÇU	R$ 17,606,347.80	72%	64%	R$ 12,676,570.42	R$ 11,224,894.21
DEMEI	R$ 11,389,358.80	70%	55%	R$ 7,972,551.16	R$ 6,278,634.63
COCEL	R$ 20,538,217.80	68%	76%	R$ 13,965,988.10	R$ 15,562,344.68
CEAL	R$ 324,088,033.60	68%	66%	R$ 220,379,862.85	R$ 213,502,896.13
ENF	R$ 35,742,431.20	66%	68%	R$ 23,590,004.59	R$ 24,426,343.38
ELETROACRE	R$ 114,888,672.40	65%	59%	R$ 74,677,637.06	R$ 67,973,348.18
URUSSANGA	R$ 6,204,356.50	63%	88%	R$ 3,908,744.60	R$ 5,451,279.94
CERON	R$ 319,505,329.10	63%	61%	R$ 201,288,357.33	R$ 194,683,135.06
CEEE	R$ 643,515,960.10	57%	52%	R$ 366,804,097.26	R$ 333,008,065.07
AMAZONAS	R$ 468,120,687.50	44%	43%	R$ 205,973,102.50	R$ 200,331,286.43
DMED	R$ 46,844,597.20	39%	48%	R$ 18,269,392.91	R$ 22,596,279.83
BOA VISTA	R$ 91,656,721.10	37%	29%	R$ 33,912,986.81	R$ 26,772,653.93
TOTAL:	R$ 20,728,123,685.20			R$ 19,192,630,061.65	R$ 18,423,515,426.48

2. REFERÊNCIAS

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

R. D. Banker, Stochastic data envelopment analysis, Pittsburgh: CarnegieMellon University (1986).