UNIVERSIDADE FEDERAL DE MINAS GERAIS

ESCOLA DE ENGENHARIA

DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

SPATIAL STATISTICAL METHODS APPLIED TO THE 2015 BRAZILIAN ENERGY DISTRIBUTION BENCHMARKING MODEL: ACCOUNTING FOR UNOBSERVED DETERMINANTS OF INEFFICIENCIES

GUILHERME DÔCO ROBERTI GIL

BELO HORIZONTE
2016

Guilherme Dôco Roberti Gil

SPATIAL STATISTICAL METHODS APPLIED TO THE 2015 BRAZILIAN ENERGY DISTRIBUTION BENCHMARKING MODEL: ACCOUNTING FOR UNOBSERVED DETERMINANTS OF INEFFICIENCIES

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Engenharia de Produção.

Área de concentração: Modelagem Estocástica e Simulação

Orientador (a): Prof. Dr. Marcelo Azevedo Costa

Belo Horizonte

Escola de Engenharia da UFMG

2016

AGRADECIMENTOS

Desde o surgimento do propósito de realizar esse mestrado, muitos fatores e pessoas intervieram para que fosse possível sua culminação. Agradeço a todos que de uma forma ou de outra participaram desta jornada e em especial:

Agradeço primeiro a Deus que permitiu essa grande oportunidade;

À minha esposa e companheira, Nara, que me estimula a cada dia ser melhor e teve que ter muita paciência com minhas privações de tempo.

Aos meus pais, Valério e Maria Lúcia, e aos meus irmãos, Camilla, Gustavo e Pietro que simplesmente o fato de existirem já contribuem para minha felicidade.

Ao Prof. Marcelo Azevedo que me orientou durante todo esse trabalho, sempre com muita paciência e tolerância com minhas incompreensões e incapacidades. Seu exemplo de dedicação e orientação sedimentaram muitos conhecimentos e meu acervo pessoal. Foi sem dúvidas determinante para a conclusão deste trabalho.

À Professora Ana Lopes e ao Professor Vinícius Mayrink que a partir de suas contribuições conseguimos robustecer todo o trabalho.

Aos meus amigos e sócios que auxiliaram e estimularam a completar esse trabalho.

" La alegría del triunfo jamás podría ser experimentada si no existiera la lucha que es la que determina la oportunidad de vencer."

Raumsol

RESUMO

Em 2015 o regulador brasileiro de energia elétrica, ANEEL, apresentou um modelo de benchmarking, baseado no DEA (Data Envelopment Analysis), para definir as metas de custos operacionais para as 61 concessionárias de distribuição de energia elétrica, a serem alcançadas em 4 anos. O modelo DEA utiliza os custos operacionais ajustados como variável de insumo, sete variáveis de produtos e restrições nos pesos. Embora variáveis não-discricionárias ou variáveis ambientais estivessem disponíveis no banco de dados, o regulador argumentou que não foram encontradas relações estatísticas significativas entre os escores de eficiência do modelo DEA e as variáveis não-discricionárias. Este estudo avalia a relação entre os escores de eficiência da DEA e as variáveis ambientais disponíveis. Além disso, métodos de estatística espacial são utilizados para mostrar que os escores de eficiência calculados pelo regulador são geograficamente correlacionados. Devido à diversidade ambiental e o grande território geográfico é improvável que apenas um componente ambiental é suficiente para corrigir as eficiências em todos os territórios brasileiros. Dessa forma, uma nova variável ambiental combinada é proposta. Por fim é apresentado um modelo de segundo estágio utilizando a variável ambiental proposta e uma estrutura espacial latente. Os resultados apresentaram grandes diferenças entre os escores de eficiência originais e os corrigidos, principalmente para as concessionárias de distribuição de energia elétrica localizadas em ambientes mais hostis e que originalmente apresentaram escores de eficiência baixo.

Existem muitas alternativas apresentadas na literatura para contabilizar o impacto das variáveis ambientais nos modelos de Benchmarking. Por exemplo, em modelos SFA e StoNED as variáveis ambientais podem ser incluídas diretamente com as variáveis de insumo e de produtos. Então, o modelo pode ser estimado em um estágio. O segundo estágio do DEA requer fortes suposições entre as variáveis no primeiro e segundo estágios. Uma suposição importante é que as variáveis ambientais e as variáveis do primeiro estágio sejam independentes, a fim de produzir estimadores não viesados. Isto evidentemente não é o caso do modelo Benchmarking das distribuidoras brasileiras de energia elétrica, uma vez que o melhor modelo de segundo estágio inclui a *e.variable* que está correlacionada com a variável consumidor-hora de energia interrompida (CHI), incluída no primeiro estágio do modelo. No entanto, Yu et al. (2009) utilizam as variáveis ambientais que foram correlacionados com as variáveis do primeiro estágio.

Além disso, é importante destacar que o atual modelo de Benchmarking brasileiro não leva em consideração qualquer componente ambiental que poderia impactar nos escores de eficiência. Embora muitas contribuições foram enviadas ao regulador, o modelo final não foi alterado. Portanto, o presente trabalho oferece novas percepções sobre o potencial impacto da localização ambiental e geográficas das empresas de distribuição de energia elétrica brasileiras nos escores corrigidos. Os resultados apresentaram evidências que um conjunto de empresas com baixos escores de eficiência na região norte, que é um ambiente hostil para o desenvolvimento do negócio de distribuição. Empresas localizadas nestas áreas podem ter seus escores de eficiência melhorados por um fator de 30% a 40%. Atualmente, uma empresa localizada nesta região apresentou o menor escore de eficiência, 22,4%. Além disso, sugerimos incluir uma estrutura espacial latente no modelo. Isto ocorre porque a maioria das variáveis ambientais representam valores médios, que não apresentam eventos extremos que realmente impactam na eficiência. Ao fazer isso, o modelo pode estimar a hostilidade do ambiente, com a estrutura latente que é compartilhada entre as áreas geograficamente mais próximas.

Vale ressaltar que as estimativas Bayesianas propostas dos escores de eficiência corrigidos apresentaram grandes intervalos de credibilidade HPD. No entanto, a maioria dos escores de eficiência originais estão fora dos intervalos HPD, o que indica uma forte correlação estatística entre os escores de eficiência e a informação ambiental. Logo, pode-se concluir que existem evidências estatísticas que as empresas brasileiras de distribuição de energia elétrica são afetadas pelo meio-ambiente em que estão localizadas.

Além disso, diferentes procedimentos estatísticos podem ser aplicados para estimar uma única variável ambiental, tais como análise de componentes principais (PCA), análise fatorial (Yu et al., 2009), entre outros. Este trabalho propõe utilizar a variável FIE, que é um indicador de performance que sumariza tanto a informação ambiental quanto a ineficiência em gestão. Assumimos que diferentes empresas são afetadas por diferentes cenários ambientais, que é representado na variável FIE. Utilizando um modelo de regressão linear múltiplo foi estimado a variável FIE associada a 11 variáveis ambientais disponíveis. Diferentes combinações de variáveis ambientais foram avaliadas. No entanto, devido ao pequeno tamanho de amostra dos dados, modelos com múltiplas variáveis ambientais não foram estatisticamente significativas.

Bogetoft & Otto (2010) apresentam três diferentes equações para estimar os escores de eficiência corrigidos. Geralmente, a classificação dos escores de eficiência corrigidos não mudam utilizando as diferentes equações. No entanto, usando a moda da distribuição condicional, as empresas que

originalmente apresentaram escores de eficiência de 100% podem também atingir escores de eficiência corrigidos de 100%. Do contrário, de modo geral utilizando as outras equações de esperança condicional esses escores de eficiência corrigidos serão menores que 100%.

Bogetoft e Lopes (2015) afirmam que o modelo de benchmarking DEA proposto pela ANEEL é impreciso e apresenta valores atípicos para algumas empresas. É importante ressaltar que o modelo de segundo estágio proposto não corrige as imprecisões do modelo original. Portanto, ainda é necessária uma investigação mais aprofundada do modelo de primeiro estágio. No entanto, o presente trabalho fornece fortes evidências estatísticas de correlações entre os escores de eficiência originais e as variáveis ambientais. Além disso, foi proposto um modelo de segundo estágio que leva em consideração a dependência espacial entre as empresas. Acreditamos que este modelo possa ser estendido à diferentes cenários de regulação, tais como transmissão de energia elétrica, água e esgoto.

Spatial statistical methods applied to the 2015 Brazilian energy distribution benchmarking model: accounting for unobserved determinants of inefficiencies

Guilherme Dôco Roberti Gil[1], Marcelo Azevedo Costa[1]

[1] Department of Production Engineering - UFMG

Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - Pampulha - Belo Horizonte - MG, Brazil.
ZIP 31270-901

Abstract

In 2015 the Brazilian regulator presented a DEA benchmarking model to set the regulatory operational cost goals, to be reached in four years for 61 electricity distribution utilities. The DEA model uses: adjusted operational cost as the input variable, seven output variables and weight restrictions. Although non-discretionary variables or environmental variables are available in the dataset, the regulator argued that no statistically significant correlation was found between the DEA efficiency scores and the non-discretionary variables. This study evaluates the statistical correlation between the DEA efficiency scores and the available environmental variables. Spatial statistic methods are used to show that the efficiency scores are geographically correlated. Furthermore, due to Brazil's environmental diversity and large territory it is unlikely that only one environmental component is sufficient to adjust inefficiencies across the Brazilian territory. Thus, a new combined environmental variable is proposed. Finally, a second stage model using the proposed environmental variable and accounting for a spatial latent structure is presented. Results show major differences between original and corrected efficiency scores, mainly for utilities located in harsh environments and which originally achieved lower efficiency scores.

Keywords: *Data Envelopment Analysis, second stage analysis, spatial statistics, Bayesian analysis*.

Introduction

The most commonly used benchmarking models in electricity distribution regulation are: Data Envelopment Analysis (DEA; Charnes et al., 1978), Stochastic Frontier Analysis (SFA; Aigner et al., 1977; Meeusen and van den Broeck, 1977), Corrected Ordinary Least Squares (COLS) (Richmond, 1974) and Stochastic Semi-nonparametric Envelopment of Data (StoNED; Kuosmanen, 2006; Kuosmanen and Kortelainen, 2012). Briefly, DEA is a non-parametric linear programming model proposed by Charnes, Cooper and Rhodes (1978) which creates the efficiency frontier using a convex linear combination of inputs and outputs of decision making units (DMUs). SFA requires a parametric equation of the efficiency frontier and assumes a compound error, which represents deviations from the frontier. The compound error is the sum of stochastic inefficiencies and stochastic noise. StoNED is similar to SFA and DEA, with a compound stochastic error and with a non-parametric, piece-wise linear frontier. Lopes and Mesquita (2015) have shown that these models are very popular among the European electricity distribution regulators.

In general, input and output variables used in DEA, SFA, COLS and StoNED models are associated with controlled factors, i.e., production variables that can be managed by the decision maker in order to improve efficiency. Another set of variables - not necessarily less important - can affect production and are, generally, non-manageable. These variables are known as environmental or contextual variables (Ray, 1988).Examples of contextual variables are climatic factors (Yu, Jamasb and Pollitt, 2009) such as temperature, precipitation; soil type, farmers' level of education (Ray and Ghose, 2014); among others. The environmental variables affect the efficiency of companies but are, generally, beyond the scope of company's decisions.

Many alternatives have been proposed to adjust efficiency using environmental factors, such as one stage or second stage analysis. Benchmarking models such as SFA and StoNED allow the inclusion of environmental variables with the input and output variables, using one stage. If the efficiencies of DMUs are estimated using DEA, then second stage analysis is the most common approach. Second stage is based on regression models in which independent variables are the environmental variables.

The analysis of environmental variables was first introduced in DEA models by Banker and Morey (1986), which included the environmental variables in the model as a regular input/output variable. Ray (1988) introduced the second stage analysis, i.e., the efficiency scores are first estimated using the DEA model and then are correlated to the environmental variables. Ray (1991) included linear regression modeling to evaluate the statistical significance between the efficiency scores and the environmental variables. Since the efficiency scores are within the range 0 – 1, different statistical regression models such as Tobit regression (Tobin, 1958), maximum likelihood models (Aigner et al, 1977), Truncated regression (Johnson and Kuosmanen, 2012), ordinary least squares (OLS) (Montgomery, Peck, and Vining, 2012), among others, can be applied. The second stage analysis is useful to assist management decisions: the impact of significant environmental factors that negatively affect productivity can be minimized. For example, Ray and Ghose (2014) identified that farmers with higher levels of education and greater access to cutting-edge technologies have better productivity scores. Therefore, public policies could be implemented to increase the levels of farmers' education. Yu et al. (2009b) identified statistical significance between weather, cost and quality performance in electricity distribution companies.

The foundation of second stage analysis is that the estimated efficiency scores using input and output controlled variables can be updated based on the impact of environmental variables. That is, companies located in a favorable environment must have their efficiency scores decreased, in general, since the environment partially contributes to a higher efficiency score. On the contrary, companies located in a harsh environment must have their efficiency scores increased, in general, since the harsh environment prevents the companies from achieving higher efficiency scores. Second stage modeling to adjust efficiency scores are proposed by Simar and Wilson (2007), Banker and Natarajan (2008) and elsewhere. Second stage analysis depends on the nature of the problem being analyzed. If DMUs are subject to environmental settings, it is convenient to use second stage analysis. That seems to be the case for most published studies, including those concerned with regulatory purposes. Furthermore, it can be assumed that the geographic location of DMUs can also

be seen as a proxy of the environment, i.e., geographically closer DMUs may be subject to the same environmental setting. This is the foundation of spatial statistical analysis.

In the specific case of Brazilian regulation, second stage analysis may change significantly the efficiency scores of the distribution service operators (DSOs). Brazil is a country as large as a continent with 8.5 million km² and it is the 5th largest country in the world with 27 states, most of them larger than some European countries. It covers several climatic zones such as the humid tropics in the north, the semi-arid northeast and temperate areas in the south. These climatic differences lead to major ecological diversity, forming distinct biogeographic zones or biomes: the Amazon Rainforest, the largest tropical rainforest in the world; the Pantanal, the largest floodplain; the Cerrado, savannas and woodlands; the Caatinga semi-arid forests; the fields of the Pampas; and the tropical Atlantic rain forest. For instance, the dry season is very strong in the northeast, in which some municipalities face lack of rain for a few months, or even years. On the contrary, the north, south and southeast of Brazil face critical problems in the raining season like floodings, landslides, etc. Therefore, it is unlikely that the geographic location of the energy distribution companies does not impact their operational costs.

This study applies spatial statistics to evaluate whether estimated 2015 DEA efficiency scores of electricity distribution companies are geographically clustered in the Brazilian territory. A second stage based on stochastic frontier analysis (Aigner et al, 1977) with a latent spatial structure, to account for possible unknown geographical variation of the outputs is proposed. Corrected efficiency scores are estimated using environmental variables and the spatial latent structure. Results show major differences between original and corrected efficiency scores, mainly for companies that originally achieved lower efficiency scores. In addition, the electricity distribution companies located in risky areas, such as areas with flooding, dry regions, or poor regions, have their final efficiency scores increased; whereas electricity distribution companies with higher scores and located in wealthier regions have their final efficiency scores slightly decreased. On average, the new efficiency scores are higher than the original scores.

This paper is organized as follows. Section 2 reviews the second stage analysis for the DEA model and some elements of spatial statistics. It also introduces a new combined environmental analysis and presents the proposed second stage model with non-discretionary and geographically latent variables. Section 3 shows the results. Discussion and conclusion are found in section 4.

2. Materials and Methods

2.1 Background

On June 4, 2014, the Brazilian National Electricity Energy Agency (ANEEL) began a debate with Brazilian society regarding rules and methodologies for defining the revenues of electricity distribution utilities for the 4th Periodic Tariff Review Cycle (4PTRC) through public hearings 023/2014 (AP023). On December 4, 2014, ANEEL presented in Technical Note (TN) 407/2014, the proposed model to calculate regulatory operational costs. The technical note introduces an input oriented, non decreasing return to scale (NDRS), Data Envelopment Analysis - DEA model. This

model uses adjusted operational cost as the input variable and seven output variables: high voltage network extension, overhead network extension, underground network extension, weighted power consumption, total number of consumers, estimated number of consumer-hours with interrupted energy, and total amount of non-technical losses (Mega-Watt). The database consists of mean values for the most recent three years, from 2011 to 2013. A total of 61 distribution companies are evaluated, therefore the sample size is n = 61. Due to the small data size and the large number of variables, in general, the DEA model generates a larger number of companies with efficiency scores equals to one. To overcome this limitation, weight restrictions are included in the model. Furthermore, non-discretionary variables or environmental variables are available in the dataset. Nevertheless, the TN 407 argues that no statistically significant correlation was found between the efficiency scores and the non-discretionary variables. In addition, it argues that the quality variables defined as the estimated number of customer-hours with interrupted energy and total amount of non-technical losses were able to capture any underlying correlation between efficiency and non-discretionary variables.

Among the contributions, Bogetoft and Lopes (2015) suggest some improvements to ANEEL DEA model. The first suggestion is to include the number of distribution transformers as a new output variable.An extensive simulation study identified this variable as one important output, which is missing in the original model. The second suggestion is to exclude two distribution companies which were identified as outliers. The third suggestion is to evaluate two environmental variables: rain precipitation and frequency of interrupted energy (FEQ) in the second stage. The environmental variables were evaluated using univariate Tobit regression models (Tobin, 1958). Nevertheless, on April 24, 2015, ANEEL presented the final model in Technical Note 66/2015, in which the model presented previously (TN 407/2014) was not changed. That is, the effects of non-discretionary variables were not accounted for in the model, possible outliers were not evaluated and the number of distribution transformers was not included in the model.

2.2 The Brazilian regulator data set

A public data set is available from the Brazilian regulator. The data set comprises average values for seven output variables and one input variable, for the 61 energy distribution companies. Average values were calculated using yearly data from 2011 to 2013. The input variable is the mean operational cost for each company and the output variables are: underground network, overhead network, high voltage network, total number of consumers, weighted energy market, non-technical losses and consumer-hour interrupted energy. The last two output variables are, in fact, non-discretionary variables, which were included in the model as negative outputs (Bogetoft and Oto, 2010). In addition to input and output variables, 13 non-discretionary, or environmental variables are available: density of consumers, network density, complexity index, precipitation index, lightning rate, low vegetation index, medium vegetation index, high vegetation index, mean declivity index, proportion of paved roads, concession area (km^2), average duration of interrupted energy (DIE) and frequency of interrupted energy (FIE).

2.3 Data Envelopment Analysis

The DEA methodology was first introduced by Charnes et al. (1978) and extended by Banker et al. (1984). The method is widely used to estimate technical efficiencies of DMU. It calculates efficiency scores which range between zero and 1, for each DMU, using a mathematical programming method. Briefly, the method calculates the best practice frontier using a set of inputs and outputs, previously defined. The relative efficiency of each DMU is measured based on its distance from the efficiency frontier. The more inefficient the DMU, the farther is its distance from the efficiency frontier.

The Brazilian regulator (ANEEL) has applied DEA since 2011 (Costa et al, 2015), and the model has been revised recently. The current DEA model assumes non-decreasing returns to scale (NDRS), has seven output variables and one input variable. The efficiency scores are calculated using the following linear programming problem:

$$\max_{\gamma,\alpha,\varphi} \theta^{ref} = \sum_{j=1}^{m} \gamma_j y_j^{ref} + \varphi$$

$$\begin{aligned} &\text{subjectto:} \\ &\sum_{d=1}^{n} \alpha_d x_d^{ref} = 1, \\ &\sum_{i=1}^{m} \gamma_i y_i^k + \varphi - \sum_{d=1}^{n} \alpha_d x_d^k \leq 0, \quad (k = 1,2,\dots,K), \\ &\alpha_d, \gamma_i \geq 0; \ \varphi \geq 0, \end{aligned} \tag{1}$$

where θ^{ref} is the efficiency score estimated for the DMU ref, y_i^k are the outputs ($i = 1, \dots, m$), and x_d^k are the inputs ($d = 1, \dots, n$) for each DMU k; m is the total number of outputs, n is the total number of inputs, and K is the total number of DMUs. The α_d's are the input parameters, the γ_i's are the output parameters, and φ is the scale parameter.

Further details of the DEA Brazilian benchmarking model for energy distribution regulation can be found in Bogetoft and Lopes (2015). As previously mentioned, the current model does not include any efficiency score correction based on non-discretionary variables.

2.4 Second stage analysis for DEA models

This study relies on the second stage model presented by Banker and Natarajan (2008) and Johnson and Kuosmanen (2011). Johnson and Kuosmanen (2011) present the following data generating model, similar to the model presented by Banker and Natarajan (2008) and Johnson and Kuosmanen (2012):

$$x = \phi(\mathbf{y}) \cdot e^{\mathbf{z\delta}+u+v} \tag{2}$$

where x is the input variable, $\mathbf{y}$is the output vector, $\mathbf{z}$ is the vector of non-discretionary variables, $\boldsymbol{\delta}$ is the vector of non-negative weights associated with the discretionary variables, $\phi(\cdot)$ is the best practice frontier, u is a positive random variable representing technical inefficiency and v is a random variable representing stochastic noise. It is assumed that the random variables u and v are independent. The density distribution of v, $\varphi_v(v)$, is symmetric with a mean of zero. Banker and Natarajan (2008) uses a two-sided truncated normal distribution for v, $|v| \leq V^M$. The $\mathbf{z}\boldsymbol{\delta}$ component represents the technical inefficiency which is explained by the environmental variables. Thus, u is the unexplained inefficiency or the corrected inefficiency.

Equation (2) can also be written as:

$$\log \frac{x_i}{\phi(\mathbf{y}_i)} = \mathbf{z}_i\boldsymbol{\delta} + u_i + v_i, \tag{3}$$

where i,hereafter, is the DMU index. The first stage uses DEA to estimate the frontier, $\hat{\phi}(\mathbf{y}_i)$. Thus, equation (3) can be rewritten using the estimated efficiency scores,

$$-\log \hat{\theta}_i = \mathbf{z}_i\boldsymbol{\delta} + u_i + v_i. \tag{4}$$

Equation (4) is a linear regression equation in which the error component is written as $\varepsilon_i = u_i + v_i$. Banker & Natarajan (2008) suggest using ordinary least squares to estimate the vector of weights, $\boldsymbol{\delta}$. Nevertheless, in order to estimate the technical inefficiency, the parameters of the compound error (ε) must be estimated using maximum likelihood.

Different density distributions have been investigated for u_i; Meeusen and van den Broeck (1977) use an exponential distribution, Aigner et al (1977) use half-normal distribution, Stevenson (1980) uses truncated normal distributions, Greene (1990) uses a Gamma distribution, Mignon and Mignon (2005) use a log-normal distribution. Nevertheless, assuming v_i as a normal distribution, $v_i \sim N(0, \sigma_v^2)$, and u_i as a half-normal distribution, $u_i \sim N^+(0, \sigma_u^2)$, the distribution of $\varepsilon_i = v_i + u_i$ can be written as

$$\varphi_{u+v}(\varepsilon) = \frac{\sqrt{2}}{\sqrt{\pi\sigma^2}} \Phi\left(\frac{\lambda\varepsilon}{\sqrt{\sigma^2}}\right) exp\left(-\frac{1}{2}\frac{\varepsilon^2}{\sigma^2}\right), \tag{5}$$

where $\lambda = \sqrt{\sigma_u^2/\sigma_v^2}$, $\sigma^2 = \sigma_v^2 + \sigma_u^2$, and $\Phi(\cdot)$ is the distribution function of a standard normal distribution. Equation (5) represents a particular case of a Skew-normal distribution (Azzalini, 2013). Aigner et al (1977) claim that this representation is very convenient, since λ represents the ratio between inefficiency and noise; $\lambda^2 \to 0$ implies that $\sigma_v^2 \to \infty$ and/or $\sigma_u^2 \to 0$, i.e., the noise component dominates the error ε_i. Otherwise, if $\sigma_v^2 \to 0$, the error is dominated by the inefficiency and, therefore, deviations from the efficiency frontier are solely due to technical inefficiency.

From Equation 5, the log-likelihood equation for the second stage model can be written as

$$l(\boldsymbol{\delta}, \sigma^2, \lambda) = -\frac{1}{2}N\log\left(\frac{\pi}{2}\right) - \frac{1}{2}N\log\sigma^2 + \sum_{i=1}^{N}\log\Phi\left(\frac{\lambda\left(-\log\hat{\theta}_i - \mathbf{z}_i\boldsymbol{\delta}\right)}{\sqrt{\sigma^2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(-\log\hat{\theta}_i - \mathbf{z}_i\boldsymbol{\delta}\right)^2 \quad (6)$$

Thus, using Equation 6, maximum likelihood estimates for $\boldsymbol{\delta}$, σ^2 and $\lambda$are found. In addition, using Equation 4, it can be shown that the corrected efficiency scores, adjusted by the environmental variables, is $\tilde{\theta}_i = \mathrm{e}^{-u_i}$ (Bogetoft and Otto, 2010). Therefore, a proper estimate for $\hat{u}_i$ is required. Sinceε is a random variable which carries information about u, using Bayes Rule, the conditional density of u given ε can be written as

$$\varphi(u|\varepsilon) = \frac{\varphi_v(\varepsilon - u)\varphi_u(u)}{\varphi_{u+v}(\varepsilon)}. \quad (7)$$

Using Equation 7, an optimal estimator for the adjusted efficiency scores, and most often used (Bogetoft & Otto, 2010) is

$$\tilde{\theta} = E[\mathrm{e}^{-u}|\varepsilon] = \frac{1}{\Phi\left(\frac{\lambda}{\sigma}\varepsilon\right)}\Phi\left(\frac{\mu}{\sigma_*}\right)\mathrm{e}^{-\left(\mu+\frac{1}{2}\sigma_*^2\right)}, \quad (8)$$

where $\sigma_*^2 = \frac{\sigma_v^2\sigma_u^2}{\sigma^2}$, $\mu = \frac{\varepsilon\sigma_u^2 - \sigma_v^2\sigma_u^2}{\sigma^2}$ and $\hat{\varepsilon}_i = -\log\hat{\theta}_i - \mathbf{z}_i\hat{\boldsymbol{\delta}}$.

2.5 Estimation of a combined environmental variable

As described in Section 2.2, there are 13 potential environmental variables in the data set, which has a limited sample size. Furthermore, the FIE variable has the largest Spearman correlation coefficient ($\rho = -0.4705$) with the efficiency scores. The FIE is correlated to most of the remaining environmental variables**.** Nevertheless, it can be argued that duration of interrupted energy (DIE) and frequency of interrupted energy (FIE) carry both environmental and management inefficiency information.

In order to create a second stage model, which accounts only environmental information, we propose a new environmental variable using a multiple linear regression model (Montgomery et al., 2001). The proposed model uses FIE as the dependent variable (y) and the following 11 environmental variables as the independent environmental variables: density of consumers, network density, complexity index, precipitation index, lightning rate, low vegetation index, medium vegetation index, high vegetation index, mean declivity index, proportion of paved roads and concession area (km^2). The estimate of the new variable is:

Let $\mathbf{Y}$ be the vector of the dependent observations, $\mathbf{Y}^{\mathrm{T}} = [\mathrm{y}_1, \ldots, \mathrm{y}_{61}]$ and $\mathbf{X}$ be the regression matrix, $\mathbf{X} = [\mathbf{1}\ \mathbf{x}_1 \ldots \mathbf{x}_{11}]$, where $\mathbf{1}$ is the unit vector and $\mathbf{x}_i$is the vector of the i^{th}-environmental

variable ($i = 1, \ldots, 11$). The new environmental variable, hereafter named *e.variable* is calculated as:

$$e.variable = \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}, \tag{9}$$

where λ, $0 \leq \lambda < \infty$, is known as the biasing parameter (Montgomery et al., 2001, pg. 350). Equation (9) is known as the Ridge Regression estimator (Hoerl and Kennard, 1970) and it is used to control linear dependence among the independent variables. The value of λ can be chosen in order to maximize the *Prediction Error Sum of Squares* (PRESS) statistic. That is, Equation (9) provides the best predictive linear model of the FIE using only environmental variables. Further information about Ridge regression and PRESS statistic are found in Montgomery et al. (2001).

2.6 Spatial Statistics

2.6.1 Spatial autocorrelation: Moran's index.

In spatial studies, it is of interest to evaluate the existence of spatial autocorrelation and the strength of the correlation between small area units. In practice, spatial autocorrelation measures the similarities between area units. It takes into account the distance between them or any spatial information, such as geographic adjacency. Therefore, it is expected that closer area units share similar measures, for example, rain rates.

The most commonly used statistic to measure the intensity of spatial autocorrelation between area units is the Moran index, or simply Moran's I (Moran, 1950), shown in Equation (10). Given a region divided into n areas, consider the random variables $Y_1, Y_2, \ldots, Y_n$, associated with the units $1,2, \ldots, n$, the Moran's *I* statistic is given by

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,(Y_i - \bar{Y})\left(Y_j - \bar{Y}\right)}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)\sum_{i=1}^{n}(Y_i - \bar{Y})^2}, \tag{10}$$

where w_{ij} is the spatial weight of the link between$i$and j. Spatial weights represent proximities between areas. By definition,$w_{ii} = 0$, for $i = 1, 2, \ldots, n$. One alternative is to define $w_{ij} = 1$ if areas i and j share geographical boundaries, and to define $w_{ij} = 0$, if not. This is the most common choice, and is used in our analysis. The k-nearest neighbors and the closest areas within a circle with radius $r$is another possibility; however, are also common choices. Nevertheless, the optimal values for the parameter$k$ and the radius r must be estimated.

Moran's I ranges from -1 to 1, as similar to the Pearson linear correlation index. A value of -1 indicates perfect negative spatial correlation; while the value of 1 indicates perfect positive spatial correlation; a value of 0 indicates spatial randomness.

It is of interest to investigate the evidence of spatial randomness. In this case, the null hypothesis is defined as the condition in which the random variables $Y_1, Y_2, \ldots, Y_n$, are independent and identically distributed. Under the null hypothesis the distribution of the Moran's I statistic is asymptotically and

normally distributed (Cliff and Ord, 1981; Moran, 1950) with mean of $-1/(n-1)$and variance given by:

$$Var(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2 S_0^2}$$

with $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$, $S_1 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(w_{ij} + w_{ji}\right)^2$ and $S_2 = \sum_{k=1}^{n}\left[\left(\sum_{j=1}^{n} w_{kj}\right) + \left(\sum_{i=1}^{n} w_{ik}\right)\right]$.

Alternative measures of spatial correlations are found in the literature, for example, Geary's **C** index (Geary, 1954) and the Gamma index (Mantel, 1967). Nevertheless, Moran's I is the most commonly used statistic to evaluate spatial correlation.

2.6.2 Conditionally autoregressive models

Although the conditional autoregressive models (CAR) were first introduced by Besag (1974), they became popular after the advent of Markov Chain Monte Carlo (MCMC) methods; in particular the Gibbs sampler algorithm (Geman and Geman, 1984; Gelfand and Smith, 1990). Given a vector of random variables, $\mathbf{\Psi} = (\psi_1, \psi_2, \dots, \psi_n)^T$, the CAR models define the full conditional distribution of ψ_k given the remaining random variables, ψ_{-k}, as $f_\psi(\psi_k | \psi_{-k})$ (Lee, 2011). The intrinsic autoregressive (IAR) (Besag, 1991) model assumes the following full conditional normal distribution:

$$\psi_k \psi_k \sigma_I^2 \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\psi_j , \frac{\sigma_I^2}{n_k}\right), \tag{11}$$

where the conditional mean is the mean value of the random effects in the neighbors, except ψ_k, and the conditional variance is proportional to the inverse of the number of neighbors, n_k. Expression (11) can be written as:

$$f_{\mathbf{\Psi}}(\psi_1, \dots, \psi_n) \propto exp\left\{-\frac{1}{2\sigma_I^2}\sum\nolimits_{i\sim j} w_{ij}\left(\psi_i - \psi_j\right)^2\right\}, \tag{12}$$

This formulation does not provide a proper distribution (Banerjee et al. 2004); however it leads to a proper posterior distribution in a Bayesian analysis. It is worth noting that Equation (12) is written as a function of the spatial weights w_{ij}.Therefore, it represents a probabilistic model which accounts the effects of the neighbors of a random variable ψ_k. Further information about CAR models can be found in Banerjee et al. (2004).

2.7 Proposed second stage model with non-discretionary and geographically latent variables

The CAR model has been previously applied to account for environmental and spatially structured components in a stochastic frontier benchmarking model (Schmidt et al, 2009). The stochastic frontier model (SFA) estimates the efficiency frontier using input, output, and environmental variables, simultaneously. Nonetheless, further investigation is required to include weight restrictions in the SFA model. In addition, the DEA model is the current benchmarking model used

by the Brazilian regulator. Therefore, we propose a second stage model in order to adjust the efficiency scores which were primarily estimated by the regulator. The proposed second stage is similar to the SFA model in which the input is the negative of the logarithm of the DEA efficiency score, as shown in Equation (4). The proposed model is written as

$$
\begin{aligned}
-\log \hat{\theta}_i &= \mathbf{z}_i\boldsymbol{\delta} + u_i + v_i, \\
v_i &\sim N(0, \sigma_v^2), \\
u_i|\psi_i &\sim N^+(\psi_i, \sigma_u^2), \\
\psi_i\,|\,\psi_{-i}, \sigma_I^2 &\sim N\left(\frac{1}{n_i}\sum_{j\sim i}\psi_j\,, \frac{\sigma_I^2}{n_i}\right).
\end{aligned}
\tag{13}
$$

In this case, the density distribution of $\varepsilon = v + u$, given ψ, can be written as

$$
\varphi_{\varepsilon|\psi}(\varepsilon) = \frac{1}{\Phi\left(\frac{\psi}{\sigma_u}\right)}\Phi\left(\frac{\lambda}{\sigma}\varepsilon + \frac{1}{\lambda\sigma}\psi\right)\phi\left(\frac{\psi-\varepsilon}{\sigma}\right)\frac{1}{\sigma}. \tag{14}
$$

The optimal estimator for the adjusted efficiency scores, similar to Equation (8), is

$$
\tilde{\theta} = E[\mathrm{e}^{-u}|\varepsilon,\psi] = \frac{1}{\Phi\left(\frac{\lambda}{\sigma}\varepsilon + \frac{1}{\lambda\sigma}\psi\right)}\Phi\left(\frac{\mu}{\sigma_*}\right)\mathrm{e}^{-\left(\mu+\frac{1}{2}\sigma_*^2\right)}, \tag{15}
$$

where $\sigma_*^2 = \frac{\sigma_v^2\sigma_u^2}{\sigma^2}$, $\mu = \frac{\varepsilon\sigma_u^2+\psi\sigma_v^2-\sigma_v^2\sigma_u^2}{\sigma^2}$ and $\hat{\varepsilon}_i = -\log\hat{\theta}_i - \mathbf{z}_i\hat{\boldsymbol{\delta}}$.

In addition, alternative estimators (Bogetoft and Otto, 2010) are

$$
\tilde{\theta}^{[2]} = \mathrm{e}^{-E[\mathrm{u}|\varepsilon,\psi]} = \mu + \frac{\Phi\left(-\frac{\mu}{\sigma_*}\right)}{\Phi\left(\frac{\mu}{\sigma_*}\right)}\sigma_*, \tag{16}
$$

$$
\tilde{\theta}^{[3]} = \mathrm{e}^{-M[\mathrm{u}|\varepsilon,\psi]} = \begin{cases} 0, & if\ \mu \le 0 \\ \mu, & if\ \mu > 0 \end{cases}, \tag{17}
$$

where $M[\mathrm{u}|\varepsilon,\psi]$ is the mode of the conditional distribution.

2.7.1 Bayesian estimates of the proposed model

The following equation must be solved in order to estimate parameters $\boldsymbol{\delta}$, σ_u^2, σ_v^2 and σ_I^2 using the likelihood function:

$$L(\boldsymbol{\psi}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2) = \int_{\boldsymbol{\psi}} P(\boldsymbol{\varepsilon}|\boldsymbol{\psi}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2) \cdot P(\boldsymbol{\psi}|\sigma_I^2) d\,\boldsymbol{\psi}, \tag{18}$$

where $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_n]$ and $\varepsilon_i = -log\hat{\theta}_i - \mathbf{z}_i\boldsymbol{\delta}$. The likelihood function shown, in Equation 18, does not have a closed form. One alternative is to use numerical integration methods combined with numerical optimization methods. Nonetheless, this approach is computationally intensive and may not converge to the maximum likelihood solution.

Another approach is to use Bayesian statistics and generate samples from the *posterior* distribution, shown in Equation (19), using Gibbs sampling.

$$P(\boldsymbol{\psi}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2, \sigma_I^2|\boldsymbol{\varepsilon}) \propto P(\boldsymbol{\varepsilon}|\boldsymbol{\psi}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2) \times P(\boldsymbol{\psi}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2|\sigma_I^2) \times P(\sigma_I^2). \tag{19}$$

According to the Bayesian paradigm, the unknown parameters $\boldsymbol{\delta}$, σ_u^2, σ_v^2 and σ_I^2 are treated as random variables. As a consequence, *prior* distributions are specified to describe the researcher's initial uncertainties about their values. If the random variables $\boldsymbol{\delta}$, σ_u^2, σ_v^2 and σ_I^2 are assumed as independent, then the left term in Equation (19) can be rewritten as

$$P(\boldsymbol{\psi}, \boldsymbol{\delta}, \sigma_u^2, \sigma_v^2|\sigma_I^2) \times P(\sigma_I^2) = P(\boldsymbol{\psi}|\sigma_I^2) \times P(\sigma_I^2) \times P(\boldsymbol{\delta}) \times P(\sigma_u^2) \times P(\sigma_v^2), \tag{20}$$

where $P(\boldsymbol{\delta})$, $P(\sigma_I^2)$, $P(\sigma_v^2)$ and $P(\sigma_u^2)$ are *prior* distributions and $P(\boldsymbol{\psi}|\sigma_I^2)$ is defined in Equation (13). If the *prior* distributions are flat, say $P(\boldsymbol{\psi}|\sigma_I^2) \times P(\sigma_I^2) \times P(\boldsymbol{\delta}) \times P(\sigma_u^2) \times P(\sigma_v^2) \propto 1$, then the *posterior* distribution is proportional to the likelihood function. Thus, samples from the *posterior* distribution are, in fact, generated from the likelihood function.

Following Schmidt et al. (2009) and Kelsall and Wakefield (2002) the *prior* distributions below are chosen for the parameters:

$$\frac{1}{\sigma_u^2} \sim Gamma(\alpha = 0.001, \beta = 0.001),$$

$$\frac{1}{\sigma_v^2} \sim Gamma(\alpha = 0.001, \beta = 0.001),$$

$$\frac{1}{\sigma_I^2} \sim Gamma(\alpha = 0.5, \beta = 0.0005),$$

$$\delta \sim Normal(\mu = 0, \sigma^2 = 10^3).$$

Therefore, the prior distribution for the variance parameters (σ_u^2, σ_v^2, σ_I^2) has mean of 1 and variance of 1,000.

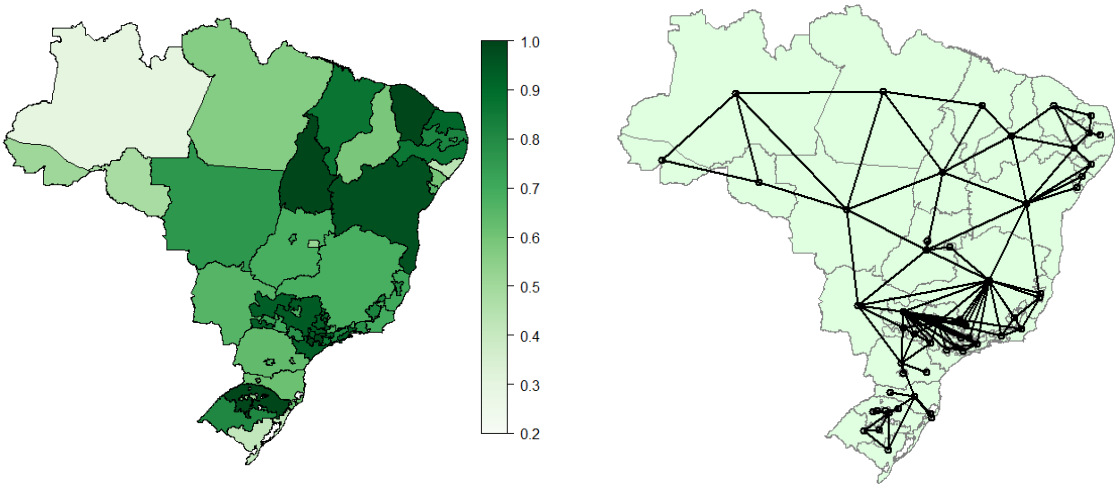
The **rbugs** package (Yan and Prates, 2013) available in the R software (R-project.org) was used to implement the proposed Bayesian model. 1,000 samples of the posterior distribution where drawn using MCMC (Markov Chain Monte Carlo) methods (Gamerman, 2006). Briefly, MCMC methods

create a sequential Markov chain of the parameters, which converges to the posterior distribution. In order to guarantee convergence of the chain and an independent sample of the parameters, the first 10^6 samples of the Markov chain were discarded (burn-in), and each sample was collected at every 500 sequential samples of the Markov chain (lag). In addition, convergence analysis of the chain was evaluated using the Geweke convergence test (Geweke, 1992), available in the R package **coda** (Plummer et al., 2006). Point estimates of the parameters were calculated using the median values of the posterior sample. High Probability Density (HPD) intervals with 95% of confidence level were also estimated using the posterior sample. The rbugs code of the proposed model is found in the Appendix.

Models with different environmental variables were compared using Watanabe-Akaike Information Criterion (WAIC), Deviance Information Criterion (DIC) and Log-Pseudo Marginal Likelihood (LPML) statistics. The best model fit is related to: lower value of DIC and larger values of WAIC and LPML. See Gelman el al. (2013) and Gamerman and Lopes (2006) for further information about these criteria.

3. Results

Figure 2(a) shows the distribution of the DEA efficiency scores in the Brazilian map, divided by the 61 energy distribution companies. There is evidence that the efficiency scores are geographically correlated. The lowest efficiency scores are mostly concentrated in the northern part of the country, where the equatorial forest, constant flooding of the Amazon river and its effluents, and other climate and vegetation aspects represent greater challenges to the efficiency of energy companies. There is also evidence of clusters of companies with larger efficiency scores, such as the companies located in the state of São Paulo (southeastern region of Brazil), as well as companies further south and in the northeastern regions. Therefore, in general, there is empirical evidence that the efficiency scores are geographically correlated. Table 1 presents the Moran's Index results for the efficiency scores and for the available environmental variables. The weight structure which was used in the Moran's equation (Equation 10) is presented in Figure 2 (b). The geographic adjacency structure is as follows: $w_{ij} = 1$ if energy distribution companies i and j share geographical border and $w_{ij} = 0$, if not. It is worth noticing that distribution companies in the northern region have large territories and are connected to fewer companies as compared to the distribution companies in southeast and south, which have smaller companies densely connected. From Equation (13), it can be seen that companies with lower numbers of connections have greater variances with respect to their latent variables (ψ_i), which might lead to larger confidence intervals for these companies.

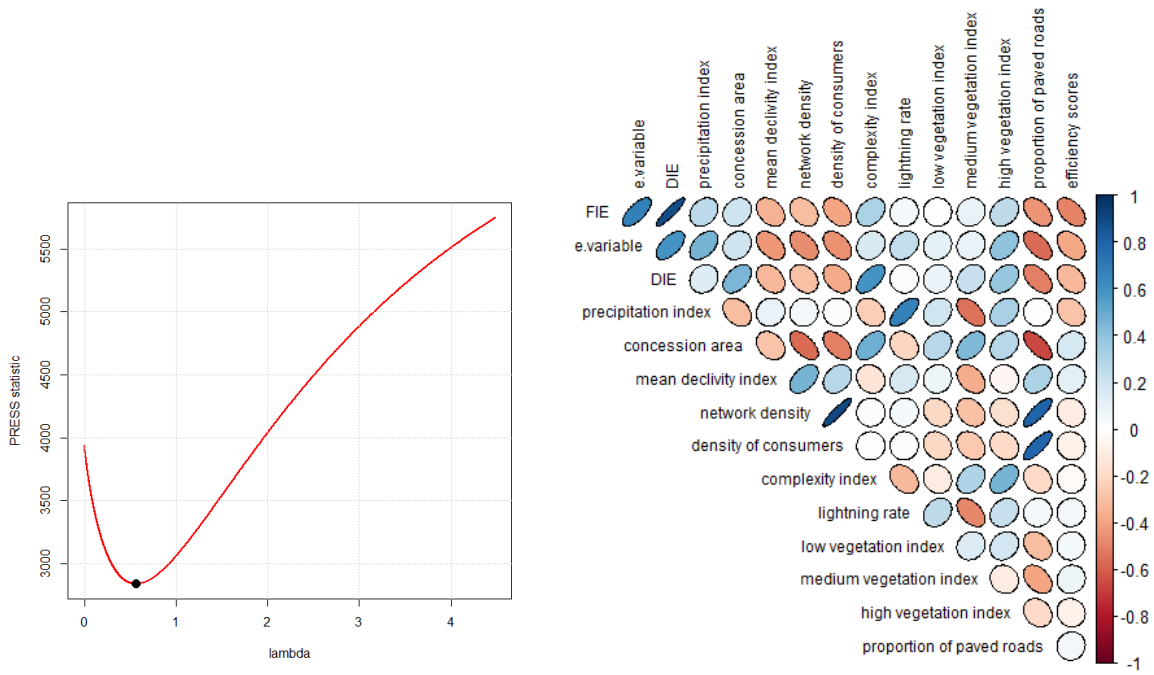


(a) Brazilian map of the DEA efficiency scores (TN 407) divided by the energy distribution companies.

(b) Weight structure generated based on geographical adjacency of the distribution companies.

Figure 2. Brazilian maps showing: (a) the distribution of the efficiency scores and (b) the weight structure generated based on geographical adjacency.

Figure 3(a) shows the PRESS statistic and the estimated λ parameter ($\hat{\lambda} = 0.5644$), which was used to create the *e.variable* (see Equation 9). Figure 3 (b) shows the Spearman (Spearman, 1904) correlation matrix using the environmental variables, FIE, DIE, DEA efficiency scores and the proposed *e.variable*. Positive correlations are represented by ellipses with positive slopes. Negative correlations are represented by ellipses with negative slopes. The larger (or smaller) the correlation the narrower is the ellipse. The second row shows the correlation among the proposed *e.variable* and the remaining variables. Note that the environmental variables are strongly correlated to the *e.variable*. Furthermore, the last column shows the correlation among DEA efficiency scores and the remaining variables. The *e.variable* and the efficiency scores have a large correlation coefficient ($\rho = -0.3757$). This correlation is smaller than the correlation between FIE and the efficiency scores ($\rho = -0.4705$). This is because the proposed *e.variable* carries only estimated environmental information from FIE.



(a) *Prediction Error Sum of Squares* (PRESS) statistic curve for different values of λ and the optimal value.

(b) Spearman correlation coefficients among environmental variables, FIE, DIE, e.variable and efficiency scores.

Figure 3. Curve of the PRESS statistic for different values of λ and the final estimate as the minimum value (a). Spearman correlation matrix among environmental, FIE, DIE, e.variable and the DEA efficiency scores (b).

Table 1 shows Moran's Index results sorted in decreasing order. The efficiency scores and most of the environmental variables are geographically correlated, i.e., P-values are smaller than 0.05 (5%), as expected. Note that medium vegetation index, precipitation index and the *e.variable* have the largest values of Moran's I, which indicates that these variables are strongly geographically correlated. One may argue that the inclusion of one of these variables in the second stage model is sufficient to handle the spatial distribution of the efficiency scores. To evaluate such a statement, the proposed second stage model was evaluated including one environmental variable and the spatial structure (latent component), i.e., one model for each environmental variable. Results are shown in Table 2.

Table 1. Moran's index for the efficiency scores and proposed non-discretionary variables.

Variable	Moran's Index	P-value
medium vegetation index	0.70	**0.000**
precipitation index	0.67	**0.000**
e.variable	0.64	**0.000**
high vegetation index	0.57	**0.000**
frequency of interrupted energy (FIE)	0.56	**0.000**
average duration of interrupted energy (DIE)	0.55	**0.000**
mean declivity index	0.48	**0.000**
lightning rate	0.46	**0.000**
complexity index	0.43	**0.000**
lowvegetation index	0.36	**0.000**
proportionof paved roads	0.24	**0.001**
efficiency scores	0.21	**0.004**
concession area (km2)	0.17	**0.008**
density of consumers	0.08	**0.043**
network density	-0.04	0.673

Table 2 shows estimated coefficients with the respective HPD intervals, and estimated density parameters for the model proposed in Equation (13). The *e.variable* is the only statistically significant variable, i.e., the zero is not included in the HPD interval. The model with the *e.variable* has a λ ratio of 1.040, which shows that the estimated values of σ_u and σ_v are very close. The ratio between σ_l and σ_u is 0.2573 (25.7%), i.e., the latent spatial structure gives a smaller contribution to the inefficiency variability.

Table 2. Results of the proposed model considering one environmental variable and the spatial latent structure.

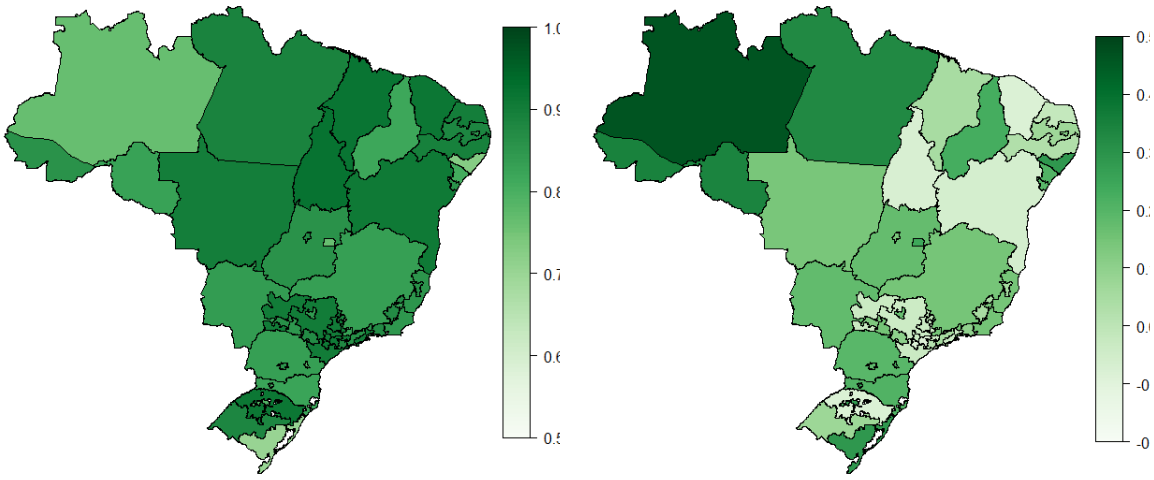
Environmental variables	Coefficient	Coefficient		σ_u	σ_v	σ_l	$\lambda=\sigma_u/\sigma_v$
		LowerLimit (95%)	UpperLimit (95%)				
medium vegetation index	-0.1447	-0.5294	0.1174	0.445	0.045	0.057	9.778
precipitation index	0.0000653	-0.0000334	0.0002150	0.313	0.161	0.059	1.887
e.variable	**0.0145**	**0.0051**	**0.0242**	**0.206**	**0.200**	**0.053**	**1.040**
high vegetation index	0.4884	-0.3362	1.3320	0.382	0.100	0.062	3.822
mean declivity index	-0.003063	-0.009379	0.006053	0.457	0.038	0.055	11.905
lightning rate	-0.00277	-0.00948	0.00534	0.447	0.037	0.063	12.195
complexity index	-0.053	-0.368	0.381	0.444	0.046	0.064	9.726
lowvegetation index	-0.339	-1.216	0.440	0.441	0.043	0.078	10.280

Table 3 shows the WAIC, DIC and LPML statistics for the selected models. The model with the *e.variable* has the largest WAIC and LPML statistics and the lowest value of the DIC statistic. Therefore, the proposed model with the *e.variable* achieves the best fit of the data.

Table 3. Model comparison criteria for each proposed model with one environmental variable and the spatial latent structure. The best model is shown in boldface.

Environmental variables	WAIC	DIC	LPML
medium vegetation index	-2.605	20.019	-2.796
precipitation index	-5.550	32.281	-5.745
e.variable	**-2.515**	**11.807**	**-2.605**
high vegetation index	-4.084	20.037	-4.242
mean declivity index	-2.285	16.856	-2.716

Figure 4 (a) shows the corrected efficiency scores map and Figure 4 (b) shows the absolute changes (corrected minus original) of the efficiency scores using the proposed model with the *e.variable* and the spatial structure. Corrected efficiency scores were calculated using Equation (15). Corrected scores are greater than original scores in the northern region, with differences close to 40%. Some energy companies had their efficiency scores reduced, mainly in the northeastern, southeastern and southern regions. Nevertheless, absolute differences between original and corrected scores for these companies were less than 10%. In general, energy companies with the lowest efficiency scores, which were located in the north, have the greatest increase in their corrected efficiency scores. These companies operate in the harshest environment: with high temperatures, high rain index, high humidity and high vegetation. Figure 5 compares the original efficiency scores with the corrected scores using our proposed model. Among the energy distribution companies, 14 companies had their efficiency scores reduced as compared to the original model. The remaining companies had their efficiency scores increased as compared to the original model.



(a) Brazilian map of the corrected scores.

(b) Brazilian map of the absolute changes in the efficiency scores.

Figure 4. Brazil maps showing: (a) the distribution of the corrected efficiency scores and (b) the map of the absolute changes in the efficiency scores.

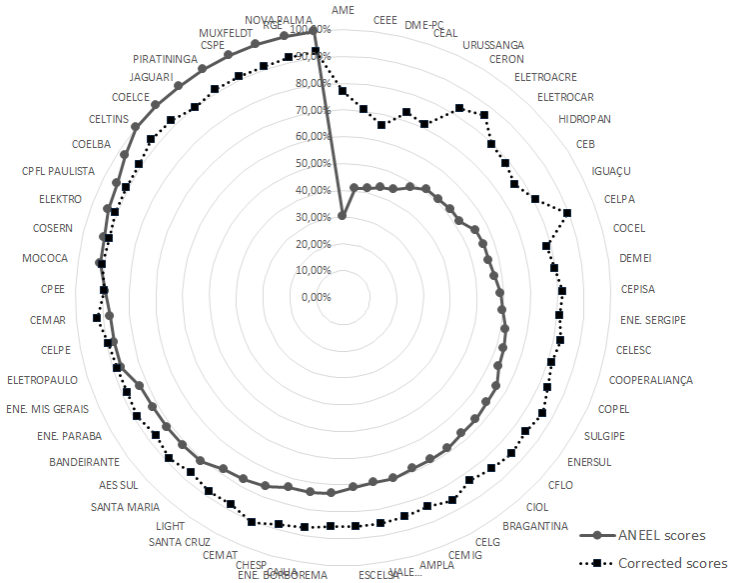
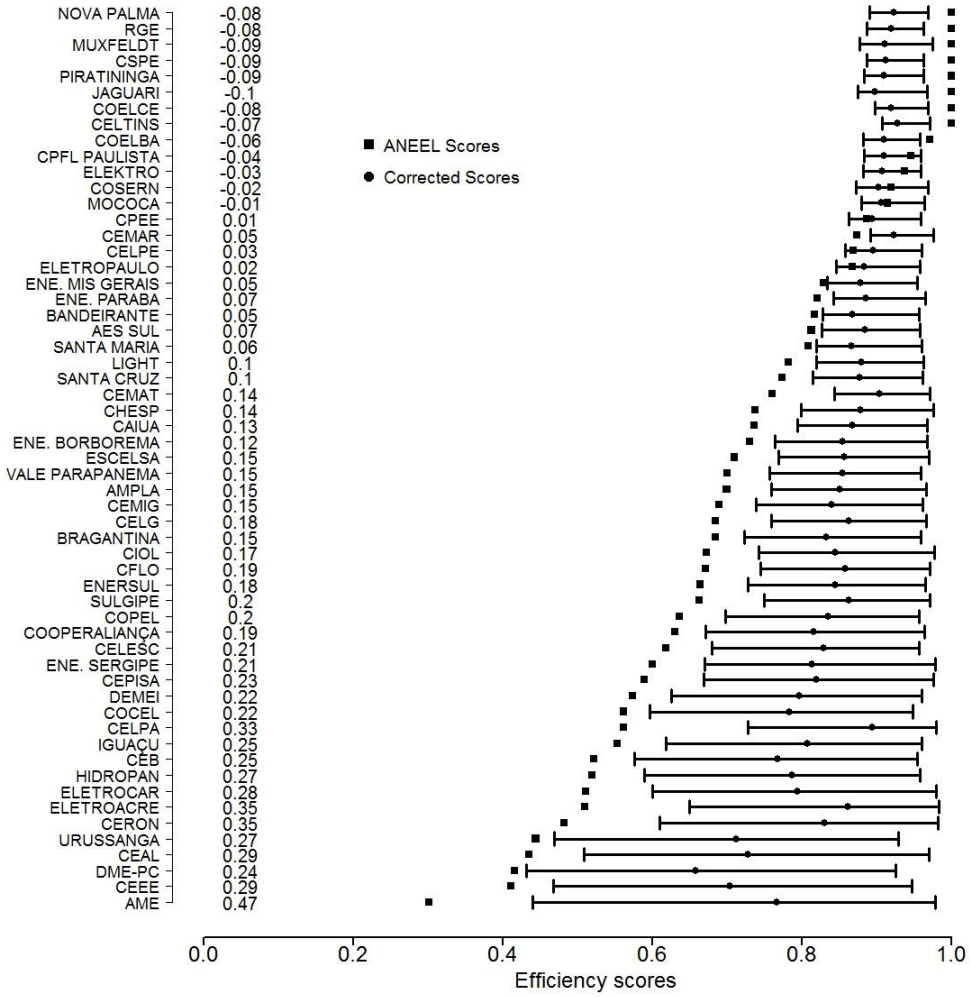


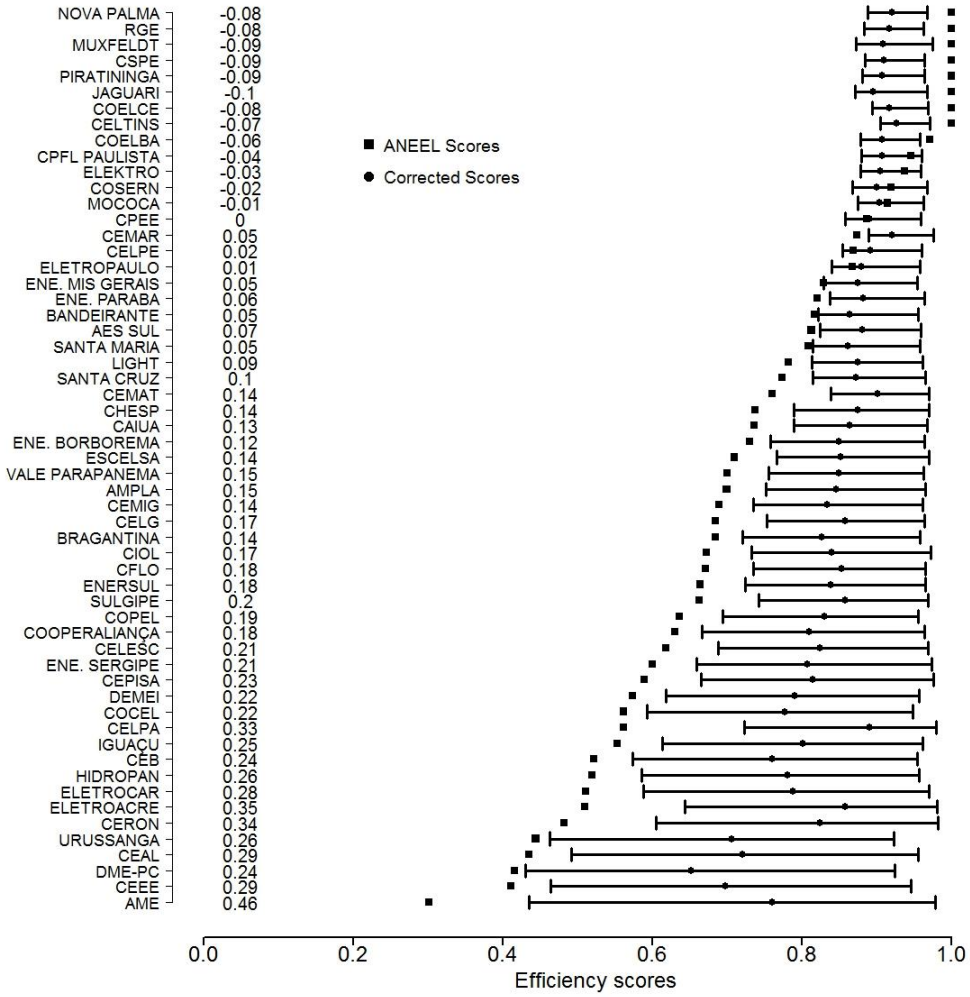
Figure 5. Radar plot comparing the original efficiency scores sorted in ascending order (clockwise), and the corrected efficiency scores using the proposed model.

Figure 6 shows the estimated corrected efficiency scores and the respective HPD intervals using Equations (15), (16) and (17). Results show that Equation (15), $\tilde{\theta} = E[\mathrm{e}^{-u}|\varepsilon, \psi]$, generates corrected scores which are less than 100%. As a consequence, companies which originally achieved efficiency scores of 100% had their corrected efficiency scores decreased. Equations (16) and (15) generate similar corrected efficiency scores. On the contrary, Equation (17), $\tilde{\theta}^{(3)} = \mathrm{e}^{-M[\mathrm{u}|\varepsilon,\psi]}$, generates corrected scores of 100%. That is, companies with original efficiency scores of 100% did not have their corrected efficiency scores changed.

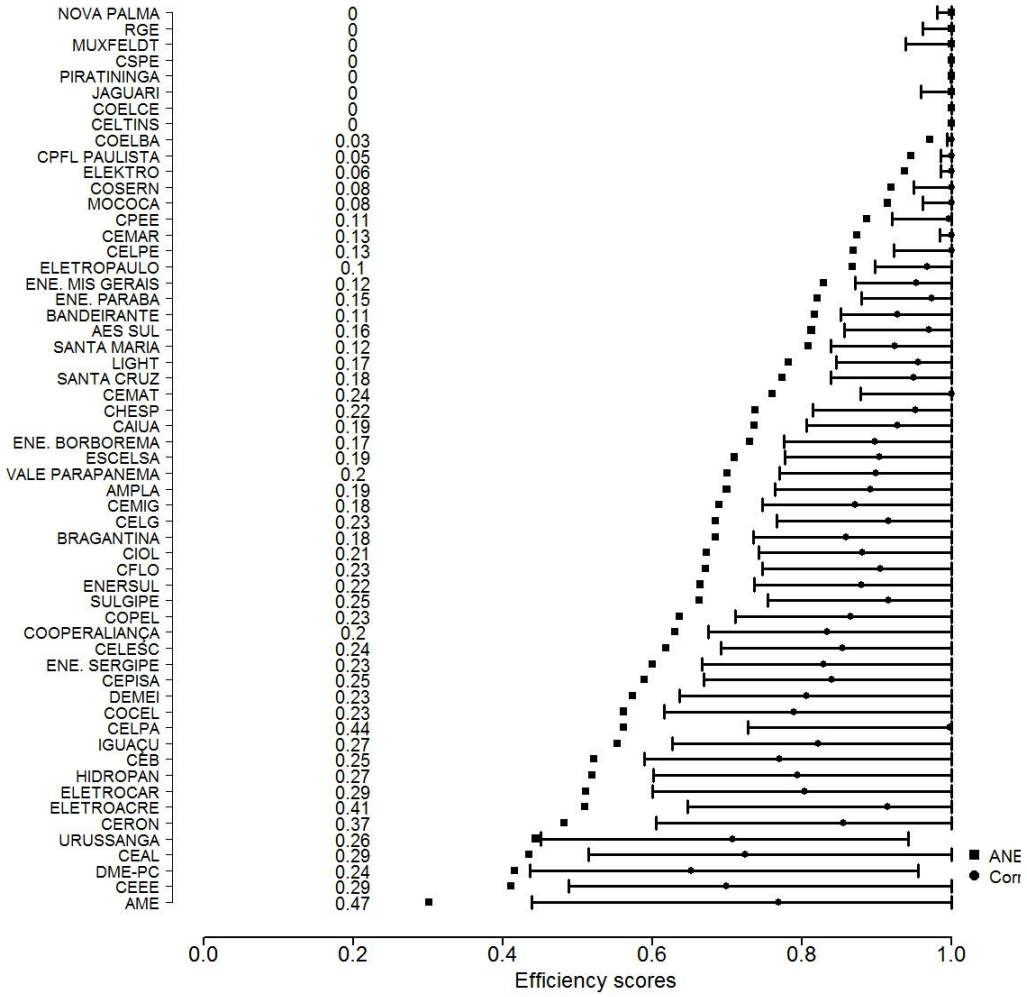
Finally, Figure 7 shows the corrected efficiency scores using the model with the spatial latent structure and the model without the spatial latent structure. This figure illustrates the impact of the latent spatial structure on the corrected efficiency scores, using Equation (15). Both models use the *e.variable*. The parameters of the model without the spatial structure were estimated using maximum likelihood. Figure 7 (a) shows that there is a slight difference between corrected efficiency scores with and without the spatial structure. As previously shown in Table 1, the *e.variable* is geographically correlated. Therefore, the estimated contribution of the spatial latent structure is small, compared to the model without the spatial structure. Differences between the corrected scores of the two models are shown in Figure 7 (b). It can be seen that the spatial latent structure contributed with small changes to the corrected efficiency scores.



(a) Corrected efficiency scores and HPD intervals using Equation 15.

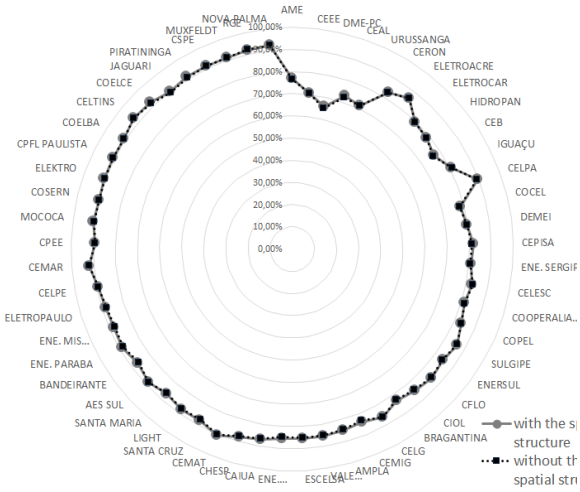


(b) Corrected efficiency scores and HPD intervals using Equation 16.

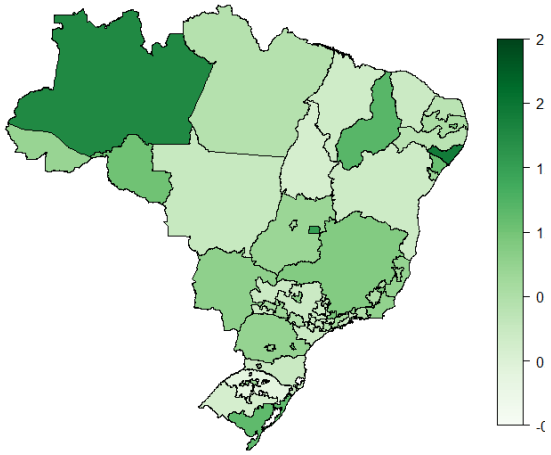


(b) Corrected efficiency scores and HPD intervals using Equation 17.

Figure 6. Corrected efficiency scores and HPD intervals using Equations 15 (a), 16 (b) and 17 (c). Energy companies are sorted in decreasing order of their original efficiency scores (black squares)



(a) Differences between corrected efficiency scores with and without the latent spatial structure for the energy distribution companies.



(b) Map of the differences between corrected efficiency scores with and without the latent spatial structure.

Figure 7. Comparison of the corrected efficiency scores with and without the latent spatial structure.

4. Discussion and conclusion

There are many alternatives in the literature to account for environmental variables in Benchmarking models. For example, in SFA and StoNED models the environmental variables can be included along with the input and output variables. Then, the model can be estimated in one stage. The DEA second stage requires strong assumptions among the variables in the first and second stages. One important assumption is that the environmental variables and the first stage variables are independent, in order to produce unbiased estimates. This is clearly not the case of the Brazilian distribution Benchmarking model, since the best second stage model includes the *e.variable* which is correlated with the consumer-hour of interrupted energy (CHI) variable, included in the first stage.

It is worth noticing that the current Brazilian benchmarking model does not account for any environmental component that could possibly impact the efficiency scores. Although many contributions were submitted to the regulator, the final model was not changed. Therefore, this work provides new insights into the potential impact of environmental and geographical location of the Brazilian DSOs in the corrected efficiency scores. Results provide evidence of a cluster of companies with lower efficiency scores in the northern region, which is a region with a harsh environment. DSOs located in this area can potentially have their scores increased by a factor of 30% to 40%. Currently, one company located in this region has the lowest efficient score (22.4%). In addition, we suggest including a latent spatial structure in the model. This is because most of the environmental variables represent average values, which do not include extreme events that truly impact efficiency. In this case, the model can estimate the environmental harshness, which is the latent structure shared between geographically closer areas.

It is worth mentioning that the proposed Bayesian estimates of the corrected efficiency scores achieve large HPD intervals, as shown in Figure 5. Nonetheless, most of the original efficiency scores are outside the HPD intervals, which indicates a strong statistical correlation between efficiency scores and environmental information. As a conclusion, there is statistical evidence that the Brazilian DSOs are affected by the environment where they are located.

Different statistical procedures could have been applied to estimate a single environmental variable, such as principal component analysis, factorial analysis (Yu et al., 2009), among others. This work proposes to use the FIE variable, a key performance indicator (KPI) variable which summarizes both environmental and management inefficiency information. We assume that different DMUs are affected by different environmental settings, which is represented in the FIE variable. Using a multiple regression model the FIE component, which is associated with the 11 available environmental variables, can be estimated. Different combinations of the environmental variables were also evaluated. However, due to the small sample size of the data, models with multiple environmental variables were not statistically significant.

Bogetoft & Otto (2010) present three different equations for estimating the corrected efficiency scores. In general, the rank of the corrected efficiency scores does not change using the different equations. However, using the mode of the conditional distribution (Equation 17), companies which originally achieved efficiency scores of 100% may also achieve corrected efficiency scores of 100%. On the contrary, using conditional expectation equations (15 and 16) the corrected efficiency scores are, in general, lower than 100%.

Bogetoft and Lopes (2015) claim that the 2014 Brazilian DEA benchmarking model is inaccurate, and has outliers. We highlight the fact that the proposed second stage model does not correct for inaccuracies in the first DEA stage. Therefore, further investigation in the first stage model is still required. Nonetheless, the present work provides strong statistical evidence of correlations among DEA efficiency scores and environmental variables. In addition, a second stage model accounting for spatial dependence is proposed. We believe this model can be extended to different regulation settings such as electricity transmission, water and sewage.

The proposed model can be further explored. One approach for future work is to investigate different spatial weight structures such as k-nearest neighbors, among others, as suggested by Schmidt et al. (2009). A second approach is to consider different conditionally autoregressive models (CAR), such as the Leroux model (Leroux, et al., 1999).

Acknowledgements

The authors thank CAPES, CNPq, FAPEMIG and FAPEMIG/CEMIG [Grant number:APQ-03165-11] for financial support.

References

AIGNER, Dennis; LOVELL, CA Knox; SCHMIDT, Peter. Formulation and estimation of stochastic frontier production function models. **Journal of Econometrics**, v. 6, n. 1, p. 21-37, 1977.

AZZALINI, Adelchi. **The skew-normal and related families**. Cambridge university press, 2013.

BANERJEE, Sudipto; CARLIN, Bradley P.; GELFAND, Alan E. **Hierarchical modeling and analysis for spatial data**. Crc Press, 2014.

BANKER, Rajiv D.; CHARNES, Abraham; COOPER, William Wager. Some models for estimating technical and scale inefficiencies in data envelopment analysis. **Management science**, v. 30, n. 9, p. 1078-1092, 1984.

BANKER, Rajiv D.; MOREY, Richard C. Efficiency analysis for exogenously fixed inputs and outputs. **Operations research**, v. 34, n. 4, p. 513-521, 1986.

BANKER, Rajiv D.; NATARAJAN, Ram. Evaluating contextual variables affecting productivity using data envelopment analysis. **Operations research**, v. 56, n. 1, p. 48-58, 2008.

BATTESE, George E.; CORRA, Greg S. Estimation of a production frontier model: with application to the pastoral zone of Eastern Australia. **Australian journal of agricultural economics**, v. 21, n. 3, p. 169-179, 1977.

BESAG, Julian. Spatial interaction and the statistical analysis of lattice systems. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 192-236, 1974.

BESAG, Julian; YORK, Jeremy; MOLLIÉ, Annie. Bayesian image restoration, with two applications in spatial statistics. **Annals of the institute of statistical mathematics**, v. 43, n. 1, p. 1-20, 1991.

BOGETOFT, Peter.; LOPES, Ana L. M. Comments on the Brazilian benchmarking model for energy distribution regulation: Fourth cycle of tariff review - Technical Note 407/2014. url: http://www.aneel.gov.br/aplicacoes/audiencia/dspListaContribuicao.cfm?attAnoAud=2014&attIdeFasAud=938&attAnoFasAud=2015&id_area=13, 2015.

BOGETOFT, Peter; OTTO, Lars. **Benchmarking with Dea, Sfa, and R**. Springer Science & Business Media, 2010.

CHARNES, Abraham; COOPER, William W.; RHODES, Edwardo. Measuring the efficiency of decision making units. **European journal of operational research**, v. 2, n. 6, p. 429-444, 1978.

CLIFF, Andrew David; ORD, J. Keith. **Spatial processes: models & applications**. London: Pion, 1981.

COSTA, Marcelo Azevedo; LOPES, Ana Lúcia Miranda; DE PINHO MATOS, Giordano Bruno Braz. Statistical evaluation of Data Envelopment Analysis versus COLS Cobb–Douglas benchmarking models for the 2011 Brazilian tariff revision. **Socio-Economic Planning Sciences**, v. 49, p. 47-60, 2015.

COOPER, William W.; SEIFORD, Lawrence M.; TONE, Kaoru. Discretionary, non-discretionary and categorical variables. **Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software**, p. 183-219, 2000.

CRESSIE, Noel. Statistics for spatial data: Wiley series in probability and statistics. **Wiley-Interscience New York**, v. 15, p. 16, 1993.

FARRELL, Michael James. The measurement of productive efficiency. **Journal of the Royal Statistical Society. Series A (General)**, v. 120, n. 3, p. 253-290, 1957.

GAMERMAN, Dani; LOPES, Hedibert F. **Markov chain Monte Carlo: stochastic simulation for Bayesian inference**. CRC Press, 2006.

GEARY, R. The contiguity ratio and statistical mapping. **The Incorporated Statistician** 5: pp115-45, 1954.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. **Bayesian Stat.**, V. 4, P. 169-188, 1992.

GELMAN, A., CARLIN, J., STERN, H., DUNSON, D., VEHTARI, A., E RUBIN, D. (2014), Bayesian Data Analysis, Chapman and Hall/CRC, 3 edn.

GEMAN, Stuart; GEMAN, Donald. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, n. 6, p. 721-741, 1984.

GREENE, William H. A gamma-distributed stochastic frontier model. **Journal of econometrics**, v. 46, n. 1, p. 141-163, 1990.

HOERL, Arthur E.; KENNARD, Robert W. Ridge regression: applications to nonorthogonal problems. **Technometrics**, v. 12, n. 1, p. 69-82, 1970.

JOHNSON, Andrew L.; KUOSMANEN, Timo. One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent StoNED method. **Journal of productivity analysis**, v. 36, n. 2, p. 219-230, 2011.

JOHNSON, Andrew L.; KUOSMANEN, Timo. One-stage and two-stage DEA estimation of the effects of contextual variables. **European Journal of Operational Research**, v. 220, n. 2, p. 559-570, 2012.

YAN, J.; PRATES, M. rbugs: Fusing R and OpenBugs and Beyond. **R package version 0.5-9, URL http://CRAN. r-project. org/package= rbugs**, 2013.

KELSALL, Julia; WAKEFIELD, Jonathan. Modeling spatial variation in disease risk: a geostatistical approach. **Journal of the American Statistical Association**, v. 97, n. 459, p. 692-701, 2002.

KUOSMANEN, Timo. Stochastic nonparametric envelopment of data: combining virtues of SFA and DEA in a unified framework. 2006.

KUOSMANEN, Timo, and KORTELAINEN, Mika. Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. **Journal of Productivity Analysis** 38, 2012, p. 11-28.

LEE, Duncan. A comparison of conditional autoregressive models used in Bayesian disease mapping. **Spatial and Spatio-temporal Epidemiology**, v. 2, n. 2, p. 79-89, 2011.

LEROUX, Brian G.; LEI, Xingye; BRESLOW, Norman. Estimation of disease rates in small areas: A new mixed model for spatial dependence. In: **Statistical models in epidemiology, the environment, and clinical trials**. Springer New York, 2000. p. 179-191.

LOPES, A.L.M.; MESQUITA, R.B. Tariff regulation of electricity distribution: a comparative analysis of regulatory benchmarking models. In: The 14th European Workshop on Efficiency and Productivity Analysis, 2015, Helsinki. Proceedings of the 14th European Workshop on Efficiency and Productivity Analysis, 2015.

MANTEL, Nathan. The detection of disease clustering and a generalized regression approach. **Cancer research**, v. 27, n. 2 Part 1, p. 209-220, 1967.

MEEUSEN, Wim; VAN DEN BROECK, Julien. Efficiency estimation from Cobb-Douglas production functions with composed error. **International economic review**, p. 435-444, 1977.

MESQUITA, R. B.; LOPES, A. L. M. ; CARDOSO, M. L. . Regulação Tarifária dos Custos da Distribuição de Energia Elétrica: uma Comparação entre os Modelos Europeus e o Brasileiro. In: XXXIX EnANPAD - Encontro Nacional da Associação de Pós-Graduação e Pesquisa em Administração, 2015, Belo Horizonte. Anais do XXXIX EnANPAD. Brasília: ANPAD, 2015.

MIGON H.S., MIGON M.N. Hierarchical bayesian models for stochastic frontier. **Estadística,** 2005, p. 57:27–52

MONTGOMERY, Douglas C. et al. Introduction to linear regression analysis. Wiley series in probability and mathematical statistics., 2001.

MORAN, Patrick AP. Notes on continuous stochastic phenomena. **Biometrika**, v. 37, n. 1/2, p. 17-23, 1950.

PLUMMER, Martyn et al. CODA: Convergence diagnosis and output analysis for MCMC. **R news**, v. 6, n. 1, p. 7-11, 2006.

TEAM, R. Core. Vienna (Austria): R foundation for statistical computing; 2012. **R: A language and environment for statistical computing**, 2013. URL http://www.R-project.org/.

RAY, Subhash C. Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation. **Socio-Economic Planning Sciences**, v. 22, n. 4, p. 167-176, 1988.

RAY, Subhash C. Resource-use efficiency in public schools: a study of Connecticut data. **Management Science**, v. 37, n. 12, p. 1620-1628, 1991.

RAY, Subhash C.; GHOSE, Arpita. Production efficiency in Indian agriculture: An assessment of the post green revolution years. **Omega**, v. 44, p. 58-69, 2014.

RICHMOND, James. Estimating the efficiency of production. International economic review, p. 515-521, 1974.

SCHMIDT, Alexandra M. et al. Spatial stochastic frontier models: accounting for unobserved local determinants of inefficiency. **Journal of Productivity Analysis**, v. 31, n. 2, p. 101-112, 2009.

SIMAR, Leopold; WILSON, Paul W. Estimation and inference in two-stage, semi-parametric models of production processes. **Journal of econometrics**, v. 136, n. 1, p. 31-64, 2007.

SPEARMAN, Charles. The proof and measurement of association between two things. **The American journal of psychology**, v. 15, n. 1, p. 72-101, 1904..

STEVENSON, Rodney E. Likelihood functions for generalized stochastic frontier estimation. **Journal of econometrics**, v. 13, n. 1, p. 57-66, 1980.

TOBIN, James. Estimation of relationships for limited dependent variables. **Econometrica: journal of the Econometric Society**, p. 24-36, 1958.

YU, William; JAMASB, Tooraj; POLLITT, Michael. Does weather explain cost and quality performance? An analysis of UK electricity distribution companies. **Energy Policy**, v. 37, n. 11, p. 4177-4188, 2009.

Appendix

The JAGS script implementing the proposed Bayesian model is shown below.

```
model
{
for ( k in 1:N ) {
      dummy[k] <- 0
      dummy[k] ~ dloglik(logLike[k])
      mu[k]    <- beta*data[k]
      logLike[k] <- -log(1.0000001 - phi(-(theta[k])/sig.u))+log(phi(((
                  sig.u/sig.v)/sqrt((sig.u*sig.u)+(sig.v*sig.v)))*(y[k]-
                  mu[k])+(theta[k]/((sig.u/sig.v) *sqrt((sig.u*sig.u) +
                  (sig.v*sig.v)))))) - 0.5* log(2*3.1416)-0.5*pow((theta[k]
                  -(y[k]-mu[k]))/sqrt((sig.u*sig.u)+(sig.v*sig.v)),2) -
                  log(sqrt((sig.u*sig.u) + (sig.v*sig.v)))
}

    ## Priors
theta[1:N] ~ car.normal(adj[], weights[], num[], tau)
beta  ~dnorm(0.0, 1.0E-06)
tau.u ~ dgamma(0.001, 0.001)
tau.v ~ dgamma(0.001, 0.001)
tau  ~dgamma(0.5, 0.0005)
lambda<- sig.u/sig.v
sig.u<- sqrt(1/tau.u)
sig.v<- sqrt(1/tau.v)
    sig2   <- (sig.u*sig.u) + (sig.v*sig.v)
    sig    <- sqrt(sig2)
sigmatheta<- sqrt(1/tau)
}
```

ANEXO A

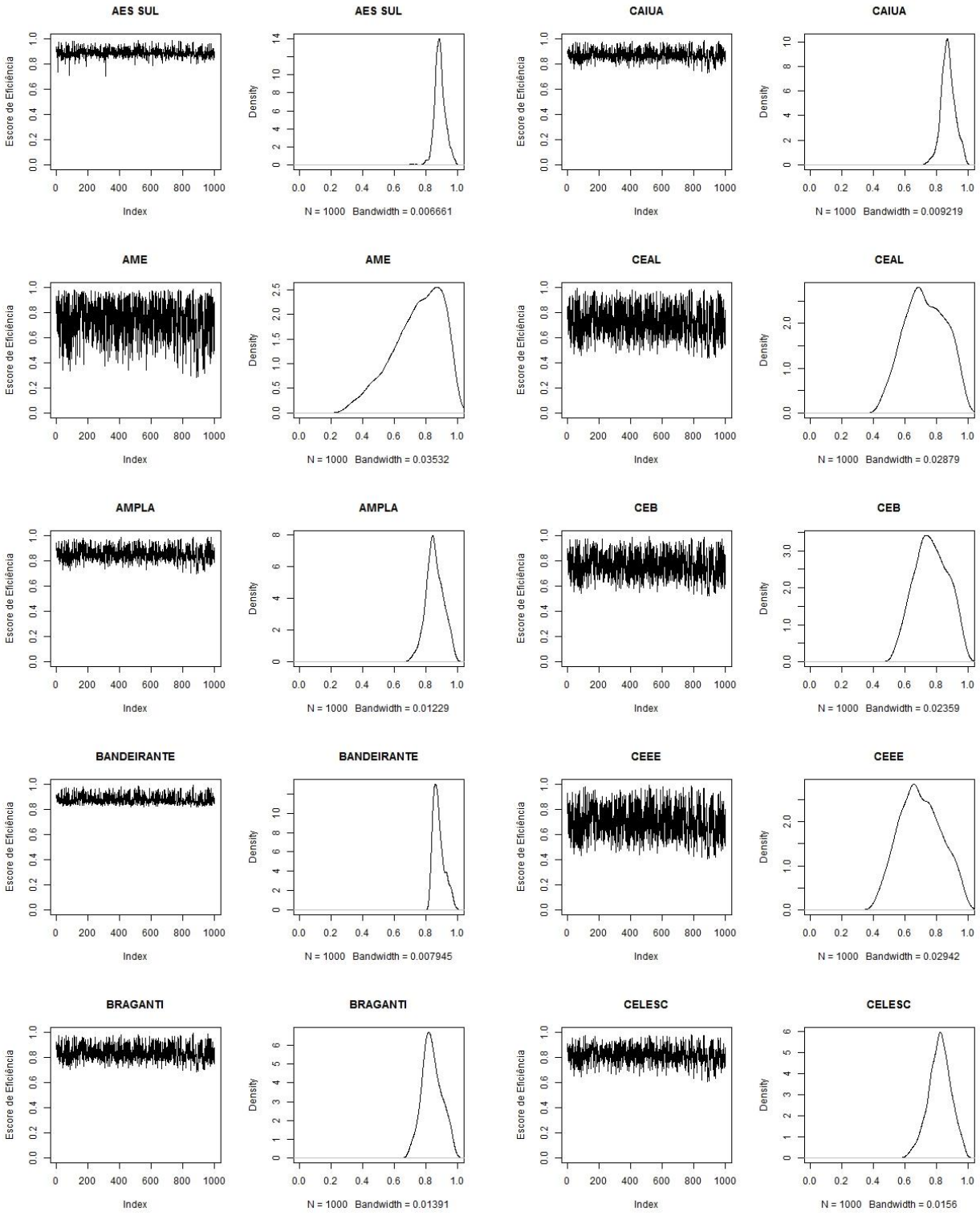
ANÁLISE DE CONVERGÊNCIA

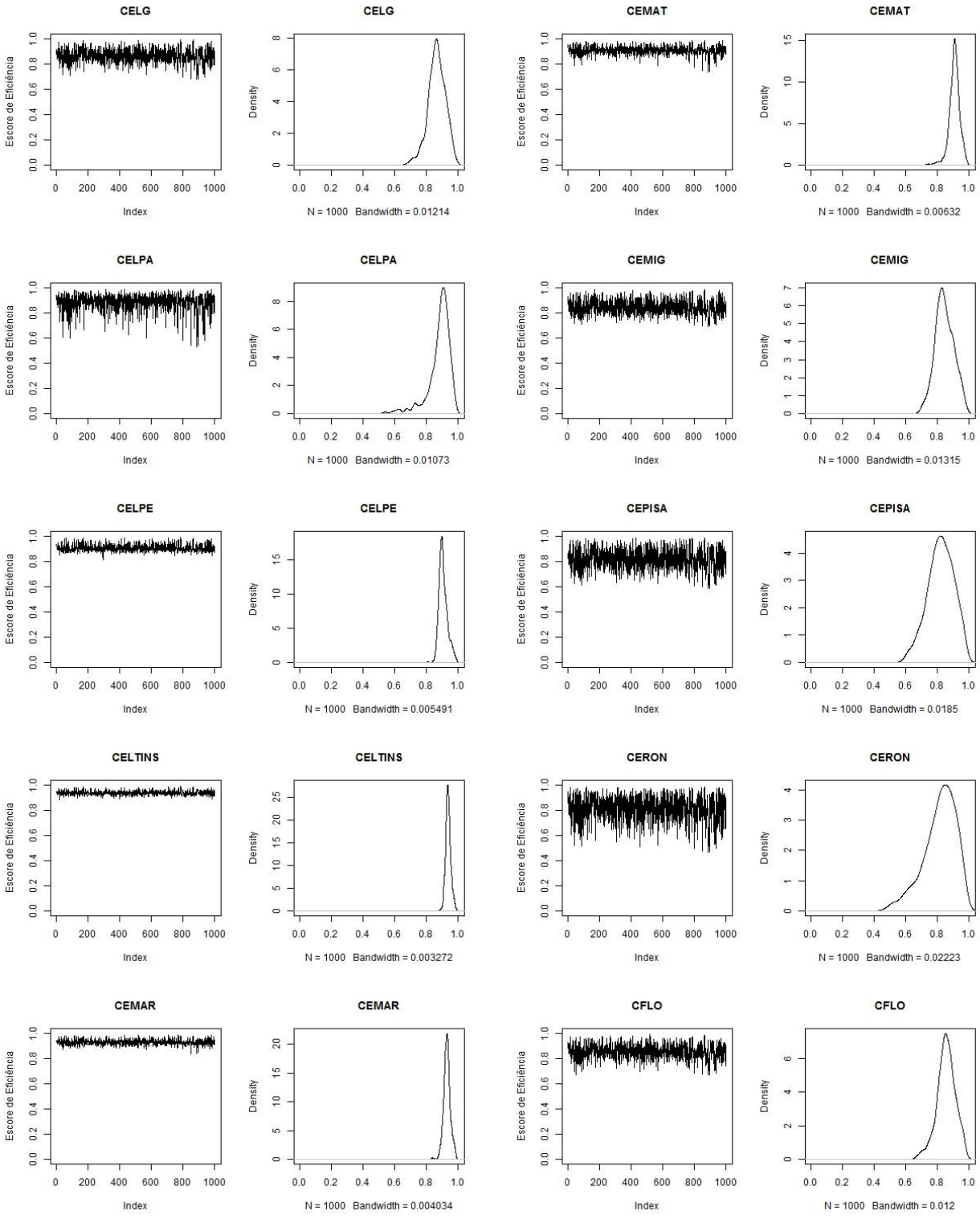
Na tabela a seguir são apresentadas as estatísticas Z do teste de Geweke (1992) para as cadeias dos escores corrigidos pela variável ***e.variable*** para as empresas Brasileiras de distribuição de energia elétrica. Pode-se verificar que os valores das estatísticas estão contidas no intervalo -1,96 e 1,96, indicando que as cadeias para os escores corrigidos convergiram ao nível de 95% de confiança.

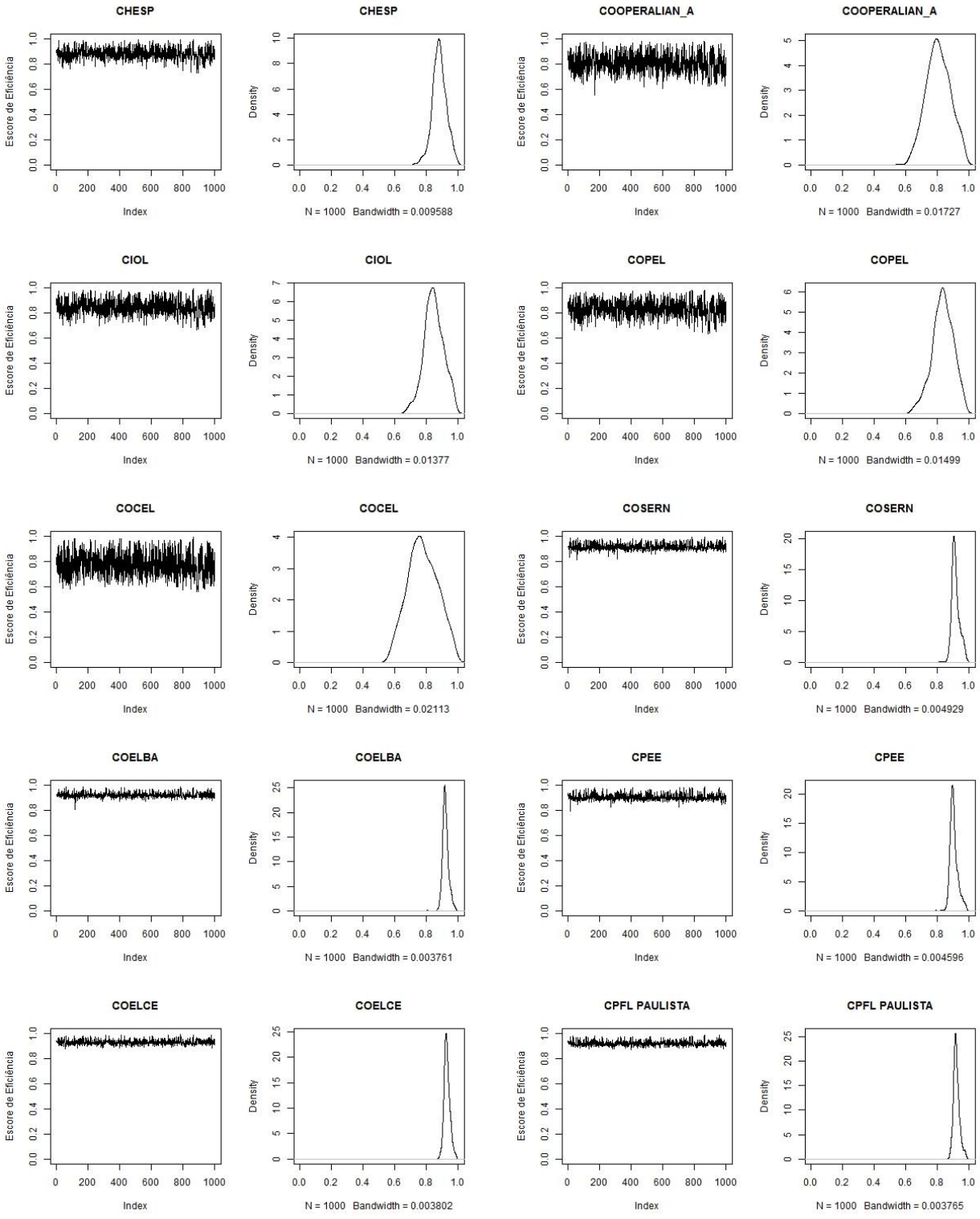
TABELA A. Estatística Z do teste de Geweke para as cadeias dos escores corrigidos pela variável *e.variable* para as empresas de distribuição de energia elétrica.

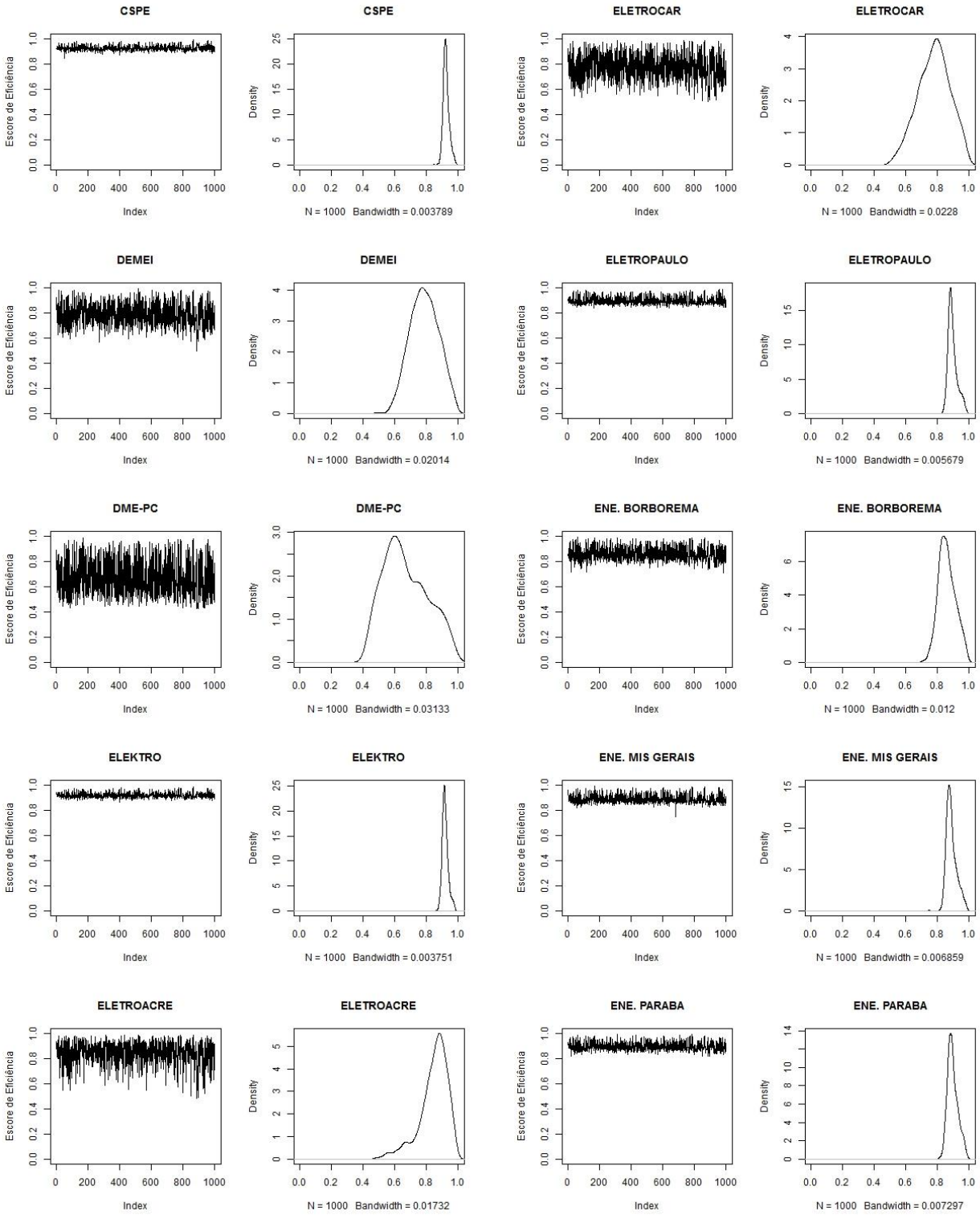
Empresa	Teste de Geweke	Empresa	Teste de Geweke	Empresa	Teste de Geweke
AES SUL	-1,271	COELCE	-0,375	ELETROACRE	0,196
AMPLA	0,263	COOPERALIANÇA	0,061	CEAL	0,374
BANDEIRANTE	0,648	COPEL	0,045	CEPISA	0,086
CAIUA	-0,119	COSERN	-0,609	CERON	0,100
CEB	0,136	JAGUARI	-0,679	ELETROCAR	-0,560
CEEE	-0,471	CPEE	-0,550	ELETROPAULO	0,032
CELESC	-0,320	MOCOCA	-0,926	SANTA MARIA	0,447
CELG	0,294	CPFL PAULISTA	-0,496	ENE. MIS GERAIS	-0,345
CELPA	0,016	PIRATININGA	-0,317	ENERSUL	0,185
CELPE	-0,459	SANTA CRUZ	-0,208	ENE. PARABA	-0,199
CELTINS	0,141	CSPE	-0,602	ESCELSA	0,244
CEMAR	0,173	DEMEI	-0,101	ENE. SERGIPE	0,308
CEMAT	0,164	DME-PC	0,021	HIDROPAN	-0,139
CEMIG	0,133	ENE. BORBOREMA	-0,970	IGUAÇU	-0,714
CFLO	-0,169	VALE PARAPANEMA	-0,136	LIGHT	0,117
CHESP	-0,421	BRAGANTINA	0,032	MUXFELDT	-1,373
CIOL	-0,042	URUSSANGA	-0,635	RGE	-0,835
COCEL	0,013	ELEKTRO	-0,239	SULGIPE	0,383
COELBA	0,445	AME	0,150	NOVA PALMA	-0,445

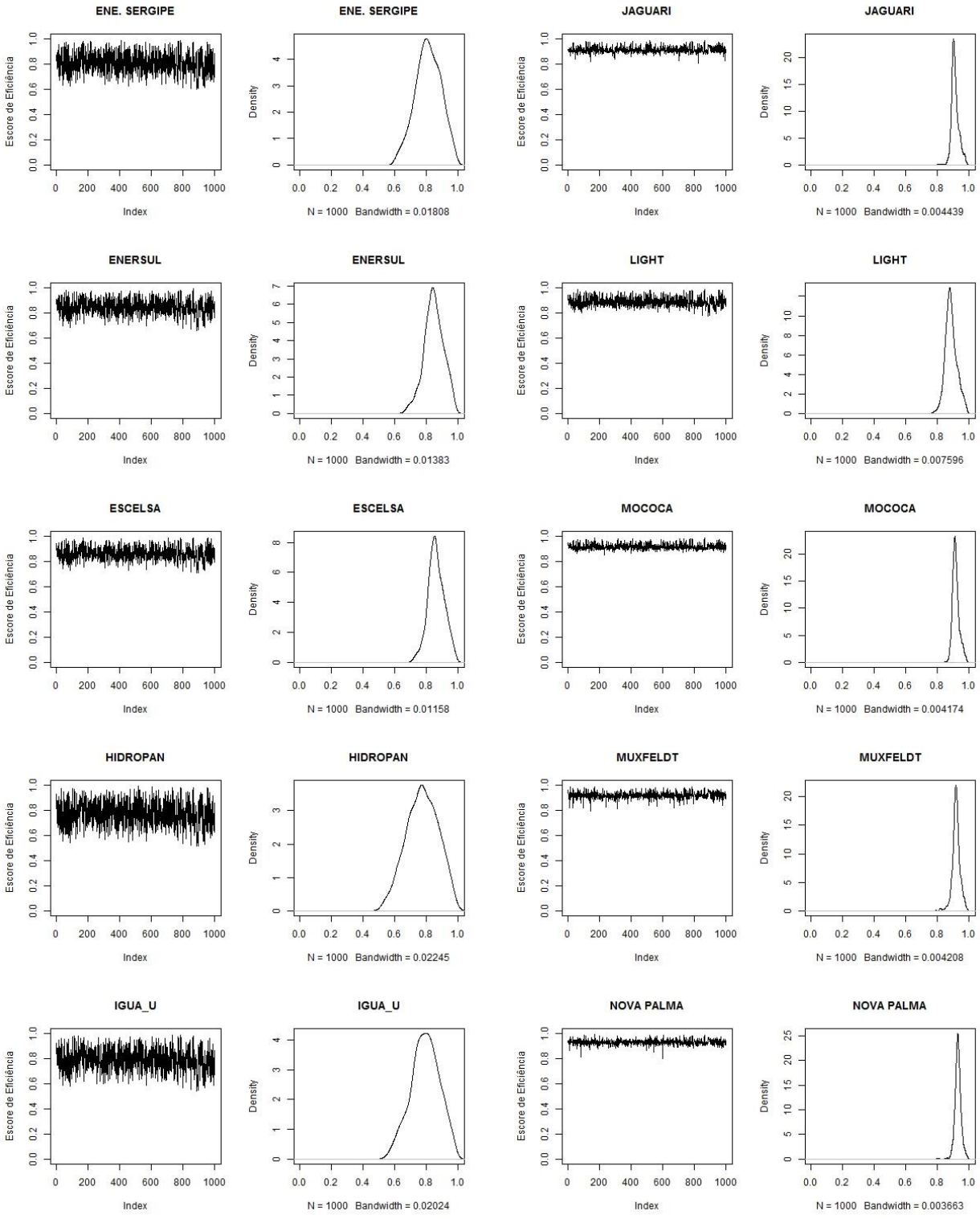
A análise de convergência também pode ser verificada de forma visual a partir dos gráficos das cadeias e das densidades estimadas para os escores de eficiência, para cada uma das empresas distribuidoras de energia elétrica. Nota-se, para todas as empresas, que os valores de seus escores de eficiência tendem a variar em torno de um valor constante, mas com variabilidades diferenciadas. Os gráficos corroboram com a hipótese de convergência das cadeias.


AES SUL
Escore de Eficiência
Index
AES SUL
Density
N = 1000 Bandwidth = 0.006661
CAIUA
Escore de Eficiência
Index
CAIUA
Density
N = 1000 Bandwidth = 0.009219
AME
Escore de Eficiência
Index
AME
Density
N = 1000 Bandwidth = 0.03532
CEAL
Escore de Eficiência
Index
CEAL
Density
N = 1000 Bandwidth = 0.02879
AMPLA
Escore de Eficiência
Index
AMPLA
Density
N = 1000 Bandwidth = 0.01229
CEB
Escore de Eficiência
Index
CEB
Density
N = 1000 Bandwidth = 0.02359
BANDEIRANTE
Escore de Eficiência
Index
BANDEIRANTE
Density
N = 1000 Bandwidth = 0.007945
CEEE
Escore de Eficiência
Index
CEEE
Density
N = 1000 Bandwidth = 0.02942
BRAGANTI
Escore de Eficiência
Index
BRAGANTI
Density
N = 1000 Bandwidth = 0.01391
CELESC
Escore de Eficiência
Index
CELESC
Density
N = 1000 Bandwidth = 0.0156


CELG
Escore de Eficiência
Index
CELG
Density
N = 1000 Bandwidth = 0.01214
CEMAT
Escore de Eficiência
Index
CEMAT
Density
N = 1000 Bandwidth = 0.00632
CELPA
Escore de Eficiência
Index
CELPA
Density
N = 1000 Bandwidth = 0.01073
CEMIG
Escore de Eficiência
Index
CEMIG
Density
N = 1000 Bandwidth = 0.01315
CELPE
Escore de Eficiência
Index
CELPE
Density
N = 1000 Bandwidth = 0.005491
CEPISA
Escore de Eficiência
Index
CEPISA
Density
N = 1000 Bandwidth = 0.0185
CELTINS
Escore de Eficiência
Index
CELTINS
Density
N = 1000 Bandwidth = 0.003272
CERON
Escore de Eficiência
Index
CERON
Density
N = 1000 Bandwidth = 0.02223
CEMAR
Escore de Eficiência
Index
CEMAR
Density
N = 1000 Bandwidth = 0.004034
CFLO
Escore de Eficiência
Index
CFLO
Density
N = 1000 Bandwidth = 0.012


CHESP
Escore de Eficiência
Index
CHESP
Density
N = 1000 Bandwidth = 0.009588
COOPERALIAN_A
Escore de Eficiência
Index
COOPERALIAN_A
Density
N = 1000 Bandwidth = 0.01727
CIOL
Escore de Eficiência
Index
CIOL
Density
N = 1000 Bandwidth = 0.01377
COPEL
Escore de Eficiência
Index
COPEL
Density
N = 1000 Bandwidth = 0.01499
COCEL
Escore de Eficiência
Index
COCEL
Density
N = 1000 Bandwidth = 0.02113
COSERN
Escore de Eficiência
Index
COSERN
Density
N = 1000 Bandwidth = 0.004929
COELBA
Escore de Eficiência
Index
COELBA
Density
N = 1000 Bandwidth = 0.003761
CPEE
Escore de Eficiência
Index
CPEE
Density
N = 1000 Bandwidth = 0.004596
COELCE
Escore de Eficiência
Index
COELCE
Density
N = 1000 Bandwidth = 0.003802
CPFL PAULISTA
Escore de Eficiência
Index
CPFL PAULISTA
Density
N = 1000 Bandwidth = 0.003765


CSPE
Escore de Eficiência
Index
CSPE
Density
N = 1000 Bandwidth = 0.003789
ELETROCAR
Escore de Eficiência
Index
ELETROCAR
Density
N = 1000 Bandwidth = 0.0228
DEMEI
Escore de Eficiência
Index
DEMEI
Density
N = 1000 Bandwidth = 0.02014
ELETROPAULO
Escore de Eficiência
Index
ELETROPAULO
Density
N = 1000 Bandwidth = 0.005679
DME-PC
Escore de Eficiência
Index
DME-PC
Density
N = 1000 Bandwidth = 0.03133
ENE. BORBOREMA
Escore de Eficiência
Index
ENE. BORBOREMA
Density
N = 1000 Bandwidth = 0.012
ELEKTRO
Escore de Eficiência
Index
ELEKTRO
Density
N = 1000 Bandwidth = 0.003751
ENE. MIS GERAIS
Escore de Eficiência
Index
ENE. MIS GERAIS
Density
N = 1000 Bandwidth = 0.006859
ELETROACRE
Escore de Eficiência
Index
ELETROACRE
Density
N = 1000 Bandwidth = 0.01732
ENE. PARABA
Escore de Eficiência
Index
ENE. PARABA
Density
N = 1000 Bandwidth = 0.007297


ENE. SERGIPE
Escore de Eficiência
Index
ENE. SERGIPE
Density
N = 1000 Bandwidth = 0.01808
JAGUARI
Escore de Eficiência
Index
JAGUARI
Density
N = 1000 Bandwidth = 0.004439
ENERSUL
Escore de Eficiência
Index
ENERSUL
Density
N = 1000 Bandwidth = 0.01383
LIGHT
Escore de Eficiência
Index
LIGHT
Density
N = 1000 Bandwidth = 0.007596
ESCELSA
Escore de Eficiência
Index
ESCELSA
Density
N = 1000 Bandwidth = 0.01158
MOCOCA
Escore de Eficiência
Index
MOCOCA
Density
N = 1000 Bandwidth = 0.004174
HIDROPAN
Escore de Eficiência
Index
HIDROPAN
Density
N = 1000 Bandwidth = 0.02245
MUXFELDT
Escore de Eficiência
Index
MUXFELDT
Density
N = 1000 Bandwidth = 0.004208
IGUA_U
Escore de Eficiência
Index
IGUA_U
Density
N = 1000 Bandwidth = 0.02024
NOVA PALMA
Escore de Eficiência
Index
NOVA PALMA
Density
N = 1000 Bandwidth = 0.003663

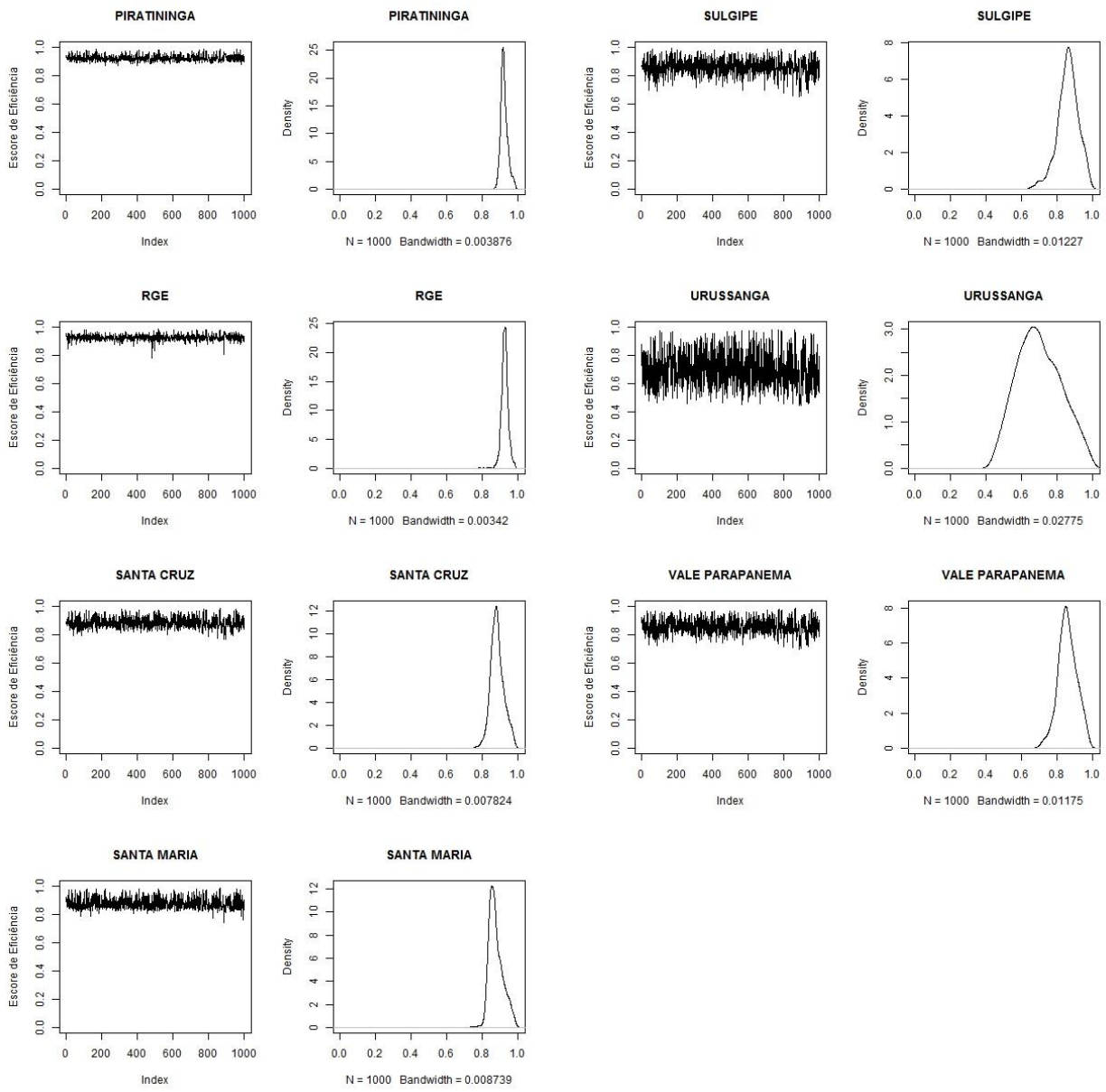


Figura A. Amostras a posteriori dos escores corrigidos pela variável ambiental *e.variable*, para as empresas Brasileiras distribuidoras de energia elétrica.