



Avaliação das eficiências operacionais das gerências CEMIG-D: Uma abordagem utilizando *Data Envelopment Analysis* e regressão truncada em segundo estágio

Lucas Cardoso Queiroz, Marcelo Azevedo Costa

Departamento de Engenharia de Produção
Universidade Federal de Minas Gerais
lqueiroz@ufmg.br, macosta@ufmg.br

Ana Lúcia Miranda Lopes

Faculdade de Ciências Econômicas
Departamento de Ciências Administrativas
Universidade Federal de Minas Gerais
analopes.ufmg@gmail.com

RESUMO

Neste trabalho, as eficiências das 16 gerências regionais da empresa da Companhia de Energia de Minas Gerais, CEMIG-D, foram estimadas utilizando a metodologia de análise envoltória de dados (DEA). Utilizando dados no período 2010 a 2012, foi aplicado o modelo DEA-NDRS em um primeiro estágio e a abordagem de Simar e Wilson (2007) de análise em segundo estágio. O segundo estágio compreende um modelo de regressão truncada e um procedimento de duplo-bootstrap para correção de viés e teste da significância estatística das variáveis ambientais. Este trabalho também propõe uma forma de corrigir os escores de eficiência, homogeneizando o efeito do ambiente nas eficiências. Por se tratar de uma base de dados relativamente pequena, são discutidas maneiras eficientes de agregação das informações como, por exemplo, o uso dos percentis para agregação dos fatores ambientais.

PALAVRAS CHAVE. Análise Envoltória de Dados (DEA), Segundo estágio, regressão truncada, bootstrap.

Tópicos: DEA - Análise Envoltória de Dados.

ABSTRACT

In this work, the efficiencies of the 16 management units of the Minas Gerais Energy Company, CEMIG-D, were estimated using the Data Envelopment Analysis (DEA) methodology. Using data from 2010 to 2012, the DEA-NDRS model was applied in a first stage, followed by Simar and Wilson's (2007) second stage analysis. The second stage comprises a truncated regression model and a double-bootstrap procedure for bias correction and statistical testing of environmental variables. This work also proposes a new approach to correct the efficiency scores, homogenizing the effect of the environment on the first stage efficiencies. Because it is a relatively small data set, efficient approaches for aggregating information such as the use of percentiles for aggregation of environmental factors, are discussed.

KEYWORDS. Data Envelopment Analysis, Second stage, truncated regression, bootstrap.

Paper topics: DEA – Data Envelopment Analysis.



1. Introdução

O setor energético, devido aos altos custos de implantação e manutenção de usinas geradoras e redes de distribuição, é considerado um monopólio natural, uma vez que a existência de concorrência poderia inviabilizar a exploração da atividade. Como consequência de um mercado monopolizado, estas empresas tendem à ineficiência, fazendo com que a regulação se torne necessária para garantir a qualidade dos serviços prestados. No Brasil, a agência reguladora – ANEEL (Agência Nacional de Energia Elétrica), adota o modelo de regulação por incentivos em sua avaliação das concessionárias prestadoras de serviços. Esta é uma forma específica de regulação voltada para a remuneração com vistas à eficiência da redução dos custos e das tarifas.

Segundo consta nos Procedimentos de Regulação Tarifária (Proret) - Submódulo 2.2, publicada em 28 de abril de 2015, a Agência Nacional de Energia Elétrica (ANEEL) estimou um índice de eficiência de 68.64% (LI=64.62%, LS=73.17%) para o custo operacional da CEMIG-D. Como consequência, a CEMIG-D deverá reduzir os seus custos operacionais em 31.36% nos próximos anos. O modelo utilizado para o cálculo dos custos eficientes foi *Data Envelopment Analysis* (DEA), conforme apresentado na Nota Técnica 66/2015. Este modelo estima, utilizando modelos de programação linear, uma fronteira de eficiência que é utilizada para estimar as eficiências relativas das unidades avaliadas (Decision Making Units - DMU), neste caso um conjunto de 61 empresas de distribuição de energia elétrica.

Existem diferentes metodologias de benchmarking que fazem uso de análise de fronteiras de produção para estimação de índices de eficiência relativa. Esses métodos podem ser divididos em paramétricos e não-paramétricos. Os métodos paramétricos requerem que a forma funcional da função de produção seja conhecida a priori, além de outros elementos estocásticos. Os métodos não-paramétricos, como é o caso dos modelos DEA, são mais flexíveis e não requerem nenhum parâmetro ou função de fronteira de produção a priori. Além disso, nos modelos estocásticos são consideradas duas variáveis aleatórias que representam ruído e ineficiência técnica. Nos modelos determinísticos não há nenhum tipo de tratamento estocástico e toda pequena variação com relação à fronteira de eficiência é considerada como ineficiência técnica (Bogetoft e Otto, 2010).

Os principais métodos para o cálculo de índices de eficiência relativa são: Data Envelopment Analysis, proposto por Charnes, Cooper e Rhodes (1978); Stochastic Frontier Analysis (SFA), proposto por Aigner, Lovell e Schmidt (1977) e Meeusen e Van Den Broeck (1977) e Corrected Ordinary Least Squares (COLS), método apresentado por Richmond (1974). Na categoria estocástico e não-paramétrico, também existe o método Stochastic Data Envelopment Analysis (SDEA), desenvolvido por Land, Lovell e Thore (1993) e Olesen e Petersen (1995). Recentemente, um modelo semi-paramétrico, combinando características de DEA e SFA, foi desenvolvido por Kuosmanen e Kortelainen (2012) e chamado de Stochastic Non-smooth Envelopment of Data (StoNED).

Uma forte suposição para o uso dos modelos mencionados é a suposição de que as empresas em estudo são comparáveis. Utilizam os mesmos insumos para produzirem os mesmos produtos e estão em contextos ou ambientes semelhantes. No caso de empresas distribuidoras de energia elétrica, o contexto está associado ao ambiente na qual estão inseridas. Por este motivo, é natural imaginar que características ambientais inerentes ao local de atuação de uma empresa de distribuição de energia deveriam ser consideradas no modelo de cálculo de eficiência, uma vez que suas influências, embora estejam fora do escopo produtivo, produzem impacto no processo produtivo.

2. Materiais e Métodos

2.1 Data Envelopment Analysis

O modelo de análise de eficiência por envelopamento de dados (Data Envelopment Analysis - DEA) tem sua origem no trabalho de Charnes, Cooper e Rhodes (1978). Em seu



trabalho, Charnes, Cooper e Rhodes (1978) apresentam um modelo não paramétrico no qual, unidades tomadoras de decisões (Decision Making Units - DMUs) que produzam o(s) mesmo(s) produto(s) ou serviço(s) (outputs), utilizando o(s) mesmo(s) insumo(s) (inputs), são comparadas entre si de acordo com o conceito de índice de eficiência de Farrel (Forsund, 2002). Este índice consiste na razão entre a soma ponderada dos outputs Y e a soma ponderada dos inputs X de cada DMU. Este modelo também é conhecido como DEA-CRS (constant returns to scale), pois os autores assumem uma economia de retorno constante de escala em seu trabalho, o que pode não ser uma suposição satisfatória para algumas economias com retornos crescente/decrescente de escala.

Para atender a necessidade de modelar outros tipos de economia de escalas, Banker, Charnes e Cooper (1984) publicaram um novo modelo conhecido como DEA-VRS (variable returns to scale). Em ambos os casos, a modelagem deve ser feita a partir de uma orientação a insumos ou a produtos. A eficiência calculada no modelo orientado a insumos assumirá que a quantidade de produtos produzidos não pode ser aumentada e, para a DMU se tornar mais eficiente, deverá reduzir a quantidade de insumos utilizados na produção desses produtos. Analogamente, no modelo orientado a produtos, os valores de insumos são mantidos os mesmos e aquelas DMUs consideradas ineficientes deverão aumentar a sua produção para se tornarem mais eficientes.

2.2 O Modelo DEA da ANEEL

O mecanismo utilizado pela ANEEL para o cálculo do custo operacional eficiente das empresas brasileiras de distribuição de energia elétrica está descrito nas NTs 407/2014-SRE/ANEEL (ANEEL, 2014) e 66/2015-SMR/SGT/ANEEL (ANEEL, 2015). O insumo é o custo operacional e os produtos são: extensão da rede da concessionária em km; mercado ponderado, que mostra a média de MWh fornecidos nas estruturas tarifárias de baixa tensão, média tensão e alta tensão por concessionária, ponderada pelas extensões de cada estrutura; número total de consumidores atendidos pela concessionária; perdas não técnicas (PNT), definida como a quantidade de energia distribuída sem autorização da companhia (ilegalmente desviadas das linhas de distribuição), em MWh; consumidor hora interrompido (CHI), variável criada para controlar a qualidade do serviço, mede a duração das interrupções na distribuição em horas.

O direcionamento escolhido da ferramenta DEA é estimar o custo operacional eficiente, ou seja, o modelo é direcionado a insumos. A ANEEL caracterizou a tecnologia do setor de distribuição de energia elétrica como retornos de escala não decrescentes (NDRS) e a sua formulação no modelo de multiplicadores é apresentada a seguir.

$$\begin{aligned} \text{Max } h_0 &= \sum_{j=1}^s v_j y_{j0} + \varphi \\ \text{sujeito a:} \quad & \sum_{i=1}^r u_i x_{ik} = 1 \\ & \sum_{j=1}^s v_j y_{jk} + \varphi - \sum_{i=1}^r u_i x_{ik} \leq 0, k = 1, 2, \dots, n \\ & v_j \text{ e } u_i \geq 0 \forall j, i \\ & \varphi \geq 0 \end{aligned} \tag{1}$$

onde h_0 é a eficiência da DMU de referência, DMU₀. u_i e v_j são pesos atribuídos aos insumos e produtos, respectivamente. x_{ik} e y_{jk} são os i insumos e j produtos da k -ésima DMU. x_{i0} e y_{j0} são os i insumos e j produtos da DMU de referência, DMU₀ e φ é o fator de escala. O modelo utilizado pela ANEEL também apresenta restrições aos pesos (u_i e v_j). Uma descrição detalhada do modelo DEA-ANEEL pode ser encontrada em da Silva (2015).



2.3 Ajuste em segundo estágio segundo Simar e Wilson (2007)

Segundo Simar e Wilson (2007), sendo $\mathbf{x} \in \mathbb{R}_+^p$ um vetor ($1 \times p$) de insumos, $\mathbf{y} \in \mathbb{R}_+^q$ um vetor ($1 \times q$) de produtos e $\mathbf{z} \in \mathbb{R}^r$ um vetor ($1 \times r$) de variáveis ambientais, o conjunto de observações disponíveis é denominado $\mathcal{L}_n = \{(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i)\}_{i=1}^n$. Assume-se que as observações $(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i)$ são realizações de variáveis aleatórias independentes e identicamente distribuídas com função densidade de probabilidade $f(\mathbf{x}, \mathbf{y}, \mathbf{z})$. Porém, ocorre que em diversas situações, \mathbf{z} não é independente em relação a (\mathbf{x}, \mathbf{y}) , ou seja $f(\mathbf{x}, \mathbf{y}|\mathbf{z}) \neq f(\mathbf{x}, \mathbf{y})$, sendo essa a motivação da análise em segundo estágio.

A fronteira de produção é definida como $\mathcal{P} = \{(\mathbf{x}, \mathbf{y}) | \mathbf{x} \text{ produz } \mathbf{y}\}$ e, pela definição adotada de Shephard (1970) e Färe (1988), \mathcal{P} é um conjunto fechado e convexo. O envoltório convexo dos pares $(\mathbf{x}_i, \mathbf{y}_i)$ do conjunto \mathcal{L}_n pode ser utilizado para estimar a fronteira de produção \mathcal{P} do seguinte modo:

$$\hat{\mathcal{P}} = \{(\mathbf{x}, \mathbf{y}) | \mathbf{y} \leq \mathbf{Y}\boldsymbol{\lambda}, \mathbf{x} \geq \mathbf{X}\boldsymbol{\lambda}, \boldsymbol{\lambda} \in \mathbb{R}_+^n\} \quad (2)$$

onde $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_n]$, $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n]$ e $\boldsymbol{\lambda}$ é um vetor ($n \times 1$) de variáveis de intensidades. A influência de uma variável ambiental \mathbf{Z} é tipicamente especificada em diversos estudos como sendo uma função $\psi(\mathbf{z}_i, \boldsymbol{\beta})$. Simar e Wilson (2007) assume uma estrutura linear na forma $\psi(\mathbf{z}_i, \boldsymbol{\beta}) = \mathbf{z}_i\boldsymbol{\beta}$. Dessa forma, o escore de eficiência pode ser associado à variável ambiental assumindo o modelo descrito na equação (3).

$$\delta_i = \mathbf{z}_i\boldsymbol{\beta} + \varepsilon_i \geq 1. \quad (3)$$

A proposta do ajuste em segundo estágio consiste em utilizar os pares $(\mathbf{x}_i, \mathbf{y}_i)$ em \mathcal{L}_n para estimar δ_i para todos $i = 1, \dots, n$, gerando um conjunto de estimadores $\{\hat{\delta}_i\}_{i=1}^n$ definido por

$$\hat{\delta}_0 = \delta(\mathbf{x}_0, \mathbf{y}_0 | \hat{\mathcal{P}}) = \max\{\phi | \phi\mathbf{y}_0 \leq \mathbf{Y}\boldsymbol{\lambda}, \mathbf{x}_0 \geq \mathbf{X}\boldsymbol{\lambda}, \boldsymbol{\lambda} \in \mathbb{R}_+^n\}. \quad (4)$$

Então, substituindo o lado esquerdo da equação (3) pelo estimador $\hat{\delta}_i$ obtém-se

$$\hat{\delta}_i = \mathbf{z}_i\boldsymbol{\beta} + \xi_i \geq 1 \quad (5)$$

onde $\varepsilon_i \sim N(0, \sigma_\varepsilon)$.

Simar e Wilson (2007) apontam que existem alguns problemas na proposta apresentada acima. Primeiramente, alegam que a variável dependente em (3) não é conhecida e por isso deve ser substituída por um estimador utilizando o modelo DEA. Os $\hat{\delta}_i$'s que são usados na estimação de (5) são serialmente correlacionados. Outro problema apresentado por Simar e Wilson refere-se ao vício dos estimadores $\hat{\delta}_i$'s que não foram levados em consideração em estudos anteriores. Para solucionar o problema do vício dos estimadores, Simar e Wilson (2007) apresentam dois algoritmos utilizando técnicas de reamostragem bootstrap. O algoritmo #2, mais completo, tem o seu passo-a-passo apresentado a seguir:

1. Utilizando os valores originais de \mathcal{L}_n , computar $\hat{\delta}_i = \delta(\mathbf{x}_i, \mathbf{y}_i | \hat{\mathcal{P}}) \quad \forall i = 1, \dots, n$, utilizando a equação (4).
2. Usar o método de máxima verossimilhança para obter $\hat{\boldsymbol{\beta}}$ bem como o estimador $\hat{\sigma}_\varepsilon$ da regressão truncada de $\hat{\delta}_i$ em \mathbf{z}_i conforme equação (5), usando as $m < n$ observações nas quais $\hat{\delta}_i > 1$.
3. Repetir os passos 3.1 a 3.4 L_1 vezes, a fim de obter n conjuntos de estimadores bootstrap $\mathcal{B}_i = \{\hat{\delta}_{ib}^*\}_{b=1}^{L_1}$:
 - 3.1. Para cada $i = 1, \dots, n$ simular ε_i a partir da distribuição $N(0, \hat{\sigma}_\varepsilon)$ truncada à esquerda em $(1 - \mathbf{z}_i\hat{\boldsymbol{\beta}})$.
 - 3.2. Para cada $i = 1, \dots, n$ computar $\delta_i^* = \mathbf{z}_i\hat{\boldsymbol{\beta}} + \varepsilon_i$.
 - 3.3. Definir $\mathbf{x}_i^* = \mathbf{x}_i$, $\mathbf{y}_i^* = \mathbf{y}_i \hat{\delta}_i / \delta_i^*$ para todos $i = 1, \dots, n$



- 3.4. Computar $\hat{\delta}_i^* = \delta(\mathbf{x}_i, \mathbf{y}_i | \hat{\mathcal{P}}^*) \forall i = 1, \dots, n$, onde $\hat{\mathcal{P}}^*$ é obtido substituindo \mathbf{X}, \mathbf{Y} na equação (2) por $\mathbf{Y}^* = [\mathbf{y}_1^* \dots \mathbf{y}_n^*]$ e $\mathbf{X}^* = [\mathbf{x}_1^* \dots \mathbf{x}_n^*]$.
4. Para cada $i = 1, \dots, n$, computar o estimador com viés corrigido $\hat{\delta}_i = \hat{\delta}_i - \widehat{BIAS}(\hat{\delta}_i)$.
 5. Usar o método de máxima verossimilhança para estimar a regressão truncada de $\hat{\delta}_i$ em \mathbf{z}_i , obtendo novos estimadores $\hat{\beta}, \hat{\sigma}$.
 6. Repetir os próximos três passos (6.1 a 6.3) L_2 vezes a fim de obter um conjunto de estimadores bootstrap $\mathcal{C} = \left\{ (\hat{\beta}^*, \hat{\sigma}_\varepsilon^*)_b \right\}_{b=1}^{L_2}$:
 - 6.1. Para cada $i = 1, \dots, n$ simular ε_i a partir da distribuição $N(0, \hat{\sigma})$ truncada a esquerda em $(1 - \mathbf{z}_i \hat{\beta})$.
 - 6.2. Novamente, para cada $i = 1, \dots, n$ computar $\delta_i^{**} = \mathbf{z}_i \hat{\beta} + \varepsilon_i$.
 - 6.3. Usar o método de máxima verossimilhança para estimar a regressão truncada de δ_i^{**} em \mathbf{z}_i , obtendo estimadores $(\hat{\beta}^*, \hat{\sigma}_\varepsilon^*)$.
 7. Usar os valores bootstrap em \mathcal{C} e os estimadores originais $\hat{\beta}, \hat{\sigma}$ para obter intervalos de confiança estimados para cada elemento de β e para σ_ε .

Para o estudo proposto neste artigo, estamos interessados em fazer um ajuste em segundo estágio em um modelo DEA orientado a insumos. Algumas alterações são necessárias no procedimento descrito anteriormente. No modelo DEA orientado a insumos, os estimadores dos escores de eficiência são obtidos na forma $0 \leq \hat{\theta}_i \leq 1 \forall i = 1, \dots, n$, onde:

$$\hat{\theta}_0 = \theta(\mathbf{x}_0, \mathbf{y}_0 | \hat{\mathcal{P}}) = \min\{\phi \mid \phi \mathbf{x}_0 \geq \lambda \mathbf{X}, \mathbf{y}_0 \leq \lambda \mathbf{Y}, \lambda \in \mathbb{R}_+^n\}. \quad (6)$$

Para o ajuste em segundo estágio, o algoritmo passo-a-passo é descrito a seguir.

1. Utilizando os valores originais de \mathcal{L}_n , computar $\hat{\theta}_i = \theta(\mathbf{x}_i, \mathbf{y}_i | \hat{\mathcal{P}}) \forall i = 1, \dots, n$ utilizando a equação (6).
2. Computar $\hat{\delta}_i = \frac{1}{\hat{\theta}_i} \forall i = 1, \dots, n$.
3. Usar o método de máxima verossimilhança para obter $\hat{\beta}$ de β bem como os estimadores $\hat{\sigma}_\varepsilon$ de σ_ε da regressão truncada de $\hat{\delta}_i$ em \mathbf{z}_i em (5) usando as $m < n$ observações nas quais $\hat{\delta}_i > 1$.
4. Repetir os passos 4.1 a 4.4 L_1 vezes, a fim de obter n conjuntos de estimadores bootstrap $\mathcal{B}_i = \left\{ \hat{\delta}_{ib}^* \right\}_{b=1}^{L_1}$:
 - 4.1. Para cada $i = 1, \dots, n$ simular ε_i a partir da distribuição $N(0, \hat{\sigma}_\varepsilon)$ truncada a esquerda em $(1 - \mathbf{z}_i \hat{\beta})$.
 - 4.2. Novamente, para cada $i = 1, \dots, n$ computar $\delta_i^* = \mathbf{z}_i \hat{\beta} + \varepsilon_i$
 - 4.3. Definir $\mathbf{y}_i^* = \mathbf{y}_i, \mathbf{x}_i^* = \mathbf{x}_i \frac{\delta_i^*}{\hat{\delta}_i}$ para todos $i = 1, \dots, n$.
 - 4.4. Computar $\hat{\delta}_i^* = \frac{1}{\theta(\mathbf{x}_i, \mathbf{y}_i | \hat{\mathcal{P}}^*)} \forall i = 1, \dots, n$, onde $\hat{\mathcal{P}}^*$ é obtido substituindo \mathbf{X}, \mathbf{Y} em (2) por $\mathbf{Y}^* = [\mathbf{y}_1^* \dots \mathbf{y}_n^*], \mathbf{X}^* = [\mathbf{x}_1^* \dots \mathbf{x}_n^*]$.
5. Para cada $i = 1, \dots, n$, computar o estimador com viés corrigido $\hat{\delta}_i = \hat{\delta}_i - \widehat{BIAS}(\hat{\delta}_i)$.
6. Usar o método de máxima verossimilhança para estimar a regressão truncada de $\hat{\delta}_i$ em \mathbf{z}_i , obtendo estimadores $(\hat{\beta}, \hat{\sigma})$.
7. Repetir os próximos três passos (7.1 a 7.3) L_2 vezes a fim de obter um conjunto de estimadores bootstrap $\mathcal{C} = \left\{ (\hat{\beta}^*, \hat{\sigma}_\varepsilon^*)_b \right\}_{b=1}^{L_2}$:
 - 7.1. Para cada $i = 1, \dots, n$ simular ε_i a partir da distribuição $N(0, \hat{\sigma})$ truncada a esquerda em $(1 - \mathbf{z}_i \hat{\beta})$.



- 7.2. Novamente, para cada $i = 1, \dots, n$ computar $\delta_i^{**} = \mathbf{z}_i \widehat{\boldsymbol{\beta}} + \varepsilon_i$.
- 7.3. Usar o método de máxima verossimilhança para estimar a regressão truncada de δ_i^{**} em \mathbf{z}_i , obtendo estimadores $(\widehat{\boldsymbol{\beta}}^*, \widehat{\sigma}_\varepsilon^*)$.
8. Usar os valores obtidos por bootstrap em \mathcal{C} e os estimadores originais $\widehat{\boldsymbol{\beta}}, \widehat{\sigma}_\varepsilon$ para obter intervalos de confiança estimados para cada elemento de $\boldsymbol{\beta}$ e para σ_ε .

2.4 Proposta de correção dos escores de eficiência

O modelo de segundo estágio proposto por Simar e Wilson (2007) mostra que, como descrito em (3), $\delta_i = \psi(\mathbf{z}_i, \boldsymbol{\beta}) + \varepsilon_i \geq 1$, e que, portanto, $\varepsilon_i \sim N^T(0, \sigma^2)$. Com o interesse de realizar a correção dos escores, temos que retirar o efeito da variável ambiental Z , para obtermos um escore corrigido que se comporta segundo a Equação (7).

$$\delta_i^{cor} = \alpha + \varepsilon_i. \quad (7)$$

No entanto, observa-se que a componente ε_i está relacionada com a variável Z , uma vez que a mesma é obtida a partir de uma normal truncada em $1 - \psi(\mathbf{z}_i, \boldsymbol{\beta})$. Para contornar este problema, a solução proposta é, inicialmente, remover a influência de Z , padronizando os erros do modelo:

$$\varepsilon_i^P = \frac{\varepsilon_i - E(\varepsilon)}{\sqrt{VAR(\varepsilon)}} \quad (8)$$

Considerando que ε_i segue uma distribuição truncada em $1 - \psi(\mathbf{z}_i, \boldsymbol{\beta})$:

$$E(\varepsilon) = \frac{\phi\left(\frac{a}{\sigma_\varepsilon}\right)}{1 - \Phi\left(\frac{a}{\sigma_\varepsilon}\right)} \sigma_\varepsilon \quad (9)$$

$$VAR(\varepsilon) = \sigma_\varepsilon^2 \left[1 + \frac{\frac{a}{\sigma_\varepsilon} \phi\left(\frac{a}{\sigma_\varepsilon}\right)}{1 - \Phi\left(\frac{a}{\sigma_\varepsilon}\right)} - \left(\frac{\phi\left(\frac{a}{\sigma_\varepsilon}\right)}{1 - \Phi\left(\frac{a}{\sigma_\varepsilon}\right)} \right)^2 \right] \quad (10)$$

onde $a = 1 - \psi(\mathbf{z}_i, \boldsymbol{\beta})$ é o truncamento à esquerda da distribuição normal. Ao realizar a padronização, observa-se que ε_i^P é um vetor com $\mu = 0$ e $\sigma^2 = 1$ e por consequência, os escores corrigidos utilizando $\alpha + \varepsilon_i^P$ também teriam o mesmo desvio padrão, $\sigma^2 = 1$. No entanto, estamos interessados em manter algumas propriedades de δ_i (como σ_δ^2 , desvio padrão dos escores não-corrigidos) mesmo após a correção da influência da componente ambiental. Para este fim, definimos então:

$$\zeta_i = \frac{\varepsilon_i - E(\varepsilon)}{\sqrt{VAR(\varepsilon)}} \sigma_\delta \quad (11)$$

Finalmente, para que os valores dos escores corrigidos tenham a propriedade $\delta^{cor} \geq 1$, é necessário definir:

$$\alpha = \max(1 - \zeta_i) \quad (12)$$

Assim, substituindo as equações (11) e (12) na equação (7), obtemos os valores corrigidos, sem influência da componente ambiental:

$$\delta_i^{cor} = \max(1 - \zeta_i) + \zeta_i \quad (13)$$

O algoritmo #2 de Simar e Wilson (2007) propõe, sumariamente, estimar o efeito do viés na fronteira de produção por meio de bootstrap, obtendo assim estimadores $(\widehat{\delta}_i, \widehat{\boldsymbol{\beta}}, \widehat{\sigma}_\varepsilon)$. Podemos, portanto, corrigir a influência da componente ambiental nos escores com viés corrigidos.



Portanto, $\hat{\zeta}_i = \frac{\hat{\varepsilon}_i - E(\hat{\varepsilon})}{\sqrt{VAR(\hat{\varepsilon})}} \hat{\sigma}$, onde $\hat{\sigma}$ é o desvio padrão dos escores sem viés estimado pelo método de máxima verossimilhança. Os valores corrigidos serão, então, obtidos pela equação 14,

$$\hat{\delta}_i^{cor} = \max(1 - \hat{\zeta}_i) + \hat{\zeta}_i. \quad (14)$$

2.5 Base de dados

A base de dados é composta por uma variável de insumo PMSO (Custos de Pessoal, Manutenção, Serviço e Operação) e pelas variáveis de produtos: comprimento de rede, número de consumidores, potencia entregue (MWh) e número de postes para 16 gerências da CEMIG-D. Os dados representam valores mensais no período de Janeiro de 2010 a Dezembro de 2013. É possível a abertura nas classes rurais e urbanas para o comprimento de rede, número de consumidores, potência entregue e número de postes.

Inicialmente, os dados foram agregados para os respectivos anos e gerências. Para as variáveis PMSO e energia entregue, foi utilizada a agregação pela soma dos valores mensais. Para a variável rede foi utilizado o último valor mensal do ano, no mês de Dezembro. A mesma operação de agregação foi realizada para a variável comprimento de rede, número de postes e número de consumidores. Os escores de eficiência foram obtidos para cada um dos três anos que compõem a base de dados.

Além das variáveis de insumos e produtos, há ainda o registro de variáveis ambientais, a saber: índice de precipitação, velocidade de vento e taxa de incidência de raios. O índice de precipitação está disponível em valores médios mensais para as gerências da CEMIG-D no período de Janeiro de 2010 a Dezembro de 2013. O mesmo formato está disponível para os dados referentes à densidade de raios. Os dados de velocidades de ventos (rajadas de vento) estão disponíveis em séries de valores medidos a cada 8 horas em estações anemométricas localizadas nas áreas de abrangência das gerências. As séries de vento, precipitação e densidade de raios foram agregadas para valores anuais utilizando uma análise de correlação entre as eficiências estimadas pelo modelo DEA-CRS (Urbano-Rural) e os percentis das séries. Para isso, foi utilizado o coeficiente de correlação de Spearman. Esta análise foi realizada sob a suposição de que as eficiências (ou ineficiências) apresentam uma correlação mais forte com valores extremos das medidas ambientais do que com os valores médios ou medianos. Como resultado, para a série de velocidade de ventos os valores anuais representam o percentil 65% dos valores diários, e apresentam um coeficiente de correlação de Spearman de -0.4233 (-42.33%) com os escores de eficiência. Para a série de densidade de raios e precipitação, os valores das correlações com os escores ficaram em valores positivos e/ou próximos de zeros para todos os valores de percentis acima de 50%, de forma que a análise em segundo estágio para estas variáveis ambientais não apresentaram significância estatística.

As gerências da CEMIG-D são centros de serviços dotados de um gerente responsável pela coordenação dos processos de execução da manutenção e pelo processo de execução de serviços comerciais, as quais estão localizadas estrategicamente nas cidades-pólo de cada uma das regiões que atendem. Para facilitar, as gerências regionais são indicadas ao longo do trabalho por siglas conforme mostra a Tabela 1:

Tabela 1 - Siglas das Gerências Regionais.

| Sigla | Gerências Regionais | Quantidade Municípios |
|-------|----------------------|-----------------------|
| DV | Divinópolis | 67 |
| GV | Governador Valadares | 87 |
| IP | Ipatinga | 52 |
| JF | Juiz de Fora | 43 |
| MT | Metalúrgica | 45 |
| MP | Metropolitana | 16 |



| | | |
|----|------------------|----|
| MC | Montes Claros | 77 |
| PR | Paracatu | 23 |
| PS | Passos | 50 |
| PM | Patos de Minas | 26 |
| PA | Pouso Alegre | 73 |
| SJ | São João Del Rey | 67 |
| TO | Teófilo Otoni | 74 |
| UR | Uberaba | 29 |
| UL | Uberlândia | 17 |
| VR | Varginha | 58 |

O mapa do estado de Minas Gerais dividido por estas gerências é apresentado na Figura 1. Os pontos identificam as estações meteorológicas que fornecem os dados de velocidade do vento. As variáveis de densidade dos raios e índice pluviométrico foram fornecidas pela empresa.

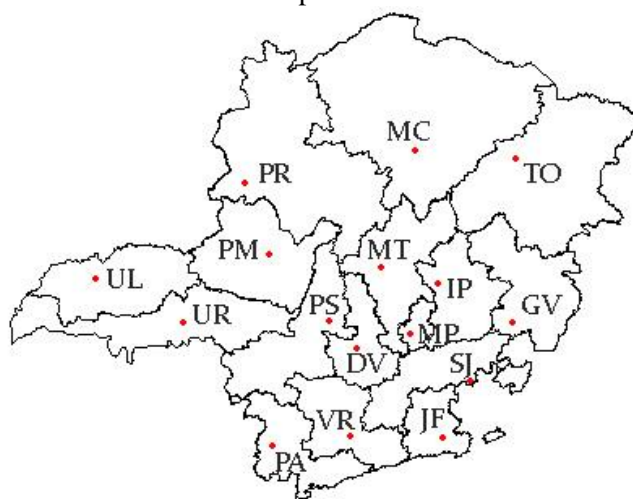


Figura 1 - Gerências regionais e estações meteorológicas referentes à variável velocidade do vento.

Após a agregação dos dados por gerência e por ano, os dados foram previamente padronizados, retirando o efeito médio do ano para todas as variáveis de insumo, produto e ambientais. As variáveis foram padronizadas conforme equação 15.

$$x_{it}^* = \bar{x}_{..} + (x_{it} - \bar{x}_{.t}) \quad (15)$$

onde $\bar{x}_{..} = \frac{1}{n} \sum_{it} x_{it}$ e $\bar{x}_{.t} = \frac{1}{n_t} \sum_i x_{it}$. O objetivo foi homogeneizar os valores médios das variáveis para os diferentes anos, sem reduzir o tamanho da amostra. Nos casos em que o número de DMUs comparadas é maior, como no modelo utilizado na ANEEL, esta homogeneização não se faz necessária.

3 Resultados

Foram ajustados dois modelos DEA em um primeiro estágio. O modelo DEA-UR apresenta a estratificação de algumas variáveis nos subgrupos urbano e rural. O modelo DEA-Total não apresenta abertura das variáveis. Em ambos os modelos, a variável PMSO é o insumo (input). No modelo DEA-Total, os produtos são comprimento total da rede, número total de consumidores e potência total fornecida MWh. Já no modelo DEA-UR, os produtos são comprimento da rede urbana, comprimento da rede rural, número de consumidores urbanos, número de consumidores rurais, potência fornecida urbana e potência fornecida rural. Na figura 2 podemos observar como estas variáveis estão correlacionadas. As variáveis “Postes Urbanos”,



“Postes Rurais” e “Postes Totais” não foram incluídas nos modelos pois estão fortemente correlacionadas com outras variáveis mais significativas.

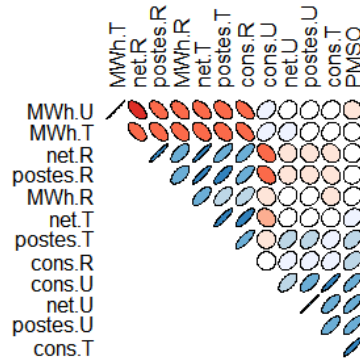


Figura 2 - Correlação de Spearman entre as variáveis disponíveis para o modelo DEA. Elipses azuis indicam uma correlação positiva enquanto elipses vermelhas indicam correlação negativa

As variáveis ambientais foram agrupadas em valores anuais de acordo com o percentil que maximiza a correlação dos seus valores com os escores de eficiência. A ideia por trás desta abordagem é a de que valores baixos de velocidade de ventos, por exemplo, exercem uma influência insignificante na eficiência da empresa. No entanto, a partir de certos valores, seu efeito já não é mais desprezível. O gráfico abaixo mostra o valor do Coeficiente de Spearman entre a variável “Velocidade de Ventos”, agrupada em valores de percentil de 0.05 a 1.00, e o escore de eficiência calculado com os modelos DEA Total e UR. O mesmo estudo foi feito para as outras variáveis ambientais.

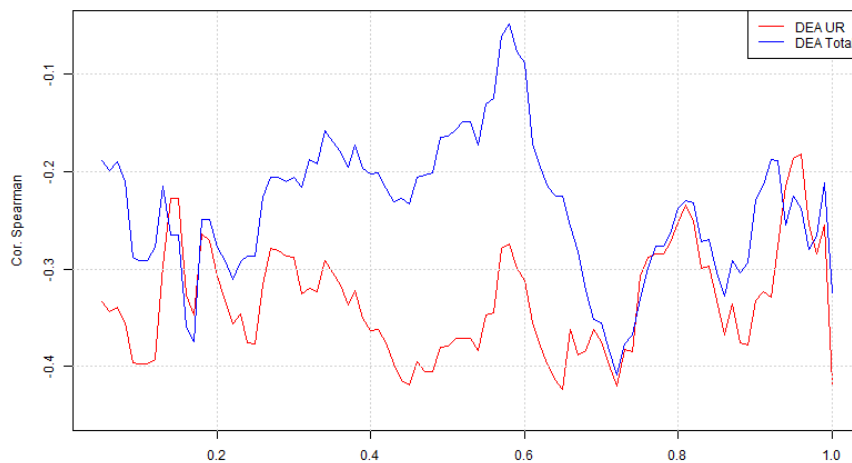


Figura 3 - Correlação de Spearman entre os escores de eficiência pelo percentil utilizado na agregação

Como podemos observar na figura 3, a correlação dos escores de eficiência com a variável “velocidade de ventos” varia de acordo com o percentil pelo qual a mesma é agregada. A agregação realizada pelo percentil de 65% maximiza o efeito negativo desta variável no cálculo de eficiência. Esta mesma análise foi feita para outras variáveis ambientais.

Considerando então esta nova variável “velocidade de ventos .65”, a tabela 2 apresenta os valores obtidos para $\hat{\beta}$, bem como o intervalo de confiança bootstrap de 95%, calculados pelo algoritmo #2, proposto por Simar e Wilson (2007).



Tabela 2 - Valores estimados para os parâmetros do modelo de regressão truncada e respectivos intervalos de confiança (95%)

| | $\hat{\beta}$ | LI | LS |
|-----------------|---------------|--------|--------|
| Intercept | 0.9161 | 0.6676 | 1.0635 |
| beta_v.vento.65 | 0.1226 | 0.0601 | 0.2109 |

Uma vez verificada a significância estatística da variável ambiental e após a correção viés as eficiências estimadas no primeiro estágio, o método proposto na sessão 2.4 foi aplicado para corrigir os escores quanto ao efeito da variável ambiental.

A figura 4 apresenta os escores corrigidos agrupados anualmente. Após a homogeneização dos dados pela variável ano, não se observa nenhuma tendência crescente ou decrescente ao longo do período dos três anos. A figura 5 apresenta as eficiências estimadas pelo modelo DEA-UR e posteriormente corrigidas pela variável ambiental, agrupadas para as gerências da CEMIG-D. Cada um dos boxplots representa as três eficiências corrigidas, uma para cada ano. Os boxplots estão ordenados segundo ordem crescente da eficiência média. As linhas sobrepostas aos boxplots representam os escores originais médios obtidos pelo DEA-UR e os escores médios obtidos após a correção do viés pelo método bootstrap. Os resultados mostram que o ranking dos escores sofreu alteração devido à correção da variável ambiental. É importante destacar a correção do viés pelo método bootstrap provoca uma redução no valor dos escores. Este efeito está de acordo com a teoria (Simar e Wilson, 1998) que relata um número maior DMUs na fronteira quando na ausência da correção do viés.

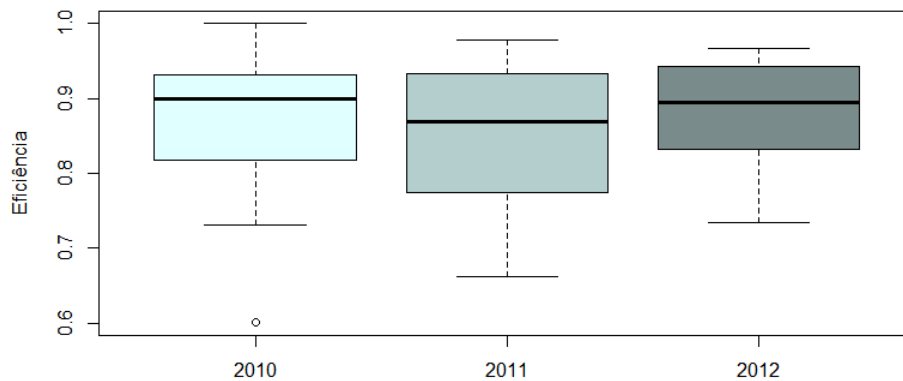


Figura 4 - Escores de eficiência corrigidos pela variável ambiental ao longo dos anos

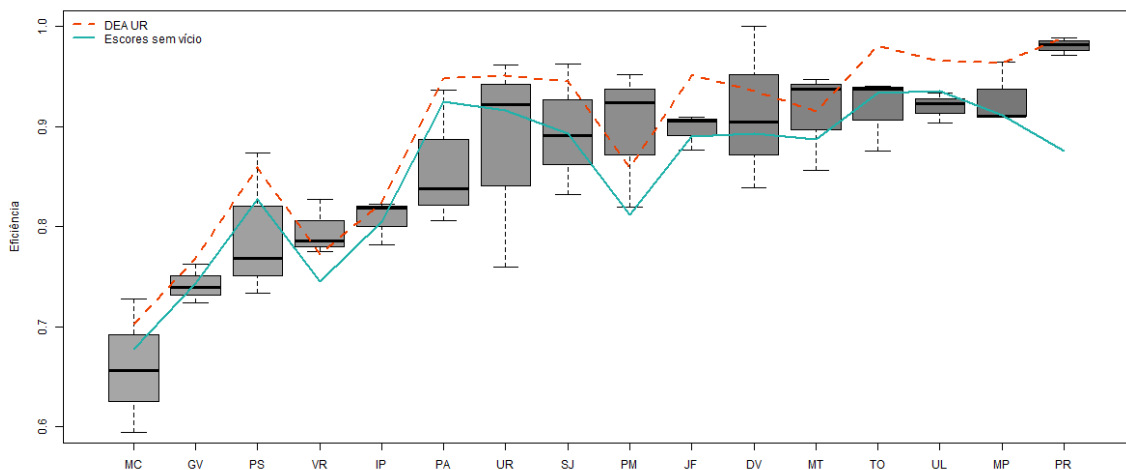


Figura 5 - Escores corrigidos ordenados pela média ao longo dos três anos



Conclusões

A construção de um modelo de Benchmarking para o cálculo de índices de eficiência para um número pequeno de DMUs é desafiador. O tamanho limitado da amostra restringe a abertura ou mesmo a inclusão de um número maior de variáveis de insumos e/ou produtos, como é o caso da base das gerências da CEMIG-D. Para contornar este problema optou-se pelo uso de dados de painel, corrigindo as médias anuais das variáveis de insumo e produto. Dessa forma, mantendo a abertura de grande parte das variáveis em suas categorias urbana e rural. Isso porque a CEMIG-D possui um amplo programa de assistências às comunidades rurais sendo algumas de suas gerências predominantemente rurais. Além disso, devido à alta resolução das variáveis ambientais disponíveis, foi possível avaliar a agregação pelos percentis. Neste trabalho, a agregação da variável ambiental “velocidade de ventos” pelo percentil 65% apresentou alta correlação estatística com os escores obtidos pelo modelo DEA-UR, mas não com o modelo DEA Total.

A correção do efeito do viés proposta por Simar e Wilson (2007) retira algumas DMUs da fronteira no modelo, o que na prática resulta em escores mais baixos. Também é possível observar na análise em segundo estágio proposta por Simar e Wilson (2007) que a variável ambiental “velocidade de ventos .65” é estatisticamente significativa. Portanto, para efeito de regulação se faz necessário a correção dos escores pelo efeito desta variável. É importante destacar que Simar e Wilson (2007) não apresentam a forma final para estimação dos escores de eficiência corrigidos, isto é, sem o efeito da componente ambiental. Este trabalho propõe uma formulação para a estimação dos escores corrigidos, homogeneizando o efeito do ambiente, apesar da distribuição truncada do erro.

Agradecimentos

Os autores reconhecem o apoio financeiro da CAPES, CNPq e FAPEMIG/CEMIG(APQ-03165-11).

Referências

- ANEEL. (2014). Nota Técnica, Metodologia de Custos Operacionais - AP 023/2014 - 2a. Fase.
- ANEEL. (2015). Nota Técnica, Metodologia de Custos Operacionais - AP 023/2014 - Final.
- Aigner, D. J., Lovell, C. A. K. e Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6: 21-37.
- Banker, R. D., Charnes, A. e W. W. Cooper (1984). Some models for estimating technical and scale inefficiencies. *Management Science*, 39: 1261-1264.
- Bogetoft, P. (2012). *Performance Benchmarking: Measuring and Managing Performance*. Springer Science & Business Media.
- Bogetoft, P., Lopes, A. L. M. (2015). Comments on the Brazilian benchmarking model for energy distribution regulation: Fourth cycle of tariff review - Technical Note 407/2014.
- Bogetoft, P., Otto, L. (2010). *Benchmarking with Dea, Sfa, and R*. Springer Science & Business Media.
- Charnes, A., Cooper, W. W. e Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2: 429-444.



Chowdhury, H., Zelenyuk, V. (2016). Performance of hospital services in Ontario: DEA with truncated regression approach. *Omega*, 63: 111-122.

da Silva, A. V. (2015). Estimação paramétrica de escores de eficiência em 2 estágios: Impacto das variáveis ambientais no ajuste das eficiências regulatórias das empresas brasileiras de distribuição de energia elétrica para o 4º Ciclo de Revisão Tarifária Periódica. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal de Minas Gerais. Orientador: Marcelo Azevedo Costa.

da Silva, A. V., Costa, M. A., Lopes, A. L. M. e do Carmo, G. M. (2017). A close look at second stage Data Envelopment Analysis using compound error models and the Tobit model. Submitted to *Socio Economic Planning Sciences*.

Färe, R. (1988). *Fundamentals of production theory*. Springer, Berlin.

Farrell, M. J. (1957). The Measurement of Productive Efficiency of Production. *Journal of the Royal Statistical Society, Series A*, 120(III), 253–281.

Forsund, F. R., Sarafoglu, N. (2002). On the Origins of Data Envelopment Analysis. *Journal of Productivity Analysis*, 17: 23 – 40.

Kuosmanen, T., Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, 38(1): 11-28.

Land, K. C., Lovell, C. A. K. e Thore, S. (1993). Chance-constrained data envelopment analysis. *Managerial and Decision Economics*, 14: 541-554.

Meeusen, W., Van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composite error. *International Economic Review*, 18: 435-444.

Olesen, O.B., Petersen, N. C. (1995). Chance constrained efficiency evaluation. *Management Science*, 41(3): 442-457.

Richmond, J. (1974). Estimating the efficiency of production. *International Economic Review*, 15(2): 515-521.

Simar, L., Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science* 44.1: 49-61.

Simar, L., Wilson, P. W. (2007). Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes. *Journal of Econometrics*, 136(1): 31-64.

Shephard, R. W. (1970). *Theory of cost and production function*. Princeton University Press, Princeton.