**Análise do modelo TOBIT no ajuste em segundo estágio dos escores de eficiências das empresas Brasileiras distribuidoras de energia elétrica**

**Marcelo Azevedo Costa, Pedro Hargreaves Passagli, Thiago Fiche Gosling**

Departamento de Engenharia de Produção

Universidade Federal de Minas Gerais

macosta@ufmg.br, pedrohargreaves@gmail.com, thiagofiche@hotmail.com

**Ana Lúcia Miranda Lopes**

Faculdade de Ciências Econômicas

Departamento de Ciências Administrativas

Universidade Federal de Minas Gerais

analopes.ufmg@gmail.com

**RESUMO**

A Agência Nacional de Energia Elétrica (ANEEL), órgão regulador do setor elétrico no Brasil, vêm utilizando a metodologia de *Benchmarking* *Data Envelopment Analysis − DEA* no cálculo da eficiência dos custos operacionais das empresas de transmissão e distribuição de energia elétrica desde 2011. No 3º ciclo de revisão tarifária periódica (3CRTP) a ANEEL aplicou ajuste em segundo estágio utilizando informações de variáveis ambientais, o que não ocorreu no 4CRTP. O ajuste em segundo estágio é aplicado em situações onde existem componentes não gerenciáveis que impactam os escores DEA de eficiência das empresas que estão sendo avaliadas. Este trabalho apresenta uma análise das variáveis ambientais e propõe o ajuste dos índices de eficiência utilizando o modelo de regressão censurada Tobit. Uma análise das propriedades do modelo Tobit e das equações de ajuste dos índices de eficiência é apresentada. Os resultados indicam mudanças significativas nos escores de eficiência das distribuidoras de energia, especialmente as distribuidoras localizadas na região norte do país.

**PALAVRAS CHAVE. Data Envelopment Analysis - DEA, Tobit, Segundo estágio.**

**Tópicos: DEA - Análise Envoltória de Dados**

**ABSTRACT**

The Brazilian National Electricity Energy Agency (ANEEL), which is the regulatory agency of the electricity sector in Brazil. It has been using the benchmarking model Data Envelopment Analysis (DEA) for calculating the operational cost efficiencies of transmission and distribution electricity companies since 2011. During the third periodic tariff revision cycle (3CRTP), ANEEL applied second stage models using information from environmental variables. This approach was not used in the following 4CRTP. Second stage modeling is applied in situations in which non-manageable factors impact the efficiencies of companies. This paper presents statistical analysis of the environmental variables and proposes adjustments of the efficiency scores using the Tobit regression model. Statistical analysis of the Tobit regression model and the adjusted efficiency equation are presented. Results show significant changes in the efficiency scores of the Brazilian electricity distribution companies, especially those located in the northern region.

**KEYWORDS. Data Envelopment Analysis - DEA, Tobit, Second stage.**

**Paper topics: DEA – Data Envelopment Analysis**

**1. Introdução**

Em 26 de outubro de 2011, a Agência Nacional de Energia Elétrica (ANEEL) publicou a nota técnica (NT) no. 294/2011 definida durante a Audiência Pública 040/2010(AP 40/2010), descrevendo a metodologia final para o cálculo dos custos operacionais regulatórios das empresas brasileiras de distribuição de energia elétrica para o terceiro ciclo de revisão tarifária periódica (3CRTP). Dois modelos foram utilizados para o cálculo dos custos operacionais eficientes*:* *Data Envelopment Analysis* (DEA)[Charnes, Cooper e Rhodes, 1978] e *Corrected Ordinary Least Squares* (COLS). Uma comparação entre DEA e COLS é apresentada em Banker et al. (1993). Cada modelo estimou um escore ou índice de eficiência para as distribuidoras e a média dos mesmos foi utilizada para o cálculo dos custos operacionais eficientes. Uma vez obtidos os escores de eficiência das distribuidoras, foi aplicada uma metodologia para sua correção pela influência do ambiente, chamada de ajuste em segundo estágio. As variáveis ambientais analisadas foram: tendência, salário médio, densidade de consumidores, índice de precipitação e índice de complexidade. Foram utilizados três métodos de estimação de regressão: (1) Simar e Wilson (2007) (Algoritmo I adaptado); (2) modelo proposto em Banker e Natarajan (2008) e (3) um modelo mais tradicional aplicando simplesmente o modelo Tobit no segundo estágio. Os resultados finais apresentaram grandes distorções, sendo alvo de inúmeras críticas por parte das distribuidoras e da sociedade.

Em 24 de abril de 2015, por meio da NT 66/2015-SEM/SGT/ANEEL, o regulador apresentou o modelo que iria vigorar para o cálculo dos custos operacionais eficientes durante o quarto ciclo de revisão tarifária periódica (4CRTP). A nota técnica define o modelo DEA-NDRS (*non-decreasing returns to scale*) com restrição aos pesos, utilizando o custo operacional ajustado como variável de insumo e sete variáveis de produtos: extensão da rede de alta tensão, extensão de rede aérea, extensão de rede subterrânea, mercado ponderado (MWh), número total de consumidores, número estimado de consumidores-horas de energia interrompido e montante das perdas não-técnicas (MWh)). O banco de dados consiste em valores médios no período 2011 a 2013 para as 61 empresas de distribuição. Diferente do modelo do terceiro ciclo (3CRTP), neste momento as variáveis não discricionárias ou variáveis ambientais, apesar de constantes do banco de dados, não foram utilizadas sob alegação de que o regulador não encontrou correlação estatisticamente significativa entre os escores de eficiência e as variáveis não discricionárias.

Em geral, as variáveis de insumos e produtos utilizadas nos modelos DEA estão associadas a fatores controláveis, ou seja, as variáveis de produção que podem ser gerenciadas pelo tomador de decisão a fim de ser eficiente. Um outro conjunto de variáveis − não necessariamente menos importantes − podem afetar a produção e são, em geral, não gerenciáveis. Essas variáveis são conhecidas como variáveis não-discricionárias, ambientais ou contextuais [Ray, 1988]. Exemplos de variáveis contextuais são fatores climáticos [Yu et al. 2009], tais como temperatura, precipitação; tipo de solo, nível educacional [Ray e Ghose, 2014]; entre outras.

O ajuste dos índices de eficiência utilizando variáveis ambientais tem como objetivo compensar o efeito do ambiente, não gerenciável, no índice de eficiência das empresas. Ou seja, empresas localizadas em um ambiente favorável à sua produção devem ter seus escores de eficiência reduzidos, uma vez que o ambiente contribui, em parte, para o índice de eficiência alcançado pois DEA faz uma análise comparativa. Por outro lado, empresas localizadas em um ambiente hostil devem ter seus escores de eficiência incrementados, uma vez que o mesmo pode impedir as empresas de atingir índices de eficiência mais elevados. Embora seja possível incluir variáveis ambientais como insumos ou produtos adicionais nos modelos DEA, a análise mais comum consiste em avaliar o efeito ambiental no índice de eficiência em uma etapa posterior. Esta abordagem é conhecida como ajuste em segundo estágio [Ray, 1988].

Diferentes modelos estatísticos de regressão podem ser aplicados no ajuste dos índices de eficiência, tais como regressão Tobit [Tobin, 1958], modelos de máxima verossimilhança [Aigner et al, 1977], regressão truncada [Johnson e Kuosmanen, 2012], Mínimos Quadrados Ordinários (OLS) [Montgomery et al. 2015], entre outros. Entretanto, nesses modelos, a variável resposta corresponde ao índice de eficiência, cujo espaço amostral está definido no intervalo unitátio 0−1. Soma-se a isso o fato de que os modelos DEA permitem estimar eficiências iguais a 100% para mais de uma empresa, o que caracteriza os índices de eficiência como variáveis aleatórias contínuas e censuradas à direita em 1 (100%). Nesses casos, sob a ótima da regressão estatística, o modelo de regressão Tobit é o mais indicado.

O objetivo deste artigo é o avaliar a real influência de variáveis ambientais nos escores de eficiência das empresas de distribuição de energia elétrica Brasileiras, utilizando os dados do 4CRTP. Busca-se, também, propor um modelo de 2º estágio que corrija adequadamente estes escores fazendo com que os mesmos reflitam a correta eficiência/ineficiência gerencial de cada empresa.

Este artigo está dividido em seis seções iniciado por esta introdução. A seção 3 faz uma revisão de regressão linear simples, enquanto a Seção 3 aprofunda no modelo de regressão censurada Tobit e a Seção 4 apresenta a base de dados utilizada. A Seção 5 apresenta os resultados obtidos com o modelo proposto e a Seção 6 encerra este artigo.

**2. Análise de Regressão Linear Simples**

A análise de regressão é um método estatístico de análise de dados que permite mensurar o efeito de uma variável independente (ou regressora) no comportamento médio de uma variável dependente (ou variável resposta). No caso estudado, deseja-se avaliar o efeito marginal de variáveis ambientais no comportamento dos escores de eficiência das empresas de distribuição de energia elétrica, avaliadas no 4º ciclo de revisão tarifária periódica (4CRTP).

Um modelo de regressão linear simples assume uma equação de primeira ordem, ou equação de uma reta, para estabelecer a relação entre a variável independente () e a variável resposta ():

|  |  |
| --- | --- |
| , onde | (1) |

onde () é a variável resposta (ou dependente), é a variável preditora, é o intercepto do modelo de regressão linear, é o coeficiente angular do modelo de regressão linear e é uma variável aleatória, ou erro, que se comporta segundo uma distribuição Normal com média 0 e variância .

Os parâmetros da Equação 1, e , são estimados a partir da minimização da soma dos quadrados dos erros:

|  |  |
| --- | --- |
|  | (2) |

onde é o tamanho amostral. De forma equivalente, os parâmetros podem ser estimados a partir da maximização da função de Verossimilhança:

|  |  |
| --- | --- |
|  | (3) |

***2.1 Decomposição da Soma dos Quadrados Totais***

É de interesse deste trabalho, estimar uma nova variável resposta, , criada a partir dos valores observados da variável resposta . A variável deve possuir correlação linear nula para com a variável preditora . Ou seja, onde é a componente de que é explicada pela variável preditora . Para proceder à especificação da componente é necessário revisar o princípio da **decomposição da Soma dos Quadrados dos Totais** [Montgomery et al. 2015].

A **Soma dos Quadrados Totais** () é definida pela Equação 3 e representa a soma dos quadrados das diferenças entre os valores observados e a média amostral () da variável resposta. Ou seja, mede a dispersão dos valores observados em relação à média amostral.

|  |  |
| --- | --- |
|  | (3) |

Quando do ajuste do modelo de regressão linear simples, com intercepto, é possível demonstrar a seguinte expressão da decomposição da :

|  |  |
| --- | --- |
|  | (4) |

onde é a estimativa da equação de regressão. Definem-se os termos à direita da Equação 4 como a **Soma dos Quadrados de Regressão**, , e a **Soma dos Quadrados dos Resíduos**, .

é a soma dos quadrados das diferenças entre o valor estimado pela reta de regressão e a média amostral da resposta. representa a componente da soma dos quadrados totais que está associada ao modelo de regressão. é a soma dos quadrados das diferenças entre o valor observado da resposta e o valor estimado pelo modelo de regressão. também pode ser definida como a componente da soma dos quadrados totais que não é explicada pelo modelo de regressão. No caso particular do modelo de regressão linear simples, a componente de variância do erro é estimada a partir da :

|  |  |
| --- | --- |
|  | (5) |

A Figura 1 ilustra a decomposição da soma dos quadrados totais. Em síntese, a variabilidade dos dados com relação à média amostral (Fig. 1(a)) pode ser decomposta pela variabilidade que não é explicada pelo modelo de regressão (Fig. 1(b)) mais a variabilidade que é explicada pelo modelo de regressão (Fig. 1(c)). Como consequência da decomposição da soma dos quadrados totais, é possível definir o coeficiente de determinação, ou R2:

|  |  |
| --- | --- |
|  | (6) |

O coeficiente de determinação () indica a proporção ou percentagem da soma dos quadrados totais que é atribuído ao modelo de regressão. Seu valor encontra-se no intervalo 0 a 1 (100%). onde 1 indica que o modelo de regressão explica 100% da variabilidade dos dados em relação à média amostral.

|  |  |
| --- | --- |
| fig05a.png  **(a) Soma dos Quadrados Totais (SQT)** é calculada a partir das diferenças entre os valores observados e a média amostral, ou seja, é uma medida da dispersão dos dados com relação à média amostral. Mensura o quanto da dispersão dos dados a média amostral não é capaz de explicar | fig05b.png  **(b) Soma dos Quadrados dos Resíduos (SQres)** é calculada a partir das diferenças entre os valores observados e os valores estimados pelo modelo de regressão linear simples. Mensura o quanto da dispersão dos dados o modelo de regressão não é capaz de explicar. |
| fig05c.png  **(c) Soma dos Quadrados de Regressão (SQReg)** é calculada a partir das diferenças entre os valores estimados pelo modelo de regressão linear simples e a média amostral. Mensura o quanto da dispersão dos dados com relação ao modelo da média é explicada pelo modelo de regressão linear simples. | |

**Figura 1.** Decomposição da soma dos quadrados totais em soma dos quadrados de regressão e soma dos quadrados dos resíduos.

Utilizando os princípios da decomposição da soma dos quadrados totais, em particular a soma dos quadrados de regressão, é possível definir a componente como:

|  |  |
| --- | --- |
|  | (7) |

onde .

**3. Modelo de Regressão Censurada Tobit**

A análise dos escores de eficiência das empresas de distribuição de energia elétrica requer uma reestruturação do modelo de regressão linear. A suposição de que a variável resposta se comporta segundo uma distribuição Normal não pode ser considerada. No caso estudado, a variável resposta, defina pelos escores de eficiência, está definida no intervalo . Para tratar esta característica, o modelo Tobit define uma variável latente compatível com o modelo apresentado na Equação (1):

|  |  |
| --- | --- |
|  | (8) |

Os valores observados estão associados às respectivas variáveis latentes na forma:

|  |  |
| --- | --- |
|  | (9) |

Segundo a Equação (9), para os escores de eficiência abaixo de 1 e maiores que zero, a variável latente é igual à variável observada. Para os escores de eficiência iguais a 1, a variável latente é desconhecida e maior que 1, enquanto que para valores iguais a zero a mesma pode tomar valores negativos ou iguais a zero. No caso estudado, os escores de eficiências não assumem valores iguais a zero. Assim, a partir das Equações (8) e (9), é possível definir a distribuição de probabilidade da variável resposta:

|  |  |
| --- | --- |
|  | (10) |

onde é uma função indicadora, , , é a função densidade da normal padrão e é a função acumulada da distribuição normal padrão.

A Equação (10) representa uma distribuição de probabilidade mista, que contempla o comportamento contínuo da variável resposta , quando , e o seu comportamento discreto, quando .

A partir da Equação (10) é possível escrever a equação log-verossimilhança na forma:

|  |  |
| --- | --- |
|  | (11) |

Os novos estimadores do modelo de regressão Tobit são definidos como .

***3.1 Cálculo dos escores de eficiência ajustados pelo modelo Tobit***

O objetivo final da análise de regressão Tobit é retirar a componente do escore de eficiência associada à variável ambiental, ou regressora (). Embora a decomposição da Soma dos Quadrados Totais, apresentada na seção 2.2, não possa ser diretamente aplicada à variável resposta , o princípio da decomposição pode ser aplicado à variável latente . Então, a componente da variabilidade da resposta latente associada à variável preditora pode ser definida por:

|  |  |
| --- | --- |
|  | (12) |

onde . Como consequência, uma nova estimativa da componente latente, sem o efeito da variável preditora, pode ser calculada como:

|  |  |
| --- | --- |
|  | (13) |

Finalmente, o escore de eficiência corrigido é obtido aplicando-se a Equação (9) à componente latente corrigida:

|  |  |
| --- | --- |
|  | (14) |

Entretanto, um passo intermediário é necessário, pois a Equação (12) pressupõe que todos os valores latentes são conhecidos, o que não é verdade quando . Nesses casos, é possível definir estimativas pontuais utilizando a densidade condicional, . Utilizando teoria de probabilidades [Sheldon, 2002], é possível demonstrar que a densidade condicional de interesse é uma distribuição normal truncada à esquerda em 1,

|  |  |
| --- | --- |
|  | (15) |

Um possível estimador pontual para as variáveis latentes desconhecidas, utilizado neste trabalho, é a esperança da densidade condicional:

|  |  |
| --- | --- |
|  | (16) |

onde .

**4. A base de dados das distribuidoras Brasileiras de energia elétrica**

Durante o 4CRTP, o regulador disponibilizou, publicamente, uma base de dados para as 61 empresas brasileiras de distribuição de energia elétrica. A base contém valores médios calculados no período de 2011 a 2013. Para o modelo DEA, foi considerada a variável custo operacional ajustado médio como insumo e as seguintes variáveis de produto: rede subterrânea, rede aérea, rede de alta tensão, número total de consumidores, mercado de energia ponderado (MWh), perdas não técnicas de energia (MWh) e consumidor-hora interrompido. As duas últimas variáveis foram incluídas no modelo DEA como produtos negativos [Bogetoft e Otto, 2010]. Além das variáveis de insumo e produtos, estão disponíveis 11 variáveis ambientais: densidade de consumidores, densidade de rede, índice de complexidade, índice de precipitação, taxa de incidência de raios, índice de vegetação de grande porte, índice de vegetação de médio porte, índice de vegetação de pequeno porte, declividade média, proporção de estradas pavimentadas e área de concessão (km²). Embora não possam ser consideradas variáveis puramente ambientais, também estão incluídas as variáveis: duração média de interrupção de energia (DEC) e freqüência de interrupção de energia (FEC), totalizando 13 variáveis não discricionárias para serem avaliadas em um possível ajuste dos escores de eficiência do modelo DEA em segundo estágio

A partir de combinações lineares das 13 variáveis, citadas anteriormente, foram propostas três novas variáveis: (i) utilizando um modelo de regressão linear múltiplo, no qual foi considerada a variável FEC como dependente e as 11 variáveis ambientais, foi estimada a variável ***z.ambiental***; e (ii) utilizando Análise de Componentes Principais e Análise Fatorial [Johnson e Wichern, 1992] foram estimados os dois primeiros fatores, ou construtos ortogonais, considerando a informação contida nas 13 variáveis não discricionárias. Essas duas variáveis foram denominadas: ***Fator 1*** e ***Fator 2***.

**5. Resultados**

A Figura 2 (a) mostra o mapa do Brasil em função das áreas de concessão das distribuidoras de energia elétrica e das eficiências estimadas pelo modelo DEA. O mapa indica um grupo de empresas com eficiências baixas na região norte, grupos de empresas com eficiências elevadas no estado de São Paulo, região sul e nordeste. Lembrando que quanto mais próximo de 1, maior é a eficiência de cada empresa. A Figura 2(b) apresenta o gráfico de correlação de *Spearman* [Fieller et al. 1957] entre as 16 variáveis ambientais e os escores de eficiência do modelo DEA. Correlações positivas são indicadas por elipses com inclinação positiva e correlações negativas são indicadas por elipses com inclinações negativas. Quanto mais estreita for a elipse, maior o valor absoluto da correlação. A última coluna apresenta a correlação de *Spearman* entre os escores de eficiência e cada variável ambiental. Em geral, as correlações são negativas e as variáveis que apresentam as maiores correlações com as eficiências do modelo DEA são: FEC, z.ambiental, precipitação, Fator 1, DEC, Fator 2 e vegetação de grande porte.

|  |  |
| --- | --- |
|  | fig_ellipse.png |
| (a) Mapa das eficiências estimadas pelo modelo DEA (primeiro estágio) | (b) Matriz de correlação de Spearman |

**Figura 2.** Mapa das eficiências estimadas pelo modelo DEA no primeiro estágio (a). Matriz de correlação de Spearman entre as variáveis ambientais e os escores de eficiência do modelo DEA (b).

A Tabela 1 apresenta os coeficientes de regressão estimados utilizando o modelo de regressão linear simples, conforme Equação (1), e o modelo Tobit. Foram ajustados modelos univariados (somente uma variável ambiental no modelo). Considerando um nível de confiança de 0,05 (5%), os resultados indicam 6 variáveis estatisticamente significativas. São elas: FEC, z.ambiental, Fator 1, precipitação, vegetação grande porte e DEC.

A Figura 3 ilustra o ajuste do modelo Tobit e as estimativas dos índices de eficiência corrigidos utilizando a variável ambiental z.ambiental. A Figura 3(a) mostra as eficiências do modelo DEA (escores observados), o ajuste do modelo Tobit e as estimativas das variáveis latentes. Destaca-se que os valores latentes são iguais aos escores observados, se os mesmos forem menores que a unidade (100%). Os valores latentes, para as empresas com eficiências iguais a 100%, são maiores que a unidade. A Figura 3(b) mostra as variáveis latentes ajustadas e os escores finais. É importante observar que, após o ajuste, não há correlação entre os escores corrigidos e a variável ambiental (z.ambiental). Portanto, a equação de regressão ajustada (reta ajustada) possui coeficiente de inclinação nulo.

**Tabela 1.** Comparação do resultado do ajuste do modelo de regressão linear simples e do modelo Tobit para cada variável ambiental.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Modelo de regressão linear** | | **Modelo Tobit** | |
| **Variável preditora** | **Coeficiente** | **p-valor** | **Coeficiente** | **p-valor** |
| Frequência de interrupção energia (FEC) | -0,00838 | 0,00024 | -0,00883 | **0,00017** |
| z.ambiental | -0,01068 | 0,00038 | -0,01101 | **0,00039** |
| Fator 1 | -0,06684 | 0,00718 | -0,07179 | **0,00635** |
| Índice de Precipitação | -0,00020 | 0,00841 | -0,00019 | **0,01614** |
| Vegetação de grande porte | -0,49652 | 0,02084 | -0,51690 | **0,02376** |
| Duração de interrupção de energia (DEC) | -0,00367 | 0,02588 | -0,00387 | **0,02787** |
| Densidade de rede | -0,01321 | 0,08469 | -0,01352 | 0,10026 |
| Fator 2 | -0,04254 | 0,09309 | -0,04054 | 0,14106 |
| Proporção de estradas pavimentadas | 0,12279 | 0,29810 | 0,16177 | 0,21702 |
| Área de concessão | 0,00000 | 0,28811 | 0,00000 | 0,34611 |
| Taxa de incidência de raios | -0,00548 | 0,51613 | -0,00467 | 0,61270 |
| Índice de complexidade | -0,10631 | 0,67510 | -0,13303 | 0,63058 |
| Densidade de consumidores | 0,00003 | 0,80473 | 0,00003 | 0,82994 |
| Vegetação de médio porte | -0,04034 | 0,80618 | -0,02203 | 0,90343 |
| Vegetação de baixo porte | -0,06812 | 0,86041 | -0,02550 | 0,95225 |
| Declividade | 0,00054 | 0,93143 | 0,00026 | 0,97038 |

|  |  |
| --- | --- |
| figura.png | figura.png |
| (a) Comparação dos escores DEA, ajuste Tobit e estimativas das variáveis latentes. | (b) Variáveis latentes ajustadas, escores ajustados e reta de regressão ajustada. |

**Figura 3.** Ilustração do modelo Tobit ajustado e das variáveis latentes estimadas (a). Estimativa final das variáveis latentes e dos escores ajustados após a retirada da componente de regressão da variável z.ambiental.

A Figura 4 mostra os mapas das mudanças nas eficiências das empresas decorrentes do ajuste em segundo estágio utilizando o modelo Tobit aonde ocorre redução na eficiência (depois < antes), aumento na eficiência (depois > antes) e sem alteração (depois = antes). Os resultados mostram que, mesmo utilizando as diferentes variáveis ambientais, as empresas localizadas na região norte sempre apresentaram um aumento nos seus índices de eficiência após a correção em segundo estágio. Por outro lado, algumas poucas empresas também foram consistentemente indicadas como empresas para as quais não há mudança nos seus índices de eficiência. O mesmo ocorreu para um grupo de empresas para as quais a correção em segundo estágio sempre resultou em uma redução nos seus índices de eficiência.

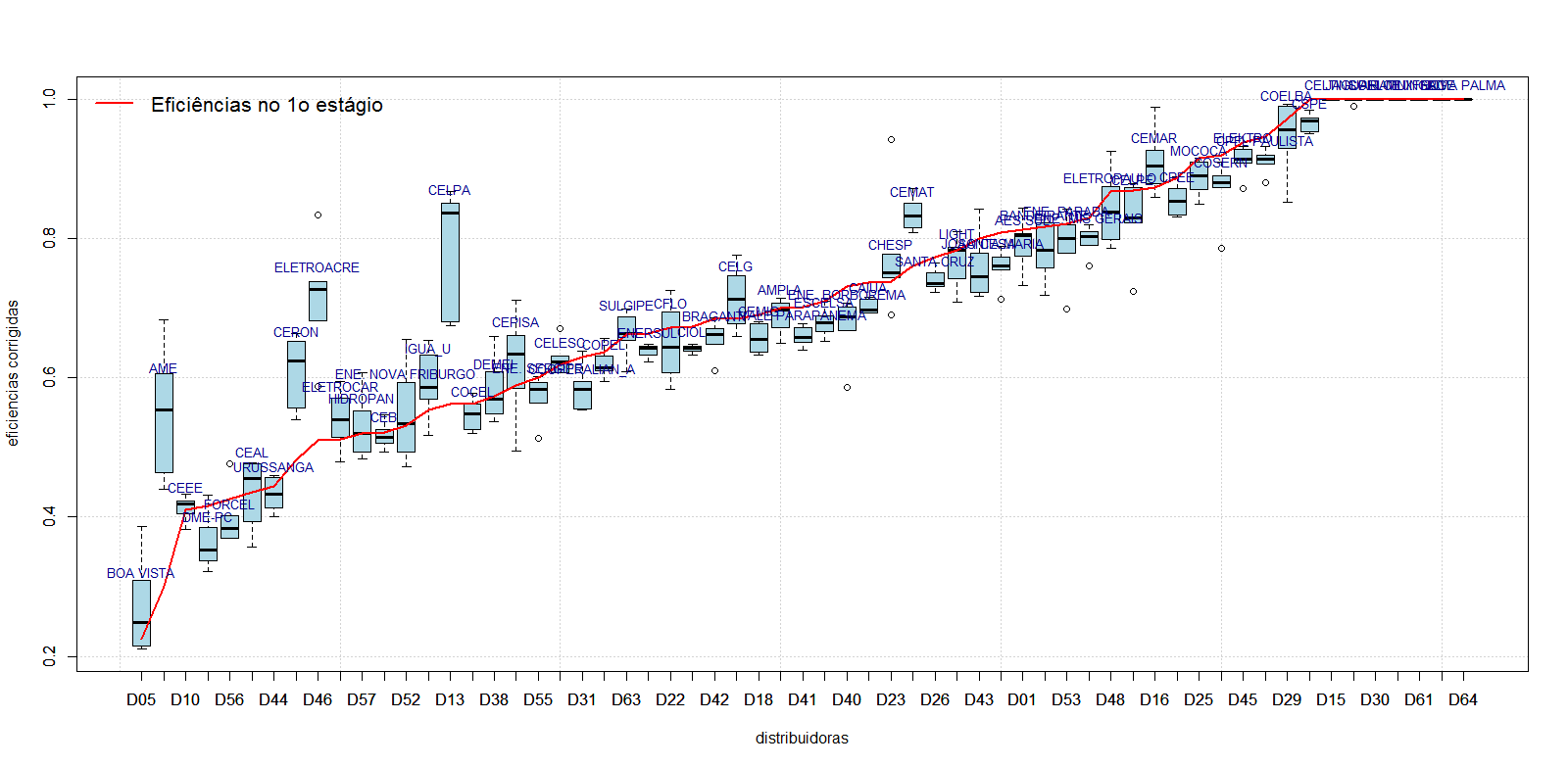
|  |  |  |
| --- | --- | --- |
| mapa_final.png  (a) FEC | mapa_final.png  (b) z.ambiental | mapa_final.png  (c) Fator 1 |
|  |  |  |
| mapa_final.png  (d) Precipitação | mapa_final.png  (e) Vegetação grande porte | mapa_final.png  (f) DEC |

**Figura 4.** Comparação das mudanças nos escores de eficiência das distribuidoras Brasileiras em função do ajuste em segundo estágio utilizando o modelo Tobit.

A Figura 5 compara os resultados das eficiências ajustadas, utilizando os modelos Tobit univariados (boxplots), com os valores originais das eficiências estimadas pelo modelo DEA (linha sólida). Cada boxplot representa os escores ajustados para cada um dos 6 modelos Tobit univariados, e estatisticamente significativos, apresentados na Tabela 1. Os boxplots são apresentados para cada empresa, ordenados de forma crescente das eficiências no primeiro estágio (DEA). Os boxplots, que estão, acima da linha representam as empresas cujas eficiências aumentaram após o ajuste pelo modelo Tobit. Como mencionado, algumas empresas também apresentaram reduções em suas eficiências. Os boxplots que se sobrepõem à linha representam empresas que apresentaram aumento e redução nos ajustes em segundo estágio, dependendo da variável ambiental utilizada. O detalhamento dos valores finais dos índices de eficiência e sua comparação com o índice de eficiência do modelo DEA é apresentado na Tabela A, no Apêndice do texto.

**6. Discussão e conclusão**

Este trabalho teve como objetivo descrever, aplicar e analisar os resultados do ajuste em segundo estágio dos índices de eficiência das empresas brasileiras de distribuição de energia elétrica, utilizando o modelo Tobit. Embora o modelo Tobit tenha sido ajustado separadamente para cada variável ambiental disponível, vale destacar que foram criadas três variáveis ambientais a partir de combinações lineares das 11 variáveis definidas como puramente ambientais. Foram analisados modelos Tobit com mais de uma variável. Mas, os resultados indicaram que somente os modelos considerando uma variável se mostraram estatisticamente significativos e, por esta razão, os resultados dos modelos com mais de uma variável ambiental não foram incluídos neste texto.

****

**Figura 5.** Comparação das eficiências corrigidas, utilizando o modelo Tobit para cada uma das 6 variáveis ambientais selecionadas, com as eficiências originalmente estimadas pelo modelo DEA.

É importante destacar que o model Tobit é um modelo muito similar ao modelo clássico de análise de regressão linear. Entretanto, o modelo Tobit permite a análise de dados censurados, que é o caso dos índices de eficiência. Por outro lado, o ajuste dos índices de eficiência implica em estimativas pontuais para as variáveis latentes desconhecidas, o que requer conhecimentos em probabilidades e decomposição da soma de quadrados totais.

As principais conclusões deste estudo são: (i) há evidência estatística de correlação entre as eficiências estimadas pelo modelo DEA no 4CRTP e as variáveis ambientais; (ii) foram identificadas seis variáveis ambientais estatisticamente significativas; e (iii) os ajustes do modelo Tobit indicaram, sistematicamente, a necessidade de correções das eficiências de empresas localizadas na região norte.

**Agradecimentos**

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

**Referências**

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1), 21-37.

Banker, R. D., Gadh, V. M., & Gorr, W. L. (1993). A Monte Carlo comparison of two production frontier estimation methods: corrected ordinary least squares and data envelopment analysis. *European Journal of Operational Research*, 67(3), 332-343.

Bogetoft, P., & Otto, L. (2010). Benchmarking with Dea, Sfa, and R (Vol. 157). Springer Science & Business Media.

Cooper, W. W., Seiford, L. M., & Zhu, J. (2004). Data envelopment analysis. Springer US.

Fieller, E. C., Hartley, H. O., & Pearson, E. S. (1957). Tests for rank correlation coefficients. I. *Biometrika*, 44(3/4), 470-481.

Johnson, A. L., & Kuosmanen, T. (2012). One-stage and two-stage DEA estimation of the effects of contextual variables. *European Journal of Operational Research*, 220(2), 559-570.

Johnson, R. A., & Wichern, D. W. (1992). Applied multivariate statistical analysis (Vol. 4). Englewood Cliffs, NJ: Prentice hall.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). Introduction to linear regression analysis. John Wiley & Sons.

Ray, S. C. (1988). Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation. *Socio-Economic Planning Sciences*, 22(4), 167-176.

Ray, S. C., & Ghose, A. (2014). Production efficiency in Indian agriculture: An assessment of the post green revolution years. *Omega*, 44, 58-69.

Sheldon, R. (2002). A first course in probability. Pearson Education India.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, 24-36.

Yu, W., Jamasb, T., & Pollitt, M. (2009). Does weather explain cost and quality performance? An analysis of UK electricity distribution companies. *Energy Policy*, 37(11), 4177-4188.

**Apêndice**

**Tabela A.** Resultado do ajuste em segundo estágio das eficiências das empresas Brasileiras de distribuição de energia elétrica.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Distribuidora** | **DEA** | **FEC** | **z.ambiental** | **Fator 1** | **Precipit.** | **Veg. alta** | **DEC** |
| BOA VISTA | 22,46% | 30,95% | 38,66% | 21,55% | 27,25% | 22,42% | 21,05% |
| AME | 30,12% | 60,55% | 68,31% | 55,90% | 44,00% | 54,89% | 46,36% |
| CEEE | 41,07% | 42,19% | 43,32% | 38,28% | 41,47% | 40,51% | 42,24% |
| DME-PC | 41,60% | 33,70% | 32,20% | 34,11% | 43,20% | 38,46% | 36,49% |
| FORCEL | 42,52% | 36,99% | 40,20% | 38,52% | 47,67% | 38,35% | 36,94% |
| CEAL | 43,51% | 47,82% | 44,54% | 46,61% | 35,74% | 39,33% | 47,59% |
| URUSSANGA | 44,43% | 40,02% | 43,64% | 42,80% | 45,63% | 45,93% | 41,28% |
| CERON | 48,24% | 62,73% | 65,21% | 62,18% | 53,94% | 66,29% | 55,69% |
| ELETROACRE | 50,96% | 83,29% | 73,75% | 73,80% | 58,77% | 71,60% | 68,14% |
| ELETROCAR | 51,09% | 55,41% | 57,06% | 51,47% | 59,43% | 47,95% | 52,59% |
| HIDROPAN | 51,98% | 54,12% | 55,25% | 49,78% | 60,70% | 48,32% | 49,37% |
| CEB | 52,19% | 54,71% | 49,32% | 51,72% | 51,08% | 50,59% | 52,52% |
| ENE. NOVA FRIBURGO | 53,19% | 49,36% | 47,28% | 56,12% | 59,31% | 65,55% | 50,58% |
| IGUA\_U | 55,35% | 65,36% | 59,04% | 58,32% | 63,21% | 51,69% | 56,88% |
| CELPA | 56,19% | 86,75% | 82,79% | 84,42% | 68,02% | 67,52% | 85,06% |
| COCEL | 56,20% | 52,62% | 52,01% | 55,36% | 56,28% | 57,71% | 54,21% |
| DEMEI | 57,37% | 60,93% | 56,63% | 54,84% | 65,87% | 53,71% | 57,16% |
| CEPISA | 58,94% | 71,08% | 58,49% | 65,72% | 49,47% | 60,97% | 66,00% |
| ENE. SERGIPE | 59,99% | 59,23% | 57,32% | 59,37% | 51,24% | 56,33% | 60,20% |
| CELESC | 61,88% | 60,74% | 63,14% | 60,73% | 67,12% | 62,87% | 61,64% |
| COOPERALIAN\_A | 63,02% | 55,36% | 59,48% | 55,51% | 63,75% | 58,33% | 58,32% |
| COPEL | 63,62% | 59,51% | 63,09% | 61,59% | 65,67% | 60,99% | 61,23% |
| SULGIPE | 66,26% | 65,79% | 69,84% | 67,01% | 60,83% | 68,80% | 65,35% |
| ENERSUL | 66,38% | 62,21% | 64,85% | 63,30% | 64,20% | 64,27% | 64,51% |
| CFLO | 67,14% | 60,76% | 69,49% | 58,34% | 72,56% | 66,07% | 62,56% |
| CIOL | 67,27% | 64,86% | 64,55% | 63,31% | 64,47% | 64,12% | 63,86% |
| BRAGANTI | 68,44% | 66,23% | 60,94% | 66,27% | 68,59% | 64,78% | 67,03% |
| CELG | 68,52% | 77,59% | 70,24% | 72,25% | 67,79% | 65,89% | 74,58% |
| CEMIG | 68,99% | 63,74% | 63,19% | 68,02% | 64,17% | 66,88% | 67,77% |
| AMPLA | 69,98% | 67,26% | 64,87% | 71,43% | 70,00% | 69,42% | 70,74% |
| VALE PARAPANEMA | 70,09% | 63,94% | 67,23% | 65,03% | 67,81% | 65,91% | 65,65% |
| ESCELSA | 71,05% | 65,28% | 67,40% | 70,93% | 66,69% | 68,94% | 68,34% |
| ENE. BORBOREMA | 73,07% | 69,18% | 66,70% | 70,08% | 58,65% | 68,37% | 70,72% |
| CAIUA | 73,63% | 69,33% | 71,61% | 69,29% | 71,24% | 69,45% | 69,88% |
| CHESP | 73,73% | 94,18% | 75,18% | 77,67% | 74,31% | 69,04% | 74,88% |
| CEMAT | 76,11% | 85,04% | 87,15% | 83,60% | 80,89% | 82,79% | 81,54% |
| SANTA CRUZ | 77,37% | 72,27% | 75,01% | 73,10% | 76,41% | 73,71% | 73,46% |
| LIGHT | 78,24% | 74,22% | 70,83% | 80,94% | 77,99% | 78,72% | 78,53% |
| JOAO CESA | 79,93% | 71,62% | 72,26% | 73,41% | 84,24% | 77,82% | 75,63% |
| SANTA MARIA | 80,87% | 75,43% | 71,29% | 78,82% | 75,50% | 76,69% | 77,36% |
| AES SUL | 81,31% | 77,46% | 80,40% | 73,30% | 84,35% | 80,75% | 80,34% |
| BANDEIRANTE | 81,73% | 75,75% | 71,80% | 78,04% | 82,25% | 82,20% | 78,65% |
| ENE. PARABA | 82,10% | 81,94% | 78,34% | 81,67% | 69,80% | 77,92% | 84,24% |
| ENE. MIS GERAIS | 82,92% | 79,06% | 76,07% | 81,02% | 81,92% | 80,30% | 80,18% |
| ELETROPAULO | 86,80% | 79,83% | 78,56% | 83,92% | 87,48% | 92,44% | 83,66% |
| CELPE | 86,93% | 82,54% | 82,26% | 87,32% | 72,39% | 83,27% | 87,85% |
| CEMAR | 87,35% | 85,95% | 98,87% | 92,64% | 87,83% | 91,96% | 88,75% |
| CPEE | 88,76% | 83,14% | 83,35% | 85,11% | 88,80% | 87,16% | 85,46% |
| MOCOCA | 91,52% | 84,96% | 89,68% | 88,17% | 90,95% | 91,48% | 87,08% |
| COSERN | 91,92% | 88,25% | 87,31% | 89,03% | 78,51% | 87,74% | 90,95% |
| ELEKTRO | 93,82% | 87,23% | 91,43% | 91,43% | 92,72% | 93,26% | 90,76% |
| CPFL PAULISTA | 94,63% | 87,96% | 91,93% | 90,70% | 93,20% | 92,00% | 90,80% |
| COELBA | 97,14% | 94,16% | 92,90% | 99,21% | 85,26% | 97,09% | 98,99% |
| CSPE | 100,00% | 95,09% | 95,28% | 96,53% | 97,33% | 98,41% | 97,17% |
| CELTINS | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| JAGUARI | 100,00% | 100,00% | 99,02% | 100,00% | 100,00% | 100,00% | 100,00% |
| COELCE | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| PIRATININGA | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| MUXFELDT | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| RGE | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
| NOVA PALMA | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |