LI Simpósio Brasileiro de Pesquisa Operacional
Limeira–SP, 02 a 06 de setembro de 2019

UMA ANÁLISE ROBUSTA DE BENCHMARKING UTILIZANDO O MÉTODO DE FRONTEIRA ESTOCÁSTICA BAYESIANO APLICADO ÀS EMPRESAS BRASILEIRAS DE DISTRIBUIÇÃO DE ENERGIA

Magno Silvério Campos
Departamento de Engenharia de Produção, Administração e Economia
Universidade Federal de Ouro Preto, Ouro Preto, Minas Gerais, Brasil
magno@ufop.edu.br

Marcelo Azevedo Costa
Departamento de Engenharia de Produção
Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil
macosta@ufmg.br

Ana Lúcia Miranda Lopes
Departamento de Ciências Administrativas
Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brasil
analopes@face.ufmg.br

RESUMO

Um modelo de *Análise Envoltória de Dados* (DEA) é aplicado pela ANEEL para definir os custos operacionais regulatórios para 61 concessionárias de distribuição de eletricidade ou DSOs. Resultados recentes mostram evidências de que tal modelo de *benchmarking* requer melhorias e carece de ajustes ambientais. Esse estudo avalia o uso da *Análise de Fronteira Estocástica* (SFA) como um modelo alternativo. Prós e contras são destacados. Os resultados mostram que o SFA é mais flexível para lidar com *outliers*. No entanto, o SFA tem grandes problemas de convergência se aplicado a amostras limitadas. Tais problemas podem ser superados utilizando métodos de computação Bayesiana ou de verossimilhança penalizada. Em particular, é proposto um modelo SFA Bayesiano que é robusto a problemas de convergência. Esse estudo defende o uso do DEA e do SFA como as melhores alternativas para definir os custos operacionais regulatórios para as DSOs brasileiras, conforme indicado pelos reguladores europeus.

PALAVRAS CHAVE. DEA, SFA, Análise Bayesiana.

Tópicos: PO na Área de Energia, Petróleo e Gás; Estatística e Modelos Probabilísticos

ABSTRACT

A *Data Envelopment Analysis* (DEA) model is applied by the ANEEL to set regulatory operational costs for 61 electricity distribution utilities or DSOs. Recent results show evidence that such a benchmarking model requires improvements and lacks environmental adjustments. This study evaluates the use of *Stochastic Frontier Analysis* (SFA) as an alternative model. Pros and cons are highlighted. Results show that the SFA is more flexible to deal with outliers. However, the SFA has major convergence problems if applied to limited samples. Such problems can be overcome using Bayesian computation or penalized likelihood methods. In particular, a Bayesian SFA model is proposed that is robust to convergence problems. This study advocates the use of both DEA and SFA as the best alternatives to set regulatory operational costs for Brazilian DSOs, as indicated by European regulators.

KEYWORDS. DEA, SFA, Bayesian Analysis.

Paper topics: OR in the Energy, Oil and Gas Area; Statistics and Probabilistic Models

https://proceedings.science/p/106652?lang=pt-br

1. Introdução

Os métodos de *benchmarking* fornecem aos analistas, em particular aos do setor de regulação de tarifas de eletricidade, ferramentas matemáticas e estatísticas para estimar o custo mínimo de produção, dada uma quantidade observada de produtos ou *outputs*. No caso das concessionárias distribuidoras (DSOs) ineficientes, modelos de *benchmarking* podem identificar as melhores práticas. Dentre estes métodos, os mais usuais, segundo a literarura, são os baseados em Análise Envoltória de Dados (DEA), em Mínimos Quadrados Corrigidos (COLS) e em Análise de Fronteira Estocástica (SFA). Além disso, [Agrell e Bogetoft, 2017] alegam que as aplicações mais proeminentes das técnicas de *benchmarking* estão relacionadas à regulação tarifária no setor elétrico. Não é à toa que a utilização de tais métodos vem crescendo significativamente nesse setor.

Especificamente, para a definição de custos regulatórios eficientes das DSOs brasileiras, a Agência Nacional de Energia Elétrica (ANEEL) se baseia na metodologia DEA, utilizando custos operacionais (OPEX) como entrada e número de consumidores, consumo ponderado de energia, extensão de rede de alta tensão, extensão de rede de baixa tensão, extensão de rede subterrânea, perdas não-técnicas e duração de interrupção de energia como variáveis de saída. As variáveis de entrada e saída estão disponíveis como valores médios de 2011 a 2013 e de 2014 a 2016 para 61 DSOs. Além disso, restrições aos pesos também são incluídas no modelo de programação linear. Utilizando o DEA, os custos operacionais de cada DSO são comparados com os custos e as saídas das DSOs restantes. Portanto, uma fronteira eficiente é estimada. O resultado final do DEA são os escores de eficiência, que indicam a capacidade de cada DSO transformar recursos (custo) em produtos (serviços prestados). As DSOs mais eficientes são nomeadas como *benchmarks*. As eficiências estimadas no último período variam de 27,8% a 100% e têm uma média de 70,6%. Mais detalhes sobre o modelo brasileiro de *benchmarking* DEA são encontrados em [ANEEL, 2015].

Algumas críticas podem ser dirigidas ao modelo DEA da ANEEL. Por exemplo, [Banker, 2011] questiona o uso de um modelo DEA - NDRS como um substituto para um modelo mais frequentemente utilizado, com retornos variáveis à escala. Em congruência, [Bogetoft, 2014], [Bogetoft e Lopes, 2015] e [Lopes et al., 2016] dissertam e sugerem que as medidas utilizadas pela ANEEL podem ser refinadas por meio de: inclusão de variáveis ambientais no modelo DEA, visando corrigir os escores de eficiência; não utilização de restrições aos pesos do modelo; remoção de variáveis com dados zerados; tratamento de *outliers*; estudo de consistência das variáveis; e, correção das distorções entre as premissas do DEA e o modelo implementado. [Gil et al., 2017] e [Veronese et al., 2019] apontam que as eficiências das DSOs brasileiras mudam significativamente se uma análise de segundo estágio for conduzida. Vale ressaltar também que, de acordo com [Tsionas, 2002], é inapropriada a utilização de um modelo de *benchmarking* isolado, sugerindo o emprego de múltiplos modelos para uma análise mais consistente.

Entretanto, mesmo com preciosas e fundadas sugestões ofertadas à ANEEL, a problemática da ausência de implementação de uma metodologia acurada ainda persiste. Com isso, a definição dos escores de eficiência das DSOs fica prejudicada, e consequentemente, tomadas de decisão se mantêm deficientes. Logo, existe uma lacuna a ser preenchida pela ANEEL: **qual metodologia utilizar e/ou ser desenvolvida, de maneira que se possa definir satisfatoriamente, uma fronteira de custos eficientes para as DSOs brasileiras?**

Também, vale a pena destacar que, nos modelos de *benchmarking* determinísticos, como é o caso do DEA utilizado pela ANEEL, os desvios em relação às fronteiras de custo são atribuídos única e exclusivamente à ineficiência técnica e/ou alocativa da DSO em questão. Porém, sabe-se que pertubações aleatórias que estão fora do controle da DSO, tais como intempéries climáticas, greves e crises políticas, têm influência significativa no cenário produtivo. Uma proposta alternativa consiste

na implementação de modelos de *benchmarking* estocásticos, tal como a *Análise de Fronteira Estocástica* (SFA), nos moldes daqueles introduzidos por [Meeusen e van den Broeck, 1977] e [Aigner et al., 1977], cujas vantagens residem justamente no fato destes reconhecerem que os desvios das fronteiras podem ser de ordem técnica, alocativa ou também aleatória.

De acordo com [Kumbhakar e Lovell, 2003], na reformulação com modelos de *benchmarking* estocáticos, as fronteiras de custo são estocásticas devido à variação aleatória no ambiente operacional. Os desvios dessas fronteiras estocásticas são unilaterais, devido a vários tipos de ineficiência. Assim, haverá componentes de erro simétrico, com a finalidade de capturar os efeitos da variação aleatória no ambiente operacional, e componentes de erro unilaterais, com a finalidade de modelar os efeitos das possíveis ineficiências técnicas e/ou alocativas. Entretanto, apesar de significativas contribuições, de acordo com [Arcoverde et al., 2005], os desdobramentos e as possibilidades ofertadas pelo SFA ainda são pouco explorados e pouco empregados no setor de regulação energética brasileira, quase sempre configurando-se em estudos pontuais, desprovidos de aprofundamentos e visão sistêmica. Tal fato também contribui para a motivação desse estudo.

Por fim, muito embora os incentivos à regulação no setor energético estejam bem disseminados no contexto global, por outro lado, de acordo com [Haney e Pollit, 2011, 2009], o mesmo não se pode dizer sobre as boas práticas na aplicação de métodos de *benchmarking*. Apenas alguns poucos países seguem os princípios dessas melhores práticas, e muitos parecem ser influenciados por práticas regionais, independentemente destas serem as melhores, ou se deparam com complicações na escolha do método mais adequado para o contexto em que estão inseridos. Prova disso pode ser encontrada no trabalho de [Haney e Pollit, 2011], que apresentam um índice capaz de traduzir o grau de utilização de boas práticas de *benchmarking* dos reguladores de 40 países amostrados. Desses, 30 países (75% da amostra) apresentam tal índice abaixo de 3, numa escala que varia de 0 a 8, em ordem crescente de boas práticas. O Brasil apresenta índice igual a 2,25. Também não há uma conclusão clara sobre qual método é melhor ou superior. Fato este que, segundo [Kuosmanen et al., 2013], levanta preocupações sobre a adequação do emprego de qualquer método único. Além disso, estes autores observam que, apesar dos escores de eficiência de diferentes métodos serem usualmente altamente correlacionados, as implicações econômicas dos resultados são substanciais.

Além dessa seção introdutória de contextualização do estudo, esse texto foi dividido em seções temáticas. São apresentados, nessa ordem: na Seção 2, conceituam-se e discutem-se as chamadas fronteiras de custo, as eficiências econômicas, apontando métodos de estimação frequentista e Bayesiana para a estimação da fronteira estocástica de custo, discutindo-se também a possibilidade da inserção de variáveis ambientais no SFA. A Seção 3 apresenta as parametrizações propostas. A Seção 4 traz a base de dados utilizada. Já a Seção 5 enumera os modelos propostos por esse trabalho. As discussões e comentários sobre os resultados estão presentes na Seção 6. Por fim, a Seção 7 traz as considerações finais do trabalho.

2. Análise de Fronteira Estocástica

As origens do SFA remetem aos trabalhos de [Meeusen e van den Broeck, 1977] e [Aigner et al., 1977], publicados com diferença de um mês entre si. Pouco tempo depois, também surgiu um terceiro artigo, publicado por [Battese e Corra, 1977], autores estes que foram revisores dos dois trabalhos anteriores. Estes três modelos de SFA originais, segundo [Kumbhakar e Lovell, 2003], compartilham uma estrutura de erros aleatórios e erros causados por ineficiências. Sendo assim, as subseções que se seguem, trazem considerações gerais, julgadas importantes, para um melhor entendimento do método de SFA e futuro embasamento teórico para as proposições que advirem.

2.1. Fronteira de Custo

A fronteira de custo indica o mínimo custo necessário para produzir um conjunto de *outputs*, a partir do preço dos *inputs* e da tecnologia de produção empregada. Também funciona

como *benchmark*, contra a qual podem ser comparados os desempenhos de todos os produtores. Matematicamente, a função fronteira de custo pode ser expressa como:

$$C(\boldsymbol{w}, y) = \boldsymbol{wx} \cdot e^{u}, \quad \text{com} \quad u \geq 0. \tag{1}$$

Nessa expressão, $\boldsymbol{w}$ é o vetor de preços dos *inputs* $\boldsymbol{x}$, y é o nível observado de *output* e u representa uma eventual ineficiência do produtor. Observa-se que, a função de custo mede o custo de produzir y quando os *inputs* são ajustados pela sua eficiência. Assim, na presença de ineficiências ($u > 0$), o custo atual (*expenditure*), $E = \boldsymbol{wx}e^{u}$, pode ser maior que o custo mínimo, $\boldsymbol{wx}$. Também, pode-se escrever a seguinte expressão para o cálculo da eficiência econômica do i-ésimo produtor:

$$EE_i = \frac{Custo^{min}}{Custo_i^{atual}} = \frac{Custo^{min}}{E_i} = \frac{\boldsymbol{wx}}{\boldsymbol{wx}e^{u_i}} = e^{-u_i} \quad \forall \quad \boldsymbol{x} \in \Re_+^n \quad \text{e} \quad y \in \Re_+. \tag{2}$$

Existem na literatura, algumas formas funcionais para a função de custo $C(\boldsymbol{w}, y)$. Dentre elas, destacam-se a função de custo *Cobb-Douglas* [Cobb e Douglas, 1928] e a função de custo *Translog* [Christensen et al., 1971].

2.2. Modelo de Custo com Fronteira Estocástica

De acordo com [Kumbhakar et al., 2015], na equação (1), pode-se: associar um vetor de parâmetros J-dimensional $\boldsymbol{\beta}$ que representa os parâmetros da tecnologia de produção; acrescentar um termo de erro aleatório e^{ν_i}, irrestrito em sinal e com média zero; escrever $\epsilon_i = \nu_i + u_i$; e, admitindo que a função de custo seja linear nos logaritmos dos parâmetros $\boldsymbol{\beta}$, chega-se ao seguinte resultado:

$$\begin{aligned} \ln E_i &= \ln C(\boldsymbol{w}_i, y_i, \boldsymbol{\beta}) + \epsilon_i, \\ \epsilon_i &= \nu_i + u_i, \quad \text{com} \quad -\infty < \nu_i < +\infty \quad \text{e} \quad u_i \geq 0. \end{aligned} \tag{3}$$

Esse resultado é conhecido na literatura como *Modelo de Custo com Fronteira Estocástica*. Nele, destacam-se: (*i*) - o termo y_i, que reprenta o nível de produção observado para o produtor i; (*ii*) - a função $C(\boldsymbol{w}_i, y_i, \boldsymbol{\beta})$, que representa a fronteira determinística de custo, comum a todos produtores, e define o mínimo custo de produção possível através da tecnologia em uso; e, (*iii*) - o termo ϵ_i, que representa o *erro composto*. Esse termo caracteriza conjuntamente, o termo ν_i, cuja finalidade é capturar o efeito dos choques aleatórios que afetam especificamente o i-ésimo produtor, e o termo u_i, que descreve a ineficiência econômica apresentada por este. Além disso, pode-se escrever a seguinte equação:

$$e^{-u_i} = \frac{C(\boldsymbol{w}_i, y_i, \boldsymbol{\beta}) \cdot e^{\nu_i}}{E_i}. \tag{4}$$

Assim sendo, como a equação (4) traz a razão entre o custo mínimo $C(\boldsymbol{w}_i, y_i, \boldsymbol{\beta}) \cdot e^{\nu_i}$ possível e o custo de produção observado, deduz-se que a eficiência econômica do i-ésimo produtor pode ser mensurada através da expressão e^{-u_i}.

https://proceedings.science/p/106652?lang=pt-br

2.2.1. Enfoque Frequentista para a Estimação de e^{-u_i}

Os métodos de estimação paramétricos objetivam estimar os parâmetros do modelo dado em (3), partindo da suposição de comportamentos probabilísticos para os erros ν_i e u_i, culminando na obtenção dos respectivos estimadores, como por exemplo, os de máxima verossimilhança. Logo, as suposições sobre esses comportamentos probabilísticos constituem o cerne desse processo de estimação. Por exemplo, a função densidade de probabilidade do erro composto, $f(\epsilon)$, é dada por:

$$f(\epsilon) = \frac{2}{\sigma}\phi\left(\frac{\epsilon}{\sigma}\right)\Phi\left(\lambda\frac{\epsilon}{\sigma}\right), \quad \text{para} \quad -\infty < \epsilon < \infty, \tag{5}$$

onde $\sigma = \sqrt{\sigma_u^2 + \sigma_\nu^2}$ é conhecido como parâmetro de escala, $\lambda = \frac{\sigma_u}{\sigma_\nu}$ representa o parâmetro de forma, $\phi(\cdot)$ é a função densidade de probabilidades normal padrão e, $\Phi(\cdot)$ é a função de distribuição de probabilidades acumuladas normal padrão. A função $f(\epsilon)$ é assimétrica, sendo um caso particular da distribuição *Skew*-Normal [Azzalini e Valle, 2013]. Porém, apesar das boas propriedades dessa distribuição, [Sartori, 2006] e [Azzalini e Valle, 2013] chamam a atenção para o fato de a estimativa de máxima verossimilhança para o parâmetro λ poder se tornar infinita, com probabilidade positiva. Se $\lambda \longrightarrow \infty$, então u domina o erro composto e as ineficiências estimadas tendem a 100%. [Veronese et al., 2019] identificou esse problema utilizando o banco de dados da ANEEL.

2.2.2. Enfoque Bayesiano para a Estimação de e^{-u_i}

As metodologias mais usuais de *benchmarking* (DEA, COLS e SFA) apresentam dificuldades quando lidam com amostras consideradas pequenas, ou quando *outliers* se fazem presentes na base de dados, ou ainda quando os produtores sob análise não operam sob as mesmas condições tecnológicas de produção e de custo. Especificamente no caso do SFA, ainda existem os já relatados problemas de convergência de estimação. Frente a esses agravantes, uma metodologia baseada no método Bayesiano vem se mostrando eficaz para a estimação de fronteiras de eficiência, sem incorrer nos problemas das outras técnicas. Trata-se do *Bayesian Stochastic Frontier Analysis* (BSFA), cujo trabalho seminal foi proposto por [van den Broeck et al., 1994].

A formulação se inicia com a especificação das distribuições de probabilidade *a priori* para os parâmetros do modelo dado em (3). Em seguida, escreve-se a distribuição de probabilidades *a posteriori* para u_i, isto é, $f(u_i|E_i, \boldsymbol{\psi})$, sendo $\boldsymbol{\psi}$, o vetor de parâmetros do modelo dado em (3). Assim, como a distribuição condicional *a posteriori* para u_i é agora conhecida, pode-se utilizar o Amostrador de Gibbs [Geman e Geman, 1984] para gerar observações de u_i. Logo, as (in)eficiências econômicas dos produtores podem ser assim estimadas.

De acordo com [Dorfman e Koop, 2005], os métodos econométricos que se baseiam em técnicas Bayesianas vêm sendo utilizados em maior escala ultimamente, em modelos que possuem vetores de grande dimensão, graças aos avanços do poder computacional nas últimas décadas e ao desenvolvimento de simuladores que aproveitam esse poder, como é o caso do *Monte Carlo via Cadeias de Markov* (MCMC) [Metropolis et al., 1953]. E devido ao fato dos modelos de SFA, muito comumente, se enquadrarem nesse contexto de grandes dimensões, há um aumento da motivação e uma crescente popularidade dos métodos Bayesianos na análise eficiência.

Apesar dessa recente popularidade, segundo [Souza et al., 2010], são poucos os estudos com enfoque Bayesiano aplicados ao setor de regulação energética brasileira.

2.3. Incorporação de Variáveis Ambientais ao SFA

Os escores de eficiência dos produtores podem ser atualizados por variáveis ambientais, uma vez que, no geral, produtores localizados em um ambiente de variáveis ambientais favoráveis devem ter seus escores de eficiência diminuídos, pois o ambiente contribui parcialmente para um maior escore de eficiência. Enquanto que aqueles produtores, localizados em um ambiente de

variáveis ambientais não tão favoráveis, devem ter seus escores de eficiência aumentados, pois esse ambiente desfavorável impede que os produtores obtenham maiores escores de eficiência. Especificamente, no caso da regulação tarifária energética brasileira, estudos recentes de [Veronese et al., 2019] e [Gil et al., 2017] demonstram a significância de variáveis ambientais para a estimação dos escores de eficiência, via DEA, das várias concessionárias de distribuição, espalhadas pelo país.

3. Parametrizações Propostas para o SFA

[Bayes e Branco, 2007], baseando-se em [Henze, 1986], propuseram uma nova reparametrização para uma variável aleatória *Skew*-Normal padrão:

$$Z = \sqrt{(1-\delta^2)}V + \delta U, \tag{6}$$

onde $V \underset{iid}{\sim} N(0,1^2)$, $U \underset{iid}{\sim} N^+(0,1^2)$ e $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. Além disso, se $\lambda > 0$ para o modelo SFA de custo, $\delta \in]0{,}1[$. Vale ressaltar aqui que, esse é um diferencial desse trabalho, uma vez que pelo princípio de Bayes-Laplace, uma escolha natural de *priori* não-informativa para δ seria uma distribuição uniforme, isto é, $\delta \sim U(-1,1)$. Como no SFA de custo, $u_i \geq 0$ implica em $\lambda > 0$, então $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in]0{,}1[$, levando a $\delta \sim U(0,1)$.

Além disso, como $\nu_i + u_i = \sigma z_i$, pode-se escrever que $\nu_i + u_i = \sigma\left(\sqrt{1-\delta^2}V_i + \delta U_i\right)$ ou $\nu_i = \sigma\sqrt{1-\delta^2}V_i$ e $u_i = \sigma\delta U_i$. Observa-se que $u_i \underset{iid}{\sim} N^+(0,\sigma_u^2)$ mede a ineficiência econômica da i-ésima DSO. O termo $\nu_i \underset{iid}{\sim} N(0,\sigma_\nu^2)$ captura a aleatoriedade sobre essa DSO. Essas manipulações oferecem maneiras alternativas de se escrever u_i e ν_i. Agora, tem-se:

$$X_i = \xi(\mathbf{y}_i) + \sigma\sqrt{1-\delta^2}V_i + \sigma\delta U_i, \tag{7}$$

com $U_i \underset{iid}{\sim} N^+(0{,}1^2)$, $V_i \underset{iid}{\sim} N(0,1^2)$, $\sigma = \sqrt{\sigma_\nu^2 + \sigma_u^2}$ e $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

Utilizando a estrutura Bayesiana e a representação estocástica para a distribuição *Skew*-Normal proposta por [Henze, 1986], o modelo SFA de custo proposto, pode ser escrito como:

$$\begin{aligned} &X_i|U_i,\xi(\mathbf{y}_i),\sigma,\delta_{(\lambda)} \underset{iid}{\sim} N\left[\xi(\mathbf{y}_i) + \sigma\delta U_i;\ \sigma^2(1-\delta^2)\right], \\ &U_i \underset{iid}{\sim} N^+\left(0,1^2\right), \quad i = 1,...,n. \end{aligned} \tag{8}$$

Seguindo [Gil et al., 2017] e [Bayes e Branco, 2007], as seguintes especificações *a priori* fracamente informativas foram escolhidas para os parâmetros do SFA:

$$\frac{1}{\sigma^2} \sim \Gamma(0{,}01;0{,}01), \quad \delta \sim U(0,1) \ \ e \ \ \beta_j \underset{iid}{\sim} N(0;100^2)$$

O modelo descrito na equação (8) pode ser implementado usando *softwares* que utilizam *Markov Chain Monte Carlo* (MCMC), como por exemplo, o JAGS [Plumer, 2003]. Esses softwares retornam os valores de u_i, σ e λ, utilizando estimação pelo contexto Bayesiano, sem precisar recorrer a métodos de verossimilhança, escapando assim das problemáticas de estimação.

4. Base de Dados

Um resumo das variáveis de entrada e saída disponíveis para as 61 DSOs brasileiras é apresentado na Tabela 1:

Tabela 1: Variáveis de entrada e saída disponíveis para as 61 DSOs brasileiras.

Variável	Sigla	Notação	Espécie	Unidade	*Proxy*
PMSO ajustado (OPEX)	PMSOa	x	Input	R$	Pessoal, material, serviços e outros
Número de consumidores	Cons	y_1	Outputs ordinários	un	Quantidade de serviços prestados
Mercado ponderado	Wmkt	y_2		MWh	Produção total
Rede aérea	Over	y_3		Km	Disseminação dos consumidores dentro da área de concessão
Rede de alta tensão	High	y_4			
Rede subterrânea	Under	y_5			
Consumidor-hora de energia interrompida	CHI	y_6	Outputs indesejados	h	Qualidade dos serviços prestados
Perdas não-técnicas	NTL	y_7		MWh	
Variável ambiental	*e.variable*	z	Variável exógena	$*$	Informações ambientais

Vale ressaltar que a decisão quanto ao modelo recai sobre o regulador. Portanto, com base no atual modelo DEA-NDRS, este trabalho propõe modelos de *benchmarking* SFA utilizando as mesmas entradas e saídas. Imita-se o uso de *outputs* negativos no SFA, utilizando distribuições *a priori Half*-Normal negativas para os coeficientes das variáveis de qualidade. No entanto, propõe-se a agregação das *perdas não-técnicas* (variável de qualidade) ao *mercado ponderado*. As perdas não-técnicas são estimadas apenas para o mercado de baixa tensão, ou seja, principalmente para consumidores residenciais. Portanto, essas perdas podem ser deduzidas do mercado de baixa tensão antes de se calcular o *mercado ponderado*. Assim, apenas uma variável de qualidade, *consumidor-hora de energia interrompida*, é avaliada. Além disso, [Gil et al., 2017] e [Veronese et al., 2019] investigaram variáveis ambientais usando o DEA de segundo estágio e concluíram que o índice ambiental (*e.variable*) é uma *proxy* razoável para informações ambientais.

5. Propostas de Modelos para o SFA

São propostos sete modelos BSFA de custo, conforme descritos a seguir.

5.1. Modelo 1

Esse modelo utiliza uma função translog, sendo o OPEX a variável de entrada e número de consumidores, mercado ponderado (variável original), rede aérea, rede de alta tensão e rede subterrânea são as variáveis de saída. Esse modelo não inclui variáveis de qualidade ou ambientais.

5.2. Modelo 2

Esse modelo de BSFA utiliza uma função translog, com as mesmas variáveis do Modelo 1. No entanto, o mercado ponderado é recalculado deduzindo-se as perdas técnicas. Além disso, o consumidor-hora de energia interrompida é utilizado como a variável de qualidade e, a *e.variable* é utilizada como a variável ambiental. A seguinte estrutura de função de custo é usada:

$$\ln x_i = \ln C(\boldsymbol{y}_i; \boldsymbol{\beta}) + g(\boldsymbol{z}_i; \boldsymbol{\delta}) + \nu_i + u_i \tag{9}$$

onde $\ln C(\boldsymbol{y}_i; \boldsymbol{\beta})$ é a função de custo translog e $g(\boldsymbol{z}_i; \boldsymbol{\delta}) = \delta_0 + \delta_1 z_i$ é a componente ambiental.

https://proceedings.science/p/106652?lang=pt-br

5.3. Modelo 3

Esse é um modelo de BSFA híbrido que utiliza uma estrutura translog para número de consumidores e mercado ponderado, juntamente com uma estrutura Cobb-Douglas para rede aérea, rede de alta tensão e rede subterrânea. Além disso, o consumidor-hora de energia interrompida é utilizado como variável de qualidade e a *e.variable* é utilizada como variável ambiental.

5.4. Modelo 4

Trata-se de um modelo BSFA híbrido que utiliza uma estrutura translog para número de consumidores, mercado ponderado e rede aérea, juntamente com uma estrutura Cobb-Douglas para rede de alta tensão e rede subterrânea. Além disso, o consumidor-hora de energia interrompida é utilizado como variável de qualidade e a *e.variable* é utilizada como variável ambiental.

5.5. Modelo 5

Esse é um modelo semelhante ao Modelo 2, porém utiliza uma distribuição de probabilidades *t-Student* para ν_i. Como indicado por [Juárez e Steel, 2010], é utilizada a seguinte distribuição *a priori* Gama para o parâmetro grau de liberdade a da distribuição *t-Student*: $a \sim \Gamma(2; 0{,}1)$.

5.6. Modelo 6

Trata-se de um modelo híbrido de BSFA, similar ao Modelo 3, porém utiliza uma distribuição de probabilidades *t-Student* para o componente de ruído e a distribuição *a priori* Gama para o parâmetro grau de liberdade proposta no modelo 5.

5.7. Modelo 7

Esse é um modelo híbrido de BSFA, similar ao Modelo 4, porém utiliza uma distribuição de probabilidades *t-Student* para o componente de ruído e a distribuição *a priori* Gama para o parâmetro grau de liberdade proposta no modelo 5.

5.8. Implementação Computacional

As parametrizações e as modelagens propostas foram implementadas a partir do *software* livre JAGS (*Just Another Gibbs Sampler*) [Plumer, 2003], em conjunto com os pacotes computacionais *rjags* [Plummer, 2016] e *mcmc* [Geyer e Johnson, 2017].

Em todos os modelos BSFA propostos, inicialmente foram realizadas 40 milhões de iterações (*burn-in*). Em seguida, executaram-se outras 40 milhões de iterações, com uma defasagem (*lag*) de 20 mil iterações entre as sucessivas amostras geradas para os parâmetros do modelo. Todos os modelos apresentaram convergência de resultados.

5.9. Critérios de Seleção de Modelos

Os modelos BSFA propostos foram comparados utilizando-se as estatísticas *Watanabe-Akaike Information Criterion* (WAIC), *Deviance Information Criterion* (DIC) e *Log-Pseudo Marginal Likelihood* (LPML). De acordo com [Gil et al., 2017], os modelos com menor valor de DIC e maiores valores de WAIC e LPML alcançam melhor ajuste aos dados. Os resultados computacionais encontrados são apresentados na próxima seção.

6. Resultados e Discussões

A Figura 1 compara as eficiências estimadas pelo DEA e pelos SFAs propostos. Os resultados mostram que as eficiências estimadas do SFA são maiores do que as eficiências do DEA. As eficiências do DEA têm média de $70{,}87\%$ e valor mínimo de $27{,}90\%$, enquanto as eficiências de SFA têm média de $86{,}87\%$ e valor mínimo de $66{,}34\%$. Menores eficiências de DEA foram fortemente criticadas. Além disso, no último Ciclo de Revisão Tarifária Periódica (CRTP), o regulador adotou uma redução máxima de custo de 5% ao ano, o que indica uma eficiência de custo mínima de 20% em cada CRTP. Portanto, os resultados do SFA estimados apoiam a decisão do regulador brasileiro.

https://proceedings.science/p/106652?lang=pt-br

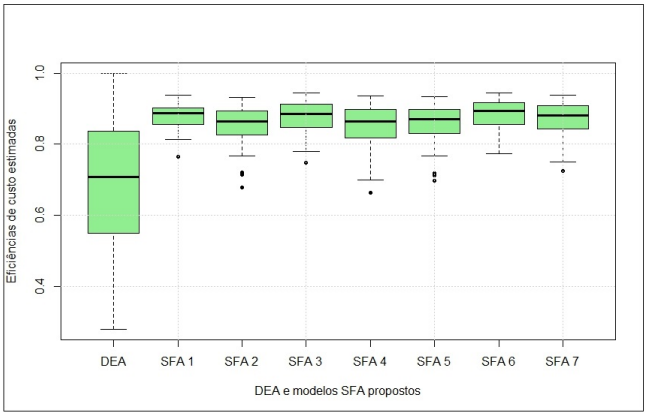
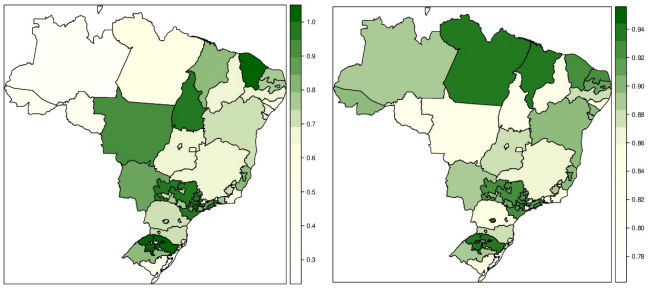


Figura 1: Boxplots comparando eficiências originais de DEA e eficiências propostas de SFA.

A Figura 2 apresenta a distribuição das eficiências DEA e SFA (Modelo 6 com melhores estatísticas) em todo o território brasileiro. Esquemas de cores foram ajustados para a faixa de eficiência DEA e SFA para minimizar as diferenças de escala. Os resultados mostram padrões semelhantes entre as eficiências DEA e SFA nas regiões sul e sudeste. Além disso, as eficiências de SFA são maiores no norte, ao contrário das eficiências de DEA que são extremamente mais baixas nessa região. Isso ocorre porque o modelo SFA inclui ajustes ambientais. Veronese et al. [2019] e Gil et al. [2017] também concluíram que a região norte do Brasil enfrenta condições ambientais adversas que exigem atualizações de eficiência de custos. Diferenças significativas entre DEA e SFA são encontradas nos estados de Mato Grosso e Tocantins. Ao contrário do DEA, as eficiências de SFA para as DSOs localizadas nesses estados foram menores em comparação com as eficiências de SFA em todo o território brasileiro. Em geral, apesar das diferenças de escala mostradas na Figura 2, tanto o DEA quanto o SFA proposto apresentaram padrões similares de maior e menor eficiência na maioria das DSOs brasileiros. Mudanças específicas são principalmente devidas a ajustes ambientais.



(a) Eficiências estimadas DEA (b) Eficiências estimadas SFA 6.

Figura 2: Eficiências estimadas DEA e SFA em todo o território brasileiro.

A Figura 3 também ilustra semelhanças e diferenças entre eficiências de DEA e SFA após ajustes de escala. As DSOs foram classificados em ordem crescente de dissimilaridade. Dois eixos y diferentes são mostrados. Os valores de eficiência de SFA estão localizados no eixo y esquerdo, enquanto os valores de eficiência DEA estão localizados no eixo y direito. Pode-se concluir que, após o escalonamento, tanto a eficiência DEA quanto a SFA são similares a um grande grupo de DSOs. Como mencionado anteriormente, grandes dissimilaridades estão associadas a algumas DSOs localizadas na região norte.

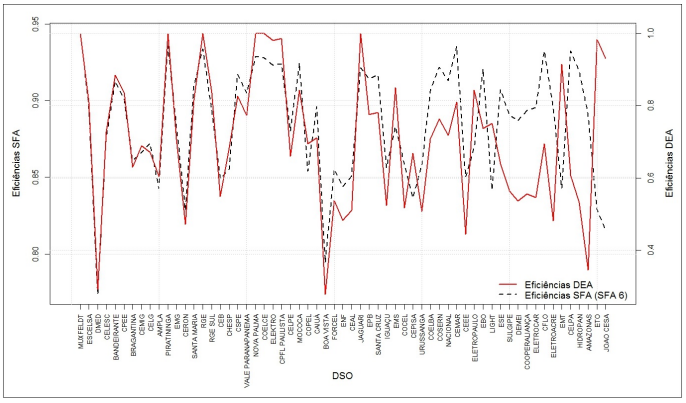


Figura 3: Similaridades e dissimilaridades entre DEA e SFA após ajustes de escala.

7. Conclusão

O objetivo desse estudo foi duplo: primeiro replicar o mais próximo possível, as eficiências estimadas para as DSOs brasileiras, utilizando uma abordagem robusta de SFA, ou seja, utilizando as mesmas entradas e saídas do modelo empregado atualmente pelo regulador. Segundo, propor uma abordagem alternativa e robusta de SFA utilizando a informação disponível, incluindo a componente ambiental. Uma correlação linear máxima de $\rho = 0{,}6531$ foi obtida usando um modelo SFA com uma função de custo translog. No entanto, o modelo original de DEA inclui restrições adicionais de peso que ainda exigem estudos adicionais para desenvolver uma proposta de SFA semelhante. Esse trabalho propôs, com sucesso, um método de *benchmarking* de SFA para estimar os custos eficientes das DSOs brasileiras.

Os resultados mostram uma grande diferença de escala entre as eficiências DEA e SFA. A eficiência média do DEA é de 70,5%, enquanto a eficiência média do SFA é de 88,5%. Essa diferença representa uma receita total de R$ 3.175.509.926,81 ou U$ 971,1 milhões (considerando uma taxa de câmbio média de R$ 3,27 por U$ 1,00 no ano de 2016), que atualmente é contabilizada como ineficiência. Como mencionado anteriormente, isso ocorre porque o SFA decompõe a distância da fronteira de custo em duas componentes, ruído e ineficiência econômica, enquanto a DEA assume apenas a ineficiência econômica.

Vale ressaltar que, embora as pontuações originais do DEA variem de 28 a 100%, o regulador aplicou *ad-hoc* restrições adicionais para forçar as eficiências no intervalo de 80 a 100%. A ANEEL aplicou um procedimento de padronização simples. Usando o SFA ou o máximo de SFA e DEA como a nova estimativa de eficiência, as eficiências estarão aproximadamente dentro do intervalo de 80 a 100%, como esperado pelo regulador.

Até onde se sabe, este é o primeiro trabalho que apresenta com sucesso um modelo robusto de SFA, utilizando técnicas Bayesianas, para as DSOs brasileiras aplicando as mesmas entradas, saídas e incluindo informações ambientais.

Apesar das significativas diferenças metodológicas entre o DEA e o SFA, independentemente de qual será escolhido pelo regulador, é interessante avaliar os dois, pois um complementa o outro e permite uma maior robustez dos resultados. Logo, como resposta para a questão-problema apresentada nesse trabalho, pode-se dizer que DEA e SFA devem ser vistas como metodologias não-excludentes e complementares, que podem ser utilizadas e desenvolvidas pelo regulador brasileiro, visando definir satisfatoriamente, uma fronteira de custos eficientes para as DSOs brasileiras.

https://proceedings.science/p/106652?lang=pt-br

Referências

Agrell, P. e Bogetoft, P. (2017). Regulatory benchmarking: models, analyses and applications. *Data Envelopment Analysis Journal*, 3:49–91.

Aigner, D., Lovell, C., e Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Economics*, 6(1):21–37.

ANEEL (2015). NT 66/2015: metodologia de custos operacionais. Technical note, SRM-SGT-ANEEL.

Arcoverde, F., Tannuri-Pianto, M., e Sousa, M. (2005). Mensuração das eficiências das distribuidoras do setor energético brasileiro usando fronteiras estocásticas. *Anais do Encontro Nacional de Economia*, XXXIII(1):1–18.

Azzalini, A. e Valle, R. (2013). Maximum penalized likelihood estimation for skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, 143(2):419–433.

Banker, R. (2011). Report on aneels proposal for electricity distribution tariff regulation. Technical note, Technical Note, 101/2011.

Battese, G. e Corra, G. (1977). Estimation of a production frontier model: with aplication to the pastoral zone of eastern Australia. *Australian Journal of Agricultural Economics*, 21(3):169–179.

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

Bogetoft, P. (2014). Comments on the brazilian benchmarking model for energy distribution regulation. forth cycle of tariff review, NT 192 2014. Technical report.

Bogetoft, P. e Lopes, A. L. M. (2015). Comments on the brazilian benchmarking model for energy distribution regulation forth cycle tariff review technical note 407/2014. Technical report.

Christensen, L., Jorgenson, R., e Lau, L. (1971). Conjugate duality and transcendental logarithmic function. *Econometrica*, 39(1):255–256.

Cobb, C. e Douglas, P. (1928). A theory of production. *The American Economic Review*, 18(1): 139–165.

Dorfman, J. e Koop, G. (2005). Current developments in productivity and efficiency measurement. *Journal of Econometrics*, 126:233–240.

Geman, S. e Geman, D. (1984). Stochastic relaxation, gibbs distributions and bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6:721—-741.

Geyer, C. J. e Johnson, L. T. *Markov Chain Monte Carlo*, 2017. URL `https://CRAN.R-project.org/package=mcmc`. R package version 0.9-5.

Gil, G., Costa, M., Lopes, A., e Mayrink, V. (2017). Spatial statistical methods applied to the 2015 brazilian energy distribution benchmarking model: accountig for unobserved determinants of inneficieies. *Energy Economics*, 64(1):373–383.

Haney, A. e Pollit, M. (2009). Efficiency analysis of energy networks: an international survey of regulators. *Energy Policy*, 37(12):5814–5830.

Haney, A. e Pollit, M. (2011). Exploring the determinants of best practice benchmarking in electricity network regulation. *Energy Policy*, 39:7739–7746.

Henze, N. (1986). A probabilistic representation of the skew-normal distribution. *Scandinavian journal of statistics*, p. 271–275.

Juárez, M. A. e Steel, M. F. (2010). Model-based clustering of non-Gaussian panel data based on skew-t distributions. *Journal of Business & Economic Statistics*, 28(1):52–66.

Kumbhakar, S. e Lovell, C. (2003). *Stochastic Frontier Analysis*. New York: Cambridge University Press.

Kumbhakar, S., Wang, H., e Horncastle, A. (2015). *Stochastic Frontier Analysis Using Stata*. New York: Cambridge University Press.

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

Lopes, A., Vilela, B., Costa, M., e Lanzer, E. (2016). Critical evaluation of the efficient costs assessment model used in the regulation of brazilian energy distribution service operator. *Revista Gestão e Tecnologia*, 16(3):5–30.

Meeusen, W. e van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2):435–444.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., e Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Plumer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In K. Hornik, F. L. e Zeileis, A., editors, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, p. 1–10.

Plummer, M. *Bayesian Graphical Models using MCMC*, 2016. URL `https://CRAN.R-project.org/package=rjags`. R package version 4-6.

Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew-normal and skew-t distributions. *Journal of Statistical Planning and Inference*, 136(12):4259–4275.

Souza, M., Souza, R., e Pessanha, J. (2010). Custos operacionais eficientes das distribuidoras de energia elétrica: um estudo comparativo dos modelos dea e sfa. *Gestão e Produção*, 17:653–667.

Tsionas, E. G. (2002). Stochastic frontier models with random coefficients. *Journal of Applied Econometrics*, 17(2):127–147.

van den Broeck, J., Koop, G., Osiewalski, J., e Steel, M. (1994). Stochastic frontier models: a baysesian perspective. *Journal of Econometrics*, 61:273–303.

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

https://proceedings.science/p/106652?lang=pt-br