

# Seleção de variáveis em DEA aplicada a uma análise do mercado de energia eléctrica

Lidia Angulo Meza †  
João Carlos Correia Baptista Soares de Mello †  
Eliane Gonçalves Gomes ‡  
Artur José Silva Fernandes †

† Universidade Federal Fluminense  
Brasil  
lidia@metal.eeimvr.uff.br  
jcsmello@producao.uff.br  
<http://www.uff.br/decisao>  
arth@predial.net.com.br

‡ Empresa Brasileira de Pesquisa Agropecuária  
Brasil  
eliane.gomes@embrapa.br

---

## Abstract

In this paper we have two main goals. From a theoretical point of view we intend to develop an improvement to the DEA variable selection models. It is in the variable normalisation step that this improvement takes place. From a practical point of view we intend to measure an energetic efficiency of the 27 Brazilian estates. A state achieves this efficiency when it has high values of some socio-economic indexes and low level of energy consumption. We used the non-parametrical approach of Data Envelopment Analysis (CCR and BCC models). We also used the inverted frontier DEA model.

## Resumo

Este artigo tem duplo objectivo: Do ponto de vista teórico pretende-se mostrar um aperfeiçoamento aos métodos de seleção de variáveis em modelos DEA. Este ocorre na etapa de normalização das variáveis. Do ponto de vista da aplicação pretende-se medir uma eficiência energética dos 27 estados brasileiros. Esta eficiência é obtida quando os estados conseguem altos valores nos indicadores socioeconómicos com baixo consumo de energia. São usados os modelos DEA clássicos CCR e BCC e os modelos de fronteira invertida.

**Keywords:** Data Envelopment Analysis; Variables selection; Electrical Sector

**Title:** Variables selection in DEA applied to an analysis of the electric energy market

---

## 1 Introdução

Uma revisão não extensa da literatura mostra que existem estudos para prever e explicar o consumo de energia eléctrica. Entretanto, as análises quantitativas para explicar o comportamento do mercado nem sempre são adequadas. Por exemplo, Lins et al. (2002) propuseram o uso da metodologia de análise condicional de demanda para estimar o consumo de energia eléctrica do sector residencial brasileiro, em função dos aparelhos usados. Estudos como este obrigam ao conhecimento sobre os hábitos dos consumidores, como as quantidades e os tipos de aparelhos eléctricos usados nas residências. Essa abordagem necessita de um grande trabalho de campo, como aquele feito pela fonte dos dados do artigo citado, qual seja, uma pesquisa realizada pelo Programa Nacional de Conservação de Energia Eléctrica. Pesquisas deste tipo nem sempre são viáveis.

O objectivo prático deste estudo é obter uma forma não paramétrica de identificar os estados com maior eficiência energética. Isto significa obter bons resultados socioeconómicos com baixo consumo de energia eléctrica. Os indicadores socioeconómicos usados são Índice de Desenvolvimento Humano – IDH, Índice de Potencial de Consumo – IPC, Produto Interno Bruto per capita – PIB per capita. Como estes indicadores não são totalmente independentes, pretende-se usar um método de selecção de variáveis para determinar os mais adequados a entrar no modelo. O método a ser usado deverá ser quantitativo e objectivo, isto é, independente de opiniões do decisor. Uma revisão de métodos de selecção de variáveis pode ser vista em Senra (2004) e Senra et al. (2007). Neste artigo será proposta uma evolução do ‘Método Multicritério Combinatório por Cenários’ (Senra et al., 2007).

Para a construção da fronteira de eficiência, duas ópticas podem ser consideradas. A primeira julga como eficiente um estado onde, mesmo com baixos valores dos índices socioeconómicos, a empresa consegue vender uma grande quantidade de energia eléctrica. Esta é considerada a óptica da empresa (Lins et al., 2007; Senra et al., 2007).

A segunda óptica considera eficiente aquele estado que, mesmo com pequeno consumo de energia eléctrica, consegue altos índices socioeconómicos. Esta é a óptica da eficiência energética adoptada, entre outros, por Angulo Meza et al. (2006), Chauhan et al. (2006), Hu e Kao (2007). No trabalho de Utlu e Hepbasli (2007) é realizada uma análise das variáveis no ponto de vista técnico, em que, mais uma vez, a energia consumida deve ser a menor possível. Além disso, pode-se encontrar também uma ampla revisão bibliográfica de todos os enfoques empregados na avaliação e análise da eficiência no uso da energia.

A óptica da eficiência energética pode ser interpretada do ponto de vista ambiental: deseja-se obter os melhores índices de desenvolvimento com o menor gasto possível de energia e, como consequência, com a menor agressão possível ao meio ambiente.

Devido ao objectivo proposto neste artigo, a óptica aqui usada é a da eficiência energética.

## 2 Análise Envoltória de Dados

### 2.1 Generalidades

A abordagem por Análise Envoltória de Dados (*Data Envelopment Analysis* – DEA) foi desenvolvida por Charnes et al. (1978) e usa a programação linear para avaliação de medidas de eficiência comparativas de Unidades de Tomada de Decisão (*Decision Making*

Units – DMUs) que utilizam os mesmos recursos (*inputs*) e geram os mesmos produtos (*outputs*).

Há dois modelos DEA clássicos: CCR e BCC. O modelo CCR (também conhecido por CRS ou *constant returns to scale*), adota como hipótese retornos constantes de escala (Charnes et al., 1978). Em sua formulação matemática considera-se que cada DMU  $k$  ( $k = 1 \dots s$ ) é uma unidade de produção que utiliza  $n$  *inputs*  $x_{ik}$ ,  $i = 1 \dots n$ , para produzir  $m$  *outputs*  $y_{jk}$ ,  $j = 1 \dots m$ .

O modelo DEA CCR maximiza o quociente entre a combinação linear dos *outputs* e a combinação linear dos *inputs*, com a restrição de que para qualquer DMU esse quociente não pode ser maior que 1. Esse problema de programação fracionária, mediante alguns artificios matemáticos, pode ser linearizado e transformado no Problema de Programação Linear (PPL) apresentado em (1), no qual  $h_o$  é a eficiência da DMU  $o$  em análise;  $x_{io}$  e  $y_{jo}$  são os *inputs* e *outputs* da DMU  $o$ ;  $v_i$  e  $u_j$  são os pesos calculados pelo modelo para *inputs* e *outputs*, respectivamente.

$$\begin{aligned} \max h_o &= \sum_{j=1}^m u_j y_{jo} \\ \text{sujeito a} \\ \sum_{i=1}^n v_i x_{io} &= 1 \\ \sum_{j=1}^m u_j y_{jk} - \sum_{i=1}^n v_i x_{ik} &\leq 0, \quad k = 1, \dots, s \\ u_j, v_i &\geq 0 \quad \forall i, j \end{aligned} \tag{1}$$

O modelo BCC (Banker et al., 1984), também chamado de VRS (*variable returns to scale*), considera situações de eficiência de produção com variação de escala e não assume proporcionalidade entre *inputs* e *outputs*. A formulação do modelo BCC usa para cada DMU o problema de programação linear (PPL) apresentado em (2).

$$\begin{aligned} \max h_o &= \sum_{j=1}^s u_j y_{jo} - u_* \\ \text{sujeito a} \\ \sum_{i=1}^m v_i x_{io} &= 1 \\ \sum_{j=1}^s u_j y_{jk} - \sum_{i=1}^n v_i x_{ik} - u_* &\leq 0, \quad k = 1, \dots, n \\ u_j, v_i &\geq 0 \quad \forall x, y \\ u_* &\in \Re \end{aligned} \tag{2}$$

Neste modelo, para a DMU  $o$  em análise, a eficiência é dada por  $h_o$ ;  $x_{ik}$  representa o *input*  $i$  da DMU  $k$ ;  $y_{jk}$  representa o *output*  $j$  da DMU  $k$ ;  $v_i$  e  $u_j$  representam os pesos dados aos *inputs*  $i$  e aos *outputs*  $j$ , respectivamente;  $u_*$  é um factor de escala (quando positivo, indica que a DMU está em região de retornos decrescentes de escala; se negativo,

os retornos de escala são crescentes). Se  $h_o$  é igual a 1, a DMU  $o$  em análise é considerada eficiente. No PPL (2) as variáveis de decisão são  $v_i$ ,  $u_j$  e  $h_o$ . De forma não matemática, no modelo BCC uma DMU é eficiente se, na escala em que opera, é a que melhor aproveita os *inputs* de que dispõe. Já no modelo CCR, uma DMU é eficiente quando apresenta o melhor quociente de *outputs* com relação aos *inputs*, ou seja, aproveita melhor os *inputs* sem considerar a escala de operação da DMU.

## 2.2 Fronteira invertida

Os modelos DEA, devido a produzirem avaliações benevolentes, geralmente resultam em um grande número de DMUs eficientes. Além disso, devido às características matemáticas do modelo BCC, as DMUs que têm o menor valor de um dos *inputs* ou o maior valor de um dos *outputs* são consideradas eficientes, mesmo que as relações com as demais variáveis não sejam as melhores (Ali, 1993). Essas DMUs são chamadas de falsamente eficientes ou eficientes à partida.

Existem vários métodos para melhorar a discriminação em modelos DEA (Adler et al., 2002; Angulo Meza e Lins, 2002; Leta et al., 2005). Um desses métodos, que tem a vantagem adicional de eliminar as DMUs falsamente eficientes no modelo BCC, é a fronteira invertida. Este tipo de fronteira, baseada na inversão de *inputs* e *outputs*, é apresentada em Yamada et al. (1994), Entani et al. (2002), Lins et al. (2005).

O uso da fronteira invertida como ferramenta para aumentar a discriminação entre as DMUs pode ser visto em Pimenta et al. (2004), Angulo Meza et al. (2005), Leta et al. (2005), Soares de Mello et al. (2005, 2006), Gomes et al. (2006). A fronteira invertida com este objectivo é utilizada na construção de um índice chamado de índice de eficiência composta. Ele é definido como sendo a média aritmética entre a eficiência em relação à fronteira DEA convencional (padrão) e o complemento da eficiência em relação à fronteira invertida, tal como visto em (3). O uso do complemento justifica-se pelo fato de a fronteira invertida gerar uma medida de ineficiência.

$$eficiência\ composta = \frac{[eficiência\ padrão + (1 - eficiência\ invertida)]}{2} \quad (3)$$

Para se obter um índice em que as unidades eficientes têm o valor de 1, é feita a normalização da eficiência composta, ao dividirem-se seus valores pela maior de todas as medidas de eficiência compostas.

## 3 Selecção de Variáveis em DEA

Na literatura DEA, pouco tem sido discutido sobre o problema da selecção das variáveis usadas na modelação. A grande maioria dos trabalhos publicados traz a abordagem da selecção de variáveis segundo a opinião de especialistas ou até mesmo da disponibilidade de dados. Thanassoulis (1996) ressaltou que a modificação do conjunto de variáveis seleccionadas poderá ter grande impacto no resultado da avaliação. Sendo assim, torna-se extremamente relevante discutir e comparar métodos de selecção de variáveis em DEA. No entanto, antes de apresentar qualquer método deve-se atentar para o que significa realmente a selecção de variáveis.

Deve-se ter em conta que o facto de uma diferente escolha de variáveis conduzir a resultados diferentes não deve ser interpretado como uma fraqueza de DEA. Na verdade, escolher variáveis diferentes significa que se pretende levar em conta uma dimensão diferente do problema, ou seja, olhar para as DMUs segundo outro ponto de vista. Assim, por exemplo, em Soares de Mello et al. (2003) diferentes escolhas dos conjuntos de variáveis levaram a 3 diferentes avaliações de empresas aéreas: operacional, de vendas e total. A análise prévia de possíveis conjuntos de variáveis é uma etapa fundamental a ser feita em conjunto por decisores, especialistas e analistas. Só após esta escolha prévia é que faz sentido pensar-se em métodos de selecção.

Caso não seja desejável usar modelos avançados, nem seja possível aumentar o número de DMUs, uma das opções é restringir as variáveis que vão entrar no modelo. Neste contexto, os métodos de selecção de variáveis devem ser vistos como instrumentos de auxílio à decisão, que orientarão a escolha final. Esta não deve ficar presa ao resultado de um modelo matemático, por mais sofisticado que seja. Sempre deve ser feita em conjunto pelos agentes de decisão, especialistas e analistas, que poderão (ou não) usar um método de selecção como ferramenta.

Os próprios métodos de selecção já existentes e o que será aqui proposto permitem diferentes graus de intervenção do agente de decisão. Aqueles que exigem pouca ou nenhuma informação subjectiva são mais adequados a decisores que estejam em grande dúvida e não consigam emitir opiniões. Outros, que exigem alguma escolha, devem ser usados por decisores que queiram impor alguma direcção ao método, sem no entanto serem capazes sozinhos de realizar uma escolha de variáveis. Em qualquer método usado, o decisor deve ser confrontado com o resultado e verificar a sua coerência. Por exemplo, deve verificar se não foi omitida nenhuma variável que considere imprescindível, e se há relação causal entre todos os pares *input-output* seleccionados. Pode ainda ser recomendável fazer a selecção por mais de um método para comparar resultados e tomar a decisão com mais conhecimento do assunto.

Um método de selecção de variáveis foi proposto por Norman e Stoker (1991), no qual aliaram a análise de correlação simples à análise envoltória de dados. Os autores propuseram um procedimento iterativo que simultaneamente identifica os *inputs* e *outputs* relevantes e calcula medidas de ineficiência dos planos de operação observados.

Já Golany e Roll (1989) enunciaram-se três estágios do processo de selecção de variáveis: (1) '*Judgmental screening*', (2) '*Non-DEA quantitative analysis*' e (3) '*DEA based analysis*'. O primeiro refere-se à distinção entre variáveis determinantes de eficiência e variáveis explicativas de ineficiência, que pode ser resolvida através de análise de causalidade. O segundo estágio sugere o uso de análise de regressão para determinar se uma variável deve ser *input* ou *output*. O terceiro estágio, sugerido por Charnes et al. (1978), pode ser usado para identificar as variáveis que deverão ser excluídas, ou seja, são aquelas que não agregam eficiência significativa ao modelo.

Em Lins e Moreira (1999) e Lins e Angulo Meza (2000) é proposto um método baseado apenas na relação causal entre *inputs* e *outputs*. Este método, chamado de 'I-O Stepwise', não é eficaz para melhorar o poder de discriminação dos modelos DEA, apesar de conduzir a modelos com forte relação causal. Em Soares de Mello et al. (2004) é proposto um modelo que leva em conta tanto a relação causal (medida pela eficiência média) quanto o poder de discriminação (medido pelo número de DMUs na fronteira eficiente), denominado 'Método Multicritério de Selecção de Variáveis em DEA'. Em Senra (2004) e Senra et al. (2007) foram apresentados modelos baseados no de Soares de Mello et al. (2004), na qual é retirada parte da subjectividade anterior.

Todos estes métodos são baseados na inserção gradativa de variáveis no modelo, sendo o ponto de paragem inteiramente arbitrário. Senra (2004) e Senra et al. (2007)

consideram que os métodos de inserção gradativa são heurísticas de um método óptimo. Este método deve considerar todas as possíveis combinações de *outputs* e *inputs*, e escolher o conjunto de variáveis com o qual seja obtido o melhor compromisso entre relação causal e poder de discriminação. No entanto, este método é, em muitos casos, extremamente moroso.

Em Senra et al. (2007) foi proposta uma forma de escolha de variáveis que reduz a morosidade do processo. O método chamado 'Multicritério Combinatório por Cenários' é um método em duas fases. Na primeira fase são construídos cenários que serão analisados na segunda. Chama-se 'cenário da primeira fase' ao melhor modelo com duas, três, quatro etc., variáveis. A segunda fase limita-se à escolha do melhor cenário e termina quando todas as variáveis tiverem sido incluídas no modelo.

O método pode ser descrito pelas Figuras 1 e 2, onde  $S_{EF}$  é a normalização das eficiências médias para cada conjunto de variáveis, e  $S_{DIS}$  é a normalização da quantidade de DMUs eficientes para cada conjunto de variáveis. As formas de normalização utilizadas podem ser vistas em Soares de Mello et al. (2002), Soares de Mello et al. (2004), Senra et al. (2007). O indicador  $S$  é a média aritmética entre  $S_{EF}$  e  $S_{DIS}$ .

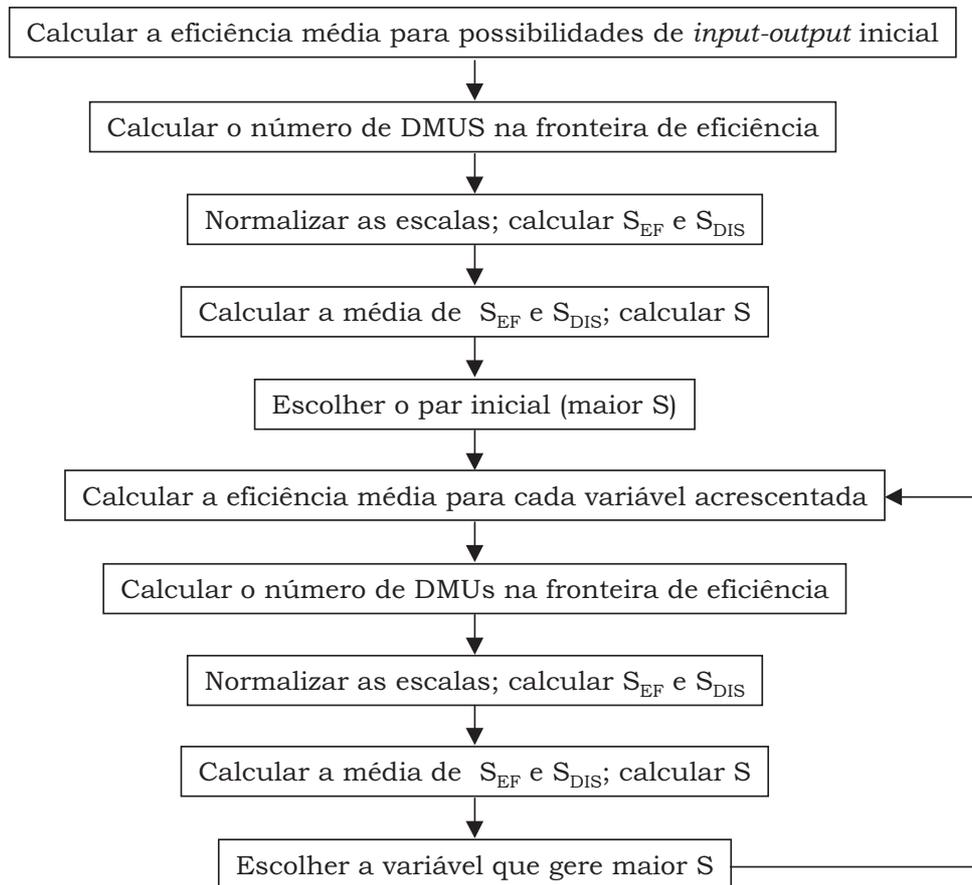


Figura 1: 1ª fase do 'Método Multicritério Combinatório por Cenários' de selecção de variáveis

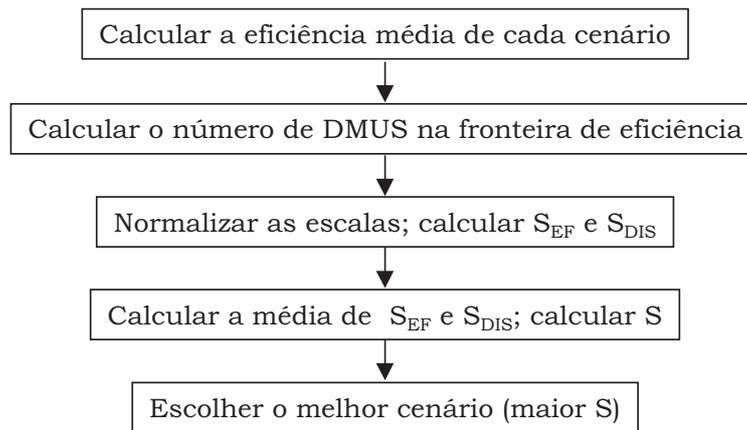


Figura 2: 2ª fase do 'Método Multicritério Combinatório por Cenários' de selecção de variáveis

Este modelo, ao contrário de outros modelos da literatura, não impõe um ponto de paragem arbitrário. Os modelos que impõem este ponto de paragem são, por sua natureza, úteis apenas quando o número de variáveis é grande comparado ao número de DMUs. Já este modelo e o modelo que dele será derivado permitem escolher não só as variáveis como o número delas a entrar no modelo. Assim, eles podem ser úteis mesmo quando há poucas variáveis e muitas DMUs.

Por outro lado, este método tem o grande inconveniente de fazer várias normalizações sucessivas com diferentes valores de referência. Esta característica pode fazer com que a combinação ideal de variáveis seja examinada, mas não escolhida.

Cabe aqui ainda mencionar o modelo de selecção de variáveis descrito por Wagner e Shimshak (2007). Os autores propuseram um procedimento tipo *stepwise* inverso, que envolve a maximização (ou minimização) sequencial da mudança média na eficiência conforme variáveis são retiradas do modelo. O algoritmo proposto pelos autores inicia com um modelo composto por todas as possíveis variáveis de *input* e *output*. O ponto de paragem é atingido quando o modelo é reduzido a uma variável de *input* e uma de *output*. Este é um modelo que só leva em conta o ajuste à fronteira. Ignora, dessa forma, o aumento de discriminação.

### 3.1 Método de selecção de variáveis proposto

Para evitar o problema das várias normalizações do método de selecção de variáveis 'Multicritério Combinatório por Cenários', propõe-se neste artigo um método alternativo que faz uso de normalização única. O método de normalização usado anteriormente considera uma eficiência média normalizada ( $S_{EF}$ ) que atinge o valor 1 na eficiência máxima e zero na eficiência mínima. Já o poder de discriminação normalizado ( $S_{DIS}$ ) apresenta o valor 1 quando há o menor número de DMUs na fronteira e zero para o maior número, para cada conjunto de variáveis usado.

Ao contrário do método anterior, a normalização aqui proposta não considera os valores máximos e mínimos efectivamente atingidos para a eficiência e o número de DMUs na fronteira para cada quantidade de variáveis. Ele leva em conta os valores extremos teoricamente alcançáveis. O número de DMUs na fronteira está no intervalo  $[1, n]$  (sendo  $n$  o número total de DMUs) e a eficiência média está no intervalo  $[0, 1]$ . Para obter  $S_{EF}$  deve-se levar em conta que quanto maior a eficiência média, maior deve ser o

valor dessa variável. Assim, para obter a normalização faz-se  $S_{EF} = \text{eficiência média}/10$ . Para obter o  $S_{DIS}$  deve-se considerar que quanto maior o número de DMUs na fronteira, menor é o valor da variável. Desta forma, para a sua normalização faz-se  $S_{DIS} = 10 \frac{n - N}{n - 1}$ , onde  $n$  é o número total de DMUs e  $N$  é o número de DMUs na fronteira.

Esta normalização pode ser usada em cada etapa do método para escolher a próxima variável a ser inserida, ou, caso o número de variáveis não seja grande, calculada para todas as combinações possíveis de *inputs* e *outputs*. O conjunto final de variáveis a ser escolhido é aquele que apresentar o maior índice  $S = S_{EF} + S_{DIS}$ , que é uma medida de compromisso entre o bom ajuste à fronteira e a boa discriminação do modelo.

## 4 Modelação e Resultados

Para avaliação do uso da energia eléctrica no Brasil foram seleccionados como DMUs os 27 estados brasileiros. Como possíveis variáveis a serem escolhidas no modelo final tem-se: Consumo Residencial per capita (medido em GWh/habitante); Produto Interno Bruto (PIB) estadual (medido em US\$/habitante, ponderado pelo poder de compra); Índice de Potencial de Consumo (IPC – índice adimensional); Índice de Desenvolvimento Humano (IDH – índice adimensional). Dado o enfoque de eficiência energética já mencionado previamente, o consumo é um *input* e as demais variáveis são candidatas a *outputs*, em uma abordagem semelhante às de Angulo Meza et al. (2006), Chauhan et al., (2006), Hu e Kao (2007).

Os dados relacionados ao consumo de energia eléctrica foram obtidos do Ministério de Minas e Energia do Brasil. O IDH e o PIB foram obtidos do Instituto de Pesquisas Económicas e Aplicadas do país e o IPC da publicação nacional *Gazeta Mercantil*. Todos os dados referem-se ao ano de 1996.

O IPC, desenvolvido pela Federação do Comércio do Estado de São Paulo em parceria com a Fundação Getúlio Vargas, mostra o perfil de consumo mensal das famílias brasileiras (Borges, 2004). Esse índice considera faixas de renda, gastos (que incluem impostos e tarifas públicas, além dos itens habituais de consumo) e variáveis socioeconómicas.

O conceito de desenvolvimento humano é a base do IDH e parte do pressuposto de que para aferir o avanço de uma população não se deve considerar apenas a dimensão económica, mas também outras características sociais, culturais e políticas que influenciam a qualidade da vida humana (Programa das Nações Unidas para o Desenvolvimento, 2004). Além de computar o PIB per capita, o IDH também considera os componentes longevidade e educação. O primeiro utiliza números de expectativa de vida ao nascer; o segundo é avaliado pelo índice de analfabetismo e pela taxa de matrícula em todos os níveis de ensino. Essas três dimensões têm a mesma importância na composição do índice, que varia de zero a um.

O PIB para cada estado é calculado a partir de estatísticas sobre o valor anual da produção, consumo intermediário e valor adicionado de cada actividade económica. A metodologia de cálculo pode ser encontrada em Instituto Brasileiro de Geografia e Estatística (2002). É importante destacar que tanto o Consumo Residencial quanto o PIB foram divididos pela população para normalizar as variáveis.

Foi usado o método proposto neste artigo e descrito na secção anterior para selecção de variáveis, chamado de 'Método Compensatório de Normalização Única'. Devido ao pequeno número de variáveis iniciais foram testadas todas as combinações possíveis, o

que caracteriza o Método Total (Senra, 2004; Senra et al., 2007). Os dados usados encontram-se na Tabela 1.

Tabela 1: Dados socioeconômicos e de consumo residencial de energia eléctrica por estado brasileiro

DMU	<i>Input</i>		<i>Outputs</i>	
	Consumo residencial per capita (GWh/hab x 10 <sup>-4</sup> )	PIB (U\$/hab.)	IPC	IDH
Acre	2,404	5.741	0,200	0,754
Alagoas	2,090	2.496	0,950	0,538
Amapá	3,131	5.370	0,170	0,786
Amazonas	2,739	5.718	1,170	0,775
Bahia	1,940	3.677	4,770	0,655
Ceará	1,998	2.667	2,400	0,590
Distrito Federal	5,692	14.854	2,230	0,869
Espírito Santo	4,222	6.251	1,450	0,836
Goiás	3,694	5.238	2,600	0,786
Maranhão	1,359	2.158	1,350	0,547
Mato Grosso	3,747	5.003	1,200	0,767
Mato Grosso do Sul	4,269	6.410	1,230	0,848
Minas Gerais	3,778	5.968	8,510	0,823
Pará	1,850	4.268	2,240	0,703
Paraíba	1,984	2.438	1,100	0,557
Paraná	4,074	6.485	5,620	0,847
Pernambuco	2,681	3.213	3,450	0,615
Piauí	1,714	2.004	0,750	0,534
Rio de Janeiro	6,196	8.653	11,630	0,844
Rio Grande do Norte	2,375	4.083	1,030	0,668
Rio Grande do Sul	4,520	7.395	7,080	0,869
Rondônia	2,816	6.448	0,580	0,820
Roraima	3,023	6.231	0,130	0,818
Santa Catarina	4,541	6.519	3,060	0,863
São Paulo	6,415	10.536	34,070	0,868
Sergipe	2,383	5.122	0,660	0,731
Tocantins	1,818	1.575	0,390	0,587

Foi usado o modelo CCR e, posteriormente, o modelo BCC, ambos orientados a *input* (modelos (1) e (2), respectivamente), o que caracteriza a eficiência energética do ponto de vista ambiental mencionado anteriormente. Isto significa que o objectivo é diminuir o consumo de energia eléctrica per capita, enquanto são mantidos os outros índices.

Os resultados para a selecção de variáveis com o uso do modelo CCR orientado a *input* são apresentados na Tabela 2. Verifica-se que o maior índice S, 16,8050, é obtido com o uso de três variáveis, Consumo per capita, PIB e IDH. No entanto, é interessante verificar que a substituição da variável IDH pela variável IPC fornece um índice S, 16,5090, muito próximo ao do modelo anterior. É importante notar que existe uma certa redundância no uso dos índices IDH e PIB, já que este é um componente do primeiro. Assim, pode ser vantajoso considerar um modelo com um índice S mais baixo, mas que evite a redundância das duas variáveis no modelo DEA.

Tabela 2: Selecção de variáveis com o 'Método Compensatório de Normalização Única' e modelo DEA CCR

DMU	Consumo e PIB	Consumo e IPC	Consumo e IDH	Consumo, PIB e IPC	Consumo, PIB e IDH	Consumo, IPC e IDH	Consumo, PIB, IDH e IPC
Acre	0,9150	0,0157	0,7789	0,9150	0,9974	0,7789	0,9974
Alagoas	0,4576	0,0856	0,6393	0,4777	0,6517	0,6393	0,6517
Amapá	0,6572	0,0102	0,6234	0,6572	0,7285	0,6234	0,7285
Amazonas	0,8001	0,0804	0,7029	0,8084	0,8763	0,7029	0,8763
Bahia	0,7263	0,4630	0,8385	0,8854	0,8778	1,0000	1,0000
Ceará	0,5115	0,2262	0,7334	0,5848	0,7452	0,7803	0,7806
Distrito Federal	1,0000	0,0738	0,3792	1,0000	1,0000	0,3807	1,0000
Espírito Santo	0,5673	0,0647	0,4918	0,5762	0,6201	0,4918	0,6201
Goiás	0,5434	0,1325	0,5285	0,5793	0,6049	0,5463	0,6072
Maranhão	0,6087	0,1871	1,0000	0,6640	1,0000	1,0000	1,0000
Mato Grosso	0,5117	0,0603	0,5084	0,5205	0,5716	0,5084	0,5716
Mato Grosso do Sul	0,5754	0,0543	0,4933	0,5799	0,6278	0,4933	0,6278
Minas Gerais	0,6054	0,4241	0,5411	0,7529	0,6648	0,7375	0,7781
Pará	0,8840	0,2280	0,9437	0,9473	1,0000	0,9708	1,0000
Paraíba	0,4708	0,1044	0,6971	0,4979	0,7056	0,6971	0,7056
Paraná	0,6100	0,2597	0,5164	0,6935	0,6643	0,6024	0,7169
Pernambuco	0,4592	0,2423	0,5697	0,5402	0,5902	0,6413	0,6431
Piauí	0,4480	0,0824	0,7737	0,4672	0,7737	0,7737	0,7737
Rio de Janeiro	0,5352	0,3534	0,3383	0,6572	0,5610	0,5261	0,6664
Rio Grande do Norte	0,6588	0,0817	0,6986	0,6717	0,7444	0,6986	0,7444
Rio Grande do Sul	0,6270	0,2950	0,4775	0,7237	0,6727	0,5925	0,7428
Rondônia	0,8774	0,0388	0,7232	0,8774	0,9519	0,7232	0,9519
Roraima	0,7898	0,0081	0,6720	0,7898	0,8608	0,6720	0,8608
Santa Catarina	0,5501	0,1269	0,4720	0,5836	0,6003	0,4923	0,6063
São Paulo	0,6294	1,0000	0,3361	1,0000	0,6480	1,0000	1,0000
Sergipe	0,8238	0,0522	0,7620	0,8238	0,9095	0,7620	0,9095
Tocantins	0,3320	0,0404	0,8020	0,3382	0,8020	0,8020	0,8020
Eficiência média	0,6361	0,1774	0,6311	0,6894	0,7574	0,6903	0,7912
N.º DMUs eficientes	1	1	1	2	3	3	5
S <sub>EFF</sub>	6,3611	1,7744	6,3114	6,8936	7,5742	6,9025	7,9120
S <sub>DIS</sub>	10,0000	10,0000	10,0000	9,6154	9,2308	9,2308	8,4615
S	16,3611	11,7744	16,3114	16,5090	16,8050	16,1333	16,3736

Apesar de não existirem mais efeitos de escala devido ao uso de variáveis per capita, foi testado o modelo BCC pelo facto de propiciar um envelopamento mais completo, sem a exigência de proporcionalidade entre *inputs* e *outputs*. Os resultados do modelo BCC orientado a *input* são apresentados na Tabela 3. Pode-se notar que, para o modelo BCC, o melhor resultado foi obtido para um modelo com duas variáveis, Consumo Residencial per capita e IDH, com um índice S de 17,0250.

Este modelo de apenas de um *input* e um *output* apresenta a vantagem de evitar a redundância mencionada no modelo CCR, ou seja, considerar apenas PIB quando incluído no IDH sem o uso explícito. Além disso, a escolha de variáveis do modelo BCC coincidiu com a opinião de especialistas do sector eléctrico que foram ouvidos. Por estes motivos, conclui-se então que, neste caso, o modelo BCC é preferível ao modelo CCR.

Maranhão é eficiente por ser o estado que apresenta menor consumo de energia residencial per capita. No entanto, por ser eficiente também no modelo CCR (tanto a que considera as mesmas duas variáveis, quanto no que foi escolhido pelo método de selecção de variáveis) ela é verdadeiramente eficiente.

O estado de Rio Grande do Sul apresenta o maior IDH. Observa-se na Tabela 3 que no modelo DEA em que seja desconsiderada essa variável, Rio Grande do Sul passa a ser ineficiente. Portanto, esta DMU deve ser considerada como falsa eficiente.

Tabela 3: Selecção de variáveis com o 'Método Compensatório de Normalização Única' e modelo DEA BCC

DMU	Consumo e PIB	Consumo e IPC	Consumo e IDH	Consumo, PIB e IPC	Consumo, PIB e IDH	Consumo, IPC e IDH	Consumo, PIB, IDH e IPC
Acre	0,9919	0,5651	0,9447	0,9919	1,0000	0,9447	1,0000
Alagoas	0,6876	0,6500	0,6500	0,6876	0,6876	0,6500	0,6876
Amapá	0,7186	0,4339	0,8097	0,7186	0,8097	0,8097	0,8097
Amazonas	0,8678	0,4961	0,8927	0,8678	0,9027	0,8927	0,9049
Bahia	0,8827	0,9727	0,8757	1,0000	0,8827	1,0000	1,0000
Ceará	0,7393	0,7612	0,7478	0,7735	0,7478	0,7824	0,7824
Distrito Federal	1,0000	0,2626	0,7940	1,0000	1,0000	0,7940	1,0000
Espírito Santo	0,6087	0,3254	0,7987	0,6087	0,7987	0,7987	0,7987
Goiás	0,5962	0,4201	0,6864	0,6032	0,6864	0,7254	0,7254
Maranhão	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Mato Grosso	0,5650	0,3626	0,6348	0,5650	0,6348	0,6348	0,6348
Mato Grosso do Sul	0,6155	0,3182	0,8876	0,6155	0,8876	0,8876	0,8876
Minas Gerais	0,6531	0,6525	0,7730	0,7725	0,7730	0,9498	0,9498
Pará	1,0000	0,8086	1,0000	1,0000	1,0000	1,0000	1,0000
Paraíba	0,7175	0,6846	0,7005	0,7175	0,7175	0,7005	0,7175
Paraná	0,6516	0,4954	0,9217	0,7114	0,9217	0,9471	0,9471
Pernambuco	0,5984	0,6278	0,5866	0,6489	0,5984	0,6416	0,6489
Piauí	0,7925	0,7925	0,7925	0,7925	0,7925	0,7925	0,7925
Rio de Janeiro	0,5555	0,4757	0,5892	0,6645	0,6333	0,6795	0,7047
Rio Grande do Norte	0,7609	0,5721	0,7326	0,7609	0,7609	0,7326	0,7609
Rio Grande do Sul	0,6605	0,4965	1,0000	0,7375	1,0000	1,0000	1,0000
Rondônia	0,9379	0,4824	1,0000	0,9379	1,0000	1,0000	1,0000
Roraima	0,8476	0,4494	0,9260	0,8476	0,9260	0,9260	0,9260
Santa Catarina	0,5873	0,3574	0,9493	0,6003	0,9493	0,9493	0,9493
São Paulo	0,6430	1,0000	0,6991	1,0000	0,7766	1,0000	1,0000
Sergipe	0,9066	0,5702	0,8735	0,9066	0,9109	0,8735	0,9109
Tocantins	0,7474	0,7474	0,8167	0,7474	0,8167	0,8167	0,8167
Eficiência média	0,7531	0,5845	0,8179	0,7881	0,8376	0,8492	0,8650
N.º DMUs eficientes	3	2	4	5	6	6	8
S <sub>EFF</sub>	7,5307	5,8445	8,1788	7,8805	8,3759	8,4923	8,6502
S <sub>DIS</sub>	9,2308	9,6154	8,8462	8,4615	8,0769	8,0769	7,3077
S	16,7615	15,4598	17,0250	16,3420	16,4528	16,5692	15,9579

Na Tabela 4 apresenta-se o índice de eficiência com a aplicação da fronteira invertida para o modelo composto das variáveis Consumo e IDH, obtida pelo *software* SIAD (Angulo Meza et al., 2005). É interessante ressaltar que esta fronteira corresponde, no presente caso, à óptica da empresa fornecedora de energia eléctrica, tal como mencionado na secção 1. Este índice permite ordenar as DMUs e evitar alguns dos problemas clássicos de DEA, como DMUs eficientes ao considerar apenas um conjunto restrito de variáveis e DMUs eficientes à partida.

Tal como pode-se observar, os estados com melhor índice estão situados na região amazónica. Isso pode ser explicado pelo facto de existir um grande número de comunidades isoladas, as quais ou não usam energia eléctrica ou geram a sua própria energia sem recorrer às empresas distribuidoras.

Tabela 4: Resultados que consideram a fronteira invertida

DMU	Eficiência BCC padrão	Eficiência BCC invertida	Índice composto normalizado
Acre	0,9447	0,7446	0,9391
Alagoas	0,6500	1,0000	0,5086
Amapá	0,8097	0,7832	0,8033
Amazonas	0,8927	0,7565	0,8890
Bahia	0,8757	0,8189	0,8269
Ceará	0,7478	0,9102	0,6554
Distrito Federal	0,7940	0,9280	0,6777
Espírito Santo	0,7987	0,8336	0,7552
Goiás	0,6864	0,8365	0,6650
Maranhão	1,0000	0,9762	0,8011
Mato Grosso	0,6348	0,8624	0,6044
Mato Grosso do Sul	0,8876	0,8260	0,8308
Minas Gerais	0,7730	0,8065	0,7563
Pará	1,0000	0,7617	0,9690
Paraíba	0,7005	0,9639	0,5764
Paraná	0,9217	0,8097	0,8701
Pernambuco	0,5866	0,9464	0,5010
Piauí	0,7925	1,0000	0,6202
Rio de Janeiro	0,5892	1,0000	0,4610
Rio Grande do Norte	0,7326	0,8371	0,7007
Rio Grande do Sul	1,0000	0,8275	0,9175
Rondônia	1,0000	0,7221	1,0000
Roraima	0,9260	0,7427	0,9260
Santa Catarina	0,9493	0,8351	0,8719
São Paulo	0,6991	1,0000	0,5471
Sergipe	0,8735	0,7658	0,8668
Tocantins	0,8167	0,9116	0,7083

Na Figura 3 pode-se observar a representação gráfica das duas fronteiras. A região compreendida entre as duas forma a região onde é possível uma DMU estar situada e serve para estimar intervalos de variação do consumo residencial dado o IDH, ou vice-versa.

A análise visual da Figura 3 mostra que as DMUs estão bem próximas à fronteira eficiente, sendo esta constituída por um número pequeno de DMUs, o que caracteriza que o método cumpriu com seus dois objectivos (máxima discriminação e mínimo número de DMUs eficientes na fronteira). No entanto, existe pouca aderência à fronteira invertida, já que o método selecção de variáveis considera apenas a fronteira padrão.

Caso deseje-se usar uma óptica inversa (óptica da empresa), o consumo passa a ser considerado um *output* e os três índices considerados como *inputs*, tal como ser observado na Figura 4. Ao aplicar-se o mesmo método de selecção de variáveis para este caso com o modelo BCC, foram seleccionadas as variáveis PIB e Consumo per capita.

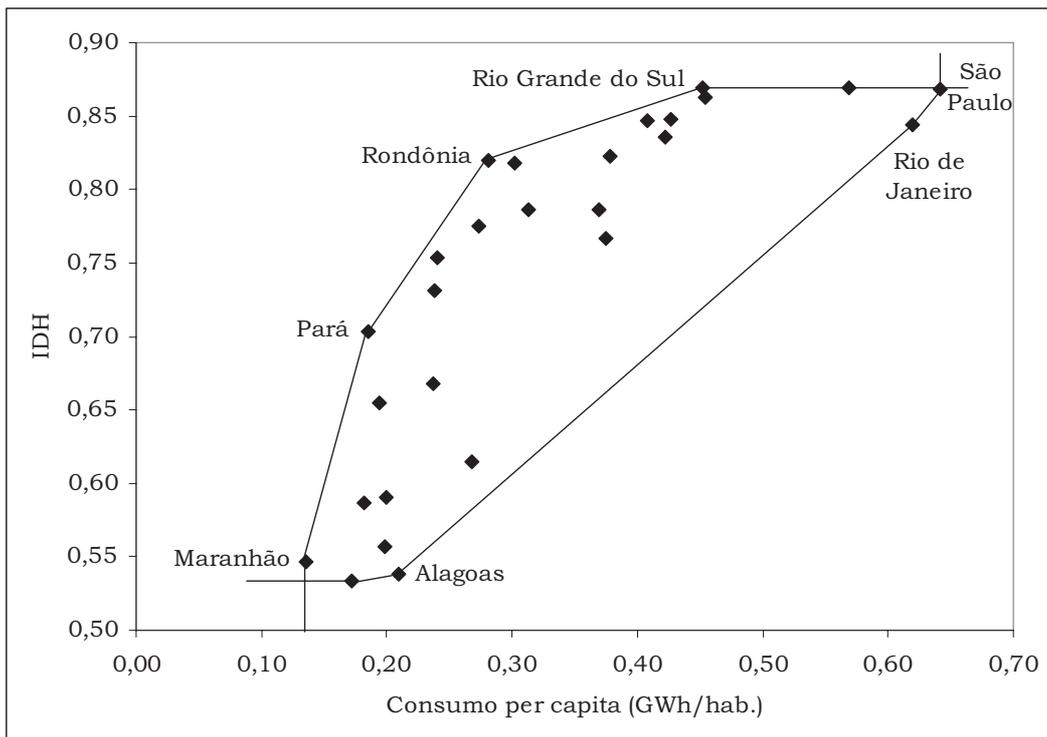


Figura 3: Fronteira DEA BCC e fronteira DEA BCC invertida para o caso investigado

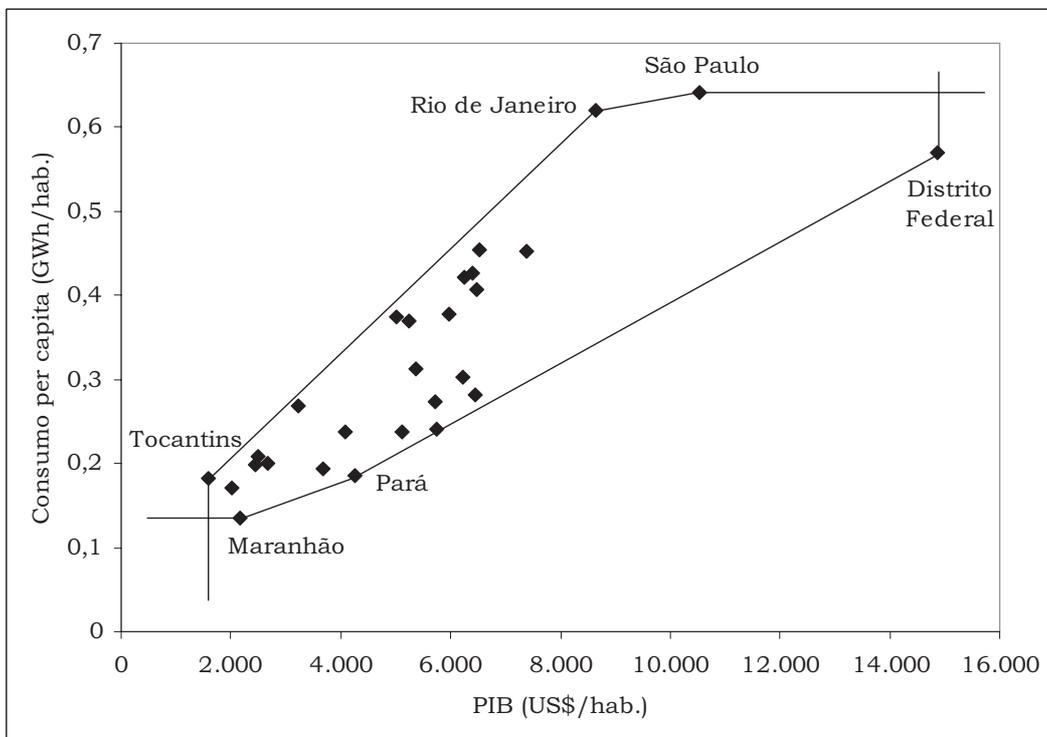


Figura 4: Fronteira DEA BCC e fronteira DEA BCC invertida para a óptica do empresário

## 4 Conclusões

Uma análise geral do método de selecção de variáveis, com o uso dos modelos CCR e BCC, mostra que em nenhum deles a variável IPC foi seleccionada. Ou seja, os outros índices usados são mais úteis para explicar o consumo do que o índice criado com esse fim.

Salienta-se o facto de os estados de região amazónica terem um bom desempenho segundo o índice composto. Uma possível explicação é o uso de geradores individuais que não entram no cálculo do consumo residencial de energia eléctrica, além da deficiente rede de energia eléctrica da região, que obriga o uso de outras fontes de energia.

Quanto ao objectivo do método de selecção de variáveis, a análise visual da Figura 3 mostra que as DMUs estão bem próximas à fronteira eficiente, sendo esta constituída por um número pequeno de DMUs, o que caracteriza que o método cumpriu com seus dois objectivos (máxima discriminação e mínimo número de DMUs eficientes na fronteira). A Figura 4 mostra a óptica do empresário, onde pode ser observado um melhor ajuste à fronteira.

## 5 Agradecimentos

Este trabalho teve o apoio do Fundo Setorial de Energia (CT-Energ), por intermédio do CNPq (CT-Energ/CNPq 01/2003), processo 400646/2003-0.

## 6 Referências

- Adler, N., Friedman, L. and Sinuany-Stern, Z. (2002) Review of ranking methods in the data envelopment analysis context, *European Journal of Operational Research*, Vol 140, pp. 249-265.
- Ali, A.I. (1993) Streamlined computation for Data Envelopment Analysis, *European Journal of Operational Research*, Vol 64, pp. 61-67.
- Angulo Meza, L., Biondi Neto, L., Soares de Mello, J.C.C.B. and Gomes, E.G. (2005) ISYDS – Integrated System for Decision Support (SIAD – Sistema Integrado de Apoio à Decisão): A Software Package for Data Envelopment Analysis Model, *Pesquisa Operacional*, Vol 25, No 3, pp. 493-503.
- Angulo Meza, L. and Lins, M.P.E. (2002) Review of methods for increasing discrimination in Data Envelopment Analysis, *Annals of Operations Research*, Vol 116, pp. 225-242.
- Angulo Meza, L., Soares de Mello, J.C.C.B., Gomes, E.G., Fernandes, A.J.S. and Biondi Neto, L. (2006) Análise da eficiência municipal no uso da energia elétrica considerando temperaturas. In: *Anais do XXXVIII Simpósio Brasileiro de Pesquisa Operacional*, Goiânia, Brasil.
- Banker, R.D., Charnes, A. and Cooper, W.W. (1984) Some models for estimating technical scale inefficiencies in Data Envelopment Analysis, *Management Science*, Vol 30, No 9, pp. 1078-1092.
- Borges, A.C. (2004) Fecomércio lança índice de potencial de consumo, Agência Estado Financeiro. Disponível em: <<http://www.aefinanceiro.com.br/artigos/2004/out/07/295.htm>>. Acesso em: 03 fev. 2005.
- Charnes, A., Cooper, W.W. and Rhodes, E. (1978) Measuring the efficiency of decision-making units, *European Journal of Operational Research*, Vol 2, pp. 429-444.
- Chauhan, N.S., Mohapatra, P.K.J. and Pandey, K.P. (2006) Improving energy productivity in

paddy production through benchmarking - An application of data envelopment analysis, *Energy Conversion and Management*, Vol 47, No 9-10, pp. 1063-1085.

Entani, T., Maeda, Y. and Tanaka, H. (2002) Dual models of interval DEA and its extensions to interval data, *European Journal of Operational Research*, Vol 136, pp. 32-45.

Golany, B. and Roll, Y. (1989) An application procedure for DEA, *Omega International Journal of Management Science*, Vol 17, No 3, pp. 237-1250.

Gomes, E.G., Soares de Mello, J.C.C.B. and Mangabeira, J.A.C. (2006) Fronteira DEA difusa na avaliação de eficiência em agricultura, *Investigação Operacional*, Vol 26, No 1, pp. 65-88.

Hu, J.-L. and Kao, C.-H. (2007) Efficient energy-saving targets for APEC economies, *Energy Policy*, Vol 35, No 1, pp. 373-382.

Instituto Brasileiro de Geografia e Estatística (2002) Contas Regionais do Brasil 2002, Contas Regionais, 2a edição, No. 13.

Leta, F.R., Soares de Mello, J.C.C.B., Gomes, E.G. and Angulo-Meza, L. (2005) Métodos de melhora de ordenação em DEA aplicados à avaliação estática de tornos mecânicos, *Investigação Operacional*, Vol 25, No 2, pp. 229-242.

Lins, M.P.E. and Angulo-Meza, L. (2000) Análise Envoltória de Dados e Perspectivas de Integração no Ambiente do Apoio à Decisão, Editora da COOPE/UFRJ, Rio de Janeiro, Brasil.

Lins, M.P.E. and Moreira, M.C.B. (1999) Método I-O Stepwise para seleção de variáveis em modelos de Análise Envoltória Dados, *Pesquisa Operacional*, Vol 19, No 1, pp. 39-50.

Lins, M.P.E., Novaes, L.F.L. and Legey, L.F.L. (2005) Real estate value assessment: a double perspective data envelopment analysis, *Annals of Operations Research*, Vol 138, No 1, pp. 79-96.

Lins, M.P.E., Silva, A.C.M. and Rosa, L.P. (2002) Regional variations in energy consumption of appliances: conditional demand analysis applied to Brazilian households, *Annals of Operations Research*, Vol 117, No 1-4, pp. 235-246.

Lins, M.P.E., Sollero, M.K.V., Calôba, G.M. and Silva, A.C.M. (2007) Integrating the regulatory and utility firm perspectives, when measuring the efficiency of electricity distribution, *European Journal of Operational Research*, Vol 181, No 3, pp. 1413-1424.

Norman, M. and Stoker, B. (1991) *Data Envelopment Analysis: The Assessment of Performance*, John Wiley and Sons, New York, USA.

Pimenta, H.L.N, Macedo, M.A. and Soares de Mello, J.C.C.B. (2004) Decisão de realização de investimentos em tecnologia da informação com análise envoltória de dados, *Produção Online*, Vol 4, No 2.

Programa das Nações Unidas para o Desenvolvimento (2004) *Desenvolvimento Humano e IDH*. Disponível em: <<http://www.pnud.org.br/idh/>>. Acesso em: 03 fev. 2005.

Senra, L.F.A.C. (2004) Métodos de seleção de variáveis em DEA: estudo de caso no setor elétrico brasileiro. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal Fluminense, Niterói, Brasil.

Senra, L.F.A.C., Nanci, L.C., Soares de Mello, J.C.C.B. and Angulo-Meza, L. (2007) Estudo sobre métodos de seleção de variáveis em DEA, *Pesquisa Operacional*, Vol 27, No 2, pp. 191-207.

Soares de Mello, J.C.C. B., Angulo-Meza, L., Gomes, E.G. and Biondi Neto, L. (2006) Evaluación de la concentración en una ruta aérea brasileira con modelo DEA y frontera invertida, *Revista de la Facultad de Ingeniería - Universidad de Tarapacá*, Vol 14, No 1, pp. 64-71.

Soares de Mello, J.C.C.B., Angulo-Meza, L., Gomes, E.G., Serapião, B.P. and Lins, M.P.E. (2003) Análise de envoltória de dados no estudo da eficiência e dos benchmarks para companhias aéreas brasileiras, *Pesquisa Operacional*, Vol 23, No 2, pp. 325-345.

Soares de Mello, J.C.C.B., Gomes, E.G., Angulo-Meza, L., Biondi Neto, L. and Sant'Anna, A.P. (2005) Fronteiras DEA difusas, *Investigação Operacional*, Vol 25, No 1, pp. 85-103.

Soares de Mello, J.C.C.B., Gomes, E.G., Angulo-Meza, L. and Lins, M.P.E. (2004) Selección de

variables para el incremento del poder de discriminación de los modelos DEA, *Revista de La Escuela de Perfeccionamiento en Investigación Operativa*, Vol 24, pp. 40-52.

Soares de Mello, J.C.C.B., Gomes, E.G., Soares de Mello, M.H.C. and Lins, M.P.E. (2002) Método multicritério para seleção de variáveis em modelos DEA, *Pesquisa Naval*, Vol 15, pp. 55-66.

Thanassoulis, E. (1996) Assessing the efficiency of schools with pupils of different ability using Data Envelopment Analysis, *Journal of the Operational Research Society*, Vol 47, No 1, pp. 84-97.

Utlu, Z. and Hepbasli, A. (2007) A review on analyzing and evaluating the energy utilization efficiency of countries, *Renewable & Sustainable Energy Reviews*, Vol 11, pp. 1-29.

Wagner, J.M. and Shimshak, D.G. (2007) Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives, *European Journal of Operational Research*, Vol 180, pp. 57-67.

Yamada, Y., Matui, T. and Sugiyama, M. (1994) New analysis of efficiency based on DEA, *Journal of the Operations Research Society of Japan*, Vol 37, No 2, pp. 158-167.