

8 Análises e Resultados

8.1. **Definindo *inputs* e *outputs*: escolha de variáveis para definição dos agrupamentos e estimação das fronteiras de eficiência**

Foi dito na seção 7.1 que um dos passos fundamentais na análise de *clusters* é a escolha de variáveis para representar cada um dos objetos. São estas variáveis que orientam a formação dos padrões. Foram obtidos vários dados³⁰ das empresas de 60 concessionárias de distribuição de energia elétrica no Brasil para o ano de 2001. Foi feito então um pré-processamento dos dados para identificar pontos discrepantes e distorções nos dados que pudessem comprometer a análise³¹. Após esta fase de tratamento dos dados, deu-se início a aplicações das metodologias de classificação de padrões e definição de fronteiras.

Em todo modelo (seja estatístico ou de programação linear), é fundamental que se faça uma discussão *a priori* de quais variáveis serão consideradas na estruturação do problema. No presente problema, sabe-se que uma empresa utiliza *inputs* para produzir *outputs*, num determinado contexto. Quais são então os *inputs* e os *outputs* a considerar na análise comparativa de concessionárias de distribuição de energia elétrica? Como caracterizar e levar em conta diferenças de contexto de operação? As respostas a estas questões constituem o que é comum designar pela fase de estruturação do problema.

³⁰ Tanto relativos a custos quanto como característicos da atividade de cada empresa como área de concessão (km²), km de rede, consumo total, número de consumidores (aberto em algumas categorias como residencial, rural), número de interrupções em um período de tempo (FEC) e sua duração média por consumidor (DEC). Estas últimas variáveis são considerados *proxy* de qualidade do serviço prestado.

³¹ Para a variável km de rede, por exemplo, havia um dado que se fosse verdadeiro, a concessionária cobriria uma área que daria para se dar um volta na Terra. Descobriu-se, no processamento, que o dado foi fornecido em metros e não em km.

Naturalmente, quanto maior for a quantidade de cada *input* usada pela empresa na distribuição de energia elétrica, maiores serão as suas despesas operacionais. O gerenciamento dos *inputs* pessoal, materiais e serviços de terceiros compete à empresa, não sendo as quantidades usadas de cada um deles nem o seu *mix* que deverão estar aqui diretamente em questão. Certo é, porém, que, dados os *outputs* do processo, quanto maiores forem as despesas em *inputs*, menor será a produtividade operacional e, portanto, menos provável é que a empresa seja eficiente quando comparada com as outras. Por tudo isto, deve tomar-se como *input* único a “despesa operacional total” na distribuição propriamente dita. É importante notar que esta “despesa operacional total” não corresponde ao total da parcela B dos “custos operacionais” tal como definidos pela ANEEL, posto que desta deve ser excluída a parcela relativa a “outros custos”. Por isso, para evitar equívocos, será adotada a sigla OPEX (do inglês *operational expenditure*).

A replicabilidade é crucial, uma vez que os excessos de despesa serão determinados por análise comparativa. Isto é, se por exemplo se concluir que uma empresa pode poupar $y\%$ do seu OPEX, é porque “existe” outra empresa B (ou combinação de outras) que produz o mesmo *output* que A, mas com uma OPEX inferior em $y\%$. Implicitamente, está a assumir-se que A pode ajustar a sua “função de produção” para replicar a de B. Se esta hipótese de replicabilidade não for realista, y não pode ser tomado como representando o excedente de OPEX de A, pois nada prova que A possa manter os mesmos *outputs* se reduzir os seus *inputs* em quantidades correspondentes a y .

Os *outputs* do processo produtivo devem servir para explicar as diferenças de despesas operacionais (OPEX) entre as distribuidoras. É importante assinalar que o objetivo é avaliar as reservas potenciais de eficiência, e não desenvolver um modelo explicativo dos níveis de despesa das empresas. Por isso, ao termo “*outputs*” deve ser dado aqui o sentido mais lato de “determinantes da OPEX” (*cost drivers*). São aspectos tais que a satisfação do cliente aumenta quando a sua quantidade aumenta ou aspectos exógenos (isto é, não gerenciáveis) que influenciam a OPEX, e devem ambos ser levados em consideração para determinar o limite de preço a fixar pelo regulador. No entanto, outros aspectos que, embora influenciem a OPEX, refletem opções de gerenciamento (como, por exemplo, as relativas a políticas próprias de pessoal e salariais), não deverão ser

considerados, sob pena de não ser possível detectar escolhas operacionais eficientes que, por exemplo, reduzam a quantidade de mão-de-obra necessária à função de distribuição.

Uma seqüência de perguntas de natureza comparativa pode ser útil para ajudar a determinar quais são os *outputs* (determinantes da OPEX). Suponha-se duas distribuidoras igualmente eficientes mas com OPEX diferentes. Primeiro:

I - Que fatores não gerenciáveis podem justificar que as duas empresas tenham OPEX diferentes?

A resposta mais óbvia para esta pergunta é: a quantidade total de energia elétrica distribuída (ou consumo total). Se duas empresas distribuem quantidades significativamente diferentes de energia elétrica, é provável que elas tenham diferentes OPEX. É importante assinalar que o consumo total é exógeno porque a empresa não pode, por exemplo, se recusar a satisfazer parte da demanda, mesmo que tal lhe fosse favorável em termos de produtividade (por exemplo, se a partir de certo nível de demanda, o custo marginal do fornecimento for superior ao preço máximo fixado); ou vice-versa.

II – Se as duas empresas distribuem a mesma quantidade de energia elétrica, existem outros fatores exógenos que justifiquem OPEX diferentes?

Nesta situação, diferentes OPEX podem ocorrer se a demanda total é a mesma mas as proporções dos consumos residencial, industrial, rural, etc divergem significativamente para as duas empresas. Estamos aqui também perante fatores não gerenciáveis, pois que a estrutura da demanda na área de concessão não é controlável pelo distribuidor.

III - Se as duas empresas distribuem a mesma quantia de energia elétrica para cada tipo de consumidores, existem outros fatores exógenos que justifiquem OPEX diferentes?

Presumivelmente o número de consumidores, pois satisfazer uma certa demanda de x consumidores acarretará mais despesas do que distribuir a mesma quantidade de energia elétrica por um número significativamente menor de clientes.

IV - Se as duas empresas distribuem a mesma quantia de energia elétrica a cada tipo de consumidores, e têm o mesmo número de clientes, existem outros fatores exógenos que justifiquem OPEX diferentes?

O número de consumidores pode ser o mesmo no total, mas proporcionalmente diferentes por cada tipo, o que poderá acarretar diferenças de OPEX.

V - Se as duas empresas distribuem a mesma quantia de energia elétrica a cada tipo de consumidores, e têm o mesmo número de clientes de cada tipo, existem outros fatores que justifiquem OPEX diferentes?

Diferentes níveis de dispersão geográfica dos consumidores podem justificar diferenças de OPEX, porque distribuir energia elétrica em uma zona de maior densidade populacional pode acarretar menores custos (de manutenção da rede, por exemplo) do que numa área de menor densidade. Como o distribuidor não controla a dispersão de seus clientes, estamos perante outro fator exógeno de natureza ambiental. O tamanho da zona de concessão poderia ser usada como medida do efeito de dispersão. Porém, características geográficas diferentes, por exemplo do relevo (ver Figura 18), podem originar diferentes tamanhos das redes em áreas de concessão idênticas e com o mesmo número de clientes por unidade de área. O tamanho da rede pode por isso ser um melhor indicador de dispersão.

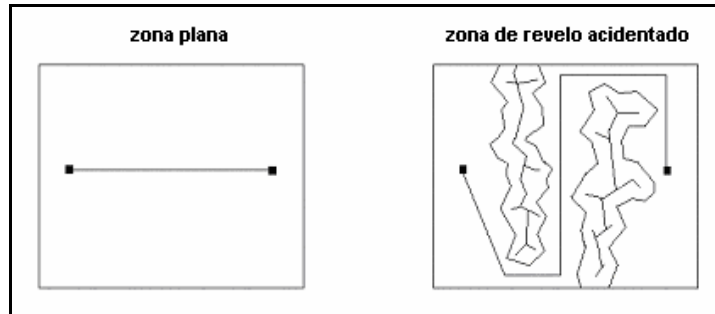


Figura 18. Rede em zona plana vs. rede em zona de relevo acidentado.

VI – Se as duas empresas forem idênticas em todos os fatores não gerenciáveis acima levantados, existem alguns outros fatores que justifiquem OPEX diferentes?

O nível de qualidade é certamente um *cost driver*: despesas operacionais maiores numa empresa que em outra com *call centers* e outras medidas para melhorar a qualidade de atendimento ao cliente e de prestação do serviço de distribuição de energia elétrica, aumentarão a OPEX em proporções diferentes.

Desta forma, não considerar a qualidade como *output*, se bem que ela seja pelo menos parcialmente gerenciável, significaria penalizar as concessionárias

com melhor esforço de qualidade, pois que o acréscimo de custos a isso devido não encontraria justificativa na análise comparativa de eficiência operacional. Estar-se-ia então a dar um incentivo perverso para reduzir custos associados à qualidade, penalizando em última análise o bem estar do consumidor. Note-se que a simples imposição de padrões mínimos de qualidade a todas as empresas não é suficiente para diferenciar entre empresas com diferentes níveis de qualidade de serviço.

Em síntese, ressalta-se então que o critério básico adotado na escolha das variáveis foi uma identificação prévia de um reduzido conjunto de dimensões com poder de explicação sobre os Custos Operacionais Gerenciáveis (OPEX)³². Especificamente, foram identificadas as quatro dimensões a seguir:

1) **Número de consumidores:** O total de consumidores está associado com custos de deslocamento, leitura, entrega de contas, planejamento, faturamento, projetos de redes, *call center*, relações com clientes etc.

2) **Consumo total:** As distribuidoras provêm dois *outputs* aos seus consumidores: o serviço e o produto. Uma *proxy* da quantidade de serviço é o número de consumidores, enquanto o produto é o montante de energia elétrica fornecida.

3) **Dispersão dos consumidores:** A dispersão geográfica das cargas determinará a necessidade de construção de redes, a quantidade de medidores necessária, a necessidade de deslocamentos para leitura e entrega de contas e para manutenção e operação das redes, além do cadastro e planejamento da expansão destas redes. Por exemplo, existe grande diferença entre distribuir energia para um consumidor de 1.000 MWh e para 1000 consumidores de 1 MWh, com reflexos diretos nos custos operacionais, apesar de terem o mesmo valor faturado.

4) **Qualidade:** O nível de qualidade é certamente um *cost driver* do OPEX, pois as despesas operacionais com *call centers* e outras medidas para melhorar a qualidade de atendimento ao cliente e de prestação do serviço de distribuição de energia elétrica contribuem para aumentar os custos operacionais.

³² Único *input* do modelo DEA, por exemplo.

Obviamente, é desejável que as variáveis usadas na estimação de fronteiras de eficiência sejam aquelas usadas na análise de agrupamentos³³. Por esta razão os seguintes princípios³⁴ devem ser observados na especificação das variáveis:

- As variáveis *inputs* e *outputs* devem mostrar a relação entre o serviço prestado pela concessionária e os insumos que ela gerencia para tal fim;
- As variáveis *inputs* devem ser elementos controláveis pela concessionária;
- As variáveis *outputs* devem refletir o gerenciamento dos *inputs* pela concessionária.

Enfim, o conjunto de variáveis *inputs* e *outputs* deve ser completo, de fácil operacionalização, não redundante e mínimo. Assim, foram escolhidas, preliminarmente, 16 variáveis³⁵ para caracterização das áreas de concessão das empresas, sendo que cada variável é uma faceta das quatro dimensões apresentadas abaixo:

Variáveis associadas com a dimensão número de consumidores

- Número de consumidores - nconsum
- Número de consumidores residenciais - nconsres
- Número de consumidores não residenciais - nconsnres

Variáveis associadas com a dimensão consumo

- Consumo total (MWh) – constot
- Consumo residencial (MWh) – consres
- Consumo não residencial (MWh) - consnres

Variáveis associadas com a dimensão dispersão dos consumidores³⁶

- Km de rede – kmderede
- Área
- Número de consumidores por km² (densidade de consumidores) – consumid/km²

³³ Novamente, no caso de um modelo DEA, estas variáveis constituem os *outputs*.

³⁴ Principalmente em se tratando de um modelo DEA.

³⁵ Os dados analisados são apresentados no Anexo A da Tese.

³⁶ A potência instalada e da quantidade de linhas e redes da área de concessão, considerando que os custos de materiais e serviço de terceiros são função destas variáveis

- Número de consumidores por km de rede (quantidade de rede necessária, em média, para atender cada consumidor) – consumid/km
- KVA por km² (densidade de carga) – kva/km²
- KVA por km de rede (quantidade de rede necessária, em média, para atender 1 MVA de demanda) – kva/km
- Número de consumidores rurais³⁷ - nconsrur
- Percentual de clientes na área rural (em relação ao total de clientes) - %rural/tot

Variáveis associadas com a dimensão qualidade

- DEC³⁸
- FEC

Com base nestas considerações, foi gerada e analisada uma matriz de correlação³⁹ (quadro 03) com as variáveis obtidas, inclusive com o OPEX. Ressalta-se que a matriz de correlação é um instrumento valioso na indicação das variáveis a serem incluídas, tanto nos modelos estatísticos (AFE), quanto nos modelos de programação linear (DEA) para definição de fronteiras de eficiência. O objetivo é o de sempre primar pelo critério da parcimônia.

A existência de correlações fortes entre *inputs* ou entre *outputs* indica redundância, ou seja, a informação contida em uma variável já está representada por outra. Portanto, seguindo o princípio de buscar um conjunto mínimo de variáveis, os *inputs* e *outputs* redundantes deveriam ser eliminados, a menos que exista alguma justificativa para mantê-las.

Os modelos de análise envoltória de dados, por exemplo, admitem um conjunto de produção convexo e que a tecnologia é monotônica, isto é, um aumento na quantidade de algum *input* não deve resultar em reduções nas quantidades dos *outputs*. Mais uma vez a matriz de correlação é útil, pois uma correlação positiva entre *inputs* e *outputs* indica que a propriedade de tecnologia

³⁷ A inclusão do % de redes rurais no total de redes é justificada pelos custos de atendimento muito maiores que as urbanas.

³⁸ Os índices DEC e FEC refletem a quantidade/qualidade das turmas de operação e manutenção e o estado atual das redes, respectivamente.

³⁹ Este procedimento é muito usado em vários trabalhos podendo-se citar como exemplo Thanassoulis (2002), (2000a), (2000b). Foram usados dados referentes ao ano de 2001.

monotônica está presente. Caso a correlação seja negativa a inclusão de *inputs* e *outputs* negativamente correlacionados deve ser reexaminada.

Quadro 03: Matriz de coeficientes de correlação

Correlations

	OPEX	CONSTOT	CONGRES	CONSNRES	NCONSUM	NCONSNRES	NCONSNRES	NCONSRUR	KMDEREDE	AREA	DEC	FEC	consumid /km2	consumid/km	kva/km2	kv/km	%rural/tot
OPEX	Pearson Correlation Sig. (2-tailed) N	1.000 .940** 60	.942** .940** 60	.939** .938** 60	.946** .943** 60	.943** .943** 60	.916** .912** 60	.563** .643** 60	.643** .643** 60	.074 .572	-.053 .687	-.228 .080	.424** .001	.350** .006	.473** .000	.370** .004	-.150 .252
CONSTOT	Pearson Correlation Sig. (2-tailed) N	.940** .000 60	1.000 .963** 60	1.000** .963** 60	1.000** .932** 60	.932** .927** 60	.912** .927** 60	.528** .600** 60	.600** .600** 60	.021 .874	-.180 .169	-.316* .014	.454** .000	.386** .002	.515** .000	.424** .001	-.143 .276
CONGRES	Pearson Correlation Sig. (2-tailed) N	.942** .000 60	.963** .000 60	1.000 .961** 60	.961** .943** 60	.943** .934** 60	.936** .936** 60	.418** .469** 60	.469** .469** 60	.015 .911	-.144 .272	-.286* .027	.557** .000	.496** .000	.629** .000	.513** .000	-.182 .164
CONSNRES	Pearson Correlation Sig. (2-tailed) N	.939** .000 60	1.000** .000 60	.961** .000 60	1.000** .931** 60	.931** .926** 60	.910** .926** 60	.528** .602** 60	.602** .602** 60	.021 .873	-.181 .167	-.316* .014	.451** .000	.393** .002	.512** .000	.422** .001	-.142 .279
NCONSUM	Pearson Correlation Sig. (2-tailed) N	.946** .000 60	.932** .000 60	.943** .000 60	1.000** .931** 60	.931** .926** 60	.912** .927** 60	.528** .600** 60	.600** .600** 60	.021 .874	-.180 .169	-.316* .014	.454** .000	.386** .002	.515** .000	.424** .001	-.143 .276
NCONSNRES	Pearson Correlation Sig. (2-tailed) N	.943** .000 60	.927** .000 60	.934** .000 60	.926** .998** 60	.998** .944** 60	1.000 .944** 60	.635** .660** 60	.660** .660** 60	.064 .629	-.078 .552	-.278* .032	.373** .003	.332** .009	.448** .000	.371** .003	-.140 .286
NCONSRUR	Pearson Correlation Sig. (2-tailed) N	.916** .000 60	.912** .000 60	.936** .000 60	.910** .964** 60	.944** .944** 60	1.000 .944** 60	.597** .605** 60	.605** .605** 60	.052 .695	-.097 .461	-.249 .055	.510** .000	.381** .003	.494** .000	.333** .009	-.080 .546
NCONSRUR	Pearson Correlation Sig. (2-tailed) N	.563** .000 60	.526** .000 60	.418** .001 60	.528** .633** 60	.633** .635** 60	.597** .597** 60	1.000 .856** 60	.856** .856** 60	.148 .258	.011 .936	-.141 .283	-.166 .206	-.250 .054	-.148 .260	-.140 .288	.222 .068
KMDEREDE	Pearson Correlation Sig. (2-tailed) N	.643** .000 60	.600** .000 60	.469** .000 60	.602** .655** 60	.655** .660** 60	.605** .605** 60	.856** .856** 60	1.000 .856** 60	.179 .172	-.019 .894	-.146 .264	-.105 .426	-.221 .089	-.082 .532	-.095 .470	.092 .482
AREA	Pearson Correlation Sig. (2-tailed) N	.074 .572	.021 .874	.015 .911	.021 .873	.062 .639	.064 .629	.052 .695	.148 .258	.179 .172	1.000 .000	.531** .000	-.565** .315	-.132 .220	-.161 .360	-.120 .319	-.131 .317
DEC	Pearson Correlation Sig. (2-tailed) N	-.053 .687	-.180 .169	-.144 .272	-.181 .167	-.083 .529	-.078 .552	-.097 .461	.011 .936	-.019 .884	1.000 .000	.778** .000	-.155 .237	-.184 .159	-.199 .127	-.215 .099	.062 .641
FEC	Pearson Correlation Sig. (2-tailed) N	-.228 .080	-.316* .014	-.286* .027	-.316* .014	-.275* .034	-.278* .032	-.249 .055	-.141 .283	-.146 .264	.565** .000	.778** .000	1.000 .282	-.141 .177	-.182 .163	-.194 .138	.199 .128
consumid/km2	Pearson Correlation Sig. (2-tailed) N	.424** .001	.454** .000	.557** .000	.451** .001	.404** .003	.373** .000	.510** .206	-.166 .426	-.105 .315	-.132 .237	-.155 .282	1.000 .000	.848** .000	.800** .000	.502** .000	-.294* .023
consumid/km	Pearson Correlation Sig. (2-tailed) N	.350** .006	.396** .002	.496** .000	.393** .002	.345** .007	.332** .009	.381** .003	-.250 .054	-.221 .089	-.161 .220	-.177 .159	.848** .000	1.000 .000	.798** .000	.672** .000	-.489** .000
kva/km2	Pearson Correlation Sig. (2-tailed) N	.473** .000	.515** .000	.629** .000	.512** .000	.461** .000	.448** .000	.494** .000	-.148 .260	-.082 .532	-.120 .360	-.199 .127	.800** .000	.798** .000	1.000 .000	.908** .000	-.278* .031
kv/km	Pearson Correlation Sig. (2-tailed) N	.370** .004	.424** .001	.513** .000	.422** .001	.367** .004	.371** .003	.333** .009	-.140 .288	-.095 .470	-.131 .319	-.215 .099	.502** .000	.672** .000	.908** .000	1.000 .000	-.283* .028
%rural/tot	Pearson Correlation Sig. (2-tailed) N	-.150 .252	-.143 .276	-.182 .164	-.142 .279	-.129 .226	-.140 .286	-.080 .088	.222 .482	.092 .482	.131 .317	.062 .641	.199 .128	-.294* .023	-.489** .000	-.278* .031	1.000 .000

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Os números do Quadro 03 revelam que existe correlação positiva e significativa do OPEX com 12 das 16 variáveis consideradas. As exceções são as variáveis: Área, os indicadores DEC e FEC e o percentual de consumidores rurais, sendo que DEC, FEC e %rural apresentam correlações negativas com OPEX e, portanto, devem ser descartadas.

Entre as variáveis que apresentam correlações positivas e significativas, o consumo e o número de consumidores (total, residencial e não residencial) são as que apresentam as maiores correlações com o OPEX (todas superiores a 0,9). Mas estas seis variáveis também apresentam correlações elevadas entre elas, assim foram selecionadas apenas duas variáveis: o consumo total e o número total de consumidores. Embora as variáveis selecionadas apresentem forte correlação entre elas, as duas variáveis serão mantidas, pois ambas tem uma forte correlação com o OPEX e cada uma delas está associada a um *output* diferente, o número de consumidores (*Nconsum*) é uma *proxy* da quantidade de serviço, enquanto o consumo total (*Constot*) é o produto fornecido.

Com relação às outras variáveis significativas, todas estão associadas com a dimensão dispersão dos consumidores, sendo que a variável km de rede é que apresenta a maior correlação com o OPEX (0,643), portanto é a variável selecionada. Apesar da menor correlação com o OPEX, a consideração da extensão da rede como um *output* é interessante, pois esta variável reflete a dispersão dos consumidores na área de concessão, um fator que não pode ser controlado pela distribuidora. A não inclusão desta variável poderia gerar distorções nas medidas de eficiências das distribuidoras, penalizando as empresas onde os consumidores se encontram dispersos na área de concessão e beneficiando as empresas onde os consumidores se encontram mais concentrados. Um ponto importante é que esta variável apresenta baixa correlação com os dois *outputs* já selecionados e por isso não há redundância.

Em resumo, as **três variáveis selecionadas**⁴⁰ foram **consumo total, número de consumidores e extensão de redes**. A única dimensão não representada foi a qualidade, pois os dois indicadores DEC e FEC apresentaram correlações negativas e não significativas com o OPEX. É importante observar que a não consideração da qualidade como *output* significaria penalizar as concessionárias com melhor esforço de qualidade, pois o acréscimo de custos, devido à qualidade, não encontraria justificativa na análise comparativa de eficiência operacional. Porém, com a consideração da qualidade, estar-se-ia dando um incentivo perverso para reduzir custos associados à qualidade, penalizando em última análise o bem estar do consumidor. Portanto, para inserir a qualidade explicitamente no modelo de análise da eficiência operacional, as despesas operacionais exclusivas das ações de qualidade deveriam ser previamente deduzidas do OPEX de todas as empresas, o que é impraticável visto que obrigaria a decompor a quantidade de cada *input*⁴¹.

⁴⁰ Estas variáveis satisfazem ainda a propriedade de tecnologia monotônica, qualificando-as como excelentes candidatas a *outputs* do modelo DEA (que também foi implementado na tese).

⁴¹ Portanto, a dimensão qualidade não será considerada nos modelos de definição de fronteira apresentados na tese.

8.2.

Encontrando grupos de empresas similares utilizando o SOM

A adoção dos *Self Organizing Maps* (SOM) no presente trabalho será feita a partir das 3 variáveis, selecionadas e abordadas na seção anterior, e obtidas para as 60 concessionárias de energia elétrica. O método possibilitará a agregação destas em padrões típicos. As variáveis⁴² escolhidas foram:

- 1) Consumo Total (MWh)
- 2) Número de Consumidores
- 3) km de rede

A análise será feita em duas etapas (“análise de sensibilidade”):

- 1) Apresentar-se-á para a Rede Neural uma matriz de dados de tamanho 60 x 2, isto é, serão apresentados os dados de consumo e números de consumidores.
- 2) Em um segundo momento, será incluída no vetor de entrada, a variável de km de rede, formando portanto um vetor 60 x 3.

Estes espaços de entrada serão projetados em um espaço de menor dimensão (espaço de saída). O mapa escolhido inicialmente foi um bidimensional composto por 6 neurônios⁴³ dispostos como representou-se na figura 19 abaixo (topologia do mapa).

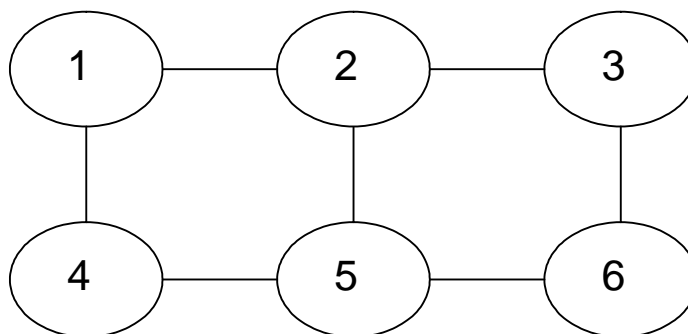


Figura 19: Topologia da rede

42 Para evitar problemas de escala das variáveis, foi feita uma padronização de todas elas usando-se a média e o desvio padrão.

43 Como dito na seção 7.1, em uma análise de *clusters*, a quantidade de neurônios (grupos) é sempre determinado pelo analista, ou seja, cabe a este definir quantos padrões previamente se quer obter. Em geral, faz-se uma abordagem *top-down*, ou seja, inicia-se com um número maior de neurônios e analisa-se a possibilidade de redução para um número menor de grupos (com base nas características dos grupos e conceitos de vizinhança na topologia da rede). Este procedimento foi adotado antes da implementação do PSOM (*Pruning Self Organizing Maps*).

Como dito, uma característica importante do algoritmo de Kohonen é que a topologia no espaço de saída reflete características semelhantes no espaço original, ou seja, entradas com características parecidas serão mapeadas em regiões próximas no espaço de saída. Isto significa que, a princípio, serão obtidos seis grupos de concessionárias. Cada neurônio se especializa em responder a estímulos (lembrar que cada concessionária será representada por um vetor de entrada) pertencentes a uma mesma classe.

Matematicamente cada neurônio receberá então um vetor que será comparado a um vetor inicializado previamente para cada neurônio. Compara-se então o vetor de entrada com o vetor de pesos inicializado previamente, sendo que o neurônio que tiver o vetor de peso mais próximo do vetor de entrada é denominado de neurônio vencedor. O treinamento da rede é realizado apresentando-se a matriz de entrada um grande número de vezes⁴⁴ até que cada neurônio se especialize, ou seja, represente um dado agrupamento de concessionárias. Durante a execução do algoritmo, ajustes nos vetores de peso iniciais são feitos. Ao final, a cada neurônio no espaço de saída corresponde um vetor-peso que pode ser visto como um protótipo de características do espaço original.

8.2.1. Duas Variáveis no Vetor de Entrada

Foram obtidos também 6 agrupamentos que são representados por um protótipo. Importante salientar que fez-se uma redução de dimensionalidade de uma matriz de 60 x 2 para 6 x 2. A análise será feita, a partir deste momento, para seis concessionárias típicas. Os protótipos (vetor representativo)⁴⁵ de cada grupo podem ser visualizados no quadro 04 a seguir.

⁴⁴ Refere-se ao número de épocas da rede neural. Neste trabalho foram utilizadas 50.000 épocas (apresentações).

⁴⁵ Importante: Não se refere ao centróide de cada grupo.

Quadro 04: Protótipos

		Protótipo	
		1	2
Neurônios	1	1.457	294
	2	7.656	1.390
	3	14.839	2.585
	4	637	169
	5	2.497	618
	6	9.370	1.939

Nota: valores em mil.

Analisando o quadro 04, vê-se a caracterização de cada um dos grupos. É importante analisar o protótipo juntamente com a visualização da disposição dos neurônios no mapa. Atenta-se também que o protótipo, como o próprio nome já diz, traz uma “representação em linhas gerais” do neurônio. Os valores para cada agrupamento representam o último vetor ajustado pelo algoritmo de Kohonen podendo ser utilizado para caracterizar, em uma primeira análise, cada padrão. Isto é, o protótipo deve ser usado apenas para entender como os grupos foram caracterizados na topologia do mapa.

Feitas estas considerações, verifica-se que o neurônio 3 apresenta os valores maiores para consumo total e número de consumidores. Similar ao neurônio 3, tem-se o neurônio 6, com o detalhe que este apresenta menor valor para consumidores. Similar aos neurônios 3 e 6, tem-se o neurônio 2.

Coerentemente, no extremo oposto do Mapa, vemos características também opostas. O neurônio 4, por exemplo, apresenta valores menores para consumo total e número de consumidores. Os neurônios 1 e 5, por sua vez, são também similares com o fato de que o neurônio 5 apresenta maiores valores para as variáveis consideradas.

No quadro 05 a seguir, vê-se a distribuição das empresas nos neurônios.

Quadro 05: Distribuição de concessionárias por grupo (utilizando duas variáveis no vetor de entrada)

GRUPO	NÚMERO DE CONCESSIONÁRIAS	%
Grupo 1	4	6,7
Grupo 2	7	11,7
Grupo 3	5	8,3
Grupo 4	29	48,3
Grupo 5	9	15,0
Grupo 6	6	10,0
TOTAL	60	100,0

Já no quadro 06 a seguir, vê-se a distribuição nominal das empresas nos neurônios:

Quadro 06: Alocação das empresas nos neurônios

CONCESSIONARIA	Neurônios					
	1	2	3	4	5	6
AES-SUL		x				
BANDEIRANTE			x			
BOA VISTA				x		
BRAGANTINA				x		
CAIUA				x		
CAT-LEO	x					
CEA				x		
CEAL					x	
CEB					x	
CEEE		x				
CELB				x		
CELESC						x
CELG		x				
CELPA					x	
CELPE						x
CELTINS				x		
CEMAR					x	
CEMAT					x	
CEMIG			x			
CENF				x		
CEPISA					x	
CERJ						x
CERON	x					
CFLO				x		
CHESP				x		
COCEL				x		
COELBA						x
COELCE		x				
COPEL						x
COSERN					x	
CPEE				x		
CPFL			x			
CSPE				x		
DEMEI				x		
ELEKTRO						x
ELETROACRE				x		
ELETROCAR				x		
ELETROPAULO			x			
ENERGIPE	x					
ENERSUL					x	
ESCELSA		x				
JAGUARI				x		
JOÃO CESA				x		
LIGHT			x			
MANAUS	x					
MOCOCA				x		
MUXFELDT				x		
NACIONAL				x		
NOVA PALMA				x		
PANAMBI				x		
PIRATININGA		x				
POÇOS DE CALDAS				x		
RGE		x				
SAELPA					x	
SANTA CRUZ				x		
SANTA MARIA				x		
SULGIPE				x		
URUSSANGA				x		
V. PARANAPANEMA				x		
XANXERÊ				x		
Total	4	7	5	29	9	6

Obviamente, se as empresas foram alocadas em neurônios distintos é porque elas têm particularidades que justificam esta alocação. Entretanto, se houver necessidade de se compactar mais os grupos obtidos, uma junção possível, **ignorando-se todas as dissimilaridades levantadas anteriormente**, é proposta a seguir, ou seja, analisando os protótipos, poder-se-ia fazer uma “separação” do Mapa em 3 grupos (neurônio 4, neurônios 1 e 5; neurônios 2, 3 e 6).

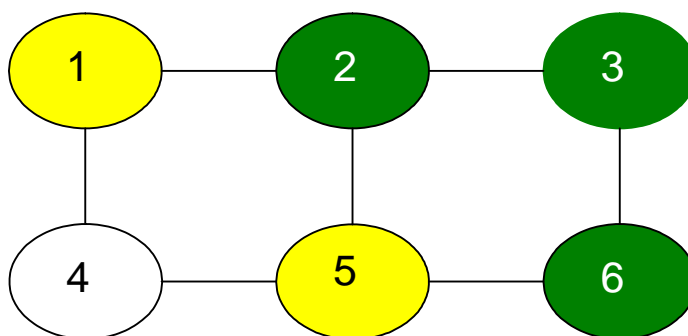


Figura 20: Redução para três grupos

8.2.2. Três Variáveis no Vetor de Entrada

Como dito anteriormente, a princípio foram obtidos também 6 agrupamentos que são representados por um protótipo. Importante salientar que fez-se uma redução de dimensionalidade de uma matriz de 60 x 3 para 6 x 3. A análise será feita novamente, a partir deste momento, para seis concessionárias típicas. Os protótipos (vetor representativo) de cada grupo podem ser visualizados no quadro 07 a seguir.

Quadro 07: Protótipos

		Protótipo		
		1	2	3
Neurônios	1	1.848	300	7
	2	8.292	1.428	34
	3	14.014	2.549	70
	4	601	161	5
	5	2.698	599	17
	6	8.719	1.904	60

Nota: Valores em mil.

Analisando o quadro 07, vê-se a caracterização de cada um dos grupos. Verifica-se que o neurônio 3 apresenta os valores maiores para consumo total, consumo número de consumidores e km de rede. Similar ao neurônio 3, tem-se o neurônio 6, com o detalhe que este apresenta valor menor para km de rede. Similar aos neurônios 3 e 6, tem-se o neurônio 2, com a especificidade deste apresentar valor menor ainda para consumidores e km de rede.

Coerentemente, no extremo oposto do Mapa, vemos características também opostas. O neurônio 4, por exemplo, apresenta valores menores para consumo total, número de consumidores e km de rede. Os neurônios 1 e 5, por sua vez, são bastante similares com o detalhe de que o 1 apresenta valores menores para consumo, consumidores e km de rede.

No quadro 08 a seguir, vê-se a distribuição das empresas nos neurônios.

Quadro 08: Distribuição de concessionárias por grupo (utilizando três variáveis no vetor de entrada)

GRUPO	NÚMERO DE CONCESSIONÁRIAS	%
Grupo 1	4	6,7
Grupo 2	7	11,7
Grupo 3	5	8,3
Grupo 4	29	48,3
Grupo 5	8	13,3
Grupo 6	7	11,7
TOTAL	60	100,0

Já no quadro 09 a seguir, vê-se a distribuição nominal das empresas nos neurônios:

Quadro 09: Alocação das empresas nos neurônios

CONCESSIONARIA	Neurônios					
	1	2	3	4	5	6
AES-SUL		x				
BANDEIRANTE		x				
BOA VISTA				x		
BRAGANTINA				x		
CAIUA				x		
CAT-LEO	x					
CEA				x		
CEAL					x	
CEB					x	
CEEE		x				
CELB				x		
CELESC						x
CELG						x
CELPA					x	
CELPE						x
CELTINS				x		
CEMAR					x	
CEMAT					x	
CEMIG			x			
CENF				x		
CEPISA					x	
CERJ		x				
CERON	x					
CFLO				x		
CHESP				x		
COCEL				x		
COELBA						x
COELCE						x
COPEL			x			
COSERN					x	
CPEE				x		
CPFL			x			
CSPE				x		
DEMEI				x		
ELEKTRO						x
ELETROACRE				x		
ELETROCAR				x		
ELETROPAULO			x			
ENERGIPE	x					
ENERSUL						x
ESCELSA		x				
JAGUARI				x		
JOÃO CESA				x		
LIGHT			x			
MANAUS	x					
MOCOCA				x		
MUXFELDT				x		
NACIONAL				x		
NOVA PALMA				x		
PANAMBI				x		
PIRATININGA		x				
POÇOS DE CALDAS				x		
RGE		x				
SAELPA					x	
SANTA CRUZ				x		
SANTA MARIA				x		
SULGIPE				x		
URUSSANGA				x		
V. PARANAPANEMA				x		
XANXERÉ				x		
Total	4	7	5	29	8	7

Pensando novamente em uma redução dos grupos, novamente as diferenças peculiares serão relevadas podendo-se, desta forma chegar em uma configuração com 3 grupos (neurônio 4, neurônios 1 e 5; neurônios 2, 3 e 6).

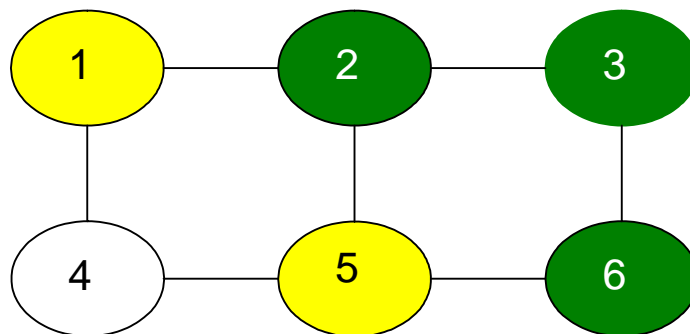


Figura 21: Redução para três grupos

Como se sabe, esta redução do número de grupos é feita através da análise dos protótipos. Neste caso, como são analisadas apenas três variáveis, é fácil organizar os protótipos em um diagrama de dispersão tridimensional, conforme apresentada na figura 22, e assim visualizá-los.

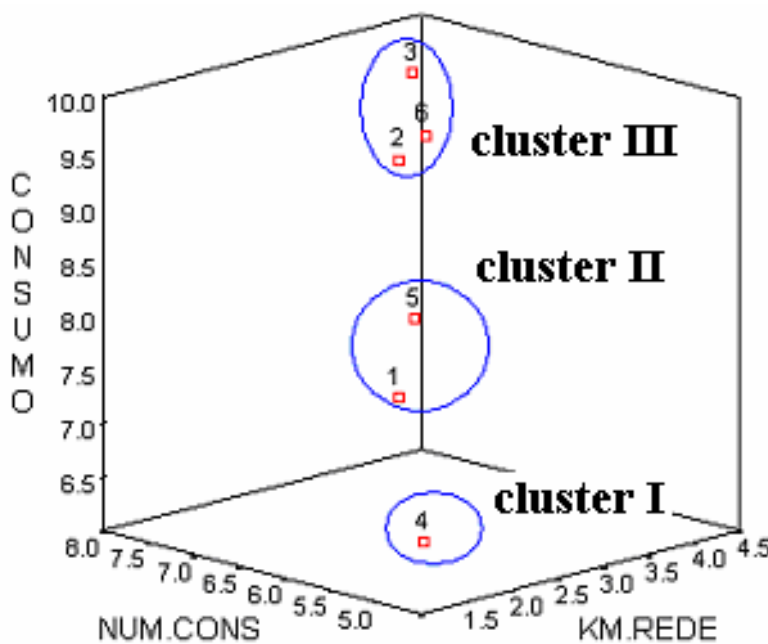


Figura 22: Diagrama de dispersão dos protótipos (construído a partir dos logaritmos das coordenadas apresentadas no quadro 7)

A partir da figura 22, é fácil observar as semelhanças entre alguns protótipos e que os 6 protótipos podem ser agregados para formar 3 agrupamentos, cujas composições são apresentadas na tabela 4.

Tabela 4 – *Clusters* de distribuidoras

Cluster I	Cluster II	Cluster III
29 distribuidoras	12 distribuidoras	19 empresas
BOA VISTA	CAT- LEO	AES-SUL
BRAGANTINA	GEAL	BANDEIRANTES
CAIUÁ	CEB	CEEE
CEA	CELPA	CELESC
CELB	CEMAR	CELG
CELTINS	CEMAT	CELPE
CENF	CEPISA	CEMIG
CFLO	CERON	CERJ
CHESP	COSERN	COELBA
COCEL	ENERGIPE	COELCE
CPEE	MANAUS	COPEL
CSPE	SAELPA	CPFL
DEMEI		ELEKTRO
ELETROACRE		ELETROPAULO
ELETROCAR		ENERSUL
JAGUARI		ESCELSA
JOÃO CESA		LIGHT
MOCOCA		PIRATININGA
MUXFELDT		RGE
NACIONAL		
NOVA PALMA		
PANAMBI		
POÇOS DE CALDAS		
SANTA CRUZ		
SANTA MARIA		
SULGIPE		
URUSSANGA		
V. PARANAPANEMA		
XANXERÊ		

A divisão das empresas em três agrupamentos é conveniente, pois permite associar os agrupamentos com **empresas de pequeno (*cluster I*), médio (*cluster II*) e grande (*cluster III*) porte**. Desta forma, respeita-se um importante aspecto para a definição de fronteiras de eficiência, e que diz respeito às dissimilaridades entre as empresas distribuidoras de energia. Isto é, as heterogeneidades deste mercado no Brasil, reforçam a necessidade de que se faça uma análise de reconhecimento de padrões antes de se definir uma fronteira de eficiência. Portanto, identificados os grupos de similaridade, é que serão estimadas as fronteiras de eficiência. Como serão usados modelos econométricos, é muito

importante que se evite problemas de micronumerosidade, ou seja, “poucos” dados para “muitos” parâmetros⁴⁶.

Em síntese, uma vantagem da divisão do conjunto de empresas em três agrupamentos é que nesta solução são obtidos agrupamentos com quantidade de empresas suficientes para aplicação de técnicas paramétricas e não paramétricas⁴⁷.

8.2.3. Resultados da Análise de Sensibilidade

A análise de *clusters* feita via Redes Neurais Artificiais indica alguns resultados importantes. Viu-se que a análise se deu através de uma estratégia *bottom-up*, isto é, partiu-se de um vetor de entrada com 2 variáveis e adicionou-se mais 1 variável em momentos distintos.

Observou-se, entretanto, que determinadas empresas sempre eram alocadas em um mesmo neurônio ou em neurônios próximos, mantendo as mesmas características gerais. A migração das empresas entre os grupos foi reduzida ao se considerar a inclusão da terceira variável (km de rede) no vetor característico. No quadro 10 a seguir pode-se visualizar o deslocamento das empresas nos neurônios.

Quadro 10: Deslocamento das empresas com a inclusão da variável km de rede

Neurônio	Entraram	Origem (Neurônio)	Saíram	Destino (Neurônio)
1	-	-	-	-
2	BANDEIRANTE	3	CELG	6
	CERJ	6	COELCE	6
3	COPEL	6	BANDEIRANTE	2
4	-	-	-	-
5	-	-	ENERSUL	6
6	CELG	2	CERJ	2
	COELCE	2	COPEL	3
	ENERSUL	5		

Nota: Observar que, com exceção da ENERSUL, os demais deslocamentos se deram dentro de um mesmo padrão.

⁴⁶ Atente-se que foram selecionadas três variáveis para as análises e, trabalhando-se com seis neurônios, existem grupos com quatro empresas (ver quadro 09).

⁴⁷ Na formulação de modelos DEA, por exemplo, recomenda-se que o número de empresas comparadas seja no mínimo o dobro do número de variáveis especificadas, uma condição que é satisfeita pelos três agrupamentos, pois neste caso o dobro de *inputs* mais *outputs* é 8, e o menor *cluster* tem 12 concessionárias. É fácil perceber que na solução com seis agrupamentos esta condição não é satisfeita em alguns agrupamentos.

Neste momento é importante ressaltar um aspecto interessante da análise via Redes Neurais Artificiais e a consistência dos resultados. É importante então atentar para a leitura do Mapa de Kohonen. Uma das grandes vantagens desta metodologia perante às outras é que a topologia no espaço de saída reflete as características originais. Isto é, empresas que são próximas no espaço de entrada (de maior dimensão o que, por sua vez, dificulta a visualização) aparecerão próximas no espaço de saída (de fácil visualização por ser de menor dimensão).

Portanto, em analogia à distância cartográfica, sabemos que as empresas que caíram no neurônio 1, por exemplo, são completamente diferentes daquelas que foram alocadas no neurônio 6. O neurônio 3 por sua vez é próximo ao 6 assim como o 1 é similar ao 4. Outro detalhe importante: mantendo-se as variáveis de entrada, o fato de uma empresa ser alocada no neurônio 1 em uma etapa, e depois ser alocada no neurônio 6 em outra fase, não traz problemas desde que toda a configuração do mapa mude. Isto é, empresas que estavam juntas precisam se deslocar em bloco (é como se estivesse rotacionando todo o mapa). Esta característica implica no fato de que os neurônios não devem ser vistos como caracterizadores de padrões fixos, ou seja, é importante sempre analisar qual o padrão foi captado por ele a partir de quais empresas nele foram alocadas.

Esta característica da análise via Rede Neural mostra ainda um resultado importante obtido no problema em estudo. Apesar da alteração das variáveis de entrada, os padrões (grupos de empresas similares) encontrados se mantiveram⁴⁸. Em síntese, a análise de sensibilidade mostrou uma consistência satisfatória para os padrões obtidos.

8.3. Encontrando grupos de empresas similares utilizando o PSOM

Como apresentado na seção 7.1.1, este trabalho de tese implementou uma técnica para auxiliar na determinação do número de padrões (ou grupos) a serem trabalhados para a definição das fronteiras de eficiência. O objetivo é, no treinamento da rede neural, achar e “podar” neurônios que não representam um número pré-especificado de padrões.

⁴⁸ Obviamente com apenas algumas poucas alterações.

O PSOM foi implementado para topologias unidimensionais (ver figura 15 no capítulo 5). Desta forma, implementou-se o PSOM partindo-se de uma topologia unidimensional composta por 12 neurônios⁴⁹ (m : quantidade inicial de neurônios ou unidades de saída). Definiu-se então os demais parâmetros de aprendizagem:

- N_c : região de vizinhança
- α : taxa inicial de aprendizagem
- α_{\min} : taxa final de aprendizagem
- χ : fator de decrescimento da taxa de aprendizagem
- \max_{ep} : número máximo de épocas
- ξ : pruning threshold
- $Inic$: intervalo de inicialização do vetor de pesos
- τ : termo de penalidade

Um resultado muito satisfatório foi que, implementado o PSOM, reduziu-se o número de grupos de 6 para 3 neurônios, isto é, o mesmo número de padrões identificados pela heurística *top-down* apresentada anteriormente. Aconteceram apenas algumas migrações de poucas empresas nos grupos apresentados na tabela 4. A nova conjugação dos grupos é apresentada na tabela 5 a seguir. Pode-se observar que apenas 6 empresas (portanto apenas 10% do universo trabalhado) se deslocaram entre os padrões, migrando todas para o grupo de “empresas de porte médio”. Uma das empresas que se deslocaram foi a Celtins, que veio do grupo de “empresas menores”. As outras 5 empresas são: AES-Sul, Enersul, Escelsa, Piratininga e RGE (que vieram do grupo de “empresas de porte grande”). Ressalta-se que este fato não constitui problema algum tendo em vista que, como dito anteriormente, os Mapas de Kohonen funcionam como um mapa cartográfico mesmo, isto é, empresas que caíram no neurônio 6 são próximas das do neurônio 3 (considerando, por exemplo, a figura 19) e diferentes das do neurônio 1. Em síntese, trabalha-se a similaridade entre os objetos (empresas) e também a proximidade entre os neurônios (grupos). Feitas estas considerações, ressalta-se

⁴⁹ Existe grande diversidade entre as opiniões de especialistas do setor elétrico sobre a quantidade de “padrões” existentes entre as empresas. Conversa com estes especialistas indicam um número que vai de 3 a 10 “tipos” de empresas. Optou-se então por definir $m = 12$ (o dobro de neurônios utilizados na análise *top-down* e mais próximo do número máximo de tipos sobre que se tem idéia).

que estas poucas empresas que se deslocaram encontravam-se em neurônios próximos na topologia antes de se executar o PSOM. Outra vantagem de se utilizar a técnica de *pruning* é que o número “adequado” de neurônios é obtido como resultado do próprio treinamento da rede (ajustes dos vetores de pesos) e não feita, a posteriori, pela análise de protótipos.

Tabela 5 – *Clusters* de distribuidoras após a implementação do PSOM

Cluster I	Cluster II	Cluster III
28 distribuidoras	18 distribuidoras	14 distribuidoras
BOA VISTA	AES-SUL	BANDEIRANTES
BRAGANTINA	CAT- LEO	CEEE
CAIUÁ	CEAL	CELESC
CEA	CEB	CELG
CELB	CELPA	CELPE
CENF	CEMAR	CEMIG
CFLO	CELTINS	CERJ
CHESP	CEMAT	COELBA
COCEL	CEPISA	COELCE
CPEE	CERON	COPEL
CSPE	COSERN	CPFL
DEMEI	ENERGIPE	ELEKTRO
ELETROACRE	ENERSUL	ELETROPAULO
ELETROCAR	ESCELSA	LIGHT
JAGUARI	MANAUS	
JOÃO CESA	PIRATININGA	
MOCOCA	RGE	
MUXFELDT	SAELPA	
NACIONAL		
NOVA PALMA		
PANAMBI		
POÇOS DE CALDAS		
SANTA CRUZ		
SANTA MARIA		
SULGIPE		
URUSSANGA		
V. PARANAPANEMA		
XANXERÊ		

8.4. Estimando os Modelos de Fronteira Estocástica

8.4.1. Análise da Hipótese de Retornos de Escala Constante

De forma a examinar se a hipótese de CRS é sustentável no conjunto de todas as concessionárias, há que se analisar o comportamento dos custos em função do consumo total. Como é sabido, a existência de retornos de escala constantes pode ser representada (ver Figura 23) por uma linha de custos médios horizontal, tendo no eixo das abcissas a quantidade produzida (o ‘consumo total’, no caso presente). Por sua vez, retornos de escala crescentes (IRS) corresponderão a um declive negativo da mesma função de custos, acontecendo o inverso no caso de retornos de escala decrescentes (DRS). O modelo VRS usual corresponde à curva em U desenhada na Figura 23, em que a uma fase inicial de custos decrescentes (de economias com o aumento de escala), se sucedem custos crescentes (deseconomias), fruto do aumento da complexidade no gerenciamento das unidades produtivas de grande dimensão. Nesta configuração, a dimensão ótima das empresas situa-se obviamente à volta do mínimo da curva.

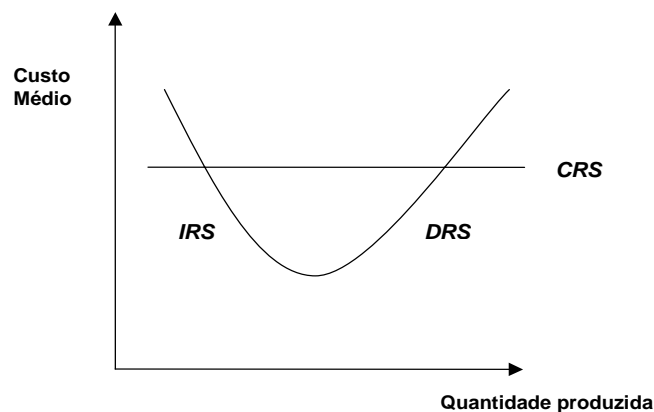


Figura 23: Retornos de escala

Como se comportam, então, os custos operacionais unitários (custos médios expressos em Reais por MWh de eletricidade distribuída, integrando exclusivamente as despesas compreendidas em OPEX: pessoal, materiais e serviços de terceiros) no caso das 60 empresas estudadas? Para ajudar nesta análise, observe-se na tabela 6, os valores médios destes custos, por escalas de 10

empresas, devidamente ordenadas por ordem crescente de acordo com a sua dimensão de operação (consumo total).

Tabela 6 – Média dos Custos Operacionais Unitários (\$R/MWh) por escalas de 10 empresas ordenadas pelo seu consumo total.

Empresas	Custo médio
Emp 1-10	31,43
Emp 11-20	32,87
Emp 21-30	37,90
Emp 31-40	35,03
Emp 41-50	30,72
Emp 51-60	18,76
Média 1-60	31,12

Pode-se observar que esta série de valores é crescente inicialmente, passando depois a decrescer. Este comportamento parece sugerir que se está perante a retornos variáveis de escala. Contudo, há que ser prudente nesta primeira análise e atentar em dois importantes fatos:

i) Em primeiro lugar, é de referir que as médias do segundo e terceiro escalões são fortemente influenciadas por um valor exorbitante de custos em cada um deles (respectivamente 70,5 \$R/MWh da empresa Boa Vista, no segundo escalão, e 65,55 \$R/MWh da empresa Celtins, no terceiro). Sem esses valores, as médias baixam para 28,69 \$R/MWh e 34,83 \$R/MWh, passando a aproximar-se da média geral e, sobretudo, passam a traduzir uma oscilação em termos de tendência. Por sua vez, também os dois últimos escalões vêm os seus valores influenciados (agora negativamente) pelos dois valores extremos mais baixos (respectivamente 9,9 \$R/MWh da AES-SUL e 10,8 \$R/MWh da Piratininga). Há pois (mesmo tendo-se retirado da análise a rubrica de “outros custos”) uma grande dispersão dos valores dos custos operacionais médios por MWh. A relação entre o máximo e o mínimo dos custos unitários é ainda de 7,15, passando a cerca de 4 (4,25) quando se retiram os quatro referidos valores extremos (os dois maiores e os dois menores). Concomitantemente, registram-se grandes diferenças entre empresas com consumos próximos. Isso é visível na Figura 24, que inclui também o ajustamento de uma série de médias móveis de passo 10, a fim de melhor se visualizar o andamento tendencial dos valores do custo médio.

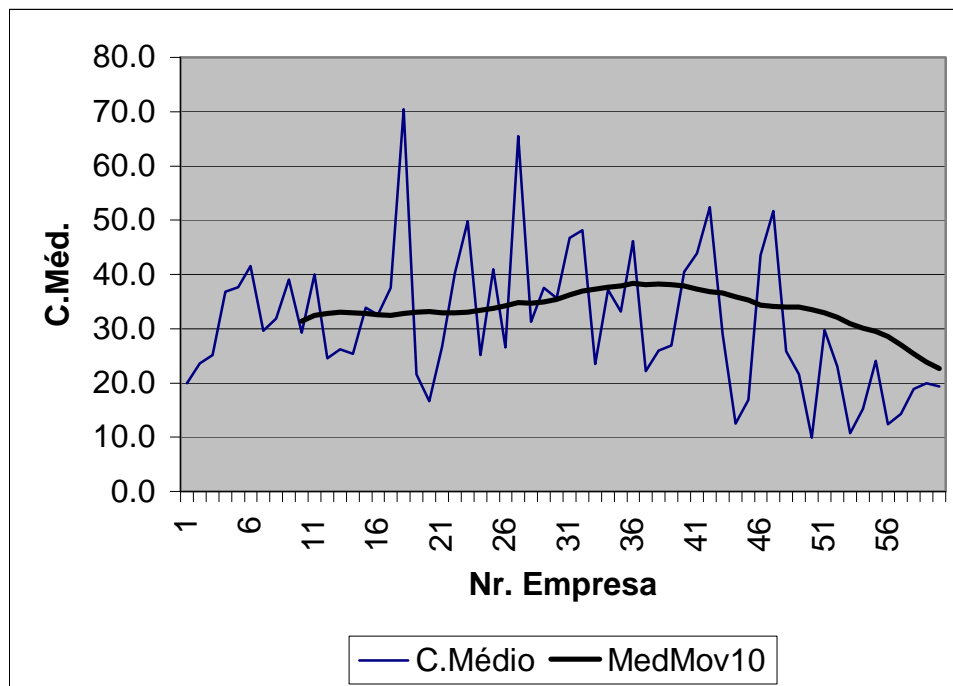


Figura 24: Custos operacionais por MWh das empresas ordenadas pelo seu consumo total

ii) É precisamente a observação desta curva de médias móveis que motiva a segunda reflexão. É que, independentemente da oscilação da mesma, esta não traduz a realidade retratada pelo modelo VRS usual. Como dito anteriormente, contempla a usual situação em que os retornos de escala decrescentes antecedem retornos crescentes. Ora, aqui temos antes uma situação mais próxima do contrário (um **U** invertido), em que os valores mais altos de custos se concentram sobretudo nas empresas de média dimensão.

Feita esta análise prévia, atenta-se agora para a análise do custo médio em função dos consumos totais, de forma a se testar a hipótese de retornos de escala constantes ilustrada na Figura 23. Para isso, analise-se a Figura 25, na qual é ainda apresentado o resultado do ajustamento de uma tendência linear.

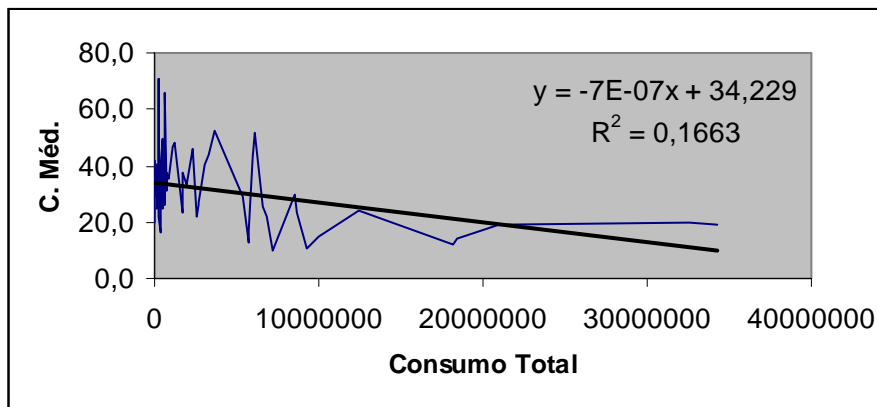


Figura 25: Custos operacionais por MWh em função do consumo total

Da análise da Figura 25 várias conclusões ressaltam. Em primeiro lugar, a difícil visualização dos valores das diferentes empresas, uma vez que há uma enorme diferença na escala do eixo horizontal, em que, por exemplo, o valor do consumo total da maior empresa é mais de 1.700 vezes maior do que o da menor.

Em segundo lugar, resalta da série de valores que qualquer ajustamento de uma função de tendência terá uma baixa aderência aos pontos da série, dada a enorme volatilidade da mesma, conforme já antes referido. Isso justifica o baixo valor encontrado para o indicador R^2 (0,166), o qual não sofre melhorias dignas de registro com o ajustamento de funções não lineares. Ainda assim, há que se inferir sobre o sinal da respectiva tendência, atendendo ao valor do respectivo coeficiente de regressão. Pela análise efetuada parece ser de concluir pela negatividade da tendência (-7E-07), influenciada pelo peso em termos de escala do grupo das maiores empresas, o qual, como já dito, apresenta valores médios de custos unitários inferiores ao conjunto das restantes empresas.

Contudo, sendo esse fato inegável, há que registrar o baixo valor (-3,4) da estatística de t-Student associada ao coeficiente de regressão relativo à mesma tendência, o qual passa no entanto a ser não significativamente diferente de zero (a um nível de confiança estatística de 95%) quando se retiram as 8 maiores empresas (empresas acima de 9 milhões de MWh de consumo). Desta forma, a negatividade na tendência geral dos custos unitários, a que se poderia associar retornos de escala crescentes, tem pouca aderência (em termos de tendência local) para as empresas de menor dimensão, como vimos anteriormente.

Em conclusão, a análise efetuada aponta para que seja válido adotar modelos de definição de *frontier benchmarking* sob a hipótese de retornos de escala constantes⁵⁰.

8.4.2. Tratamento de Pontos Discrepantes e Estimação dos Modelos de Fronteira Estocástica

Para a especificação dos modelos de fronteira estocástica nesta tese, foi analisada uma *cross-section* com 60 concessionárias (*Decision Making Units: DMU's*), sendo que cada uma delas é descrita por quatro variáveis: custo operacional (OPEX), montante de energia distribuída, número de consumidores atendidos e extensão total da rede de distribuição. Porém, antes de ajustar a equação de regressão (fronteira) deve-se garantir que as DMU's analisadas sejam comparáveis. Por esta razão fez-se uma análise exploratória nos dados, visando verificar a existência de empresas com dimensões atípicas (*outliers*) em relação ao conjunto das 60 empresas. Os resultados desta análise preliminar indicaram a necessidade de incluir algumas variáveis *dummies* nos modelos que estimam a fronteira de eficiência.

A seguir, os gráficos das figuras 26, 27 e 28 mostram os diagramas de dispersão do logaritmo das variáveis: número de consumidores, consumo e extensão de redes. Em cada gráfico, os quadrados que envolvem os pontos extremos indicam as concessionárias consideradas atípicas.

⁵⁰ Para fins de comparação, mesmo com este resultado da análise sobre os retornos de escala, serão estimadas fronteiras também com retornos de escala variáveis.

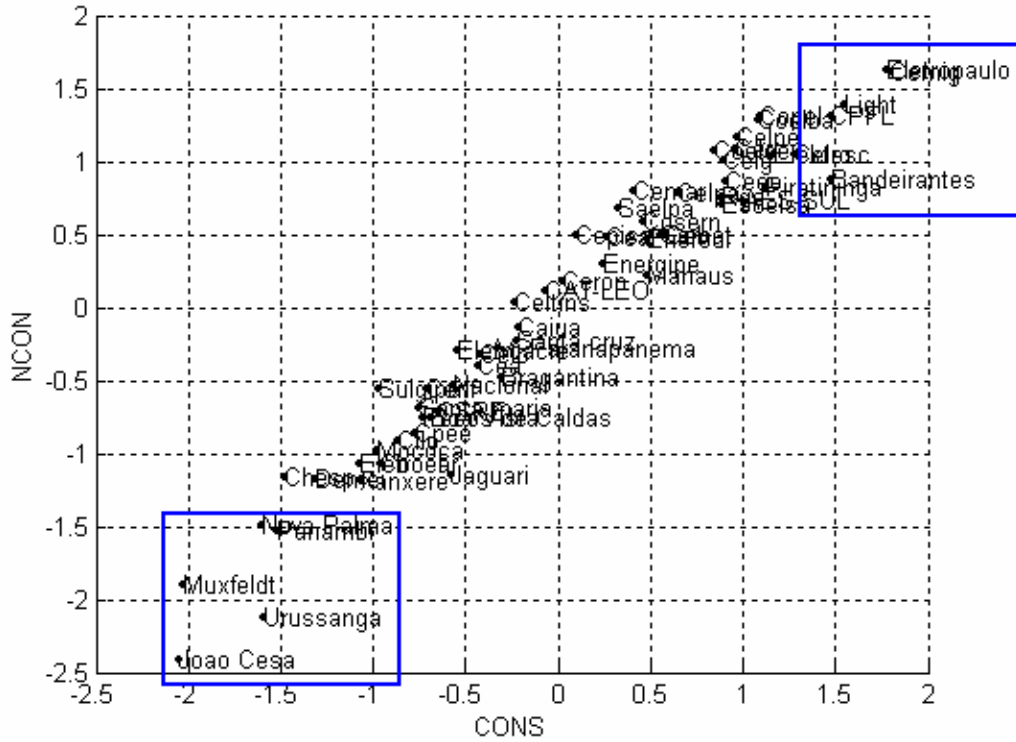


Figura 26: Diagrama de dispersão das variáveis número de consumidores (NCON) e energia distribuída (CONS)

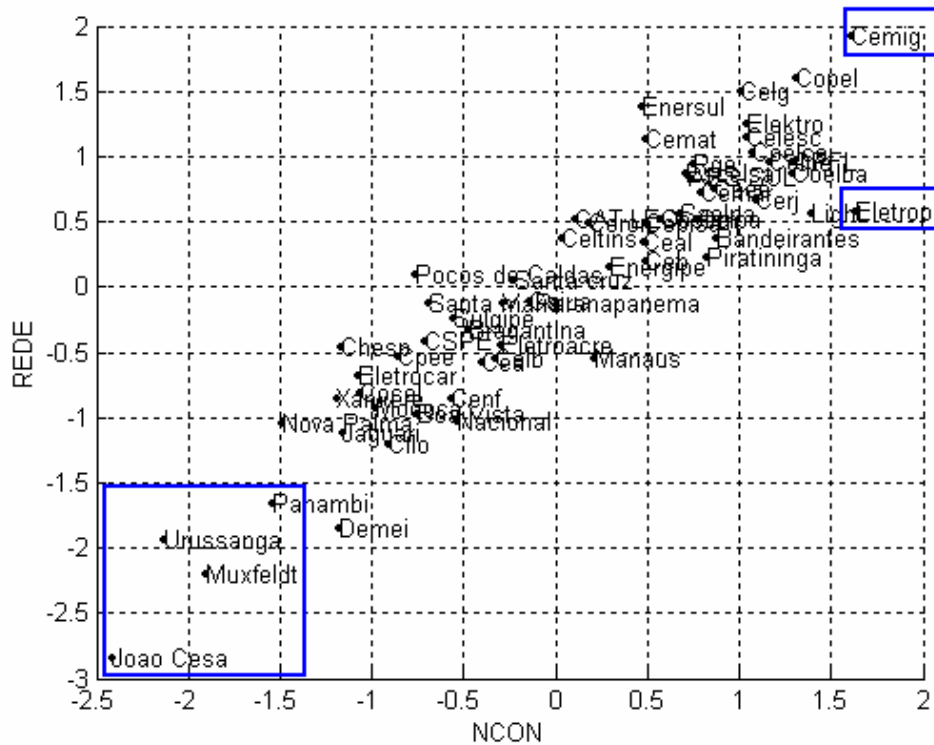


Figura 27: Diagrama de dispersão das variáveis extensão de rede (REDE) e número de consumidores (NCON)

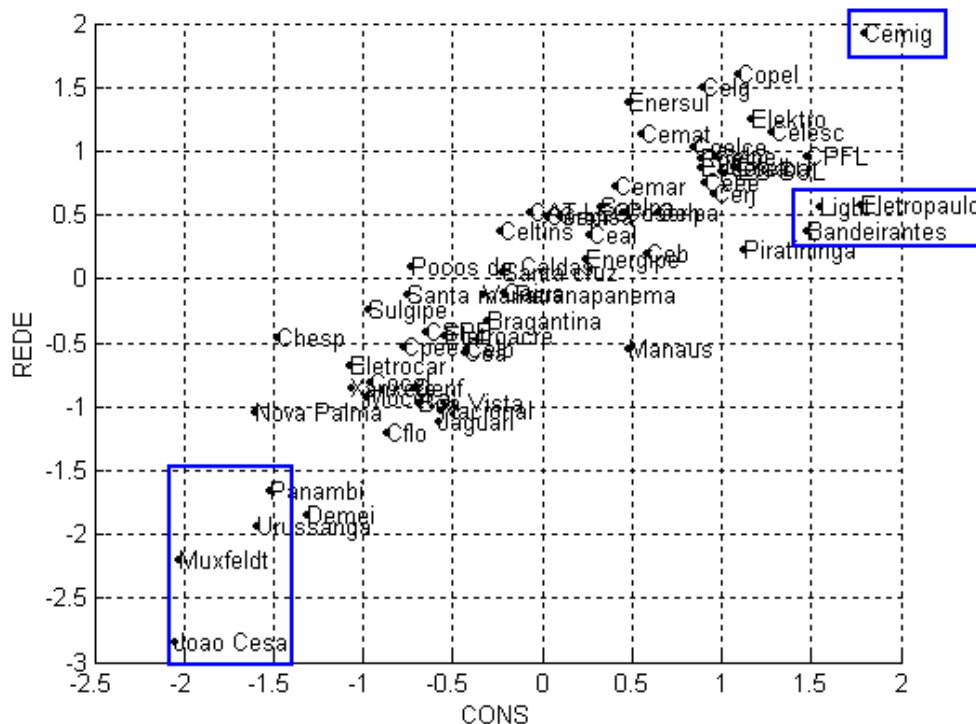


Figura 28: Diagrama de dispersão das variáveis extensão de rede (REDE) e energia distribuida (CONS)

Uma outra forma utilizada para identificar possíveis concessionárias atípicas foi a análise do gráfico qui-quadrado⁵¹, apresentado na figura 29, onde as empresas indicadas são aquelas onde a distância generalizada⁵² é maior que o percentil de 85% de uma distribuição qui-quadrado com três graus de liberdade. Este gráfico confirma algumas das concessionárias identificadas na análise anterior.

⁵¹ Johnson & Wichern (1998).

⁵² Distância generalizada $= (x_i - \bar{X})^T S^{-1} (x_i - \bar{X})$, onde x_i é o vetor da i -ésima concessionária, para todo $i=1$ a 60.

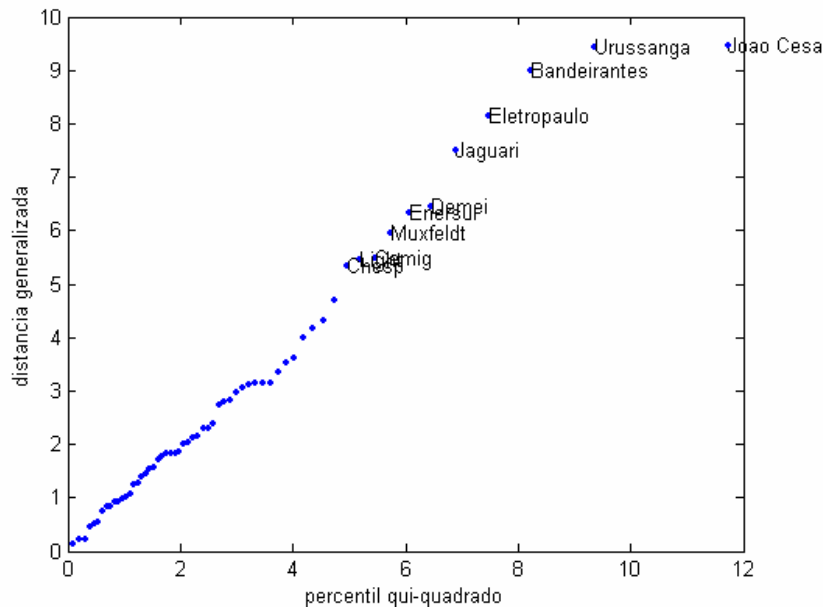


Figura 29: Gráfico Qui-Quadrado

A partir das concessionárias sugeridas pelos gráficos das figuras 26 a 29, foram testadas diversas possibilidades de inclusão das variáveis *dummies* com o objetivo de capturar as especificidades das concessionárias atípicas e assim não contaminar os indicadores de eficiência. Entre as diversas possibilidades testadas, a melhor combinação encontrada é formada pelas quatro variáveis *dummies* a seguir:

- D_1 = variável *dummy* que assume o valor 1 nas empresas CPFL, Eletropaulo e Light
- D_2 = variável *dummy* que assume o valor 1 apenas na Cemig
- D_3 = variável *dummy* que assume o valor 1 nas concessionárias João Cesa, Muxfeldt, Nova Palma, Panambi e Urussanga.
- D_4 = variável *dummy* que assume o valor 1 apenas na Bandeirante

As empresas Bandeirantes, Cemig, CPFL, Light e Eletropaulo, juntas distribuem 47,96 % de toda energia elétrica consumida no Brasil e concentram 36,68% do número de consumidores. Por outro lado, as empresas João César, Muxfeldt, Nova Palma, Panambi e Urussanga, juntas distribuem 0,07 % de toda energia distribuída a fornecem energia para apenas 0,07% dos consumidores. Portanto, as *dummies* capturam as especificidades das cinco menores e cinco maiores concessionárias de distribuição do setor elétrico brasileiro.

Na especificação do modelo de fronteira estocástica, o custo operacional de cada concessionária é explicado pelos níveis de três tipos de *outputs*: montante de energia distribuída, número de consumidores e extensão da rede de distribuição em cada concessionária.

Embora o número de consumidores e a energia distribuída sejam variáveis correlacionadas, a inclusão delas no modelo é explicada pelo fato de que o número de consumidores é uma *proxy* da quantidade de serviço, enquanto o montante de energia distribuída é uma *proxy* do produto (Feng-Yu, Chen-Fu & James, 2001). A consideração da extensão da rede de distribuição como um *output* é interessante, pois esta variável reflete a dispersão dos consumidores na área de concessão, um fator que não pode ser controlado pela distribuidora. A não inclusão desta variável poderia gerar distorções nas medidas de eficiência das distribuidoras, penalizando as empresas onde os consumidores se encontram dispersos na área de concessão e beneficiando as empresas onde os consumidores se encontram mais concentrados.

Assim, foram estimados dois modelos de fronteira estocástica de custo: um modelo Log-Linear (Cobb-Douglas) e um Translog, apresentados a seguir nas equações 8.1 e 8.2.

$$\begin{aligned} \ln OPEX_i = & \beta_0 + \beta_1 \ln CONS_i + \beta_2 \ln NCONS_i + \beta_3 \ln REDE_i + \\ & \beta_4 D_1 + \beta_5 D_2 + \beta_6 D_3 + \beta_7 D_4 + \varepsilon_i \end{aligned} \quad (8.1)$$

$$\begin{aligned} \ln OPEX_i = & \beta_0 + \beta_1 \ln CONS_i + \beta_2 \ln NCON_i + \beta_3 \ln REDE_i + \\ & \beta_4 (\ln CONS_i)^2 + \beta_5 (\ln NCON_i)^2 + \beta_6 (\ln REDE_i)^2 + \\ & \beta_7 \ln CONS_i \cdot \ln NCON_i + \beta_8 \ln CONS_i \cdot \ln REDE_i + \\ & + \beta_4 D_1 + \beta_5 D_2 + \beta_6 D_3 + \beta_7 D_4 + \varepsilon_i \end{aligned} \quad (8.2)$$

Onde :

$CONS_i$ = energia distribuída pela i -ésima concessionária (MWh)

$NCON_i$ = total de consumidores da i -ésima concessionária

$REDE_i$ = extensão da rede de distribuição da i -ésima concessionária (km)

$OPEX_i$ = custo operacional da i -ésima concessionária (R\$)

D_1 = variável *dummy* que assume o valor 1 nas empresas CPFL, Eletropaulo e Light

D_2 = variável *dummy* que assume o valor 1 apenas na Cemig

D_3 = variável *dummy* que assume o valor 1 nas empresas João Cesar, Muxfeldt, Nova Palma, Panambi e Urussanga

D_4 = variável *dummy* que assume o valor 1 apenas na Bandeirante

$\varepsilon_i = v_i + u_i$ (erro composto da fronteira estocástica de custo)

Com o auxílio do programa Frontier 4.1 (Coelli, 1996), cujas rotinas de entrada e saída foram ligeiramente modificadas de forma a facilitar a manipulação dos dados e resultados, obtiveram-se as estimativas das equações 8.1 e 8.2, considerando distribuições *Half-Normal* e Normal truncada para a componente u_i do erro composto, conforme mostra a tabela 6.

Os resultados da tabela 6, em particular a estatística γ e o teste da razão de verossimilhança (*LR test*), ambas significativas nos quatro modelos, confirmam a existência da ineficiência na *cross-section* analisada e justificam o emprego de modelos de fronteira estocástica.

A seguir, a figura 30 mostra os *boxplots* dos indicadores de eficiência obtidos pelos modelos DEA-CRS⁵³ e DEA-VRS, e modelos SFA (modelos 1 a 4). A comparação dos *boxplots* revela que os modelos DEA obtiveram indicadores ligeiramente maiores que os estimados pelos modelos SFA. Os *boxplots* também mostram que os indicadores definidos pelos modelos DEA, em particular o DEA-CRS, são mais próximos dos resultados dos modelos SFA com forma funcional Translog (modelos 2 e 4), enquanto, que os modelos Cobb-Douglas (modelos 1 e 3) obtiveram os menores indicadores. Nota-se também que para uma dada especificação, os modelos SFA com distribuição Normal truncada para u_i , resultaram em indicadores ligeiramente maiores que os encontrados pelos modelos com distribuição *Half-Normal*. Estas conclusões estão resumidas nas estatísticas apresentadas na tabela 7.

⁵³ *Constant return of scale* (CRS) e *Variable return of scale* (VRS)

Tabela 6 – Modelos de fronteira estocástica (SFA)

	Modelo 1	Modelo 2	Modelo 3	Modelo 4
	Cobb-Douglas Normal Half- Normal	Translog Normal Half- Normal	Cobb-Douglas Normal truncada	Translog Normal truncada
Intercepto	-1,2634 (-18,7751*)	-12,0219 (-11,7029*)	-1,2753 (-4,3163*)	-6,9737 (-2,0928**)
<i>Ln CONS</i>	0,1830 (48,0741*)	3,1522 (5,1083*)	0,2541 (2,4175*)	1,7942 (5,5769*)
<i>Ln NCON</i>	0,6954 (4,9772*)	-1,6867 (-2,3195*)	0,6097 (4,6864*)	-0,8946 (-6,9339*)
<i>Ln REDE</i>	-0,0023 (-0,0595)	1,1257 (4,7600*)	0,0154 (0,3667)	1,0279 (2,3983*)
(<i>Ln CONS</i>) ²		-0,1253 (-1,1393)		0,0736 (1,8787**)
(<i>Ln NCON</i>) ²		0,0202 (0,0878)		0,2503 (2,3832*)
(<i>Ln REDE</i>) ²		0,0861 (3,8102*)		0,0973 (3,4462*)
<i>LnCONS</i> *		0,1553 (0,4963)		-0,2450 (-1,7040**)
<i>LnNCON</i> *		-0,1490 (-1,0280)		-0,0644 (-0,5038)
<i>LnREDE</i> *		-0,0442 (-0,2541)		-0,1471 (-1,8409**)
<i>D1</i>	0,2744 (9,2998*)	0,5010 (3,5836*)	0,1898 (1,1949**)	0,4707 (2,3145*)
<i>D2</i>	0,6342 (4,3347*)	0,9746 (5,2232*)	0,4793 (1,8291**)	0,8959 (2,5302*)
<i>D3</i>	-0,4000 (-4,9512*)	0,2028 (2,5263*)	-0,3837 (-3,1656*)	0,0021 (0,0155*)
<i>D4</i>	0,6658 (4,8888*)	0,8518 (3,6053*)	0,4936 (1,7267**)	0,7016 (5,3233*)
σ^2	0,3498 (12,0233*)	0,2510 (7,8320*)	0,4365 (1,1106)	0,7282 (2,5003*)
γ	1,0000 (46908,9420*)	1,0000 (2867197,9000*)	0,9831 (40,6591*)	1,0000 (64959,6240*)
<i>LogLikelihood</i>	-11,4926	1,8533	-13,4898	8,2763
<i>LR test</i>	12,6957*	26,7591*	8,7014*	39,6050*
μ			-0,3726 (-0,3492)	-1,6503 (-2,1363*)
Iterações	21	27	25	44

Notas: O valor entre parêntesis é o *t-valor*

*significativamente diferente de zero ao nível de 5%

** significativamente diferente de zero ao nível de 10%

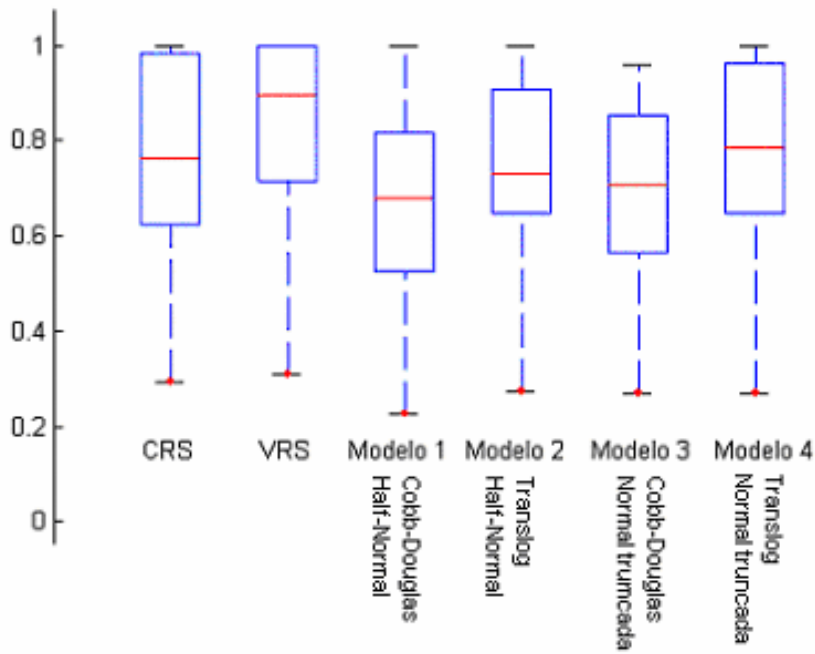


Figura 30: *Boxplots* dos indicadores de eficiência obtidos em cada modelo

Tabela 7 – Estatísticas descritivas dos indicadores de eficiência

	DEA		SFA Normal / Half-Normal		SFA Normal / Normal Truncada	
	CRS	VRS	Cobb-Douglas	Translog	Cobb-Douglas	Translog
Média	0,755	0,836	0,6639	0,7353	0,6943	0,7635
Desvio-padrão	0,211	0,199	0,2039	0,1929	0,1919	0,2089
Mínimo	0,294	0,308	0,2258	0,2742	0,2689	0,2687
Máximo	1,000	1,000	0,9998	0,9999	0,9612	0,9994

Nas tabelas 8, 9 e 10 a seguir, podem ser visualizados os indicadores de eficiência para cada um dos grupos definidos na seção 8.3 (implementação do PSOM).

Tabela 8 – Indicadores de eficiência das concessionárias do grupo 1 (empresas de grande porte)

Concessionárias	DEA		SFA Normal/Half Normal		SFA Normal/Normal Truncada	
	CRS	VRS	Cobb- Douglas	Translog	Cobb- Douglas	Translog
Bandeirantes	1,000	1,000	0,8908	0,9055	0,8573	0,9973
CEEE	0,313	0,465	0,2643	0,2742	0,2915	0,2687
Celesc	0,592	0,595	0,4059	0,3229	0,4599	0,3553
Celg	0,619	0,700	0,3889	0,4608	0,4269	0,4507
Celpe	1,000	1,000	0,8741	0,9512	0,8924	0,9551
Cemig	0,840	1,000	0,8956	0,6934	0,8573	0,8070
Cerj	0,769	0,855	0,6704	0,7508	0,7036	0,7538
Coelba	0,735	0,758	0,6229	0,6788	0,6492	0,7011
Coelce	0,795	0,950	0,6983	0,7866	0,7324	0,7787
Copel	1,000	1,000	0,8151	0,9083	0,8589	0,9029
CPFL	1,000	1,000	0,9307	0,9351	0,9064	0,9906
Elektro	1,000	1,000	0,7642	0,6753	0,8428	0,7076
Eletropaulo	0,742	1,000	0,6548	0,8644	0,6443	0,9712
Light	0,769	0,856	0,7120	0,9146	0,7054	0,9707
Média	0,798	0,870	0,6849	0,7203	0,7020	0,7579
Desvio	0,201	0,177	0,207	0,226	0,192	0,243
Mínimo	0,313	0,465	0,2643	0,2742	0,2915	0,2687
Máximo	1,000	1,000	0,9307	0,9512	0,9064	0,9973

Tabela 9 – Indicadores de eficiência das concessionárias do grupo 2 (empresas de médio porte)

Concessionárias	DEA		SFA Normal/Half Normal		SFA Normal/Normal Truncada	
	CRS	VRS	Cobb- Douglas	Translog	Cobb- Douglas	Translog
AES-Sul	1,000	1,000	0,9998	0,9524	0,9612	0,9826
CAT-LEO	0,692	1,000	0,6502	0,8251	0,6946	0,7951
Ceal	0,665	0,771	0,6180	0,7256	0,6451	0,7167
Ceb	0,294	0,342	0,3055	0,3652	0,3354	0,3474
Celpa	0,394	0,425	0,3520	0,4085	0,3750	0,3977
Celtins	0,427	0,744	0,4137	0,4944	0,4386	0,4940
Cemar	0,742	1,000	0,6015	0,7222	0,6173	0,7491
Cemat	0,464	0,490	0,3600	0,4601	0,4055	0,4543
Cepisa	0,721	0,831	0,6276	0,7319	0,6432	0,7694
Ceron	0,477	0,679	0,4701	0,5805	0,5049	0,5557
Cosern	0,898	0,982	0,8476	0,9899	0,8822	0,9462
Energipe	0,744	1,000	0,7716	0,9223	0,8193	0,8748
Enersul	1,000	1,000	0,5840	0,8977	0,6550	0,8939
Escelsa	0,682	0,695	0,6814	0,6789	0,7598	0,6895
Manaus	0,382	0,630	0,4340	0,7287	0,4755	0,6860
Piratinga	0,913	1,000	0,8497	0,9683	0,9059	0,9991
RGE	1,000	1,000	0,9837	0,9937	0,9554	0,9964
Saelpa	0,970	1,000	0,8183	0,9667	0,8335	0,9994
Média	0,693	0,811	0,632	0,745	0,662	0,742
Desvio	0,239	0,225	0,215	0,211	0,205	0,217
Mínimo	0,294	0,342	0,306	0,365	0,335	0,347
Máximo	1,000	1,000	1,000	0,994	0,961	0,999

Tabela 10 – Indicadores de eficiência das concessionárias do grupo 3 (empresas de pequeno porte)

Concessionárias	DEA		SFA Normal/Half Normal		SFA Normal/Normal Truncada	
	CRS	VRS	Cobb- Douglas	Translog	Cobb- Douglas	Translog
Boa Vista	0,297	0,308	0,2727	0,3072	0,297	0,3077
Bragantina	0,759	1,000	0,5807	0,7205	0,6424	0,6954
Caiuá	0,626	1,000	0,5476	0,6333	0,5869	0,6115
Cea	0,450	0,489	0,4132	0,4669	0,4430	0,4607
Celb	0,942	1,000	0,9018	0,9878	0,9105	0,9963
Cenf	0,663	0,705	0,6806	0,6634	0,7089	0,7130
Cflo	0,803	0,835	0,7784	0,8315	0,8257	0,8605
Chesp	1,000	1,000	0,9181	0,6918	0,9186	0,8922
Cocel	0,851	0,852	0,7816	0,8236	0,8491	0,8603
Cpee	0,844	0,845	0,7504	0,8369	0,8189	0,8518
Cspe	1,000	1,000	0,8721	0,9999	0,9137	0,9994
Demei	0,817	0,861	0,9757	0,7636	0,9322	0,9812
Eletoacre	0,741	0,914	0,7041	0,6867	0,7276	0,7411
Eletoacar	0,724	0,734	0,7135	0,6939	0,7737	0,7491
Jaguari	1,000	1,000	0,5604	0,9084	0,6507	0,9772
João César	0,888	1,000	0,5828	0,9389	0,6512	0,9990
Mococa	0,753	0,773	0,7582	0,7227	0,8105	0,7650
Muxfeldt	1,000	1,000	0,9549	0,6918	0,9339	0,8795
Nacional	0,835	0,882	0,7981	0,9349	0,8326	0,9565
Nova Palma	1,000	1,000	0,7927	0,9626	0,8523	0,9935
Panambi	0,569	0,579	0,4359	0,5728	0,4751	0,5639
Poços de Caldas	1,000	1,000	0,6174	0,8591	0,6865	0,8925
Santa Cruz	0,728	1,000	0,5982	0,7234	0,6513	0,6941
Santa Maria	0,818	0,863	0,6816	0,7880	0,7395	0,8097
Sulgipe	1,000	1,000	0,9820	0,8064	0,9347	0,9834
Urussanga	0,460	0,499	0,2258	0,5067	0,2689	0,5124
V. Paranapanema	0,570	0,722	0,5013	0,5777	0,5413	0,5608
Xanxerê	0,534	0,535	0,4949	0,5105	0,5471	0,5448
Média	0,774	0,836	0,674	0,736	0,712	0,780
Desvio	0,195	0,197	0,201	0,170	0,187	0,192
Mínimo	0,297	0,308	0,226	0,307	0,269	0,308
Máximo	1,000	1,000	0,982	1,000	0,935	0,999

As correlações entre os indicadores de eficiência, obtidos pelos diferentes modelos, são apresentadas na tabela 11. Observe que os indicadores obtidos pelos modelos SFA são ligeiramente mais correlacionados com os indicadores DEA-CRS. Outro resultado importante, revelado pela tabela 11, é que as correlações entre os indicadores obtidos por modelos SFA com a mesma forma funcional

(modelos 1 e 3 ou modelos 2 e 4) são maiores do que as verificadas entre modelos SFA com especificações diferentes.

Tabela 11 – Matriz de correlações entre os indicadores de eficiência

	CRS	VRS	Modelo 1	Modelo 2	Modelo 3	Modelo 4
CRS	1	0,8818	0,8612	0,8431	0,8885	0,8950
VRS	0,8818	1	0,7824	0,8408	0,8047	0,8558
Modelo 1	0,8612	0,7824	1	0,7872	0,9881	0,8671
Modelo 2	0,8431	0,8408	0,7872	1	0,8196	0,9630
Modelo 3	0,8885	0,8047	0,9881	0,8196	1	0,8802
Modelo 4	0,8950	0,8558	0,8671	0,9630	0,8802	1

A seguir, a tabela 12 apresenta o *ranking* das concessionárias em ordem crescente de eficiência, segundo cada um dos modelos analisados. Na tabela 12, a área em cinza indica as DMU's eficientes identificadas pelos modelos DEA. Tais DMU's ocupam a mesma posição no *ranking*, pois os modelos DEA não conseguem discriminar entre as DMU's eficientes.

Posteriormente, a tabela 13 apresenta as correlações entre os *rankings* das concessionárias na tabela 12. Os resultados da tabela 13 confirmam, mais uma vez, a maior correlação entre os modelos SFA com mesma especificação funcional (modelos 1 e 3 ou modelos 2 e 4) e também confirmam a maior correlação dos modelos SFA com o modelo DEA-CRS. As correlações com os modelos DEA, apresentadas na tabela 13, devem ser vistas com ressalva, pois nestes modelos as DMU's eficientes ocupam a mesma posição no *ranking*.

Tabela 12 – Concessionárias em ordem crescente de eficiência

CRS	VRS	Cobb-Douglas Normal Half- Normal	Cobb-Douglas Normal truncada	Translog Normal Half- Normal	Translog Normal truncada
Ceb	Boa_Vista	Urussanga	CEEE	Urussanga	CEEE
Boa_Vista	Ceb	CEEE	Boa_Vista	CEEE	Boa_Vista
CEEE	Celpa	Boa_Vista	Celesc	Boa_Vista	Ceb
Manaus	CEEE	Ceb	Ceb	Ceb	Celesc
Celpa	Cea	Celpa	Celpa	Celpa	Celpa
Celtins	Cemat	Cemat	Cemat	Cemat	Celg
Cea	Urussanga	Celg	Celg	Celg	Cemat
Urussanga	Xanxere	Celesc	Cea	Celtins	Cea
Cemat	Panambi	Cea	Celtins	Cea	Celtins
Ceron	Celesc	Celtins	Urussanga	Celesc	Urussanga
Xanxere	Manaus	Manaus	Xanxere	Panambi	Xanxere
Panambi	Ceron	Panambi	Panambi	Manaus	Ceron
VParanapanema	Escelsa	Ceron	VParanapanema	Ceron	VParanapanema
Celesc	Celg	Xanxere	Ceron	VParanapanema	Panambi
Celg	Cenf	VParanapanema	Caiua	Xanxere	Caiua
Caiua	VParanapanema	Caiua	Cenf	Caiua	Manaus
Cenf	Eletrocar	Jaguari	Elektro	Cemar	Escelsa
Ceal	Celtins	Bragantina	Coelba	Bragantina	Ssanta_Cruz
Escelsa	Coelba	Joao_Cesa	Escelsa	Cepisa	Bragantina
CAT-LEO	Ceal	Enersul	Eletroacre	Eletropaulo	Coelba
Cepisa	Mococa	Ssanta_Cruz	Chesp	Ceal	Elektro
Eletrocar	Cepisa	Cemar	Muxfeldt	Coelba	Cenf
Santa_Cruz	Cflo	Pocos_Caldas	Cemig	Jaguari	Ceal
Coelba	Cpee	Ceal	Eletrocar	Joao_Cesa	Eletroacre
Eletroacre	Cocel	Coelba	Bragantina	Santa_Cruz	Cemar
Eletropaulo	Cerj	Cepisa	Cemar	Enersul	Eletrocar
Cemar	Light	CAT-LEO	Mococa	Pocos_Caldas	Cerj
Energipe	Demei	Eletropaulo	Ssanta_Cruz	CAT-LEO	Mococa
Mococa	Santa_Maria	Cerj	Ceal	Cerj	Cepisa
Bragantina	Nacional	Cenf	Manaus	Light	Coelce
Cerj	Eletroacre	Escelsa	Cepisa	Cenf	CAT-LEO
Light	Coelce	Santa_Maria	Cerj	Eletroacre	Cemig
Coelce	Cosern	Coelce	Demei	Coelce	Santa_Maria
Cflo	Bandeirantes	Eletroacre	Coelce	Santa_Maria	Cpee
Demei	Celpe	Light	Santa_Maria	Escelsa	Cocel
Santa_Maria	Cemig	Eletrocar	Sulgipe	Eletrocar	Cflo
Nacional	Copel	Cpee	Cocel	Mococa	Energipe
Cemig	CPFL	Mococa	CAT-LEO	Cpee	Muxfeldt
Cpee	Elektro	Elektro	Cflo	Energipe	Chesp
Cocel	Eletropaulo	Energipe	Cpee	Cflo	Pocos_Caldas
Joao_Cesa	AES-Sul	Cflo	Pocos_Caldas	Nacional	Enersul
Cosern	CAT-LEO	Cocel	Eletropaulo	Saelpa	Copel
Piratininga	Cemar	Nova_Palma	Enersul	Elektro	Cosern
Celb	Energipe	Nacional	Bandeirantes	Cocel	Celpe
Saelpa	Enersul	Copel	Copel	Nova_Palma	Nacional
Bandeirantes	Piratininga	Saelpa	Jaguari	Bandeirantes	Light
Celpe	RGE	Cosern	Light	Cemig	Eletropaulo
Copel	Saelpa	Piratininga	Energipe	Copel	Jaguari
CPFL	Bragantina	CSPE	AES-Sul	Cosern	Demei
Elektro	Caiua	Celpe	Nacional	Celpe	AES-Sul
AES-Sul	Celb	Bandeirantes	CPFL	Piratininga	Sulgipe
Enersul	Chesp	Cemig	Joao_Cesa	CPFL	CPFL
RGE	CSPE	Celb	Celpe	Celb	Nova_Palma
Chesp	Jaguari	Chesp	Nova_Palma	CSPE	Celb
CSPE	Joao_Cesa	CPFL	Saelpa	Chesp	RGE
Jaguari	Muxfeldt	Muxfeldt	Piratininga	Demei	Bandeirantes
Muxfeldt	Nova_Palma	Demei	Celb	Muxfeldt	Joao_Cesa
Nova_Palma	Pocos_Caldas	Sulgipe	Cosern	Sulgipe	Piratininga
Pocos_Caldas	Ssanta_Cruz	RGE	RGE	RGE	Saelpa
Sulgipe	Sulgipe	AES-Sul	CSPE	AES-Sul	CSPE

Tabela 13 – Matriz de correlações entre os *rankings* das concessionárias

	CRS	VRS	Modelo 1	Modelo 2	Modelo 3	Modelo 4
CRS	1	0,2241	-0,0315	0,3020	0,1578	0,1102
VRS	0,2241	1	-0,1245	-0,0638	-0,0894	-0,1205
Modelo 1	-0,0315	-0,1245	1	-0,2347	0,4824	-0,1718
Modelo 2	0,3020	-0,0638	-0,2347	1	-0,1388	0,2447
Modelo 3	0,1578	-0,0894	0,4824	-0,1388	1	0,0845
Modelo 4	0,1102	-0,1205	-0,1718	0,2447	0,0845	1

Do ponto de vista da regulação, variações substanciais nos indicadores de eficiência e nos *rankings* obtidos pelos diferentes métodos não é tranquilizador. Na ausência de um consenso acerca da abordagem mais apropriada, a diminuição das disparidades na definição da eficiência das concessionárias pode ser conseguida pela combinação dos resultados dos diferentes modelos, por exemplo, a média geométrica dos indicadores de eficiência obtidos pelos melhores modelos (Coelli e Peleman, 1999).

Neste trabalho, os resultados mais consistentes foram obtidos pelos modelos DEA-CRS, modelos SFA com forma funcional Translog (modelos 2 e 4) e pelo modelo SFA com forma funcional Cobb-Douglas e Normal truncada (modelo 3). Então, seguindo a sugestão de utilizar a média geométrica, obtiveram-se os indicadores de eficiência apresentados na tabela 14.

Tabela 14 – Média geométrica dos indicadores de eficiência dos modelos DEA-CRS e SFA (2, 3 e 4)

Concessionária	Eficiência	Grupo
Bandeirantes	0,9380	1
CEEE	0,2863	1
Celesc	0,4204	1
Celg	0,4840	1
Celpe	0,9489	1
Cemig	0,7967	1
Cerj	0,7439	1
Coelba	0,6903	1
Coelce	0,7728	1
Copel	0,9161	1
CPFL	0,9572	1
Elektro	0,7966	1
Eletropaulo	0,7959	1
Light	0,8330	1
AES-Sul	0,9669	2
CAT-LEO	0,7494	2
Ceal	0,6873	2
Ceb	0,3344	2
Celpa	0,3936	2
Celtins	0,4625	2
Cemar	0,7055	2
Cemat	0,4453	2
Cepisa	0,7149	2
Cosern	0,9281	2
Energipe	0,8374	2
Enersul	0,8515	2
Escelsa	0,7018	2
Manaus	0,5489	2
Piratininga	0,9458	2
RGE	0,9862	2
Saelpa	0,9401	2
Boa Vista	0,3022	3
Bragantina	0,7030	3
Caiuá	0,6142	3
Cea	0,4551	3
Celb	0,9585	3
Cenf	0,6867	3
Cflo	0,8299	3
Chesp	0,8677	3
Cocel	0,8459	3
Cpee	0,8378	3
CSPE	0,9775	3
Demei	0,8691	3
Eletroacre	0,7238	3
Eletrocar	0,7346	3
Jaguari	0,8718	3
João César	0,8582	3
Mococa	0,7622	3
Muxfeldt	0,8682	3
Nacional	0,8880	3
Nova Palma	0,9502	3
Panambi	0,5436	3
Poços de Caldas	0,8518	3
Santa Cruz	0,6985	3
Santa Maria	0,7882	3
Sulgipe	0,9279	3
Urussanga	0,4233	3
V. Parapanema	0,5623	3
Xanxerê	0,5339	3

Os indicadores de eficiência obtidos pelos diferentes modelos servem de referência para o regulador estabelecer a componente operacional X^O . Recente pesquisa realizada por Jasmab e Pollit (2001) mostra que existem diversas formas de converter a medida de eficiência na componente operacional X^O . Em geral, pode-se definir a componente X^O como uma parcela da distância entre o vetor representativo da concessionária e a fronteira de eficiência, ou seja, uma parcela de $1-\theta^*$ (θ^* é o índice de eficiência da concessionária)⁵⁴, combinada previamente ou imposta pela agência reguladora⁵⁵. Por este critério, as concessionárias eficientes teriam um valor nulo para a componente X^O .

⁵⁴ Como foi proposto no capítulo 6 desta tese.

⁵⁵ Por exemplo, Chercye e Post (2001) mostram que na Holanda, durante o período 2001-2003, cada concessionária de distribuição de eletricidade tiveram que reduzir seus preços pelo grau de ineficiência obtido por um modelo DEA-CRS, respeitando um limite de 8% por ano, após a correção pelo índice de inflação.