

3

Descrição da Metodologia

A proposta deste capítulo é discutir o referencial analítico que auxiliou o desenvolvimento deste trabalho. Foram utilizados dois métodos diferentes: DEA e *tobit*.

3.1

Estratégia de Análise

Para construir uma boa estratégia de análise optou-se por aplicar um conjunto de metodologias matemáticas e estatísticas. A técnica consiste em calcular, em um primeiro estágio, uma medida de eficiência com a metodologia DEA e, com o resultado, iniciar o segundo estágio, que consiste na regressão *tobit*. A metodologia DEA deverá ser aplicada com o objetivo de eliminar o efeito de escala dos operadores, para isso definiu-se como insumo as variáveis relacionadas com o tamanho, ou seja, foram utilizadas as variáveis de Receita Bruta e Crescimento que representam os resultados (*outputs*) e número de funcionários e número de clientes com contrato representam os recursos (*inputs*). Com o resultado deste *score* será feita uma regressão *tobit*, com finalidade de identificar quais os fatores que causam impacto na eficiência dos operadores no mercado brasileiro.

3.2

Análise Envoltória de Dados (DEA)

O desenvolvimento do método da Análise por Envoltório de Dados teve início na tese de doutorado de Edward Rhodes, apresentada a *Carnegie Mellon University* no ano de (1978), com a orientação de W.W. Cooper que resultou na formulação do modelo Charnes, Cooper e Rhodes de Análise por Envoltório de Dados.

De acordo com Figueiredo (2009) a técnica de construção de fronteiras de produção e indicadores de eficiência produtiva relativa teve origem

com o trabalho de Farrel (1957) e foi generalizada por Charnes, Cooper e Rhodes em (1978), no sentido de trabalhar com múltiplos recursos e múltiplos produtos.

3.2.1

Vantagens do DEA

Desde os primeiros estudos a técnica de DEA é considerada uma forte ferramenta de análise criteriosa.

Segundo Figueiredo (2005), a metodologia pode ser aplicada para avaliar qualquer tipo de empresas ou setores, mas as comparações entre as unidades tomadoras de decisão devem ser de grupos homogêneos. Desta forma, é possível identificar as causas da ineficiência das unidades e obter o índice de eficiência, tomando as unidades eficientes como o conjunto de referências para as demais unidades estudadas.

Uma das vantagens da técnica é a não exigência de nenhuma forma funcional das variáveis relacionadas na análise envolvendo os *inputs* e os *outputs*, isto é, trata-se de um método não paramétrico. Outra vantagem proporcionada pela metodologia é o fato de se poder trabalhar com vários *outputs*, embora a escolha das variáveis no modelo deva ser feita de maneira cautelosa, pois, quanto mais variáveis presentes no modelo menor será o poder discriminatório.

O estudo de Cardoso e Costa (2007) descreve que DEA indica a eficiência da produção relacionada com a Tecnologia de Informação. As autoras citam o estudo de Thanassoulis (1993) e Athanassoupoulos & Curram (1996) sobre a comparação com outros métodos como a análise de regressão.

De acordo com Charnes, Cooper e Rhodes (1978) e Guerreiro (2006), uma desvantagem da DEA em relação às técnicas de estimação paramétricas está nos testes que comprovam as hipóteses, no qual a priori, não é possível testar com a rigidez estatística, bem como o erro relativo à estimação da fronteira de eficiência, já que os insumos e os produtos são variáveis aleatórias.

Pode ser definida como um método quantitativo que gera um único indicador de desempenho para cada unidade avaliada, considerando a relação ponderada entre insumos e produtos. Na dissertação de Figueiredo (2005) são destacadas algumas vantagens do DEA como:

- ✓ Não exige conhecimento dos pesos, insumos e produtos a priori.

- ✓ Os insumos e produtos podem ser mensurados em diferentes unidades, é invariante em relação à escala.
- ✓ Considera-se qualquer tipo de função de produção.
- ✓ Pode integrar opinião de especialista.
- ✓ Resultam estimativas específicas das mudanças almejadas nos insumos e produtos para projeção das DMUs abaixo da fronteira de eficiência.
- ✓ Destaque nas melhoras práticas identificadas e não em medidas de tendência central das fronteiras.

Também destacou algumas limitações do DEA, tais como:

- ✓ Por ser uma técnica não paramétrica é difícil de formular hipóteses estatísticas.
- ✓ Pode levar muito tempo computacional, pois é um programa linear para cada unidade em análise.
- ✓ Uma metodologia de ponto extremo, como erros de medição, pode atrapalhar na análise dos resultados.
- ✓ A sensibilidade da metodologia em relação ao número de *inputs*, *outputs* utilizados e o tamanho da amostra de DMUs. Aumentando o número de DMUs existe uma tendência de reduzir a média das pontuações de eficiência da amostra, pois quanto maior o número de DMUs maior poderá ser o número de DMUs na fronteira. Outro ponto importante, quando o tamanho da DMUs é pequeno em relação à soma do número de *inputs* e *outputs* a média de eficiência da amostra tende a aumentar. Nunamaker (1985) e Bowlin (1998) recomendam que o tamanho de DMUs observadas da amostra seja pelo menos três vezes maior que a soma dos *inputs* e dos *outputs*.
- ✓ E, por último, a metodologia só resulta em medidas de eficiência relativas dentro de uma amostra particular. Não se pode comparar a pontuação entre dois resultados diferentes, quando as práticas são desconhecidas.

Em suma, o foco da metodologia DEA está nas observações individuais representadas por n otimizações, uma para cada DMU, não tendo foco nas médias e nas estimativas de parâmetros que estão associadas com as técnicas e aproximações estatísticas (Charnes et al. (1994)).

Segundo Charnes, Cooper e Rhodes (1978), enquanto a Análise de Regressão solicita uma função de produção, na qual variáveis independentes se relacionam com as variáveis dependentes por meio desta função e assumindo hipóteses específicas sobre distribuições de erros, o método

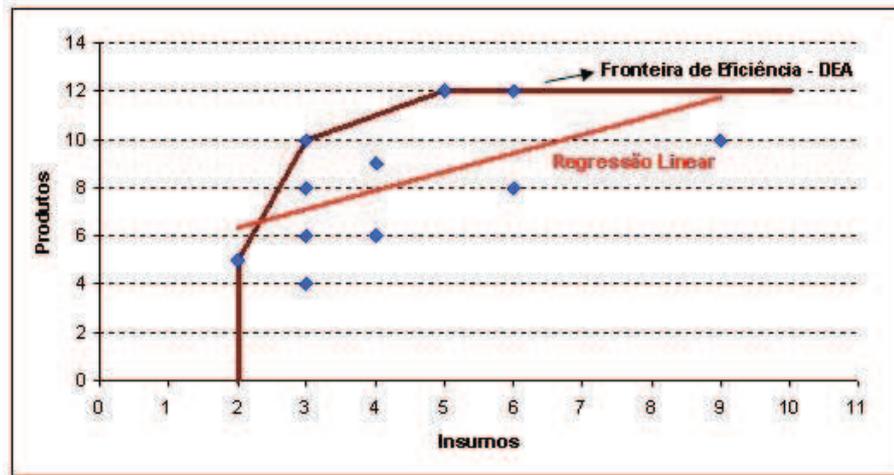


Figura 3.1: Regressão linear

DEA calcula a máxima medida de desempenho para cada DMU relativa a todas as outras DMU's, tendo como princípio que cada DMU está na fronteira de eficiência ou abaixo dela.

3.2.2

O conceito da metodologia DEA

Segundo Guerreiro (2006), uma referência sobre a metodologia DEA é M.J. Farrel, pois é considerado um dos pioneiros no assunto. Seu estudo caracterizou a eficiência global como composta de duas componentes distintas de eficiência: preços e técnica.

A Análise por Envoltório de Dados (DEA) avalia a eficiência técnica, obtendo a soma ponderada de produtos dividida pela soma ponderada de recursos. Trata-se de uma técnica matemática que permite avaliar o grau de eficiência produtiva das empresas. As empresas são nomeadas unidades tomadoras de decisão (DMU - *Decision Making Unit*), considerando recursos disponíveis (*inputs*) com os resultados obtidos (*outputs*).

Segundo Charnes, Cooper e Rhodes (1978) e Guerreiro (2006), o método paramétrico tem como objetivo otimizar um plano de regressão simples, já o modelo DEA permite otimizar individualmente cada uma das observações, uma em relação às outras, formando uma fronteira de eficiência.

A figura 3.2 mostra um conjunto de dados colocados como exemplo para ilustração do princípio básico do DEA. Cada ponto plotado é como uma

DMU. As DMU's estudadas devem ser autossuficientes na tomada de decisão para a aplicação proposta pelo método.

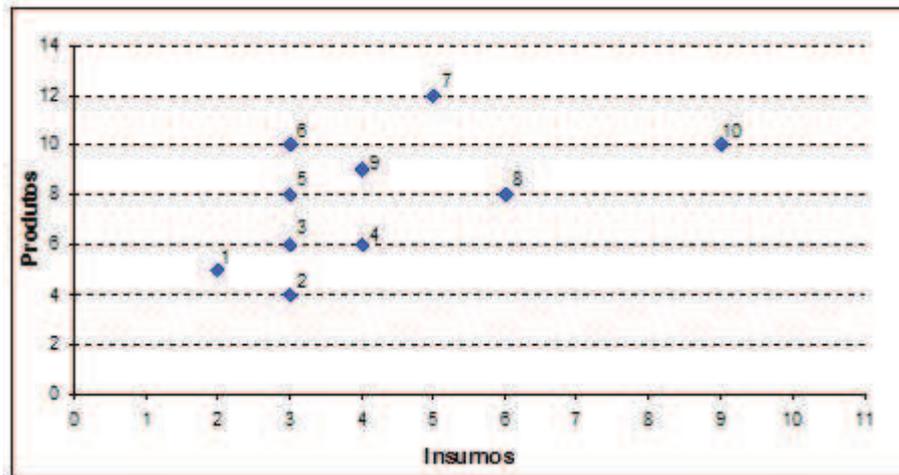


Figura 3.2: Conjunto de dados

Observando a figura 3.2, a DMU 6 produz mais produtos que a DMU 5 utilizando a mesma quantidade de insumos. Desta forma, consideramos a DMU 6 mais eficiente que a DMU 5. Comparando as DMU's 3 e 4, para a mesma quantidade de produtos, mas a DMU 4 utiliza mais insumos que a DMU 3. Neste caso, a DMU 3 é mais eficiente que a DMU 4.

Assim, as DMU's mais eficientes são aquelas que conseguem aproveitar ao máximo a quantidade de insumos disponíveis, produzindo um maior nível de produtos.

Estas observações introduzem o fundamento da metodologia: medir a eficiência das DMU's considerando a quantidade de insumos, ou *input*, para atingir uma dada quantidade de produto, ou *output*.

Assim, a abordagem DEA maximiza a soma ponderada das quantidades de produtos, isto é, uma combinação linear das quantidades dos produtos buscando os pesos para cada produto de forma que a combinação linear dos produtos seja máxima.

De acordo com a definição de Pareto-Koopmans, a DMU será eficiente quando nenhum dos produtos possa ser adicionado sem que algum insumo necessite de acréscimo, ou que outro produto seja reduzido. Ou ainda, nenhum dos insumos possa ser atenuado sem que algum produto necessite de aumento, ou que outro produto seja reduzido.

A Figura 3.3 mostra a fronteira de eficiência. Foram considerados eficientes, no sentido de Pareto, os pontos com maior aproveitamentos dos

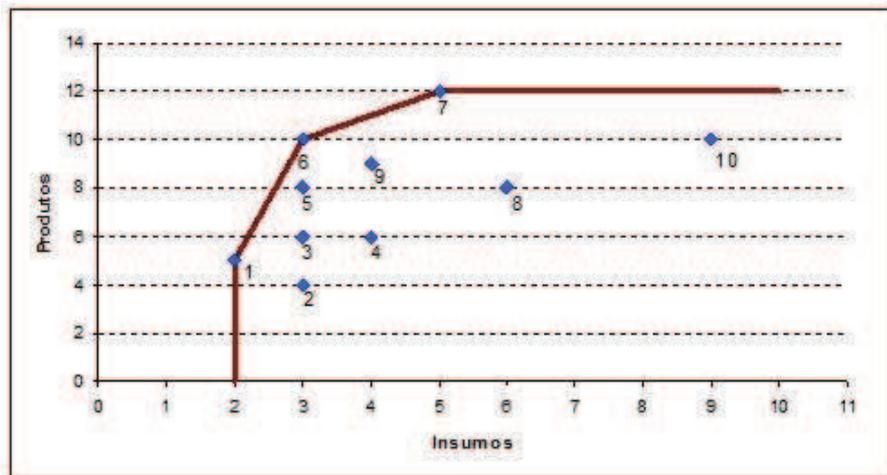


Figura 3.3: Fronteira de Eficiência

insumos, isto é, o nível máximo de produção para um dado nível de insumo. A fronteira assim é formada pelas unidades eficientes determinadas pela otimização de cada ponto ou unidade individualmente.

Na curva de fronteira, temos uma das principais características da programação linear que é a convexidade, permitindo um resultado viável.

Quando na curva de fronteira tem concavidade, a solução de um PPL é inviável, não sendo o caso desse problema. Assim, podemos projetar ortogonalmente as DMU's ineficientes, calculando a ineficiência relativa de cada ponto. A figura 3.4 mostra a relação entre a quantidade de insumos e a quantidade produzida por cada DMU.

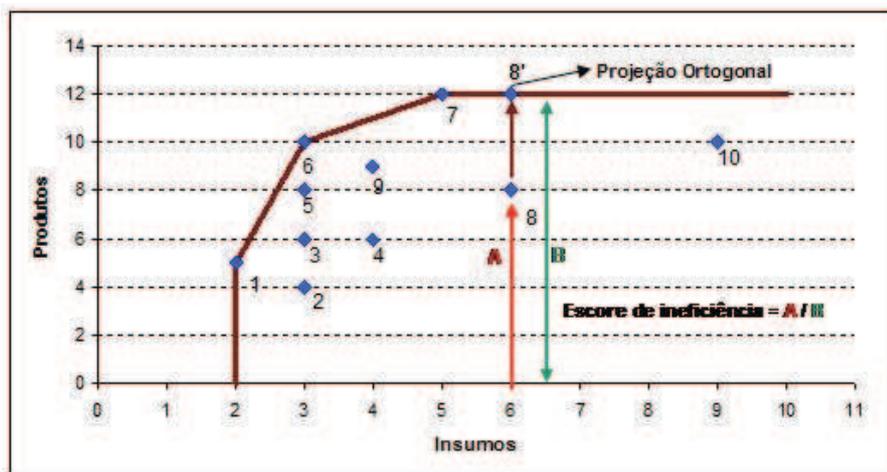


Figura 3.4: Fronteira de Eficiência

Temos que as unidades consideradas eficientes formam uma superfície côncava, no qual as DMU's ineficientes são projetadas ortogonalmente na fronteira, assim gerando o índice. O índice é calculado pela forma da

projeção das ineficientes na fronteira, ou seja, calcula-se a distância da unidade até a fronteira.

A projeção das ineficientes até a fronteira pode ser medida por dois modelos:

- (a) Modelos orientados para a maximização de *output* (produto): calculado pela máxima expansão do *output* dado uma quantidade de *input* utilizada.
- (b) Modelos orientados para a minimização de *input* (insumo): calculado pela máxima redução de *input* para uma mesma produção de *output*.

3.2.3

Modelos básicos do DEA

A Análise Envoltória de Dados (DEA) pode ser considerada como um corpo de conceitos e metodologias que está incorporada a uma coleção de modelos com possibilidades interpretativas diversas (Charnes, Cooper, Lewin, Seldorf: 1997). Apresenta-se abaixo dois modelos DEA:

- (a) Modelo CCR - Charnes, Cooper e Rhodes (1978);
- (b) Modelo BCC - Banker, Charnes e Cooper (1984).

Para melhor compreender e estabelecer a diferença entre os modelos, apresentaremos os conceitos de eficiência abordados pela metodologia. Segundo Belloni (2000), os conceitos de eficiência são classificados como:

- ✓ Eficiência produtiva: refere-se à habilidade de evitar desperdícios produzindo tantos resultados quanto os recursos utilizados permitem ou o mínimo de recursos para a produção. Pode-se decompor em dois componentes: a eficiência de escala e a eficiência técnica.
- ✓ Eficiência de escala: é o componente de eficiência produtiva associado às variações de produtividade decorrentes de mudanças na escala de operações.
- ✓ Eficiência técnica: é o componente da eficiência produtiva que resulta quando são isolados os efeitos da eficiência de escala. Desta forma, a ineficiência técnica está associada à habilidade gerencial dos administradores.

3.2.4

Modelo CCR

O primeiro modelo, CCR, define eficiência como uma soma ponderada dos *outputs* dividido pela soma ponderada dos *inputs*. É um modelo, também considerado como CRS (*Constant Returns to Scale*), que trabalha com retornos constantes de escala, ou seja, qualquer variação nos insumos leva a uma variação proporcional nos produtos.

O CCR exige que um conjunto de pesos seja atribuído a todas DMU's, o que é uma tarefa com grande grau de complexidade. Para resolução deste problema, Charnes, Cooper e Rhodes (1978) argumentaram que cada unidade possui um sistema de valores próprios e que, por si só, tem vigor para definir seu conjunto de pesos no sentido de maximizar sua eficiência.

O modelo pode ser definido como um problema de programação fracionária e que pode ser transformado em um problema de programação linear (PPL), tendo como única restrição que as unidades tenham eficiência menor ou igual a 1. Abaixo seguem as formulações dos modelos primal e dual, sua representação gráfica, e a decisão da notação usada no modelo.

Maximização de *Outputs*

CCR-O

Primal (Multiplicadores)

$$\text{Min } Eff_0 = \sum_{i=1}^r v_i x_{io}$$

Sujeito a:

$$\sum_{j=1}^s u_j y_{jk} = 1, \forall k = 1, 2, \dots, n$$

$$\sum_{i=1}^r v_i x_{ik} - \sum_{j=1}^s u_j y_{jk} \leq 0, \forall k = 1, 2, \dots, n$$

$$u_j \text{ e } v_i \geq 0 \forall j, i$$

Dual (Envelope)

Max θ

Sujeito a:

$$-\theta_{j0} + \sum_{k=1}^n y_{jk} \lambda_k \geq 0, \forall j = 1, 2, \dots, s$$

$$x_{i0} + \sum_{k=1}^n y_{ik} \lambda_k \geq 0, \forall i = 1, 2, \dots, r$$

$$u_j \text{ e } v_i \geq 0 \forall j, i$$

$$\lambda_k \geq 0 \forall k$$

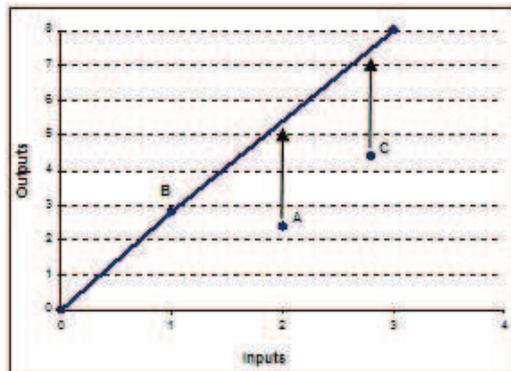


Figura 3.5: Representação Gráfica CCR-O

onde:

Eff_0 - eficiência da DMU_0 ;

u_j, v_i - pesos de *outputs* e *inputs* respectivamente;

x_{ik}, y_{jk} - *inputs* i e *outputs* j da DMU_k ;

x_{i0}, y_{j0} - *inputs* i e *outputs* j da DMU_0 ;

θ - eficiência;

λ_k - k -ésima coordenada da DMU_0 em uma base formada pelas DMU 's de referência.

Minimização de *Inputs*

CCR-I

Primal (Multiplicadores)

$$\text{Max } Eff_0 = \sum_{j=1}^s u_j y_{j0}$$

Sujeito a:

$$\sum_{i=1}^r v_i x_{ik} = 1, \forall k = 1, 2, \dots, n$$

$$\sum_{j=1}^s u_j y_{jk} - \sum_{i=1}^r v_i x_{ik} \leq 0, \forall k = 1, 2, \dots, n$$

$$u_j \text{ e } v_i \geq 0 \forall j, i$$

Dual (Envelope)

Min θ

Sujeito a:

$$\theta x_{i0} - \sum_{k=1}^n x_{ik} \lambda_k \geq 0, \quad i = 1, 2, \dots, r$$

$$-y_{j0} + \sum_{k=1}^n y_{jk} \lambda_k \geq 0, \quad j = 1, 2, \dots, s$$

$$\lambda_k \geq 0 \quad \forall k$$

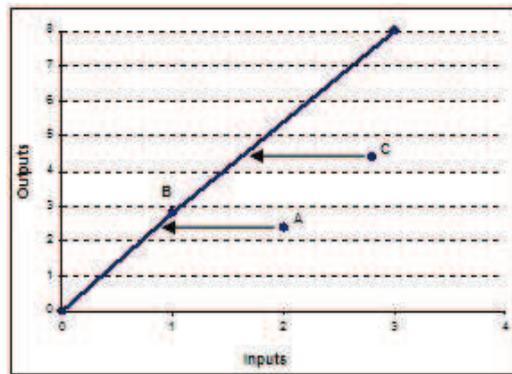


Figura 3.6: Representação Gráfica CCR-I

onde:

$Ef f_0$ - eficiência da DMU_0 ;

u_j, v_i - pesos de *outputs* e *inputs* respectivamente;

x_{ik}, y_{jk} - *inputs* i e *outputs* j da DMU_k ;

x_{i0}, y_{j0} - *inputs* i e *outputs* j da DMU_0 ;

θ - eficiência;

λ_k - k -ésima coordenada da DMU_0 em uma base formada pelas DMU 's de referência.

3.2.5

Modelo BCC

O modelo BCC é obtido pelo acréscimo no qual garante que a DMU será comparada com uma combinação convexa das restantes DMU 's, o que não ocorre com o modelo CCR que possui uma combinação linear. A convexidade abrevia um conjunto de possíveis produções viáveis e converte uma tecnologia de retorno de escala variável. O modelo BCC,

também é conhecido como VRS (*Variable Return Scale*), implica que as DMU's avaliadas apresentam retornos variáveis de escala.

O modelo determina uma fronteira VRS que considera retornos crescentes ou decrescentes de escala na fronteira eficiente. Considera que um acréscimo no *input* poderá promover um acréscimo no *output*, não necessariamente proporcional, ou até mesmo um decréscimo. A medida de eficiência técnica identifica a utilização correta dos recursos à escala de operações da DMU. A eficiência de escala é igual ao quociente da eficiência BCC pela a eficiência CCR. Otemos uma medida da distância da DMU estudada até uma DMU fictícia, no qual atua com o tamanho da escala mais produtivo.

BCC não assume crescimento constante, o que torna mais próxima da realidade das funções de produção dos processos a serem modelados. Abaixo seguem a formulação do modelo e sua representação gráfica e notação do problema.

Maximização de *Outputs*

BCC-O

Primal (Envelope)

Max θ

Sujeito a:

$$x_{i0} + \sum_{k=1}^n x_{ik} \lambda_k \geq 0, \forall i = 1, 2, \dots, r$$

$$-\theta y_{j0} + \sum_{k=1}^n y_{jk} \lambda_k \geq 0, \forall j = 1, 2, \dots, s \quad u_j \text{ e } v_i \geq 0 \quad \forall j, i$$

$$\lambda_k \geq 0 \quad \forall k$$

Dual (Envelope)

$$\text{Min } h_0 = \sum_{i=1}^r v_i x_{i0} - V_*$$

Sujeito a:

$$\sum_{j=1}^s u_j y_{jk} = 1, \forall k = 1, 2, \dots, n$$

$$\sum_{i=1}^r v_i x_{ik} - \sum_{j=1}^s u_j y_{jk} - V_* \leq 0, \forall k = 1, 2, \dots, n$$

$$u_j \text{ e } v_i \geq 0 \quad \forall j, i$$

$$V_* \in \mathfrak{R}$$

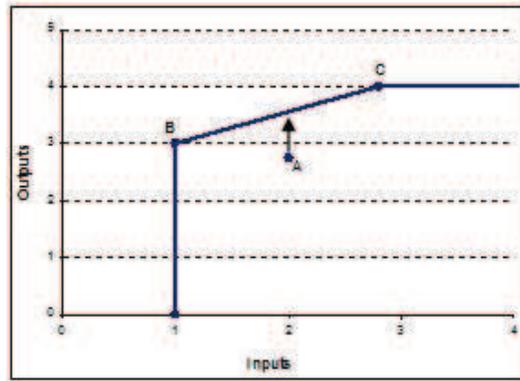


Figura 3.7: Representação Gráfica BCC-O

onde:

u_j, v_i - pesos de *outputs* e *inputs* respectivamente;

x_{ik}, y_{jk} - *inputs* i e *outputs* j da DMU_k ;

x_{i0}, y_{j0} - *inputs* i e *outputs* j da DMU_0 ;

θ - eficiência;

λ_k - k -ésima coordenada da DMU_0 em uma base formada pelas DMU 's de referência.

Minimização de *Inputs*

BCC-I

Primal (Envelope)

Min θ

Sujeito a:

$$\theta x_{i0} - \sum_{k=1}^n x_{ik} \lambda_k \geq 0, \quad i = 1, 2, \dots, r$$

$$-y_{j0} + \sum_{k=1}^n y_{jk} \lambda_k \geq 0, \quad j = 1, 2, \dots, s$$

$$\lambda_k \geq 0 \quad \forall k$$

Dual (Multiplicadores)

$$\text{Max } h_0 = \sum_{j=1}^s u_j y_{j0} - u_*$$

Sujeito a:

$$\sum_{i=1}^r v_i x_{i0} = 1$$

$$\sum_{j=1}^s u_j y_{jk} - \sum_{i=1}^r v_i x_{ik} - u_* \leq 0$$

$$k = 1, 2, \dots, n$$

$$u_j \text{ e } v_i \geq 0 \quad \forall j, i$$

$$u_* \in \mathfrak{R}$$

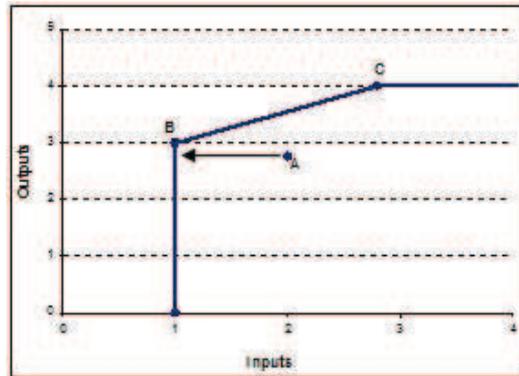


Figura 3.8: Representação Gráfica BCC-I

onde:

u_j, v_i - pesos de *outputs* e *inputs* respectivamente;

x_{ik}, y_{jk} - *inputs* i e *outputs* j da DMU_k ;

x_{i0}, y_{j0} - *inputs* i e *outputs* j da DMU_0 ;

θ - eficiência;

λ_k - k -ésima coordenada da DMU_0 em uma base formada pelas DMU 's de referência.

Desta forma, o que difere os modelos multiplicadores BCC e os modelos multiplicadores CCR são as variáveis u_* para orientação a *input* e v_* para orientações a *output*. Elas são interpretadas como fatores de escala, ou seja, variáveis positivas significam retornos decrescentes de escala, variáveis negativas retornos crescentes de escala, e variáveis nulas retornos constantes de escala.

3.2.6

DEA aplicado a operadores de logística

Gonçalves *et al.* (2005) utilizou DEA e técnicas de agrupamento para avaliar o desempenho de operadores logísticos, um estudo de caso na entrega domiciliar de jornais. Os operadores foram subdivididos em áreas de atuação e classificados pela técnica de agrupamento. A metodologia DEA

foi utilizada para avaliar a eficiência nas áreas de atuação. O modelo proposto foi o CCR (*Constant Returns to Scale*), as variáveis entregadores e reclamações diárias foram usadas como *inputs*, e jornais entregues e ponto de entrega como *outputs*. Baseado nesse modelo foi calculado um índice final comparativo de desempenho entre os operadores. Os autores concluíram que algumas características geográficas e socioeconômicas alteram os padrões de operações e influenciam a eficiência dos operadores, assim devem ser analisados em grupos homogêneos. Esse estudo teve como base a dissertação de mestrado de Gonçalves (2005).

Alguns pesquisadores da área de Transporte e Logística utilizaram a metodologia DEA com outro foco. Um bom exemplo é o estudo de Novaes (2001), que aplicou a metodologia DEA para analisar a eficiência e os ganhos de escala em 21 metrôes que operam no país e utilizar como instrumento auxiliar no *benchmarks* dos não eficientes, com destaque para o metrô de São Paulo.

Assim, a metodologia DEA foi pouco utilizada para analisar os operadores de logística. Porém, podemos dizer que essa metodologia está começando a ser utilizada como uma ferramenta para organizar e analisar de forma independente ou de forma complementar os operadores de logísticas.

3.3

Modelo tobit

De acordo com Greene (2002), o modelo de regressão *tobit* é utilizado para o estudo em que a variável dependente está concentrada entre faixas de valores, ou seja, pontos iguais a um valor limite. Nestes casos, o modelo *tobit* é o mais adequado já que soluciona esse problema valendo-se de técnicas estatísticas que possibilitam realizar inferências para toda população sem perda de qualidade.

Johnston e DiNardo (2001) descrevem o modelo *tobit* como uma extensão do modelo *probit* que, na realidade é uma aproximação para lidar com dados censurados.

Seguindo uma cronologia será descrito brevemente o modelo de regressão, e dois de seus casos particulares: o modelo *probit* e *tobit*. Essa cronologia serve para uma melhor compreensão dos modelos estatísticos. Para mais informações consultar Johnston e DiNardo (2001).

3.3.1

Modelo de Regressão

Modelos de regressão são modelos matemáticos que relacionam o comportamento de uma variável Y com outra X . Quando a função f que relaciona duas variáveis é do tipo $f(X) = \alpha + \beta X$ temos o modelo de regressão linear. A variável X é a variável independente da equação enquanto $Y = f(X)$ é a variável dependente das variações de X .

O modelo de regressão é chamado de simples quando a relação causal envolve apenas duas variáveis. O modelo de regressão é multivariado quando o comportamento de Y é explicado por mais de uma variável independente X_1, X_2, \dots, X_n .

Os modelos simples e multivariados simulam relacionamentos entre as variáveis do tipo linear, equação da reta ou do plano, ou não linear, equação exponencial, geométrica etc.

Esses modelos servem para verificar se há relações entre as variáveis independentes e a variável dependente, e como essa relação é descrita.

O modelo de regressão pode ser descrito da seguinte forma:

$$Y^* = \alpha + \beta X + \epsilon \quad (3-1)$$

onde Y^* representa um vetor com as observações da variável dependente Y ou uma transformação dessa variável. α representa o coeficiente linear do modelo, β o coeficiente angular e X um vetor de variáveis independentes (no caso do modelo simples) ou uma matriz (no caso do modelo múltiplo). O parâmetro ϵ é a estrutura do erro que, de acordo com o modelo especificado segue uma determinada distribuição estatística.

Dada a especificação do modelo e a estrutura do erro, os coeficientes α , β e os parâmetros da distribuição de ϵ são estimados através dos dados de Y e X . Existem alguns métodos para estimar esses parâmetros, dentre os mais usuais estão o método de máxima verossimilhança e os mínimos quadrados.

O método de máxima verossimilhança consiste em estimar os parâmetros desconhecidos do modelo de forma que a função de verossimilhança seja maximizada. A função de verossimilhança é definida por

$$L(\theta_1, \dots, \theta_n; x_1, \dots, x_n) = f(x_1; \theta_1, \dots, \theta_n) \dots f(x_n; \theta_1, \dots, \theta_n) \quad (3-2)$$

que deve ser encarada como uma função de $\theta_1, \dots, \theta_n$. O estimador de máxima verossimilhança de $\theta_1, \dots, \theta_n$ é o valor $\hat{\theta}_1, \dots, \hat{\theta}_n$ que maximiza $L(\theta_1, \dots, \theta_n; x_1, \dots, x_n)$.

O método de estimação de mínimos quadrados é uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados, tentando minimizar a soma dos quadrados das diferenças entre a curva ajustada e os dados (tais diferenças são chamadas resíduos).

Para exemplificar, retornemos ao modelo 3-1, entretanto, pensaremos nele com apenas uma variável independente, ou seja, X deixa de ser uma matriz para ser um vetor. Com isso, serão estimados os parâmetros α e β . Tendo esse modelo como base, as estimativas por mínimos quadrados seriam obtidas da seguinte forma

$$\min \sum_{i=1}^n \epsilon_i = \min \sum_{i=1}^n [y_i - (\alpha + \beta x_i)]^2 \quad (3-3)$$

onde *min* significa o mínimo dessa equação. Para encontrar esse mínimo basta derivar parcialmente a equação em relação aos parâmetros α e β e igualar a zero.

O procedimento de inferência dos parâmetros é baseado nas seguintes hipóteses

$$\begin{aligned} H_0 : \theta_i &= 0 \\ H_1 : \theta_i &\neq 0 \end{aligned} \quad (3-4)$$

onde H_0 é chamada de hipótese nula e H_1 é a hipótese alternativa.

Essas hipóteses são testadas de acordo com o tipo de modelo, mas sempre baseadas na sua estimativa ($\hat{\theta}_i$) e na variância da sua estimativa ($VAR(\hat{\theta}_i)$). Esse teste baseia-se em comparar o valor dessa estimativa, chamada de calculado, e compará-lo com a distribuição para essa estimativa. Nessa comparação é obtida uma probabilidade chamada de *p-valor*. O *p-valor* é a probabilidade de se obter um valor do teste estatístico tão ou mais extremo quanto o que foi observado realmente, dado que a hipótese nula é verdadeira.

Resumindo, se encontramos um *p-valor* = $\gamma\%$, dizemos que temos evidências para rejeitar H_0 num nível de $1 - \gamma\%$. Exemplificando, se obtivermos um *p-valor* = 0,01 para um parâmetro θ_i , afirmamos que temos evidências para dizer que este parâmetro é diferente de zero ao nível de 99,9%. Com isso, temos que quanto menor o *p-valor* mais

evidências temos para rejeitar H_0 . O usual é rejeitarmos H_0 ao nível de 90% ou acima, ou quando o p – valor $\leq 0,1$.

Rejeitar H_0 significa dizer que o parâmetro θ_i é diferente de zero, portanto temos que essa variável influencia de alguma forma a variável dependente. Entretanto, se não rejeitamos a hipótese nula não podemos dizer que a estimativa do parâmetro é diferente de zero, logo não há evidências para afirmar que esta variável influencia no modelo proposto.

Probit

O Modelo *probit* é uma alternativa do modelo *logit*, para maiores detalhes ver Johnston e Dinardo (2001), que admite a função de distribuição normal para expressar a relação não linear entre as probabilidades estimadas da variável dependente e as variáveis explicativas.

O modelo probit é utilizado para descrevermos a relação entre uma variável binária ou dicotômica e uma ou mais variáveis explicativas.

A construção do modelo *probit* é dada quando se possui uma variável y que toma um dos dois valores, 0 e 1. Defina-se uma variável latente y^* tal que

$$y_i^* = X_i\beta + \epsilon_i \quad (3-5)$$

Entretanto, nós não observamos y^* , só observamos y que toma valores 0 ou 1 de acordo com a seguinte regra:

$$y_i = \begin{cases} 1 & \text{se } y_i^* > 0 \\ 0 & \text{caso contrário} \end{cases} \quad (3-6)$$

Também se assume que $\epsilon \sim N(0, \sigma^2)$.

Os parâmetros do modelo *probit* são estimados pelo método da Máxima Verossimilhança, método de estimação não linear.

Os coeficientes estimados do modelo probit não têm uma interpretação fácil e direta. Para serem interpretados os coeficientes devem ser marginalizados. Para maiores detalhes sobre o procedimento de estimação dos parâmetros e a construção dos procedimentos de inferência, ver Johnston e DiNardo (2001).

Tobit

O modelo foi desenvolvido em 1958 por James Tobin, economista e Prêmio Nobel.

O modelo é uma extensão dos modelos *Probit* e *Logit*. Pela semelhança com as formulações desses dois modelos, o modelo ficou conhecido como *tobit*.

Este modelo consiste em utilizar amostras censuradas. Uma amostra censurada é quando a informação sobre a variável resposta está disponível somente para algumas observações, por isso, o modelo *tobit* também é conhecido como regressão censurada.

Muitos autores têm utilizado o modelo DEA para analisar eficiências das organizações. Contudo, no modelo DEA os escores de eficiência encontram-se limitados entre 0 e 1, o que dificulta a aplicação dos métodos de mínimos quadrados ordinários (MQO), devendo-se utilizar uma regressão do tipo *tobit*.

O modelo *tobit* pode ser descrito da seguinte forma

$$y^* = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + \epsilon_i \quad (3-7)$$

onde,

$$y_i = \begin{cases} y^* & \text{se } y_i^* > 0 \\ 0 & \text{caso contrário} \end{cases} \quad (3-8)$$

e

$$\epsilon \sim N(0, \sigma^2) \quad (3-9)$$

A equação do modelo *tobit* pode ser considerada como a de um modelo de regressão linear, no qual y^* (variável latente) substitui o y . Esta substituição é necessária devido ao modelo *tobit* possuir dados censurados ou truncados. Este y^* pode assumir valores à direita, à esquerda ou em um determinado intervalo, podendo ser reescrito como

$$y_i = \max(0, \beta X_i + \epsilon_i) \quad (3-10)$$

Em relação à estimação dos parâmetros o método de máxima verossimilhança é preferível por ter melhores características nesse tipo de modelo.

Assim como no modelo *probit*, os coeficientes estimados do modelo *tobit* também não são interpretáveis. Para ter uma interpretação dos coeficientes é necessário marginalizar os coeficientes.

O procedimento de inferência é baseado nas hipóteses apresentadas em (3-4) e a significância da variável, a rejeição da hipótese nula, será feita analisando o p – valor do parâmetro.

Neste trabalho, os procedimentos de estimação e inferência sobre os parâmetros são realizados utilizando o *software STATA*, versão 10 *SE*.

Abaixo, segue uma tabela da saída, apresentada pelo *software STATA*, e a sua interpretação. Esse exemplo foi retirado do *link* http://www.ats.ucla.edu/stat/stata/output/Stata_tobit.htm.

Tabela 3.1: Exemplo de saída do modelo *tobit*

tobit gre gpa topnotch, ll(200) ul(800)

Number of obs = 400

LR chi2(2) = 70,93

Prob > chi2 = 0,0000

Pseudo R2 = 0,0150

Log likelihood = -2331,4314

gre	Coefficiente	Erro	t	$P > t $	Intervalo de Confiança (95%)	
gpa	111,3085	15,19665	7,32	0	81,43273	141,1842
topnotch	46,65774	15,75356	2,96	0,003	15,68716	77,62833
_cons	205,8515	51,24073	4,02	0	105,1152	306,5879
σ	111,4882	4,143727			103,3419	119,6345

Obs. summary:

0 left-censored observations

375 uncensored observations

25 right-censored observations at $gre \geq 800$

Nessa saída temos que:

- Number of obs é o número de observações no conjunto de dados em que existe informação para todas as variáveis preditoras.
- LR chi2 (2) é a Razão de Verossimilhança (LR) do teste quiquadrado em que pelo menos um dos coeficientes da regressão é diferente de zero. O número entre parênteses indica os graus de liberdade da distribuição utilizados para testar a $LRchi2$, que é definido pelo número de preditores no modelo.
- $Prob > chi2$ é a probabilidade de se obter uma estatística de teste LR tão extrema quanto, ou mais, do que a estatística observada sob a hipótese nula. A hipótese nula é que todos os coeficientes de regressão são simultaneamente iguais a zero. Em outras palavras, esta é a probabilidade de se obter uma estatística quiquadrado (70,93), ou ainda, se de fato não há nenhum efeito das variáveis preditoras. Esta probabilidade é o p -valor, já descrito acima desse teste. Quanto menor mais evidências temos para dizer que pelo menos um dos coeficientes é diferente de zero.
- Pseudo R2 é pseudo R2 descrito por McFadden. A regressão *Tobit* não possui um R^2 equivalente ao R^2 que se encontra em modelos de

regressão por mínimos quadrados. No entanto, muitas pessoas têm tentado chegar a um. Há uma grande variedade de estatísticas para o Pseudo R². Como esta estatística não possui o mesmo sentido do R² da regressão de mínimos quadrados (a proporção da variância da variável resposta explicada pelos preditores), sugere-se interpretá-la com cautela.

- Log likelihood é a verossimilhança do modelo ajustado. Ela é utilizada no teste de probabilidade quiquadrado.
- Coeficiente de fato são os coeficientes estimados para cada variável independente pertencente ao modelo.
- gre é a variável dependente do modelo ajustado.
- gpa é uma das variáveis independentes do modelo ajustado.
- topnotch é uma das variáveis independentes do modelo ajustado.
- _const é o intercepto, ou coeficiente angular, do modelo ajustado.
- σ é o desvio padrão da distribuição do erro do modelo
- Erro de fato consiste nos erros padrão dos coeficientes de regressão. Ele é utilizado no cálculo da estatística de teste t e no intervalo de confiança do coeficiente de regressão.
- t é a estatística t, que é a divisão do Coeficiente pelo Erro. Este t é utilizado para realizar o teste bilateral de que o parâmetro é diferente de zero.
- $P > |t|$ é a probabilidade da estatística de teste t (ou um teste estatístico mais extremo) seja observado sob a hipótese nula de que um particular coeficiente preditor da regressão é zero, dado que o resto dos preditores estão no modelo. Para um dado nível α , $P > |t|$ determina se a hipótese nula pode ser rejeitada ou não. Se $P > |t|$ é menor que α , então a hipótese nula pode ser rejeitada.
- Intervalo de Confiança (95%) é o Intervalo de Confiança (IC) para o coeficiente, dado que todos os outros preditores estão no modelo ao nível de 95%.
- Obs. summary indica a quantidade de observações no conjunto de dados que são censurados. Aqui vemos que nenhum dos registros são censurados à esquerda (todos são superiores a 200) e 25 dos registros são censuradas à direita (maior ou igual a 800).

3.4

Aplicação de DEA e tobit

A aplicação dos dois métodos é bastante utilizada na literatura em diversas áreas: o modelo DEA para gerar o *score* da eficiência e a regressão *tobit* para analisar o impacto dos fatores na eficiência.

Barbosa e Wilhelm (2009) analisaram a influência de fatores sociais e econômicos no desempenho das escolas de ensino fundamental e médio do Núcleo Regional de Educação de Paranavaí. Eles utilizaram o modelo DEA para "avaliação cruzada" que determinou os índices de desempenho das escolas, e a regressão *tobit* para destacar os fatores sociais e econômicos dos familiares dos alunos no desempenho das escolas.

Santos *et al.* (2009) analisaram a eficiência técnica, através do modelo DEA, de 228 talhões de café das regiões do Cerrado e Sul de Minas Gerais, identificando seus principais determinantes pelo modelo *tobit*.

Silva (2009) estudou eficiência técnica de pequenos e médios produtores de camarão no Rio Grande do Norte, obtendo escores de eficiência por meio da estimação das fronteiras de produção não paramétricas, usando DEA e FDH. Em seguida determinou, pelo modelo *tobit* determinou a associação entre os escores de eficiência e indicadores de gestão dos carcinicutores, assim como da localização de suas fazendas.

Ferreira *et al.* (2007) investigaram o desempenho das Cooperativas de Economia e Crédito Mútuo de Minas Gerais. Os fatores condicionantes da eficiência foram identificados por meio do modelo *tobit*.

Souza (2006) estudou os modelos estatísticos censurados e truncados na avaliação de efeitos técnicos que afetam a fronteira de produção dos centros de pesquisa da Embrapa (Empresa Brasileira de Pesquisa Agropecuária). As medidas de eficiência técnica foram calculadas com base na DEA e na Análise de Fronteiras Estocásticas de Produção. O autor conclui que as medidas de eficiência DEA são mais informativas para a avaliação dos efeitos técnicos e que a geração de renda, as ações de parceria, a qualidade técnica dos projetos e o tamanho são fatores significantes. Já a análise *tobit* foi consistente com o ajuste de uma distribuição normal truncada aos resíduos obtidos de uma função de produção DEA.

Na dissertação de Rios (2005) analisou-se a eficiência relativa das operações dos terminais de contêineres do Mercosul nos anos de 2002 até 2004, utilizando os modelos CCR e BCC e a regressão *tobit* para

verificar as variáveis que mais influenciam a eficiência. O autor resumiu alguns trabalhos com DEA e regressão *tobit*, como mostra a figura 3.9 que foi retirada da dissertação.

Autores	Amostra	Objetivo
Kirjavainen e Loikkanen (1998)	291 escolas secundárias da Finlândia	Relação da eficiência com o tipo de escola
Shao e Lin (2002)	370 firmas	Efeitos da TI na eficiência
Nasierowski e Arcelus (2003)	45 países	Efeitos da eficiência e produtividade no esforço tecnológico
Marinho (2003)	73 municípios do rio de Janeiro	Relações entre eficiência e indicadores econômicos
Watcharasriroj e Tang (2004)	92 hospitais sem fins lucrativos na Tailândia	Efeitos do tamanho e TI na eficiência
Grosskopf et al. (2004)	254 hospitais de ensino	Relação da eficiência com a competitividade
Scheraga (2004)	38 companhias de aviação	Efeitos da eficiência na mobilidade financeira
Krasachat (2004)	74 fazendas de arroz da Tailândia	Relação da eficiência com fatores específicos das fazendas
Boame (2004)	30 sistemas de trânsito do Canadá	Relação entre eficiência e outras variáveis
Turner et al. (2004)	26 terminais de contêineres dos Estados Unidos e Canadá	Relação da eficiência com variáveis de infraestrutura

Figura 3.9: Resumo dos estudos com DEA e regressão *tobit* até 2004