



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

**PREDIÇÃO DA EFICIÊNCIA BANCÁRIA
UTILIZANDO A TÉCNICA DE FRONTEIRA
ESTOCÁSTICA**

por

Guilherme Maia Rodrigues Gomes 09/95771

Brasília

2012

Guilherme Maia Rodrigues Gomes 09/95771

**PREDIÇÃO DA EFICIÊNCIA BANCÁRIA
UTILIZANDO A TÉCNICA DE FRONTEIRA
ESTOCÁSTICA**

Relatório apresentado à disciplina Estágio Supervisionado II do curso de graduação em Estatística, Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: Prof. Dr. Alan Ricardo da Silva

Brasília

2012

Dedico este trabalho à quem me ensinou a sorrir pelo simples fato de poder viver mais um dia, minha avó Iramy Rodrigues Gomes (*in memoriam*).

Guilherme Maia Rodrigues Gomes

“Success is the ability to go from one failure to another with no loss of enthusiasm.”

Sir Winston Churchill

Agradecimentos

Aos meus mentores Alan Ricardo da Silva e Benjamin Tabak, esse por todas as oportunidades dadas e por toda confiança que sempre foi oferecida, e aquele por toda ajuda, por ter sido orientador desse trabalho e, após ter aceito minha proposta de tema, tomo-a como um desafio se tornando um constante incentivador.

Aos meus pais Solange e Walmir, pelo apoio incondicional, tanto no âmbito financeiro quanto emocional, por serem exemplos de pessoas e por terem me ensinado o que jamais poderei aprender em livro nenhum: ser íntegro, destemido, gentil e persistente.

Aos meus irmãos Gustavo e Giovanna, pela compreensão, pelo amor, por toda grandeza que têm, apesar de ainda não saberem.

Aos meus familiares e amigos, em especial à minha avó Antônia Maia, por ter me garantido oportunidades únicas em minha vida.

À Fernanda Duarte de Lima, pelo amor e carinho, pelo incentivo e compreensão, principalmente, por ser a sonhadora que é e ter me ensinado a acreditar nos meus sonhos. Agradeço também por ter se tornado a melhor companheira que alguém pode ter, em todos os sentidos que essa palavra possui.

Resumo

Um dos principais objetivos de uma grande firma de qualquer setor econômico é ser eficiente. A alocação certa do recursos bem como a utilização de insumos para produção máxima são *trade – offs* que os gestores fatalmente enfrentam diariamente. Objetivando saber se as decisões tomadas são eficientes com relação a outras firmas desse mesmo setor econômico foi criada metodologia de determinação da eficiência, *Data Envelopment Analysis* (DEA) . Sabe-se, porém, que nenhuma firma utiliza apenas um insumo para produzir seu produto e é bem comum as firmas de um mesmo setor utilizam insumos diferentes em sua produção, sendo assim esse método determinístico passa ser demasiadamente relativo ao que se está estudando. A Fronteira Estocástica de Produção foi estruturada devido a busca de uma metodologia probabilística que, além de ser mais suave, pudesse ser mais explicativa em termos econômicos. O intuito desse trabalho é descrever, implementar, comparar e aplicar a metodologia de predição de eficiência técnica por meio de uma Função Fronteira Estocástica. Foi utilizado dados de bancos comerciais brasileiros para mostrar a eficácia desse método comparado ao DEA.

Abstract

One of the main goals of a large firm of any economic sector is to be efficient. Allocating the right resources and using inputs for maximum production are trade-offs that managers inevitably face daily. Aiming to know whether the decisions taken are effective or not against others same sector's firms, it was established a methodology for determining the efficiency, Data Envelopment Analysis (DEA). It is known, however, that no firm uses only one input to produce your product and it is quite common for firms from the same industry to use different inputs in production, so this deterministic method is being overly relative to what is being studied. The Stochastic Frontier Production was structured due an aiming of a probabilistic methodology that, besides smoother, could be more explanatory in economic terms. The purpose of this paper is to describe, implement, compare, and apply the methodology to predict technical efficiency through a Stochastic Frontier Function. Brazilian commercial banks data was used to show the effectiveness of this method compared to DEA.

Lista de Tabelas

2.1	Rendimento de Escala vs Tecnologia.	10
2.2	Funções de Produção.	13
3.1	Rotinas de maximização que podem ser usadas	33
3.2	Comparação das estimativas entre os algoritmos de estimação <i>Cross- Sectional</i>	34
3.3	Comparação das estimativas entre os algoritmos para dados em painel	37
4.1	Análise descritiva dos dados - base anual	43
4.2	Análise descritiva dos dados - base semestral	44
4.3	Média das Eficiências Utilizando o DEA por controle	46
4.4	Estimativas dos parâmetros utilizando a base anual e semestral	47
4.5	Média das Eficiências Utilizando o Fronteira Estocástica por controle	50

Lista de Figuras

2.1	Função de produção para um insumo	8
2.2	Isoquantas de CES_{12}	12
2.3	Estrutura da Fronteira Estocástica de Produção	18
2.4	Comparação entre modelagens	22
2.5	Fronteiras: DEA vs Estocástica	26
3.1	Comparação entre o modelo Pooled e Paineis	38
4.1	Gráficos de Dispersão	45
4.2	Eficiências médias por ano para cada tipo de controle bancário	48
4.3	Eficiências médias por semestre para cada tipo de controle bancário	49
4.4	Gráficos das Eficiências	50

Sumário

RESUMO	iv
ABSTRACT	v
1 INTRODUÇÃO	1
1.1 OBJETIVOS	3
2 FUNÇÃO FRONTEIRA ESTOCÁSTICA DE PRODUÇÃO	4
2.1 INTRODUÇÃO	4
2.2 FUNÇÕES DE PRODUÇÃO	5
2.2.1 Propriedades da Função de Produção	6
2.2.2 Medidas quantitativas de interesse	8
2.2.3 Retornos de Escala	9
2.2.4 Exemplos	10
2.3 TRANSFORMAÇÃO DAS TECNOLOGIAS DE PRODUÇÃO	12
2.3.1 Custo	14
2.3.2 Receita	14
2.3.3 Lucro	15
2.4 ANÁLISE POR ENVELOPAMENTO DE DADOS (DEA)	15

2.5	FUNÇÃO FRONTEIRA ESTOCÁSTICA DE PRODUÇÃO	17
2.5.1	Modelo <i>Half-Normal</i>	19
2.5.2	Outros modelos comuns	22
2.5.3	Estimação de Modelos Não-Lineares e Testes de Hipóteses . . .	23
2.5.4	Predizendo a eficiência técnica	25
2.6	MODELO DE DADOS EM PAINEL	27
2.6.1	Modelos de Ineficiência Invariantes no tempo	28
2.6.2	Modelos de Ineficiência Variantes no tempo	28
3	ALGORITMO DESENVOLVIDO	31
3.1	DADOS EM CORTE TRANSVERSAL	31
3.2	DADOS EM PAINEL PARA UM EFEITO ALEATÓRIO	34
4	EFICIÊNCIA ESTOCÁSTICA DOS BANCOS BRASILEIROS	40
4.1	INTRODUÇÃO	40
4.2	METODOLOGIA	40
4.3	BASE DE DADOS	41
4.4	RESULTADOS EMPÍRICOS	42
4.4.1	Análise Descritiva	42
4.4.2	Construção da Fronteira Estocástica de Custo	46
4.4.3	Comparação com o DEA	50
5	CONCLUSÕES	52
5.1	Limitações do Trabalho	53

5.2	Trabalhos Futuros	53
	Referências	54
	APÊNDICES	57
.1	Códigos para construção das bases de dados	57
.2	Derivações da densidade Normal truncada	59

Capítulo 1

INTRODUÇÃO

O estudo da função de fronteira estocástica de produção, proposto por Aigner et al. (1977) e Meeusen and van den Broeck (1977), tem obtido significativas contribuições da modelagem econométrica, em particular da estimação de eficiência técnica de firmas em um determinado setor econômico. A função de fronteira estocástica está envolvida com dois componentes aleatórios, um associado com a ineficiência técnica (compreendida entre 0 e 1) e o outro com um erro aleatório. O termo ineficiência está relacionado a quanto determinada firma é menos eficiente comparada com a fronteira estimada no modelo. Sendo assim, a eficiência propriamente dita é dada por $1 - \textit{ineficiência}$. Estudos de funções de fronteira podem envolver tanto dados em um determinado tempo como dados em painel. Defini-se então, uma função de produção como:

$$y_{it} = f(x_{it}, p_{it}; \beta) e^{V_{it} - U_{it}} \quad (1.1)$$

e

$$U_{it} = z_{it}\delta + W_i t \quad (1.2)$$

onde:

- y_{it} denota a variável de produção (ou Lucro) no t -ésimo tempo para a i -ésima firma;
- x_{it} é um vetor $(1 \times k)$ de variáveis de insumos;
- p_{it} é um vetor $(1 \times m)$ de variáveis de produtos;
- β é o vetor $((k + m) \times 1)$ de parâmetros não conhecidos a serem estimados;
- V_{it} é o termo do erro aleatório assumido ser iid $N(0, \sigma_v^2)$ e independente de U_{it} ;
- Considera-se os U_{it} 's variáveis aleatórias associadas a ineficiência técnica de produção com U_{it} seguindo uma distribuição normal truncada no 0 com média $z_{it}\delta$ e variância σ^2 ;
- z_{it} é um vetor $(1 \times m)$ de variáveis aleatórias associadas a ineficiência técnica da produção das firmas no tempo;
- δ é um vetor $(m \times 1)$ dos coeficientes não conhecidos; e
- $W_{it} \sim IID \text{ half-normal}(0, \sigma^2)$

A estimação dos parâmetros da função de fronteira estocástica de produção será feita por meio da função de máxima verosimilhança em equações simultâneas (Battese and Coelli, 1995). Desta forma, temos que a eficiência técnica, proposta por Battese and Coelli (1993), é dada por:

$$TE_{it} = e^{-U_{it}} = e^{z_{it}\delta + W_{it}} \quad (1.3)$$

Há pacotes com essa metodologia implementada, no entanto com algumas limitações. No SAS, por exemplo, o procedimento QLIM só estima a fronteira para

dados em *cross-sectional* e não prediz a eficiência individual. Já os *softwares* R e STATA predizem a eficiência e possuem a metodologia para dados em painel, no entanto o R utiliza um como base de estruturação do algoritmo o artigo Battese and Coelli (1992), sendo que o STATA já utiliza Battese and Coelli (1995). Dessa forma, seus resultados são diferentes. A desvantagem desse último é o pouquíssimo número de alternativas de otimizações numéricas.

Nos últimos anos, verificou-se um crescimento no número de estudos sobre eficiência direcionados ao setor bancário (Marra et al., 2011), buscando melhor compreender as razões pelas quais alguns bancos são menos eficientes do que outros. Nesse sentido buscar-se-á avaliar o sistema bancário brasileiro utilizando esta metodologia para estimar valores individuais de eficiência.

1.1 OBJETIVOS

O objetivo geral do trabalho é implementar a metodologia de estimação da eficiência estocástica usando a função de fronteira estocástica de produção.

Os objetivos específicos são:

- conduzir sua implementação via PROC IML no *software* SAS 9.2;
- mostrar o poder e a aplicabilidade da metodologia, comparando-a com o DEA (Análise por Envolvimento de Dados); e
- prever a eficiência estocástica de custo dos bancos brasileiros.

Capítulo 2

FUNÇÃO FRONTEIRA ESTOCÁSTICA DE PRODUÇÃO

2.1 INTRODUÇÃO

A função de produção de uma determinada firma relaciona os recursos usados e a produção no final de um determinado período. Considera-se assim um determinado setor econômico em que os insumos de cada firma são variáveis explicativas e a produção de um determinado produto é a variável resposta. Maximizando esse modelo teórico de produção, obtém-se um comportamento de referência, conhecido como fronteira. A primeira estimação de fronteira de produção foi proposta por Aigner and Chu (1968) em que considerava uma função de produção Cobb-Douglas da forma:

$$\ln(q_i) = \mathbf{x}_i\boldsymbol{\beta} - u_i \quad i = 1, \dots, n \quad (2.1)$$

onde q_i representa o produto da i -ésima firma; \mathbf{x}_i é um vetor ($1 \times k$) contendo o logaritmo dos insumos; $\boldsymbol{\beta}$ um vetor ($k \times 1$) de parâmetros desconhecidos; e u_i uma variável aleatória não-negativa associada a ineficiência técnica. Muitas metodologias, como a *Data Envelopment Analysis* (DEA), são usados para predizer a eficiência da firma a partir da Equação 2.1. No entanto, um problema que segue dessa é que

todos os desvios da fronteira são dados como resultado da ineficiência técnica, ou seja além de possuir ruídos assimétricos, prejudicando a estimação dos parâmetros, esse modelo não considera os desvios naturais relacionados a omissão de outras variáveis importantes (Coelli et al., 2005).

Para minimizar esse problema Aigner et al. (1977) e Meeusen and van den Broeck (1977) propuseram a função fronteira estocástica de produção que incluía o termo estocástico v_i , uma variável aleatória simétrica, relacionado aos resíduos do modelo e assumindo o papel de conter os desvios dados pela omissão de variáveis importantes. O modelo é dado da seguinte forma:

$$\ln(q_i) = \mathbf{x}_i\beta + v_i - u_i \quad (2.2)$$

Sendo assim, o foco deste estudo será a fronteira estocástica de produção dada pela Equação 2.2.

2.2 FUNÇÕES DE PRODUÇÃO

Esta seção foi destinada a formalização da Função de Produção. Do jargão econômico, alguns termos serão previamente conceituados dado que serão amplamente utilizados ao longo do texto. Os termos são:

- produtividade: é a razão de produtos sobre insumos, sendo que dada uma função de produção sua derivada representa a produtividade;
- eficiência técnica: utilização, por parte de uma firma, de seus insumos para produção máxima de produtos possível;

- eficiência alocativa: seleção correta dos insumos para obtenção máxima de produtos;
- economia de escala: ocorre quando a produtividade de uma firma é igual a 1, ponto ótimo de escala. Firms eficientes operam em economia de escala ou muito próximo disso;
- mudança técnica: está relacionada a inovações tecnológicas que podem tanto aumentar como diminuir a eficiência ao longo do tempo;
- isoquantas: são as curvas de níveis da função de produção, representando a relação entre os insumos 2 a 2.
- fronteira de produção: comportamento de referência em um determinado setor econômico que determina o quanto certa firma é ineficiente comparada a fronteira.

Esses conceitos são fundamentais para compreender as próximas formalizações.

2.2.1 Propriedades da Função de Produção

Considerando uma firma com N insumos (matéria-prima, trabalho, maquinaria etc) que produz um único produto, temos a seguinte função da produção de q :

$$q = f(\mathbf{x}) \tag{2.3}$$

onde q representa o produto e $\mathbf{x} = (x_1, x_2, \dots, x_N)$ um vetor(1 x N) de insumos.

Tendo em vista viabilizar a análise econômica dessa firma, algumas propriedades são dadas. A função $f(\mathbf{x})$ deve ser:

1. Não negativa: $0 < f(\mathbf{x}) < \infty$ tal que $f(\mathbf{x}) \in \mathbb{R}$.
2. Essencialmente Fraca: A produção de um produto é impossível sem ao menos um insumo.
3. Não decrescente em \mathbf{x} : $f(\mathbf{x})$ é monotonicamente crescente, ou seja, para $\mathbf{x}_0 \geq \mathbf{x}_1$, tem-se $f(\mathbf{x}_0) \geq f(\mathbf{x}_1)$. Isso implica que $f'(\mathbf{x}) \geq 0 \forall \mathbf{x} \in \Theta$, sendo Θ suporte de \mathbf{x} .
4. Côncava em \mathbf{x} : $f''(\mathbf{x}) < 0$, implicando que todos os produtos marginais são não crescentes (lei da produtividade marginal decrescente).

Essas propriedades, no entanto, não são universais (Coelli et al., 2005). Por exemplo, o caso em que há um grande número de insumos, esse problema, chamado de congestionamento de recursos, leva ao relaxamento da propriedade de monotonicidade. Já para o caso em que temos um insumo produzindo mais de um produto, a propriedade de concavidade em \mathbf{x} não é satisfeita.

A Figura 2.1 mostra um exemplo de uma função de produção definida sobre um único insumo. Verifica-se violação das propriedades enunciadas anteriormente. Primeiramente, na região 0-D há problemas com a concavidade, tendo em vista que está voltada para cima, $f''(x) > 0$, já em GR, temos que a propriedade da monotonicidade é violada. É interessante destacar o que representam os pontos D, E e G: o primeiro representa a mudança de concavidade, $f''(x) = 0$, onde começa a região chamada de economicamente factível; o ponto E, por sua vez, representa o ponto ótimo de escala, ou seja, produtividade = 1, $f'(x) = 1$; e, finalmente, o ponto G representa o final da região econômica factível dado que apartir desse ponto

$f'(x) < 0$, ou seja, a função passa a ser decrescente contrapondo a propriedade de monotonicidade.

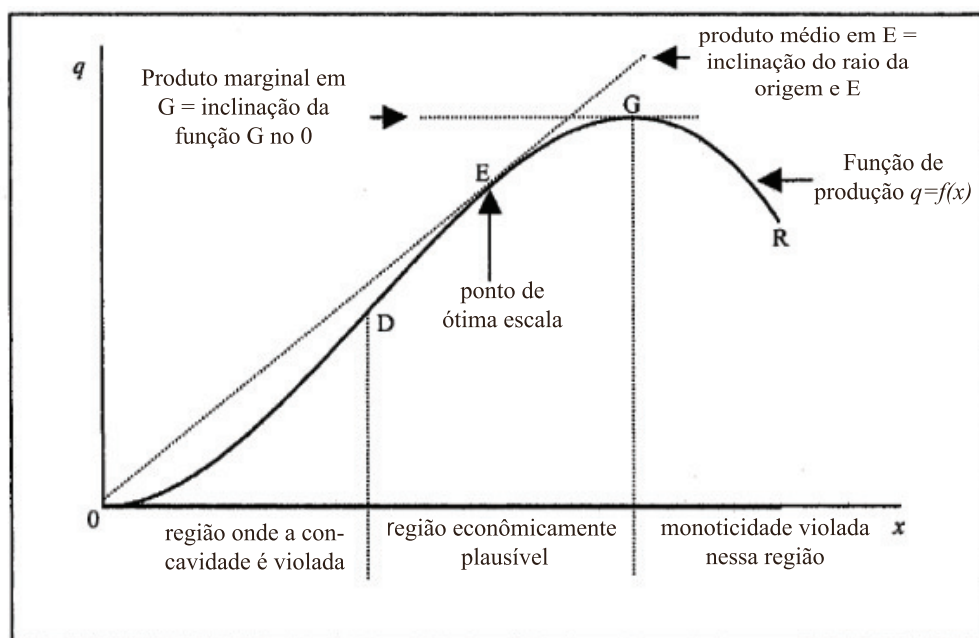


Figura 2.1: Função de produção para um insumo
 Fonte: Coelli et al. (2005).

Pode-se dizer que para o caso de um único insumo, a região economicamente factível, a visualização é direta, no entanto, na prática, a produção de um produto jamais será função de apenas um recurso. É essencial a utilização de medidas quantitativas para entender o processo de produção de determinada firma e assim determinar suas características.

2.2.2 Medidas quantitativas de interesse

Considera-se uma função de produção da forma $f(\mathbf{x})$, com $\mathbf{x} = (x_1, x_2, \dots, x_n)$, temos as seguintes medidas:

- Produto marginal (ou produtividade):

$$PM_i = \frac{\partial f(\mathbf{x})}{\partial x_i} \quad (2.4)$$

mede o quanto a produção aumentará se aumentarmos uma unidade do insumo i , mantendo a quantidade dos outros insumos fixa;

- Taxa Técnica de Substituição:

$$TTS_{ij} = -\frac{PM_j}{PM_i} \quad (2.5)$$

mede o quanto a firma deve abrir mão do insumo i e acrescentar em j para continuar produzindo a mesma quantidade;

- Elasticidade Produto:

$$E_i = \frac{\partial f(\mathbf{x})}{\partial x_i} \frac{x_i}{f(\mathbf{x})} \quad (2.6)$$

mede qual a percentagem de mudança no produto variando 1% o insumo i ; e

- Elasticidade de Substituição:

$$DES_{ij} = \sigma_{ij} = \frac{d(x_j/x_i)}{d(PM_i/PM_j)} \frac{PM_i/PM_j}{x_j/x_i} \quad (2.7)$$

mede a percentagem de mudança na razão dos insumos relacionada com a percentagem de mudança na taxa de substituição, em termos formais mede a curvatura da isoquanta dos insumos i e j .

2.2.3 Retornos de Escala

Como foi visto anteriormente o produto marginal de um insumo mede quanto a produção aumenta ao aumentar em uma unidade esse insumo, considerando fixos

todos os outros, no entanto é interessante saber a resposta da produção variando os insumos simultaneamente, leia-se proporcionalmente. Desta forma, a Tabela 2.1 mostra dado o tipo de rendimento de escala qual deve ser o tipo de tecnologia a ser usado. Seja $f(\mathbf{x})$ um função de produção, onde \mathbf{x} denota o vetor de insumos:

Tabela 2.1: Rendimento de Escala vs Tecnologia.

Tipo de Rendimentos de escala	Formato da Função de Produção	Tipo de tecnologia
Constantes	$f(t\mathbf{x}) = tf(\mathbf{x}), t > 0, \forall \mathbf{x}$	replica
Decrescente	$f(t\mathbf{x}) < tf(\mathbf{x}), t > 1, \forall \mathbf{x}$	algum fator fixo
Crescente	$f(t\mathbf{x}) > tf(\mathbf{x}), t > 1, \forall \mathbf{x}$	ganhos de escala

As medidas vistas nessas seção não somente descrevem as características da produção específicas de cada firma como também indicam a violação ou não das propriedades enunciadas anteriormente.

2.2.4 Exemplos

Para ilustrar o cálculo de produto marginal e das elasticidades, considera-se um firma com sua produção perfeitamente descrita por dois insumos - capital(k) e trabalho (l) - com base na função Cobb-Douglas, dada por:

$$q = 2x_k^{0,5}x_l^{0,4}$$

Temos então os seguintes produtos marginais:

$$PM_k = \frac{\partial q}{\partial x_k} = x_k^{-0,5}x_l^{0,4}$$

$$PM_l = \frac{\partial q}{\partial x_l} = 0,8x_k^{0,5}x_l^{-0,6}$$

Consequentemente, a Taxa de Substituição é:

$$TTS_{kl} = -0,8 \left(\frac{x_k}{x_l} \right)$$

Temos também as seguintes elasticidades produto:

$$E_k = \frac{\partial q}{\partial x_k} \frac{x_k}{q} = 0,5$$

$$E_l = \frac{\partial q}{\partial x_l} \frac{x_l}{q} = 0,4$$

E, finalmente, a elasticidade de substituição é dada por:

$$\sigma_{kl} = \frac{d(x_l/x_k)}{d(PM_k/PM_l)} \frac{PM_k/PM_l}{x_l/x_k} = 0,8 \frac{1}{0,8} \left(\frac{x_l}{x_k} \right) \left(\frac{x_k}{x_l} \right) = 1 \quad (2.8)$$

Pode-se assim mostrar que se esta firma aumentar em 1% o capital, a produção aumentará 0,5%, já se aumentar em 1% o trabalho a produção só aumentará 0,4% para qualquer nível de capital ou trabalho. Outro resultado interessante é a elasticidade de substituição igual 1, ou seja, a razão dos insumos é igual a taxa de substituição. Ambos resultados são propriedades restritivas da tecnologia Cobb-Douglas.

Outro exemplo é dado pela função CES (Elasticidade de Produção Constante). As Figuras em 2.2 são isoquantas de casos especiais da função de produção CES para dois insumos. Sendo assim, podemos determinar sua curvatura da seguinte forma:

$$CES(x_1, x_2; \gamma) = (x_1^\gamma + x_2^\gamma)^{\frac{1}{\gamma}}$$

Os casos especiais são:

- $\gamma \rightarrow -\infty$ (Figura 2.2a)

$$\lim_{\gamma \rightarrow -\infty} (x_1^\gamma + x_2^\gamma)^{\frac{1}{\gamma}} = \min x_1, x_2 \Rightarrow \lim_{\gamma \rightarrow -\infty} \sigma_{12} = 0$$

- $\gamma \rightarrow 0$ (Figura 2.2b)

$$\lim_{\gamma \rightarrow 0} (x_1^\gamma + x_2^\gamma)^{\frac{1}{\gamma}} = x_1^\alpha x_2^{1-\alpha} \Rightarrow \lim_{\gamma \rightarrow 0} \sigma_{12} = 1$$

- $\gamma \rightarrow 1$ (Figura 2.2c)

$$\lim_{\gamma \rightarrow 1} (x_1^\gamma + x_2^\gamma)^{\frac{1}{\gamma}} = x_1 + x_2 \Rightarrow \lim_{\gamma \rightarrow 1} \sigma_{12} = \infty$$

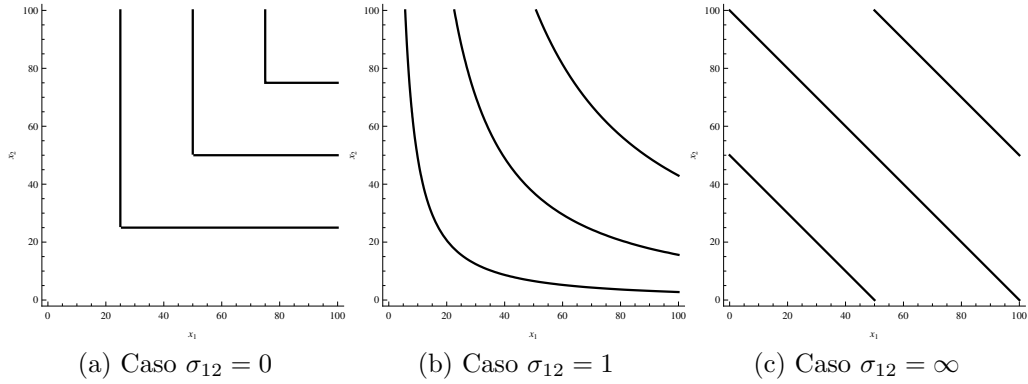


Figura 2.2: Isoquantas de CES_{12}

Além da função Cobb-Douglas e da CES, outros modelos são usadas para descrever a produção da firma. Abaixo algumas funções comuns:

2.3 TRANSFORMAÇÃO DAS TECNOLOGIAS DE PRODUÇÃO

Considerando o problema da firma em decidir a melhor alocação de seus insumos, pode-se tomar como base a minimização dos custos, a maximização da receita ou levando em conta os dois anteriores simultaneamente, ou seja, a maximização do lucro. Assim, a função de produção, vista anteriormente, será usada nessa seção para estender o raciocínio a custos, receitas e finalmente o lucro. Passaremos a trabalhar em termos monetários (preços), deixando a relação física (quantidade) como suporte. Para começar, uma forma geral para as transformações das tecnologias de produção podem ser dadas da seguinte forma:

$$T(\mathbf{x}, q) = q - f(\mathbf{x}) = 0 \tag{2.9}$$

Tabela 2.2: Funções de Produção.

Linear	$y = \beta_0 + \sum_{n=1}^N \beta_n x_n$
Quadrática	$y = \beta_0 + \sum_{n=1}^N \beta_n x_n + \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \beta_{nm} x_n x_m$
Quadrática Normalizada	$y = \beta_0 + \sum_{n=1}^{N-1} \beta_n \frac{x_n}{x_N} + \frac{1}{2} \sum_{n=1}^{N-1} \sum_{m=1}^{N-1} \beta_{nm} \frac{x_n}{x_N} \frac{x_m}{x_N}$
Translog	$y = \exp \left(\beta_0 + \sum_{n=1}^N \beta_n \ln(x_n) + \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \beta_{nm} \ln(x_n) \ln(x_m) \right)$
Cobb-Douglas	$y = \beta_0 \prod_{n=1}^N x_n^{\beta_n}$
Generalização de Leontief	$y = \sum_{n=1}^N \sum_{m=1}^N \beta_{nm} (x_n x_m)^{\frac{1}{2}}$
Elasticidade de substituição constante (CES)	$y = \beta_0 \left(\sum_{n=1}^N \beta_n x_n^\gamma \right)^{\frac{1}{\gamma}}$

onde q é o produto e $f(\mathbf{x})$ a função de produção. Destaca-se que, dado ser uma transformação linear, as propriedades da função de produção permanecem inalteradas. Para mais detalhes ver (Coelli et al., 2005).

2.3.1 Custo

Considera-se um firma pequena o suficiente para ser uma tomadora de preços no mercado em que atua. Com isso temos que o problema da firma em minimizar os custos é dado por:

$$c(\mathbf{w}, q(\mathbf{x})) = \min_{\mathbf{x}} \mathbf{w}'\mathbf{x} \quad \text{tal que} \quad T(\mathbf{x}, q) = 0 \quad (2.10)$$

onde q representa a função de produção, \mathbf{x} um o vetor de insumos e $\mathbf{w} = (w_1, w_2, \dots, w_N)'$ é o vetor de preços dos insumos \mathbf{x} . Na Equação 2.10 a notação usada na função $c(\mathbf{w}, q)$ foi proposital para enfatizar que a minimização dos custos está relacionada com a variação correta em \mathbf{w} e q .

2.3.2 Receita

As funções teóricas de receita se assemelham muito com as de produção, no entanto dado que trabalha-se em termos monetários ao invés das quantidades, temos uma equação de preços de produtos combinada com as quantidades produzidas. Assim, se uma firma vende cada unidade do seu produto pelo preço p , sua receita é:

$$r(p, \mathbf{x}) = \max_{\mathbf{q}} \mathbf{p}'\mathbf{q} \quad \text{tal que} \quad T(\mathbf{q}, \mathbf{x}) = 0 \quad (2.11)$$

onde $\mathbf{p} = (p_1, p_2, \dots, p_M)'$ é um vetor de preços dos produtos sobre os quais a firma não tem influência, ou seja, considera-se um setor perfeitamente competitivo no mercado de produtos.

2.3.3 Lucro

Para estudar como é feita a transformação da função Lucro, deve-se pensar em simultaneamente minimizar os custos (2.10) e maximizar a produção (Receita) (2.12). Esse pensamento é razoável tendo em vista que uma firma faz suas decisões nos dois sentidos, maximizando suas receitas e minimizando seus custos. Sendo assim, temos a seguinte Equação:

$$\pi(\mathbf{p}, \mathbf{w}) = \max_{\mathbf{q}, \mathbf{x}} \mathbf{p}'\mathbf{q} - \mathbf{w}'\mathbf{x} \quad \text{tal que} \quad T(\mathbf{q}, \mathbf{x}) = 0 \quad (2.12)$$

onde \mathbf{p} e \mathbf{w} são os preços dos produtos e insumos, respectivamente. Pelo princípio da dualidade, Coelli et al. (2005) mostrou que as transformações, Custo, Receita e Lucro, preservam as informações do da função de produção. Uma outra importante implicação desse Teorema é que podem ser usados muitos tipos diferentes de função para representar todas as características relevantes de um determinado setor econômico.

2.4 ANÁLISE POR ENVELOPAMENTO DE DADOS (DEA)

O DEA é uma alternativa não paramétrica da função fronteira estocástica de produção. Estruturada de forma a criar uma superfície de comportamento ótimo sobre os dados, todas as firmas são comparadas a essa fronteira e tal relação é denominada de eficiência (Coelli et al., 2005).

Além de ser um método fácil de trabalhar, é construído por uma ideia intuitiva, pois não necessita de conhecimento específico da estrutura dos preços dos insumos,

sendo assim, por muito, o dispositivo preferido para predição de eficiência (Staub et al., 2009) e (Fries and Taci, 2005). A maximização da produção e por consequente a eficiência técnica é dada na forma envelopada a partir da Equação 2.13.

$$\begin{aligned}
 & \min_{\theta, \lambda} \theta, \\
 \text{Sujeito a} \quad & -\mathbf{q}_i + \mathbf{Q}\lambda \geq 0, \\
 & \theta \mathbf{x}_i - \mathbf{X}\lambda \geq 0, \\
 & \lambda \geq 0,
 \end{aligned} \tag{2.13}$$

onde θ é um escalar a ser estimado que representa a eficiência da i -ésima firma e λ é um vetor ($I \times 1$) de constantes, \mathbf{x}_i e \mathbf{q}_i representam o insumo e o produto, respectivamente, da i -ésima firma e $(\mathbf{X}\lambda, \mathbf{Q}\lambda)$ representa esses pontos projetados na fronteira.

Apesar da facilidade em predizer a eficiência, o DEA possui algumas falhas que podem prejudicar a análise. A mais importante, talvez seja que os postos de cada firma segundo a eficiência podem variar muito de acordo com a amostra que o pesquisador possui, dado que a fronteira é calculada sobre os dados e não se considera a amostragem feita. Outro ponto que pesa contra o DEA é que não há nesse dispositivo uma variável que acumularia os erros relacionados a omissão de variáveis, ou seja, a eficiência é calculada com base somente na combinação de produtos e insumos dada. Por fim, ao considerar dados longitudinais, o DEA por si só é inoperante. Há, no entanto, uma sugestão dada por Souza and Staub (2007) em que realiza-se dois procedimentos estatísticos que considerariam correlação serial, podendo assim, sob algumas restrições, predizer a eficiência.

2.5 FUNÇÃO FRONTEIRA ESTOCÁSTICA DE PRODUÇÃO

Proposta por Aigner et al. (1977) e Meeusen and van den Broeck (1977) a função fronteira estocástica de produção é dada da seguinte forma:

$$q_i = \underbrace{\exp(\mathbf{x}_i\beta)}_{\text{termo determinístico}} \times \underbrace{\exp(v_i)}_{\text{erro aleatório}} \times \underbrace{\exp(-u_i)}_{\text{ineficiência}} \quad (2.14)$$

onde q_i representa o produto da i -ésima firma; \mathbf{x}_i é um vetor ($1 \times k$) contendo o logaritmo dos insumos; β um vetor ($k \times 1$) de parâmetros desconhecidos; u_i uma variável aleatória não-negativa associada a ineficiência técnica; e $v_i \sim N(0, \sigma_v^2)$ o erro aleatório do modelo, relacionado a omissão de variáveis explicativas. O termo fronteira estocástica é devido a existir no modelo dado na Equação 2.14 uma fronteira superior dos produtos dada por uma variável aleatória $\exp(\mathbf{x}_i\beta + v_i)$, conforme a Figura 2.4. Para melhor entender os dois termos estocásticos, a Figura 2.3 mostra a estrutura da fronteira de produção q . Destaca-se duas firmas, A e B, onde a primeira possui um erro positivo e a segunda um ruído negativo, no entanto ao considerar o segundo termo aleatório, relacionado a ineficiência, percebe-se que a firma B é mais eficiente que a firma A.

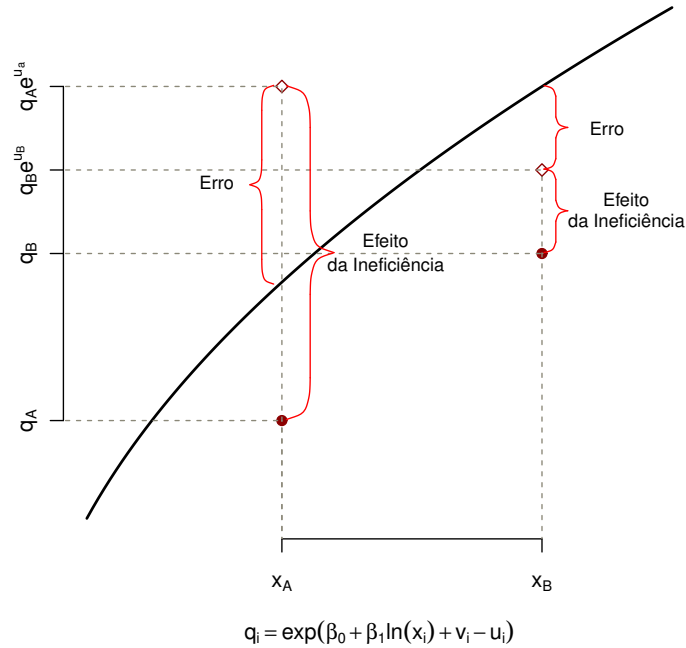


Figura 2.3: Estrutura da Fronteira Estocástica de Produção

Compreendida entre 0 e 1, a medida de eficiência é dada da seguinte forma:

$$TE_i = \frac{q_i}{\exp(\mathbf{x}_i \beta + v_i)} = \frac{\exp(\mathbf{x}_i \beta + v_i - u_i)}{\exp(\mathbf{x}_i \beta + v_i)} = \exp(-u_i) \quad (2.15)$$

Cada TE_i mede a relação do produto da i -ésima firma com o quanto poderia ser produzido caso esta firma fosse completamente eficiente. Claramente, a predição da eficiência é o passo seguinte a estimação dos parâmetros, tendo em vista que está em função dos parâmetros. O termo predição da eficiência será usado no lugar de estimação tendo em vista que trata-se de uma variável aleatória, não de um parâmetro.

Percebe-se pela Equação 2.14 que o único termo que não foi definido é o u_i , dado já que temos: uma parte determinística; e um erro aleatório que segue uma distribuição *Normal*. Desta forma, os diferentes modelos de estimação da fronteira

estocástica estão relacionados no que seja considerado a melhor suposição da distribuição desse termo aleatório.

Tendo em vista que o modelo possui dois termos estocásticos não conhecidos, a estimação torna-se um pouco mais complicada e estimadores comuns não poderão ser usados. O estimador de mínimos quadrados ordinários (MQO), por exemplo, é consistente para a estimação da inclinação das variáveis explicativas da fronteira de produção, no entanto o intercepto é viesado para baixo, tornando a predição da eficiência impossível. Há uma correção do viés utilizando o estimador *corrected ordinary least squares* (COLS) sugerido por Winston (1957), no entanto a estimação via máxima verossimilhança é assintoticamente mais eficiente (Coelli et al., 2005). Dentre os modelos mais recorrentes, suas funções de verossimilhança possuem em comum a não-linearidade e a complexidade das derivadas de primeira ordem.

2.5.1 Modelo *Half-Normal*

Modelo desenvolvido por Aigner et al. (1977) assume a distribuição assimétrica *half-normal* para o termo relacionado a eficiência técnica. Assim, um modelo de produção de um determinado setor econômico é dado por:

$$\mathbf{y} = \mathbf{x}_i\beta + \varepsilon_i \quad i = 1, \dots, N \quad (2.16)$$

onde $\varepsilon_i = v_i - u_i$, sendo $v_i \sim iidN(0, \sigma_v^2)$ independente de $u_i \sim iidN^+(0, \sigma_u^2)$. A função conjunta de v e u é dada da seguinte forma:

$$f(u, v) = \frac{1}{\pi\sigma_u\sigma_v} \exp\left(-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right) \quad (2.17)$$

Sendo que $v = \varepsilon + u$, temos:

$$f(u, \varepsilon) = \frac{1}{\pi\sigma_u\sigma_v} \exp\left(-\frac{u^2}{2\sigma_u^2} - \frac{(\varepsilon + u)^2}{2\sigma_v^2}\right) \quad (2.18)$$

A função de probabilidade de ε é dada pela função de probabilidade marginal da Equação 2.18, da seguinte forma:

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty f(u, \varepsilon) du \\ &= \int_0^\infty \frac{1}{\pi\sigma_u\sigma_v} \exp\left(\frac{-u^2(\sigma_v^2) - (\varepsilon + u)^2(\sigma_u^2)}{2\sigma_u^2\sigma_v^2}\right) du \end{aligned}$$

Utilizando a reparametrização proposta por Battese and Corra (1977), temos que $\sigma^2 = \sigma_v^2 + \sigma_u^2$ e $\lambda^2 = \sigma_u^2/\sigma_v^2$. Segue a seguinte expressão:

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty \frac{1}{\frac{\pi\lambda\sigma^2}{\lambda^2+1}} e^{-\left(\frac{u^2(\sigma^2)+\varepsilon^2\left(\frac{\lambda^2\sigma^2}{\lambda^2+1}\right)+2u\varepsilon\left(\frac{\lambda^2\sigma^2}{\lambda^2+1}\right)}{2\left(\frac{\lambda\sigma^2}{\lambda^2+1}\right)^2}\right)} du \\ &= \left(\frac{\lambda^2+1}{\pi\lambda\sigma^2}\right) \int_0^\infty e^{-\frac{1}{2\sigma^2}\left(\frac{\lambda^2+1}{\lambda}\right)^2\left[u^2+2u\varepsilon\left(\frac{\lambda^2}{\lambda^2+1}\right)+\varepsilon^2\left(\frac{\lambda^2}{\lambda^2+1}\right)^2-\varepsilon^2\left(\frac{\lambda^2}{\lambda^2+1}\right)^2+\varepsilon^2\left(\frac{\lambda^2}{\lambda^2+1}\right)\right]} du \\ &= \left(\frac{\lambda^2+1}{\pi\lambda\sigma^2}\right) e^{-\frac{\varepsilon}{2\sigma^2}} \int_0^\infty e^{-\frac{1}{2\sigma^2}\left(\frac{\lambda^2+1}{\lambda}\right)^2\left[u+\varepsilon\left(\frac{\lambda^2}{\lambda^2+1}\right)\right]^2} du \end{aligned}$$

Seja $z = \left(\frac{\lambda^2+1}{\lambda\sigma}\right) \left[u + \varepsilon\left(\frac{\lambda^2}{\lambda^2+1}\right)\right]$, temos:

$$\begin{aligned} f(\varepsilon) &= \left(\frac{\lambda^2+1}{\lambda\sigma}\right) \left(\frac{\sqrt{2\pi}}{\pi\sigma}\right) e^{-\frac{\varepsilon}{2\sigma^2}} \int_{\frac{\varepsilon\lambda}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\left(\frac{\lambda^2+1}{\lambda}\right)^2\left[z\left(\frac{\lambda\sigma}{\lambda^2+1}\right)\right]^2} \left(\frac{\lambda\sigma}{\lambda^2+1}\right) dz \\ &= \left(\sqrt{\frac{2}{\pi\sigma^2}}\right) e^{-\frac{\varepsilon}{2\sigma^2}} \underbrace{\int_{\frac{\varepsilon\lambda}{\sigma}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz}_{1-\Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)} \end{aligned}$$

Desta forma, temos a seguinte função de probabilidade para ε :

$$f_\varepsilon(\varepsilon_i) = \left(\sqrt{\frac{2}{\pi\sigma^2}}\right) \Phi\left(-\frac{\varepsilon_i\lambda}{\sigma}\right) e^{-\frac{\varepsilon_i}{2\sigma^2}} \quad (2.19)$$

Para o caso do custo onde $v = \varepsilon - u$, temos:

$$f_\varepsilon(\varepsilon_i) = \left(\sqrt{\frac{2}{\pi\sigma^2}} \right) \Phi \left(\frac{\varepsilon_i\lambda}{\sigma} \right) e^{-\frac{\varepsilon_i}{2\sigma^2}} \quad (2.20)$$

onde $\Phi(\cdot)$ representa a distribuição acumulada de uma normal padrão. Segue de imediato a função de log-verossimilhança para o modelo da Equação 2.16:

$$\ln L(\mathbf{y}; \beta, \sigma, \lambda) = -\frac{N}{2} \ln \left(\frac{\pi\sigma^2}{2} \right) + \sum_{i=1}^N \ln \Phi \left(-\frac{\varepsilon_i\lambda}{\sigma} \right) - \frac{1}{2\sigma^2} \sum_{i=1}^N \varepsilon_i^2 \quad (2.21)$$

onde $\varepsilon_i = y_i - \mathbf{x}_i\beta$. Destaca-se que, caso $\lambda = 0$, não haverá efeito de ineficiência técnica.

Para estimar os parâmetros do modelo, será utilizada a técnica de maximizar a função de log-verossimilhança, ou seja, derivando para cada parâmetro e igualando a 0. No entanto, tendo em vista que a Equação 2.21 envolve termos altamente não lineares, é extremamente complexo estabelecer estimadores analíticos para β , σ e λ . Assim, os estimadores serão obtidos através de métodos iterativos de otimização, ou seja, tanto o gradiente quanto a matriz hessiana serão obtidas por meio de derivações numéricas.

Na Figura 2.4 temos a comparação entre a fronteira estocástica de produção baseada modelo *half-normal* com o modelo de regressão comum e o DEA. A fronteira

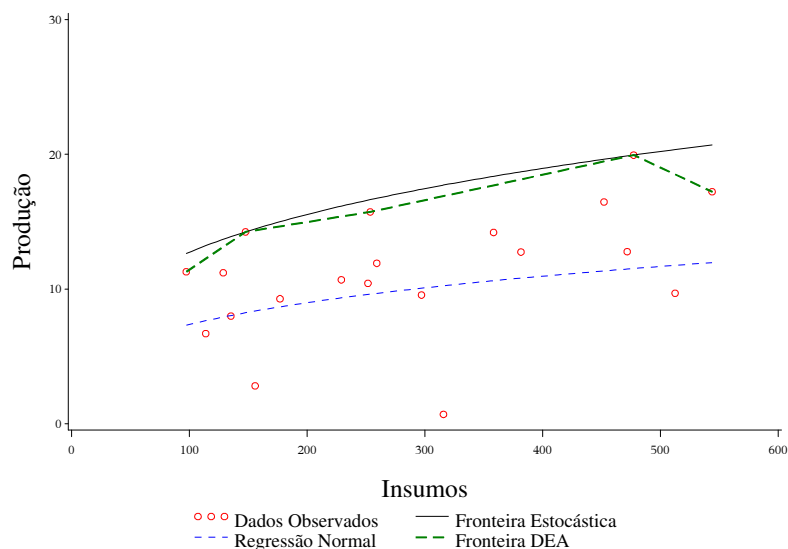


Figura 2.4: Comparação entre modelagens

foi calculada com base em um dado setor econômico supondo retornos decrescentes de escala para somente um insumo x .

2.5.2 Outros modelos comuns

Não é incomum trocar a suposição de *half*-normalidade em (2.16) por algumas das seguintes:

$$u_i \sim iidN^+(\mu, \sigma_u^2) \quad (Normal\ truncada) \quad (2.22)$$

$$u_i \sim iidG(\lambda, 0) \quad (Exponencial\ com\ média\ \lambda) \quad (2.23)$$

$$u_i \sim iidG(\lambda, m) \quad (Gamma\ com\ média\ \lambda\ e\ m\ graus\ de\ liberdade) \quad (2.24)$$

Da mesma forma que no modelo *half-normal*, esses modelos só podem ser estimados por meio de rotinas iterativas de otimização. A função de verossimilhança desses modelos pode ser encontrada em Kumbhakar and Lovell (2000).

A escolha da distribuição pode ser feita por meio de conveniências computacionais, ou por considerações teóricas. No primeira caso, temos que algumas di-

tribuições exigem menos tempo para serem calculadas do que outras. Já para as considerações teóricas, temos que alguns pesquisadores evitam modelos *half-normal* ou *exponencial*, tendo em vista que ambos tem a moda no zero. Isso implica que a maior parte das ineficiências estão na vizinhança de zero, ou seja, a medida de eficiência das firmas está sempre próxima de um. As distribuições *Normal truncada* e *Gamma*, por exemplo, possuem uma flexibilidade maior de modelagem, em contrapartida são computacionalmente desvantajosas. A escolha da distribuição pode ser crucial para a predição das eficiências em termos absolutos, no entanto os postos das firmas são, em geral, robustos a escolha da distribuição (Coelli et al., 2005).

2.5.3 Estimação de Modelos Não-Lineares e Testes de Hipóteses

Da mesma forma que em modelos lineares, podemos considerar derivar a função soma de quadrados (ou verossimilhança) produção com respeito a β , igualar a zero e teríamos a uma solução para condições de primeira ordem. Tanto para soma de quadrados como para verossimilhança temos o estimador: *Mínimos Quadrados Não-Lineares (NLS)*. Infelizmente, não é sempre que conseguimos formas analíticas para os parâmetros do modelo, é comum encontrarmos funções altamente não lineares. Nessas situações a solução só pode ser obtida por métodos numéricos iterativos (Coelli et al., 2005). Detalhes dessas otimizações podem ser obtidas em Souza (1998), Grenne (2003) e Judge et al. (1985).

Encontrando o máximo global da função de verossimilhança tem-se consequentemente os parâmetros estimados. No entanto, falta ainda verificar se os parâmetros estimados não são significativos, ou seja, se não são diferentes de 0. Para verifi-

car se há evidências para rejeitar essa hipótese, pode-se usar o teste de razão de verossimilhança, teste de Wald e Multiplicadores de Lagrange.

Em especial, no estudo sobre eficiência buscar-se-á saber se $\sigma_u^2 \neq 0$ (Coelli et al., 2005), pois caso σ_u^2 não seja significativo há indicação de não haver ineficiências. No modelo *Half-Normal* há interesse no seguinte teste:

$$\begin{aligned} H_0 : \lambda &= 0 \\ H_1 : \lambda &\neq 0 \end{aligned} \tag{2.25}$$

Infelizmente, tendo em vista que o erro dos modelos de fronteira são assimétricos, não é possível aplicar o teste t nem o teste F . Em contrapartida, pelo seguinte teorema de estimadores de máxima verossimilhança temos uma solução:

2.5.1. Teorema. *O estimador de máxima verossimilhança $\hat{\theta}_n$ é tal que $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converge em distribuição para a $N_p(\tilde{0}, I(\theta_0)^{-1})$.*

onde $I(\theta_0)^{-1}$ é o inverso da matriz de informação de Fisher para θ_0 . Mais detalhes em Souza (1998). Portanto, como $I(\theta_0) = -H(\theta_0)$, temos que a matriz Hessiana converge assintoticamente para a matriz de variância e covariância do estimador, $-H(\theta_0) \Rightarrow Var(\hat{\theta}_n)$. Além disso, pelo Teorema 2.5.1 temos que estimadores dados pela Máxima Verossimilhança convergem para uma distribuição normal, garantindo a possibilidade de aplicar o teste z para testar hipóteses sobre os parâmetros. Por exemplo, sobre λ , temos que a estatística do teste é dada por:

$$z = \frac{\hat{\lambda}}{DP(\hat{\lambda})} \sim N(0, 1) \tag{2.26}$$

onde $DP(\hat{\lambda})$ é o desvio padrão do estimador $\hat{\lambda}$. Coelli (1995) descreve usando o método de Monte Carlo, no entanto, uma baixa eficiência do teste z em amostras

pequenas, sendo esse o motivo de alguns pesquisadores utilizarem testes alternativos como o Wald e Razão de Verossimilhança.

2.5.4 Predizendo a eficiência técnica

Como foi definido em (2.15), a eficiência da i -ésima firma é dada em função de um termo u_i , em que a informação desse termo estocástico assimétrico dado valores observados de produção, q_i , pode ser resumido da seguinte forma, considerando *half*-normalidade:

$$p(u_i|q_i) = \frac{1}{\sqrt{2\pi\sigma_*^2}} \exp\left\{-\frac{1}{2\sigma_*^2}(u_i - u_i^*)^2\right\} / \Phi\left(\frac{u_i^*}{\sigma_*}\right) \quad (2.27)$$

onde $u_i^* = -(\ln q_i - \mathbf{x}\beta)\sigma_u^2/\sigma^2$ e $\sigma_*^2 = \sigma_v^2\sigma_u^2/\sigma^2$. Entre outras coisas, Jondrow et al. (1982) utilizou a função de densidade de probabilidades para derivar o seguinte preditor de u_i :

$$\hat{u}_i \equiv E\{u_i|q_i\} = u_i^* + \sigma_* \left[\frac{\phi(u_i^*/\sigma_*)}{\Phi(u_i^*/\sigma_*)} \right] \quad (2.28)$$

E dessa forma, Horrace and Schmidt (1995) construíram os seguintes intervalos de confiança para a predição de u_i :

$$L_i = u_i^* + \sigma_* \Phi^{-1}\{(1 - \alpha/2)\Phi(u_i^*/\sigma_*)\} \quad (2.29)$$

e

$$U_i = u_i^* + \sigma_* \Phi^{-1}\{(\alpha/2)\Phi(u_i^*/\sigma_*)\} \quad (2.30)$$

Alternativamente, Battese and Coelli (1988) usaram (2.27) para derivar o seguinte preditor de u_i :

$$\widehat{TE}_i \equiv E\{\exp(-u_i)|q_i\} = \left[\Phi\left(\frac{u_i^*}{\sigma_*} - \sigma_*\right) / \Phi\left(\frac{u_i^*}{\sigma_*}\right) \right] \exp\left\{\frac{\sigma_*^2}{2} - u_i^*\right\} \quad (2.31)$$

Os intervalos para u_i são dados por (2.29) e (2.30) e assim podemos prever a eficiência técnica específica de cada firma. Para estimar a eficiência média da indústria estudada, um estimador consistente é a média amostral.

A Figura 2.5 compara o comportamento Fronteira Estocástica com o método determinístico DEA. A comparação consiste em ir retirando firmas e verificando as eficiências encontradas. Os dados utilizados foram simulados considerando que 1 insumo produz um produto.

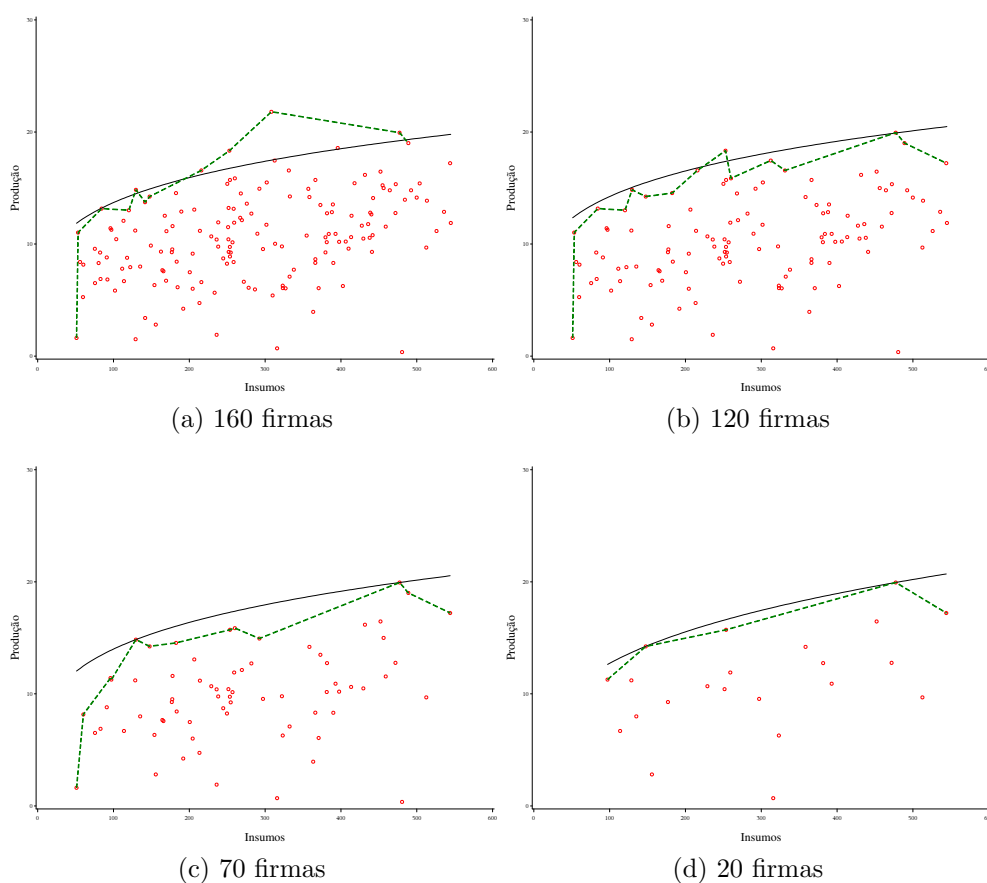


Figura 2.5: Fronteiras: DEA vs Estocástica

Como já foi discutido anteriormente a determinação da eficiência por meio do DEA é relativo as firmas que estão sendo comparadas. Por outro lado, a Fronteira Estocástica possui um comportamento mais suave e foi visivelmente mais robusta,

ou seja os dados foram sendo retirados e curva manteve-se praticamente inalterada.

2.6 MODELO DE DADOS EM PAINEL

A predição de eficiência discutida até esse ponto foi voltada para dados cortados no tempo (*cross – sectional*). Esse tópico tem como objetivo estender o estudo de fronteiras para dados em painel, nesse tipo de análise considera-se que as firmas possuem dados ao longo do tempo. A motivação para fazer tal generalização decorre do fato que dados em painel são mais informativo por possuírem muito mais dados e assim é razoável considerar que: os preditores serão mais eficientes; algumas suposições relacionadas as distribuições poderão ser relaxadas; e passa a ser possível fazer inferência sobre a mudança de eficiência da mesma firma ao longo do tempo (Coelli et al., 2005).

Considere o modelo de fronteira estocástica de produção para os dados em painel.

$$y_{it} = f(x_{it}; \beta) e^{V_{it} - U_{it}} \quad (2.32)$$

onde:

- y_{it} denota a variável de produção no t -ésimo tempo para a i -ésima firma;
- x_{it} é um vetor $(1 \times k)$ de variáveis de insumos;
- β é o vetor $(k \times 1)$ de parâmetros não conhecidos a serem estimados;
- V_{it} é o termo do erro aleatório assumido ser iid $N(0, \sigma_v^2)$ e independente de U_{it} ;
- U_{it} é o termo assimétrico que deverá controlar o efeito da heterogeneidade individual não observada.

Assim como no caso de dados em corte no tempo Coelli et al. (2005), o termo U_{it} assume o papel de definir os diferentes tipo de modelagem de fronteira para dados em painel. Considera-se, então, dois tipo de predição de eficiência: invariante e variante no tempo.

2.6.1 Modelos de Ineficiência Invariantes no tempo

É a modelagem mais simples para fronteira estocástica de dados em painel. O tempo assimétrico é tratado da seguinte forma:

$$U_{it} = U_i \quad (2.33)$$

onde U_i pode ser um parâmetro fixo ou uma variável aleatória. Em modelos invariantes no tempo a eficiência é predita para cada firma e fixa no tempo (Battese and Coelli (1988) e Pitt and Lee (1981)), não permitindo a análise da mudança de eficiência da firma ao longo tempo. Caso seja necessário esse tipo de análise deve-se utilizar modelos variantes no tempo.

2.6.2 Modelos de Ineficiência Variantes no tempo

Modelos de ineficiência que consideram a variação no tempo são utilizados em situações que se considera que os gestores das firmas aprendem com resultados anteriores, buscando melhorar a eficiência da firma no ano corrente. Esse modelos seguem a seguinte estrutura:

$$U_{it} = f(t).u_i \quad (2.34)$$

onde $f(.)$ é uma função que determina como a ineficiência técnica varia ao longo do tempo. Os dois modelos mais utilizados são:

Proposto por Kumbhakar (1990)

$$f(t) = [1 + \exp(\alpha t + \beta t^2)]^{-1} \quad (2.35)$$

e o modelo proposto por Battese and Coelli (1993)

$$f(t) = \exp[\eta(t - T)], \quad t \in \varphi(i); \quad i = 1, \dots, N \quad (2.36)$$

em ambos modelos foi considerado que $U_i \sim iidN^+(\mu, \sigma_u^2)$. Além disso, as definições anteriores sobre estimação dos parâmetros, teste de hipótese e predição de eficiência são válidas para a extensão em dados em painel. No algoritmo que será apresentado no Capítulo 3 foi implementado a função dada pela Equação 2.36. Todas as derivações estão presentes no Apêndice B. Sobre o teste de razão de verossimilhança, há um em especial que busca verificar evidências de devemos usar o modelo to tipo invariante no tempo ou variante, o teste é dado por: $H_0 : \eta = 0$. Caso aceite a hipótese nula o modelo sugere que há dificuldade de perceber mudanças ao longo do tempo.

A predição da eficiência no modelo da Equação 2.36 é dada pela Equação 2.38.

$$E(U_i | \varepsilon_i = e_i) = \mu_i^* + \sigma_i^* \left[\frac{\phi\left(\frac{-\mu_i^*}{\sigma_i^*}\right)}{1 - \Phi\left(\frac{-\mu_i^*}{\sigma_i^*}\right)} \right] \quad (2.37)$$

onde μ_i^* e σ_i^* estão definidos nas Equações 5 e 6 dos apêndices. Sendo assim, para predizer a eficiência utilizaremos $\hat{U}_i = E(U_i | \varepsilon_i = e_i)$ e a eficiência técnica é dada por:

$$TE_{it} = 1 - e^{-\eta t U_i} \quad (2.38)$$

Há modelos que consideram u_i como sendo um parâmetro fixo a ser estimado aproximando-se da predição por DEA. Esses modelos são denominados efeitos fixos.

Infelizmente, assim como no DEA, esses modelos não são plausíveis se o número de firmas for pequeno. Um algoritmo muito bem implementado dessa metodologia pode ser encontrado em Wang and Ho (2010).

Capítulo 3

ALGORITMO DESENVOLVIDO

3.1 DADOS EM CORTE TRANSVERSAL

Seguindo as sugestões apresentados no capítulo anterior, foi desenvolvido no SAS/IML um algoritmo capaz de estimar, em um corte no tempo, a fronteira estocástica e prever a eficiência técnica para cada firma de um determinado setor econômico. O algoritmo é chamado da seguinte forma:

```
%SFA(  
  base=,  
  y=,  
  x=,  
  intercept =s,  
  LR=n,  
  type=1,  
  maxiter=3000,  
  maxfun=6000,  
  nlp=NLPQN,  
  Predict=n,  
  IC=n,  
  alpha=.05  
);
```

onde cada objeto recebe:

1. `base` - a base de dados onde estão as variáveis;
2. `y` - a variável dependente;

3. `x` - as variáveis explicativas;
4. `intercept` - a variável booleana (n ou s) para ter ou não intercepto;
5. `LR` - a variável booleana (n ou s) para fazer o teste de razão de verossimilhança;
6. `type` - a variável booleana (0 ou 1) para maximizar ou não a função de produção, caso 0, será minimizado a função custo;
7. `maxiter` - número máximo de iterações para calcular a máxima verossimilhança;
8. `maxfun` - número de máximas funções calculadas durante a otimização;
9. `nlp` - o nome da otimização a ser usada para maximizar a função de verossimilhança, as opções estão presentes na Tabela 3.1;
10. `Predict` - a variável booleana (n ou s) para prever ou não a eficiência;
11. `IC` - a variável booleana (n ou s) para construir ou não intervalo de confiança para cada predição de eficiência; e
12. `alpha` - o nível de significância a ser usado para construir o intervalo de confiança da predição de eficiência.

A função de log-verossimilhança foi implementada da seguinte forma:

```
s=&type;
start HalfNormal(parm) GLOBAL(s);
  beta=parm[1:p];
  _sigma2=parm[p+1];
  lambda=parm[p+2];
```

```

phi=CDF('Normal', (-1)**s*(y-x*beta)*lambda/sqrt(_sigma2));
do i=1 to nrow(phi);
if phi[i]=0 then phi[i]=10e-16;
end;
L =-(N/2)*Log(pi*_sigma2/2)
+ (Log(phi))'*j(N,1,1)-(1/(2*_sigma2))*(y-x*beta)'*(y-x*beta);
Return(L);
finish HalfNormal;

```

Vale ressaltar que foram feitas restrições sobre as variâncias não permitindo que essas assumam valores negativos. A maximização da função de verossimilhança pode ser feita por meio das seguintes otimizações já presentes no IML/SAS:

Tabela 3.1: Rotinas de maximização que podem ser usadas

Nome da Rotina	Como deve ser chamada
Conjugate-Gradient	NLPCG
Quasi-Newton	NLPQN
Newton-Raphson	NLPNRA
Double Dogleg	NLPDD
Nelder-Mead Simplex	NLPNMS
Newton-Raphson Ridge	NLPNWRR
Quadratic	NLPQUA
Trust-Region	NLPTR

Quanto a predição da eficiência individual foi usado o preditor proposto por Battese and Coelli (1988) e implementado da seguinte forma:

```

TE=(CDF('Normal', (_u_E/_sigma_E) - _sigma_E)/ CDF('Normal', (_u_E/_sigma_E)))
#exp((_sigma_E**2/2)-_u_E);

```

Essa metodologia para dados em corte transversal já está bem implementada no SAS pelo procedimento QLIM e no pacote frontier do *software* STATA. Na Tabela 3.2 temos a comparação das estimativas obtidas pelos diferentes algoritmos em uma base de dados simulada de apenas um insumo e um produto.

Tabela 3.2: Comparação das estimativas entre os algoritmos de estimação *Cross – Sectional*

Parâmetro	PROC QLIM	macro SFA	STATA frontier
intercepto	3.9270 (11,362)	3,9293*** (1,1588)	3,932525*** (0.7475)
ln(x)	0,9533*** (0,152)	0,9533*** (0,1519)	0,9533021*** (0,1519)
σ	0.5698	0,5699*** (0,0354)	0,5699*** (0,0354)
λ	0.0000	0.0048 (2,0441)	0.0121 (0,3572)
σ_u	0.0000 (14,21)	0.0027	0.0069
σ_v	0,56983*** (0,0309)	0.5698	0.5698
log Verossimilhança	-145.61097	-145.6110	-145.6110
AIC	299.22193	295.2219	295.2220

Vale ressaltar que o procedimento QLIM do SAS estima a fronteira, no entanto não é capaz de prever a eficiência. Por outro lado, a macro SFA e o pacote frontier do STATA permitem essa opção.

3.2 DADOS EM PAINEL PARA UM EFEITO ALEATÓRIO

A motivação para a elaboração do algoritmo para dados em painel vem em grande parte da idéia de que em um cenário econômico real a utilização de dados acumulados no tempo é essencial, pois são mais informativos, possuem menor variabilidade e mais graus de liberdade.

A construção desse algoritmo, chamado aqui de macro SFPA, seguiu os mesmo critérios do algoritmo para dados não correlacionados no tempo, a diferença crucial nesse para aquele é a função de log-verossimilhança, além das funções gradientes que foram derivadas por Battese and Coelli (1992). Foi implementado a utilização do termo assimétrico considerando a densidade normal truncada.

A macro SFPA é chamada praticamente da mesma forma que a macro SFA, no entanto deve-se definir qual é a variável que possui a identificação da unidade observacional e a variável que possui a identificação temporal. Sendo assim temos:

```
%SFPA(
    base=,
    id=,
    time=,
    y=,
    x=,
    ...);
```

onde `id` recebe a variável identificadora da firma e `time` recebe a variável identificadora do tempo.

A estimação do modelo de custo foi feito seguindo uma manipulação dos termos aleatórios. A fronteira de produção tem a forma dada por:

$$y_{it} = x_{it}\beta + v_{it} - u_{it} \quad (3.1)$$

Já a fronteira de custo é dada por:

$$y_{it} = x_{it}\beta + v_{it} + u_{it} \quad (3.2)$$

Para a forma da Equação 3.2, basta multiplicar a Equação 3.1 por -1 . Assim, temos:

$$-y_{it} = -x_{it}\beta - v_{it} + u_{it}$$

Como o termo v_{it} é simétrico temos que $v_{it} = -v_{it}$ e assim construímos a fronteira de custo.

Da mesma forma que foi feito na seção anterior, temos na Tabela 3.3 uma comparação de algoritmos implementados utilizando dados simulados. Nesse caso, como não há implementação feita no SAS foi usado o pacote frontier do R e o pacote xt-frontier do STATA. Além disso, será feito uma estimação do tipo Pooled, em que não considera que os dados estão estruturados conforme um painel, utilizando a macro SFA. Os 4 algoritmos seguindo essa instrução adequarão a Equação 3.3

$$\ln(q_{it}) = \beta_0 + \beta_1 \ln(x_{it}) + v_{it} - \eta_{it} U_i \quad (3.3)$$

Os dados simulados seguem a estrutura de uma base balanceada de 100 firmas, novamente com apenas um insumo, ao longo de 6 anos construída levando em conta os seguintes parâmetros:

- $\beta_0 = \ln(2)$
- $\beta_1 = 0,3$
- $\sigma^2 = 0,2$
- $\sigma_v^2 = 1$
- $\eta = -2$
- $\mu = 0$

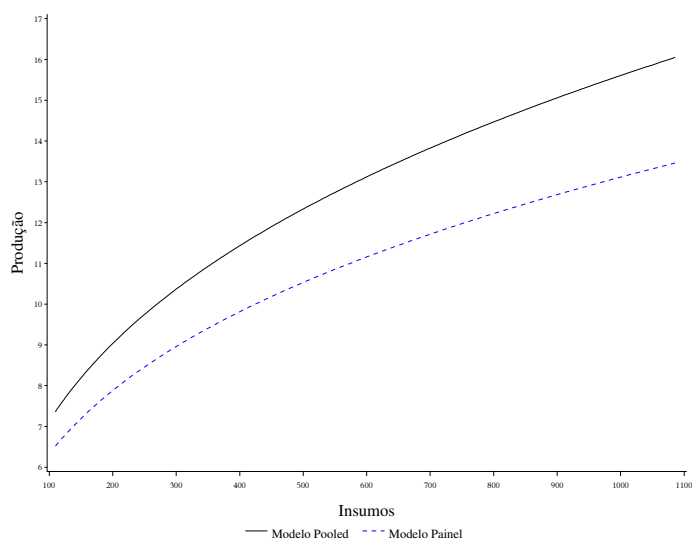
Dessa forma, seguem as estimativas na Tabela 3.3:

Tabela 3.3: Comparação das estimativas entre os algoritmos para dados em painel

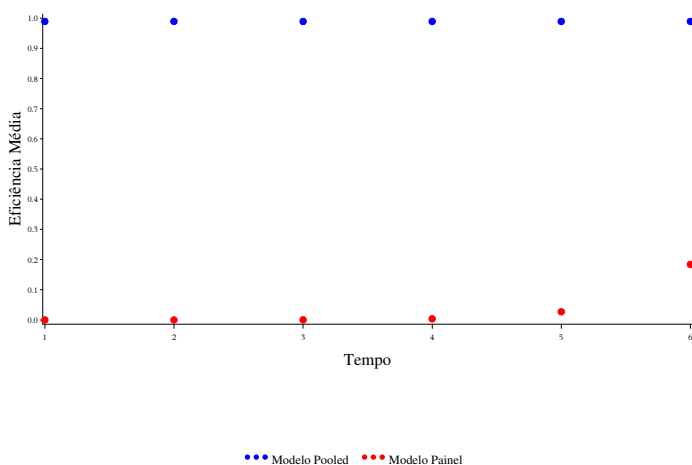
Parâmetro	macro SFA	Painel	macro SFPA	xtfrontier STATA	frontier R
<i>intercepto</i>	0,495 (3,165)	<i>intercepto</i>	0,489 (0,422)	0,6058 (0,4173)	0,5978 (0,4195)
$\ln(x)$	0,330*** (0,066)	$\ln(x)$	0,335*** (0,067)	0,3293*** (0,066)	0,3289*** (0,067)
σ	0,981*** (0,073)	η	-2,170*** (0,001)	-0,7436 (0,570)	-0,8108* (0,4895)
λ	0,009 (3,972)	σ_s^2	1,032*** (0,004)	1,012*** (0,2390)	1,1474*** (0,1264)
		γ	0,001*** (0,000)	0,056*** (0,2372)	0.171* (0,0883)
		μ	0,290*** (0,002)	0,3288 (0,4404)	
log Verossimilhança	-845,56		-842,279	-840,623	-840.6673

Tendo em vista que foi escolhido um $\eta < 0$ e todos os algoritmos tentaram adequar a um modelo de produção, há em quase todos um erro de convergência e um aviso de que os resíduos são assimétricos para o lado do custo, ou seja, clara indicação de que a fronteira foi escolhida equivocadamente. Sobre as estimativas dos βs não há discrepância entre os algoritmos nem no modelo pooled, no entanto os parâmetros de definição do modelo possuem diferentes estimativas e significancias para os algoritmos estruturados para os dados em painel. Caso surja então dúvida do motivo de não usar o pooled tendo em vista que é mais rápido e simples, a resposta é ilustrada na Figura 3.1 em que as eficiências estimadas no modelo pooled praticamente são constantes, já no modelo de painel são bem mais razoáveis.

É importante salientar que a base de dados foi simulada para verificar qual algo-



(a) Fronteira



(b) Eficiência Média por Tempo

Figura 3.1: Comparação entre o modelo Pooled e Painel

ritmo consegue estimar melhor o que foi construído. Considerando que a prioridade é o estudo da eficiência (termos aleatórios) e não da fronteira de produção propriamente dita podemos afirmar que a estimativa do algoritmo da macro SFPA chegou mais perto do determinante de mudança por período (η) que foi construído. Evidentemente que com apenas essa base de dados não podemos afirmar, de fato, qual é o melhor algoritmo. Para o próximo capítulo utilizaremos dados reais de uma base

de dados financeiros dos bancos comerciais do Brasil, para utilização do algoritmo SFPA em uma fronteira de custo.

Capítulo 4

EFICIÊNCIA ESTOCÁSTICA DOS BANCOS BRASILEIROS

4.1 INTRODUÇÃO

O Brasil tem sua história baseada em ciclos econômicos, que ajudaram na evolução do sistema financeiro. Atualmente, controlado por instituições financeiras governamentais e privadas, o setor bancário brasileiro baseia sua gestão na consolidação de políticas monetárias eficientes e no equilíbrio do mercado nacional com o externo. A preocupação com o aumento da confiança dos investidores estrangeiros no Brasil se torna cada vez mais importante para a consolidação e expansão do mercado financeiro. Com a criação de bancos mais sólidos, resultante do processo de fusões e aquisições, o mercado financeiro brasileiro aumentou sua importância no sistema financeiro global, assim como os países em desenvolvimento citados acima (Marra et al., 2011). Dessa forma, estudaremos o comportamento da fronteira de *custo* do sistema bancário brasileiro.

4.2 METODOLOGIA

Para mensurar a eficiência individual de cada banco em cada período de tempo,

consideraremos o modelo dado pela função fronteira translog de custo apresentado pela Equação 4.1.

$$y_{it}^* = \beta_0 + \beta_q \ln q_{it} + \beta_{qq} (\ln q_{it})^2 + \sum_{n=2}^3 \beta_n z_{n_{it}} + \sum_{n \neq m} \beta_{nm} z_{nm_{it}} + \sum_{n=2}^3 \beta_{qn} z_{qn_{it}} + Tipo_{it} + v_{it} + u_{it} \quad (4.1)$$

onde $y_{it}^* \equiv \ln c_{it} - \ln w_{1_{it}}$,

$$z_{n_{it}} \equiv \ln w_{n_{it}} - \ln w_{1_{it}},$$

$$z_{nm_{it}} \equiv \ln w_{n_{it}} \ln w_{m_{it}} - 0,5(\ln w_{n_{it}})^2 - 0,5(\ln w_{m_{it}})^2,$$

$$z_{qn_{it}} \equiv \ln q_{it} z_{n_{it}} \text{ e}$$

$Tipo_{it}$ representa as variáveis dummies para o tipo do banco: público, privado ou estrangeiro.

Temos também que c_{it} representa os custos totais, $w_{n_{it}}$ representa o n -ésimo preço do insumo para o i -ésimo banco na t -ésima unidade de tempo e q_{it} representa um produto do banco, no caso utilizamos Empréstimos. Para os custos totais foi usado a variável Despesa Total. Foram construídos 3 preços: preço de fundos (w_1) dado pela razão entre a despesa da intermediação financeira sobre depósitos totais, preço de capital (w_2) dado pela razão entre a despesa total menos despesa com pessoal sobre ativos fixos e o preço de pessoal (w_3) construído na razão despesa de pessoal sobre ativo total (Panzar and Rosse, 1987).

4.3 BASE DE DADOS

Será utilizado uma base de dados não balanceada dos bancos comerciais brasileiros, individuais e conglomerados, levando em conta os meses de janeiro de 2001 a

junho de 2011. Os dados são de domínio público e disponíveis em BaCen (2011a), já para montar as informações de conglomerados foi feito uma fusão dos bancos controlados pela mesma instituição, essa informação pode ser encontrada em BaCen (2011b).

Tendo em vista que são dados financeiros as contas de resultados são acumuladas semestralmente, já as contas do Balanço Patrimonial são uma espécie de fotografia do Banco no período em questão. Nesse estudo foi montado duas bases de dados e comparados seus respectivos resultados: uma com informações anuais e outra com informações semestrais, 10 e 21 períodos, respectivamente.

4.4 RESULTADOS EMPÍRICOS

4.4.1 Análise Descritiva

Antes de determinar a fronteira de custo dos bancos e suas respectivas eficiências, foi feita uma análise descritiva dos dados financeiros que compõe o modelo. Para as variáveis selecionadas para entrar no modelo não houve valores *missing* encontrados. Pelas Tabelas 4.1 e 4.2, verificamos que não existem problemas nas contas, quanto as suas construções, ou seja, contas devedoras (Despesa Total, Despesa da Intermediação e Despesa de Pessoal) com valores positivos e contas credoras com valores negativos (Ativo, Empréstimos e Depósitos). Além disso, as Figuras 4.1 são gráficos de dispersão da variável y_{it}^* com relação às variáveis preditoras da fronteira de custo.

Tendo em vista que pretende-se ajustar a um modelo de fronteira estocástica, há algumas medidas descritivas que podem ser consideradas antes de utilizar o modelo

Tabela 4.1: Análise descritiva dos dados - base anual

Variável	Média	Desvio Padrão	Mínimo	Máximo
Ativo Circulante	1,7146E10	6,78403E10	16970624,1	8,46364E11
Ativo Permanete	2370977318	1,66023E10	8.136,250	2,87987E11
Empréstimos	4887807188	1,9748E10	24.704,720	2,72081E11
Depósitos	8587295537	3,70123E10	1.499,480	4,6343E11
Despesa Total	-3,52185E9	1,22331E10	-1,8569E11	-1641803,0
Despesa da Intermediação	-2,50119E9	9250406556	-1,6177E11	-320,400
Despesa de Pessoal	-358617961	1223802988	-1,1173E10	-363116,66
y_{it}^*	20,513	2,659	7,505	27,416
$\ln q$	19,642	2,515	10,115	26,329
$(\ln q)^2$	392,113	96,984	102,308	693,235
z_2	1,242	2,561	-8,465	11,209
z_3	-2,649	2,465	-14,723	5,043
z_{12}	-4,047	6,942	-62,823	-0,000
z_{13}	-6,545	10,714	-108,388	-0,000
z_{23}	-9,218	9,848	-104,508	-0,004
z_{q2}	24,288	47,544	-149,677	182,118
z_{q3}	-52,061	45,858	-260,342	73,078

Tabela 4.2: Análise descritiva dos dados - base semestral

Conta	Média	Desvio Padrão	Mínimo	Máximo
Ativo Circulante	1,73792E10	6,94488E10	12992146,3	8,81957E11
Ativo Permanete	2345326333	1,64466E10	6.398,600	2,87987E11
Empréstimos	4944285874	2,03196E10	13.937,570	2,90581E11
Depósitos	8551399396	3,72519E10	189,330	4,80082E11
Despesa Total	-3,27667E9	1,09048E10	-1,8569E11	-1242880,8
Despesa da Intermediação	-2,28351E9	7905649045	-1,6177E11	-268,390
Despesa de Pessoal	-347550308	1197597603	-1,1173E10	-334865,74
y_{it}^*	20,487	2,690	5,412	27,416
$\ln q$	19,643	2,500	9,542	26,395
$(\ln q)^2$	392,080	96,720	91,056	696,704
z_2	1,182	2,607	-10,758	13,939
z_3	-2,664	2,482	-15,468	5,043
z_{12}	-4,094	7,137	-97,141	-0,000
z_{13}	-6,627	11,098	-119,625	-0,000
z_{23}	-9,046	9,510	-104,508	-0,001
z_{q2}	23,134	48,273	-194,983	186,245
z_{q3}	-52,419	46,329	-286,678	73,078

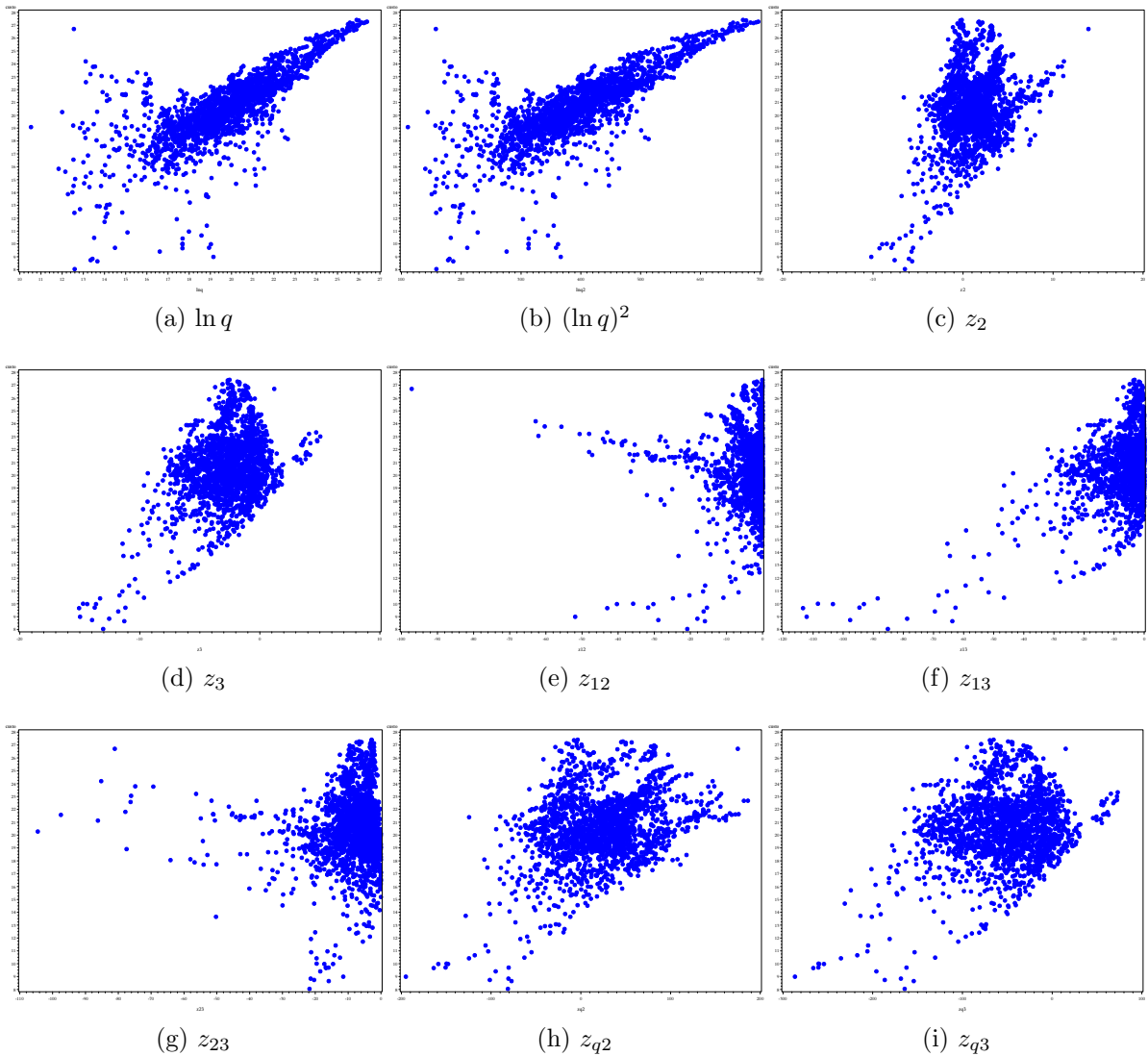


Figura 4.1: Gráficos de Dispersão

pretendido. Por exemplo, verificar se os erros são assimétricos à direita, já que será estimado uma fronteira de custo. Sabe-se, porém, que tal análise é dispensável tendo em vista que já está automatizado no algoritmo. Outra análise descritiva seria determinar as eficiências pelo DEA e assim teríamos uma idéia de como deve se comportar as eficiências estocásticas. Pela Tabela 4.3 temos as médias das eficiências técnicas determinadas para tipo de controle bancário.

Tabela 4.3: Média das Eficiências Utilizando o DEA por controle

Base de dados	Público	Privado	Estrangeiro
Anual	0,5536	0,5105	0,4672
Semestral	0,5621	0,4995	0,4481

4.4.2 Construção da Fronteira Estocástica de Custo

Esse tópico dedica-se ao ajustamento ao modelo, sendo que nessa fase utilizaremos o algoritmo proposto para, além de predizer as eficiências dos bancos, estruturar a fronteira estocástica. Utilizando a Equação 4.1 foi estimado os parâmetros da fronteira e os resultados seguem na Tabela 4.4.

Podemos afirmar que as estimativa das bases de dados (anual e semestral) se confirmam mutualmente, como era de se esperar. Além disso, como foi usado a função translog de custo deve-se interpretar os parâmetros como elasticidades, ou seja, como foi visto na Equação 2.8 o efeito do parâmetro é em termos percentuais. Por exemplo, no caso anual se aumetarmos 1% o produto Empréstimos, aumentamos em média 0,564% a fronteira do custo. Pode-se perceber que as estimativas do parâmetro relacionado a z_3 são, geral, não significativas, ou seja, o preço de pessoal é pouco influente na fronteira de custo bancário. A utilização da *dummy* de tipo de controle (público, privado ou estrangeiro) possibilitou fazermos a inferência de que bancos públicos possuem a fronteira de custo deslocada para baixo em relação aos bancos privados e estrangeiros, ou seja, bancos públicos possuem em média custo mais baixos, indicando que para serem eficientes devem produzir mais do que os outros tipos de bancos. Vale ressaltar que não houve diferença significativa entre a fronteira dos tipos de controle bancário privado ou estrangeiro. Nas Figuras 4.2 e

Tabela 4.4: Estimativas dos parâmetros utilizando a base anual e semestral

Parâmetro	Base Anual	Base Semestral
<i>intercepto</i>	16,699*** (1,459)	16,081*** (1,060)
$\ln q$	0,564*** (0,144)	0,529*** (0,108)
$(\ln q)^2$	-0,037*** (0,003)	-0,036*** (0,003)
z_2	-1,168*** (0,078)	-1,182*** (0,112)
z_3	-0,271* (0,118)	-0,080 (0,130)
z_{12}	0,049*** (0,009)	0,054*** (0,006)
z_{13}	-0,041*** (0,008)	-0,062*** (0,006)
z_{23}	0,009 (0,005)	-0,007 (0,009)
z_{q2}	0,066*** (0,005)	0,064*** (0,006)
z_{q3}	0,002 (0,007)	-0,004*** (0,007)
<i>dummy público</i>	-1,132*** (0,046)	-1,063*** (0,274)
<i>dummy privado</i>	0,024 (0,031)	0,085 (0,061)
η	-0,00016*** (0,000)	-0,006*** (0,000)
σ_s^2	1,656*** (0,032)	2,242*** (0,007)
γ	0,525*** (0,005)	0,738*** (0,001)
μ	0,000 (0,109)	0,000 (0,141)
Número de Bancos	107	136
Número de "tempos"	10	21
Eficiência Média	0,5172	0,7022
log Verossimilhança	-1255,68	-2402,12
AIC	2535,36	4828,24
BIC	2593,71	4896,4

4.3 temos o gráficos das eficiências médias para cada tipo controle de banco ao longo do tempo.

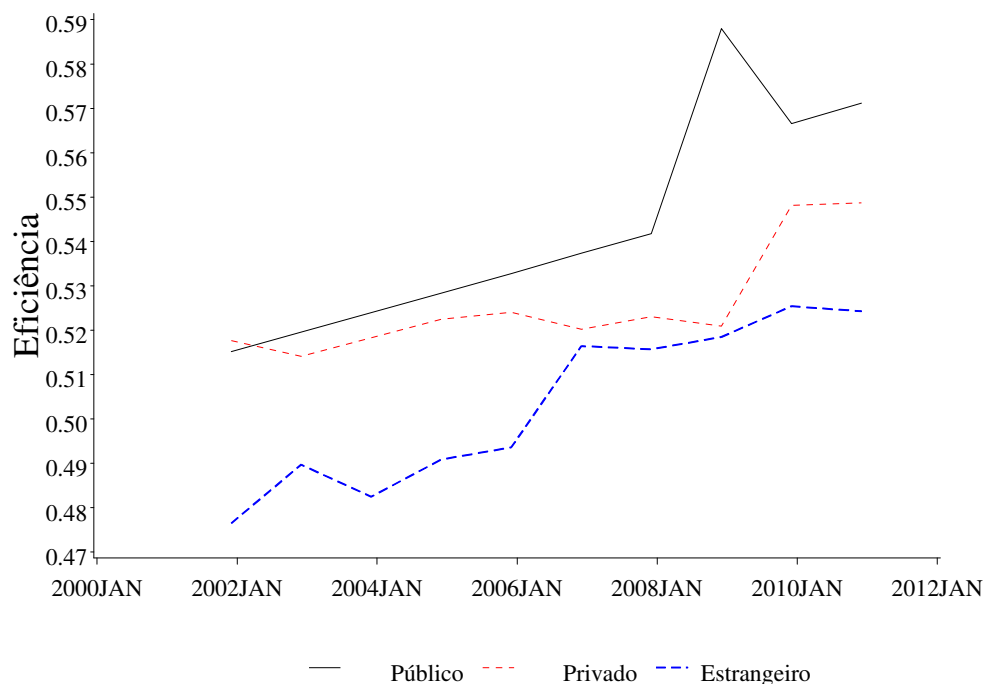


Figura 4.2: Eficiências médias por ano para cada tipo de controle bancário

Pelas Figuras 4.2 e 4.3 percebe-se que os bancos públicos possuem um crescimento quase linear, com um aumento repentino nos anos de 2008 e 2009. Já os bancos privados e estrangeiros possuem curvas um pouco mais variáveis, mas também é visível um aumento repentino, nesses o aumento acontece um pouco mais defasado, nos anos de 2009 e 2010. Sabe-se que o mundo passou por um momento de crise financeira iniciando em meados de 2008 a 2009, sendo assim há forte indicação que tal momento turbulento teve efeito na eficiência dos bancos comerciais brasileiros. Para tentar verificar tal efeito, foi feito considerou-se uma variável *dummy* para crise, ou seja, valor 1 se a observação é do ano é de 2008 a 2009, 0 caso contrário. E constatou-se que em ambas bases de dados, anual e semestral, o efeito da crise foi de

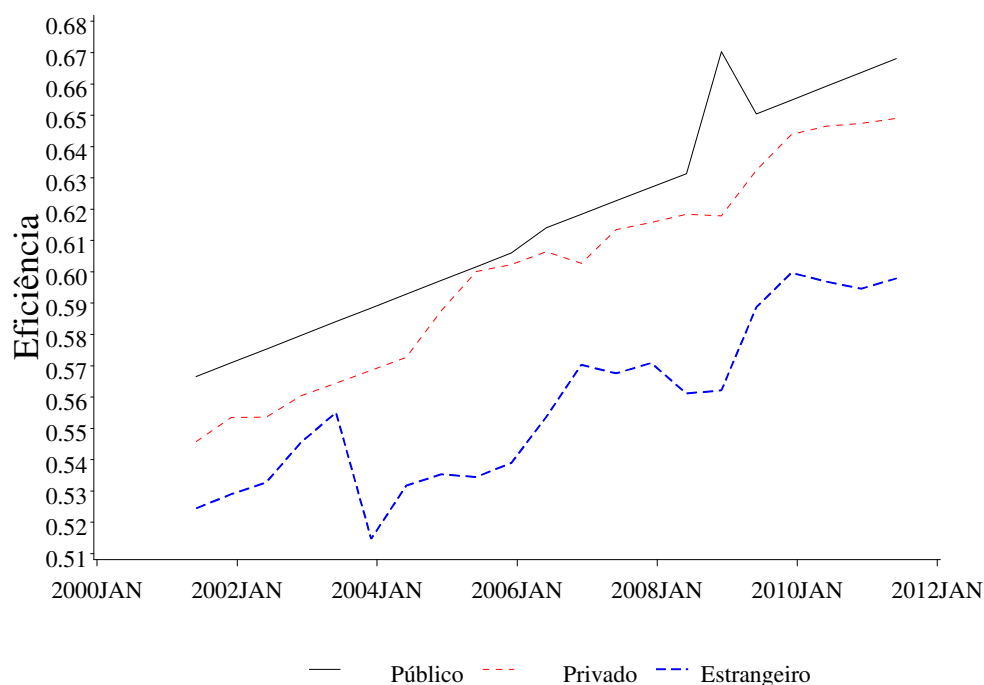


Figura 4.3: Eficiências médias por semestre para cada tipo de controle bancário

diminuir a fronteira de custo em 0,2166% na base anual e 0,254% na base semestral. O motivo da diminuição confirma uma políticas econômicas adotadas pelo Banco Central em outubro de 2008 que previa uma regra para forçar os bancos a liberarem o crédito obtido com o alívio no compulsório, além é claro da redução do IOF (imposto sobre Operações Financeiras) para 0. É razoável pensar que tais medidas deveriam ser percebidas com aumento de eficiência ao longo dos próximos meses e, de fato, foi assim que ocorreu para os bancos privados e estrangeiros. O súbito aumento da eficiência dos bancos públicos ainda em 2008, pode ser reflexo de medidas político/econômica autorizando o Banco do Brasil e a Caixa Econômica Federal a comprarem ações e participações de instituições financeiras nacionais. Além de que o Banco do Brasil anunciou em 1 de outubro de 2008 que liberaria R\$ 5 bilhões para financiamento da safra agrícola, aumentando consideravelmente seus produtos e por

consequência sua eficiência.

4.4.3 Comparação com o DEA

Em comparação com a Tabela 4.3, temos na Tabela 4.5 a predição média das eficiência por controle bancário.

Tabela 4.5: Média das Eficiências Utilizando o Fronteira Estocástica por controle

Base de dados	Público	Privado	Estrangeiro
Anual	0,5063	0,495	0,4781
Semestral	0,6119	0,6005	0,559

Para mostrar o efeito do determinismo do DEA e assim eventuais problemas de determinação da eficiência, escolheu-se na base de dados semestral um banco de características mediana, ou seja, ativos medianos, e determinou sua eficiência por DEA com e sem o insumo pessoal e foi predito sua eficiência por meio da Fronteira Estocástica e comparado o seu comportamento ao longo dos anos.

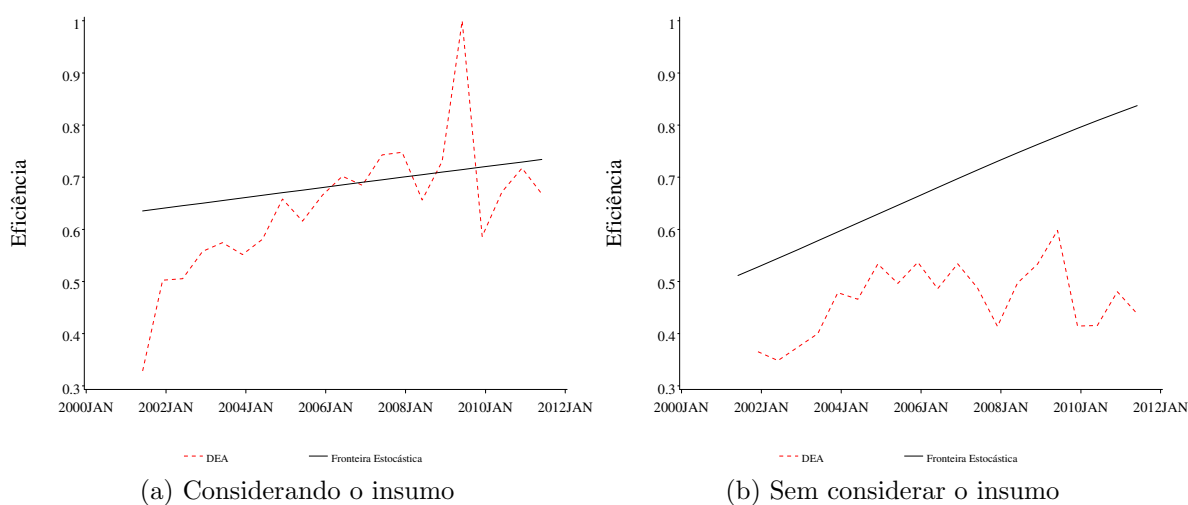


Figura 4.4: Gráficos das Eficiências

A Figura 4.4 ilustra bem como é sensível a metodologia DEA, tendo em vista

que todo desvio da fronteira é considerado ineficiência para o método (Atkinson et al., 2003), sendo assim não há tratamento para eventuais erros aleatórios. Por outro lado, o DEA se aproximou bastante das médias encontrados pelo método estocástico, porém a determinação foi feita de forma a utilizar os mesmo produtos e insumos e, tendo em vista, que os dados são de quase toda população bancos. Isso significa que o erro aleatório se tornou insignificante, e assim as eficiência médias foram bem próximas, no caso anual. Já para o caso semestral onde a correlação dentro do ano é bem grande, o tratamento dado em painel considerou tal efeito, e assim a diferença para o DEA foi bem maior.

Além disso, com o avanço computacional é natural que a utilização de Fronteira Estocástica, para mensurar eficiência seja mais utilizada deixando o DEA como forma descritiva. Tendo em vista que o método estocástico possibilita inferência sobre os parâmetros e é certamente mais robusto aos diversos tipos de dados.

Capítulo 5

CONCLUSÕES

O modelo de predição de eficiência estocástica é complexo e deve ser utilizado com alguns cuidados, como, por exemplo, a escolha correta dos valores iniciais. Certamente, esse é um dos motivos que eventualmente essa metodologia é preterida ao DEA. No entanto, é certo que o método determinístico de Envelopamento de dados, além de ser extremamente relativo as observações presentes e aos produtos e insumos que foram utilizados, é pouco inferencial e não admite o tratamento de dados correlacionados no tempo. No estudo de caso, por exemplo, seria custoso e pouco significativo as afirmações de que nos anos de crise o custo diminuiu para os bancos e os bancos públicos aumentaram seus produtos. Provavelmente, não teríamos tido tal inferência.

Pequenas e grande firmas buscam sempre a eficiência máxima e o DEA é sem dúvida uma ótima ferramenta descritiva, mas dificulta conhecer as causas e prever futuras eficiências, concluí-se então que a utilização da fronteira estocástica é uma análise mais profunda sobre eficiência técnica.

5.1 Limitações do Trabalho

Uma das limitações principais do trabalho foi a de haver poucos pacotes estatísticos robustos que fossem compatíveis com a metodologia implementada para assim comparar os resultados. O pacote *frontier* e *xtfrontier* são os que mais se aproximam do que foi implementado, no entanto possuem algoritmos de otimização limitados impedindo em alguns casos que se pudesse atingir a convergência e assim comparar o resultado com o que foi encontrado no algoritmo. Em outros ambientes, como o *R*, há o algoritmo de eficiência estocástica, no entanto foram construídos utilizando outros artigos como base.

5.2 Trabalhos Futuros

Foi implementado nesse trabalho duas metodologias de fronteira estocástica: uma para dados em *cross – sectional* e outra para dados em painel para 1 efeito aleatório. Fica indicado que ainda há diversas outras modalidades de dados em painel que podem ser implementados para utilizar o método de fronteira, como, por exemplo, 2 efeitos fixos ou 2 aleatórios.

Referências Bibliográficas

- Aigner, D. J. & Chu, S. F. (1968). On estimating the industry production function. *American Economic Review*, 58:826–839.
- Aigner, D. J., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6:21–37.
- Atkinson, S. E., Cornwell, C., & Honerka, O. (2003). Measuring and decomposing productivity change: Stochastic distance function estimation versus data envelopment analysis. *Journal of Business and Economic Statistics*, 21(2):284–94.
- BaCen (2011a). Balancetes. URL <http://www4.bcb.gov.br/fis/cosif/balancetes.asp>. Acesso em 01 de junho de 2012.
- BaCen (2011b). Informação de conglomerados. URL <http://www4.bcb.gov.br/fis/cosif/principal.asp>. Acesso em 01 de junho de 2012.
- Battese, G. E. & Coelli, T. (1988). Prediction of firm-level technical efficiencies with a generalised frontier production function and panel data. *Journal of Econometrics*, (38):387–399.
- Battese, G. E. & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in india. *Journal of Productivity Analysis*, (56).
- Battese, G. E. & Coelli, T. J. (1993). A stochastic frontier production function incorporating a model for technical inefficiency effects. *Working Papers in Econometrics and Applied Statistics*, (69).
- Battese, G. E. & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20:325–332.

- Battese, G. E. & Corra, G. S. (1977). Estimation of a production frontier model: With application to the pastoral zone of eastern australia. *Australian Journal of Agricultural Economics*, (21):169–179.
- Coelli, T. (1995). Estimators and hypothesis tests for a stochastic frontier function: A monte carlo analysis. *Journal of Productivity Analysis*, 6:247–268.
- Coelli, T. J., Rao, D. S., O'Donnell, C. J., & Battese, G. E. (2005). *An Introduction To Efficiency And Productivity Analysis*, (2th ed.). Springer.
- Fries, S. & Taci, A. (2005). Evolution of bank efficiency in brazil: A dea approach. *Journal of Banking and Finance*, 29:55–81.
- Grenne, W. (2003). *Econometrics Analysis*, (5th ed.). Prentive Hall, NJ.
- Horrace, W. & Schmidt, P. (1995). Multiple comparisons with the best, with applications to the efficiency measurement problem. *Working Paper*.
- Jondrow, J., Lovell, C., Materov, I., & Schmidt, P. (1982). On estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19:233–238.
- Judge, G., Griffiths, W., Hill, R., Lutkepohl, H., & Lee, T.-C. (1985). *Introduction to The Theory and Practice of Econometrics*, (2th ed.). John Wiley, New York.
- Kumbhakar, S. & Lovell, C. (2000). *Stochastic Frontier Analysis*. Cambridge, Cambridge UK.
- Kumbhakar, S. C. (1990). Production frontiers, panel data and time-varying technical inefficiency. *Journal of Econometrics*, 46:201–211.
- Marra, L., Gomes, G. M. R., Guerra, S. M., & Tabak, B. M. (2011). Comparação da eficiência de custo para brics e américa latina. Working Papers Series 252, Central Bank of Brazil, Research Department.
- Meeusen, W. & van den Broeck, J. (1977). Efficiency estimation from cobb-douglas production function with composed error. *International Economic Review*, 18:435–444.
- Panzar, J. & Rosse, J. (1987). Testing for "monopoly" equilibrium. *Journal of Industrial Economics*, 35(4):443–56.
- Pitt, M. & Lee, L.-F. (1981). Measurement and sources of technical inefficiency in the indonesian weaving industry. *Journal of Development Economics*, 9:43–64.

- Souza, G. (1998). *Introdução aos Modelos de Regressão Linear e Não-Linear*.
- Souza, G. & Staub, R. (2007). Two-stage inference using data envelopment analysis efficiency measurements in univariate production models. *Transactions in Operations Research*, 14:245–258.
- Staub, R. B., Souza, G., & Tabak, B. M. (2009). Evolution of bank efficiency in brazil: A dea approach. Working Papers Series 200, Central Bank of Brazil, Research Department.
- Wang, H.-J. & Ho, C.-W. (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics*, 157:286–296.
- Winston, C. (1957). Discussion on mr. farrell's papper. *Journal of the Royal Statistical Society, Series A, General*(120):282–84. Part 3.

APÊNDICES

.1 Códigos para construção das bases de dados

Segue o código usado para a construção da Tabela 3.2:

```
data example;
  do firm=1 to 170;
    x=ranuni(120)*100+50;
    lnx=log(x);
    lnq=lnx+abs(rannor(150));
    output;
  end;
run;
%sfa(base=example,
  y=lnq,
  x=lnx);
proc qlim data=example;
  model lnq=lnx;
  endogenous lnq ~ frontier (type=half production);
run;
```

Segue os códigos usados para a construção da Tabela 3.3.

Pelo SAS:

```
data example2;
  do firm=1 to 100;
    x0=ranuni(120)*1e3+100;
    r=ranuni(10)*5;
    ui=abs(rannor(2))*0.2;
  do t=1 to 6;
```

```

x=x0+rannor(20)*r;
lnx=log(x);
output;
end;
end;
drop x0 r;
run;
data example2; set example2;
eit=exp(rannor(1)-exp(-(-2)*(t-6))*ui);
q=2*(x**.3)*eit;
lnq=log(q);
run;
proc export data=example2 dbms=dta replace
outfile="~/example2.dta";
run;
proc export data=example2 dbms=csv replace
outfile="~/example2.csv";
run;

```

Pelo STATA:

```

use "~/example2.dta"
xtset firm t
xtfrontier lnq lnx, tvd

```

Pelo R:

```

library(frontier)
data<-read.csv("~/example2.csv", header = T, sep = ",")
panel <- plm.data(data,c("firm","t"))
eff <- sfa(lnq ~ lnx, data = panel, ineffDecrease = TRUE, maxit=1000, timeEffect=
summary(eff)

```

.2 Derivações da densidade Normal truncada

As derivações estão presentes com mais detalhes em Battese and Coelli (1992). Considere a função fronteira de produção para dados em painel.

$$Y_{it} = \mathbf{x}_{it} + \varepsilon_{it} \quad (1)$$

onde

$$\varepsilon = V_{it} - \eta_{it}U_i$$

com

$$\eta_{it} = \exp(-\eta(t - T)), \quad t \in \varphi(i); \quad i = 1, \dots, N$$

Assume-se que $V_{it} \sim iidN(0, \sigma_v^2)$ e U_i segue uma distribuição normal $N(\mu, \sigma^2)$ não negativa truncada cuja densidade é dada por:

$$f_{U_i}(u_i) = \frac{\exp\left(-\frac{(u_i - \mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma[1 - \Phi(-\mu/\sigma)]}, \quad u_i \geq 0 \quad (2)$$

onde $\Phi(\cdot)$ representa a função distribuição da normal padrão.

Segue então a função conjunta de U_i e $\varepsilon_i = V_{it} - \eta_{it}U_i$, onde ε_i é o vetor de dimensão $(T_i \times 1)$, sendo T_i o número de observações da i -ésima firma:

$$f_{U_i, \varepsilon_i}(u_i, e_i) = \frac{\exp\left[-\frac{1}{2}\left(\frac{(u_i - \mu)^2}{\sigma^2} + \frac{(e_i + \eta_i u_i)'(e_i + \eta_i u_i)}{\sigma_v^2}\right)\right]}{\sqrt{(2\pi)^{T_i+1}}\sigma\sigma_v^{T_i}[1 + \Phi(-\mu/\sigma)]} \quad (3)$$

Integrando em U_i podemos mostrar que a densidade marginal de ε_i é dada por:

$$f_{\varepsilon_i}(e_i) = \frac{\left[1 - \Phi\left(\frac{-\mu_i^*}{\sigma_i^*}\right)\right] \exp\left[-\frac{1}{2}\left(\frac{e_i' e_i}{\sigma_v^2} + \left(\frac{\mu}{\sigma}\right)^2 - \left(\frac{\mu_i^*}{\sigma_i^*}\right)^2\right)\right]}{\sqrt{(2\pi)^{T_i}}\sigma_v^{T_i-1}\sqrt{(\sigma_v^2 + \eta_i' \eta_i \sigma^2)}[1 - \Phi(-\mu/\sigma)]} \quad (4)$$

onde

$$\mu_i^* \equiv \frac{\mu\sigma_v^2 - \eta_i' e_i \sigma^2}{\sigma_v^2 + \eta_i' \eta_i \sigma^2} \quad (5)$$

e

$$\sigma_i^{*2} \equiv \frac{\sigma^2 \sigma_v^2}{\sigma_v^2 + \eta_i' \eta_i \sigma^2} \quad (6)$$

Usando a reparametrização de Battese and Corra (1977), onde $\sigma_v^2 + \sigma^2 = \sigma_s^2$ e

$\gamma = \sigma^2/\sigma_s^2$, a função log verossimilhança de $\mathbf{y} \equiv (\mathbf{y}'_1, \dots, \mathbf{y}'_n)$ é dada por:

$$\begin{aligned}
L^*(\theta; \mathbf{y}) = & -\frac{1}{2} \left(\sum_{i=1}^N T_i \right) (\ln 2\pi + \ln \sigma_s^2) - \frac{1}{2} \sum_{i=1}^N (T_i - 1) \ln [1 - \gamma] \\
& - \frac{1}{2} \sum_{i=1}^N \ln [1 + (\eta'_i \eta_i - 1)\gamma] - N \ln [1 - \Phi(-z)] - Nz^2 \\
& + \sum_{i=1}^N \ln [1 - \Phi(-z_i^*)] + \frac{1}{2} \sum_{i=1}^N z_i^{*2} - \sum_{i=1}^N \frac{(\mathbf{y}_i - \mathbf{x}_i \beta)'(\mathbf{y}_i - \mathbf{x}_i \beta)}{2(1 - \gamma)\sigma_s^2}
\end{aligned} \tag{7}$$

onde $\theta \equiv (\beta, \sigma_s^2, \gamma, \mu, \eta)'$, $z \equiv \frac{\mu}{\sqrt{\gamma\sigma_s^2}}$ e

$$z_i^* = \frac{\mu(1 - \gamma) - \gamma \eta'_i (\mathbf{y}_i - \mathbf{x}_i \beta)}{\sqrt{\gamma(1 - \gamma)\sigma_s^2 [1 + (\eta'_i \eta_i - 1)\gamma]}}$$

Temos também o vetor gradiente definido pelas seguintes Equações:

$$\frac{\partial L^*}{\partial \beta} = \sum_{i=1}^N \frac{\mathbf{x}'_i (\mathbf{y}_i - \mathbf{x}_i \beta)}{(1 - \gamma)\sigma_s^2} + \sum_{i=1}^N \frac{\left[\frac{\phi(-z_i^*)}{1 - \Phi(-z_i^*)} + z_i^* \right] \gamma \mathbf{x}'_i \eta_i}{\sqrt{\gamma(1 - \gamma)\sigma_s^2 [1 + (\eta'_i \eta_i - 1)\gamma]}} \tag{8}$$

$$\begin{aligned}
\frac{\partial L^*}{\partial \sigma_s^2} = & -\frac{1}{2\sigma_s^2} \left[\sum_{i=1}^N -N \left(\frac{\phi(-z)}{1 - \Phi(-z)} + z \right) z + \sum_{i=1}^N \left(\frac{\phi(-z_i^*)}{1 - \Phi(-z_i^*)} + z_i^* \right) z_i^* \right] \\
& - \frac{1}{2\sigma_s^2} \left[\sum_{i=1}^N \frac{(\mathbf{y}_i - \mathbf{x}_i \beta)'(\mathbf{y}_i - \mathbf{x}_i \beta)}{(1 - \gamma)\sigma_s^2} \right]
\end{aligned} \tag{9}$$

$$\begin{aligned}
\frac{\partial L^*}{\partial \gamma} = & \frac{1}{2(1 - \gamma)} \sum_{i=1}^N (T_i - 1) - \sum_{i=1}^N \frac{(\eta'_i \eta_i - 1)}{2[1 + (\eta'_i \eta_i - 1)\gamma]} + \frac{Nz}{2\gamma} \left(\frac{\phi(-z)}{1 - \Phi(-z)} + z \right) \\
& + \sum_{i=1}^N \left(\frac{\phi(-z_i^*)}{1 - \Phi(-z_i^*)} + z_i^* \right) \frac{\partial z_i^*}{\partial \gamma} - \sum_{i=1}^N \frac{(\mathbf{y}_i - \mathbf{x}_i \beta)'(\mathbf{y}_i - \mathbf{x}_i \beta)}{2(1 - \gamma)^2 \sigma_s^2}
\end{aligned} \tag{10}$$

$$\frac{\partial L^*}{\partial \mu} = -\frac{N}{\sqrt{\gamma\sigma_s^2}} \left(\frac{\phi(-z)}{1 - \Phi(-z)} + z \right) + \sum_{i=1}^N \frac{\left[\frac{\phi(-z_i^*)}{1 - \Phi(-z_i^*)} + z_i^* \right] (1 - \gamma)}{\sqrt{\gamma(1 - \gamma)\sigma_s^2 [1 + (\eta'_i \eta_i - 1)\gamma]}} \tag{11}$$

$$\frac{\partial L^*}{\partial \eta} = \sum_{i=1}^N \left(\frac{\phi(-z_i^*)}{1 - \Phi(-z_i^*)} + z_i^* \right) \frac{\partial z_i^*}{\partial \eta} - \frac{\gamma}{2} \sum_{i=1}^N \frac{\frac{\partial \eta'_i \eta_i}{\partial \eta}}{[1 + (\eta'_i \eta_i - 1)\gamma]} \tag{12}$$

onde

$$\frac{\partial z_i^*}{\partial \gamma} = - \frac{\mu + \eta'_i(\mathbf{y}_i - \mathbf{x}_i\beta)}{\sqrt{\gamma(1-\gamma)\sigma_s^2[1 + (\eta'_i\eta_i - 1)\gamma]}} \frac{[\mu(1-\gamma) - \gamma\eta'_i(\mathbf{y}_i - \mathbf{x}_i\beta)][(1-2\gamma) + (\eta'_i\eta_i - 1)\gamma(2-3\gamma)]}{(\gamma(1-\gamma)\sigma_s^2[1 + (\eta'_i\eta_i - 1)\gamma])^{3/2}} \quad (13)$$

$$\frac{\partial z_i^*}{\partial \eta} = \frac{\gamma \sum_{t \in \varphi(i)} (t-T)e^{-\eta(t-T)}(\mathbf{y}_i - \mathbf{x}_i\beta)}{\sqrt{\gamma(1-\gamma)\sigma_s^2[1 + (\eta'_i\eta_i - 1)\gamma]}} \frac{1}{2} \frac{[\mu(1-\gamma) - \gamma\eta'_i(\mathbf{y}_i - \mathbf{x}_i\beta)] \left[\frac{1}{2}\gamma^2(1-\gamma)\sigma_s^2 \frac{\partial \eta'_i\eta_i}{\partial \eta} \right]}{(\gamma(1-\gamma)\sigma_s^2[1 + (\eta'_i\eta_i - 1)\gamma])^{3/2}} \quad (14)$$

$$\text{e } \frac{\partial \eta'_i\eta_i}{\partial \eta} = -2 \sum_{t \in \varphi(i)} (t-T)e^{-2\eta(t-T)} \text{ se } \eta \neq 0.$$

Segundo a densidade normal truncada a eficiência da i -ésima firma no t -ésimo tempo pode ser predita da seguinte forma:

$$E(U_i | \varepsilon_i = e_i) = \mu_i^* + \sigma_i^* \left[\frac{\phi\left(\frac{-\mu_i^*}{\sigma_i^*}\right)}{1 - \Phi\left(\frac{-\mu_i^*}{\sigma_i^*}\right)} \right] \quad (15)$$

onde μ_i^* e σ_i^* estão definidos nas Equações 5 e 6. Sendo assim, para predizer a eficiência utilizaremos $\hat{U}_i = E(U_i | \varepsilon_i = e_i)$ e a eficiência técnica é dada por:

$$TE_{it} = 1 - e^{-\eta_{it}U_i} \quad (16)$$