

Análise Bayesiana de Modelos de Fronteiras de  
Produção Estocásticas

por

Luis Alberto Toscano Medrano

Orientador: Prof. Hélio S. Migon

Universidade Federal do Rio de Janeiro

Instituto de Matemática

Março de 2003

ANÁLISE BAYESIANA DE MODELOS DE  
FRONTEIRAS DE PRODUÇÃO ESTOCÁSTICA

Por

Luis Alberto Toscano Medrano

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO  
DE MATEMÁTICA DA UNIVERSIDADE FEDERAL DO RIO DE JA-  
NEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA OBTENÇÃO  
DO GRAU DE MESTRE EM CIÊNCIAS.

Área de concentração : Estatística.

Aprovada por :

---

Prof. Hélio dos Santos Migon  
Orientador  
IM - UFRJ

---

Prof. Hedibert Freitas Lopes  
IM - UFRJ

---

Prof. João Victor Issler  
EPGE - FGV

# Resumo

Esta dissertação se propõe em resolver três problemas referentes aos modelos de fronteira de produção estocástica. O primeiro problema trata dos aspectos numéricos do modelo de fronteira Normal-Gama, onde alguns parâmetros do modelo, não possuem a distribuição condicional conhecida e não é log-côncava. Nesta dissertação apresentaremos métodos alternativos para gerar estas distribuições condicionais e faremos apenas comparações. Os resultados mostraram que o algoritmo de Slice Sampling teve um bom desempenho para amostrar de uma distribuição condicional que não é log-côncava.

A outra parte desta monografia trata do problema da estimação empírica dos modelos de fronteira de produção estocástica, onde utiliza-se, com frequência, a função Coob-Douglas, embora esta possa ser bastante restritiva para descrever certos processos produtivos. O outro objetivo desta dissertação é comparar outras funções de produção alternativas e modelar a distribuição das ineficiências. Os resultados mostraram que o modelo de fronteira Normal-Lognormal e usando a forma funcional CES ou GPF é o mais apropriado. Também se observou que o modelo de fronteira normal-lognormal com apenas um parametro e utilizando qualquer forma funcional, conseguiu diferenciar melhor as firmas mais eficientes das menos eficientes.

# Abstract

This dissertation purpose solves three problems that refers to models for stochastic production frontier. The first problem refers to numeric aspects in the model of frontier Norma-Gamma where some parameters of this model doesn't have a conditional distribution known and it isn't log-concave. In this dissertation will be presented alternative methods to generate these conditional distributions and will be made only comparisons. The results showed that the Slice Sampling algorithm had good performance to sampling one conditional distribution that isn't log-concave.

The second part refers to the empirical estimation problems in the front models of stochastic production frontier where frequently is used the Cobb-Douglas function though this is restrictive to describe some productive processes. The other aim of this dissertation is compare other alternative production functions and model the inefficiency distributions. The results showed that the front model Normal-Lognormal and using a functional form CES or GPF is the most appropriated. Was observed that the front model Normal-Lognormal with only one parameter and using anyone functional form, achieved differences in the most efficient firms.

## Agradecimentos

- À meus pais por ter me ensinado com seu exemplo a perseverar e lutar para alcançar meus objetivos;
- Ao Professor Hélio Migon pela sugestão do assunto, e principalmente, pelo apoio e incentivo - sempre presentes - durante a realização deste trabalho;
- Às minhas amigas Aline, Madalena e Romy e meus amigos Rafael, Ralph e Ricardo pelo bom humor, amizade e o apoio que serviu para fazer-me sentir como em casa;
- Às pessoas que deixo de mencionar e que conviveram o dia a dia deste trabalho e que muitas vezes, sem saber, estavam me incentivando a continuar nesta caminhada;
- Ao Departamento de Métodos Estatísticos do Instituto de Matemática, pela oportunidade.

# Sumário

Agradecimentos	iii
Introdução	2
<b>1 Aspectos Numéricos em Fronteira de Produção Estocástica: O</b>	
<b>Caso Normal-Gama</b>	<b>5</b>
1.1 Introdução . . . . .	5
1.2 Fronteira de Produção Estocástica . . . . .	6
1.3 Fronteira de Produção Normal-Gama . . . . .	8
1.3.1 Distribuição a Priori . . . . .	9
1.3.2 Inferência Bayesiana . . . . .	11
1.4 Métodos MCMC . . . . .	13
1.4.1 Metropolis otimizado . . . . .	13
1.4.2 Slice Sampling . . . . .	15
1.4.3 Metropolis otimizado modificado . . . . .	18
1.4.4 Metropolis-Hastings . . . . .	19
1.5 Aplicação . . . . .	21
1.5.1 Dados Artificiais . . . . .	21
1.5.2 Exemplo com Dados Reais . . . . .	34
<b>2 Seleção de Modelos de Fronteira de Produção Estocástica</b>	<b>38</b>

2.1	Introdução . . . . .	38
2.2	Função de produção . . . . .	39
2.2.1	Produto Marginal . . . . .	40
2.2.2	Retorno de Escala . . . . .	41
2.2.3	Taxa Marginal de Substituição . . . . .	42
2.2.4	Elasticidade de Substituição . . . . .	43
2.2.5	Principais Tipos de Funções de Produção . . . . .	43
2.3	CrITÉrios de Seleção de Modelo . . . . .	48
2.3.1	Fator de Bayes . . . . .	48
2.3.2	Deviance Information Criterion (DIC) . . . . .	49
2.4	Comparação entre as Funções de Produção . . . . .	50
2.4.1	Implementação no WinBUGS . . . . .	52
2.4.2	Distribuição a Priori . . . . .	54
2.4.3	Análise MCMC . . . . .	54
2.4.4	Resultados . . . . .	56
2.5	Comparação entre os Modelos de Fronteira de Produção . . . . .	58
2.5.1	Distribuição a Priori . . . . .	60
2.5.2	Análise MCMC . . . . .	61
2.5.3	Resultados . . . . .	64
<b>3</b>	<b>Fronteira de Produção Estocástica para Múltiplos-Outputs</b>	<b>71</b>
3.1	Introdução . . . . .	71
3.2	Fronteira de Produção para Múltiplos- <i>Outputs</i> . . . . .	72
3.2.1	Modelo proposto por Fernández et al. (2000) . . . . .	73
3.3	Modelo Fatorial Bayesiano . . . . .	77
3.3.1	Modelo Fatorial . . . . .	78
3.4	Modelo Fatorial e Fronteira de Produção Estocástica . . . . .	81
3.4.1	Distribuição a Priori . . . . .	83

3.4.2	Métodos MCMC . . . . .	84
3.5	Aplicação . . . . .	87
3.5.1	Dados Artificiais . . . . .	87
3.5.2	Dados Reais . . . . .	93
<b>4</b>	<b>Conclusão e Extensões</b>	<b>99</b>



# Lista de Figuras

1.1	Duas alternativas para a densidade da eficiência: Gama e Exponencial . . . . .	9
1.2	Ilustração do algoritmo de Slice Sampling . . . . .	18
1.3	Densidade a posteriori para alguns parâmetros: $P, \theta, \sigma^2, \beta_2$ usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 1: $(P, \theta) = (0.8, 1)$ e $N=100$ . . . . .	24
1.4	Densidade a posteriori para os parâmetros $P, \theta, \sigma^2, \beta_2$ usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 1: $(P, \theta) = (0.8, 1)$ e $N=500$ . . . . .	25
1.5	Densidade a posteriori para alguns parâmetros: $P, \theta, \sigma^2, \beta_2$ usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 2: $(P, \theta) = (2, 1)$ e $N=100$ . . . . .	27
1.6	Densidade a posteriori para alguns parâmetros: $P, \theta, \sigma^2, \beta_2$ usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 2: $(P, \theta) = (2, 1)$ e $N=500$ . . . . .	28
1.7	Histograma da amostra gerada de $P$ usando o algoritmo Metropolis otimizado considerando $P = 2$ e $N=100$ . . . . .	29
1.8	Traço da sequência dos valores gerados de $P$ usando o algoritmo Metropolis otimizado considerando $P = 2$ e $N=100$ . . . . .	30

1.9	Densidade a posteriori para os parâmetros: $P$ usando o método Metropolis Otimizado para o parâmetro $P$ e Slice Sampling para o parâmetro $U$ considerando $P = 2$ e $N=100$ . . . . .	31
1.10	Densidade a posteriori para o parâmetro $P$ usando os métodos: Metropolis-Hastings e Slice Sampling considerando $P = 2$ e $N=100$	31
1.11	Análise de sensibilidade da média e do desvio padrão a posteriori de $\theta$ , $P$ e $VF$ versus os 15 valores definidos no intervalo de $[0.2, 4]$ para $P$ para os casos: Caso 1 - $VF = 0.95$ , Caso 2 - $VF = 0.75$ e Caso 3 - $VF = 0.5$ . . . . .	33
1.12	Densidade a posteriori para os parâmetros: $\theta$ , $P$ , $\sigma^2$ , $\beta_1$ usando o método de Slice Sampling . . . . .	36
2.1	Distribuição a posteriori da medida de eficiência e do rank para a melhor (rho[10]), pior (rho[4]) e uma mediana (rho[25]) firma segundo o modelo de fronteira Normal-Gama usando a função de produção Cobb-Douglas e CES. . . . .	67
2.2	Distribuição a posteriori da medida de eficiência e do rank para a melhor (rho[10]), pior (rho[4]) e uma mediana (rho[25]) firma segundo o modelo de fronteira Normal-Gama usando a função de produção Trannlog e GPF. . . . .	68
2.3	Distribuição a posteriori da medida de eficiência e do rank para a melhor (rho[10]), pior (rho[4]) e uma mediana (rho[25]) firma segundo o modelo de fronteira Normal-LogNormal usando a função de produção Cobb-Douglas e CES. . . . .	69
2.4	Distribuição a posteriori da medida de eficiência e do rank para a melhor (rho[10]), pior (rho[4]) e uma mediana (rho[25]) firma segundo o modelo de fronteira Normal-Lognormal usando a função de produção Trannlog e GPF. . . . .	70

3.1	Densidade a posteriori para todos os parâmetros do modelo de fronteira de produção fatorial usando o conjunto de dados simulados	91
3.2	Densidade a posteriori para todos os parâmetros do modelo proposto por Fernández et al. (2000) usando o conjunto de dados simulados . . . . .	92
3.3	Mediana da distribuição a posteriori dos Rank segundo o modelo proposto por Fernández et al. (2000) vs. a mediana da distribuição a posteriori dos Rank usando o modelo de fronteira de produção fatorial . . . . .	93
3.4	<i>Ranking</i> das firmas segundo o modelo proposto por Fernández et al. (2000) vs. o <i>ranking</i> das firmas segundo o modelo fronteira de produção fatorial . . . . .	94
3.5	Densidade a posteriori para todos os parâmetros do modelo de fronteira de produção fatorial usando o conjunto da Embrapa . . .	97
3.6	Histograma do <i>ranking</i> da eficiência para da melhor firma (rho[34]), pior firma (rho[4]) e uma firma mediana (rho[14]) segundo o modelo de fronteira de produção fatorial e o modelo proposto por Fernández et al. (2000) usando o conjunto de dados da Embrapa .	98

# Lista de Tabelas

1.1	Sumário da distribuição a posteriori para todos os parâmetros do modelo usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 1: $(P, \theta) = (0.8, 1)$ e $N=100$ (o verdadeiro valor dos parâmetros está entre colchetes) . . . . .	23
1.2	Sumário da distribuição a posteriori para todos os parâmetros do modelo usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 2: $(P, \theta) = (2, 1)$ e $N=100$ (o verdadeiro valor dos parâmetros está entre colchetes) . . . . .	26
1.3	Sumário da distribuição a posteriori para todos os parâmetros do modelo usando o método Slice Sampling e as abordagens de van den Broeck et al (1994) e Greene (1990) para o conjunto de dados da industria de serviço elétrico dos EUA . . . . .	35
1.4	Sumário da distribuição a posteriori das eficiência de cinco firma usando o método Slice Sampling e as abordagens de van den Broeck et al (1994) e Greene (1990) para o conjunto de dados da indústria de serviço elétrico dos EUA . . . . .	37
2.1	Resultado da Deviance Information Criterion para os quatro modelos de produção . . . . .	57
2.2	Sumário do pseudo fator de Bayes para os quatro modelos de produção . . . . .	57

2.3	Resultado do Deviance Information Criterion para os oito modelos de fronteira de produção . . . . .	65
2.4	Sumário do pseudo fator de Bayes para os oito modelos de fronteira de produção . . . . .	66
3.1	Sumário da distribuição a posteriori para todos os parâmetros do modelo de fronteira de produção fatorial usando o conjunto de dados simulados . . . . .	88
3.2	Sumário da distribuição a posteriori para todos os parâmetros do modelo proposto por Fernández et al. (2000) usando o conjunto de dados simulados . . . . .	89
3.3	Sumário da distribuição a posteriori para todos os parâmetros do modelo de fronteira de produção fatorial usando o conjunto de dados da Embrapa . . . . .	95



# Introdução

A idéia de fronteira de produção está intimamente ligada ao comportamento otimizador da firma. Ou seja, as firmas procuram, dada uma certa quantidade de insumos, maximizar seus lucros ou minimizar seus custos. Para isso, precisam atingir um nível máximo de produto utilizando um nível mínimo de insumos necessário para atingir tal quantidade do produto.

A noção de fronteira de produção é consistente com os princípios da otimalidade da teoria econômica em termos de otimização, onde cada desvio da fronteira possui uma interpretação clara e pode ser considerada como uma medida de ineficiência técnica.

Os modelos de fronteira produção estocástica consideram uma representação paramétrica de uma tecnologia e um termo de erro decomposto em duas componentes. Uma parte representa a flutuação aleatória que não depende do comportamento da firma. A outra, a medida de ineficiência passível do controle das firmas. Estes modelos são flexíveis o suficiente para englobar as fronteiras determinísticas.

Esta dissertação se propõe em resolver três problemas referentes aos modelos de fronteira de produção estocástica. O primeiro problema trata dos aspectos numéricos do modelo de fronteira Normal-Gama. Neste modelo alguns parâmetros não possuem a distribuição condicional a posteriori conhecida e assim, não é possível implementar o amostrador de Gibbs. Como as distribuições

condicionais não são log-côncava, não é possível, também, utilizar-se o método de amostragem de rejeição adaptativa (Gilks and Wild, 1992). Recentemente, Tsionas (2000) apresentou uma proposta para gerar destas distribuições condicionais, o qual denominamos de Metropolis otimizado. Nesta dissertação apresentaremos métodos alternativos, como o Slice Sampling (Neal, 1997) e Metropolis-Hastings (Metropolis et al., 1953, Hastings, 1970) e faremos apenas comparações dos resultados obtidos.

A outra parte desta monografia trata do seguinte problema: na maioria dos estudos a estimação empírica dos modelos de fronteira de produção estocástica utiliza-se, com frequência, a função Coob-Douglas, embora esta possa ser bastante restritiva para descrever certos processos produtivos. Outras especificações de funções de produção encontradas na literatura incluem: a Translog, CES (Constant Elasticity of Substitution) e GPF (Generalized Production Functions). O outro objetivo desta dissertação é comparar essas funções de produção alternativas e modelar a distribuição das ineficiências, como por exemplo: a Gama ou a Lognormal. As comparações serão feitas através dos critérios: pseudo fator de Bayes (Geisser e Eddy, 1979) e Deviance Information Criterion (Spiegelhalter et al., 2001).

O última questão abordada nesta dissertação trata do problema de múltiplos *outputs*. A maior parte dos estudos, para a utilização dos múltiplos outputs, constroem um único índice de produção, com os pesos fixos ou com pesos estimados (Fernandez *et al.*, 2000). Finalmente, o último objetivo desta dissertação é apresentar uma forma de estimar a fronteira de produção para caso de múltiplos *outputs*, utilizando a técnica de análise fatorial.

Essa dissertação está organizada da seguinte maneira: no capítulo 2 tratamos dos aspectos numéricos da fronteira de produção Normal-Gama, fazendo uma comparação do desempenho de alguns métodos através de um exemplo com



dados artificiais e, também, aplicamos à dados reais. Modelos de fronteira de produção com diferentes formas funcionais e erros assimétricos modelados através da distribuição Gama e Lognormal são comparadas usando-se dados reais anteriormente analisadas na literatura da área. No capítulo 3 apresentaremos um modelo de fronteira de produção para o caso de múltiplos *outputs*, utilizando a técnica de análise fatorial. As conclusões e as possíveis extensões do estudo estão apresentados no quarto capítulo.

# Capítulo 1

## Aspectos Numéricos em Fronteira de Produção Estocástica: O Caso Normal-Gama

### 1.1 Introdução

Este primeiro capítulo refere-se aos aspectos numéricos do modelo de fronteira de produção Normal-Gama. A fronteira Normal-Gama apresenta o problema de alguns dos seus parâmetros não possuírem a distribuição condicional a posteriori conhecida. Como a distribuição condicional não é log-côncava, não é possível utilizar algum método de amostragem de rejeição adaptativa (ARS). Recentemente, Tsionas (2000) apresentou uma proposta para gerar destas distribuições.

O objetivo deste capítulo é avaliar a performance dos métodos Metropolis otimizado, Slice Sampling (Neal, 1997) e o Metropolis-Hastings (Metropolis *et al.*, 1953 e Hastings, 1970) para realizar a inferência Bayesiana em modelos de fronteira estocástica Normal-Gama.

O restante do capítulo é organizado da seguinte forma. Na seção 2 é apre-

sentado o modelo de fronteira estocástica. O modelo de fronteira Normal-Gama, incluindo suas distribuições a priori e as condicionais completas são apresentado na seção 3. Na seção 4 alguns métodos proposto para fazer inferência Bayesiana quando a distribuição a posteriori não é tratável analiticamente são introduzidos. Finalmente, na seção 5 são descritas aplicação com dados artificiais e reais.

## 1.2 Fronteira de Produção Estocástica

Os modelos de fronteira de produção estocástica foram propostos inicialmente por Aigner, Lovell e Schmidt (1977) e Meeusen e van den Broeck (1977). Est abordagem consistiu, basicamente, em uma tentativa de superar as limitações das fronteiras determinísticas que não permitiam a presença de erros aleatórios, considerando todos os resíduos como ineficiência técnica das mesmas, controlados pelas firmas.

Neste modelo, a estimação das fronteiras utiliza tecnologias que admitem que o termo relativo ao erro seja dividido em duas partes: uma que meça a eficiência técnica, passível de controle pelas firmas; e outra que capture os erros aleatórios que estejam fora do controle das firmas. O modelo de fronteira estocástica possui a seguinte a forma geral:

$$y_i = f(x_i, \beta) + \varepsilon_i \quad \varepsilon_i = v_i - u_i \quad i = 1, \dots, N \quad (1.1)$$

onde  $y_i$  é geralmente o logaritmo do *output* (ou valor negativo do logaritmo do custo),  $x_i$  denota o logaritmo do vetor de *inputs*,  $\beta$  é um vetor  $k \times 1$  de coeficientes relacionados com a tecnologia adotada,  $f$  é a função de produção determinada,  $\varepsilon_i$  é o erro que está dividido em duas componentes: uma dada por  $v_i$  que assume-se ter distribuição normal com média 0 e variância  $\sigma^2$ , e outra dada por  $u_i$ , que assume-se que tenha uma distribuição assimétrica. Essa última parcela do erro  $\varepsilon$  mede a eficiência técnica através da diferença entre o *output* observado e o

*output* na fronteira. Assume-se que  $v_i$  e  $u_i$  sejam independentes. A eficiência que corresponde a firma  $i$  é definida por  $r_i = \exp(-u_i)$ .

O distúrbio não negativo  $u_i$  revela que o *output* de cada firma deve estar localizado sobre a fronteira ou abaixo dela. Qualquer desvio é devido a fatores que estão dentro do controle das firmas.

No contexto clássico são definidos diferentes distribuições para a parcela relativa à ineficiência, como a exponencial proposta por Meeusen e van den Broeck (1977), a half-normal por Aigner *et al.* (1977), a normal truncada proposta por Stevenson (1980) e a gama proposta por Greene (1990). No contexto Bayesiano, Migon em 2002 propõe usar a distribuição log-normal para descrever a ineficiência.

Um dos primeiros artigos sobre modelos de fronteira estocástica utilizando uma abordagem Bayesiana é apresentado em van den Broeck *et al.* (1994) que modelou a componente de ineficiência utilizando uma distribuição gama com parâmetro de forma conhecido. Em particular concentra-se no caso Erlang (valores inteiros no parâmetro de forma da gama) e utilizando o método de amostragem por importância para a realização da inferência Bayesiana. No caso em que o parâmetro de forma não é fixo, este método requer a avaliação de integrais que não podem ser aproximadas com uma precisão satisfatória. Este mesmo problema também foi encontrado por Greene (1990).

Koop, Osiewalski e Steel (1995) têm usado amostrador de Gibbs para analisar o modelo de fronteira normal-gama (com parâmetro de forma também suposto conhecido). Infelizmente, essas técnicas não se generalizam para o caso em que o parâmetro de forma é real positivo. Utilizando este mesmo modelo, outras contribuições são de Fernandez, Osiewalski e Steel (1997) e Koop, Osiewalski e Steel (1997) nas quais modelos para dados de painel são apresentados no contexto Bayesiano. O artigo de Osiewalski e Steel (1998) discute aspectos numéricos na análise Bayesiana para modelos de fronteira estocástica.

O modelo de fronteira Normal-Gama sem restrições no parâmetro de forma apresenta o problema de alguns dos seus parâmetros não possuírem a distribuição condicional a posteriori conhecida. Neste caso, recentemente, Tsionas(2000) propôs um método de aceitação para gerar valores dessas distribuições condicionais completas. Através de uma ilustração com dados artificiais, Tsionas(2000) mostra que o algoritmo que ele propõe, não tem estimativas precisas quando o parâmetro de forma do termo de ineficiência é maior que um. Mais detalhes serão apresentada na proxima seção.

A proposta neste capítulo é utilizar o método Slice Sampling (Neal, 1997) para realizar a inferência Bayesiana em modelos de fronteira estocástica com distribuição gama no termo de ineficiência e não se supõe que o parâmetro de forma seja conhecido ou que tenha que ser restrito a valores inteiros.

### 1.3 Fronteira de Produção Normal-Gama

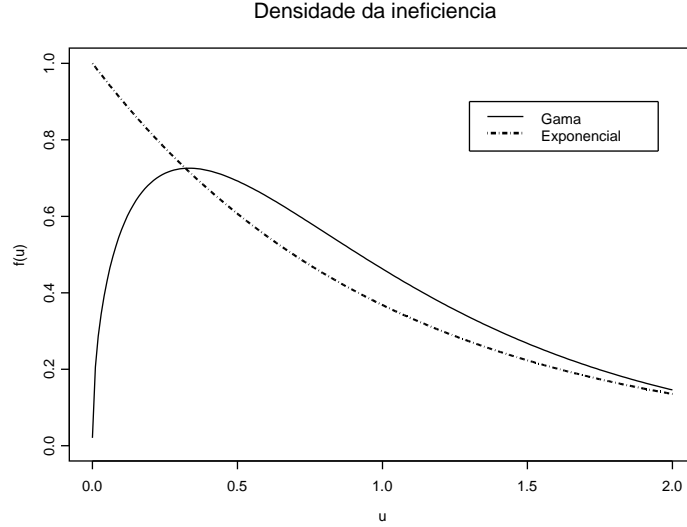
O modelo de fronteira normal-gama provém de uma extensão do modelo normal-exponencial. A Figura 1.2 ilustra o gráfico da distribuição Gama com parâmetro de escala  $\theta = 1.5$  e os parâmetros de forma  $P = 1.5$  e  $P = 1$  (distribuição exponencial). Observe que ambas as distribuições tem média 1 mas com formas diferentes na densidade. O valor de  $P$  maior que 1 a massa da distribuição se move da origem. Assim se assumimos o modelo de fronteira normal-gama temos várias alternativas na forma da densidade da eficiência.

Na abordagem Bayesiana, o modelo de fronteira estocástica normal-gama pode ser escrito da seguinte forma :

$$(y_i|\beta, \sigma^2, u_i) \sim N(f(x_i, \beta) - u_i, \sigma^2) \quad i = 1, \dots, N \quad (1.2)$$

onde  $(u_i|P, \theta) \sim G(P, \theta)$  para  $P > 0$  e  $\theta > 0$ . Neste capítulo vamos assumir que  $f(x_i, \beta) = x_i'\beta$ , ou seja, a função de produção Cobb-Douglas que é uma das

Figura 1.1: Duas alternativas para a densidade da eficiência: Gama e Exponencial



formas funcionais mais utilizadas em modelos de fronteiras estocástica. A função Cobb-Douglas será apresentada com mais detalhes no capítulo 2.

A densidade amostral das observações  $y_i$  dado  $\psi = (\beta, \sigma^2, P, \theta)$  pode ser vista como uma combinação de normal com o termo de ineficiência(gama) , isto é,

$$p(y_i|x_i, \psi) = \int_{\mathfrak{R}_+^N} f_N(y_i|f(x_i, \beta) - u_i, \sigma^2) f_G(u_i|P, \theta) du \quad (1.3)$$

Assumindo-se independência entre as firmas, a função de verossimilhança é dada pelo produto das densidades expressas em (1.3):

$$l(\psi|y, X) = p(y|X, \beta, \sigma^2, P, \theta) = \prod_{i=1}^N p(y_i|x_i, \psi), \quad (1.4)$$

### 1.3.1 Distribuição a Priori

Para realizar a inferência Bayesiana, é preciso complementar a função de verossimilhança com a distribuição a priori dos parâmetros  $(\beta, \sigma^2, P, \theta)$ . Foi escolhido

uma distribuição a priori própria com a seguinte estrutura:

$$p(\beta, \sigma^2, P, \theta) = p(\beta)p(\sigma^2)p(P)p(\theta) \quad (1.5)$$

Neste estudo, foram escolhidos valores para os hiperparâmetros levaram a uma priori relativamente vaga.

### **Priori para $\beta$**

A priori assumida no parâmetro de fronteira tem f.d.p de uma distribuição Normal N dimensional com vetor de média  $\mu$  e matriz de variância e covariância  $\Sigma$ .

$$p(\beta) = f_N^k(\beta|b_0, H_0^{-1}) \quad (1.6)$$

Na ilustração empírica foi definido  $b_0 = 0_k$ <sup>1</sup> e  $H_0 = 10^{-4} \times I_k$ <sup>2</sup>. Consequentemente, a priori escolhida é não informativa.

### **Priori para $\sigma$**

Seja  $\tau = \sigma^{-2}$ , assim a distribuição a priori para  $\tau$  tem f.d.p.

$$p(\tau) = f_G(\tau|n_0/2, a_0/2). \quad (1.7)$$

Na aplicação, foi considerado  $n_0/2 = 1$  (ou seja, uma priori exponencial para  $\tau$ ) e  $a_0/2 = 10^{-6}$ . Esses valores implicam uma priori com bastante incerteza.

### **Priori para $\theta$**

A priori definida para o parâmetro de escala tem a f.d.p. dada por

$$p(\theta) = f_G(\theta|v_0, \omega_0). \quad (1.8)$$

Uma priori não informativa foi usada na ilustração empírica, cujos os valores são  $v_0 = 0.1$  e  $\omega_0 = 0.01$ .

---

<sup>1</sup> $0_k$  denota um vetor de zero de dimensão  $k$ .

<sup>2</sup> $I_k$  denota uma matriz identidade de dimensão  $k$ .

## Priori para P

A distribuição a priori para P tem f.d.p. igual a

$$p(P) = f_G(P|d_0, \varepsilon_0), \quad (1.9)$$

Definiu-se uma priori vaga para P cujos hiperparâmetros foram  $d_0 = 1$  e  $\varepsilon_0 = 0.1$ .

### 1.3.2 Inferência Bayesiana

A função de verossimilhança juntamente com a distribuição a priori definida em (1.6)-(1.9) definem o modelo Bayesiano, ou seja, a distribuição conjunta das observações e dos parâmetros.

Para aplicar MCMC as distribuições condicionais completas devem ser avaliadas. Nesta seção, serão descritas todas as distribuições condicionais completas envolvidas na implementação do amostrador de Gibbs.

#### Distribuição condicional de $\beta$

Multiplicando (1.6) por (1.4) obtém-se a seguinte f.d.p para  $\beta$

$$P(\beta|y, X, \sigma^2, P, \theta, u) = f_N^k(\beta|\beta_*, H_*) \quad (1.10)$$

onde

$$\beta_* = (X'X)^{-1}[X'(y + u)]$$

e

$$H_* = \sigma^2(X'X)^{-1}$$



### Distribuição condicional de $\sigma^2$

Foi avaliada a distribuição condicional posteriori de  $\tau = \sigma^{-2}$ . A f.d.p desta distribuição é proporcional ao produto de (1.7) e (1.4), o qual resulta em

$$p(\tau|y, X, \beta, P, \theta, u) = f_G\left(\tau \mid \frac{(N+n_0)}{2}, \frac{((y+u-X\beta)'(y+u-X\beta)+a_0)}{2}\right) \quad (1.11)$$

### Distribuição condicional de $\theta$

De (1.8) e (1.4) podemos obter a seguinte f.d.p para  $\theta$ :

$$p(\theta|y, X, \beta, \sigma^2, P, u) = f_G\left(\theta \mid NP + v_0, \sum_{i=1}^N u_i + \omega_0\right) \quad (1.12)$$

### Distribuição condicional de $\mathbf{u}$

Fazendo-se uma amostragem separadamente para cada uma das  $N$  componentes de  $\mathbf{u}$ , a distribuição condicional a posteriori de  $u_i$ ,  $i = 1, \dots, N$  tem a f.d.p em  $(0, \infty)$  expressa por:

$$p(u_i|y, X, \beta, \sigma^2, \theta, P) \propto u_i^{P-1} \exp\left[\frac{-(u_i + e_i)^2}{2\sigma^2} - \theta u_i\right] \quad (1.13)$$

onde

$$e_i = y_i - x_i' \beta.$$

Em contraste com todas as distribuições condicionais apresentadas até agora, amostrar da distribuição condicional em (1.13) não é possível. Infelizmente o método ARS(Adaptive Rejection Sampling) proposto por Gilks (1992) não pode ser aplicado, pois quando  $P < 1$  a distribuição condicional de  $\mathbf{u}$  não é log-côncava. A maioria dos autores se fixam em  $P = 1$ , ou seja, assume-se uma fronteira de produção normal-exponencial, e assim a distribuição condicional a posteriori de  $\mathbf{u}$  é log-côncava e tem distribuição Normal Truncada. A estratégia adotada aqui nesta dissertação é de gerar  $\mathbf{u}$  univariadamente componente a componente através dos métodos Slice Sampling, Metropolis-Hastings e Metropolis otimizado. Estes métodos serão descritos na seção 1.4.2.

## Distribuição condicional posteriori de $P$

Do produto de (1.9) por (1.4), obtém-se a seguinte f.d.p para  $P$ :

$$p(P|Y, X, \beta, \sigma^2, \theta) \propto P^{d_0-1} \Gamma(P)^{-N} \exp(QP) I_{(0,\infty)}(P) \quad (1.14)$$

onde

$$Q = \sum_{i=1}^N \log u_i - \varepsilon + N \log \theta$$

Como a distribuição condicional de  $P$  não tem uma forma fechada, novamente algum método de amostragem aleatória deve ser implementada. Como também a distribuição condicional de  $P$  não é log-côncava quando  $d_0 < 1$ , um dos métodos que se esta propondo em utilizar é o algoritmo de Slice Sampling. Os outros métodos a serem utilizados são os algoritmos de Metropolis-Hastings e o proposto por Tsionas(2000), que sugere um método de aceitação e rejeição semelhante ao algoritmo Metropolis-Hastings, a diferença está na construção da distribuição proposta, onde seu parâmetro é alterado a cada passo do amostrador de Gibbs, e este parâmetro é determinado com o objetivo de maximizar a probabilidade de aceitação.

Para uma melhor compreensão do algoritmo proposto por Tsionas (2000)(Metropolis otimizado), do método Slice Sampling (Neal, 1997) e do algoritmo de Metropolis-Hastings serão descritos a seguir, com o objetivo de poder avaliar o desempenho dos métodos através de uma ilustração com dados artificiais.

## 1.4 Métodos MCMC

### 1.4.1 Metropolis otimizado

Conforme foi mencionado anteriormente, no modelo de fronteira normal-gama as distribuições condicionais do termo de ineficiência e também do parâmetro de forma ( $P$ ) da distribuição Gama não estão disponíveis para amostragem. Neste

caso, Tsionas(2000) propõe um método de aceitação e rejeição. A seguir será apresentado o esquema do algoritmo Metropolis otimizado para gerar valores da distribuição condicional de  $P$ .

Considere  $\pi(P)$  a distribuição condicional completa de  $P$  que está descrita na Eq.(1.14). Suponha que a distribuição proposta tenha distribuição exponencial com parâmetro  $\lambda$ , ou seja,

$$g(P, \lambda) = \lambda \exp(-\lambda P) \quad (1.15)$$

A seleção do parâmetro  $\lambda$  é feita com a finalidade de maximizar a probabilidade de aceitação  $R(P, \lambda)$  onde  $R(P, \lambda) = \pi(P)/g(P, \lambda)$ . Se denotamos  $r(P, \lambda) = \log R(p, \lambda)$ , segue que:

$$r(P, \lambda) = (Q + \lambda)P - \log \lambda - N \log \Gamma(P) + (d_0 - 1) \log P \quad (1.16)$$

onde  $Q = \sum_{i=1}^N \log u_i - \varepsilon + N \log \theta$  e  $\Gamma(\cdot)$  denota a função gama.

O problema da seleção de  $\lambda$  resume-se em encontrar o  $\max_P \text{mim}_\lambda : r(P, \lambda)$ , ou seja, primeiro aplicando a derivada em  $r(P, \lambda)$  com relação a  $\lambda$  e igualando a zero, obtemos a seguinte expressão:

$$\lambda = 1/P \quad (1.17)$$

A seguir aplicando a derivada em  $r(P, \lambda)$  com relação a  $P$  e igualando a zero, obtém-se a seguinte equação

$$Q + \lambda - N\psi(P) + (d_0 - 1)P^{-1} = 0 \quad (1.18)$$

e substituindo  $\lambda$  pela expressão obtida em (1.17), obtemos a seguinte equação não linear, que será resolvida em relação a  $P$ :

$$Q - N\psi(P) + d_0P^{-1} = 0 \quad (1.19)$$

onde  $\psi(x) \equiv \frac{d \log \Gamma(x)}{dx}$  é a função psi (ou digama). Assim, obtido o valor ótimo de  $P$  que será denotado por  $P^*$  e o valor ótimo de  $\lambda$  será  $\lambda = 1/P^*$ .

Após definir o valor ótimo de  $\lambda$ , o algoritmo segue 2 etapas:

1. Gerar um valor de  $P$  a partir de  $g(P, \lambda)$ .
2. Aceitar  $P$  com probabilidade  $\alpha(P) = R(P)/R(P^*)$ , ou seja,  $P$  será aceito se  $\alpha(P) \geq V$ , onde  $V$  é uniformemente distribuída em  $(0, 1)$ .

A equação não linear definida em (1.19) pode ser resolvida através do método de Newton-Raphson. Dado um valor inicial para  $P$ , as interações de Newton são como segue:

$$\Delta P = \frac{Q + d_0 p^{-1} - N\psi(P)}{d_0 P^{-2} + N\psi'(P)} \quad (1.20)$$

O mesmo procedimento é feito para gerar valores da distribuição condicional completa para cada  $u_i$ , com  $i = 1, \dots, N$ , sendo que neste caso,  $g(u_i, S)$  tem distribuição Gama. Para mais detalhes destes procedimentos, ver Tsionas (2000).

Como a componente de ineficiência  $u$  é um parâmetro de dimensão  $N$ , logo para implementar o algoritmo proposto por Tsionas (2000) seria preciso resolver o problema de otimização  $N$  vezes a cada passo do amostrador de Gibbs e isso levaria um certo custo computacional para ser realizado.

Como o algoritmo proposto por Tsionas (2000) não está implementado em nenhum pacote, o algoritmo teve que ser implementado na linguagem Fortran com o objetivo de poder avaliar a desempenho do método em dados artificiais.

### 1.4.2 Slice Sampling

No caso em que a distribuição condicional não é conhecida, alguns outros métodos podem ser usados, como por exemplo o Adaptive Rejection Sampling (ARS) (Gilks e Wid 1992; Gilks 1992), o qual só pode ser usado quando a distribuição condicional é log-côncava.

Já quando a distribuição condicional não é log-côncava um método muito usado é o algoritmo de Metropolis-Hastings (Metropolis, et al 1953, Hastings

1970), onde é necessário encontrar um candidato apropriado para ser a distribuição proposta.

Uma outra alternativa utilizada neste estudo quando a distribuição condicional não é log-côncava é o método de *Slice Sampling* proposto por Neal (1997). As principais vantagens do método Slice Sampling são:

- Requerer menos esforço na implementação;
- E, em particular, requerer um menor tempo de afinamento para se obter uma performance satisfatória do que outros métodos mais utilizados, como o algoritmo de Metropolis.

Para introduzir a idéia do *Slice Sampling*, suponha que se deseja amostrar da função de densidade  $f(x)$  onde  $x \in X \subseteq \Re$ . A idéia do método é formalizado introduzindo uma variável auxiliar  $y$  e definindo a distribuição conjunta de  $x$  e  $y$  que é uniforme sobre a região  $U = \{(x, y) : 0 < y < f(x)\}$  debaixo da curva definida por  $f(x)$ . Assim a densidade conjunta de  $(x, y)$  é

$$p(x, y) = \begin{cases} 1/Z, & \text{se } 0 < y < f(x) \\ 0, & \text{c.c.} \end{cases} \quad (1.21)$$

onde  $Z = \int f(x)dx$ . Consequentemente a densidade marginal para  $x$  é

$$p(x) = \int_0^{f(x)} (1/Z)dy = f(x)/Z \quad (1.22)$$

Para amostrar de  $x$ , devemos amostrar conjuntamente de  $(x, y)$ , e então ignorar  $y$ . Muitas vezes gerar pontos de  $(x, y)$  conjuntamente amostrando uniformemente de  $U$ , não pode ser fácil. Uma alternativa é o esquema de Gibbs Sampling. Para tanto é preciso definir as seguintes distribuições condicionais:

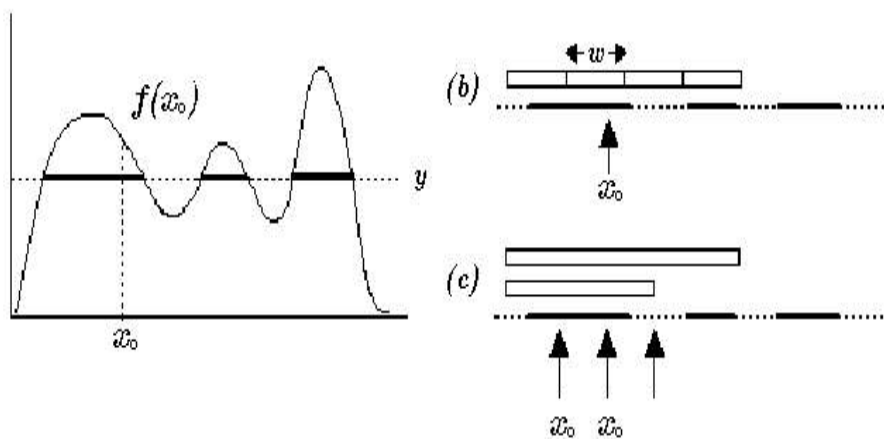
- $(y|x) \sim U(0, f(x))$
- $(x|y) \sim U(S(y))$  onde  $S(y) = \{x : y < f(x)\}$

Gerar um ponto de  $x$  amostrando uniformemente em  $S(y)$ , ainda pode ser difícil. Uma alternativa é proposta por Neal (1997) para amostrar um ponto de  $x$  segue 3 etapas:

- (a) Amostra-se um valor de  $y$  de uma uniforme de  $(0, f(x_0))$ , onde  $x_0$  é o valor inicial. Assim define-se a fatia ('slice') horizontal:  $S(y) = \{x : y < f(x)\}$ .
- (b) Encontra-se um intervalo,  $I = (L, R)$ , em torno de  $x_0$  que contenha a maior parte da fatia. Este segundo passo poder implementado de vários modos. A Figura 1.2 ilustra umas das formas, proposto por Neal (1997), para de construir o intervalo  $I$ . O esquema consiste em criar um intervalo de largura  $w$  que é posicionado aleatoriamente em torno de  $x_0$  e então esse intervalo é ampliado com incrementos de tamanho  $w$  até que os extremos do intervalos saiam fora da fatia.
- (c) O ultimo passo consiste em amostrar um novo ponto  $x_1$ , de uma parte da fatia pertencente ao intervalo, ou seja, de  $S \cap I$ . Assim um novo ponto  $x_1$  é amostrado de uma  $U(L,R)$  e este valor é aceito se  $y < f(x_1)$ , caso contrário se  $x_1 < x_0$  então  $L = x_1$  se não  $R = x_1$

Apesar que o método de Slice Sampling esta disponível no pacote WinBUGS, o algoritmo foi implementado em Fortran para poder fazer uma comparação com o método proposto por Tsionas (2000) e o algoritmo Metropolis-Hastings *random walk*.

Figura 1.2: Ilustração do algoritmo de Slice Sampling



### 1.4.3 Metropolis otimizado modificado

Uma modificação intuitiva para se poder melhorar o desempenho do algoritmo de Metropolis Otimizado, é otimizar até uma certa número de iterações, ou seja, se otimizamos até a  $i$ th replicação do algoritmo de MCMC e no restante das iterações seria utilizado a média dos valores encontrados até a  $i$ th iteração.

Para uma melhor compreensão a seguir será apresentado os passos do algoritmo do Metropolis otimizado modificado para gerar valores da distribuição condicional de  $P$ .

Novamente considerando  $\pi(P)$  seja a distribuição condicional completa de  $P$  que está descrita na Eq.(1.14). Suponha uma distribuição proposta com distribuição exponencial com parâmetro  $\lambda$

A primeira parte do algoritmo de Metropolis otimizado modificado consiste que na  $1$ th até a  $i$ th iteração o valor do parâmetro  $\lambda$  é escolhido com a finalidade de maximizar a probabilidade de aceitação, ou seja, o mesmo esquema do

algoritmo de Metropolis otimizado descrito na seção 1.4.1.

Após  $i$ th iteração, no algoritmo de Metropolis otimizado modificado o valor de  $\lambda$  será a média dos valores de  $\lambda$  encontrado na  $1$ th até a  $m$ th iteração  $(\lambda^{(1)}, \dots, \lambda^{(m)})$ , ou seja, o valor de  $\lambda$  será fixo até a ultima iteração. A média dos valores de  $\lambda$  será denotada por  $\bar{\lambda}$ . Assim, o algoritmo de Metropolis otimizado modificado segue duas etapas:

1. Gerar um valor de  $P$  a partir de  $g(P, \bar{\lambda})$ .
2. Aceitar  $P$  se  $\alpha(P) = R(P)/R(P_{ult}) \geq V$ , onde  $V$  é uniformemente distribuída em  $(0, 1)$  e  $P_{ult}$  é o último valor aceito de  $P$ .

O mesmo procedimento é realizado para gerar cada valor de  $u_i$  para  $i = 1, \dots, N$ . Este algoritmo também foi implementado na linguagem Fortran.

#### 1.4.4 Metropolis-Hastings

Uma outra forma muito utilizada para gerar amostrar da distribuição condicional quando não é tratável analiticamente, é o algoritmo de Metropolis-Hastings (Metropolis, et al 1953, Hastings 1970).

Seja  $\pi$  uma distribuição conhecida e suponha que desejamos gerar uma de  $\pi$  usando cadeias de Markov. Neste caso, temos que construir um núcleo de transição  $p(\theta, \phi)$  tal que  $\pi$  seja a distribuição de equilíbrio da cadeia. Uma forma fácil de fazer isto é quando  $p$  satisfaz a condição de reversibilidade da cadeia.

$$\pi(\theta)p(\theta, \phi) = \pi(\phi)p(\phi, \theta), \quad \forall \theta, \phi \quad (1.23)$$

que é conhecida como equação de equilíbrio detalhado (Green, 1995). Est é uma condição suficiente para que  $\pi$  seja a distribuição de equilíbrio da cadeia, pois o processo o processo de integração implica que

$$\int \pi(\theta)p(\theta, \phi)d\theta = \pi(\phi) \quad \forall \phi \quad (1.24)$$



O núcleo  $p(\theta, \phi)$  pode ser construído em duas partes: um núcleo arbitrário de transição  $q(\theta, \phi)$ , onde  $\int q(\theta, \phi)d\phi = 1$ , e uma probabilidade de aceitação  $\alpha(\theta, \phi)$  tal que

$$p(\theta, \phi) = q(\theta, \phi)\alpha(\theta, \phi), \quad \theta \neq \phi \quad (1.25)$$

O núcleo de transição  $q(\theta, \phi)$  propõe o movimento da cadeia e o processo está no ponto  $\theta$ , este gera um novo valor  $\phi$  a partir de  $q(\theta, \phi)$ . A expressão para a probabilidade de aceitação é

$$\alpha(\theta, \phi) = \min \left\{ 1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)} \right\}. \quad (1.26)$$

O quociente nesta expressão foi chamado de razão de teste por Hastings (1970). Como  $\alpha$  definido deste modo podemos ver que  $p(\theta, \phi)$  satisfaz a condição de reversibilidade.

A simulação de uma amostra de  $\phi$  usando método de Metropolis-Hastings poder descrito como segue

1. Inicialize o contador de iterações da cadeia em  $j = 1$  e arbitre o valor inicial  $\theta^{(0)}$ ;
2. mova a cadeia para um novo valor  $\phi$  gerado da densidade  $q(\theta^{(j-1)}, \cdot)$ .
3. Calcule a probabilidade de aceitação do movimento,  $\alpha(\theta^{(j-1)}, \phi)$ . Se  $u \leq \alpha$  onde  $u$  é uniformemente distribuída no intervalo  $(0, 1)$ , ou seja, se o movimento é aceito, então  $\theta^{(j)} = \phi$ . Se  $u > \alpha$  o movimento é rejeitado, então  $\theta^{(j)} = \theta^{(j-1)}$  e a cadeia não se movimenta.
4. Mude o contador de  $j$  pra  $j+1$  e retorne ao passo 2 até a convergência.

Um caso particular do algoritmo de Metropolis que será usado neste capítulo, é o algoritmo de Metropolis-Hastings *random walk* onde a distribuição proposta  $q(\theta, \cdot)$  tem distribuição Normal.

Como a distribuição da proposta tem densidade normal e distribuição condicional de  $P$  está definida para  $p \in \mathfrak{R}^+$ , o mais conveniente fazer uma transformação em  $P$ , para que a distribuição esteja definida nos reais. Assim seja  $\pi(P)$  a distribuição condicional de  $P$  e  $Z = \log P$ , temos que a distribuição de  $Z$  é igual a

$$f(Z) = \pi(\exp Z)|J| \tag{1.27}$$

onde  $J = \exp Z$

O mesmo procedimento é feito para gerar uma amostra do parâmetro  $u_i$  para  $i = 1, \dots, N$ . O algoritmo de Metropolis-Hastings *random walk* também foi implementado na linguagem Fortran.

## 1.5 Aplicação

### 1.5.1 Dados Artificiais

A primeira análise tem como objetivo em avaliar a performance dos métodos apresentados na seção 1.4, através de uma análise com dados artificiais. Para está análise, deve-se levar em consideração alguns cuidados:

1. Que a influência da priori não seja muito forte e não domine a informação obtida na amostra.
2. Que a convergência dos parâmetros seja alcançada.

As duas primeiras aplicações são feitas a partir de dados que foram gerados artificialmente através de uma rotina feita em S-plus. No primeiro exercício será apresentado uma comparação dos algoritmos de Slice Sampling e Metropolis Otimizado. No outro exercício apresentaremos uma comparação entre os métodos Slice Sampling, Metropolis otimizado, Metropolis otimizado modificado e o Metropolis *random walk* em amostras onde o parâmetro  $P > 1$ .

## Exercício 1

Neste exercício foram gerados dois conjuntos de dados: um com  $N = 100$  firmas e outro com  $N = 1000$  firmas e  $k = 3$  regressores com os coeficientes iguais a  $\beta = (1, 1, 1)$ . A matriz  $X$  foi obtida da seguinte forma: Na primeira coluna colocou-se um vetor de 1's correspondendo ao intercepto. A segunda e terceira colunas, foram geradas de uma distribuição  $N(0, 1)$  para cada uma das  $N$  firmas.

Com relação a  $P$  e  $\theta$  serão examinados dois casos. O primeiro, **caso 1**, quando  $(P, \theta) = (0.8, 1)$ . A escolha destes valores é sugerida no artigo do Tsionas(2000) onde é examinado o caso de  $P < 1$ . Nesta situação existe uma dificuldade relativa, pois como já foi comentado anteriormente, a distribuição condicional de  $u_i$  não é log-côncava. O segundo, **caso 2**, é quando  $(P, \theta) = (2, 1)$ , ou seja,  $P \geq 1$ . E  $\sigma^2 = 0.05$  nos dois casos.

Três funções dos parâmetros serão examinadas, a ineficiência média, definida por  $m = P/\theta$ ; o desvio padrão da ineficiência, o qual é dado por  $S = (P/\theta^2)^{1/2}$ , e a parte da variância da fronteira sobre a variância total, que é dado por  $VF = \frac{P/\theta^2}{P/\theta^2 + \sigma^2}$ .

Na ilustração empírica, foram realizadas 15.000 iterações do algoritmo de MCMC e foram descartadas as 5.000 iterações iniciais para o período de *burn-in*.

A convergência dos parâmetros foram cheçadas observando o gráfico da sequência dos valores gerados da saída do amostrador de Gibbs para duas cadeias que iniciam de valores diferentes. Assim, pode-se ter uma razoável confiança de que a convergência foi alcançada.

Os resultados empíricos para o **caso 1** utilizando o método Metropolis Otimizado e o Slice Sampling para uma amostra de tamanho 100, são apresentados na Tabela 1.1 a qual mostra a média, desvio padrão, mediana e os percentis de 2.5 e 97.5 a posteriori de todos os parâmetros do modelo. Observa-se, em ambos os métodos, que as médias a posteriori estão todas razoavelmente concentradas em

torno do valor real. Além disso, os valores reais estão todos contidos no intervalo de 95% de credibilidade.

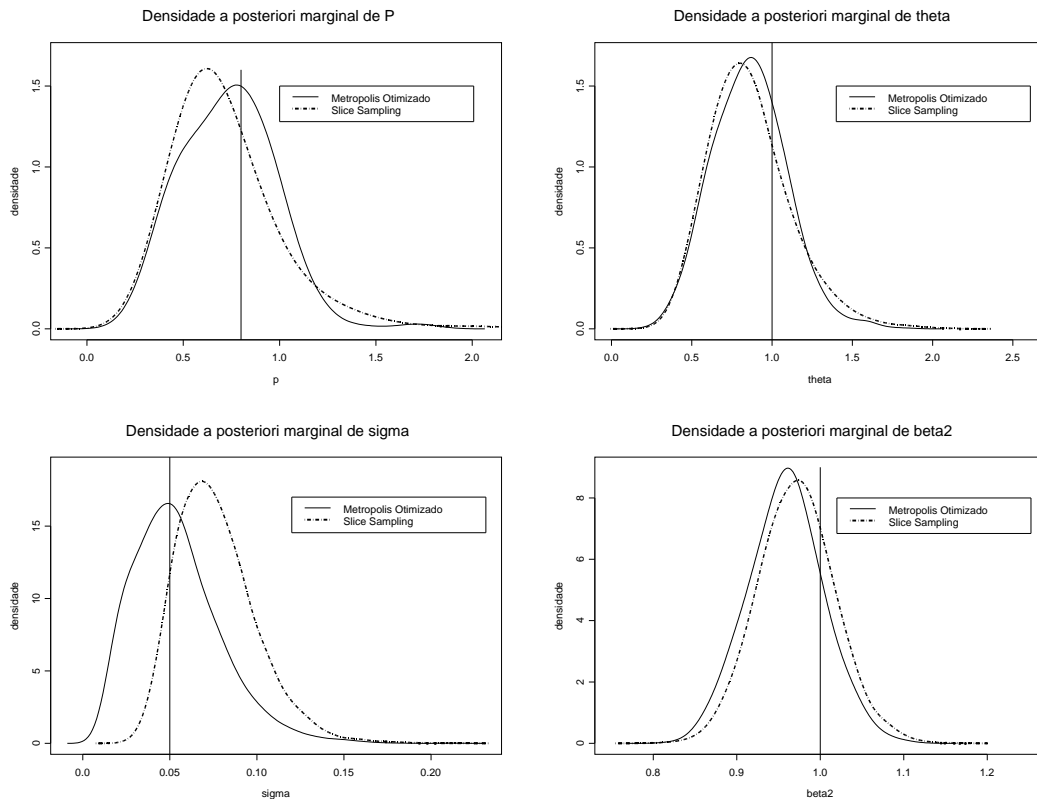
Tabela 1.1: Sumário da distribuição a posteriori para todos os parâmetros do modelo usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 1:  $(P, \theta) = (0.8, 1)$  e  $N=100$  (o verdadeiro valor dos parâmetros está entre colchetes)

Caso 1	Métodos									
N=100	Slice Sampling					Metropolis Otimizado				
Parâmetros	Media	d.p.	Med	2.5%	97.5%	Media	d.p.	Med	2.5%	97.5%
$\beta_1[1]$	1.026	0.103	1.016	0.960	1.082	1.050	0.107	1.063	0.973	1.125
$\beta_2[1]$	0.972	0.046	0.972	0.940	1.001	0.959	0.045	0.959	0.929	0.988
$\beta_3[1]$	0.914	0.048	0.913	0.883	0.945	0.910	0.048	0.911	0.882	0.941
$\sigma[0.05]$	0.077	0.023	0.074	0.060	0.090	0.054	0.025	0.051	0.036	0.068
$\theta[1]$	0.862	0.255	0.834	0.681	1.003	0.861	0.234	0.857	0.695	1.011
$P[0.8]$	0.716	0.276	0.662	0.533	0.838	0.733	0.243	0.738	0.554	0.892
$M[0.8]$	0.831	0.181	0.816	0.702	0.945	0.847	0.148	0.847	0.747	0.944
$s[0.89]$	1.009	0.201	0.978	0.867	1.117	1.017	0.171	0.994	0.900	1.108
$VF [0.94]$	0.923	0.036	0.930	0.904	0.950	0.948	0.029	0.953	0.934	0.968

Na Figura 1.3, são apresentadas as densidades marginais a posteriori dos parâmetros  $P, \theta, \sigma^2, \beta_2$ , utilizando os métodos de Metropolis otimizado e Slice Sampling, para o Caso 1 e  $N = 100$ . Observa-se que, independente do método utilizado, a densidade a marginal a posteriori de todos os parâmetros estão razoavelmente concentrados em torno dos verdadeiros valores utilizados para gerar os dados.

Para uma amostra de tamanho  $N = 500$ , a performance encontrada também é satisfatória em ambos os algoritmos. As densidades marginais a posteriori para os parâmetros  $P, \theta, \sigma^2, \beta_2$ , utilizando os dois métodos e o Caso 1 para  $N = 500$ , estão apresentados na Figura 1.4. Observa-se que, os dois algoritmos tiveram

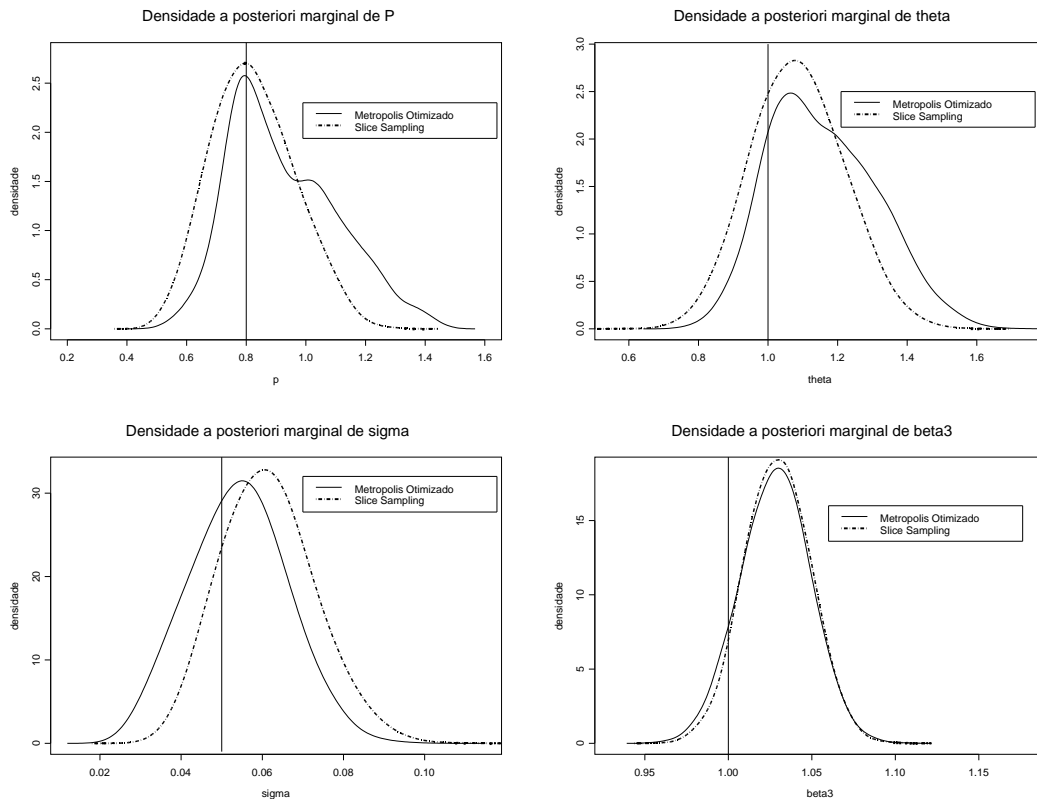
Figura 1.3: Densidade a posteriori para alguns parâmetros:  $P, \theta, \sigma^2, \beta_2$  usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 1:  $(P, \theta) = (0.8, 1)$  e  $N=100$



uma performance aceitável, pois as densidades a posteriori estão centralizadas em torno do valor verdadeiro. Nota-se também que as densidade a posteriori de  $P, \theta, \sigma^2, \beta_3$  estão mais concentrados em torno do valor real do que as densidades a posteriori obtidas na amostra de tamanho 100, o que já era esperado que o desvio padrão a posteriori diminua quando se aumenta o tamanho da amostra.

Na Tabela 1.2 apresenta-se os momentos a posteriori de todos os parâmetros para o caso 2, onde  $P = 2$  e  $\theta = 1$ , usando um tamanho amostral igual a 100. Percebe-se que as médias a posteriori estão todas concentradas em torno do valor

Figura 1.4: Densidade a posteriori para os parâmetros  $P, \theta, \sigma^2, \beta_2$  usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 1:  $(P, \theta) = (0.8, 1)$  e  $N=500$



real quando o algoritmo Slice Sampling é utilizado. Já os resultados obtidos utilizando-se o algoritmo Metropolis Otimizado são diferentes, pois observa-se que a média a posteriori do parâmetro  $P$  não está concentrada em torno do valor real.

As densidades marginais a posteriori de todos parâmetros usando os dois métodos são apresentadas na Figura 1.5 para o Caso 2 e  $N = 100$ . Observa-se claramente que utilizando o algoritmo de Slice Sampling, todos os parâmetros estão bem concentrados em torno dos valores reais. Um quadro um pouco dife-

Tabela 1.2: Sumário da distribuição a posteriori para todos os parâmetros do modelo usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 2:  $(P, \theta) = (2, 1)$  e  $N=100$  (o verdadeiro valor dos parâmetros está entre colchetes)

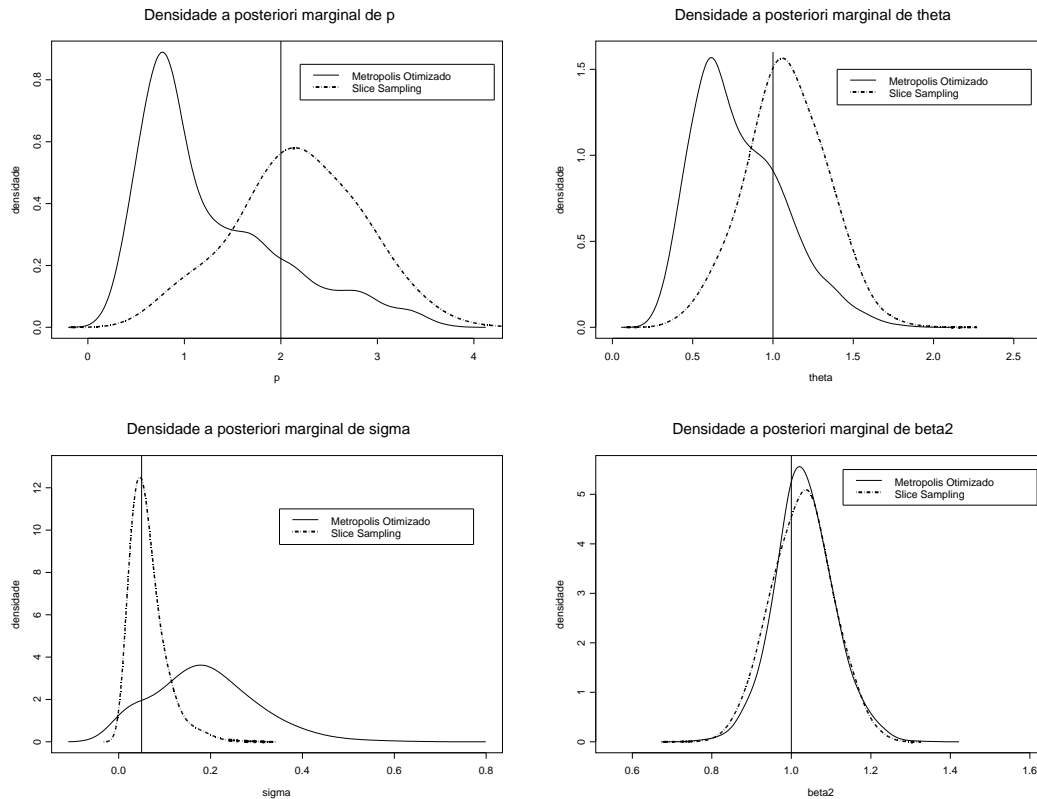
Caso 2	Métodos									
N=100	Slice Sampling					Metropolis Otimizado				
Parâmetros	Media	d.p.	Med	2.5%	97.5%	Media	d.p.	Med	2.5%	97.5%
$\beta_1[1]$	0.944	0.245	1.001	0.817	1.122	0.469	0.369	0.362	0.194	0.800
$\beta_1[1]$	1.027	0.077	1.028	0.974	1.078	1.033	0.075	1.029	0.986	1.079
$\beta_3[1]$	0.983	0.083	0.986	0.927	1.039	0.977	0.085	0.973	0.917	1.033
$\sigma^2[0.05]$	0.064	0.039	0.055	0.037	0.080	0.188	0.111	0.181	0.111	0.255
$\theta[1]$	1.082	0.253	1.079	0.919	1.256	0.794	0.284	0.742	0.580	0.982
$P[2]$	2.174	0.668	2.175	1.745	2.644	1.276	0.738	0.994	0.746	1.717
$m[2]$	1.985	0.322	2.012	1.790	2.203	1.515	0.393	1.430	1.222	1.806
$s[1.41]$	1.376	0.178	1.353	1.253	1.470	1.418	0.214	1.391	1.268	1.528
$VF[0.975]$	0.966	0.021	0.971	0.957	0.982	0.912	0.057	0.920	0.882	0.951

rente é observado quando  $P > 1$ . Quando o método proposto por Tsionas(2000) é utilizado, a densidade marginal a posteriori de  $P$  fica assimétrica à direita. Os resultados obtidos utilizando o algoritmo Metropolis Otimizado são similares aos relatados por Tsionas(2000).

Finalmente para o segundo caso com  $N = 500$ , ambos os algoritmos tiveram um desempenho bom. As densidades marginais a posteriori para os parâmetros  $P, \theta, \sigma^2, \beta_2$  utilizando os dois métodos, para o Caso 2 e  $N = 500$  estão apresentadas na Figura 1.6. Nota-se que os dois algoritmos produziram resultados satisfatórios, pois todas as densidades estão centradas em torno dos valores verdadeiros.

O algoritmo de Metropolis Otimizado usado para amostrar da distribuição condicional de  $P$  teve uma taxa de aceitação muito baixa. Por exemplo, nas

Figura 1.5: Densidade a posteriori para alguns parâmetros:  $P, \theta, \sigma^2, \beta_2$  usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 2:  $(P, \theta) = (2, 1)$  e  $N=100$



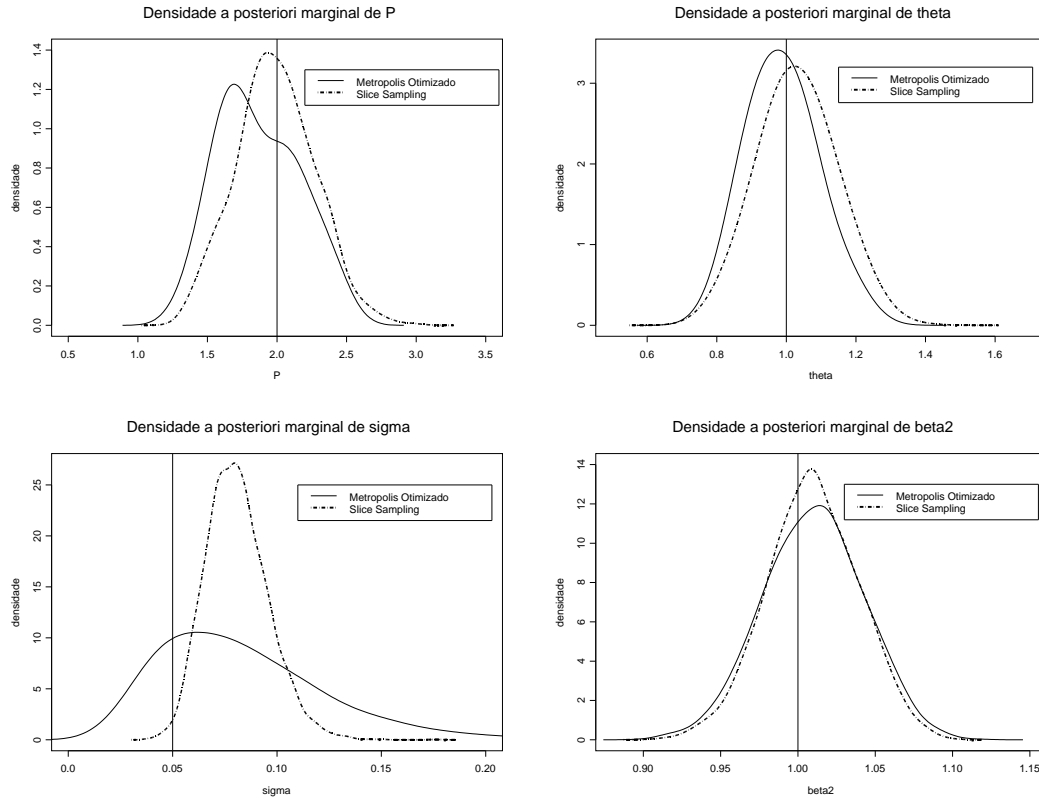
15000 iterações realizadas apenas foram aceitos em torno de 951 valores de  $P$ , para  $N=100$ . Para uma amostra de tamanho 500, o número de valores aceitos diminui para cerca de 446.

O tempo gasto no Fortran rodar 15.000 iterações com uma amostra de tamanho 100 usando o método de Metropolis otimizado foi aproximadamente 12 minutos e utilizando o método de Slice Sampling o tempo foi em torno de 7 minutos. Os programas foram executados em um Pentium II com 200 MHz, com 256 Mb de memória RAM.

Com relação a influência das prioris, foram experimentados vários outros va-



Figura 1.6: Densidade a posteriori para alguns parâmetros:  $P, \theta, \sigma^2, \beta_2$  usando os métodos: Slice Sampling e Metropolis Otimizado, considerando o caso 2:  $(P, \theta) = (2, 1)$  e  $N=500$



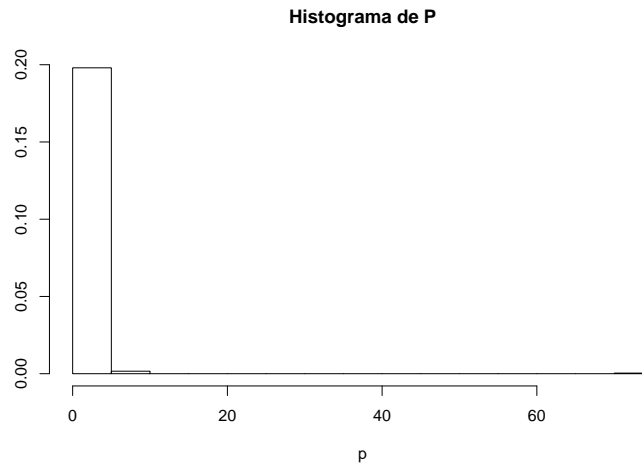
lores para os hiperparâmetros das priors em (1.6)-(1.9) e percebe-se que os resultados a posteriori não se alteram, o que indica que as priors não influenciaram nos resultados obtidos. Por exemplo, se definimos que  $H_0 = I_k$  e multiplicamos  $a_0, \omega_0$  e  $\varepsilon_0$  por 100, os resultados não se alteram.

## Exercício 2

Os resultados obtidos no exercício anterior mostraram que o algoritmo proposto por Tsionas(2000) não obteve uma boa performance no conjunto de dados artificiais onde  $P = 2$  e  $N = 100$ , assim neste segundo exercício será apresentado uma comparação entre os métodos Slice Sampling, Metropolis otimizado, Metropolis otimizado modificado e Metropolis-Hastings , através de um conjunto de dados artificial com as mesmas características do **caso 2**, ou seja:  $N = 100$ ,  $(P, \theta) = (2, 1)$ ,  $\beta = (1, 1, 1)$ ,  $X \sim N(0, 1)$ , e  $\sigma^2 = 0.05$ .

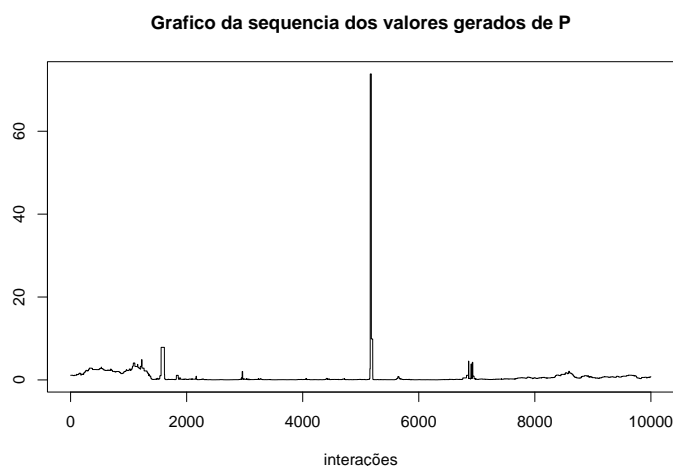
As estimativas das densidades a posteriori marginais de  $P$  utilizando o algoritmo de Metropolis otimizado foram muito ruins. Pelas Figuras 1.7 e 1.8, podemos notar a presença de valores absurdos na amostra gerada do parâmetro  $P$ , como por exemplo o valor 60. Estes mesmos resultados foram encontrados utilizando outros conjunto de dados artificiais.

Figura 1.7: Histograma da amostra gerada de  $P$  usando o algoritmo Metropolis otimizado considerando  $P = 2$  e  $N=100$



Uma estratégia adotada para obter-se um melhor resultado, é usar conjuntamente os algoritmos Metropolis otimizado e o Slice Sampling, para gerar da distribuição condicional de  $P$  e  $U$  respectivamente. Está estratégia tem a fina-

Figura 1.8: Traço da sequência dos valores gerados de  $P$  usando o algoritmo Metropolis otimizado considerando  $P = 2$  e  $N=100$



lidade de observar se o mal desempenho do algoritmo de Metropolis otimizado apresentado nas Figuras 1.7 e 1.8, é devido a mal geração de valores da distribuição condicional de  $\mathbf{u}$ . A estimativa da densidade a posteriori de  $P$  usando conjuntamente o algoritmo de Metropolis otimizado e o Slice Sampling está apresentada na Figura 1.9. Pela Figura 1.9 observa-se que a estimativa da densidade a posteriori de  $P$  está centrado em torno do valor verdadeiro de  $P(P=2)$ . Este resultado mostra que o método proposto por Tsionas (2000) está gerando valores absurdos de  $\mathbf{u}$  e por consequência está influenciado diretamente na geração dos valores de  $P$ .

Para finalizar o exercício 2, a Figura 1.10 apresenta a densidade a posteriori de  $P$  usando os algoritmos de Slice Sampling e Metropolis-Hastings para gerar uma amostra de  $P$  e  $U$ . Pela Figura 1.10 podemos notar que ambos os algoritmos tiveram um bom desempenho, pois as densidades estão bem centradas em torno do valor verdadeiro de  $P$ .

Figura 1.9: Densidade a posteriori para os parâmetros:  $P$  usando o método Metropolis Otimizado para o parâmetro  $P$  e Slice Sampling para o parâmetro  $U$  considerando  $P = 2$  e  $N=100$

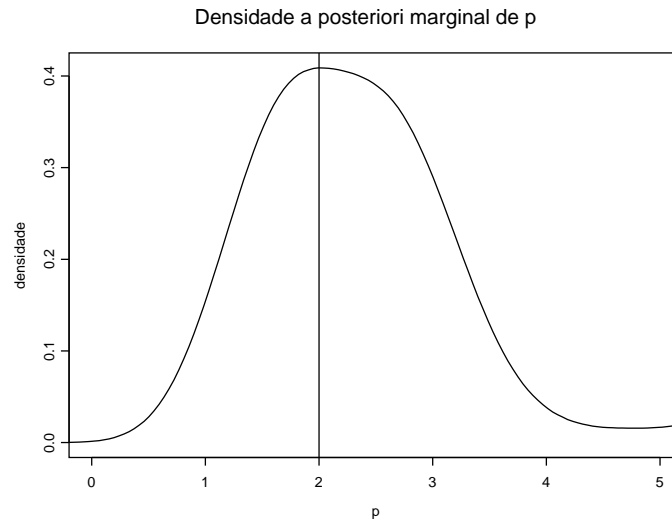
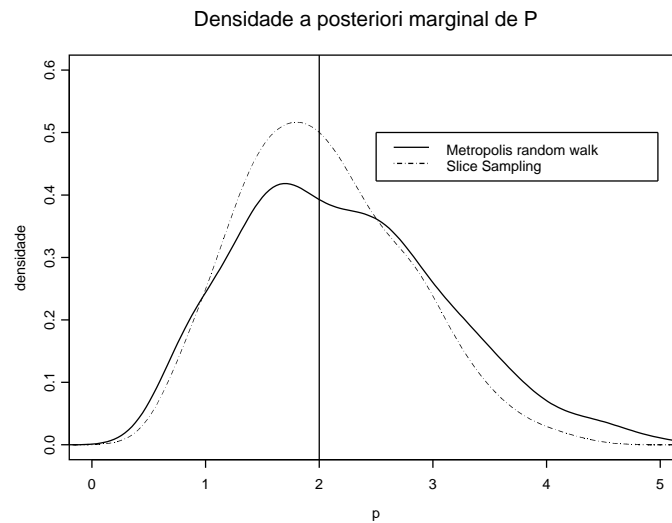


Figura 1.10: Densidade a posteriori para o parâmetro  $P$  usando os métodos: Metropolis-Hastings e Slice Sampling considerando  $P = 2$  e  $N=100$



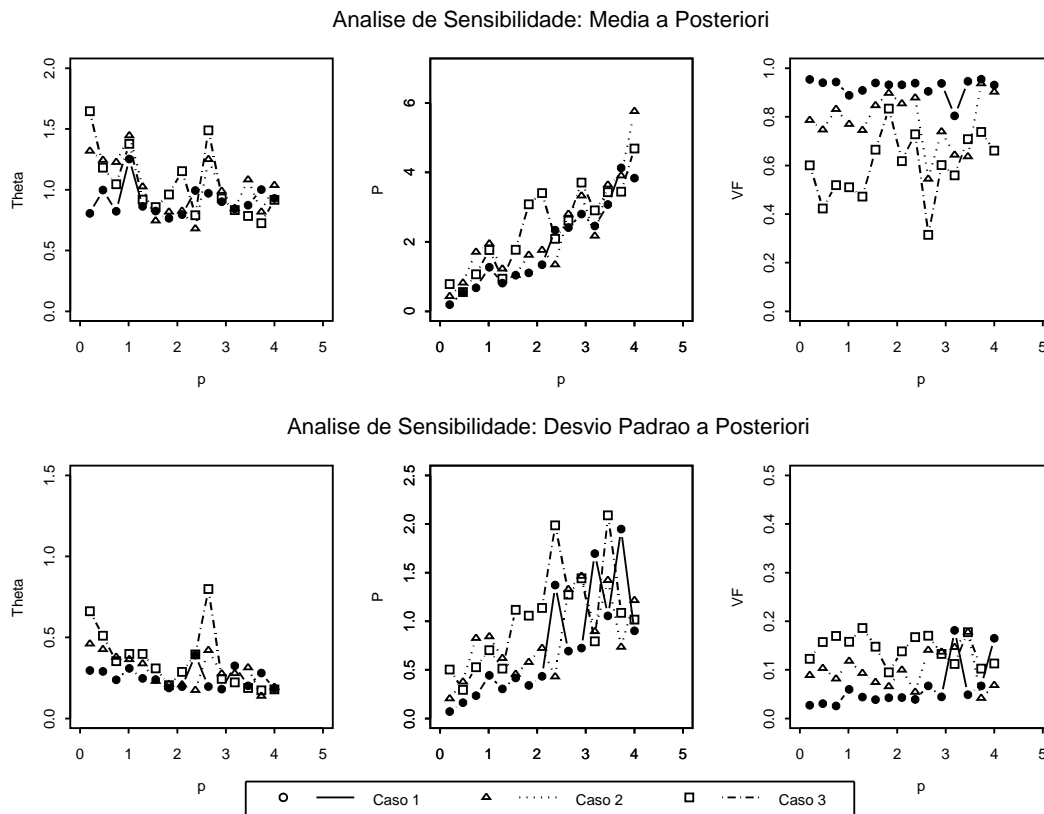
## Análise de Sensibilidade

A proposta deste experimento é verificar se os resultados encontrados são robustos a valores alternativos de  $P$  usando o algoritmo de Slice Sampling. Seja  $\sigma^2 = \frac{P(1-VF)}{\theta VF}$  e considere três casos: Caso 1 -  $VF = 0.95$ , Caso 2 -  $VF = 0.75$  e Caso 3 -  $VF = 0.5$ . Com relação a  $P$ , para os três casos foram definidos 15 valores no intervalo de  $[0.2,4]$ . Para tanto, foi necessário executar 45 vezes o amostrador de Gibbs e foram realizadas 5.000 iterações descartando as 1.500 primeiras para diminuir os possíveis efeitos iniciais. Em todos esses casos, o valor verdadeiro de  $\theta$  é 1.

A Figura 1.11 apresenta o gráfico da média e desvio padrão a posteriori de  $\theta$ ,  $VF$  e  $P$  versus os 15 valores definidos para  $P$  para os três casos. Pela Figura 1.11 observa-se que quanto mais  $VF$  se aproxima de 1, ou seja, a maior parte da variabilidade do modelo é referente ao termo de ineficiência ( $var(v) \ll var(u_i)$ ), mais robustas ficam as estimativas. Já no caso 3, quando  $VF = 0.5$ , ou seja,  $\sigma^2 = var(u_i)$ , percebe-se uma certa sensibilidade nas estimativas.

Assim, podemos concluir que quanto maior é a parcela da variabilidade do modelo de fronteira Normal-Gama que vem da variância da componente Normal, menor é a precisão das estimativas dos parâmetros.

Figura 1.11: Análise de sensibilidade das média e do desvio padrão a posteriori de  $\theta$ ,  $P$  e  $VF$  versus os 15 valores definidos no intervalo de  $[0.2, 4]$  para  $P$  para os casos: Caso 1 -  $VF = 0.95$ , Caso 2 -  $VF = 0.75$  e Caso 3 -  $VF = 0.5$



## 1.5.2 Exemplo com Dados Reais

Aqui será apresentada uma ilustração com dados reais. Para tanto foram considerados os dados coletados por Christensen e Greene (1976) para 123 companhias de serviço elétrico nos Estados Unidos em 1970. Estes dados estão listados no apêndice do artigo de Greene(1990), que também foram usados por Van den Broeck et al. (1994).

Christensen e Greene (1976) e Greene (1990) ajustaram uma função de custo sugerido por Nerlove (1963) baseando-se na função de produção Cobb-Douglas, porém generalizaram incluindo o termo quadrático (log do output  $Q$ ), passando a permitir que o retorno de escala <sup>3</sup> varie com  $Q$ . Nestes dados temos três fatores de produção: trabalho, capital e combustível com os respectivos preços  $P_l$ ,  $P_k$  e  $P_f$ ; e com a especificação da função de custo dada por

$$\begin{aligned} y_i = & - \beta_0 - \beta_1 \ln Q_i - \beta_2 \ln^2 Q_i - \beta_3 \ln(P_l/P_f) \\ & - \beta_4 \ln(P_k/P_f) + v_i - u_i, \end{aligned} \quad (1.28)$$

onde  $y_i = -\ln(\text{custo da firma } i/P_f)$

O objetivo principal nesta aplicação é mostrar que o algoritmo Slice Sampling tem um bom desempenho em aplicações com dados reais.

Uma das vantagens da abordagem Bayesiana é que a f.d.p a posteriori da eficiência de determinada firma específica  $r_i = \exp(-u_i)$  pode ser calculada.

As estatísticas obtidas foram baseadas em 20.000 iterações, das quais foram descartadas as 10.000 primeiras para o período de *burn-in*. Na Tabela 1.3 são apresentadas as médias, os desvios padrão, as medianas e os percentis de 2.5 e 97.5 a posteriori de todos os parâmetros usando o algoritmo *Slice Sampling* juntamente com os resultados apresentados por Greene(1990) e Van den Broeck et al. (1994) que utilizam uma abordagem Clássica e Bayesiana respectivamente.

---

<sup>3</sup>A definição de retorno de escala se encontra no capítulo 2

Tabela 1.3: Sumário da distribuição a posteriori para todos os parâmetros do modelo usando o método Slice Sampling e as abordagens de van den Broeck et al (1994) e Greene (1990) para o conjunto de dados da industria de serviço elétrico dos EUA

Parâmetros	Fronteira de Normal-Gama								
	Slice Sampling					van den Broeck et al 1994		Greene 1990	
	Média	d.p.	Med	2.5%	97.5%	Média	d.p.	Média	d.p
$\beta_1$	-7.4712	0.3624	-7.4709	-7.7139	-7.2260	-7442	0.343	-7.810	0.378
$\beta_2$	0.4132	0.0432	0.4132	0.3848	0.4418	0.407	0.042	0.473	0.043
$\beta_3$	0.0302	0.0029	0.0302	0.0283	0.0322	0.030	0.003	0.026	0.003
$\beta_4$	0.2477	0.0699	0.2473	0.2007	0.2938	0.257	0.068	0.291	0.068
$\beta_5$	0.0588	0.0650	0.0582	0.0155	0.1020	0.060	0.064	0.025	0.066
$\sigma^2$	0.0152	0.0034	0.0148	0.0128	0.0172	0.016	0.004	0.018	0.002
$\theta$	17.4316	6.5614	16.8275	12.2361	21.6908	23.446	11.368	21.347	3.944
$P$	2.8410	2.0125	2.1096	1.1299	4.1924	2	-	2.450	1.102
$m$	0.1492	0.0661	0.1407	0.0945	0.2013	-	-	0.115	-
$s$	0.0930	0.0198	0.0928	0.0805	0.1052	-	-	0.005	-
$VF$	0.3673	0.1182	0.3658	0.2872	0.4462	-	-	0.226	-

Pela Tabela 1.3 observa-se que os resultados não são muito diferentes dos encontrados por Greene (1990) e Van den Broeck et al. (1994). A Figura 1.12 mostra as densidades marginais a posteriori para os parâmetros  $\theta$ ,  $P$ ,  $\sigma^2$ ,  $\beta_1$  utilizando o algoritmo de *Slice Sampling*.

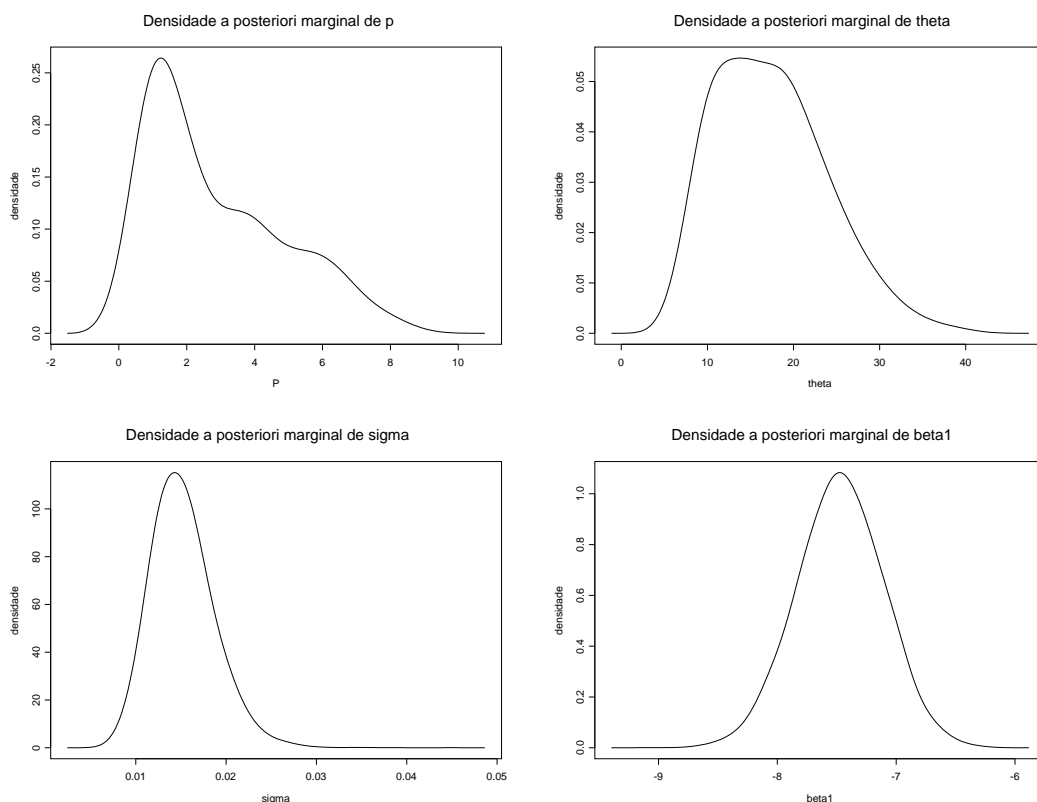
Os momentos a posteriori para as eficiências individuais  $r_i$  ( $i = 1, \dots, 5$ ), das cinco primeiras firmas da amostra são apresentados na Tabela 1.4 juntamente com as estimativas apresentadas por van den Broeck et al (1994)<sup>4</sup> e Greene (1990)<sup>5</sup>.

<sup>4</sup>Os resultados que se encontram na Tabela 1.4 são os mesmos apresentados em van den Broeck et al (1994), o qual foram apenas a média a posteriori e rank da firma.

<sup>5</sup>Os resultados que se encontram na Tabela 1.4 são os mesmos apresentados em Greene



Figura 1.12: Densidade a posteriori para os parâmetros:  $\theta$ ,  $P$ ,  $\sigma^2$ ,  $\beta_1$  usando o método de Slice Sampling



Observando a Tabela 1.4 e fazendo uma comparação com os resultados encontrados por Greene (1990) e Van den Broeck et al. (1994), percebe-se algumas diferenças nas estimativas da medida de eficiência. Por exemplo, observa-se que as estimativas nas eficiência das cinco firmas selecionadas utilizando o método de Slice Sampling são menores em relação com às encontradas por Greene (1990) e Van den Broeck et al. (1994). Também podemos observar que os *ranking* das cinco firma selecionadas são parecidas tanto usando o método de Slice Sampling quanto as obtidas por Van den Broeck et al. (1994)

---

(1990), o qual foram apenas a média e o desvio padrão.

Tabela 1.4: Sumário da distribuição a posteriori das eficiência de cinco firma usando o método Slice Sampling e as abordagens de van den Broeck et al (1994) e Greene (1990) para o conjunto de dados da indústria de serviço elétrico dos EUA

	Fronteira de Normal-Gama						
	Slice Sampling			van den Broeck et al 1994		Greene 1990	
Parâmetros	Média	d.p.	Rank	Média	Rank.	Média	d.p
$r_1$	0.7232	0.1121	120	0.8261	121	0.7435	0.1007
$r_2$	0.9304	0.0547	2	0.9980	3	0.9392	0.0360
$r_3$	0.8871	0.0769	48	0.9121	36	0.9058	0.0591
$r_4$	0.8724	0.0824	68	0.8967	77	0.8960	0.0654
$r_5$	0.9174	0.0634	4	0.9277	4	0.9277	0.0432

## Capítulo 2

# Seleção de Modelos de Fronteira de Produção Estocástica

### 2.1 Introdução

Do ponto de vista econômico, a necessidade de sobreviver em um ambiente competitivo na maioria das unidades econômicas induz a uma convicção que muitos deles estão perto da fronteira, isto é, eficiência completa. Porém devido a vários fatores as firmas podem estar longe da sua fronteira. Embora algumas considerações econômicas nós guiam formando alguma idéia relativo a eficiência, não se tem uma forma funcional exata para a distribuição das eficiências. Esta incerteza pode ser resolvida comparando os modelos de fronteira com alternativas distribuições nas eficiências.

Na estimação empírica dos modelos de fronteira estocástica de produção utiliza-se, com frequência, a função Coob-Douglas, dada suas características simples e de fácil estimação. No entanto, ela impõe algumas restrições como elasticidades e retornos de escala constante além de apresentar elasticidades de substituição iguais a um. Outras especificações incluem: a Translog, CES (Constant Elasticity

of Substitution) e GPF's (Generalized Production Functions).

O objetivo principal deste capítulo é comparar essas funções de produção alternativas e selecionar a distribuição das ineficiências, como por exemplo: a Gama ou a Lognormal. As comparações serão feitas através dos critérios: pseudo fator de Bayes (Geisser e Eddy, 1979) e Deviance Information Criterion (Spiegelhalter et al. , 2002).

O restante do capítulo é organizado da seguinte forma: Na segunda seção são apresentadas as funções de produção com suas propriedades. Os critérios de seleção de modelos são apresentados na terceira seção. A quarta seção apresenta uma comparação entre as funções de produção com dados reais. Uma comparação entre os modelos de fronteira estocástica com dados reais é apresentada na quinta seção.

## 2.2 Função de produção

Dá-se o nome de função de produção à relação entre as quantidades produzidas e os insumos, sendo que o conhecimento desta relação é fundamental para um planejamento empresarial. A função de produção indica o máximo de produção que se pode obter a partir de uma dada quantidade de inputs, fornecendo também um perfil das condições de operação das empresas individuais ou dos setores como um todo, gerando portanto, importantes subsídios para o processo de tomada de decisão. Assim, a função de produção descreve os esquemas de máxima produção que podem ser definidos a partir de diversas combinações dos insumos.

Desta forma, a função de produção pode ser convenientemente visualizada como a **fronteira** entre as regiões viáveis e não viáveis no espaço dos *inputs* e *outputs*.

De uma forma geral, a função de produção é o lugar geométrico de todas as possíveis combinações eficientes dos diversos insumos e das diversas variáveis

tecnológicas e operacionais de um determinado sistema produtivo.

Tradicionalmente, a função de produção pode ser representada por:

$$y = f(x_1, x_2) \quad (2.1)$$

onde  $y$  representa o output, e  $x_1$  e  $x_2$  são os inputs.

### 2.2.1 Produto Marginal

Um elemento importante associado à função de produção num ponto, é o produto marginal (PM). O produto marginal mede a variação de outputs quando o valor de um dos inputs é alterado. O produto marginal está associado às derivadas parciais da função de produção, se elas existirem. Suponha que uma função de produção com um único output  $y$  e um input qualquer  $x_j$ . Define-se como produto marginal em relação ao input  $j$ , à relação:

$$MP_j = \frac{\partial y}{\partial x_j} \quad (2.2)$$

A produção marginal está ligada ao comportamento da variação da produção quando varia-se um dos inputs de cada vez, mantendo constante os demais, ou seja, aumentando-se a quantidade de um insumo e permanecendo a quantidade dos demais fatores fixa, a produção aumentará inicialmente à taxas crescentes (produto marginal positivo); a seguir, num certo ponto ocorrerão taxas decrescentes (produto marginal negativo).

Como exemplo, supõe-se que numa indústria o input fixo é representado pelo número de máquinas. O fator variável é representado pelo número de operários. Se várias combinações de máquinas e mão-de-obra forem utilizadas para produzir um certo produto e se a quantidade de máquinas for constante, o aumento da produção dependerá do aumento da mão-de-obra utilizada. Quando isso ocorrer, se alterarão as proporções de combinações entre fatores fixos e variáveis. Nesse

caso a produção aumentará até certo ponto e depois decrescerá. Isto quer dizer que de início poderão ocorrer rendimentos crescentes, enquanto os acréscimos de utilização do fator variável provocarem incrementos na produção. Todavia, essa fase, quando ocorre, é passageira, passando a existir logo em seguida rendimentos decrescentes.

## 2.2.2 Retorno de Escala

Outra forma de analisar a função de produção é observar a taxa de variação do output, quando todos os inputs variam nas mesmas proporções. O retorno de escala está ligado ao incremento obtido na produção quando todos os inputs são igualmente incrementados.

Retorno de escala:  $\Delta y = f[(1 + \Delta)x_1, \dots, (1 + \Delta)x_n] - f(x_1, \dots, x_n)$

Uma forma mais prática de analisar o retorno de escala é através da seguinte expressão:

$$y = f(\lambda x_1, \lambda x_2, \dots, \lambda x_n) = \lambda^k f(x_1, x_2, \dots, x_n) \quad (2.3)$$

Uma função de produção é dita homogênea, se puder ser expressa na forma indicada em 2.3. Por exemplo, a função de produção

$$y = f(x_1, x_2) = [a_1 x_1^{-b} + a_2 x_2^{-b}]^{-1/c} \quad (2.4)$$

é homogênea pois,

$$\begin{aligned} y = f(\lambda x_1, \lambda x_2) &= [a_1 (\lambda x_1)^{-b} + a_2 (\lambda x_2)^{-b}]^{-1/c} = \\ &= \lambda^{b/c} [a_1 x_1^{-b} + a_2 x_2^{-b}]^{-1/c} = \lambda^{b/c} f(x_1, x_2) \end{aligned} \quad (2.5)$$

Ao passo que a função de produção,

$$y = f(x_1, x_2) = [a_1 x_1^{-b_1} + a_2 x_2^{-b_2}]^{-1/c} \quad (2.6)$$

com  $b_1 \neq b_2$  não é homogênea.

De acordo com a variação da quantidade produzida em função da variação da quantidade utilizada dos inputs é possível identificar três tipos de ganhos de escala: (1) retornos crescentes de escala; (2) retornos constantes de escala e (3) retornos decrescentes de escala. O primeiro ocorre quando  $k > 1$ , o segundo quando  $k = 1$  e o terceiro quando  $k < 1$ .

Os retornos crescentes de escala ocorrem quando a variação na quantidade do produto final é proporcionalmente maior que a variação da quantidade utilizada dos inputs. Por exemplo, aumentando-se a utilização dos inputs em 10%, o output cresce em 20%. Os retornos constantes de escala ocorrem quando a variação total do produto é igual à variação da quantidade utilizada dos insumos. Por exemplo, aumentando-se a utilização dos fatores em 10%, o produto total também aumenta em 10%.

Finalmente, os retornos decrescentes de escala ocorrem quando a variação da produção é menor que a variação proporcional dos fatores. Por exemplo, aumentando-se a utilização dos fatores em 10%, o produto decresce em 5%.

### 2.2.3 Taxa Marginal de Substituição

Na análise de produção, é muito importante o conhecimento das tendências de variação de uma variável em função das outras. Em particular, é interessante analisar tais variações de forma adimensional. Assim, define-se a elasticidade da produção  $y$  em função do input  $x_i$ , num certo ponto  $x_i^A$ , através da relação:

$$\epsilon_i = \frac{\frac{\Delta y}{y}}{\frac{\Delta x_i}{x_i}} \longrightarrow \frac{x_i}{y} \vartheta y$$
 (2.7)

Suponha que deseja-se medir a variação no input  $x_i$  em função da variação no input  $x_j$ , mantendo-se o nível de produção fixo e os demais inputs permanecendo inalterados.

Assim, é possível escrever

$$\Delta y = \frac{\partial y}{\partial x_1} \Delta x_1 + \frac{\partial y}{\partial x_2} \Delta x_2 + \dots + \frac{\partial y}{\partial x_i} \Delta x_i + \dots + \frac{\partial y}{\partial x_j} \Delta x_j + \dots \quad (2.8)$$

Sendo fixos os inputs diferentes de  $i$  e  $j$ , as derivadas parciais  $\frac{\partial y}{\partial x_k}$ , com  $k \neq i$  e  $k \neq j$ , serão nulas. Similarmente, se  $y$  é mantido invariável, então  $\Delta y = 0$ .

Assim, a relação (2.8) se simplifica em:

$$\frac{\partial y}{\partial x_i} \Delta x_i + \frac{\partial y}{\partial x_j} \Delta x_j = 0 \quad (2.9)$$

e

$$TMS = -\frac{\Delta x_j}{\Delta x_i} = \frac{MP_i}{MP_j} \quad (2.10)$$

A relação (2.10) é denominada taxa marginal de substituição do insumo  $i$  pelo insumo  $j$ .

## 2.2.4 Elasticidade de Substituição

Um outro conceito é a de "elasticidade de substituição" que mede a taxa de mudança da taxa marginal de substituição, ou seja, é uma medida de quão boa foi substituição de um input por outro.

$$\varepsilon = \frac{TMS(x_1 TMS + x_2)}{x_2 x_1 \left( TMS \frac{\partial TMS}{\partial x_2} - \frac{\partial TMS}{\partial x_1} \right)} \quad (2.11)$$

onde  $0 \leq \varepsilon < \infty$ . Sendo que quanto maior o valor de  $\varepsilon$ , melhor será a substituição de um input pelo outro. Se  $\varepsilon = 0$ , então a substituição não é possível. Já quando  $\varepsilon = \infty$ , os inputs são substitutos perfeitos.

## 2.2.5 Principais Tipos de Funções de Produção

A seguir será descrita uma classe de funções de produção que apresentam a elasticidade de substituição constante: a função Cobb-Douglas e a função CES (Constant Elasticity of Substitution). Também será apresentada a função Translog, que



é uma função que não impõe restrições. Por último será descrita a função GPF, que possui uma generalização em duas direções: na elasticidade de substituição e no retorno de escala.

### Função Cobb-Douglas

A função de produção Coob-Douglas, dada suas características simples e de fácil estimação, tem sido bastante usado em análise microeconômica. No entanto, ela impõe algumas restrições como elasticidades e retornos de escala constante além de apresentar elasticidades de substituição iguais a um.

Em geral, a forma desta função para um output  $Q$ , que está relacionado aos dois inputs: capital e trabalho ( $K$  e  $L$ ), é expressa por:

$$Q = \gamma K^{\beta_1} L^{\beta_2} \quad (2.12)$$

ou, também pode ser escrita como

$$\log Q = \log \gamma + \beta_1 \log K + \beta_2 \log L \quad (2.13)$$

onde  $\gamma$ ,  $\beta_1$  e  $\beta_2$  são parâmetros tais que  $\beta_1 > 0$  e  $\beta_2 > 0$

As produções marginais de  $KeL$  são

$$MP_K = \frac{\beta_1 y}{L} \quad e \quad MP_L = \frac{\beta_2 y}{K}$$

Logo, pela Eq.(2.10) a taxa marginal de substituição é expressa por

$$\tau = \frac{MP_L}{MP_K} = \frac{\beta_2 L}{\beta_1 K}$$

e pela Eq. (2.11) a elasticidade de substituição é expressa por

$$\varepsilon = \frac{\tau(K\tau + L)}{LK \left( r \frac{\partial \tau}{\partial L} - \frac{\partial \tau}{\partial K} \right)} = 1 \quad (2.14)$$

onde  $\frac{\partial \tau}{\partial L} = \frac{\beta_2}{\beta_1 K}$  e  $\frac{\partial \tau}{\partial K} = -\frac{\beta_2 L}{\beta_1 K^2}$

Para mais detalhes sobre as propriedades da função de produção Cobb-Douglas, ver Wynn e Holden(1974).

### Função C.E.S (Constant Elasticity of Substitution)

Como a função de produção de Cobb-Douglas é uma função bastante restritiva, pois possui a elasticidade de substituição igual a 1 em qualquer ponto da função, uma generalização é permitir que elasticidade de substituição seja constante, mas diferente de 1. A forma da função de produção C.E.S para dois *inputs* é:

$$Q = \gamma[(1 - \delta)K^{-\rho} + \delta L^{-\rho}]^{-v/\rho} \quad (2.15)$$

ou

$$\log Q = \log \gamma + v \log\{[(1 - \delta)K^{-\rho} + \delta L^{-\rho}]^{-1/\rho}\} \quad (2.16)$$

onde  $\gamma$ ,  $v$ ,  $\delta$  e  $\rho$  são parâmetros tais que  $0 < \gamma < \infty$ ,  $0 < \delta < 1$ ,  $-\infty < v < \infty$  e  $-1 < \rho < \infty$ .

Fazendo uma expansão de Taylor em torno de  $\rho = 0$  tem-se:

$$\begin{aligned} y &= \log \gamma + v(1 - \delta) \log K + v\delta \log L + \\ &\quad \rho v \delta (1 - \delta) \left\{ -\frac{1}{2} \log^2(K/L) \right\} \\ &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \end{aligned} \quad (2.17)$$

onde  $y = \log Q$ ,  $x_1 = 1$ ,  $x_2 = \log K$ ,  $x_3 = \log L$ ,  $x_4 = -1/2 \log^2(K/L)$ ,  $\beta_1 = \log \gamma$ ,  $\beta_2 = v(1 - \delta)$ ,  $\beta_3 = v\delta$  e  $\beta_4 = \rho v \delta (1 - \delta)$ .

Portanto, se  $\rho = 0$ , a função CES recai na forma da função Cobb-Douglas.

A produção marginal de K e L são expressas por

$$MP_K = \frac{(1 - \delta)}{K^{1+\rho}} A y^{1+\rho/v} \quad e \quad MP_L = \frac{\delta}{L^{1+\rho}} A y^{1+\rho/v}$$

onde  $A = \frac{v}{\gamma^{\rho/v}}$

Logo, pela Eq.(2.10) a taxa marginal de substituição é dada por

$$\tau = \frac{MP_L}{MP_K} = \left( \frac{\delta}{1 - \delta} \right) \left( \frac{K}{L} \right)^{1+\rho} \quad (2.18)$$

e pela Eq.(2.11) a elasticidade de substituição é dada por

$$\varepsilon = \frac{\tau(L\tau + K)}{KL \left( \tau \frac{\partial \tau}{\partial K} - \frac{\partial \tau}{\partial L} \right)} = \frac{1}{1 + \rho} \quad (2.19)$$

onde  $\frac{\partial \tau}{\partial K} = \frac{(1+\rho)R}{K}$  e  $\frac{\partial \tau}{\partial L} = -\frac{(1+\rho)R}{L}$

### Função Translog

Muitas vezes, a formulação Cobb-Douglas não é satisfatória para representar um certo processo produtivo por impor algumas restrições, como retornos de escala constante e elasticidades de substituição iguais a um. Nestes casos, é comum adotar-se uma formulação mais flexível, ajustando uma função do tipo translogarítmica, ou abreviadamente translog, a saber:

$$\log Q = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \left( 1/2x_1^2 \right) + \beta_4 \left( 1/2x_2^2 \right) + \beta_5 x_1 x_2 \quad (2.20)$$

onde  $x_1 = \log L$ ,  $x_2 = \log K$ ,  $\beta_0 = \log A$  e  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$  e  $\beta_5$  são parâmetros tais que  $0 < \beta_0 < 1$ ,  $\beta_1 > 0$ ,  $\beta_2 > 0$ ,  $\beta_3 > 0$ ,  $\beta_4 > 0$  e  $\beta_5 > 0$ .

Observa-se na Eq. 2.20, que se  $\beta_3 = \beta_4 = \beta_5 = 0$ , tem-se a forma da função Cobb-Douglas. A função de produção Translog serve como uma aproximação local de segunda ordem para qualquer função de produção.

### Função GPF's (Generalized Production Functions)

Zellner e Revankar (1970) propõem a função de produção Generalizada (GPF's). Esta função foi introduzida com o objetivo de permitir uma generalização em duas direções. Primeiro, permite que a elasticidade de substituição possa ser constante ou variável. A segunda generalização é permitir que os retornos de escala variem com o nível do output. A seguir será brevemente descrito, a forma descrita em Zellner e Revankar (1970) e Zellner (1971) para obter a função produção Generalizada .

Em termos determinísticos, considere a seguinte equação diferencial:

$$\frac{dQ}{df} = \frac{Q}{f} \frac{\alpha(Q)}{\alpha_f} \quad (2.21)$$

com solução

$$Q = g(f) \quad (2.22)$$

onde  $\alpha(Q)$  é o retorno de escala, que é função do output  $Q$ ,  $f(L, K)$  é qualquer função de produção onde  $L$  e  $K$  são os inputs capital e trabalho, respectivamente,  $\alpha_f$  é parâmetro do retorno de escala associado a função de produção  $f$  e  $g$  é uma transformação de  $f$ . A elasticidade de substituição associada a  $Q = g(f)$  é a mesma relacionada a função  $f$ .

Uma particular função GPF é determinada quando  $\alpha(Q)$  tem a seguinte forma:

$$\alpha(Q) = \frac{\alpha}{1 + \lambda Q} \quad (2.23)$$

onde  $\lambda > 0$  e  $\alpha > 0$ . Inserindo  $\alpha(Q)$ , dado em (2.23), em (2.21), o resultado da equação diferencial é

$$\frac{dQ}{df} = \frac{Q}{f} \frac{\alpha}{\alpha_f(1 + \lambda Q)}, \quad (2.24)$$

com solução  $Qe^{\lambda Q} = C f^{\alpha/\alpha_f}$ , onde  $C$  é uma constante de integração. Se a função  $f$  tem a forma da função de produção Cobb-Douglas, então  $f = AK^{\alpha_f(1-\delta)}L^{\alpha_f\delta}$ . Desta maneira, a forma da função de produção Generalizada é expressa por:

$$Qe^{\lambda Q} = \gamma K^{\alpha(1-\delta)}L^{\alpha\delta} \quad (2.25)$$

com  $\gamma = CA$ ,  $0 < \delta < 1$  e  $\gamma$ . Ou tomando o logaritmo em ambos os lados, tem-se:

$$z = y + \lambda \exp(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \quad (2.26)$$

onde  $y = \log Q$ ,  $x_1 = \log K$ ,  $x_2 = \log L$ ,  $\beta_1 = \log \gamma$ ,  $\beta_2 = \alpha(1 - \delta)$  e  $\beta_3 = \alpha\delta$

Pela Eq. (2.26) observa-se que, se  $\theta = 0$ , a função GPF's recai na forma da função Cobb-Douglas.

## 2.3 Critérios de Seleção de Modelo

Esta seção introduz duas medidas de diagnósticos que foram utilizados para medir o desempenho dos diferentes modelos estudados.

A escolha de modelos é uma atividade fundamental que vem tornando-se cada vez mais importante nas análises estatísticas, uma vez que, devido aos avanços computacionais, é possível construir modelos cada vez mais complexos. Tal complexidade normalmente aumenta de acordo com a estrutura imposta pelos modelos que requerem especificações em cada um de seus níveis.

A seguir serão apresentados os critérios de Fator de Bayes e o de DIC (Deviance Information Criterion) para a seleção de modelos.

### 2.3.1 Fator de Bayes

Uma forma de comparar modelos utilizando a teoria Bayesiana, baseia-se no fator de Bayes. Assim, para dois modelos  $M_1$  e  $M_2$ , define-se uma probabilidade a priori para cada modelo  $p(M_1)$  e  $p(M_2)$ . A probabilidade a posteriori desses modelos satisfaz:

$$\frac{p(M_2|\underline{y})}{p(M_1|\underline{y})} = \frac{f(\underline{y}|M_2)p(M_2)}{f(\underline{y}|M_1)p(M_1)} \quad (2.27)$$

onde  $f(\underline{y}|M_i)$  é a verossimilhança preditiva do modelo  $M_i$ ,  $i = 1, 2$ . Assim, a verossimilhança preditiva é dada por

$$f(\underline{y}|M_i) = \int f(\underline{y}|\theta_i M_i) f(\theta_i|M_i) d\theta_i \quad (2.28)$$

A razão  $f(\underline{y}|M_2)/f(\underline{y}|M_1)$  de verossimilhanças preditivas é chamada de fator de Bayes. Quando a razão de chances a priori for 1 indicando indiferença a priori, a razão chance a posteriori coincide com o fator de Bayes. Assim, o fator de Bayes fornece uma medida relativa de avaliação de um modelo. Para mais detalhes sobre o fator de Bayes, veja em Kass e Raftery (1995).

O cálculo de  $f(\underline{y}|M_i)$  não pode ser realizado analiticamente devido á complexidade do integrando em (2.28), ou pode ser em alguns casos indefinido. A seguir, será apresentado o método de cálculo aproximado de (2.28) usando essencialmente técnica de simulação.

Uma densidade preditiva que foi utilizada extensivamente em trabalhos recentes de Gelfand (1995) e Gelfand, Dey e Chang (1992) foi a densidade preditiva de validação cruzada  $f(y_i|y_{-i})$ <sup>1</sup> (Stone, 1974) Geisser e Eddy (1979) sugeriram que o produto  $\prod_{i=1}^n f(y_i|y_{-i})$  dessas densidades fosse utilizado como um indicador relativo do valor da verossimilhana preditiva  $f(y)$  através do pseudo fator de Bayes

$$\frac{\prod_{i=1}^n f(y_i|y_{-i}, M_2)}{\prod_{i=1}^n f(y_i|y_{-i}, M_1)} \quad (2.29)$$

que aproximaria o fator de Bayes. Outro estimadores de  $f(y)$  pode ser visto em Gamerman, (1997).

A implementação de uma aproximação da densidade preditiva de validação cruzada  $f(y_i|y_{-i})$  é apresentado na seção 2.4.1

### 2.3.2 Deviance Information Criterion (DIC)

Do ponto de vista frequentista, a avaliação do modelo é baseado na *deviance*. Analogamente, Dempster (1974) sugere examinar a distribuição a posteriori da deviance clássica definido por:

$$D(\theta) = -2 \ln f(y|\theta) + 2 \ln g(y) \quad (2.30)$$

onde  $f(y|\theta)$  é a função de verossimilhança e  $\ln g(y)$  é um termo que só depende dos dados. Dempster (1974) propõe comparar a média à posteriori de  $D(\theta)$  e Spiegelhalter et al. (2002) segue essa sugestão no desenvolvimento do DIC como um critério de escolha de modelos. Baseado na distribuição à posteriori de  $D(\theta)$ ,

---

<sup>1</sup> $y_{-i}$  é o restante dos dados excluindo  $y_i$ .

o DIC consiste em duas componentes: um termo que mede a bondade do ajuste e outro termo de penalidade para o crescimento da complexidade do modelo.

$$DIC = \bar{D} + P_D \quad (2.31)$$

1. O primeiro termo é uma medida Bayesiana de ajuste do modelo, que é definido como a média a posteriori da *deviance*

$$\bar{D} = E_{\theta|y}[D] = E_{\theta|y}[-2 \ln f(y|\theta)]. \quad (2.32)$$

2. A segunda componente mede a complexidade do modelo através do *número efetivo de parâmetros*  $p_D$ , definida como a diferença entre a média a posteriori da deviance e a deviance avaliada em  $\bar{\theta}$ , que é a média a posteriori do parâmetro  $\theta$ :

$$P_D = \bar{D} - D(\bar{\theta}) = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]). \quad (2.33)$$

Através da Eq.(2.33), pode-se obter  $\bar{D} = D(\bar{\theta}) + p_D$ . Então, o DIC definido em (2.31) pode ser reescrito da forma:

$$DIC = D(\bar{\theta}) + 2P_D \quad (2.34)$$

## 2.4 Comparação entre as Funções de Produção

Antes de realizar uma comparação entre os modelos de fronteira de produção estocástica, faz-se uma comparação entre as funções de produção, ou seja, não considera-se o termo de ineficiência ( $u_i$ ). Assim, os modelos ajustados tem a seguinte forma:

$$y_i = f(x_i, \beta) + v_i \quad i = 1, \dots, N \quad (2.35)$$

onde  $y_i$  é usualmente o logaritmo do output,  $x_i$  denota o logaritmo do vetor de inputs,  $f(\cdot, \cdot)$  é a função de produção determinada,  $v_i$  é assumido que tem distribuição normal com media 0 e variância  $\sigma^2$  representando flutuações aleatórias que não dependem da firma.

Desta forma, os quatro modelos ajustados foram:

MODELO 1: Modelo usando a função de produção Cobb-Douglas.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + v_i \quad (2.36)$$

com  $v_i \sim N(0, \sigma^2)$

MODELO 2: O modelo com a função de produção CES.

$$y_i = \beta_0 + v \log\{[(1 - \delta) \exp(x_1)^{-\rho} + \delta \exp(x_2)^{-\rho}]^{-1/\rho}\} + v_i \quad (2.37)$$

com  $v_i \sim N(0, \sigma^2)$

MODELO 3: usando a função de produção Translog.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \frac{1}{2} \beta_3 x_1^2 + \frac{1}{2} \beta_4 x_2^2 + \beta_5 x_1 x_2 + v_i \quad (2.38)$$

com  $v_i \sim N(0, \sigma^2)$

MODELO 4: Modelo utilizando a função de produção GPF's.

$$z_i = y_i + \theta \exp(y_i) = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + v_i \quad (2.39)$$

com  $v_i \sim N(0, \sigma^2)$ .

A única função de produção que não tem uma relação direta entre  $Y$  e  $X$  é a função GPF. Assim, para completar o modelo é necessário definir a transformação Jacobiana de  $Z$  para  $Y$ , dada por

$$J = \prod_{i=1}^N \frac{\partial z_i}{\partial y_i} = \prod_{i=1}^N (\lambda + 1/y_i) \quad (2.40)$$



Então a função de verossimilhança do modelo 4 é:

$$f(y_i|\beta, \sigma^2, \lambda) = f_N(z|x'_i\beta, \sigma^2) | J_i | \quad (2.41)$$

onde  $z = y + \lambda \exp(y)$  e  $J_i = \lambda + 1/y_i$

### 2.4.1 Implementação no WinBUGS

Todos os modelos usados neste capítulo foram ajustados dentro do pacote BUGS (Bayesian Inference Using Gibbs Sampling). O uso do BUGS é atrativo devido ao fato de ser fácil a implementação dos modelos.

Como foi dito anteriormente, o modelo utilizando a função de produção GPF não tem uma relação direta entre  $y$  e  $x$ , logo, para implementar esta distribuição amostral que não esta incluída na lista de distribuições do WinBUGS. A seguir será apresentado uma exemplo para ilustrar a implementação de uma densidade.

Suponha que  $y$  tenha distribuição Normal com média  $\mu$  e variância  $\sigma^2$ , logo a função de verossimilhança de  $y$  é

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right) \quad (2.42)$$

E seja  $\alpha \sim Poisson(\theta)$  com  $\alpha = 0$  a verossimilhança de  $\alpha$  será

$$P(\alpha = 0) = \exp(-\theta) \quad (2.43)$$

Assim, o "truque" é definir  $\theta = -\log(f(y))$ , ou seja,

$$\theta = \frac{1}{2\sigma^2}(y - \mu)^2 + \log(\sqrt{2\pi\sigma^2}) \quad (2.44)$$

Desta forma temos

$$P(\alpha = 0) = f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y - \mu)^2\right) \quad (2.45)$$

Portanto com estes procedimentos, podemos escrever a função de verossimilhança de  $y$  de outra forma no WinBUGS. O código no WinBUGS seria descrito

desta forma:

```
C <- 10000 # para garantir que theta>0
for (i in 1:N) {
alpha[i] <- 0
theta[i] <- log(sigma[1]) + 0.5 * pow((y[i] - mu[1]) / sigma[1], 2) + C
alpha[i] ~ dpois(theta[i])
}
```

Para mais detalhes, ver o manual de WinBUGS versão 1.4 (Spiegelhalter et al., 1999).

As estatísticas DIC são calculadas automaticamente no WinBUGS 1.4. Apenas com o MODELO 4, que tem a função de verossimilhança descrita anteriormente, o valor da estatística DIC foi calculada fora do WinBUGS.

Infelizmente, o cálculo exato densidade preditiva de validação cruzada  $p(y_i|y_{-i})$  não pode ser realizado no WinBUGS. Isso porque no WinBUGS não é possível separar os dados para realizar o cálculo. Porém, é possível calcular a preditiva ordenada  $p(y_i)$ . Gelfand e Dey (1994) apresentam o seguinte fato interessante:

$$\frac{1}{p(y_i|y_{-i})} = \int \frac{1}{p(y_i|y_{-i}, \theta)} p(\theta|\underline{y}) d\theta \quad (2.46)$$

e conseqüentemente a estimativa de Monte Carlo de  $p(y_i|y_{-i})$  é obtido como uma média harmônica de  $p(y_i|y_{-i}, \theta)$ .

A idéia em ?? pode ser implementado no WinBUGS, para obter-se a estimativa de  $p(y_i)$ . Para isso é necessário construir uma variável que avalie o valor da inversa da verossimilhança a cada iteração, ou seja,  $p.inv_i = p(y_i|\theta)^{-1}$ . Havendo completado de rodar todas as iterações, a estimativa preditiva ordenada  $p(y_i)$  é realizado tomando a inversa da média a posteriori de  $p.inv_i$ . Mais detalhes pode ser visto no manual do BUGS versão 0.5 (Thomas et al.,1996)

A seguir, são descritas as distribuições a priori assumidas para os quatro modelos.

## 2.4.2 Distribuição a Priori

As distribuições a priori definidas para os quatro modelos foram:

- Modelo 1:  $\beta_j \sim N(b_0, H_0^{-1})$  para  $j = 1, \dots, 3$  e  $\sigma^2 \sim GI(n_0/2, a_0/2)$
- Modelo 2:  $\beta_0 \sim N(b_0, H_0^{-1})$ ,  $v \sim N(m_0, s_0^{-1})$ ,  $\delta \sim \text{Beta}(w_0, \varphi_0)$ ,  $\sigma^2 \sim GI(n_0/2, a_0/2)$  e  $\rho \sim NT_{[-1, \infty]}(r_0, g_0)$
- Modelo 3:  $\beta_j \sim N(b_0, H_0^{-1})$  para  $j = 1, \dots, 5$  e  $\sigma^2 \sim GI(n_0/2, a_0/2)$
- Modelo 4:  $\beta_j \sim N(b_0, H_0^{-1})$  para  $j = 1, \dots, 3$ ,  $\sigma^2 \sim GI(n_0/2, a_0/2)$  e  $\lambda \sim U(0, 1)$

Os valores escolhidos dos hiperparâmetros que levaram a uma distribuição a priori não informativa foram:

$$\begin{bmatrix} b_0 & H_0^{-1} \\ m_0 & s_0^{-1} \\ w_0 & \varphi_0 \end{bmatrix} = \begin{bmatrix} 0 & 0.001 \\ 0 & 0.001 \\ 1 & 1 \end{bmatrix}$$

Como a distribuição a priori de  $\rho$  do modelo 2 tem distribuição Normal Truncada, logo não faz parte da lista de distribuições do WinBUGS. Assim, a implementação desta priori no WinBUGS é realizada da mesma forma mencionada na seção 2.4.1.

## 2.4.3 Análise MCMC

Para aplicar MCMC, as distribuições condicionais completas devem ser avaliadas. As distribuições condicionais completas nos quatro modelos estão brevemente descritas a seguir:

- Modelo 1:

$$- (\beta|Y, X, resto) \sim N^k \left[ (H_0 + \sigma^{-2} X' X)^{-1} (H_0 b_0 + X' Y \sigma^{-2}), H_0 + \sigma^{-2} X' X \right]$$

$$- (\sigma^{-2}|Y, X, resto) \sim G \left[ \frac{(N+N_0)}{2}, \frac{((y-X\beta)'(y-X\beta)+a_0)}{2} \right]$$

- Modelo 2:

$$- (\beta_0|Y, X, resto) \sim N \left[ (H_0 + \sigma^{-2})^{-1} (H_0 b_0 + \mathbf{1}' Y \sigma^{-2}), H_0 + \sigma^{-2} \right]$$

$$- (\sigma^{-2}|Y, X, resto) \sim G \left[ \frac{(N+N_0)}{2}, \frac{((y-X^*\pi)'(y-X^*\pi)+a_0)}{2} \right]$$

$$- p(\delta|Y, X, resto) \propto \exp \left[ -\frac{(y-X^*\pi)'(y-X^*\pi)}{2\sigma^2} \right]$$

$$- p(v|Y, X, resto) \propto \exp \left[ -\frac{(y-X^*\pi)'(y-X^*\pi)}{2\sigma^2} - \frac{s_0(v-m_0)^2}{2} \right]$$

$$- (\rho|resto) \propto \exp \left[ -\frac{(y-X^*\pi)'(y-X^*\pi)}{2\sigma^2} - \frac{(\rho-\tau_0)^2}{2g_0} \right] I_{(-1,\infty)}(\rho)$$

$$\text{onde } X^* = \left[ 1 \quad \log\{[(1-\delta)\exp(X_1)^{-\rho} + \delta\exp(X_2)^{-\rho}]^{-1/\rho}\} \right] \text{ e}$$

$$\pi = (\beta_0 \quad v)'$$

Como a função de produção CES tem uma forma complexa, a distribuição condicional completa de  $\delta$ ,  $v$  e  $\rho$  não tem uma forma fechada. Assim, o método utilizado pelo WinBUGS para gerar uma amostra de  $\delta$  e  $v$  é o algoritmo de Slice Sampling. Como mencionado anteriormente, a distribuição a priori de  $\rho$  não faz parte da lista de distribuições do WinBUGS, por este motivo, o método utilizado pelo WinBUGS para gerar uma amostra de  $\rho$  é o algoritmo de Metropolis-Hastings.

- Modelo 3:

$$- (\beta|Y, X, resto) \sim N^k \left[ (X^{*'} X^*)^{-1} (X^{*'} y), \sigma^2 (X^{*'} X^*)^{-1} \right]$$

$$- (\sigma^{-2}|Y, X, resto) \sim G \left[ \frac{(N+N_0)}{2}, \frac{((y-X^*\beta)'(y-X^*\beta)+a_0)}{2} \right]$$

onde <sup>2</sup>  $X^* = \begin{bmatrix} 1 & X_1 & X_2 & X_3 & 1/2X_2 & 1/2X_3 & X_2X_3 \end{bmatrix}$

- Modelo 4:

$$- (\beta|Y, X, resto) \sim N^k \left[ (H_0 + \sigma^{-2}X'X)^{-1}(H_0b_0 + X'Z\sigma^{-2}), H_0 + \sigma^{-2}X'X \right]$$

$$- (\sigma^{-2}|Y, X, resto) \sim G \left[ \frac{(N+N_0)}{2}, \frac{((Z-X\beta)'(Z-X\beta)+a_0)}{2} \right]$$

$$- p(\lambda|resto) \propto \exp \left[ \frac{1}{2\sigma^2}(Z - X\beta)'(Z - X\beta) + \sum_{i=1}^N \log(y_i\lambda + 1) \right]$$

onde  $Z = Y + \lambda \exp Y$

Embora as distribuições condicionais de  $\beta$  e  $\sigma^2$  estejam disponíveis na forma fechada, os métodos utilizado pelo WinBUGS para amostrar uma amostra de  $\beta$  e  $\sigma^2$  foram os algoritmos de Metropolis-Hastings e Slice Sampling respectivamente. Isto porque a função de verossimilhança foi construída da forma mencionada na seção 2.4.1. Como a distribuição condicional de  $\lambda$  não apresenta uma forma conhecida, o WinBUGS utiliza o algoritmo de Slice Sampling para gerar uma amostra de  $\lambda$ .

## 2.4.4 Resultados

Nesta aplicação, foram utilizados os dados usados por Zellner e Revankar (1970) e Zellner (1971) para estudar a produção da indústria de equipamentos de transportes de 25 estados, nos Estados Unidos, no ano de 1957. Estes dados estão listados no apêndice do artigo de Zellner e Revankar (1970). Nestes dados foram definidas as seguintes variáveis:  $y = \log(\text{produção}/m)$ ,  $x_1 = \log(\text{capital}/m)$  e  $x_2 = \log(\text{trabalho}/m)$  onde  $m$  é o número de estabelecimentos em cada estado.

Os resultados para os MODELOS 1-4 são baseados em 10.000 iterações. Foram realizadas 20.000 iterações das quais foram descartadas as 10.000 primeiras para

---

<sup>2</sup> $X_i$  é a coluna  $i$  da matriz  $X$

o período de "burn-in". A convergência foi checada através do critério de Gelman e Rubin (1992a) que está implementado no WinBUGS.

Na Tabela 2.1 são apresentados os resultados da estatística DIC juntamente com  $\bar{D}$  e  $p_D$  para cada um dos quatro modelos.

Tabela 2.1: Resultado da Deviance Information Criterion para os quatro modelos de produção

Modelo	$\bar{D}$	$D(\theta)$	$p_D$	DIC
MODELO 1- Cobb-Douglas	-11.284	-15.411	4.127	-7.157
MODELO 2- CES	-11.472	-15.039	3.567	-7.905
MODELO 3- Translog	-15.411	-22.770	7.360	-8.051
MODELO 4- GPF	-15.2	-19.894	4.694	-10.506

Na Tabela 2.1 observa-se que o modelo mais adequado para descrever os dados baseando-se no DIC é o modelo usando a função de produção Generalizada (MODELO 4), seguido do modelo com a função CES (MODELO 3). Como os modelos analisados não têm o mesmo número de parâmetros, o uso do critério de seleção DIC é adequado para estes modelos.

A Tabela 2.2 apresenta os valores aproximados do pseudo fator de Bayes confrontando os quatro modelos.

Tabela 2.2: Sumário do pseudo fator de Bayes para os quatro modelos de produção

Fator de Bayes	MODELO 1 Cobb-Douglas	MODELO 2 CES	MODELO 3 Translog
MODELO 2 - CES	0.8662		
MODELO 3 - Translog	1.0467	0.8275	
MODELO 4 - GPF	5.1069	5.8958	4.8789

Na Tabela 2.2 observa-se que o valor aproximado do pseudo fator de Bayes do MODELO 4 versus todos os outros modelos, indica uma preferência pelo modelo usando a função de produção Generalizada (MODELO 4).

## 2.5 Comparação entre os Modelos de Fronteira de Produção

A análise seguinte apresenta uma comparação entre os modelos de fronteira de produção com diferentes formas funcionais, assumindo diferentes distribuições para o termo de ineficiência.

Inicialmente, os modelos de fronteira estocástica foram propostos por Aigner, Lovell e Schmidt (1977) e Meeusen e Van Den Broecker (1977). Em sua forma geral tem-se:

$$y_i = f(x_i, \beta) - u_i + v_i \quad i = 1, \dots, N \quad (2.47)$$

onde  $y_i$  é, usualmente, o logaritmo do output,  $x_i$  denota o logaritmo do vetor de inputs,  $f(\cdot, \cdot)$  é a função de produção determinada,  $v_i$  assume-se ter distribuição normal com media 0 e variância  $\sigma^2$ , representando a flutuação aleatória, que não depende da firma e  $u_i$  representando a eficiência da firma e assume-se ter uma distribuição assimétrica. As distribuições assumidas para a componente de ineficiência foram as distribuições Gama e a Lognormal. A distribuição Lognormal com um (com o parâmetro de locação igual a zero) ou dois parâmetros pode ter uma forma similar a Gama através da escolha conveniente dos parâmetros. Para mais detalhes sobre as equivalências entre a distribuição Lognormal com a Gama ver Migon(2002). Na aplicação será apresentada na seção 2.5.3, assumiu-se que a componente de ineficiência tem distribuição Lognormal com o parâmetro de locação igual a zero.

Assim, a estrutura dos oito modelos que serão comparados são:

MODELO 1: Modelo usando a função de produção Cobb-Douglas considerando a distribuição do termo de ineficiência Gama.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 - u_i + v_i \quad (2.48)$$

com  $v_i \sim N(0, \sigma^2)$  e  $u_i \sim G(P, \theta)$ .

MODELO 2: Modelo com a função de produção CES e o termo de ineficiência com distribuição Gama.

$$y_i = \beta_0 + v \log\{[(1 - \delta) \exp(x_1)^{-\rho} + \delta \exp(x_2)^{-\rho}]^{-1/\rho}\} - u_i + v_i \quad (2.49)$$

com  $v_i \sim N(0, \sigma^2)$  e  $u_i \sim G(P, \theta)$ .

MODELO 3: Usando a função de produção Translog e assumindo-se que o termo de ineficiência tenha distribuição Gama.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \frac{1}{2} \beta_3 x_1^2 + \frac{1}{2} \beta_4 x_2^2 + \beta_5 x_1 x_2 - u_i + v_i \quad (2.50)$$

com  $v_i \sim N(0, \sigma^2)$  e  $u_i \sim G(P, \theta)$ .

MODELO 4: Modelo utilizando a função de produção GPF com distribuição Gama no termo de ineficiência.

$$z_i = y_i + \lambda \exp(y_i) = \beta_0 + \beta_1 x_1 + \beta_1 x_2 - u_i + v_i \quad (2.51)$$

com  $v_i \sim N(0, \sigma^2)$  e  $u_i \sim G(P, \theta)$ .

MODELO 5: Modelo usando a função de produção Cobb-Douglas e assumindo-se que o termo de ineficiência tenha distribuição Lognormal.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 - u_i + v_i \quad (2.52)$$

com  $v_i \sim N(0, \sigma^2)$  e  $u_i \sim LN(\mu, \psi)$ .

MODELO 6: Modelo com a função de produção CES e o termo de ineficiência Lognormal.

$$y_i = \beta_0 + v \log\{[(1 - \delta) \exp(x_1)^{-\rho} + \delta \exp(x_2)^{-\rho}]^{-1/\rho}\} - u_i + v_i \quad (2.53)$$



com  $v_i \sim N(0, \sigma^2)$  e  $u_i \sim LN(\mu, \psi)$ .

MODELO 7: usando o a função de produção Translog e o termo de ineficiência Lognormal.

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \frac{1}{2} \beta_3 x_1^2 + \frac{1}{2} \beta_4 x_2^2 + \beta_5 x_1 x_2 - u_i + v_i \quad (2.54)$$

com  $v_i \sim N(0, \sigma^2)$  e  $u_i \sim LN(\mu, \psi)$ .

MODELO 8: Modelo utilizando a função de produção GPF's e assumindo-se que termo de ineficiência tenha distribuição Lognormal.

$$z_i = y_i + \lambda \exp(y_i) = \beta_0 + \beta_1 x_1 + \beta_1 x_2 - u_i + v_i \quad (2.55)$$

com  $v_i \sim N(0, \sigma^2)$  e  $u_i \sim LN(\mu, \psi)$ .

Novamente, a função de produção generalizada (GPF) usada nos MODELOS 4 E 8, não tem uma relação direta entre  $Y$  e  $X$ , logo a transformação Jacobiana de  $Z$  para  $Y$ , é dada por:

$$J = \prod_{i=1}^N \frac{\partial z_i}{\partial y_i} = \prod_{i=1}^N (\lambda + 1/y_i) \quad (2.56)$$

Assim a função de verossimilhança dos MODELO 4 E 8 são da forma:

$$f(y_i | \beta, \sigma^2, \lambda, u_i) = f_N(z | x_i' \beta - u_i, \sigma^2) | J_i | \quad (2.57)$$

onde  $z = y + \lambda \exp(y)$

A seguir, são descritas as distribuições a priori assumidas para os oito modelos.

### 2.5.1 Distribuição a Priori

A distribuição a priori para o parâmetro  $\tau = \sigma^{-2}$  é a mesma para todos os modelos, assim a priori selecionada foi:  $\tau \sim G(n_0, a_0)$

A seguir, serão definidas as distribuições a priori para os parâmetros das duas alternativas da componente de ineficiência, que é a mesma para todos os modelos:

- Gama:  $p \sim G(d_0, \varepsilon_0)$  e  $\theta \sim G(v_0, \omega_0)$ .
- Lognormal:  $\mu \sim N(a_0, A_0)$  e  $\psi \sim G(d_0, D_0)$ .

As distribuições a priori definidas para o restante dos parâmetros foram:

- MODELO 1 E 5: para  $j = 1, \dots, 3$ ,  $\beta_j \sim N(b_0, H_0)$
- MODELO 2 E 6:  $\beta_0 \sim N(b_0, H_0)$ ,  $v \sim N(m_0, s_0)$ ,  $\delta \sim \text{Beta}(w_0, d_0)$  e  $\rho \sim NT_{[-1, \infty]}(r_0, g_0)$
- MODELO 3 E 7: para  $j = 1, \dots, 5$ ,  $\beta_j \sim N(b_0, H_0)$
- MODELO 4 E 8: para  $j = 1, \dots, 3$ ,  $\beta_j \sim N(b_0, H_0)$  e  $\lambda \sim U(0, 1)$

Da mesma forma mencionada anteriormente a distribuição a priori de  $\rho$  do modelo 2 e 6, não faz parte da lista de distribuições do WinBUGS, assim a distribuição a priori foi implementada da forma mencionada na seção 2.4.1.

Os valores escolhidos para os hiperparâmetros que levaram a uma distribuição a priori relativamente vaga nos oito modelos foram:

$$\begin{bmatrix} d_0 & \varepsilon_0 \\ v_0 & \omega_0 \\ b_0 & H_0^{-1} \\ m_0 & s_0^{-1} \\ w_0 & \varphi_0 \end{bmatrix} = \begin{bmatrix} 1 & 0.1 \\ 0.01 & 0.01 \\ 0 & 0.001 \\ 0 & 0.001 \\ 1 & 1 \end{bmatrix}$$

## 2.5.2 Análise MCMC

Para aplicar o método MCMC, necessitamos definir as distribuições condicionais completas.

A seguir vamos definir a distribuição condicional completa para o termo de ineficiência  $u_i$  para  $1, \dots, N$  e dos parâmetros associados ao termo de ineficiência Gama e Lognormal:

- Quando  $u_i \sim G(P, \theta)$

$$- (\theta | \mathbf{Y}, \mathbf{X}, resto) \sim G \left( NP + v_0, \sum_{i=1}^N u_i + \omega_0 \right)$$

$$- p(P | \mathbf{Y}, \mathbf{X}, resto) \propto P^{d_0-1} \Gamma(P)^{-N} \exp(QP) I_{(0,\infty)}(P)$$

$$\text{onde } Q = \sum_{i=1}^N \log u_i - \xi + N \log \theta$$

$$- \text{para } 1, \dots, N, p(u_i | \mathbf{Y}, \mathbf{X}, resto) \propto u_i^{P-1} \exp \left[ \frac{-(u_i + e_i)^2}{2\sigma^2} - \theta u_i \right]$$

Como mencionado no capítulo 1.4.2 na fronteira Normal-Gama as distribuições condicionais de  $P$  e  $u_i$  não estão disponíveis para amostragem, porém as distribuições são log-concavas em certos casos. O método utilizado pelo WinBUGS para gerar uma amostra de  $P$  e  $u_i$  é o algoritmo de Slice Sampling.

- Quando  $u_i \sim LN(0, \phi)$

$$- (\mu | Y, X, resto) \sim N \left[ \left( \sum_{i=1}^N N \log u_i / \psi + A^{-1} \right) a (N\psi + A^{-1})^{-1} \right]$$

$$- (\psi | Y, X, resto) \sim G \left[ d + N, D + \sum_{i=1}^N (\log(u_i - \mu)^2) - 1 \right]$$

$$- \text{para } 1, \dots, N, p(u_i | Y, X, resto) \propto N[y_i - x'_i \beta, \sigma^2] LN[\mu, \psi]$$

Neste caso, as distribuições condicionais de  $\mu$  e  $\psi$  estão disponíveis na forma fechada. Já a distribuição condicional de  $u_i$  não está disponível para amostragem, neste caso o método utilizado pelo WinBUGS foi o algoritmo de Slice Sampling.

As outras distribuições condicionais para os oito modelos, estão brevemente descritos a seguir:

- MODELO 1 E 5:

$$- (\beta | Y, X, resto) \sim N^k \left[ (H_0 + \sigma^{-2} X'X)^{-1} (H_0 b_0 + X'(Y + u)\sigma^{-2}), H_0 + \sigma^{-2} X'X \right]$$

$$- (\sigma^{-2}|Y, X, resto) \sim G \left[ \frac{(N+N_0)}{2}, \frac{((y+u-X\beta)'(y+u-X\beta)+a_0)}{2} \right]$$

- MODELO 2 E 6:

$$- (\beta_0|Y, X, resto) \sim N \left[ (H_0 + \sigma^{-2})^{-1}(H_0 b_0 + \mathbf{1}'(Y + u)\sigma^{-2}), H_0 + \sigma^{-2} \right]$$

$$- (\sigma^{-2}|Y, X, resto) \sim G \left[ \frac{(N+N_0)}{2}, \frac{((y+u-X^*\pi)'(y+u-X^*\pi)+a_0)}{2} \right]$$

$$- p(\delta|Y, X, resto) \propto \exp \left[ -\frac{(y+u-X^*\pi)'(y+u-X^*\pi)}{2\sigma^2} \right]$$

$$- p(v|Y, X, resto) \propto \exp \left[ -\frac{(y+u-X^*\pi)'(y+u-X^*\pi)}{2\sigma^2} - \frac{s_0(v-m_0)^2}{2} \right]$$

$$- (\rho|resto) \propto \exp \left[ -\frac{(y+u-X^*\pi)'(y+u-X^*\pi)}{2\sigma^2} - \frac{(\rho-\tau_0)^2}{2g_0} \right] I_{(-1,\infty)}(\rho)$$

$$\text{onde } X^* = \left[ 1 \quad \log\{[(1-\delta)\exp(X_1)^{-\rho} + \delta\exp(X_2)^{-\rho}]^{-1/\rho}\} \right] \text{ e}$$

$$\pi = (\beta_0 \quad v)'$$

Pelo mesmo motivo mencionado na seção 2.4.2, o método utilizado pelo WinBUGS para amostrar da distribuição condicional de  $\delta$  e  $v$  foi o algoritmo de Slice Sampling e da distribuição condicional de  $\rho$  o algoritmo de Metropolis-Hastings.

- MODELO 3 E 7:

$$- (\beta|Y, X, resto) \sim N^k \left[ (H_0 + \sigma^{-2}X'X)^{-1}(H_0 b_0 + X'(Y + u)\sigma^{-2}), H_0 + \sigma^{-2}X'X \right]$$

$$- (\sigma^{-2}|Y, X, resto) \sim G \left[ \frac{(N+N_0)}{2}, \frac{((y+u-X^*\beta)'(y+u-X^*\beta)+a_0)}{2} \right]$$

$$\text{onde } X^* = \left[ 1 \quad X_1 \quad X_2 \quad X_3 \quad 1/2X_2 \quad 1/2X_3 \quad X_2X_3 \right]$$

- MODELO 4 E 8:

$$- (\beta|Y, X, resto) \sim N^k \left[ (H_0 + \sigma^{-2}X'X)^{-1}(H_0 b_0 + X'(Z + u)\sigma^{-2}), H_0 + \sigma^{-2}X'X \right]$$

$$\begin{aligned}
& - (\sigma^{-2}|Y, X, resto) \sim G \left[ \frac{(N+N_0)}{2}, \frac{((z+u-X\beta)'(z+u-X\beta)+a_0)}{2} \right] \\
& - p(\lambda|resto) \propto \exp \left[ \frac{1}{2\sigma^2}(z+u-X\beta)'(z+u-X\beta) + \sum_{i=1}^N \log(y_i\lambda + 1) \right] \\
& \text{onde } z = Y + \lambda \exp Y
\end{aligned}$$

Como a distribuição condicional de  $\lambda$  não se encontra na forma fechada, o método utilizado pelo WinBUGS para amostra condicional de  $\lambda$  foi o algoritmo de Slice Sampling.

### 2.5.3 Resultados

Para ilustrar uma comparação entre os modelos de fronteira de produção novamente serão utilizados os dados usados por Zellner e Revankar(1970) para estudar a produção da indústria de equipamentos de transportes de 25 estados, nos Estados Unidos, no ano de 1957. Foram definidas as seguintes variáveis:  $y = \log(\text{produção}/m)$ ,  $x_1 = \log(\text{capital}/m)$  e  $x_2 = \log(\text{trabalho}/m)$  onde  $m$  é o numero de estabelecimentos em cada estado.

Todos os resultados encontrados foram baseados em amostras nas quais a convergência foi testada através do critério de Gelman e Rubin (1992) que está implementado no WinBUGS.

Os resultados para os MODELOS 1-8 são baseados em 10.000 iterações. Foram realizadas 20.000 iterações das quais foram descartadas as 10.000 primeiras para o período de "burn-in".

Na Tabela 2.3 são apresentados os resultados da estatística DIC juntamente com  $\bar{D}$  e  $p_D$  para cada um dos oito modelos.

Pela Tabela 2.3 podemos observar que os modelos de fronteira de produção mais adequado para descrever dados baseando-se no DIC foram os modelos usando forma funcional Translog e a função de produção Generalizada com termo de ineficiência Lognormal (Modelo 7 e 8).

Tabela 2.3: Resultado do Deviance Information Criterion para os oito modelos de fronteira de produção

Modelo	$\bar{D}$	$D(\theta)$	$p_D$	DIC
MODELO 1: Cobb-Douglas e Gama	-20.485	-31.626	11.141	-9.344
MODELO 2: CES e Gama	-20.999	-31.035	10.037	-10.962
MODELO 3: Translog e Gama	-23.448	-37.120	13.672	-9.776
MODELO 4: GPF e Gama	-22.450	-33.671	11.221	-11.229
MODELO 5: Cobb-Douglas e Lognormal	-37.6881	-60.218	22.530	-15.158
MODELO 6: CES e Lognormal	-38.881	-57.957	19.077	-19.804
MODELO 7: Translog e Lognormal	-40.291	-64.352	24.061	-16.231
MODELO 8: GPF e Lognormal	-38.9	-60.214	21.314	-17.586

A Tabela 2.4 apresenta os valores aproximados do pseudo fator de Bayes confrontando os oito modelos.

Na Tabela 2.4 observamos que o valor aproximado do pseudo fator de Bayes do modelo 8 e do modelo 6 versus todos os modelos indicam uma preferência pelos dois modelos, ou seja, usando a fronteira de produção normal-lognormal usando a forma funcional GPF (Generalized Production Functions) ou CES (Constant Elasticity of Substitution) .

Para concluir a compração dos modelos, as distribuições dos ranks associada a medida de eficiência (Goldstain e Spiegelhalter, 1996) são apresentadas. Essas distribuições são diretamente obtidas usando as cadeias de MCMC. Se denotarmos por  $e_j^{(m)} = \exp(-u_j^{(m)})$  como a medida de eficiência para a  $j^{th}$  firma na  $m^{th}$  iteração da cadeia de MCMC. E colocando os  $e_j^{(m)}$  em ordem para  $j = 1, \dots, N$ , o rank associado à  $j^{th}$  firma na  $m^{th}$  iteração é a posição ocupada por  $e_j^{(m)}$  na ordenação.

As Figuras 2.1 e 2.2 apresentam distribuição a posteriori da medida de eficiência

Tabela 2.4: Sumário do pseudo fator de Bayes para os oito modelos de fronteira de produção

Fator de Bayes	Mod 1	Mod 2	Mod 3	Mod 4	Mod 5	Mod 6	Mod 7
Modelo 2	1,437						
Modelo 3	1,206	0,839					
Modelo 4	1,567	1,091	1,299				
Modelo 5	4,359	3,034	3,614	2,782			
Modelo 6	14,829	10,321	12,294	9,463	3,402		
Modelo 7	11,823	8,228	9,802	7,544	2,712	0,797	
Modelo 8	49,084	34,161	40,693	31,321	11,260	3,310	4,152

e do rank para a melhor ( $\rho[10]$ ), pior ( $\rho[4]$ ) e uma mediana ( $\rho[25]$ ) firma segundo os modelos de fronteira Normal-Gama usando a função de produção: Cobb-Douglas, CES, Translog e GPF.

As Figuras 2.3 e 2.4 apresentam a distribuição a posteriori da medida de eficiência e do rank para a melhor ( $\rho[10]$ ), pior ( $\rho[4]$ ) e uma mediana ( $\rho[25]$ ) firma segundo os modelos de fronteira Normal-Lognormal usando a função de produção: Cobb-Douglas, CES, Translog e GPF.

O objetivo desta última análise é comparar a capacidade dos modelos de classificar as firmas. Pelas Figuras 2.1-2.4 nota-se que o modelo de fronteira normal-lognormal com apenas um parâmetro e utilizando qualquer forma funcional, conseguiu diferenciar melhor as firma mais eficientes das menos eficientes. Por exemplo, se comparamos a Figura 2.3 com a Figura 2.1 podemos observar a variabilidade da distribuição dos postos para uma considerada firma eficiente ( $\rho[10]$ ) utilizando o modelo de fronteira Normal-Lognomal com a função de produção Cobb-Douglas é bem menor do que utilizando o modelo de fronteira Normal-Gama com a mesma forma funcional.

Figura 2.1: Distribuição a posteriori da medida de eficiência e do rank para a melhor (rho[10]), pior (rho[4]) e uma mediana (rho[25]) firma segundo o modelo de fronteira Normal-Gama usando a função de produção Cobb-Douglas e CES.

### Modelo 1: Cobb-Douglas e Gama

rho[10] -> Melhor firma   rho[4] -> Pior firma   rho[25] -> Firma Mediana



### Modelo 2: CES e Gama

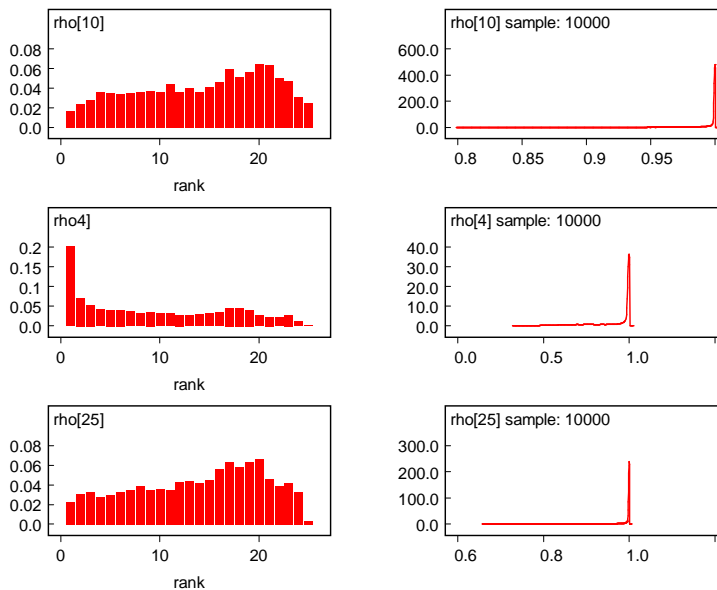
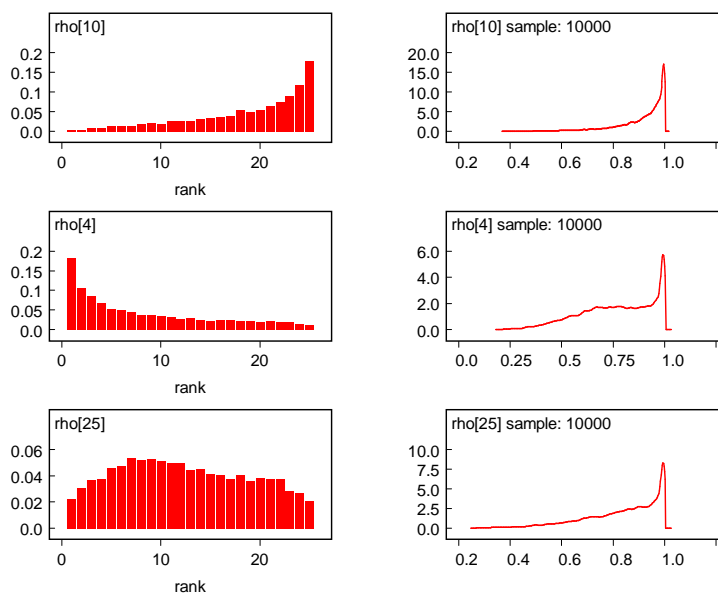




Figura 2.2: Distribuição a posteriori da medida de eficiência e do rank para a melhor (rho[10]), pior (rho[4]) e uma mediana (rho[25]) firma segundo o modelo de fronteira Normal-Gama usando a função de produção Translog e GPF.

### Modelo 3: Translog e Gama

rho[10] -> Melhor firma   rho[4] -> Pior firma   rho[25] -> Firma Mediana



### Modelo 4: GPF e Gama

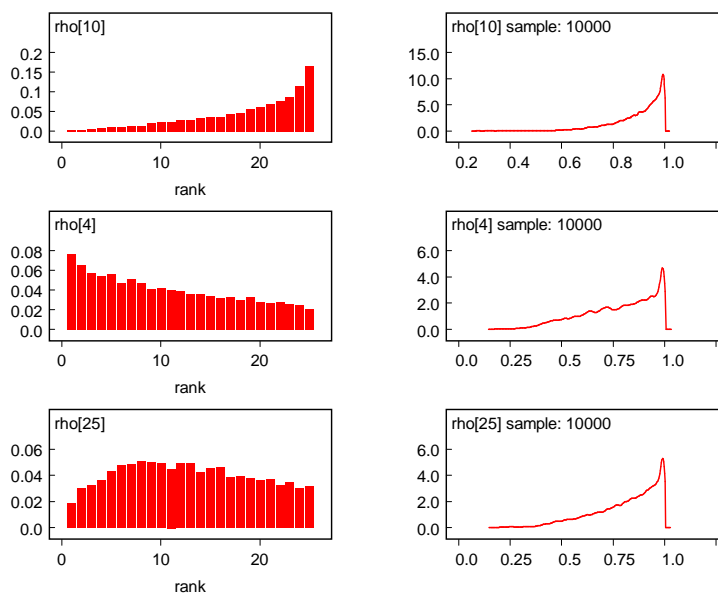
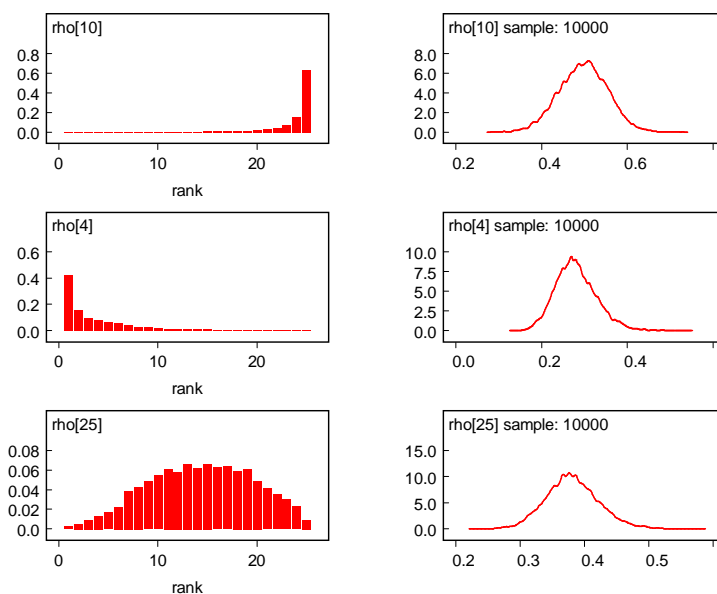


Figura 2.3: Distribuição a posteriori da medida de eficiência e do rank para a melhor ( $\rho[10]$ ), pior ( $\rho[4]$ ) e uma mediana ( $\rho[25]$ ) firma segundo o modelo de fronteira Normal-LogNormal usando a função de produção Cobb-Douglas e CES.

### Modelo 5: Cobb-Douglas e Lognormal

$\rho[10]$  -> Melhor firma     $\rho[4]$  -> Pior firma     $\rho[25]$  -> Firma Mediana



### Modelo 6: CES e Lognormal

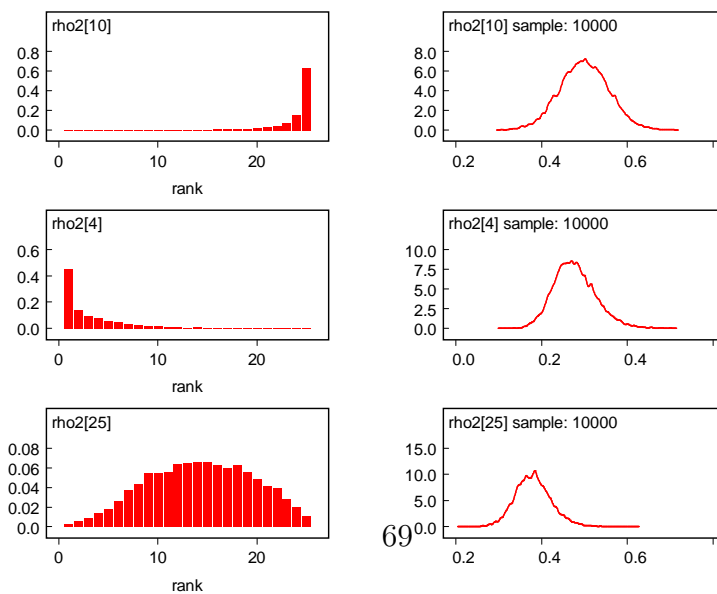
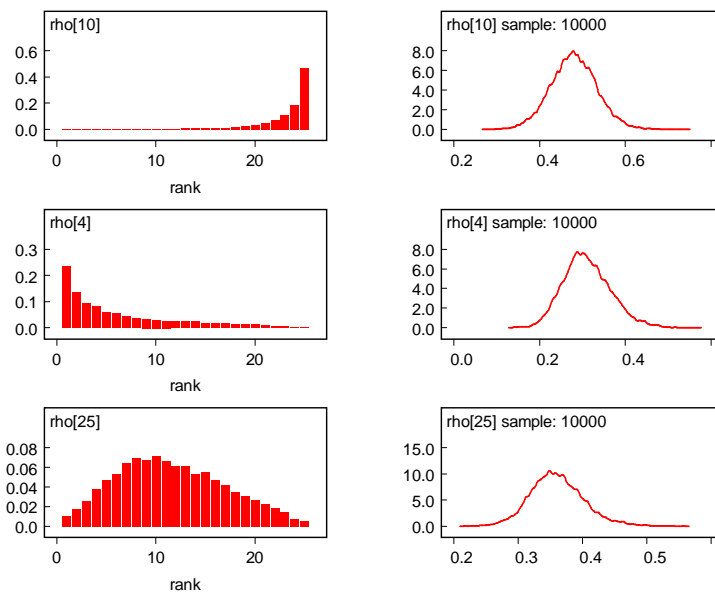


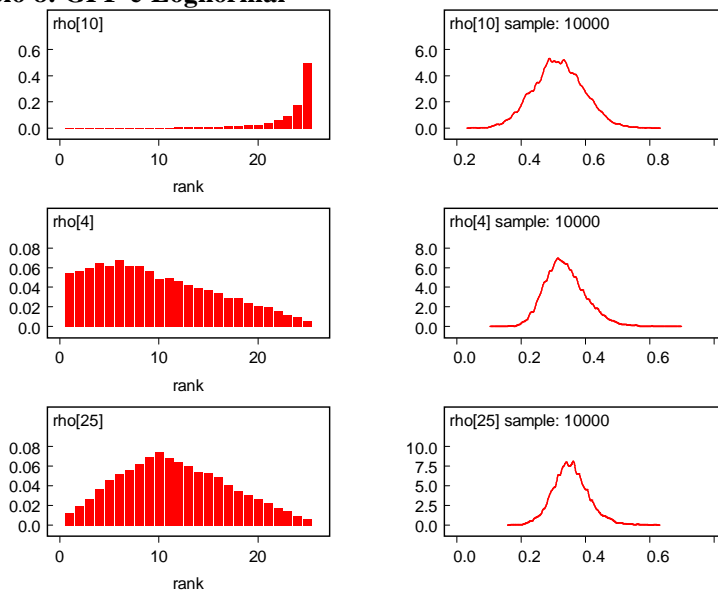
Figura 2.4: Distribuição a posteriori da medida de eficiência e do rank para a melhor (rho[10]), pior (rho[4]) e uma mediana (rho[25]) firma segundo o modelo de fronteira Normal-Lognormal usando a função de produção Trannlog e GPF.

**Modelo 7: Translog e LogNormal**

rho[10] -> Melhor firma   rho[4] -> Pior firma   rho[25] -> Firma Mediana



**Modelo 8: GPF e Lognormal**



# Capítulo 3

## Fronteira de Produção

## Estocástica para

## Múltiplos-Outputs

### 3.1 Introdução

Em muitas situações as firmas produzem mais de um único *output*. Nesta caso para utilizar todos os outputs na maioria dos estudos é construído um índice dos outputs com pesos fixo ou com pesos estimados (Fernandez et al, 2000), para poder aplicar a fronteira de produção neste índice.

Neste capítulo desenvolvemos ferramentas Bayesianas para estimar a fronteira de produção para múltiplos *outputs*. Neste presente trabalho será utilizado a técnica de Análise Fatorial para a construção de um índice e neste será modelado uma fronteira de produção.

O restante do capítulo está organizado da seguinte forma: A segunda seção descreve alguns trabalhos existentes na literatura para múltiplos *outputs*. Na terceira seção é apresentado o modelo fatorial Bayesiano. A quarta seção apresenta

e motiva o modelo proposto neste capítulo. Finalmente, na seção 5 são descritas a aplicação com dados artificiais e reais.

## 3.2 Fronteira de Produção para Múltiplos-*Outputs*

Para motivar este capítulo, vale a pena discutir, brevemente, algumas das literaturas relevantes no caso de múltiplos-*outputs*. O ponto de partida teórico na maioria das análises de tecnologia com múltiplos *outputs* é a função de transformação, que descreve a tecnologia de produção através de uma relação da forma:

$$f(y, x) = 0,$$

onde  $y$  é uma vetor de  $p$  *outputs* e  $x$  é um vetor de *inputs*. Se a função de transformação é separável então podemos escrevê-la da seguinte forma:

$$g(y) = h(x),$$

Para mais informações sobre a função de transformação e as dificuldades na estimação deste modelo pode ser vista em Fernández et al. (2000). A maioria dos investigadores trabalham com a função de custo quando se confrontam com problemas de múltiplos *outputs* e assim evitam os problemas causados ao utilizar a função de transformação. Porém, esta solução requer que o pesquisador tenha coletado os dados dos custos e/ou dos preços. No caso de apenas  $y$  e  $x$  serem observados, o pesquisador é obrigado a lidar diretamente com a função de transformação.

Existe uma extensa literatura (Fare e Primont, 1990) que utiliza uma abordagem não econométrica, assumindo uma função de transformação determinística (sem medida de erro nos dados) e usando técnicas de programação linear.

Os trabalhos que se encontram na literatura que utilizam apenas uma abordagem econométrica são os de Adams et al.(1999) e Lothgren(1997), todos assumindo a função de transformação separável.

O trabalho mais recente encontrado na literatura utilizando uma abordagem Bayesiana é o da Fernández et al. (2000) onde usa métodos econométricos assumindo uma função de transformação separável. Na próxima seção vamos descrever brevemente o modelo proposto por Fernández et al.(2000) com o objetivo de fazer um paralelo com o modelo proposto nesta dissertação.

Neste presente trabalho, vamos utilizar a técnica da análise multivariada que nos permite descrever a variabilidade de um conjunto de *outputs* em termos de uma única variável aleatória não observável chamado de fator. E assim, podendo-se construir uma fronteira de produção em cima deste fator.

Gostaríamos de enfatizar que o propósito deste capítulo não é discutir se a abordagem Bayesiana é invariavelmente superior para a econometria clássica ou abordagens determinísticas. Cada método possui suas vantagens e desvantagens.

### 3.2.1 Modelo proposto por Fernández et al. (2000)

Considere um conjunto de  $N$  observações correspondendo aos outputs de  $N$  firmas diferentes. O output da firma  $i$  ( $i = 1, \dots, N$ ) é  $p$ -dimensional e é dado pelo vetor  $y_i = (y_{i,1}, \dots, y_{i,p})' \in \mathfrak{R}_+^p$ .

A estrutura do modelo é desenvolvido pela seguinte transformação do vetor de output  $p$ -dimensional:

$$\theta_i = \left( \sum_{l=1}^p \alpha_l^q y_{i,l}^q \right)^{1/q} \quad (3.1)$$

com  $\alpha_j \in (0, 1)$  para  $\forall j = 1, \dots, p$  e de modo que  $\sum_{j=1}^p \alpha_j = 1$  e com  $q > 1$ .

Na literatura, a transformação em (3.1) é chamado de 'elasticidade de transformação constante' (ver Powell e Gruen, 1968; Kumbhakar, 1987), onde a elasticidade de transformação de quaisquer dois outputs é dado por  $1/(1 - q)$ .

Deste modo, a Eq. (3.1) define a transformação de vetor do output multivariado  $y_i$  para a quantidade univariada  $\theta_i$ . Se interpretamos o valor de  $\theta_i$  como um

tipo de 'output agregado', então definindo que  $\delta_i = \log(\theta_i)$  e  $\delta_i$  é modelado pelo seguinte modelo de fronteira estocástica:

$$\delta_i = x_i' \beta - u_i + v_i \quad (3.2)$$

onde  $x_i$  denota vetor de *input*  $m$ -dimensional e o vetor correspondente ao coeficiente de regressão é denotado por  $\beta \in \mathfrak{R}^k$ . O termo de ineficiência  $u_i$  assume-se ter distribuição exponencial com média  $1/\lambda$  e  $v_i$  tem distribuição Normal com média 0 e variância  $\sigma^2$ . A função de densidade de probabilidade (f.d.p) de  $\delta$  dado  $(\beta, u, \sigma^2)$  será denotado como segue:

$$P(\delta|\beta, u, \sigma) = f_N^N(\delta|X\beta - u, \sigma^2 \mathbf{I}_N) \quad (3.3)$$

Como a Eq. (3.3) não é suficiente para definir a densidade amostral da observação  $y_i$ . Para completar o modelo amostral defini-se a seguinte variável:

$$\eta_{i,j} = \frac{\alpha_j^q y_{i,j}^q}{\sum_{l=1}^p \alpha_l^q y_{i,l}^q}, \quad j = 1, \dots, p \quad e \quad \eta_i = (\eta_{i,1}, \dots, \eta_{i,p}) \quad (3.4)$$

e assume-se amostras independentes ( $i = 1, \dots, N$ ) de

$$p(\eta_i|s) = f_D^{p-1}(\eta_i|s), \quad (3.5)$$

onde  $s = (s_1, \dots, s_p)' \in \mathfrak{R}_+^p$  e  $f_D^{p-1}$  denota a f.d.p de uma distribuição Dirichlet  $(p-1)$ -dimensional com parâmetro  $s$ . Note que  $\eta_i$  pode ser interpretado como um vetor da participação de cada componente de  $y_i$ .

Através das Eqs (3.3)-(3.5) obtém-se a f.d.p. da matriz  $N \times p$  de  $Y$ :

$$P(\mathbf{Y}|\beta, z, \sigma^2, \alpha, q, s) = f_N^N(\delta|X\beta - u, \sigma^2 \mathbf{I}_N) \prod_i f_D^{p-1}(\eta_i|s) \prod_{i,j} q^{1-(1/p)} \frac{\eta_{i,j}}{y_{i,j}}, \quad (3.6)$$

onde o último termo é a transformação Jacobiana.

Se tomarmos o logaritmo em (3.4) temos a seguinte expressão:

$$\log y_{i,j} = \frac{\log \eta_{i,j}}{q} - \log \alpha_j + \log \theta_i \quad (3.7)$$

E como  $\log \theta_i = \delta_i \sim N(x_i\beta - u_i, \sigma^2)$  temos que

$$(y_{i,j} | \eta_{i,j}, \beta, z, \sigma^2, \alpha, q, s) \sim N \left[ \left( \frac{\log \eta_{i,j}}{q} - \log \alpha_j - x_i\beta - u_i \right), \sigma^2 \right] \quad (3.8)$$

Como a densidade amostral de  $Y$  apresentado em (3.6) não tem uma forma conhecida, a implementação deste modelo no WinBUGS pode ser realizado da mesma forma descrita na seção 2.4.1 do capítulo 2.

A seguir, serão apresentadas as distribuições a priori e as condicionais completas do modelo proposto por Fernández et al. (2000).

### Distribuição a Priori

A distribuição a priori para  $(\alpha, q, s, \beta, \lambda, \sigma)$  tem a seguinte estrutura:

$$p(\alpha, q, s, \beta, \lambda, \sigma) = p(\alpha)p(q)p(s)p(\beta)p(\lambda)p(\sigma) \quad (3.9)$$

- Priori para  $\alpha$

Como os componentes de  $\alpha$  estão todos no intervalo  $(0, 1)$  e a soma é igual a um, a distribuição a priori é uma Dirchlet com f.d.p.

$$p(\alpha) = f_D^{p-1}(\alpha|a) \quad (3.10)$$

- Priori para  $s_j$

Assumindo independência nas componentes de  $s$  temos:

$$p(s) = \prod_{j=1}^p p(s_j) = \prod_{j=1}^p f_G(b_j, c_j). \quad (3.11)$$

- Priori para  $q$

Para  $q$  é definido uma priori exponencial truncada no intervalo  $(1, \infty)$ ,

$$p(q) \propto f_G(q|1, d)I_{(1, \infty)}(q) \quad (3.12)$$



- Priori para  $\beta$

A priori assumida para o parâmetro de fronteira tem p.d.f.

$$p(\beta) = f_N^k(\beta|b_0, H_0^{-1}) \quad (3.13)$$

- Priori para  $\sigma^2$

Definindo a precisão  $\tau = \sigma^{-2}$ , a distribuição a priori para  $\tau$  é

$$p(\tau) = f_G(h|n_0/2, a_0/2) \quad (3.14)$$

- Priori para  $\lambda$

A priori para  $\lambda$  tem f.d.p.

$$p(\lambda) = f_G(\lambda|v_0, \omega_0) \quad (3.15)$$

### Distribuições Condicionais

A seguir, serão descritas as distribuições condicionais a posteriori dos parâmetros  $(\alpha, q, s, \beta, \lambda, \sigma, u)$ .

- Distribuição condicional a posteriori de  $\alpha$

$$p(\alpha|resto) \propto \prod_j \alpha_j^{a_j + s_j q^{N-1}} \prod_i \left( \sum_j \alpha_j^q y_{i,j}^q \right)^{-\sum_j s_j} \exp \left\{ -\frac{1}{2\sigma^2} (\delta - X\beta + u)' (\delta - X\beta + u) \right\}, \quad (3.16)$$

- Distribuição condicional a posteriori de  $q$

$$p(q|resto) \propto q^{N(p-1)} \exp(-dq) \exp\{-A(q)\} I(1, \infty)(q), \quad (3.17)$$

$$\text{onde } A(q) = \frac{1}{2\sigma^2} (\delta - X\beta + u)' (\delta - X\beta + u) + \sum_{i,j} s_j \log \left( \frac{\sum_{l=1}^p \alpha_l^q y_{i,l}^q}{\alpha_j^q y_{i,j}^q} \right),$$

- Distribuição condicional a posteriori de  $s_j$  para  $j = 1, \dots, p$

$$p(s_j|resto) \propto \frac{\Gamma(\sum_l s_l)^N}{\Gamma(s_j)^N} s_j^{b_j-1} \exp \left[ -s_j \left\{ c_j + \sum_i \log \left( \frac{\sum_l \alpha_l^q y_{i,l}^q}{\alpha_j^q y_{i,j}^q} \right) \right\} \right], \quad (3.18)$$

- Distribuição condicional a posteriori de  $\beta$

$$p(\beta|resto) = f_N^k(\beta|b_*, H_*^{-1}) \quad (3.19)$$

onde  $b_* = H_*^{-1}\{H_0 b_0 + \sigma^{-2} X'(\delta + u)\}$  e  $H_* = H_0 + \sigma^{-2}(X'X)^{-1}$

- Distribuição condicional a posteriori de  $\sigma^2$

$$p(\tau|resto) = f_G \left( \tau \mid \frac{n_0 + N}{2}, \frac{a_0 + (\delta - X\beta + u)'(\delta - X\beta + u)}{2} \right). \quad (3.20)$$

- Distribuição condicional a posteriori de  $\lambda$

$$p(\lambda|resto) = f_G \left( \lambda \mid N + v_0, \sum_{i=1}^N u_i + \omega_0 \right) \quad (3.21)$$

- Distribuição condicional a posteriori de  $u_i$  para  $i = 1, \dots, N$

$$p(u_i|resto) = f_{NT}(u_i|m, R) \quad (3.22)$$

onde  $m = x'_i \beta - \delta - \sigma^2 \lambda$  e  $R = \sigma^2$

Podemos observar que o modelo proposto por Fernández et al. (2000), as distribuições condicionais dos parâmetros  $\alpha, q$  e  $s$  não tem uma forma fechada, logo algum método de amostragem aleatória deve ser implementada.

### 3.3 Modelo Fatorial Bayesiano

A idéia básica desenvolvida e apresentada nesta seção é a técnica da análise multivariada, que tem como principal objetivo descrever a variabilidade original de um conjunto de variáveis aleatórias observáveis em termos de um número

menor de variáveis aleatórias não observáveis chamada de fatores. Estes fatores contêm a informação das características comuns entre as variáveis originais e a complexa relação entre elas.

Especificamente no modelo fatorial, parte da variabilidade do conjunto de variáveis aleatórias observáveis é atribuída aos fatores comuns sendo que o restante da variabilidade é atribuída a variáveis que não foram incluídas no modelo.

Algumas aplicações Bayesianas em modelos de análise fatorial, que podem ser vistos em, por exemplo, Geweke e Zhou(1996), Lopes (2000) e Aguilar e West(2001).

A seguir, introduziremos, formalmente, o modelo fatorial com algumas propriedades básicas.

### 3.3.1 Modelo Fatorial

Assumindo que  $y_i$  é um vetor aleatório  $p$ -dimensional com media  $0_p$  e matriz de covariância  $\Delta$ . O modelo  $p$ -fatorial básico para as observações  $\{y_i, i = 1, \dots, N\}$  considera que as  $p$  variáveis são representadas através de uma combinação linear dos  $k$  fatores comuns com  $k \ll p$  da seguinte forma,

$$y_i = Lf_i + \varepsilon_i \quad (3.23)$$

onde  $L$  é uma matriz  $p \times k$  de coeficientes usualmente chamada de matriz de fatores de cargas. Os elementos do vetor aleatório  $k$ -dimensional  $f_i$  são os fatores comuns e  $\varepsilon_i$  é vetor aleatório  $p$ -dimensional usualmente denotado por erro ou termo de distúrbio. Neste modelo é assumido que:

1. Os fatores  $f_i$  são independentes com  $f_i \sim N(\mathbf{0}, \mathbf{I}_k)$ ,
2. Os  $\varepsilon_i$  são independentes com  $\varepsilon_i \sim N(\mathbf{0}, \Psi)$  com  $\Psi = \text{diag}(\psi_1, \psi_2, \dots, \psi_p)$ ,
3.  $\varepsilon_t$  e  $f_s$  são mutuamente independentes para todo  $t$  e  $s$ ,

Por estas suposições, o modelo  $k$ -fatorial pode ser expresso em termos de condições simples na matriz de covariância  $\Delta$ ,

$$\Delta = LL' + \Psi, \quad (3.24)$$

onde os elementos da diagonal da matriz  $LL'$  são tradicionalmente chamados de comunalidade,  $l_i^2 = \sum_{j=1}^k l_{ij}^2$ , para  $i = 1, \dots, p$  e os elementos de  $\Psi$  são chamados de unicidade ou de variância específica.

### Identificação e Restrições

O modelo  $k$ -fatorial como é apresentado em (3.23) é altamente parametrizado e conseqüentemente não identificado pelos dados. Por esses problemas o modelo  $k$ -fatorial deve ser restringido para definir um único modelo livre de problemas de identificação.

#### 1. Posto na matriz de cargas

Uma primeira restrição é que a matriz  $L$  seja de posto completo, para evitar problemas de identificação. Geweke e Singleton(1980) mostram que se  $L$  é de posto incompleto, então o modelo não é identificável.

Se assumimos que  $rank(L) = r$  com  $r < k$ , então existe uma matriz  $Q$  de dimensão  $k \times (k - r)$  de modo que  $LQ=0$  e  $Q'Q=I_{k-r}$ , e para qualquer matriz ortogonal  $M$  temos que,

$$\begin{aligned} \Delta = LL' + \Psi &= (LL' + MM') + \Psi - MM' \\ &= (L + MQ')(L + MQ')' + \Psi - MM'. \end{aligned}$$

isso implica que  $\Delta = \hat{L}\hat{L}' + \hat{\Psi}$  onde  $\hat{L} = L + MQ'$  e  $\hat{\Psi} = \Psi - MM'$  e conseqüentemente o modelo não é identificável.

#### 2. Parsimonia

A segunda restrição na matriz de cargas é evitar o excesso de parâmetros e isso pode ser feito simplesmente assegurando que o numero de parâmetros livres não exceda  $p(p+1)/2$  parâmetros .

### 3. Rotação ortogonal

Por último necessitamos para assegurar a invariância sobre transformações da forma  $f_i^* = Pf_i$  e  $L^* = LP'$  onde P é qualquer matriz ortogonal  $k \times k$  (Press 1982, capítulo 10).

Há muitas formas de evitar os problemas mencionados anteriormente. Uma delas é impondo restrições na matrix de cargas  $\mathbf{L}$ . A solução adotado neste estudo segue Geweke e Zhou(1996), entre outros, que adotam uma tradicional solução baseado em restrições "hierárquicas"na matriz de cargas. Assim, a forma da matriz definida para  $\mathbf{L}$  é uma matriz triangular inferior com 1's na diagonal,

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{2,1} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ l_{k,1} & l_{k,2} & l_{k,3} & \cdots & 1 \\ l_{k+1,1} & l_{k+1,2} & l_{k+1,3} & \cdots & l_{k+1,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ l_{p,1} & l_{p,2} & l_{p,3} & \cdots & l_{p,k} \end{pmatrix}$$

Esta forma imediatamente assegura que  $\mathbf{L}$  seja de posto completo k e que a matriz de cargas tenha  $d = pk - k(k+1)/2$  parâmetros livres. Para mais detalhes e outras soluções ver Geweke e Zhou(1996), Lopes (2000) e Aguilar e West(2001).

## 3.4 Modelo Fatorial e Fronteira de Produção Estocástica

A finalidade deste capítulo é apresentar uma forma de estimar fronteiras de produção para caso de múltiplos *outputs*, onde todas as informações sobre todos os *outputs* fossem levadas em consideração na construção de um índice final.

A seguir, vamos utilizar a técnica de Análise Fatorial como o objetivo de descrever a variabilidade original de um conjunto de *outputs* em termos de uma única variável aleatória não observável chamado de fator e este fator será modelado através de um modelo de fronteira estocástica.

Considere um conjunto de  $N$  observações correspondente aos outputs de  $N$  diferentes firmas. O *output* da firma  $i$  ( $i=1, \dots, N$ ) é  $p$ -dimensional e é dado através de um vetor  $y_i = (y_{i,1}, \dots, y_{i,p})' \in \mathfrak{R}^p$ .

O modelo  $k$ -fatorial considerado aqui neste estudo é o modelo fatorial com apenas um fator ( $k = 1$ ) tem a seguinte forma,

$$y_i = Lf_i + \varepsilon_i \quad (3.25)$$

onde  $L = (1, l_{2,1}, \dots, l_{p,1})'$  é o vetor de cargas com  $l_{j,1} \in \mathfrak{R}$  para  $j = 2, \dots, p$ ,  $f_i$  é fator comum e  $\varepsilon_i$  assume-se ter distribuição Normal com vetor de média  $\mathbf{0}$  e matriz de covariância  $\Psi$  com  $\Psi = \text{diag}(\psi_1^2, \dots, \psi_p^2)$ . Assim a f.d.p de  $y_i$  dado  $(L, f_i, \Psi)$  pode ser denotada por:

$$p(y_i|L, f_i, \Psi) = f_N(y_i|Lf_i, \Psi) \quad (3.26)$$

A Eq.(3.25) deste modo define a transformação de um vetor de output multivariado  $y_i$  para uma quantidade univariada  $f_i$ . Dada está transformação, o problema básico de encontrar a eficiência de uma firma específica é a mesma que no caso de um único *output*. Se interpretarmos o valor de  $f_i$  como um tipo de

'índice dos *outputs*' e assim obteremos um grupo de *outputs* transformados em um vetor N-dimensional

$$f = (f_1, \dots, f_N) \quad (3.27)$$

e  $f_i$  é modelado pelo seguinte modelo de fronteira estocástica:

$$f_i = x'_i \beta - u_i + v_i \quad (3.28)$$

Seguindo a prática usual,  $x_i$  denota o vetor de *input* m-dimensional e o vetor correspondente ao coeficiente de regressão é denotado por  $\beta \in \mathfrak{R}^m$ . A ineficiência tecnológica é capturada pelo fato das firmas poderem estar abaixo da fronteira, obtendo-se assim um vetor de ineficiência  $u_i$ , onde  $u_i$  assume-se ter particular distribuição unilateral e  $v_i$  tem distribuição Normal com média 0 e variância  $\sigma^2$ . A eficiência correspondente a firma  $i$  será definido por  $\exp(-u_i)$ .

A essência do modelo de fronteira estocástica para múltiplos *outputs* é descrita pela equação (3.28), o qual estende o conceito de fronteira e eficiência para a situação de múltiplos *outputs*.

A seguir o modelo definido em (3.26) e (3.28) será chamado de modelo de fronteira de produção fatorial.

Finalmente, o modelo em (3.28) também captura o fato de que a fronteira não é conhecida exatamente, mas pode ser estimada através dos dados. Assim a f.d.p de  $f$  dado  $(\beta, \sigma^2, u)$  é dado por:

$$p(f|\beta, \sigma^2, u) = f_N^N(f|X\beta - u, \sigma^2 I_N) \quad (3.29)$$

Através das Eqs. (3.26) e (3.28) obtém-se a seguinte densidade amostral para os *outputs* observados  $y_i$  dado  $\Theta = (L, \Psi, \beta, \lambda, \sigma)$ :

$$p(y_i|\Theta, x_i) = \int_{\mathfrak{R}} \int_{\mathfrak{R}^+} f_N^3(y|Lf, \Psi) f_N^1(f_i|x'_i \beta - u_i, \sigma^2) f_{Exp}(u_i|\lambda) du_i df_i \quad (3.30)$$

Assumindo independência entre as firmas, a função de verossimilhança de  $Y$  é o produto das densidades dadas na equação (3.30):

$$l(\Theta|y, X) = p(Y|X, \Theta) = \prod_{i=1}^N p(y_i|x_i, \Theta), \quad (3.31)$$

### 3.4.1 Distribuição a Priori

Para realizar a análise Bayesiana completa, necessitamos complementar a função de verossimilhança de (3.31) com a distribuição a priori dos parâmetros  $(L, \Psi, \beta, \lambda, \sigma)$ . A distribuição conjunta a priori para todos os parâmetros é dado por:

$$p(L, \Psi, \beta, \lambda, \sigma) = p(L)p(\Psi)p(\beta)p(\lambda)p(\sigma^2) \quad (3.32)$$

Nas aplicações todos os valores escolhidos para os hiperparâmetros levaram a uma priori relativamente não informativa, porém própria.

#### Priori para $L$

A priori assumida para o parâmetro  $L$  tem p.d.f.

$$p(L) = f_N^p(L|\mu_0, C_0) \quad (3.33)$$

Na ilustração empírica foi definido  $\mu_0 = 0_p$  e  $C_0 = 1 \times I_p$ .

#### Priori para $\Psi$

Seja  $\phi_j = \psi_j^{-2}$  para  $(j = 1, \dots, p)$  assim foi definido que a distribuição a priori para cada  $\phi_j$  seja

$$p(\phi_j) = f_G(\phi_j|v_0/2, s_0/2). \quad (3.34)$$

Nas aplicações, foram definidas  $v_0/2 = 0.001$  e  $s_0/2 = 0.001$ .



### Priori para $\beta$

A priori assumida para o parâmetro de fronteira tem p.d.f.

$$p(\beta) = f_N^k(\beta|b_0, H_0^{-1}) \quad (3.35)$$

Nas aplicações com dados reais e simulados foram definidos  $b_0 = 0_k$  e  $H_0 = 10^{-4} \times I_k$ .

### Priori para $\sigma^2$

Seja  $\tau = \sigma^{-2}$  assim foi definido que a distribuição a priori para  $\tau$  seja

$$p(\tau) = f_G(\tau|n_0/2, a_0/2). \quad (3.36)$$

Na ilustração empírica, foi definido  $n_0/2 = 1$  (ou seja uma priori exponencial para  $\tau$ ) e  $a_0/2 = 10^{-6}$ . Esses valores implicam uma priori com bastante incerteza.

### Priori para $\lambda$

A priori definida para o parâmetro de escala do termo de ineficiência foi

$$p(\lambda) = f_G(\lambda|v_0, \omega_0). \quad (3.37)$$

Uma priori não informativa foi usada na ilustração empírica, cujos os valores foram  $v_0 = 0.1$  e  $\omega_0 = 0.01$ .

## 3.4.2 Métodos MCMC

A função de verossimilhança em (3.31) junto com a distribuição a priori definida em (3.33)-(3.37) definem o modelo Bayesiano, isto é, a distribuição conjunta das observações e dos parâmetros.

Para aplicar MCMC todas as distribuições condicionais completas devem ser avaliadas. A seguir serão descritas todas as distribuições condicionais completas:

### Distribuição condicional Posteriori de $\mathbf{L}$

Multiplicando (3.31) e (3.33) podemos obter seguinte f.d.p para  $l_{j,1}$  com  $j = 2, \dots, p$ :

$$p(l_{j,1}|Y, X, f, \Psi, \sigma^2, \lambda, u) = f_N^1(l_j|\mu_j, C_j) \quad (3.38)$$

onde

$$\mu_j = C_j^{-1}(C_0^{-1}\mu_0 + \psi_j^{-2} \sum_{k=1}^N f_k y_{k,j})$$

e

$$C_j = C_0^{-1} + \psi_j^{-2} \mathbf{f}'\mathbf{f}$$

### Distribuição Condicional Posteriori de $\Psi$

A condicional completa a posteriori para os elementos de  $\Sigma$  se reduz para um conjunto de  $p$  gamas independentes. A distribuição condicional a posteriori de  $\psi_j^2$  para  $j = 1, \dots, p$  tem a seguinte f.d.p:

$$p(\psi_j^2|Y, X, \mathbf{L}, f, \beta, \lambda, u) = f_{GI} \left( \psi_j^2 \left| \frac{(v + N)}{2}, \frac{(s + \sum_{k=1}^N (y_{k,j} - f_k l_{k,1})^2)}{2} \right. \right) \quad (3.39)$$

### Distribuição Condicional Posteriori de $\mathbf{f}$

Amostraremos separadamente cada umas das  $N$  componentes de  $\mathbf{f}$ . A distribuição condicional posteriori de  $f_i$  com  $i = 1, \dots, N$ , tem f.d.p igual a

$$p(f_i|y, X, L, \Psi, \beta, \sigma^2, \lambda, u) = f_N^1(f_i|m_*, R_*^{-1}) \quad (3.40)$$

onde

$$m_* = R_*^{-1}[\sigma^{-2}(x'_i\beta + u_i) + L'\Psi^{-1}y_i]$$

e

$$R_* = \sigma^{-2} + L'\Psi^{-1}L$$

### Distribuição condicional Posteriori de $\beta$

Multiplicando (3.31) e (3.35) obtemos a seguinte f.d.p para  $\beta$

$$P(\beta|L, \Psi, f, Y, X, \sigma^2, \lambda, u) = f_N^k(\beta|b_*, H_*^{-1}) \quad (3.41)$$

onde

$$b_* = H_*^{-1}\{H_0 b_0 + \sigma^{-2} X'(Y + u)\}$$

e

$$H_* = H_0 + \sigma^{-2}(X'X)^{-1}$$

### Distribuição Condicional Posteriori de $\sigma^2$

A distribuição condicional posteriori de  $\tau = \sigma^{-2}$  tem f.d.p proporcional ao produto de (3.31) e (3.36), o qual resulta em

$$p(\tau|Y, X, L, \Psi, f, \beta, \lambda, u) = f_G\left(\tau \mid \frac{(N+N_0)}{2}, \frac{((y+u-X\beta)'(y+u-X\beta)+a_0)}{2}\right) \quad (3.42)$$

### Distribuição Condicional Posteriori de $\lambda$

A condicional a posteriori de  $\lambda$  tem a f.d.p igual a

$$p(\lambda|Y, X, L, f, \beta, \sigma^2, u) = f_G\left(\lambda \mid N + v_0, \sum_{i=1}^N u_i + \omega_0\right) \quad (3.43)$$

### Distribuição Condicional Posteriori de $u$

Como todos os elementos de  $u$  são condicionalmente independentes podemos amostrar os  $u$ 's separadamente. A distribuição condicional posteriori de  $u_i$  para  $i = 1, \dots, N$  tem f.d.p. igual a

$$p(u_i|y, X, L, \Psi, f, \beta, \sigma^2, \lambda) = f_{NT}(u_i|m, R) \quad (3.44)$$

onde  $m = x_i'\beta - f_i - \sigma^2\lambda$

e

$$R = \sigma^2$$

Uma das vantagens do modelo proposto neste trabalho, é que todas as distribuições condicionais se encontram disponíveis em uma forma fechada e assim podendo ser facilmente implementada no WinBUGS.

## 3.5 Aplicação

### 3.5.1 Dados Artificiais

Para avaliar as possibilidades para conduzir uma inferência, é conveniente considerar primeiro o desempenho do modelo que estamos propondo, com dados artificiais. Analisar os dados simulados nos permitem verificar empiricamente três importantes questões:

1. que todos os parâmetros sejam identificados e que uma inferência razoavelmente precisa possa ser conduzida com base em uma amostra de uma ordem de magnitude provável de acontecer na prática.
2. que a influência da priori não seja muito forte e não domine a informação obtida na amostra.
3. que as propriedades numéricas do algoritmo de MCMC usado seja satisfatório e que a convergência dos parâmetros seja alcançada.

A aplicação a seguir foi feita a partir de dados que foram gerados artificialmente através de uma rotina feita em S plus do modelo de fronteira de produção fatorial. O conjunto de dados artificiais contém  $N = 400$  firmas,  $p = 3$  outputs e  $\beta = (1, 1)$ . As variáveis explicativas na matriz  $400 \times 2$   $X$  são obtidas da seguinte forma: A matriz  $X$  de dimensão  $400 \times 2$  é obtida da seguinte forma: A primeira coluna de  $X$  é um vetor de 1's que corresponde ao intercepto. A segunda coluna foi retirada de uma distribuição  $N(0, 1)$  para cada uma das  $N = 400$  firmas. E

os valores dos parâmetros  $\sigma^2$  e  $\lambda$  são 0.25 e 5 respectivamente. Todos estes valores são os mesmos usados em Fernández et al.(2000). Os valores dos parâmetros relacionados ao modelo fatorial são:  $\mathbf{L} = (1, 0.57, 0.57)$  e  $diag(\Psi) = (0.8, 0.8, 0.8)$ .

Nesta ilustração com dados artificiais, foram realizadas 15000 iterações no qual foram descartadas as 5000 primeiras para o período de burn-in. A convergência de todos os parâmetros foi verificada segundo o critério de convergência de Gelman e Rubin (1992a). Os modelos de fronteira de produção fatorial e o proposto por Fernández (2000) foram implementados no WinBUGS.

A Tabela 3.1 apresenta a média, o desvio padrão, a mediana e os percentis de 2.5 e 97.5 a posteriori de todos os parâmetros do modelo de fronteira de produção fatorial. Podemos notar que as médias e medianas a posteriori estão todas em torno dos valores verdadeiros.

Tabela 3.1: Sumário da distribuição a posteriori para todos os parâmetros do modelo de fronteira de produção fatorial usando o conjunto de dados simulados

	Valor real	Media	d.p.	Mediana	2.5%	97.5%
$l_1$	1	-	-	-	-	-
$l_2$	0.577	0,4896	0,0862	0,4868	0,3299	0,6663
$l_3$	0.577	0,6160	0,0961	0,6111	0,4372	0,8193
$\psi_1^2$	0.8	0,8854	0,1061	0,8822	0,6844	1,1030
$\psi_2^2$	0.8	0,8200	0,0642	0,8173	0,7020	0,9537
$\psi_3^2$	0.8	0,7659	0,0648	0,7632	0,6462	0,9002
$\beta_1$	0.5	0,5063	0,1032	0,4994	0,3225	0,7269
$\beta_2$	0.5	0,4870	0,05741	0,4864	0,3739	0,6052
$\sigma^2$	0.25	0,2499	0,1034	0,2543	0,0512	0,4517
$\lambda$	5	5,0330	2,1780	4,560	2,2460	10,7600

A Figura 3.1 apresenta a densidade marginal a posteriori para todos os parâmetros do modelo de fronteira de produção fatorial. Claramente observa-se que todas as densidades estão concentradas em torno dos valores usados para gerar os dados.

A seguir, será apresentada uma análise usando os mesmos dados gerados, mas usando agora o modelo proposto por Fernández et al. (2000). Os resultados das estimativas obtidas usando o modelo proposto por Fernández et al. (2000) estão na Tabela 3.2 que contém a média, o desvio padrão, a mediana e os percentis de 2.5 e 97.5 a posteriori de todos os parâmetros.

Tabela 3.2: Sumário da distribuição a posteriori para todos os parâmetros do modelo proposto por Fernández et al. (2000) usando o conjunto de dados simulados

	Media	d.p.	Mediana	2.5%	97.5%
$\alpha_1$	0,2139	0,0199	0,2138	0,1757	0,2531
$\alpha_2$	0,3650	0,0280	0,3648	0,3112	0,4208
$\alpha_3$	0,4211	0,0295	0,4209	0,3623	0,4789
q	1,0250	0,0262	1,0180	1,0010	1,1030
$s_1$	1,2300	0,1040	1,2310	1,0240	1,4370
$s_2$	1,6090	0,1369	1,6040	1,3450	1,8860
$s_3$	1,7460	0,1597	1,7400	1,4360	2,0770
$\beta_1$	0,6104	0,0700	0,6078	0,4830	0,7585
$\beta_2$	0,3287	0,0350	0,3282	0,2643	0,3995
$\sigma^2$	0,6530	0,0308	0,6536	0,5920	0,7134
$\lambda$	4,7380	1,4260	4,4470	2,7800	8,7300

Fazendo um paralelo entre os dois modelos apresentados nesta dissertação, observa-se que alguns parâmetros são incomparáveis. Os parâmetros que podem ser considerados comparáveis são os parâmetros de fronteira. Pela Tabela 3.2 no-

tamos que as médias a posteriori obtidas para  $\beta_1$ ,  $\beta_2$ ,  $\sigma^2$  e  $\lambda$  estão todas próximas dos valores usados para gerar os dados. Um outro resultado interessante é que a média a posteriori do parâmetro  $q$  (1,0250) é próximo de 1 e se observarmos a Eq. (3.1) podemos dizer que os 3 *outputs* possuem uma relação próxima da linear.

As densidades marginais a posteriori para todos os parâmetros utilizando-se o modelo proposto por Fernández et al. (2000) são apresentadas na Figura 3.2.

Uma outra forma de se comparar as estimativas obtidas dos dois modelos é através das distribuições dos ranks associada a medida de eficiência. Essas distribuições são diretamente obtidas usando as cadeias de MCMC. Se denotarmos por  $e_j^{(m)} = \exp(-u_j^{(m)})$  como a medida de eficiência para a  $j^{th}$  firma na  $m^{th}$  replicação da cadeia de MCMC. E colocando os  $e_j^{(m)}$  em ordem para  $j = 1, \dots, N$ , o rank associado à  $j^{th}$  firma na  $m^{th}$  replicação é a posição ocupada por  $e_j^{(m)}$  na ordenação. A Figura 3.3 mostra o gráfico das medianas da distribuição a posteriori dos ranking baseado no modelo proposto por Fernández et al. (2000) versus as medianas da distribuição a posteriori dos *ranking* utilizando o modelo de fronteira de produção fatorial.

Figura 3.1: Densidade a posteriori para todos os parâmetros do modelo de fronteira de produção fatorial usando o conjunto de dados simulados

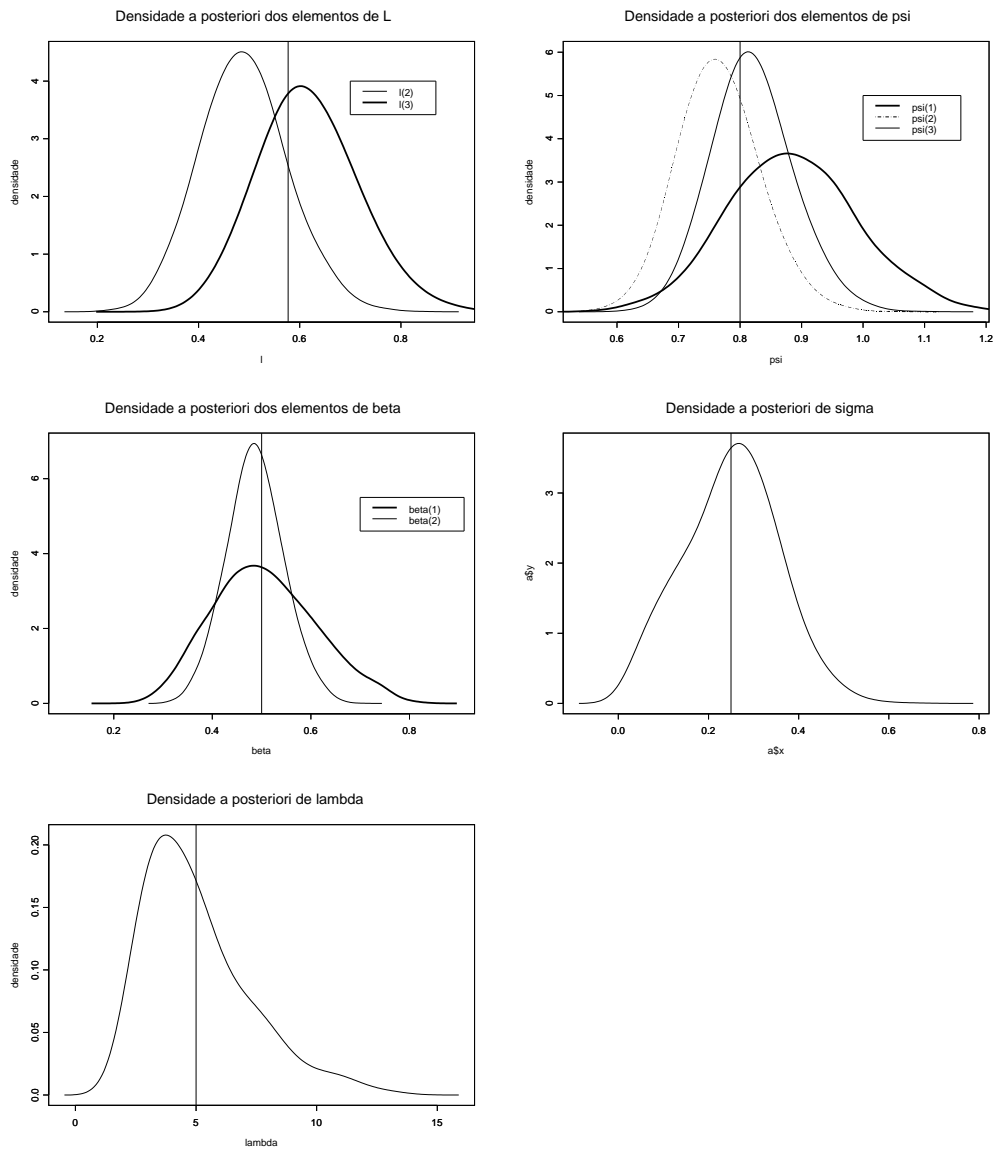




Figura 3.2: Densidade a posteriori para todos os parâmetros do modelo proposto por Fernández et al. (2000) usando o conjunto de dados simulados

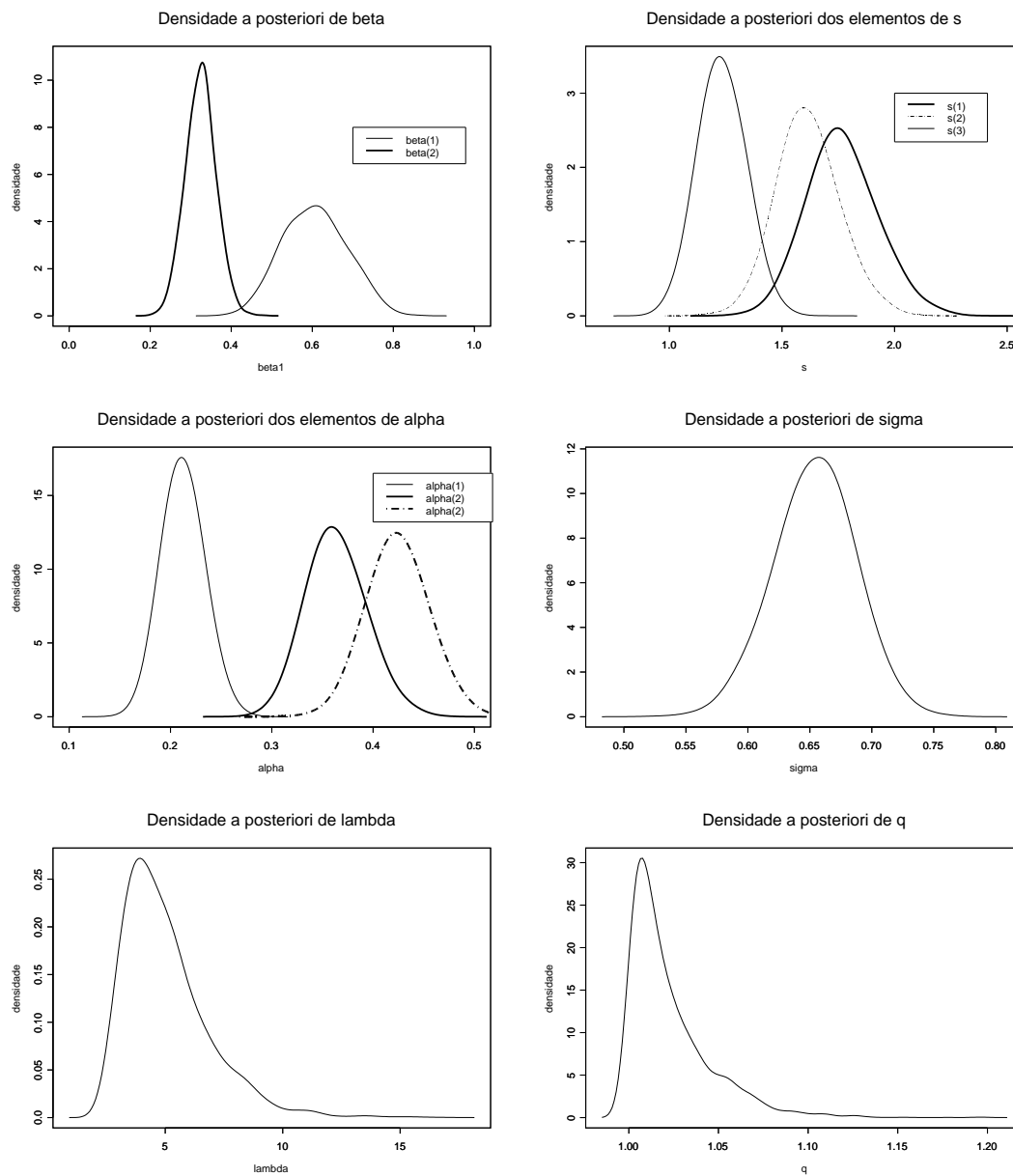
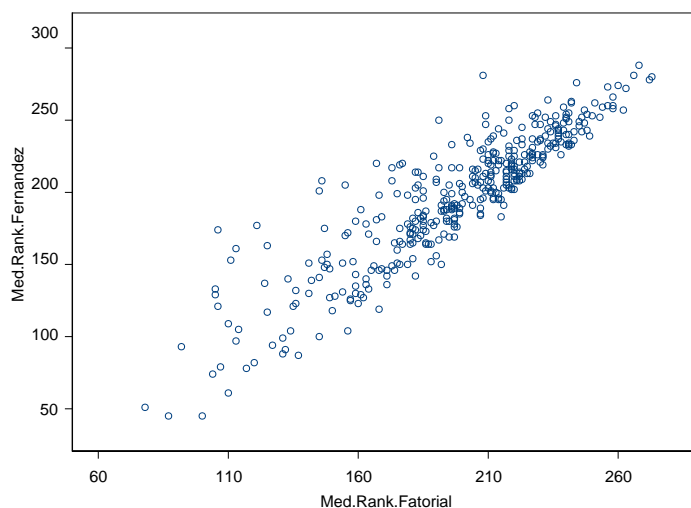


Figura 3.3: Mediana da distribuição a posteriori dos Rank segundo o modelo proposto por Fernández et al. (2000) vs. a mediana da distribuição a posteriori dos Rank usando o modelo de fronteira de produção fatorial



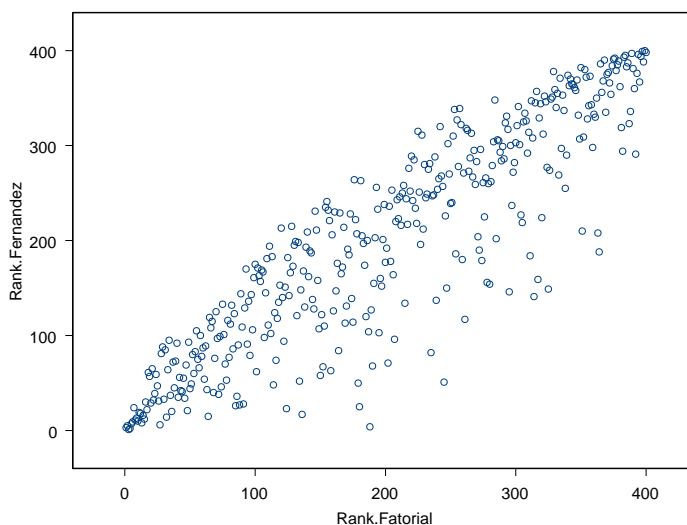
Uma outra maneira de se obter um *ranking* das firmas associado à medida de eficiência, é ordenando as médias a posteriori das eficiências ( $\exp[-u_i]$  para  $i = 1, \dots, N$ ). Assim o rank associado à  $j^{th}$  firma é a posição ocupada pela média a posteriori de  $e_j$  na ordenação. A Figura 3.4 mostra o gráfico do *ranking* das firmas utilizando o modelo proposto por Fernández et al. (2000) versus o *ranking* das firmas utilizando o modelo de fronteira de produção fatorial.

Pelas Figuras 3.3 e 3.4 observa-se uma relação linear entre os rankings das firmas, ou seja, a ordenação das firmas baseada na medida de eficiência utilizando o modelo proposto, foram similares com os resultados encontrados utilizando a abordagem proposta por Fernández et al.(2000).

### 3.5.2 Dados Reais

Agora vamos fazer uma aplicação para um conjunto de dados que foram usados em Souza et al. (1996). O conjunto de dados analisados é referente a produção

Figura 3.4: *Ranking* das firmas segundo o modelo proposto por Fernández et al. (2000) vs. o *ranking* das firmas segundo o modelo fronteira de produção fatorial



em pesquisas agrícolas de  $N = 34$  unidades (DMUs) centrais de pesquisas da Embrapa no ano de 1996. O conjunto de dados contém  $p = 3$  *outputs* (produção científica, produção de publicações técnicas, desenvolvimento de tecnologia, produtos, e processos e difusão de tecnologias e imagem), 3 *inputs* (pessoal, custo operacional e capital).

Neste estudo não temos como objetivo principal contribuir com a literatura de eficiência da produção em pesquisas agrícolas e conseqüentemente não vamos entrar em detalhes particulares dos dados. A pessoa interessada deve recorrer a Souza et al. (1996) para mais detalhes sobre os dados.

A Tabela 3.3 apresenta as médias, os desvios padrão, as medianas e os percentis de 2.5 e 97.5 a posteriori de todos os parâmetros e a distribuição da eficiência de três firmas para o modelo de fronteira de produção fatorial. Estas três firmas foram escolhidas baseado numa medida "grosseira" do tamanho do output. Isto é, foram simplesmente somados todos os três outputs adquirindo uma medida de um output agregado. Logo as três firmas escolhidas tem os níveis mínimo,

mediana e máxima do output agregado, respectivamente.

Tabela 3.3: Sumário da distribuição a posteriori para todos os parâmetros do modelo de fronteira de produção fatorial usando o conjunto de dados da Embrapa

	Media	d.p.	mediana	2.5%	97.5%
$l_1$	1	-	-	-	-
$l_2$	0,1856	0,1987	0,1845	-0,2008	0,5807
$l_3$	0,5756	0,1098	0,5717	0,3753	0,8051
$l_4$	0,6709	0,1660	0,6676	0,3496	1,0050
$\psi_1$	0,1055	0,0706	0,0896	0,0193	0,2800
$\psi_2$	0,9792	0,2475	0,9441	0,6038	1,5620
$\psi_3$	0,2195	0,0630	0,2111	0,1209	0,3654
$\psi_4$	0,6349	0,1679	0,6094	0,3834	1,0350
$\beta_1$	0,1833	0,1809	0,1848	-0,1681	0,5339
$\beta_2$	0,0871	0,3539	0,0857	-0,6124	0,7697
$\beta_3$	1,0120	0,3658	1,0080	0,3102	1,7400
$\beta_4$	0,3830	0,3259	0,3751	-0,2179	1,0400
$\sigma^2$	0,1793	0,1266	0,1586	0,0104	0,4753
$\lambda$	3,0940	1,7930	2,5730	1,2590	8,1120
Eff. da firma min.	0,3967	0,2667	0,3139	0,0716	0,9582
Eff. da firma med.	0,6865	0,2042	0,7047	0,2820	0,9876
Eff. da firma max.	0,8671	0,1143	0,8980	0,5747	0,9960

As densidades marginais a posteriori para todos os parâmetros do modelo de fronteira fatorial são apresentados na Figura 3.5.

A Figura 3.6 apresenta o histograma do *ranking* da eficiências para a melhor firma (rho[34]), pior firma (rho[4]) e uma firma mediana (rho[14]) segundo o

modelo de fronteira de produção fatorial e, também, usando o modelo proposto por Fernández et al. (2000) para o conjunto de dados da Embrapa.

Pela Figura 3.6 observamos que os dois modelos conseguiram diferenciar da mesma forma as firmas mais eficientes das menos eficientes.

Os resultados obtidos envolvendo dados simulados e reais indicou um bom desempenho do modelo de fronteira de produção para múltiplos-*outputs* proposto nesta dissertação. E também mostrou que a ordenação das firmas baseada na medida de eficiência utilizando o modelo de fronteira de produção fatorial, foram similares com os resultados encontrados utilizando a abordagem proposta por Fernández et al. (2000).

Figura 3.5: Densidade a posteriori para todos os parâmetros do modelo de fronteira de produção fatorial usando o conjunto da Embrapa

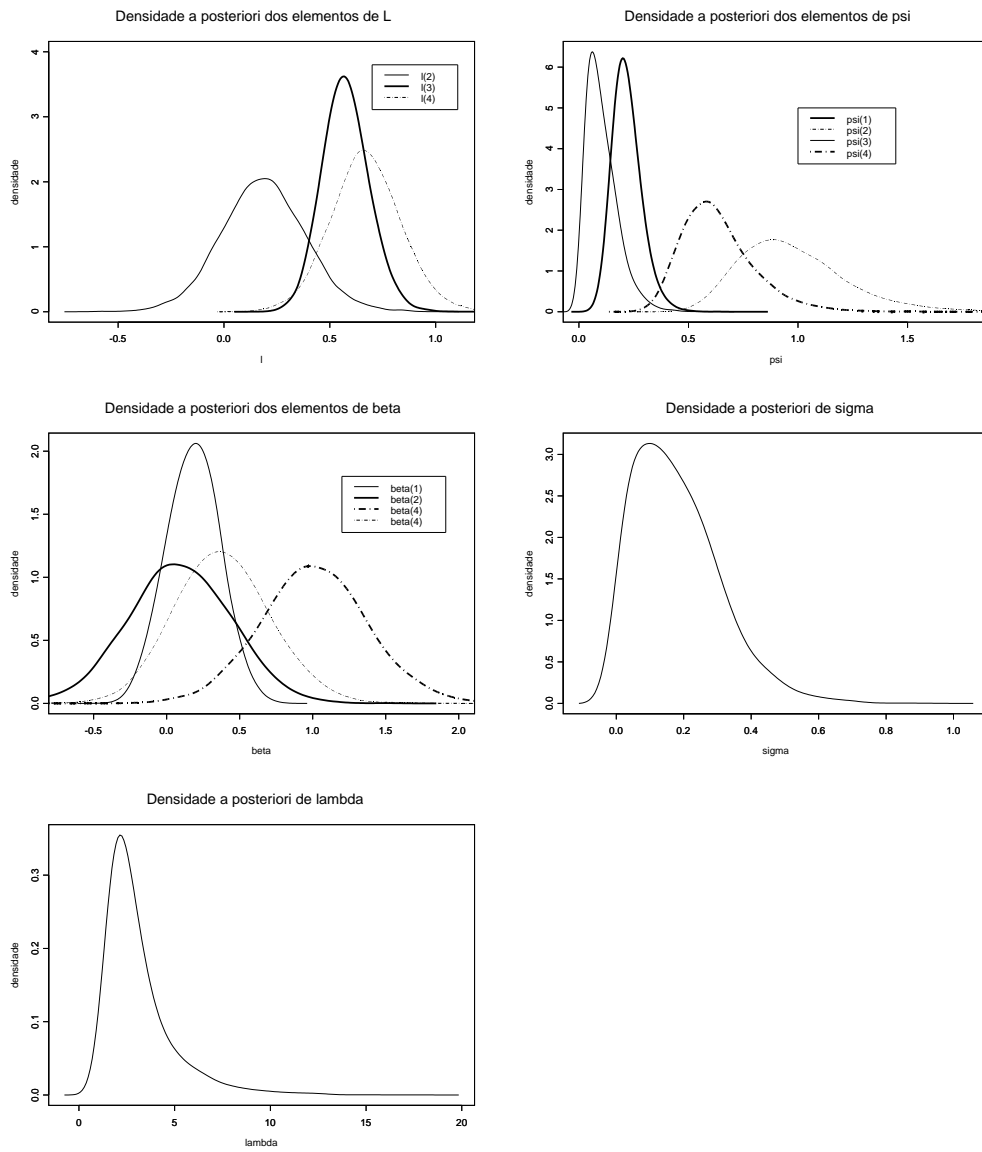
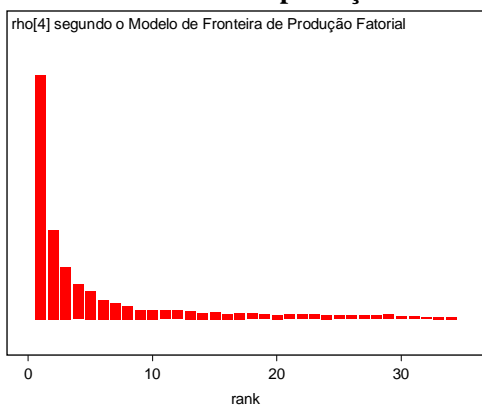


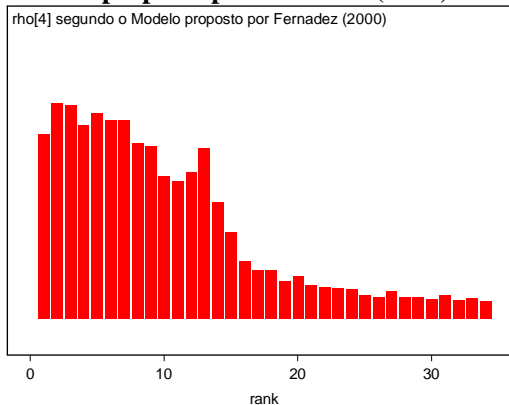
Figura 3.6: Histograma do *ranking* da eficiência para da melhor firma (rho[34]), pior firma (rho[4]) e uma firma mediana (rho[14]) segundo o modelo de fronteira de produção fatorial e o modelo proposto por Fernández et al. (2000) usando o conjunto de dados da Embrapa

**rho[4] -> Pior Firma      rho[17] -> Firma mediana      rho[34] -> Melhor Firma**

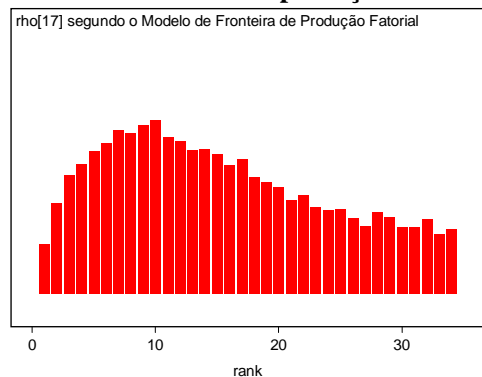
**Modelo de fronteira de produção fatorial**



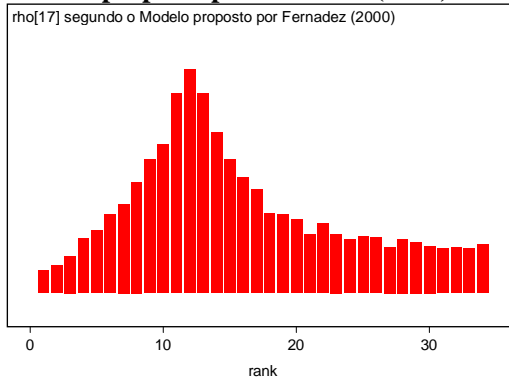
**Modelo proposto por Fernandez (2000)**



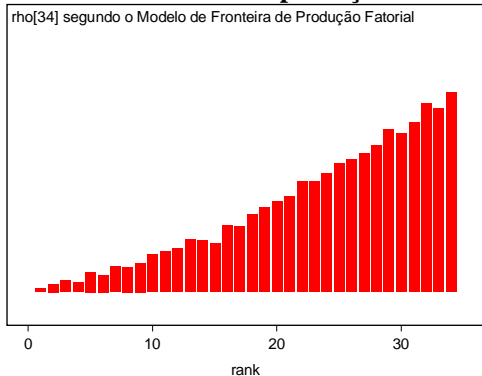
**Modelo de fronteira de produção fatorial**



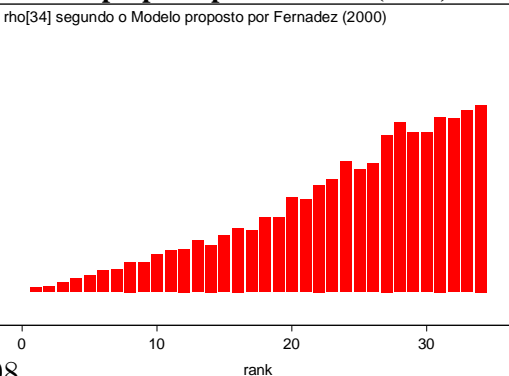
**Modelo proposto por Fernandez (2000)**



**Modelo de fronteira de produção fatorial**



**Modelo proposto por Fernandez (2000)**



# Capítulo 4

## Conclusão e Extensões

O objetivo dessa monografia era cuidar de alguns problemas relacionados com o modelo de fronteira de produção estocástica. As três questões tratadas nesta dissertação foram:

- O modelo de fronteira normal-gama apresenta o problema de que alguns parâmetros tem a distribuição condicional a posteriori não conhecida e não log-côncava.
- A maioria dos estudos limitam-se em utilizar a função de produção Cobb-Douglas para estimar as fronteiras de produção.
- Em muitas situações as firmas produzem mais de um *output* e, assim sendo, modelos de fronteira de produção são aplicados somente em combinações arbitrários dos vários *outputs*.

Segundo os resultados obtidos no primeiro capítulo, pode-se concluir através dos dados artificiais e reais, com relação ao modelo de fronteira normal-gama, que o algoritmo de Slice Sampling teve um bom desempenho para amostrar de uma distribuição condicional que não é log-côncava. Vimos também, que o método de Slice Sampling teve resultados parecidos com o algoritmo de Metropolis-Hasting,



porém foi necessário um esforço menor na implementação do algoritmo, além de um menor tempo de aquecimento. Por último, através de dados artificiais vimos que o algoritmo do Metropolis Otimizado (Tsionas, 2000), não teve uma performance satisfatória em certos casos. Observou-se, ainda, que para alguns dados simulados o algoritmo proposto por Tsionas (2000) gerou resultados fora da região esperado do parâmetro.

A comparação dos vários modelos estudados apresentado no capítulo 2, identificou que o modelo que se mostrou mais apropriado, foi o fronteira de produção com termo de ineficiência lognormal e usando a forma funcional CES (Constant Elasticity of Substitution) ou GPF (Generalized Production Functions). Notou-se, também, que o modelo de fronteira normal-lognormal com apenas um parâmetro e utilizando qualquer forma funcional, conseguiu diferenciar melhor as firma mais eficientes das menos eficientes. As comparações foram feitas utilizando uma aproximação do pseudo fator de Bayes (Geisser e Eddy, 1979) e Deviance Information Criterion (Spiegelhalter et. al. , 2001).

A ilustração empírica, apresentada no terceiro capítulo, envolvendo dados simulados e reais indicaram uma bom desempenho do modelo de fronteira de produção para múltiplos-*outputs* proposto nesta dissertação. O exemplo com dados artificiais também mostrou que a ordenação das firmas baseada na medida de eficiência utilizando o modelo proposto, foram similares com os resultados encontrados utilizando a abordagem proposta por Fernandez et al.(2000).

Ao longo do desenvolvimento do estudo surgiram algumas idéias de possíveis extensões desta dissertação. Estas idéias são apresentadas a seguir como sugestões para serem incluídas em pesquisas futuras.

No caso da fronteira Normal-Gama, este estudo utilizou o algoritmo de Slice Sampling (Neal, 1997) univariado para amostrar de cada elemento de  $\mathbf{u} = (u_1, \dots, u_N)'$ . Uma extensão a ser considerado é utilizar-se o método de Slice Sampling multi-

variado para amostrar simultaneamente os  $u's$ .

Um outra extensão interessante para o modelo proposto no capítulo 3, seria modelar a fronteira de produção para múltiplos-*outputs* em dados de painel, onde as eficiência seriam modelas para cada firma ao longo do tempo.

A metodologia utilizada nessa dissertação apresentou soluções para as questões mencionadas no início deste capítulo, fortemente viabilizado pelo uso da abordagem Bayesiana. Esta análise baseada na noção central de medida de eficiência relativa pode vir a ser um instrumento importante que auxilie a tomada de decisão.

# Referências Bibliográficas

- [1] Adams, R., Berger, A., Sickles, R., 1999. Semiparametric approaches to stochastic panel frontiers with applications in the applications in the banking industry. *Journal of Business and Economic Statistics* 17, 349-358.
- [2] Aigner, D., C. Lowell, e P. Schmidt (1977) - Formulation and Estimation of Stochastic Frontier Production Function Models, *Journal of Econometrics*, 6, 21-37.
- [3] Aguilar, O. e M. West (2001) Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics* 18, 338-357.
- [4] van den Broeck, J., G. Koop, J. Osiewalski, e M.F.J. Steel (1994) Stochastic frontier models: A Bayesian perspective, *Journal of Econometrics* 61, 273-303,
- [5] Brooks S P and Gelman A (1998) Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*. 7, 434-455.
- [6] Broeck, J. van den, Koop, G., Osiewalski, J., e Steel, M. F. J. (1994) - Stochastic frontier models - A Bayesian perspective, *Journal of Econometrics*, 61, 273-303.

- [7] Christensen, L. R. e W. H. Greene. 1976, Economies of scale in U.S. electric power generation. *Journal of Political Economy* 84. 655-676.
- [8] Fare, R., Primont, D., 1990. A distance function approach to multioutput technologies. *Southern Economic Journal* 56, 879-891.
- [9] Fernandez, C., Osiewalski, J. e Steel, M. F. J.(1997) - On the use of Panel Data in Stochastic Frontier Models with Improper Priors, *Journal of Econometrics*, 79, 169-193.
- [10] Fernandez, C., Osiewalski, J. e Steel, M. F. J. (2000) - A Bayesian Analysis of multiple-output production frontiers, *Journal of Econometrics*. V.98, 47-79
- [11] Gamerman, D. (1997) *Markov Chain Monte Carlo*, Chapman-Hall.
- [12] Geisser, S. e Eddy, W. (1979) A predictive approach to model selection. *J. Am. Statist. Ass.*, 74, 153-160.
- [13] Gelfand, A. E. e Dey, D. K. (1994). Bayesian model choice: asymptotic and exact calculations. *Journal of the Royal Statistical Society, Series B*, 56, 501-14.
- [14] Gelfand, A. E. (1995). Model determination using sampling-based methods. Em *Markov Chain Monte Carlo in Practice* (eds.: W. R. Gilks, S. Richardson e S. J. Spiegelhalter), cap. 9. New York: Chapman and Hall.
- [15] Gelfand, A. E., Dey, D. K. e Chang H. (1992). Model Determination using Predictive Distributions with Implementation via Sampling-Based Methods. Em *Bayesian Statistics 4* (eds: J. M. Bernardo et al.), pp. 147-167. Oxford: University Press.
- [16] Gelman, A., and Rubin, D. (1992a) - Inference from Iterative Simulation using Multiple Sequences, *Statistical Science*, 7, 457-511.

- [17] Geweke, J.F. and Singleton, K.J. (1980) - Interpreting the likelihood ratio statistic in factor models when sample size is small. *Journal of the American Statistical Association*, 75, 133-137.
- [18] Geweke, J.F. and Zhou, G. (1996)- Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies*, 9, 557-587.
- [19] Gilks W. (1992) Derivative-free adaptive rejection sampling for Gibbs sampling. *Appl. Statist.*, **41** 337-348.
- [20] Goldstein, H., e Spiegelhalter, D. J. (1996) - Leaguer Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance, *Journal of the Royal Statistical Society, Serie A*, 159, 3, 385-443.
- [21] Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82, 711-732.
- [22] Greene, W. H. (1990) - A Gamma Distributed Stochastic Frontier Model, 46, 141-163.
- [23] Hastings, W. K. (1970) Monte Carlo sampling methods using *Markov chain Monte Carlo with applications*. *J. Am. Statist. Ass.*, **93**, 585-595.
- [24] Kass, R. E. e Raftery, A. E. (1995) - Bayes Factor, *JASA*, 90, n. 430. 773-7796
- [25] Kumbhakar, S., 1987 - The specification of technical and allocative inefficiency of multi-product firms in stochastic production and profit frontiers. *Journal of Quantitative Economics* 3, 213-223.
- [26] Koop, G., Steel, M. F. J. e Osiewalski, J.(1995) -Posterior Analysis of Stochastic Frontier Models Using gibbs Sampling, *Computational Statistics*, 10, 353-373

- [27] Koop, G., Osiewalski, J. e Steel, M. F. J. (1997) - Bayesian Efficiency Analysis Through Individual Effects: Hospital cost frontier, *Journal of Econometrics*, 7, 77-105.
- [28] Lopes, H. (2000) Bayesian Analysis in Latent Factor and Longitudinal Models. Ph. D. thesis, Duke University.
- [29] Lothgren, M., 1997. Generalized stochastic frontier production models. *Economics Letters* 57, 255-259.
- [30] Meeusen, W. e J. van den Broeck (1977) Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 8, 435-444.
- [31] Metropolis, N. Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. e Teller, E. (1953) Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087-1091.
- [32] Migon, S., H. (2002) - A Bayesian Approach to Stochastic Production Frontier, Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro.
- [33] Neal R (1997) - Markov chain Monte Carlo methods based on ‘slicing’ the density function, Department of Statistics, University of Toronto.
- [34] Nerlove, M., 1963, Return to scale in electricity supply, in: C. F. Christ, ed., *Measurement in economics* (Stanford University Press, Stanford, CA).
- [35] Osiewalski, J. e Steel, M. F. J. (1998) - Numerical Tools for the Bayesian Analysis of Stochastic Frontier Models, *Journal of Productivity Analysis*, 10, 103-117.

- [36] Powell, A., Gruen, F., 1968. The constant elasticity of transformation production frontier and linear supply system. *International Economic Review* 9, 315-328.
- [37] Press, S.J. (1982)- *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference* (2nd edition). New York: Krieger.
- [38] Stevenson, R. (1980) Likelihood Functions for Generalized Stochastic Frontier Estimation, *Journal of Econometrics*, 13, 1980, pp. 57-66.
- [39] Spiegelhalter, D. J., Thomas, A., Best, N. G. e Gilks, W R (1996). *BUGS: Bayesian inference Using Gibbs Sampling. Manual (Version II)*. MRC Biostatistics Unit, Cambridge, UK.
- [40] Spiegelhalter, D. J., Thomas, A., Best, N. G. e Lunn, D. (1999). *WinBUGS Version 1.4 User Manual*, MRC Biostatistics Unit
- [41] Spiegelhalter, D. J., Best, N. G. Carlin, B. P. e van der Linde, A. (2002), "Bayesian measure of model complexity and fit", *Journal of the Royal Statistical Society, Series B*, 64, Parte 3, forthcoming.
- [42] Souza e Silva, G., Alves, E., Avila, A. F. D. (1996) - *Technical Efficiency of Production in Agricultural Research*.
- [43] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B*, 36, 111-147.
- [44] Tsionas, E., G. (2000) - Full Likelihood Inference in Normal-Gamma Stochastic Models, *Journal of Productivity Analysis*, 13, 183-205
- [45] Wynn, R. F., e Holden K. (1974) - *An Introduction to Applied Econometric Analysis*, The Macmillan Press Ltd.

- [46] Zellner, A. e N. Revankar (1970), Generalized Production Functions,"The Review of Economic Studies, 36, 241-250.
- [47] Zellner, A., 1971, An introduction to Bayesian inference in econometrics (Wiley, New York, NY).