

RAIMUNDO CARDOSO DE OLIVEIRA NETO

**REGRESSÃO QUANTÍLICA SOB ENFOQUE BAYESIANO COMO
ALTERNATIVA NO AJUSTE DA EFICIÊNCIA TÉCNICA:
UMA APLICAÇÃO PARA A AGRICULTURA FAMILIAR BRASILEIRA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS - BRASIL
2018

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

O48r
2018
Oliveira Neto, Raimundo Cardoso de, 1991-
Regressão quantílica sob enfoque bayesiano como
alternativa no ajuste da eficiência técnica : uma aplicação para a
agricultura familiar brasileira / Raimundo Cardoso de Oliveira
Neto. – Viçosa, MG, 2018.
x, 65f. : il. (algumas color.) ; 29 cm.

Inclui apêndices.

Orientador: Ana Carolina Campana Nascimento.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 46-50.

1. Teoria bayesiana de decisão estatística. 2. Métodos de
simulação. 3. Agricultura familiar. I. Universidade Federal de
Viçosa. Departamento de Estatística. Programa de
Pós-Graduação em Estatística Aplicada e Biometria. II. Título.

CDD 22. ed. 519.542

RAIMUNDO CARDOSO DE OLIVEIRA NETO

**REGRESSÃO QUANTÍLICA SOB ENFOQUE BAYESIANO COMO
ALTERNATIVA NO AJUSTE DA EFICIÊNCIA TÉCNICA: UMA
APLICAÇÃO PARA A AGRICULTURA FAMILIAR BRASILEIRA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

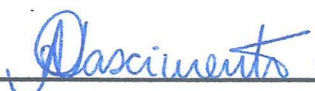
APROVADA: 23 de fevereiro de 2018.



Fernanda Maria de Almeida



Camila Ferreira Azevedo
(Coorientadora)



Ana Carolina Campana Nascimento
(Orientadora)

All models are wrong, but some are useful.

George Box

Agradecimentos

- Aos meus pais, Francisco e Maria, pelas bases que formam meu caráter e moldam minha pessoa. Em especial ao meu pai, por acreditar que os estudos levariam seus filhos a uma vida melhor e me apoiado em todos os momentos, pessoa a qual eu tenho como exemplo na vida;
- Aos meus irmãos, Antonio, Carlos e João, por sempre fazerem-se presentes, mesmo quando estamos separados por quilômetros;
- À minha esposa, Jennyfer, por aceitar fazer parte da minha vida, pelo apoio, paciência, carinho e dedicação, pessoa pela qual alimento um carinho especial;
- À minha Orientadora, Prof^a. Ana Carolina, por toda a orientação durante a produção deste, e outros, trabalhos e pela pronta disponibilidade em ajudar sempre que precisei;
- Aos demais integrantes do grupo de estudos LICAE pelas tardes de sexta em que os aprendizados foram além da Estatística, em especial aos professores Moysés e Camila, os quais também são meus coorientadores, pelos conselhos;
- Ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, professores e funcionários, pelos ensinamentos repassados e disponibilidade em sanar questões que surgiram pelo caminho;
- Aos amigos que fiz no PPESTBIO, pelos momentos de descontração e discussão;
- Ao amigo de graduação, mestrado e colega de apartamento, Ithalo, pelo companherismo e discussões sobre variados temas;
- Finalmente, à CAPES, pelo apoio financeiro fornecido para a realização do mestrado.

A todos, o meu muito obrigado.

Biografia

RAIMUNDO CARDOSO DE OLIVEIRA NETO, filho de Francisco Antonio da Silva e Maria Araujo de Oliveira, nasceu em Cocal dos Alves, Piauí, no dia 24 de janeiro de 1991.

Ingressou no curso de Bacharelado em Estatística em março de 2011, na Universidade Federal do Piauí, em Teresina-PI, graduando-se em fevereiro de 2016.

Em março de 2016 iniciou o Mestrado no Programa de Pós-Graduação em Estatística Aplicada e Biometria, na Universidade Federal de Viçosa, submetendo-se à defesa de dissertação em 23 de fevereiro de 2018.

Sumário

Lista de Figuras	vii
Lista de Tabelas	viii
Resumo	ix
Abstract	x
1 Introdução	1
2 Referencial Teórico	5
2.1 Conceitos introdutórios	5
2.2 Metodologias para ajustar a eficiência técnica	6
2.2.1 Análise Envoltória de Dados	6
2.2.2 Análise de Fronteira Estocástica	10
2.2.3 Regressão Quantílica	14
2.3 O enfoque Bayesiano para Regressão Quantílica	18
2.4 Alguns pontos em Inferência Bayesiana	19
3 Material e Métodos	23
3.1 Estudo de simulação	23
3.1.1 Simulação dos dados	23
3.1.2 Modelos ajustados	25
3.2 Aplicação aos dados da Agricultura Familiar	30
3.2.1 Descrição do conjunto de dados	30
3.2.2 Modelo ajustado	31
4 Resultados e Discussões	32
4.1 Resultados da simulação	32
4.2 Aplicação aos dados da Agricultura Familiar	36

4.2.1	Análise exploratória dos dados	36
4.2.2	Análise do modelo ajustado	38
4.2.3	Análise da eficiência técnica das microrregiões brasileiras . .	42
5	Considerações Finais	44
	Referências Bibliográficas	46
	Apêndice A: Eficiências técnicas ajustadas <i>versus</i> simuladas	51
	Apêndice B: <i>Traceplots</i> dos parâmetros da fronteira ajustada na aplicação aos dados da agricultura familiar brasileira.	53
	Apêndice C: Eficiências técnicas ajustadas para as microrregiões brasileiras	54
	Apêndice D: Códigos R e JAGS	59

Lista de Figuras

Figura 4.1:	EQM da eficiência técnica ajustada pelos modelos em estudo em relação à eficiência técnica simulada nas 300 repetições	33
Figura 4.2:	Coefficiente de Correlação de Pearson entre eficiência técnica simulada e ajustada pelos modelos nas 300 repetições.	34
Figura 4.3:	Concordância na classificação das DMUs menos eficientes.	35
Figura 4.4:	Concordância na classificação das DMUs mais eficientes. .	36
Figura 4.5:	Histogramas com estatísticas descritivas dos dados da agricultura familiar.	37
Figura 4.6:	Efeito dos insumos na produção da agricultura familiar, condicionado ao nível de produção da microrregião, ajustado por regressão quantílica frequentista.	38
Figura 4.7:	Densidades <i>a posteriori</i> da fronteira de produção da agricultura familiar.	41
Figura 4.8:	Histograma das eficiências técnicas ajustadas pela regressão quantílica no quantil 0,95	42
Figura 4.9:	Distribuição da eficiência técnica da agricultura familiar obtida por meio de regressão quantílica bayesiana	43

Lista de Tabelas

Tabela 4.1:	Resultados dos testes de convergência de Gelman e Rubin e tamanho amostral efetivo no modelo de regressão quantílica bayesiana.	39
Tabela 4.2:	Estimativas e Erros padrão dos coeficientes da regressão quantílica sob enfoque bayesiano ajustada no quantil 0,95.	40

Resumo

OLIVEIRA NETO, Raimundo Cardoso de, M.Sc., Universidade Federal de Viçosa, fevereiro de 2018. **Regressão quantílica sob enfoque bayesiano como alternativa no ajuste da eficiência técnica: uma aplicação para a agricultura familiar brasileira.** Orientadora: Ana Carolina Campana Nascimento. Coorientadores: Moysés Nascimento e Camila Ferreira Azevedo.

O presente trabalho tem como objetivo propor o uso de Regressão Quantílica sob o enfoque bayesiano para ajustar a eficiência técnica de unidades produtivas. Para tanto, é fornecida uma breve revisão compreendendo os primeiros trabalhos relevantes para a definição de eficiência técnica, métodos tradicionais utilizados para ajustar esta medida, e as propostas de Regressão Quantílica encontradas na literatura. Foi realizado um estudo de simulação consistindo do ajuste da eficiência por Análise Envoltória de Dados (DEA), Análise de Fronteira Estocástica (SFA) e Regressão Quantílica frequentista (RQ) e bayesiana (BRQ), em conjuntos de dados simulados. Os dados foram simulados por um modelo de SFA abrangendo cenários com e sem quebra de suposições quanto a homogeneidade das unidades produtivas e diferentes suposições de dominância do erro por parte de fatores aleatórios. Verificou-se que a Regressão Quantílica bayesiana fornece bom ajuste da eficiência, com relação a viés e classificação das unidades produtivas, inclusive em cenários considerados ideais para SFA e DEA, mesmo sem ter usado informação *a priori* relevante. Além disso, a metodologia foi aplicada ao estudo da eficiência da agricultura familiar brasileira, a nível de microrregião. Observou-se que a eficiência técnica média das microrregiões com base no quantil 0,95 é 41,09% com desvio padrão de 25,26%, tendo a região Nordeste a menor média de eficiência. Observou-se também, que nas microrregiões com melhores práticas de produção a mão-de-obra é o insumo mais importante, seguido de máquinas e implementos agrícolas, e ainda, que a produção destas microrregiões não responde de forma significativa à mudanças na quantidade de área utilizada. Assim, concluiu-se que a regressão quantílica sob enfoque bayesiano caracteriza-se como uma alternativa interessante para ajustar a eficiência técnica.

Abstract

OLIVEIRA NETO, Raimundo Cardoso de, M.Sc., Universidade Federal de Viçosa, February, 2018. **Bayesian quantile regression as an alternative to technical efficiency estimation: an application to family farming in Brazil.** Advisor: Ana Carolina Campana Nascimento. Co-Advisors: Moysés Nascimento and Camila Ferreira Azevedo.

This work proposes the use of Bayesian Quantile Regression to estimate the technical efficiency of productive units. For such, it is given a short review comprising from the initial relevant works for the definition of technical efficiency, traditional methods for estimating such a measure, and the Quantile Regression proposals found in the literature. Besides that, a simulation study was conducted consisting of the estimation of efficiency by Data Envelopment Analyzes (DEA), Stochastic Frontier Analyzes (SFA) and Quantile Regression, frequentist (QR) and bayesian (BQR). The datasets were simulated by a SFA model with and without the homocedasticity assumption about the productive units as well as different assumptions of error dominance by random noise. It was verified that Bayesian Quantile Regression estimates the efficiency consistently, considering bias and units ranking, including in scenarios considered ideal for DEA and SFA, although no relevant a priori information was used. As an illustration, the efficiency of family farming, at microregion level, in Brazil is estimated by Bayesian Quantile Regression. The mean efficiency in the country was 41.09%, based on the 0.95 quantile, with standard deviation of 25.26%, while the Northeastern region had the highest concentration of microregions with lower levels of efficiency. In addition, the microregions with best practices have the labor as the most important input, followed by machinery, and their production does not change significantly in response to changes in area used for production. Thus, it can be concluded that Bayesian regression quantile is an interesting alternative to estimate efficiency levels.

Capítulo 1

Introdução

A análise de eficiência tem se mostrado uma ferramenta bastante útil em diversos setores públicos e privados. Segundo Coelli *et al.* (2005), isto se deve ao fato de a mesma poder ser utilizada em diferentes níveis de agregação. Isto é, pode-se estudar a eficiência de uma indústria ao longo do tempo ou mesmo entre regiões geográficas. Com esta análise, são obtidas medidas da performance das unidades produtivas (índices ou escores de eficiência), que servem como ponto de partida para ações no intuito de melhorar o desempenho das unidades que estão sendo avaliadas.

O estudo da eficiência iniciou-se na década de 50, tendo como principal ponto de partida os trabalhos de Koopmans (1951) e Debreu (1951). Koopmans (1951) definiu como eficiente uma unidade produtiva em que somente é possível aumentar a produção de um produto se houver diminuição na produção de outro produto, ou, equivalentemente, só é possível diminuir o uso de um insumo se houver aumento no uso de outro insumo. Debreu (1951), por sua vez, propôs o coeficiente de utilização de recursos, como a primeira tentativa de medir a eficiência de uma unidade produtiva.

Farrell (1957) aprimorou os conceitos propostos anteriormente e concluiu que a eficiência econômica de uma unidade produtiva consiste da combinação de eficiência técnica e eficiência alocativa. A eficiência técnica trata-se da capacidade de produzir o máximo possível a partir de determinada quantidade de insumos, ao passo que a eficiência alocativa consiste em alocar os insumos da produção em proporções ótimas considerando-se os preços dos mesmos.

Para a realização da análise de eficiência, é necessário ainda uma metodologia para a obtenção dos níveis de eficiência. No final da década de 70 sugeriram as duas principais técnicas para o ajuste de tais níveis: a Análise de Fronteira Estocástica (SFA - *Stochastic Frontier Analysis*), proposta por Aigner *et al.* (1977) e Meeusen e van den Broeck (1977), de forma independente e paralela, e a Análise

Envoltória de Dados (DEA - *Data Envelopment Analysis*) por Charnes *et al.* (1978). Essas técnicas baseiam-se em princípios diferentes com suas vantagens e desvantagens.

A DEA ajusta uma fronteira não paramétrica, via programação matemática, isto é, sem a necessidade suposições probabilísticas ou sobre formas algébricas, que modela o máximo alcançável de produção para determinada quantidade de insumos. Por outro lado, a SFA modela a produção de forma paramétrica, com parâmetros estimados pelo método da máxima verossimilhança, e os resíduos do modelo são decompostos em erro aleatório e ineficiência. Portanto, a suposição de distribuições de probabilidade, bem como a suposição de uma forma algébrica para descrever a função produção se fazem necessárias na SFA.

Estas duas técnicas abriram campos de pesquisa em análise eficiência os quais recebem extensões até os dias atuais. Para DEA têm sido propostas alterações para considerar retornos variáveis de escala (BANKER *et al.*, 1984); introdução de *bootstrap* para corrigir o viés das estimativas de eficiência (SIMAR; WILSON, 1998); modelos híbridos com técnicas de aprendizado de máquina, como máquina de vetores suporte (FARAHMAND *et al.*, 2014) dentre outros. Por outro lado, na SFA surgiram, por exemplo, diferentes propostas de distribuição de probabilidade para a ineficiência; extensões que permitem modelar a ineficiência como função de variáveis da unidade produtiva e/ou do tempo (BATTESE; COELLI, 1988; BATTESE; COELLI, 1992; BATTESE; COELLI, 1995); estimação bayesiana (BROECK *et al.*, 1994).

Apesar disso, em amostras com heterogeneidade é possível que uma única fronteira de produção não seja adequada para explicar a produção de todas as unidades produtivas, como é feito na SFA e DEA. Assim, é possível que a tecnologia de produção de unidades produtivas próximas à média de produção, ou próximas ao máximo teórico da amostra, não seja adequada para que as demais unidades produtivas sejam comparadas com ela para que a partir dessa comparação sejam obtidos os níveis de eficiência.

Assim, mais recentemente, alguns autores recomendam a regressão quantílica, modelo proposto por Koenker e Bassett (1978), para ajustar a eficiência técnica. A regressão quantílica (RQ), ao passo que mantém a necessidade de especificação de forma funcional, um ponto criticado por alguns autores em relação à SFA (RASMUSSEN, 2012), fornece ao pesquisador a habilidade de ajustar qual-

quer quantil da produção. Isso permite escolher qual quantil considerar para o ajuste da eficiência (mais extremo que SFA ou menos extremo que DEA), bem como avaliar se o efeito de insumo é o mesmo em unidades que produzem quantidades diferentes.

Aplicações de regressão quantílica no contexto de análise de eficiência podem ser encontradas na literatura para analisar o setor hoteleiro (BERNINI *et al.*, 2004), hospitalar (KNOX *et al.*, 2007), agricultura (KADITI; NITSI, 2010), bancário (BEHR, 2010) e até mesmo estudos de cunho mais teórico como o de Liu *et al.* (2008) e dados em painel (HSU *et al.*, 2017). Nestes estudos, a regressão quantílica foi comparada com DEA e SFA mostrando-se uma alternativa interessante, até mesmo em casos em que os dados haviam sido simulados por um modelo de fronteira estocástica.

Koenker e Bassett (1978) forneceram alguns resultados assintóticos a respeito dos estimadores da regressão quantílica. Tais resultados contribuíram para o desenvolvimento de testes propostos por Koenker e Bassett (1982) e Koenker e Machado (1999) para testar a igualdade dos parâmetros em diferentes quantis, significância de parâmetros e qualidade ajuste, testes que também são usados no contexto de análise de eficiência. Entretanto, estes testes requerem suposições que nem sempre podem ser garantidas, como afirma Hao e Naiman (2007).

É comum recorrer-se então ao uso de *bootstrap*. Múltiplos métodos de *bootstrap* existem, alguns estudados no contexto de regressão quantílica, como em Feng *et al.* (2011), Parzen *et al.* (1994), Galvao e Montes-Rojas (2015). Entretanto, a literatura sobre estes métodos em regressão quantílica no contexto de análise de eficiência é escassa, além de que o emprego de *bootstrap* em pequenas amostras, em especial se a distribuição do parâmetro for assimétrica, não fornece boas aproximações (CHERNICK; LABUDDE, 2011).

Neste sentido, segundo Bolstad (2007), a adoção do enfoque bayesiano para regressão quantílica no contexto de análise de eficiência traz benefícios sobre o enfoque frequentista como a possibilidade de fazer afirmações probabilísticas sobre os parâmetros da fronteira e os níveis de eficiência, sem recorrer a propriedades assintóticas. Além disso, a inclusão a informação *a priori* a cerca da fronteira em determinado quantil no processo de estimação poderia melhorar a qualidade das estimativas. Nos casos em que pequenas amostras são utilizadas, a eficiência poderia ser ajustada sem perda de precisão, uma vez que os resultados

da estimação bayesiana não dependem do tamanho amostral. Ainda, é possível obter a distribuição *a posteriori* da eficiência de cada unidade produtiva bem como para os parâmetros estudados em diferentes quantis de interesse e realizar testes sem recorrer a *bootstrap*.

Arroyo Hinostroza (2017) propôs um modelo de fronteira estocástica quantílica, em que é possível ajustar um modelo de fronteira estocástica a qualquer quantil de interesse. Todavia, não foram encontrados trabalhos aplicados em que tenha sido usado o modelo proposto pelo autor, inclusive por tratar-se de uma proposta muito recente.

Nesse sentido, este trabalho propõe usar o modelo clássico de regressão quantílica sob o enfoque bayesiano para ajustar a eficiência técnica. Múltiplas abordagens bayesianas para regressão quantílica foram propostas, dentre as quais, o trabalho de Yu e Moyeed (2001), em que é feito o uso da distribuição de Laplace Assimétrica, proposta para regressão quantílica por Koenker e Machado (1999). Apesar do modelo ainda não ter sido aplicado em estudos de análise de eficiência técnica, é possível encontrar trabalhos utilizando regressão quantílica bayesiana em dados longitudinais (YUAN; YIN, 2010), predição de níveis de ozônio (REICH *et al.*, 2011) e em modelos mistos não-lineares (WANG, 2012), por exemplo.

Dessa forma, este trabalho tem como objetivos:

1. realizar um estudo de simulação, como forma de verificar o desempenho da regressão quantílica bayesiana em ajustar a eficiência técnica; e
2. ajustar a eficiência técnica da agricultura familiar, a nível das microrregiões brasileiras usando a metodologia proposta.

Assim, no capítulo 2 são abordados o conceito de eficiência e alguns aspectos matemáticos dos modelos DEA, SFA, regressão quantílica e estimação bayesiana. No capítulo 3 são apresentados os materiais e métodos usados. No capítulo 4, são apresentados os resultados e discussões referentes ao estudo de simulação e à aplicação aos dados da agricultura familiar. Por fim, no capítulo 5, são apresentados considerações finais com relação aos resultados.

Capítulo 2

Referencial Teórico

2.1 Conceitos introdutórios

A eficiência técnica é uma medida que refere-se ao desempenho do processo de produção de uma unidade produtiva, também conhecida como Unidade Tomadora de Decisão (DMU). Farrell (1957) definiu eficiência técnica como a capacidade de uma DMU em produzir o máximo possível a partir de uma determinada quantidade de insumos¹. Esta medida é geralmente apresentada no intervalo de 0 a 1, de modo que DMUs com alta performance na produção têm nível de eficiência próximo a 1.

Associada à medida de eficiência técnica tem-se o que é conhecido como orientação da eficiência. A eficiência pode ser orientada a insumos, neste caso, pode-se dizer em quanto o uso de insumos pode ser diminuído para que a DMU seja eficiente (sem diminuir a produção), ou orientada a produto, caso em que pode-se dizer em quanto pode ser aumentada a produção para que a DMU seja eficiente (sem aumentar o uso de insumos). Caso a DMU opere com retornos constantes de escala a eficiência orientada a insumo e a produto são iguais (COELLI *et al.*, 2005).

Os retornos de escala, mencionados anteriormente, basicamente referem-se à maneira como a produção responde a um aumento no uso de todos os insumos. Quando aumentos na produção ocorrem na mesma proporção do aumento no uso de insumos, diz-se que há retornos constantes de escala. Quando a proporção do aumento na produção é maior que a ocorrida no uso de insumos, há retornos crescentes de escala. O caso oposto ao anterior constitui retornos decrescentes de escala (RASMUSSEN, 2012). Matematicamente, definindo a produção

¹Farrell (1957) também define a eficiência alocativa, como a capacidade de alocar os insumos em proporções ótimas considerando-se seus preços. Este tipo de eficiência não será abordada aqui.

como função dos insumos (x) tem-se (COELLI *et al.*, 2005):

$$f(kx) < kf(x) \Rightarrow \text{retorno decrescente} \quad (2.1)$$

$$f(kx) = kf(x) \Rightarrow \text{retorno constante} \quad (2.2)$$

$$f(kx) > kf(x) \Rightarrow \text{retorno crescente} \quad (2.3)$$

Para medir a eficiência técnica e a produtividade faz-se necessário informações das quantidades produzidas e quantidades de insumos utilizados. Uma medida de eficiência técnica orientada a produto comumente usada é (COELLI *et al.*, 2005):

$$ET_i = \frac{y_i}{\hat{y}_i}, \quad (2.4)$$

em que y_i é produção observada e \hat{y}_i é a produção esperada caso a i -ésima DMU seja eficiente. Para obtenção da produção esperada faz-se necessário algum método para modelar a produção. Além disso, em casos em que as DMUs produzem mais de um produto, é necessário alguma forma para criar um produto agregado para que se possa usar a expressão (2.4) no cálculo da eficiência.

Acrescenta-se que, segundo Farrell (1957), a eficiência é relativa ao grupo de DMUs estudadas e aos insumos e produtos considerados. Isto é, a introdução de outras DMUs, bem como alteração nos insumos e produtos considerados, pode alterar a eficiência de DMUs já inclusas.

2.2 Metodologias para ajustar a eficiência técnica

2.2.1 Análise Envoltória de Dados

A Análise envoltória de dados (DEA) foi proposta por Charnes *et al.* (1978) como alternativa para se obter a eficiência de DMUs. Trata-se de uma técnica não paramétrica cujo ajuste é feito via programação matemática na qual cria-se uma fronteira, também chamada de envelope, que engloba todas as DMUs não eficientes.

O desenvolvimento da DEA começa com o problema de otimização (2.5):

$$\left\{ \begin{array}{l} \text{maximizar } h_i = \frac{\omega' \mathbf{y}_i}{\delta' \mathbf{x}_i} \\ \text{sujeito a:} \\ \mathbf{Y}\omega - \mathbf{X}\delta \leq \mathbf{0} \\ \omega \geq \mathbf{0}, \delta \geq \mathbf{0} \end{array} \right. \quad (2.5)$$

em que \mathbf{y}_i e \mathbf{x}_i são, respectivamente, os vetores de produtos e insumos utilizados pela a i -ésima DMU e \mathbf{Y} e \mathbf{X} são, respectivamente, as matrizes de produtos e insumos da amostra. O problema (2.5) é resolvido para $i = 1, 2, \dots, n$, em que n é o tamanho da amostra. Os vetores ω e δ tratam-se de variáveis do problema de otimização em questão cujos valores ótimos constituem a solução do problema, e são encontrados na resolução do mesmo.

O problema (2.5) tenta maximizar a eficiência da i -ésima DMU com a condição que a eficiência desta e das demais DMUs não ultrapasse o valor 1. Porém, este problema tem infinitas soluções (COELLI *et al.*, 2005; WEN, 2015). Desta forma, por meio de reformulações e uso de teoria da dualidade em problemas de programação linear, opta-se por resolver o problema (2.6), ao invés de (2.5),

$$\left\{ \begin{array}{l} \text{minimizar } \theta_i \\ \text{sujeito a:} \\ -\mathbf{y}_i + \mathbf{Y}'\lambda \geq \mathbf{0} \\ \theta_i \mathbf{x}_i - \mathbf{X}'\lambda \geq \mathbf{0} \\ \lambda \geq \mathbf{0} \end{array} \right. \quad (2.6)$$

em que λ é um vetor de constantes $n \times 1$ e θ_i trata-se da eficiência da i -ésima DMU.

Em (2.6), θ_i pertence ao intervalo $(0, 1]$ e é tal que $1 - \theta_i$ representa o quanto a i -ésima DMU precisa diminuir no uso de insumos para que a mesma seja eficiente. Um valor $\theta_i = 1$ indica que a DMU é tecnicamente eficiente, dado que neste caso $1 - \theta_i = 0$, ou seja, não é necessário reduzir o uso de insumos pois ela já é tecnicamente eficiente. Isto é, θ_i refere-se a eficiência com orientação a insumos.

A eficiência com orientação a produto, por outro lado, é obtida por meio de pequena alteração em (2.6), o que leva a (2.7) (COELLI *et al.*, 2005). Agora, $\theta_i \in [1, \infty)$ e geralmente faz-se $ET_i = 1/\theta_i$ tal que $ET_i \in (0, 1]$. Dessa forma, ET_i

indica o quanto a atual produção representa no potencial de produção da DMU. Isto é, com uma eficiência de ET_i a i -ésima DMU produz apenas $ET_i\%$ do que poderia estar produzindo e poderia aumentar a produção em $(1 - ET_i)\%$, sem aumentar o uso de insumos, para tornar-se eficiente.

$$\left\{ \begin{array}{l} \text{maximizar } \theta_i \\ \text{sujeito a:} \\ -\theta_i \mathbf{y}_i + \mathbf{Y}' \boldsymbol{\lambda} \geq \mathbf{0} \\ \mathbf{x}_i - \mathbf{X}' \boldsymbol{\lambda} \geq \mathbf{0} \\ \boldsymbol{\lambda} \geq \mathbf{0} \end{array} \right. \quad (2.7)$$

No modelo proposto por Charnes *et al.* (1978), em (2.6) e (2.7), os níveis de eficiência são iguais independente da orientação da eficiência. Isto se deve ao fato que o modelo funciona com suposição de retornos constantes de escala. Banker *et al.* (1984) propuseram uma alteração para flexibilizar o modelo de forma que as DMUs em estudo possam exibir retornos variáveis de escala. Ao impor a restrição $\mathbf{1}' \boldsymbol{\lambda} = \mathbf{1}$, em (2.7) e (2.6), cria-se um envelope que engloba as DMUs de forma mais consistente e assim, as DMUs ineficientes são então comparadas com DMUs de escalas de produção similares (COELLI *et al.*, 2005).

A DEA não fornece meios analíticos para obtenção de viés ou intervalos de confiança das eficiências ajustadas. Para tal, faz-se necessário o uso de *bootstrap*. Além disso, Simar e Wilson (1998) argumentam que os modelos DEA tendem a produzir muitos valores de eficiência iguais a 1, em especial na medida em que o número de insumos utilizados aumenta, o que causa uma aproximação pobre da distribuição da eficiência próximo a 1. Portanto, a utilização de um *bootstrap* suavizado melhora a aproximação da distribuição nas proximidades do valor 1. Os autores propõem o seguinte procedimento de *bootstrap* suavizado:

1. para cada (\mathbf{x}_{i0}, y_i) , em que \mathbf{x}_{i0} é o vetor de insumos original e y_i é a produção da i -ésima DMU, com $i = 1, 2, \dots, n$, resolva o problema (2.6) (ou (2.7)) e obtenha $\hat{\theta}_0 = [\hat{\theta}_{10}, \hat{\theta}_{20}, \dots, \hat{\theta}_{n0}]$;
2. repita para $b = 1, 2, \dots, B$
 - (a) gere $\boldsymbol{\phi} = [\phi_1, \phi_2, \dots, \phi_n]$ uma amostra aleatória com reposição de $\hat{\theta}_{b-1}$

(b) defina

$$\psi_i = \begin{cases} \phi_i + h\epsilon_i & \text{se } \phi_i + h\epsilon_i \leq 1 \\ 2 - \phi_i - h\epsilon_i & \end{cases}$$

em que $h > 0$ é um parâmetro de suavização e $\epsilon_i \sim N(0, 1)$

(c) defina $\theta_{ib}^* = \bar{\phi} + \frac{\psi_i - \bar{\phi}}{\sqrt{1 + h^2/\hat{\sigma}_{\hat{\theta}_0}^2}}$, em que $\bar{\phi} = \frac{\sum_{i=1}^n \phi_i}{n}$

(d) faça $\mathbf{x}_{ib} = (\hat{\theta}_{i0}/\theta_{ib}^*)\mathbf{x}_{i0}$, e resolva o problema (2.6), ou (2.7), para $(\mathbf{x}_{ib}, \mathbf{y}_i)$, para cada i , obtendo $\hat{\theta}_b$.

O parâmetro suavização h é bastante discutido em trabalhos sobre estimação de densidade por *kernel* e pode ser fornecido pelo pesquisador. Entretanto, como Simar e Wilson (1998) apontam, existem formas de escolher h com base nos dados, procedimento que é realizado por *default* nos pacotes estatísticos que implementam o método.

Após a realização do *bootstrap*, para cada DMU, tem-se um vetor $\hat{\theta}_{ib}$, com $b = 1, 2, \dots, B$, o qual pode ser usado para calcular o viés de $\hat{\theta}_{i0}$. Assim, a eficiência da i -ésima DMU após a correção é dada por,

$$\hat{\theta}_i = \hat{\theta}_{i0} - V_{\hat{\theta}_{i0}}, \quad \text{com} \quad V_{\hat{\theta}_{i0}} = \frac{\sum_{b=1}^B \hat{\theta}_{ib}}{B} - \hat{\theta}_{i0}, \quad (2.8)$$

em que $V_{\hat{\theta}_{i0}}$ é o viés de $\hat{\theta}_{i0}$.

Os intervalos de confiança de $(1 - \alpha)100\%$ também podem ser obtidos a partir de $\hat{\theta}_{ib}$ por meio do quantil $\alpha/2$ (isto é, $q_{\alpha/2}$) e do quantil $1 - \alpha/2$ (isto é, $q_{1-\alpha/2}$) os quais também precisam ser corrigidos. Desta forma, os limites inferior e superior do intervalo de confiança de $(1 - \alpha)100\%$ podem ser obtidos, respectivamente, por $q_{\alpha/2} - 2V_{\hat{\theta}_{i0}}$ e $q_{1-\alpha/2} - 2V_{\hat{\theta}_{i0}}$. É preciso subtrair $2V_{\hat{\theta}_{i0}}$ pois, se plotarmos um histograma com os valores de $\hat{\theta}_{ib}$, veremos que este está centrado na média de $\hat{\theta}_{ib}$. Ao subtrair $V_{\hat{\theta}_{i0}}$, o histograma ficaria centrado em $\hat{\theta}_{i0}$, a estimativa viesada da eficiência. Desta forma, é preciso subtrair $V_{\hat{\theta}_{i0}}$ mais uma vez para que a distribuição empírica da eficiência fique centrada em $\hat{\theta}_i$ e se tenha limites intervalares não viesados.

Os modelos DEA são bastante populares na literatura para obter-se níveis de eficiência. Coelli *et al.* (2005), resumem as principais extensões de modelos DEA existentes até o início dos anos 2000 de forma clara e sucinta. Uma referência muito recente e bastante atualizada, certamente muito útil para aqueles com

interesse em aprofundar-se no estudo dos modelos DEA, é o livro “*Advances in DEA Theory and Applications: With Extensions to Forecasting Models*”, editado por Kaoru Tone e publicado em maio de 2017.

2.2.2 Análise de Fronteira Estocástica

Os modelos de fronteira estocástica (SFA), também conhecidos como modelo de erro composto, foram propostos paralelamente por Aigner *et al.* (1977) e Meeusen e van den Broeck (1977). Esta classe de modelos se desenvolve em um ramo em que os pesquisadores buscam modelar a produção de forma paramétrica. Nos modelos de fronteira estocástica as oscilações que ocorrem na produção são consideradas como influência de erros aleatórios bem como da ineficiência da DMU. Com isso, tornou-se possível decompor os resíduos do modelo em dois componentes: ineficiência e fatores aleatórios.

Por tratar-se de uma técnica paramétrica, a SFA requer que a produção seja definida por meio de formas algébricas, conhecidas como formas funcionais, do tipo

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}, \xi_i) \quad (2.9)$$

em que y_i é a produção da i -ésima DMU, \mathbf{x}_i um vetor de insumos, $\boldsymbol{\beta}$ um vetor de parâmetros a ser estimado e ξ_i é uma variável aleatória (v.a.) que modela as oscilações na produção relacionadas à i -ésima DMU. Neste caso, assume-se que ξ_i é uma combinação de oscilações devidas a erros aleatórios, sobre os quais a administração da DMU não tem poder, e oscilações devidas à ineficiência da DMU.

Existe uma série de formas funcionais que podem ser usadas em (2.9). Dentre as principais é possível destacar a linear (corriqueiramente vista em modelos de regressão linear), Cobb-Douglas, quadrática, quadrática normalizada, translog, Leontief generalizada e elasticidade de substituição constante (COELLI *et al.*, 2005). Rasmussen (2012), entretanto, argumenta que tais formas funcionais têm sido usadas de forma mais ou menos subjetiva, e que a escolha depende também da área de estudo. O autor também destaca que a mais comum trata-se da

forma funcional de Cobb-Douglas, a qual é definida como: ²

$$y_i = \alpha \prod_{j=1}^p x_{ij}^{\beta_j} \xi_i. \quad (2.10)$$

Entretanto, a forma funcional de Cobb-Douglas é geralmente utilizada em sua forma linear para que ajuste do modelo seja realizado. Sua forma linear é obtida por meio de uma transformação logarítmica e é dada por:

$$\ln(y_i) = \ln(\alpha) + \sum_{j=1}^p \beta_j \ln(x_{ij}) + \ln(\xi_i). \quad (2.11)$$

Por conveniência, comumente refere-se a $\ln(\alpha)$ como β_0 . Neste caso, tem-se que $\ln(\xi_i) = v_i - u_i$ em que v_i é uma v.a. que modela o impacto de erros puramente aleatórios e u_i é uma v.a. que modela o impacto da ineficiência na produção. Tipicamente, $v_i \sim N(0, \sigma_v^2)$ e u_i assume qualquer distribuição com suporte no intervalo $[0, \infty)$.

Com relação aos parâmetros β_j , na forma funcional de Cobb-Douglas, eles fornecem a elasticidade da produção com relação ao j -ésimo insumo. Neste caso, o aumento de 1% no uso do j -ésimo insumo, mantendo os demais fixos, acarreta um aumento de $\beta_j\%$ na produção (ou uma diminuição caso $\beta_j < 0$). Além disso, se $\sum_{j=1}^p \beta_j < 1$ então as DMUs operam com retornos decrescentes e $\sum_{j=1}^p \beta_j > 1$ indica retornos crescentes enquanto $\sum_{j=1}^p \beta_j = 1$ indica retornos constantes de escala.

Com relação à v.a. u_i , diferentes propostas de distribuição existem. De tais propostas, as mais conhecidas são a de Aigner *et al.* (1977), que consideram o uso da distribuição *half*-normal, Meeusen e van den Broeck (1977), que consideram uma exponencial, Stevenson (1980), que considera uma normal truncada em zero, a qual tem a distribuição *half*-normal como caso específico, Greene (1990) que propõe a distribuição gama, Medrano e Migon (2004) que propõem a distribuição log-normal, Tsionas (2007) que sugere o uso da distribuição weibull, Griffin e Steel (2008) que sugerem a distribuição gama generalizada e ainda, Almanidis

² A forma funcional translog é também bastante popular, e pode ser vista como uma extensão da formula (2.11),

$$\ln(y_i) = \ln(\alpha) + \sum_{j=1}^p \beta_j \ln(x_{ij}) + \frac{1}{2} \sum_{k=1}^p \sum_{l=1}^p \beta_{kl} \ln(x_{ik}) \ln(x_{il}) + \ln(\xi_i).$$

et al. (2014) sugerem truncar a cauda direita da distribuição utilizada para u_i alegando que DMUs com altos níveis de ineficiência rapidamente deixam de existir devido à concorrência.

Com relação a isso, Parmeter e Kumbhakar (2014) argumentam que quando o interesse primordial está na classificação das DMUs, em mais e menos eficientes, e nas características da tecnologia de produção, fornecidas pelas estimativas de β , a distribuição de u_i usada para o ajuste tem impacto muito baixo. Por outro lado, os níveis de eficiência mudam dependendo da escolha da distribuição para u_i .

A distribuição normal centrada em zero para v_i , por outro lado é consolidada na literatura. Todavia, Horrace e Parmeter (2015) propõem o que chamaram que fronteira estocástica de Laplace, em que v_i tem distribuição de Laplace, também conhecida como exponencial dupla, e u_i tem distribuição de Laplace truncada. Os autores discutem que um diferencial do modelo proposto está no fato que DMUs com $v_i - u_i > 0$ têm mesmo valor predito de ineficiência, algo que não é possível na SFA tradicional. Eles argumentam ainda que DMUs para as quais $v_i - u_i > 0$ possuem baixos níveis de ineficiência e, portanto, apenas DMUs entre as mais eficientes teriam mesma eficiência ajustada.

Neste trabalho, adotou-se a distribuição normal truncada em zero, proposta por Stevenson (1980), para u_i e a distribuição normal para v_i . Assim, pode-se encontrar $\hat{\beta}$ pela maximização da log-verossimilhança (2.12):

$$l(\mathbf{y}|\beta, \lambda, \sigma^2, \mu) = -\frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2} - \sum_{i=1}^n [(y_i - \ln(\mathbf{x}_i)\beta) - \mu]^2 \quad (2.12)$$

$$+ \sum_{i=1}^n \ln \left(\left[1 - \Phi \left(-\frac{\mu}{\sigma\lambda} - \frac{(y_i - \ln(\mathbf{x}_i)\beta)\lambda}{\sigma} \right) \right] \right)$$

$$- n \ln \left(1 - \Phi \left(-\frac{\mu}{\sigma_u} \right) \right)$$

em que $\sigma^2 = \sigma_v^2 + \sigma_u^2$, $\lambda = \sigma_u/\sigma_v$ e μ é a média da distribuição normal que quando truncada em zero gera os valores de u_i . Além disso, $\Phi(\cdot)$ denota a distribuição acumulada da normal padrão.

O processo usual para encontrar os estimadores de máxima verossimilhança consistem em derivar a log-verossimilhança (2.12) em relação a β , σ^2 , λ e μ , igualar tais expressões a zero e isolar β , σ^2 , λ e μ . Todavia, neste caso tal

procedimento não conduz a formas analíticas de obter-se estimadores para os parâmetros citados. Assim, faz-se o uso de algoritmos iterativos para encontrar $\hat{\beta}$, $\hat{\sigma}^2$, $\hat{\lambda}$ e $\hat{\mu}$, processo em que as primeiras derivadas em relação aos parâmetros são necessárias³.

Usando, então, o método da máxima verossimilhança, as variâncias desses estimadores podem ser obtidas pelo o inverso dos elementos na diagonal da matriz hessiana de (2.12) avaliada nos pontos $\hat{\beta}$, $\hat{\sigma}^2$, $\hat{\lambda}$ e $\hat{\mu}$ (BOGETOFT; OTTO, 2010). De posse de tais variâncias, testes estatísticos podem ser conduzidos. A hipótese $h_0 : \mu = 0$, pode ser testada e caso não seja rejeitada, temos que o ajuste da SFA pode ser realizado considerando a distribuição *half*-normal para u_i , como proposto por Aigner *et al.* (1977). A hipótese $h_0 : \sigma_u^2 = 0$ também é de interesse, pois caso seja rejeitada, indica que parte significativa das oscilações na produção se devem à ineficiência. Além disso, caso a hipótese $h_0 : \beta_j = 0$ seja rejeitada, tem-se que a produção responde de forma significativa à mudanças na quantidade do j -ésimo insumos utilizado.⁴

Com relação ao cálculo da eficiência orientada a produtos da i -ésima DMU usando um modelo de fronteira estocástica, a mesma é obtida pelo uso da definição (2.4). Utilizando a forma funcional de Cobb-Douglas, sendo \hat{y}_i o valor de produção esperado caso a DMU seja eficiente, então $u_i = 0$ e temos:

$$ET_i = \frac{y_i}{\hat{y}_i} \quad (2.13)$$

$$ET_i = \frac{\exp \left\{ \hat{\beta}_0 + \sum_{j=1}^p \hat{\beta}_j \ln(x_{ij}) + \hat{v}_i - \hat{u}_i \right\}}{\exp \left\{ \hat{\beta}_0 + \sum_{j=1}^p \hat{\beta}_j \ln(x_{ij}) + \hat{v}_i \right\}} \quad (2.14)$$

$$ET_i = \exp \{ -\hat{u}_i \}. \quad (2.15)$$

No entanto, apenas quando Jondrow *et al.* (1982) propuseram um procedimento para encontrar $E(u_i | \epsilon_i)$, em que $\epsilon_i = \ln(y_i) - [\beta_0 + \sum_{j=1}^p \beta_j \ln(x_{ij})]$, foi possível obter os níveis de eficiência individuais para cada DMU, nos casos em que assume-se distribuição *half*-normal ou exponencial para u_i . Battese e Coelli (1988), por outro lado, verificaram que $E[\exp(-u_i) | \epsilon_i]$ fornece uma medida mais

³As primeiras derivadas parciais de (2.12) com relação aos parâmetros não são fornecidas aqui, mas podem ser obtidas em Stevenson (1980). Os trabalhos onde propostas diferentes para u_i são estudadas também apresentam tais derivadas referentes ao caso do estudo.

⁴Entretanto, se $h_0 : \beta_j = 0$ não for rejeitada, isso não significa que pode-se descartar o j -ésimo insumo.

adequada para a eficiência e propuseram uma generalização do procedimento anterior para o caso em que assume-se uma distribuição normal truncada em zero para u_i . Assim, a eficiência ajustada é obtida como:

$$ET_i = E[\exp(-u_i)|\epsilon_i] = \left\{ \frac{1 - \Phi[\sigma_i^* - (\mu_i^*/\sigma_i^*)]}{1 - \Phi(-\mu_i^*/\sigma_i^*)} \right\} \exp \left\{ -\mu_i^* + \frac{1}{2}\sigma_i^{*2} \right\}, \quad (2.16)$$

em que

$$\mu_i^* = \frac{\mu\sigma_v^2 - \sigma_u^2\epsilon_i}{\sigma_v^2 + \sigma_u^2} \quad \text{e} \quad \sigma_i^{*2} = \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2 + \sigma_u^2}, \quad (2.17)$$

em que $\Phi(\cdot)$ denota a distribuição acumulada de uma normal padrão.

Utilizando-se do princípio da invariância dos estimadores de máxima verossimilhança (CASELLA; BERGER, 2002), uma vez obtidos $\hat{\beta}$, $\hat{\sigma}^2$, $\hat{\lambda}$ e $\hat{\mu}$, pode-se fazer $\hat{\sigma}_u^2 = \hat{\sigma}^2(\hat{\lambda}^{-2} + 1)^{-1}$ e $\hat{\sigma}_v^2 = \hat{\sigma}^2 - \hat{\sigma}_u^2$ e utilizá-los em (2.16) e (2.17), juntamente com $\hat{\mu}$, para a obtenção das eficiências técnicas e níveis de confiança.

2.2.3 Regressão Quantílica

A Regressão Quantílica (RQ) trata-se de uma técnica introduzida por Koenker e Bassett (1978) como alternativa à regressão via mínimos quadrados ordinários. Esta metodologia permite estudar a influência de um conjunto de variáveis explicativas em qualquer quantil da variável resposta. A estrutura do modelo é bastante semelhante àquela vista para SFA. Com o uso da forma funcional de Cobb-Douglas, temos a expressão (2.18), em que o sobrescrito (τ) indica que o modelo ajusta o quantil τ da produção:

$$\ln(y_i) = \beta_0^{(\tau)} + \sum_{j=1}^p \beta_j^{(\tau)} \ln(x_{ij}) + \epsilon_i^{(\tau)}. \quad (2.18)$$

Os estimadores de $\beta^{(\tau)}$ são encontrados por meio da minimização da função de perda ponderada:

$$\begin{aligned} \Delta(\beta^{(\tau)}; \tau) &= \sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}'_i \beta^{(\tau)}) \\ &= \sum_{i=1}^n \left[\tau - I(y_i - \mathbf{x}'_i \beta^{(\tau)} < 0) \right] (y_i - \mathbf{x}'_i \beta^{(\tau)}) \end{aligned} \quad (2.19)$$

em que $I(\cdot)$ é uma função indicadora a qual associa-se valor igual a 1 caso $y_i - \mathbf{x}'_i \boldsymbol{\beta}^{(\tau)} < 0$ e 0 caso contrário. Isso indica que valores de $\epsilon_i^{(\tau)}$, isto é, de erros, positivos tem peso τ e aqueles referentes a erros negativos tem pesos $1 - \tau$. A função $\rho_\tau(\cdot)$ é conhecida como função *check*⁵.

Uma vez que a função de perda (2.19) não é diferenciável em todos os pontos, torna-se necessário o uso de algoritmos de otimização para encontrar $\hat{\boldsymbol{\beta}}^{(\tau)}$. Koenker e d'Orey (1987) propuseram uma modificação no Algoritmo de Barrodale e Roberts (BARRODALE; ROBERTS, 1974) para este fim. O algoritmo de Barrodale e Roberts é uma modificação do método *simplex* para encontrar os estimadores no modelo regressão de mínimos desvios absolutos (regressão LAD, *Least Absolute Deviation*) (LEI *et al.*, 2001).

Chen e Wei (2005) investigaram os aspectos computacionais no ajuste da regressão quantílica. Os autores verificaram que o algoritmo de Barrodale e Roberts (1974) com a alteração sugerida por Koenker e d'Orey (1987) fornece boa performance em conjuntos de dados com até 5.000 observações e 50 variáveis, o que é equivalente a uma matriz de dados com 250 mil elementos.

Alguns resultados assintóticos a respeito dos parâmetros do modelo de regressão quantílica foram fornecidos em Koenker e Bassett (1978) sob a suposição de erros independentes e identicamente distribuídos (i.i.d). Koenker e Bassett (1982) expandiram estas ideias sob a suposição assintótica de erros i.i.d., isto é, que a heterocedasticidade desaparece com $n \rightarrow \infty$, para fornecer resultados complementares. Esses resultados são úteis para o estabelecimento de testes para avaliar significância dos parâmetros no modelo de regressão quantílica e a hipótese homogeneidade dos mesmos ao longo dos quantis.

Koenker e Machado (1999), introduziram testes de qualidade de ajuste para verificar o efeito da adição de variáveis independentes na melhoria do ajuste no quantil τ baseado em testes da teoria clássica de modelos de regressão. Uma medida análoga do R^2 da teoria clássica pode ser obtida por:

$$R^1(\tau) = 1 - \frac{S(\tau)}{S_m(\tau)}, \quad (2.20)$$

⁵ A função *check* foi inicialmente expressa como:

$$\rho_\tau(x) = \begin{cases} \tau x & x \geq 0 \\ (1 - \tau)|x| & x < 0 \end{cases}$$

em que

$$S(\tau) = \min_{\boldsymbol{\beta}^{(\tau)}} \sum_{i=1}^n \rho_{\tau} \left(y_i - \mathbf{x}_i' \boldsymbol{\beta}^{(\tau)} \right) \quad (2.21)$$

$$S_m(\tau) = \min_{\boldsymbol{\beta}_m^{(\tau)}} \sum_{i=1}^n \rho_{\tau} \left(y_i - \mathbf{x}_i' \boldsymbol{\beta}_m^{(\tau)} \right) \quad (2.22)$$

tal que $\boldsymbol{\beta}^{(\tau)} = [\boldsymbol{\beta}_m^{(\tau)} \quad \boldsymbol{\beta}_q^{(\tau)}]$, com $p + 1 = m + q$.

Desta forma, pode se testar a melhoria no ajuste do quantil τ proporcionado pela inclusão de q variáveis. Assim como na interpretação de R^2 , valores de $R^1(\tau)$ muito próximos de 0 indicam que as variáveis às quais os coeficientes em $\boldsymbol{\beta}_q^{(\tau)}$ referem-se não fornecem muita informação para o ajuste do quantil τ . Por outro lado, $R^1(\tau) \rightarrow 1$ indica que tais variáveis são importantes no ajuste do referido quantil.

Um teste de hipótese individual sobre os parâmetros para testar $H_0 : \beta_j^{(\tau)} = \beta_{j0}^{(\tau)}$ contra $H_1 : \beta_j^{(\tau)} \neq \beta_{j0}^{(\tau)}$ pode ser realizado por meio de um teste t ,

$$t_{n-p} = \frac{\left(\widehat{\beta}_j^{(\tau)} - \beta_{j0}^{(\tau)} \right)}{\widehat{\sigma}_{\widehat{\beta}_j^{(\tau)}}} \quad (2.23)$$

em que t_{n-p} tem distribuição t com $n - p$ graus de liberdade e $\widehat{\sigma}_{\widehat{\beta}_j^{(\tau)}}$ é o estimador do erro padrão de $\widehat{\beta}_j^{(\tau)}$. Por outro lado, para testar $H_0 : \boldsymbol{\beta}_q^{(\tau)} = \boldsymbol{\beta}_{q0}^{(\tau)}$ contra $H_1 : \boldsymbol{\beta}_q^{(\tau)} \neq \boldsymbol{\beta}_{q0}^{(\tau)}$, isto é, sobre um grupo de parâmetros, pode ser realizado um teste Wald,

$$W_q = \left(\widehat{\boldsymbol{\beta}}_q^{(\tau)} - \boldsymbol{\beta}_{q0}^{(\tau)} \right)' \boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_q^{(\tau)}}^{-1} \left(\widehat{\boldsymbol{\beta}}_q^{(\tau)} - \boldsymbol{\beta}_{q0}^{(\tau)} \right) \quad (2.24)$$

em que W_q tem distribuição χ^2 com q graus de liberdade e $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_q^{(\tau)}}^{-1}$ é o inverso da matriz de variância e covariância de $\widehat{\boldsymbol{\beta}}_q^{(\tau)}$.

Assim como em modelos de regressão clássicos, para encontrar $\widehat{\sigma}_{\widehat{\beta}_j^{(\tau)}}$ e $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}_q^{(\tau)}}^{-1}$ são necessárias algumas suposições. Todavia, é possível também utilizar-se de *bootstrap* para calcular as estatísticas dos testes. Os dois métodos de *bootstrap* mais comuns em regressão envolvem reamostragem dos resíduos ou das observações, isto é, (y_i, \mathbf{x}_i) , sendo este último mais robusto em caso de assimetria na distribuição dos erros ou especificação errônea da parte determinística do modelo (CHER-

NICK; LABUDDE, 2011).

O método de reamostrar os resíduos, no caso de regressão quantílica funciona da seguinte forma:

1. ajustar o modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^{(\tau)}$ usando a proposta de Koenker e d'Orey (1987);
2. calcular $\hat{\boldsymbol{\epsilon}}^{(\tau)} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(\tau)}$;
3. para $b = 1, 2, \dots, B$:
 - (a) gerar uma amostra $\hat{\boldsymbol{\epsilon}}^{(\tau)*}$ de tamanho n , com reposição, de $\hat{\boldsymbol{\epsilon}}^{(\tau)}$;
 - (b) calcular $\mathbf{y}^* = \mathbf{X}\hat{\boldsymbol{\beta}}^{(\tau)} + \hat{\boldsymbol{\epsilon}}^{(\tau)*}$;
 - (c) realizar (1) e (2), salvar $\hat{\boldsymbol{\beta}}^{(\tau)}$ e, se de interesse, $\hat{\boldsymbol{\epsilon}}^{(\tau)}$, obtidos.

No caso de reamostragem das observações (y_i, \mathbf{x}_i) , no passo 3a faz-se a reamostragem das observações e o passo 3b não seria executado. Após este procedimento pode-se obter $\hat{\sigma}_{\hat{\beta}_j^{(\tau)}}$ e $\Sigma_{\hat{\boldsymbol{\beta}}^{(\tau)}}^{-1}$ com o uso das realizações de *bootstrap* e realizar os testes mencionados. Pode-se também realizar a correção de viés do estimador como:

$$\tilde{\beta}_j^{(\tau)} = \hat{\beta}_j^{(\tau)} - V_{\hat{\beta}_j^{(\tau)}} \quad \text{com} \quad V_{\hat{\beta}_j^{(\tau)}} = \bar{\beta}_{jB}^{(\tau)} - \hat{\beta}_j^{(\tau)} \quad (2.25)$$

em que $\bar{\beta}_{jB}^{(\tau)}$ é a média dos valores de *bootstrap* obtidos. Intervalos de confiança de *bootstrap* de $(1 - \alpha)100\%$ também podem ser obtidos por meio dos quantis $\alpha/2$ e $(1 - \alpha/2)$ dos valores de *bootstrap*, que também precisam ser corrigidos, assim como mencionado no caso da DEA.

Outros métodos de *bootstrap* existem na literatura e podem ser empregados no contexto de regressão quantílica. Por exemplo, Liu *et al.* (1988) propõem o *wild bootstrap* que é mais indicado em casos de heterocedasticidade dos erros, He e Hu (2002) propõem o *markov chain marginal bootstrap* que ao invés de uma sequência de valores independentes, gera uma cadeia de markov para cada parâmetro do modelo, e é mais indicado em problemas complexos, envolvendo muitos parâmetros.

Voltando ao contexto da análise de eficiência, o uso da RQ se dá pela possibilidade que esta técnica fornece em ajustar uma equação para cada quantil condicional da variável resposta (neste contexto, a produção). Esta possibilidade é de

especial interesse em casos nos quais as DMUs em estudo apresentam diferentes tecnologias de produção. Já para calcular a eficiência, considera-se que as DMUs localizadas nos quantis mais altos dominam melhores práticas de produção (KADITI; NITSI, 2010; KNOX *et al.*, 2007) e utilizamos o valor da produção ajustado em tal quantil como sendo a produção da DMU caso esta fosse eficiente. A partir disso, a eficiência da DMU é calculada com base na expressão (2.4), sendo neste caso \hat{y}_i é a produção esperada caso a i -ésima DMU operasse nas condições das DMUs localizadas nas proximidades do quantil τ ,

$$ET_i^{(\tau)} = \frac{y_i}{\hat{y}_i} \quad (2.26)$$

$$ET_i^{(\tau)} = \frac{\exp \left\{ \hat{\beta}_0^{(\tau)} + \sum_{j=1}^p \hat{\beta}_j^{(\tau)} \ln(x_{ij}) + \hat{\epsilon}_i^{(\tau)} \right\}}{\exp \left\{ \hat{\beta}_0^{(\tau)} + \sum_{j=1}^p \hat{\beta}_j^{(\tau)} \ln(x_{ij}) \right\}} \quad (2.27)$$

$$ET_i^{(\tau)} = \begin{cases} \exp \left\{ \hat{\epsilon}_i^{(\tau)} \right\} & , \text{ se } \exp \left\{ \hat{\epsilon}_i^{(\tau)} \right\} \leq 1 \\ 1 & , \text{ se } \exp \left\{ \hat{\epsilon}_i^{(\tau)} \right\} > 1 \end{cases} \quad (2.28)$$

em que $\hat{\epsilon}_i^{(\tau)} = \ln(y_i) - \widehat{\ln(y_i)}^{(\tau)}$, ou seja, é o erro observado para a i -ésima DMU em relação ao quantil τ ajustado para os níveis de insumos que ela utiliza.

A escolha do quantil τ para o qual ajustar a fronteira de produção pode ser baseada no tamanho amostral. Como discutido por Kaditi e Nitsi (2010) quanto maior amostra maior pode ser o quantil escolhido.

2.3 O enfoque Bayesiano para Regressão Quantílica

Koenker e Machado (1999) foram os primeiros a observar que minimizar a função (2.19) com relação a $\beta^{(\tau)}$ fornece resultados equivalentes a minimizar a função de verossimilhança formada a partir da distribuição de Laplace assimétrica. A distribuição de Laplace assimétrica para o caso da regressão quantílica, denotada por $ALD(\tau, \sigma, \mu)$, pode ser expressa como:

$$f(y|\tau, \sigma, \mu) = \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\frac{\rho_\tau(y-\mu)}{\sigma} \right\}, \quad (2.29)$$

em que τ é o quantil de interesse, σ é um parâmetro de escala, μ corresponde ao τ -ésimo quantil e $\rho(\cdot)$ é a função *check* já vista. Para encontrar $\hat{\beta}^{(\tau)}$, pode ser feito

$\mu = \mathbf{x}'\boldsymbol{\beta}^{(\tau)}$ e realizar o processo de minimização com relação a $\boldsymbol{\beta}^{(\tau)}$.

A função de verossimilhança com base na função densidade acima é dada por

$$L(\mathbf{y}|\boldsymbol{\beta}^{(\tau)}, \sigma) = \left[\frac{\tau(1-\tau)}{\sigma} \right]^n \exp \left\{ -\frac{1}{\sigma} \sum_{i=1}^n \rho_{\tau} \left(y_i - \mathbf{x}'_i \boldsymbol{\beta}^{(\tau)} \right) \right\}, \quad (2.30)$$

a qual pode ser utilizada para realizar o ajuste bayesiano do modelo.

O trabalho de Yu e Moyeed (2001) trata-se de um dos primeiros a propor o uso de inferência bayesiana para o ajuste de regressão quantílica usando a função de verossimilhança acima. Os autores realizaram o procedimento em três conjuntos de dados, sendo um deles artificial, obtendo resultados satisfatórios, e mostrando a viabilidade da inferência bayesiana na estimação de $\boldsymbol{\beta}^{(\tau)}$.

Com relação ao uso de distribuições *a priori*, Yu e Moyeed (2001) argumentam que mesmo distribuições *a priori* impróprias para $\boldsymbol{\beta}^{(\tau)}$ resultam em distribuições *a posteriori* próprias. Além disso, tratando-se do parâmetro de escala σ os autores citam que este não é necessário no processo de estimação e pode ser feito igual a 1, simplificando a verossimilhança (2.30).

Todavia, Luo *et al.* (2012), por exemplo, usam uma distribuição gama inversa pouco informativa para σ . O uso de uma gama inversa como *a priori* não informativa para σ é válido. Como argumentam Gelman *et al.* (2006), quando o parâmetro de escala não se trata de um hiperparâmetro, qualquer distribuição razoavelmente não informativa é aceitável.

2.4 Alguns pontos em Inferência Bayesiana

Para calcular estimativas dos parâmetros de um modelo por meio da inferência bayesiana, é necessário obter o que é conhecido nesse contexto como distribuição *a posteriori* do parâmetro. A distribuição *a posteriori* de um parâmetro θ é definida como,

$$p(\theta|\mathbf{y}) = \frac{L(\mathbf{y}|\theta) p(\theta)}{\int_{\theta} L(\mathbf{y}|\theta) p(\theta) d\theta}, \quad (2.31)$$

em que $p(\theta|\mathbf{y})$ é a distribuição *a posteriori* do parâmetro, $L(\mathbf{y}|\theta)$ é função de verossimilhança dos dados (\mathbf{y} é um vetor de dados) e $p(\theta)$ é a distribuição de pro-

babilidade *a priori* do parâmetro.

Observa-se que a integral no denominador de (2.31) resulta em uma constante. Ou seja, $p(\theta|\mathbf{y})$ é proporcional a $L(\mathbf{y}|\theta)p(\theta)$. Assim, procura-se trabalhar, para termos de simplificação dos cálculos, apenas com as partes que dependem de θ , ou seja, do parâmetro, e escreve-se:

$$p(\theta|\mathbf{y}) \propto L(\mathbf{y}|\theta)p(\theta). \quad (2.32)$$

Na prática, geralmente se trabalha com mais de um parâmetro e, então, no lugar de θ tem-se um vetor de parâmetros e $p(\theta|\mathbf{y})$ passa a se chamar distribuição *a posteriori* conjunta dos parâmetros. Todavia, usualmente o interesse está na distribuição individual de alguns dos parâmetros no vetor θ . Para tanto, é preciso obter a distribuição *a posteriori* marginal dos parâmetros. Neste procedimento fazem-se necessárias as distribuições condicionais completas *a posteriori* de cada parâmetro, isto é, a distribuição *a posteriori* de cada θ_i , $i = 1, 2, \dots, k$ dado os θ_j , $j \neq i$. Por exemplo, supondo que o vetor θ tem tamanho k , a distribuição condicional completa *a posteriori* de θ_1 é obtida considerando todos os demais parâmetros como constantes. Isso fornece a expressão:

$$p(\theta_1|\mathbf{y}, \theta_2, \dots, \theta_k) \propto L(\mathbf{y}|\theta)p(\theta_1), \quad (2.33)$$

em que $p(\theta_1|\mathbf{y}, \theta_2, \dots, \theta_k)$ é a distribuição condicional completa *a posteriori* de θ_1 .

Por outro lado, na prática, dificilmente $p(\theta_1|\mathbf{y}, \theta_2, \dots, \theta_k)$ apresenta forma conhecida ou fácil de derivar analiticamente. Assim, os algoritmos de simulação de cadeias de Markov via métodos de Monte Carlo (MCMC) são indispensáveis no contexto de inferência bayesiana. Exemplos clássicos de algoritmos de MCMC são o Amostrador de Gibbs e Metropolis-Hastings. O pesquisador tem a opção de codificar o algoritmo de MCMC que deseja utilizar por conta própria ou utilizar algum software para este fim.

Os algoritmos de MCMC usam, então, as distribuições condicionais completas *a posteriori* para simular valores da distribuição *a posteriori* de cada parâmetro de interesse. Para tanto, são necessários chutes iniciais e então novos valores são gerados sucessivamente até que se atinja um estágio de equilíbrio. Quando este estágio é atingido, diz-se que a cadeia de markov do parâmetro convergiu.

Somente após a convergência, os valores gerados são representativos da distribuição *a posteriori* do parâmetro e podem ser utilizados para fazer inferência a respeito do mesmo.

Assim, torna-se necessário descartar as iterações iniciais, isto é, os valores gerados antes da convergência. Estes valores formam o que é chamado de período de aquecimento da cadeia ou *burn-in*. Raftery e Lewis (1992), propuseram um teste que fornece estimativas para o número de iterações que devem ser descartadas bem como o número total de iterações que devem ser simuladas, para que se possa estimar um quantil q da distribuição *a posteriori* com precisão s e probabilidade p .

O teste necessita de uma cadeia piloto para que as quantidades citadas possam ser definidas. O número mínimo de iterações na cadeia piloto é definido como:

$$N_{min} = \Phi^{-1} \left(\frac{1}{2}(1 + p) \right)^2 \frac{q(1 - q)}{s^2}, \quad (2.34)$$

em que $\Phi(\cdot)^{-1}$ é a função distribuição inversa da normal padrão. Além disso, as observações geradas na cadeia de markov usualmente apresentam algum tipo de autocorrelação, e assim, é necessário armazenar apenas um valor em cada k iterações da cadeia. Este processo é conhecido como *thinning* e k é chamado *thin*. O teste de Raftery e Lewis também fornece uma estimativa de k com base na amostra piloto. O teste ainda retorna uma medida conhecida com fator de dependência, a qual trata-se de um fator de inflação do número de iterações necessárias para se estimar q que ocorre devido à autocorrelação na cadeia. Valores dessa medida abaixo de 5 indicam que já houve convergência da cadeia (NOGUEIRA *et al.*, 2004).

Também devido à autocorrelação, a informação contida em uma cadeia geralmente é equivalente a uma cadeia menor cujas observações não sejam autocorrelacionadas. É interessante, então, calcular o tamanho amostral efetivo (TAE) da cadeia simulada, que mostra quantas iterações não correlacionadas seriam necessárias para obter a mesma informação. O TAE pode ser obtido por,

$$TAE = \frac{m}{1 + 2 \sum_{t=1}^{\infty} \rho_t} \quad (2.35)$$

em que ρ_t é a correlação da observação na posição i com aquela na posição $i + t$ e m é o tamanho da cadeia de markov usada. Na medida que t cresce, ρ_t decresce e, caso ρ_t decresça de forma muito lenta, TAE se torna muito pequeno, indicando que as observações da cadeia são muito autocorrelacionadas Gelman *et al.* (2014).

Pode-se citar também como úteis na avaliação das cadeias de markov dos parâmetros os testes de propostos por Gelman e Rubin (1992). O teste de Gelman e Rubin (1992) faz uso de pelo menos duas cadeias de markov para cada parâmetro e verifica se a diferença entre a variância entre e dentro das cadeias é significativamente diferente. Valores da estatística do teste, conhecida com fator de redução de escala potencial, próximos de 1,1 são indicativos que a convergência foi atingida (NOGUEIRA *et al.*, 2004).

Uma maneira menos formal de verificar a convergência de uma cadeia e definir tamanho de *burn-in* consiste na análise gráfica da cadeia. Para tanto, cria-se um gráfico com os valores, na ordem em que foram gerados, e verifica-se a partir de qual ponto os valores começam a ser gerados dentro de uma mesma amplitude. Os valores gerados antes de tal ponto podem ser descartados com o *burn-in*.

Por fim, é preciso resumir as observações simuladas para se obter um estimador pontual (estimador de Bayes) e/ou estimativas intervalares. O estimador Bayes é tipicamente obtido pela minimização de função de perda. Uma função de perda quadrática, por exemplo, equivale a ter a média aritmética *a posteriori* das observações como estimador de Bayes. Outra alternativa é definir o estimador de Bayes como o valor que maximiza a distribuição *a posteriori*.

Entretanto, são as estimativas intervalares que têm o foco no contexto bayesiano, chamadas de intervalos de credibilidade. Um intervalo de credibilidade de $(1 - \alpha)100\%$ pode ser obtido com diferentes limites inferiores e superiores. O intervalo de credibilidade mais adequado é aquele que concentra entre seus limites a região com maior densidade de probabilidade. Estes são chamados de intervalos de credibilidade HPD (do inglês *Highest Probability Density*). Estes intervalos são geralmente os utilizados pois tendem a ser mais curtos.

Ao contrário dos intervalos de confiança na inferência frequentista, o intervalo de credibilidade é interpretado em termos de probabilidade. Em um intervalo de credibilidade de $(1 - \alpha)100\%$, diz-se que o verdadeiro valor do parâmetro está contido entre os limites do intervalo com probabilidade igual a $(1 - \alpha)$.

Capítulo 3

Material e Métodos

3.1 Estudo de simulação

3.1.1 Simulação dos dados

No estudo de simulação foram gerados conjuntos de dados segundo um modelo de fronteira estocástica utilizando a forma funcional de Cobb-Douglas, considerando para v_i uma distribuição normal e para u_i uma normal truncada em zero. As simulações foram divididas em três cenários com respeito à quebra de suposições da fronteira estocástica. Cada um dos três cenários foi subdividido em outros três subcenários, com respeito ao peso da ineficiência na variância total do erro composto. Ao todo, foram simulados nove conjuntos de dados, com 100 unidades produtivas, 3 insumos, um produto e cada conjunto de dados foi repetido 300 vezes, alterando-se apenas os valores de erro aleatório.

O uso de apenas 3 insumos na produção se baseia no fato que, segundo Samuelson e Nordhaus (2010), insumos de produção podem ser categorizados em três amplos grupos, a saber, matéria-prima, mão-de-obra e capital. A escolha por gerar dados de 100 unidades produtivas e por 300 repetições é fundamentalmente devido a custos computacionais.

Apesar disso, o tamanho amostra coincide com aquele usado por Behr (2010) e Liu *et al.* (2008). Liu *et al.* (2008) argumentam que este tamanho é condzidente com os conjuntos de dados usados em trabalhos práticos e é pouco provável que o tamanho amostral afete as estimativas de máxima verossimilhança na SFA. Os autores citados também simularam os dados usando um modelo de fronteira estocástica com forma funcional de Cobb-Douglas.

Considerando o erro composto como $\epsilon_i = v_i - u_i$, com variância $\sigma^2 = \sigma_v^2 + \sigma_u^2$ e o peso da ineficiência nas oscilações da produção dado por $\gamma = \sigma_u^2 / \sigma^2$,

os conjuntos de dados no primeiro cenário (C1) foram simulados pelo modelo:

$$Y_{ik}^{(C1,\gamma)} = 2 + 0,45X_{i1} + 0,3X_{i2} + 0,4X_{i3} + v_{ik} - u_i. \quad (3.1)$$

Neste caso, $k = 1, 2, \dots, 300, i = 1, 2, \dots, 100, u_i \sim Nt(0, 3; \gamma\sigma^2)$ e $v_{ik} \sim N[0, (1 - \gamma)\sigma^2]$. No cenário C1 não há quebras de suposição quanto a homogeneidade da tecnologia de produção nem distribuição do erro.

No segundo cenário (C2) foram simulados dados em que as DMUs são provenientes de dois grupos com tecnologias de produção distintas. Das 100 DMUs em C2, 50 tiveram produção simulada pela expressão (3.2) e as outras 50 pela expressão (3.3):

$$Y_{tk}^{(C2,\gamma)} = 2 + 0,45X_{t1} + 0,3X_{t2} + 0,4X_{t3} + v_{tk} - u_t \quad t = 1, \dots, 50 \quad (3.2)$$

$$Y_{zk}^{(C2,\gamma)} = 2 + 0,54X_{z1} + 0,32X_{z2} + 0,42X_{z3} + v_{zk} - u_z \quad z = 1, \dots, 50 \quad (3.3)$$

No terceiro cenário (C3), foi considerado a quebra de homogeneidade do erro composto. Assim como em C2, as DMUs em C3 são provenientes de dois grupos nos quais a média de ineficiência é diferente. Das 100 DMUs em C3, 50 tiveram produção simulada pela expressão (3.4) e as outras 50 pela expressão (3.5):

$$Y_{tk}^{(C3,\gamma)} = 2 + 0,45X_{t1} + 0,3X_{t2} + 0,4X_{t3} + v_{tk} - u_t \quad t = 1, \dots, 50 \quad (3.4)$$

$$Y_{zk}^{(C3,\gamma)} = 2 + 0,45X_{z1} + 0,3X_{z2} + 0,4X_{z3} + v_{zk} - u_z^* \quad z = 1, \dots, 50 \quad (3.5)$$

em que $u_t \sim Nt(0, 3; \gamma\sigma^2)$ e $u_z^* \sim Nt(0, 05; \gamma\sigma^2)$.

Cada um dos três cenários acima (C1, C2 e C3) foi subdividido em outros três, consistindo de conjuntos de dados simulados para valores de $\gamma = 0,1$, $\gamma = 0,5$ e $\gamma = 0,9$. Isto é, conjuntos de dados em que 10% da variância do erro composto se deve à ineficiência, outro em que tal peso passa a ser 50% e outro em que 90% da variância é devido à ineficiência. Por fim, os valores dos insumos X_1 , X_2 e X_3 foram gerados uniformemente no intervalo (1,3). Ainda,

todo o processo de simulação dos dados e ajuste dos modelos foi conduzido no software R (R Core Team, 2016).

3.1.2 Modelos ajustados

Análise Envoltória de Dados

O modelo DEA avaliado no estudo de simulação trata-se do modelo sugerido por Banker *et al.* (1984). A escolha se deve ao fato de que os dados não foram simulados com suposição de retornos constantes de escala. Assim, o modelo DEA ajustado neste caso trata-se do seguinte problema de otimização:

$$\left\{ \begin{array}{l} \text{maximizar } \theta_i \\ \text{sujeito a:} \\ -\theta_i y_i + \sum_{r=1}^{100} y_r \lambda_r \geq 0 \\ x_{ij} \sum_{r=1}^{100} x_{rj} \lambda_r \geq \text{para } j = 1, 2, 3 \\ \sum_{r=1}^{100} \lambda_r = 1 \\ \lambda_r \geq 0 \end{array} \right. \quad (3.6)$$

em que θ_i é a eficiência orientada a produto da i -ésima DMU.

Optou-se por ajustar um modelo DEA com orientação a produtos pois a definição de eficiência técnica usada nos modelos de fronteira estocástica, e também regressão quantílica, a razão da produção observada pela produção ajustada, é uma definição de eficiência orientada a produto (COELLI *et al.*, 2005).

Além disso, realizou-se a correção de viés via *bootstrap* proposta por Simar e Wilson (1998), como já descrita. Neste caso, foram utilizadas $B = 1.000$ realizações de *bootstrap*. Esse procedimento encontra-se implementado no pacote rDEA (SIMM; BESSTREMYANNAYA, 2016), o qual foi utilizado para o ajuste dos modelos DEA.

Análise de Fronteira Estocástica

O modelo de fronteira estocástica ajustado aos dados simulados segue a mesma estrutura do modelo utilizado para simular os dados, ou seja,

$$\ln(y_i) = \beta_0 + \beta_1 \ln(x_{i1}) + \beta_2 \ln(x_{i2}) + \beta_3 \ln(x_{i3}) + v_i - u_i, \quad (3.7)$$

com $i = 1, 2, \dots, 100$, $v_i \sim N(0, \sigma_v)$ e u_i uma normal truncada em zero com média e variância desconhecidas.

Neste caso, o ajuste da eficiência é feito pelo procedimento proposto por Battese e Coelli (1988),

$$ET_i = E[\exp(-u_i)|\epsilon_i] = \left\{ \frac{1 - \Phi[\sigma_i^* - (\mu_i^*/\sigma_i^*)]}{1 - \Phi(-\mu_i^*/\sigma_i^*)} \right\} \exp \left\{ -\mu_i^* + \frac{1}{2}\sigma_i^{*2} \right\}, \quad (3.8)$$

em que

$$\mu_i^* = \frac{\mu\sigma_v^2 - \sigma_u^2\epsilon_i}{\sigma_v^2 + \sigma_u^2} \quad \text{e} \quad \sigma_i^{*2} = \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2 + \sigma_u^2}, \quad (3.9)$$

com $\epsilon_i = v_i - u_i$, e $\Phi(\cdot)$ denota a distribuição acumulada de uma normal padrão, neste caso $i = 1, 2, \dots, 100$.

Esses procedimentos estão disponíveis no pacote `frontier` (COELLI; HENNINGSEN, 2017), o qual foi utilizado para o ajuste do modelo de fronteira estocástica via método da Máxima Verossimilhança.

Regressão Quantílica

O modelo de regressão quantílica avaliado no estudo de simulação é dado por:

$$\ln(y_i) = \beta_0^{(\tau)} + \beta_1^{(\tau)} \ln(x_{i1}) + \beta_2^{(\tau)} \ln(x_{i2}) + \beta_3^{(\tau)} \ln(x_{i3}) + \epsilon_i^{(\tau)}, \quad (3.10)$$

em que τ indica que os parâmetros referem-se ao ajuste do quantil τ da produção¹. Neste caso, optou-se por utilizar $\tau = 0,8$, uma vez que Liu *et al.* (2008), ao ajustar o quantil 0,5 e 0,8 em amostras de mesmo tamanho, verificaram melhor ajuste para este quantil.

O modelo foi ajustado com o auxílio do pacote `quantreg` (KOENKER, 2017). Optou-se por usar a versão modificada do Algoritmo de Barrodale e Roberts (BARRODALE; ROBERTS, 1974) proposta por Koenker e d'Orey (1987), uma vez que este procedimento é recomendado pelos autores para problemas com até poucos milhares de observações.

Com relação a eficiência da i -ésima DMU, esta pode ser obtida por meio

¹Os logaritmos são funções monótonas, então \ln do quantil τ de y é igual ao quantil τ de $\ln(y)$.

da expressão,

$$ET_i^{(\tau)} = \begin{cases} \exp(\hat{\epsilon}_i^{(\tau)}), & \text{se } \exp(\hat{\epsilon}_i^{(\tau)}) \leq 1 \\ 1, & \text{caso contrário.} \end{cases} \quad (3.11)$$

que resulta da razão da produção observada pela a produção esperada, caso a DMU operasse com a mesma tecnologia de produção que as DMUs próximas ao quantil τ .

Regressão Quantílica Bayesiana

Para ajustar o modelo de regressão (3.10) sob o enfoque bayesiano, é necessário a suposição que:

$$\ln(y_i) \sim ALD \left(\tau, \sigma, \mu = \beta_0^{(\tau)} + \sum_{j=1}^3 \ln(x_{ij})\beta_j^{(\tau)} \right), \quad (3.12)$$

em que $j = 1, 2$ e 3 refere-se os insumos. Yu e Moyeed (2001) argumentam que o parâmetro σ não faz-se necessário para o ajuste do modelo de regressão quantílica, podendo ser considerado igual 1 no momento do ajuste do modelo. Todavia, optou-se por ajustar o modelo com σ e outro sem tal parâmetro para averiguar se nesse caso haveria diferença.

Por conveniência, chame $\beta_0^{(\tau)} + \sum_{j=1}^3 \ln(x_{ij})\beta_j^{(\tau)}$ de $\mathbf{x}_i\boldsymbol{\beta}^{(\tau)}$. A verossimilhança com base nesta distribuição é então, dada por:

$$L(\ln(\mathbf{y})|\tau, \sigma, \boldsymbol{\beta}^{(\tau)}) = \left[\frac{\tau(1-\tau)}{\sigma} \right]^{100} \exp \left\{ \frac{-\sum_{i=1}^{100} \rho_{\tau} \left(y_i - \mathbf{x}_i\boldsymbol{\beta}^{(\tau)} \right)}{\sigma} \right\}. \quad (3.13)$$

Tomando a distribuição $N(0, 100)$ como distribuição *a priori* para $\beta_0^{(\tau)}$ e para $\beta_j^{(\tau)}$ e para σ uma distribuição gama inversa com parâmetros $\alpha = 0,001$ e $\beta = 0,001$, ou seja, distribuições *a priori* pouco informativas, a distribuição *a posteriori* conjunta é dada por:

$$p(\boldsymbol{\beta}^{(\tau)}, \sigma|\mathbf{y}) \propto L(\mathbf{y}|\tau, \boldsymbol{\beta}^{(\tau)}, \sigma) \prod_{j=0}^3 p(\beta_j^{(\tau)}) p(\sigma). \quad (3.14)$$

As distribuições condicionais completas não foram obtidas manualmente pois as cadeias de markov foram simuladas no software JAGS, versão 4.2, (PLUM-

MER *et al.*, 2003), com o auxílio do pacote `rjags` (PLUMMER, 2016) do R.

Entretanto, a distribuição de Laplace Assimétrica não está disponível diretamente no referido software. Foi necessário usar a estratégia do “*ones trick*”, descrita em Ntzoufras (2009). Desta maneira, a verossimilhança 3.13 é utilizada por meio de uma variável aleatória auxiliar (ξ) com todos os valores iguais a 1, cuja função de probabilidade é uma Bernoulli com probabilidade de sucesso igual a $C \cdot L(\mathbf{y}|\tau, \sigma, \boldsymbol{\beta}^{(\tau)})$, em que C é um valor $0 < C \ll 1$. Isto é feito para que se possa garantir que a probabilidade de sucesso seja sempre menor que 1. Em pseudo-código, isso é feito como:

para i de 1 a 100;

$$\mu_i = \mathbf{x}_i' \boldsymbol{\beta}^{(\tau)}$$

$$et_i = \begin{cases} 1 & \text{se } \exp(y_i - \mu_i) \geq 1, \\ \exp(y_i - \mu_i) & \text{caso contrário} \end{cases}$$

$$\xi_i \sim \text{bernoulli}(C \cdot L(y_i|\tau, \sigma, \boldsymbol{\beta}^{(\tau)}))$$

para j de 0 a 3;

$$\beta_j^{(\tau)} \sim N(0, 100)$$

$$\sigma \sim \text{GammaInv}(0, 001; 0, 001)$$

É possível observar pelo pseudo-código acima que a eficiência é ajustada pela expressão (3.11), da mesma forma que na RQ frequentista.

Para cada um dos parâmetros foram simuladas duas Cadeias de Markov com tamanho 200 mil. A primeira metade de cada uma das cadeias foi descartada para evitar o período de aquecimento (*burn-in*). Na segunda metade, uma de cada 5 iterações foram salvas, como forma de quebrar a correlação entre os valores simulados. Desse modo, restaram 40 mil valores para fins de inferência em cada parâmetro. Uma vez que o interesse primordial neste caso é avaliar a capacidade dos modelos em ajustar a eficiência, observou-se apenas as Cadeias de Markov referentes as eficiências técnicas, tomando como estimador bayesiano a média *a posteriori* destas.

Comparação dos modelos

Uma vez que os verdadeiros valores de eficiência técnica são conhecidos, o que na prática é impossível, estes podem ser contrapostos com os respectivos valores ajustados e, assim, comparações por meio de medidas mais simples são possíveis. Optou-se por avaliar a adequabilidade dos modelos para o ajuste da eficiência por meio da distância entre os valores simulados e ajustados, pelo uso do Erro Quadrático Médio, e pela concordância quanto à classificação das DMUs, por meio do Coeficiente de Correlação de Pearson. Os valores destas medidas são apresentados em forma de boxplots, que fornecem informação pontual e de como tais medidas variam ao longo das repetições.

O EQM pode ser obtido pela expressão,

$$EQM = \frac{\sum_{i=1}^n (et_i - \hat{et}_i)^2}{n}, \quad (3.15)$$

em que n é o número de observações no conjunto de dados, neste caso $n = 100$, et_i e \hat{et}_i são, respectivamente, a eficiência técnica simulada e ajustada da i -ésima DMU.

O Coeficiente de correlação de Pearson (\hat{r}) por ser calculado como:

$$\hat{r}_{et,\hat{et}} = \frac{cov(et, \hat{et})}{\sqrt{S_{et}^2 \cdot S_{\hat{et}}^2}}, \quad (3.16)$$

em que $cov(et, \hat{et}) = \frac{\sum_{i=1}^n (et_i - \bar{et})(\hat{et}_i - \bar{\hat{et}})}{n - 1}$ é a covariância entre a eficiência técnica simulada e ajustada, $S_{et}^2 = \frac{\sum_{i=1}^n (et_i - \bar{et})^2}{n - 1}$ e $S_{\hat{et}}^2 = \frac{\sum_{i=1}^n (\hat{et}_i - \bar{\hat{et}})^2}{n - 1}$ são as variâncias da eficiência técnica simulada e ajustada, respectivamente.

Foi utilizado também o percentual de concordância na classificação das DMUs em relação às mais eficientes e menos eficientes. Para esta medida, apenas a informação pontual foi utilizada, neste caso a média da medida. O percentual de concordância entre as mais eficientes foi calculado como:

$$PC_p = \frac{\hat{\delta}_p}{\delta_p}, \quad 0 < p < 1 \quad (3.17)$$

em que δ_p é o número de DMUs correspondente às $p\%$ mais eficientes da amostra e $\hat{\delta}_p$ é o número de DMUs classificadas corretamente entre as $p\%$ mais eficientes

pelo modelo. O percentual de concordância entre as menos eficientes é obtido pela mesma expressão, exceto que, para tal, δ_p deve ser o número de DMUs correspondente às $p\%$ menos eficiente da amostra e $\widehat{\delta}_p$ é o número de DMUs classificadas corretamente entre as $p\%$ menos eficientes. Tem-se também que quando $p \rightarrow 1$, então $PC_p \rightarrow 1$.

3.2 Aplicação aos dados da Agricultura Familiar

3.2.1 Descrição do conjunto de dados

As informações utilizadas na aplicação foram obtidas junto ao SIDRA (Sistema IBGE de Recuperação Automática) coletadas no Censo Agropecuário de 2006. O conjunto consiste das informações sobre a produção dos estabelecimentos pertencentes à grupos familiares nas 558 microrregiões brasileiras.

Especificamente, tem-se a produção da Agricultura Familiar (em mil reais, denotada por y), a área ocupada por estabelecimentos familiares (em hectares e denotada por x_1), o número de máquinas e implementos agrícolas nos estabelecimentos familiares (número de tratores, roçadeiras e etc., denotado por x_2), mão-de-obra utilizada nos estabelecimentos familiares (número de pessoas, denotada por x_3) a nível de microrregião.

Na composição da produção da agricultura familiar (y) foi considerada a renda proveniente da produção animal, vegetal e agroindústria, uma vez que omitir alguma dessas fontes de produção introduziria um viés na produção da microrregião. Na composição da variável área (x_1), foram excluídas áreas com reserva legal e áreas degradadas não próprias para produção pois a adição de tais áreas acarretaria um viés no uso deste insumo.

Uma única microrregião no conjunto de dados, Amapá-AP, apresentou zero máquinas e implementos nos estabelecimento familiares. Devido a linearização da forma funcional de Cobb-Douglas, valores iguais a zero em alguma das variáveis introduzem problemas computacionais, uma vez que $\ln(0) = -\infty$. O valor 10^{-5} foi então adicionado à variável x_2 (número de máquinas), para todas as microrregiões, de modo que tal microrregião não fosse excluída da análise.

3.2.2 Modelo ajustado

Apenas o modelo de regressão quantílica bayesiana com parâmetro de escala, equação (3.13), foi ajustado aos dados da Agricultura Familiar. Trata-se da mesma estrutura do modelo ajustado para os dados simulados, exceto que desta vez utilizou-se o quantil 0,95 uma vez que o tamanho amostral é de 558 micror-regiões. Assim como no caso do estudo de simulação, o software JAGS, com o auxílio do pacote `rjags` foi utilizado.

O teste de Raftery e Lewis foi utilizado para determinar o tamanho do *burn-in* e número total de iterações necessárias neste caso. Após isso, o teste de Gelman Rubin foi utilizado para verificar a convergência das cadeias e o TAE (tamanho amostral efetivo) foi calculado para verificar a qualidade da amostra.

Capítulo 4

Resultados e Discussões

4.1 Resultados da simulação

Na Figura 4.1 tem-se os boxplots dos valores de EQM observados para cada um dos modelos nas 300 repetições de cada cenário. Observa-se que na medida que o peso da ineficiência na variância total do erro composto aumenta ($\gamma \rightarrow 1$), melhor é o ajuste da eficiência técnica, em todos os cenários, por parte da SFA e DEA. Esse comportamento é verificado para os modelos de regressão quantílica apenas no cenário C3, onde a suposição de homogeneidade da distribuição da ineficiência não é satisfeita. Além disso, apesar do disposto por Yu e Moyeed (2001), o modelo de regressão quantílica com parâmetro de escala (BRQ1) obteve, em C3, resultados melhores para $\gamma = 0,1$ e $\gamma = 0,5$ em relação ao modelo sem tal parâmetro (BRQ).

A medida que γ aproxima-se de 1, ocorre que toda a variância do erro composto é devido à ineficiência. Ou seja, não há erros aleatórios impactando a produção de forma relevante. Nesta condição, C1 e C3 são cenários ideais para DEA. Mais ainda, nesta condição C1 é também um cenário ideal para SFA. Isso explica o fato do cenário C2 ser aquele em que DEA e SFA apresentaram resultados menos satisfatórios.

No cenário C3 a quebra na homogeneidade da distribuição da ineficiência foi introduzida apenas em termos da média de ineficiência. Este caso parece ser contornado pela SFA quando o peso da ineficiência na variância total do erro composto aumenta. Isso provavelmente ocorre pelo fato de que com variâncias mais altas para a ineficiência, os dois grupos se mesclam e podem ser percebidos pelo modelo como um único grupo, como se estivesse no caso de C1. Entretanto, é possível que a diferença introduzida seja muito pequena e isso ajude o modelo a ter resultados semelhantes àqueles em C1.

Pelo fato de que distribuições *a priori* pouco informativas foram utiliza-

das para os modelos de regressão quantílica sob enfoque bayesiano, já era esperado que os resultados fossem semelhantes aos obtidos pelo enfoque frequentista. Além disso, com base nos resultados de Liu *et al.* (2008) e Behr (2010) os valores de EQM em C1 para regressão quantílica também eram esperados. Em C2, como os modelos de regressão quantílica são indicados para casos em que a influência das variáveis independentes na variável resposta depende do quantil ajustado, o comportamento do EQM também foi como esperados por tratar-se de uma cenário onde regressão quantílica é mais indicado.

Como na regressão quantílica não se faz suposição sobre a distribuição dos erros do modelo para estimar os parâmetros, poderia se esperar que o EQM em C3 fosse semelhante àquele em C1, pois também há apenas uma tecnologia de produção. Porém, em C3 isso se confirma apenas para $\gamma = 0,9$. De fato, se $\gamma = 1$, a componente de erro aleatório deixa de existir e tem-se então um modelo com apenas a ineficiência. Se a distribuição do erro aleatório também tivesse mudando juntamente com a ineficiência, provavelmente a hipótese de que os resultados se assemelhariam com os obtidos em C1 se confirmasse também para $\gamma = 0,1$ e $\gamma = 0,5$.

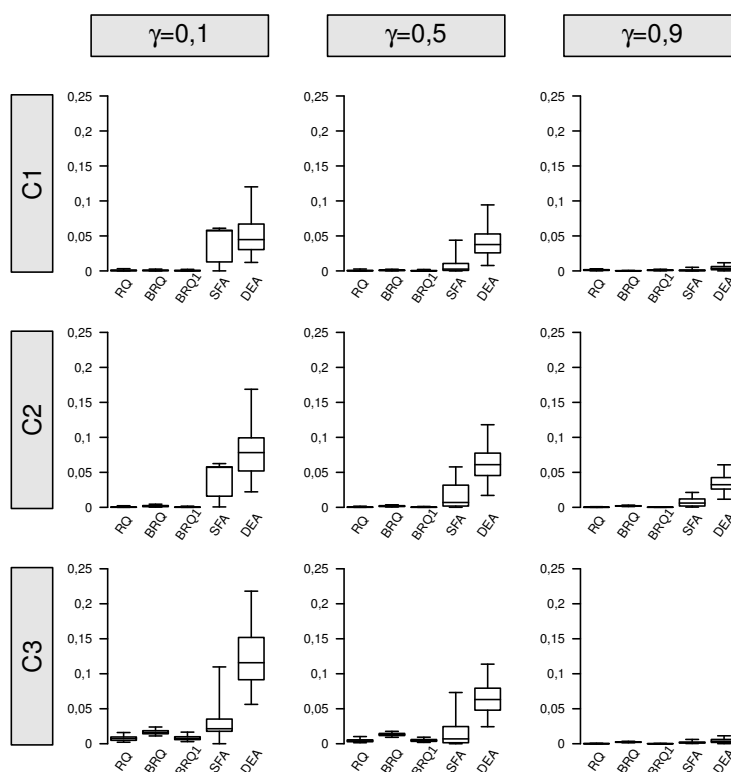


Figura 4.1: EQM da eficiência técnica ajustada pelos modelos em estudo em relação à eficiência técnica simulada nas 300 repetições. Nota: RQ = Regressão Quantílica; BRQ = Regressão Quantílica sem parâmetro de escala; BRQ1 = Regressão Quantílica com parâmetro de escala; SFA = Fronteira Estocástica; DEA = Análise Envoltória de Dados.

Na Figura 4.2 verifica-se que a medida que $\gamma \rightarrow 1$, o Coeficiente de Correlação de Pearson estimado (\hat{r}) se aproxima de 1 e os valores observados nas 300 repetições concentram-se mais em torno da mediana, para todos os modelos. Observa-se também que os modelos de regressão quantílica (BRQ, BRQ1 e RQ) e SFA apresentam níveis de correlação muito parecidos independente do cenário, indicando que os modelos classificam as DMUs de forma semelhante. Já DEA, apesar de apresentar valores de \hat{r} parecidos com aqueles da SFA e regressão quantílica quando os valores de γ são baixos, tende a classificar as DMUs de maneira menos satisfatória que SFA e regressão quantílica na medida que $\gamma \rightarrow 1$.

Para SFA e regressão quantílica, a correlação obtida foi a esperada, com base na Figura 4.1 e nos resultados de Behr (2010), de que a classificação é melhor para altos valores de γ . A correlação observada em C3 para SFA comporta-se de acordo o esperado pela observação de Parmeter e Kumbhakar (2014), de que a distribuição da ineficiência não afeta a classificação das DMUs. O comportamento da DEA também foi o esperado, considerando que esta técnica não modela erros aleatórios, portanto é indicada quando há apenas ineficiência.

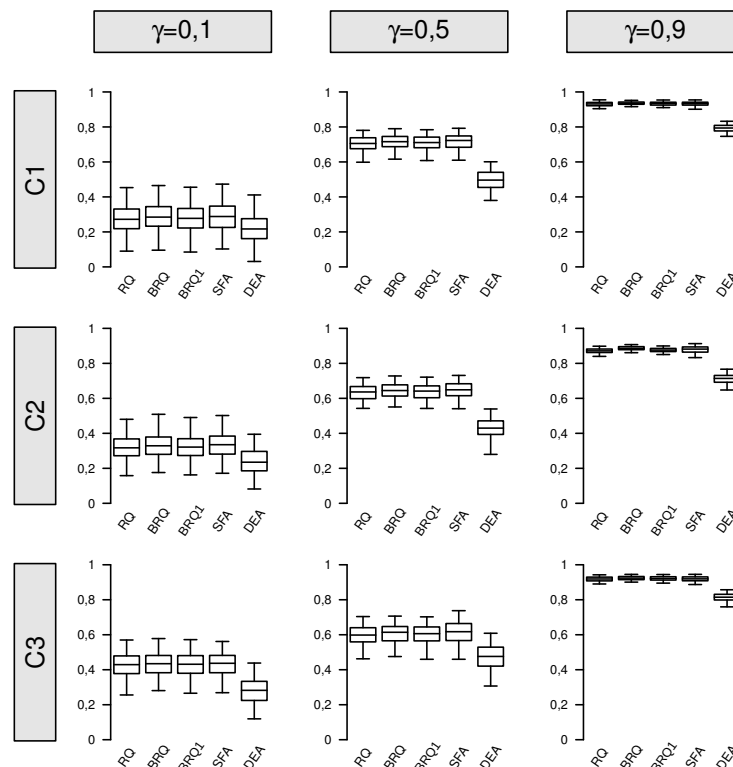


Figura 4.2: Coeficiente de Correlação de Pearson (\hat{r}) entre eficiência técnica simulada e ajustada pelos modelos nas 300 repetições. Nota: RQ = Regressão Quantílica; BRQ = Regressão Quantílica sem parâmetro de escala; BRQ1 = Regressão Quantílica com parâmetro de escala; SFA = Fronteira Estocástica; DEA = Análise Envoltória de Dados.

A Figura 4.3, mostra a qualidade da classificação em termos do percentual de DMUs entre as menos eficientes classificadas corretamente pelos modelos. Observa-se que com $\gamma \rightarrow 1$ a qualidade da classificação das DMUs entre os 70% menos eficientes melhora notavelmente. Por outro lado, o percentual de DMUs entre as mais eficientes, Figura 4.4, que foram classificadas corretamente, não parece ser influenciada por γ , mas, interessante, o modelo de regressão quantílica sob enfoque frequentista foi aquele que teve o pior desempenho em classificar os 10% mais eficientes. Em ambos os casos (Figuras 4.3 e 4.4), entretanto, a classificação nas caudas também não parece ser influenciada pela quebra de suposições.

Estes resultados também estão de acordo com o esperado pela análise do EQM e do coeficiente de correlação, para o caso da classificação das menos eficientes. Para o caso da classificação das mais eficientes, o resultado se deve ao fato que, para estas DMUs, os níveis de ineficiência são baixos. Assim, desvios devido a erros aleatórios podem ser tomados com sendo devido à ineficiência. Isso faz com que DMUs que na verdade têm eficiência próxima a 1 recebam níveis de eficiência mais baixos que os reais.

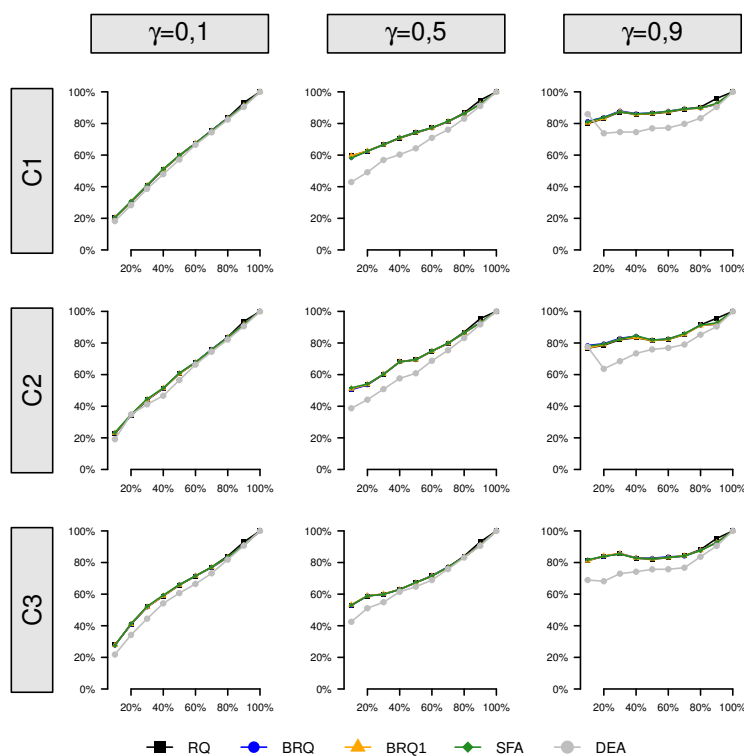


Figura 4.3: Concordância na classificação das DMUs menos eficientes tecnicamente. Nota: RQ = Regressão Quantílica; BRQ = Regressão Quantílica sem parâmetro de escala; BRQ1 = Regressão Quantílica com parâmetro de escala; SFA = Fronteira Estocástica; DEA = Análise Envoltória de Dados.

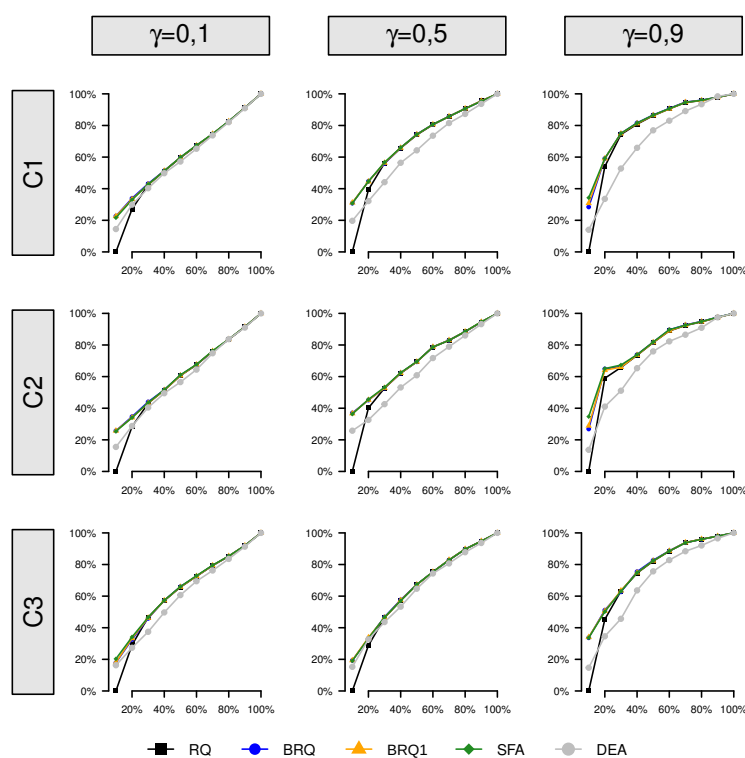


Figura 4.4: Concordância na classificação das DMUS mais eficientes tecnicamente. Nota: RQ = Regressão Quantílica; BRQ = Regressão Quantílica sem parâmetro de escala; BRQ1 = Regressão Quantílica com parâmetro de escala; SFA = Fronteira Estocástica; DEA = Análise Envoltória de Dados.

Os resultados obtidos quanto ao EQM (Figura 4.1) e correlação de Pearson (Figura 4.2) corroboram com a literatura existente em favor da regressão quantílica como modelo mais indicado em casos de heterogeneidade nas DMUs em estudo. Com relação ao modelo de regressão quantílica sob enfoque bayesiano, a parametrização da distribuição de Laplace assimétrica com parâmetro de escala é mais indicada, uma vez que esta parametrização mostrou resultados melhores ou iguais àqueles obtidos com a parametrização sem o referido parâmetro, tanto em relação ao viés quanto em relação à classificação das DMUs.

4.2 Aplicação aos dados da Agricultura Familiar

4.2.1 Análise exploratória dos dados

A Figura 4.5 traz os histogramas e algumas estatísticas descritivas dos dados da agricultura familiar brasileira, a nível de microrregião, em 2006. Observa-se que a distribuição das variáveis avaliadas apresenta uma cauda direita muito longa, isto é, possuem assimetria à direita, ou positiva, severa. A variável má-

quinas é aquela que possui a maior dispersão em relação à média (coeficiente de variação de 185,33%), enquanto a varável área possui a menor (coeficiente de variação de 94,08%). Diante de tal heterogeneidade, quantis diferentes da produção certamente sofrem efeitos diferentes por parte das variáveis independentes o que caracteriza um cenário em que a regressão quantílica é indicada.

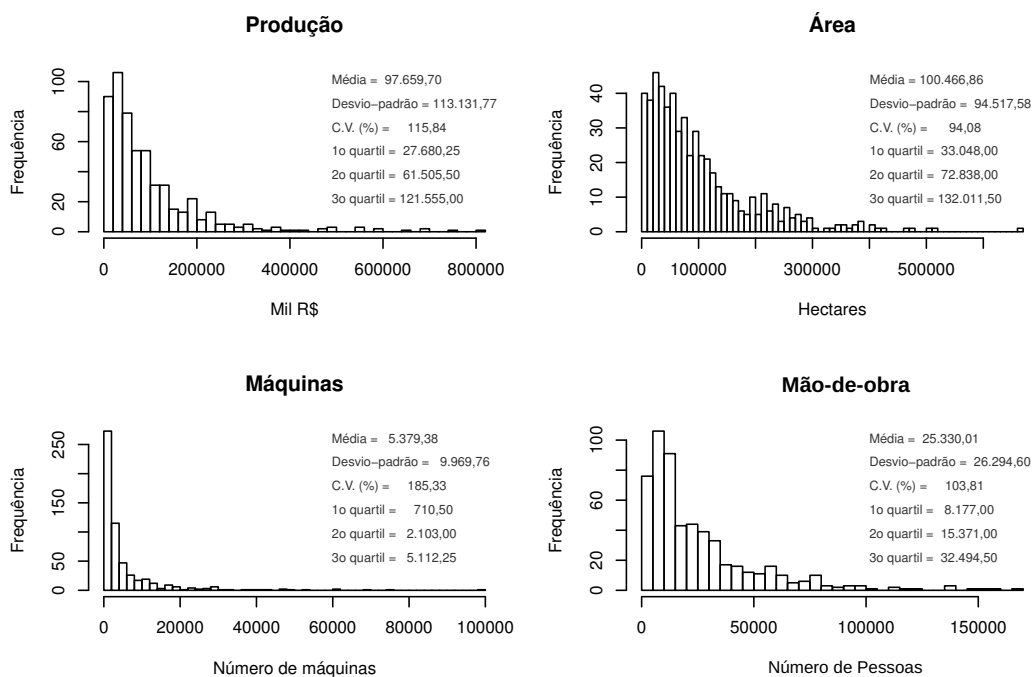


Figura 4.5: Histogramas com estatísticas descritivas dos dados da agricultura familiar, coletadas no Censo Agropecuário de 2006 pelo IBGE.

Através de um modelo de regressão quantílica frequentista, foi verificado se os efeitos dos insumos diferem ao longo dos quantis da produção para alguma das variáveis. A Figura 4.6, em que as bandas em cinza representam intervalos de confiança de *bootstrap* de 95%, traz o efeito de cada variável avaliado do quantil 0,05 ao quantil 0,95 em passos de 0,05.

Percebe-se que o efeito da área decresce na medida que se passa do quantil 0,05 para o quantil 0,95. Para microrregiões a partir do quantil 0,60 incrementos na área não apresentam efeito significativo na produção. Para a variável mão-de-obra, verifica-se o oposto. Na medida que se passa do quantil 0,05 para o quantil 0,95 o efeito desta variável cresce, com efeito positivo em todos os quantis. A variável máquinas, por outro lado, a medida que se passa do quantil 0,05 para o quantil 0,95 apresenta a forma de uma parábola com concavidade voltada para baixo. Esta variável possui efeito positivo em todos os quantis, sendo o pico registrado para o quantil 0,3 (Figura 4.6).

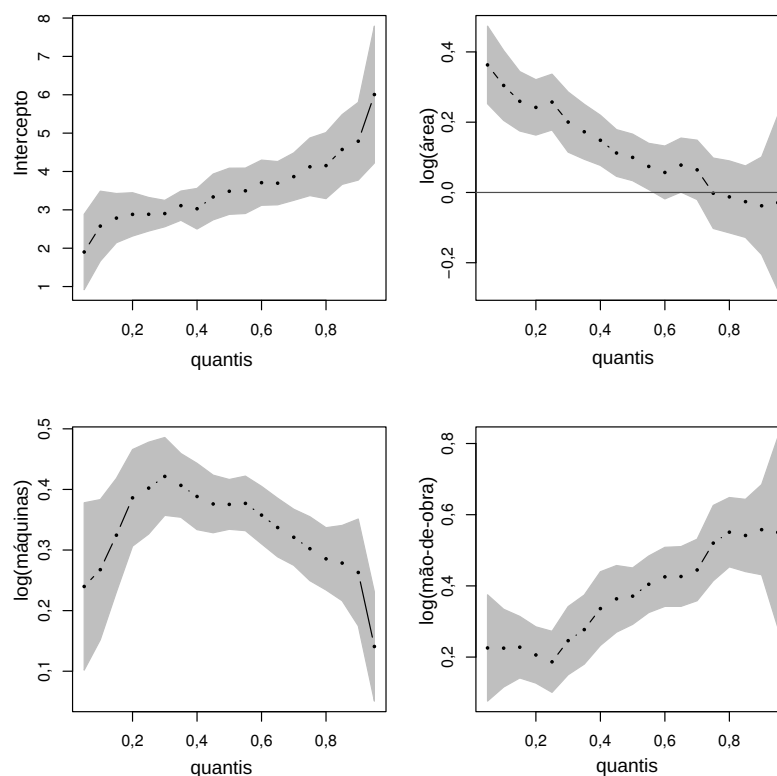


Figura 4.6: Efeito dos insumos na produção da agricultura familiar, condicionado ao nível de produção da microrregião, ajustado por regressão quantílica frequentista.

Percebe-se então que a tecnologia de produção, dada pelos coeficientes da função de Cobb-Douglas, daqueles com maior produção é diferente daquela presente nas microrregiões com menores níveis de produção. Isso justifica o uso de regressão quantílica para ajustar uma fronteira de produção mais flexível uma vez que neste caso a tecnologia de produção existente nas microrregiões com produção próximas à média não representa as melhores práticas. Com base nos resultados do estudo de simulação, optou-se por ajustar o modelo de regressão quantílica com parâmetro de escala, uma vez que este mostrou viés menor em todos os conjuntos de dados simulados e classificação semelhante aos demais.

4.2.2 Análise do modelo ajustado

Com base no teste de Raftery e Lewis, um *burn-in* de 423 iterações, e um total de 455.070 iterações, armazenando uma a cada 17 iterações, seria o necessário para que o quantil 0,025 da distribuição *a posteriori* dos parâmetros da fronteira fosse estimado com precisão de 0,05 e probabilidade de 95%. Cada cadeia de markov simulada com base nestes números geraria uma amostra de 26.743 valores da distribuição *a posteriori*. Todavia, optou-se por simular as cadeias com um

número maior de iterações e, então, foi tomado como *burn-in* 50 mil iterações, e um total de 1,7 milhão de iterações, armazenando-se uma a cada 17. Duas cadeias foram simuladas e mescladas após a análise convergência, resultando em uma amostra de 200 mil observações da distribuição *a posteriori* de cada parâmetro.

A Tabela 4.1 traz o teste de convergência de Gelman Rubin para os parâmetros do modelo, o qual mostra que houve convergência das duas cadeias geradas, pois os valores do fator de redução de escala potencial em cada parâmetro foi igual a 1. Todavia, pelo tamanho amostral efetivo (TAE) percebe-se que ainda existe autocorrelação nas cadeias, mesmo após o *thin* de tamanho 17 indicado pelo teste de Raftery e Lewis. Isso ocorre pois, quanto mais próximo do número de valores armazenados for o TAE, menor é a autocorrelação na amostra simulada, e neste caso estes dois números não estão próximos. Por exemplo, as 100 mil observações obtidas para estimar os efeitos da mão-de-obra, em cada cadeia, tem a mesma informação que 1.869 observações não correlacionadas. Entretanto, a autocorrelação é uma característica inerente aos algoritmos de MCMC e, como alguns autores argumentam, por exemplo Raftery e Lewis (1992), Link e Eaton (2012) e Gelman *et al.* (2014), uma vez que as cadeias tenham atingido convergência, as inferências podem ser realizadas.

Tabela 4.1: Resultados dos testes de convergência de Gelman e Rubin e tamanho amostral efetivo no modelo de regressão quantílica bayesiana.

	GR	TAE	
		Cadeia 1	Cadeia 2
intercepto	1,00	5.124	5.124
área	1,00	2.693	2.693
máquinas	1,00	10.365	10.365
mão-de-obra	1,00	1.869	1.869

GR - Teste de Gelman e Rubin; TAE - Tamanho Amostral Efetivo.

Tomando a média *a posteriori* como estimador de bayes, a Tabela 4.2 traz algumas informações sobre as estimativas dos parâmetros da fronteira ajustada no quantil 0,95. Através do erro de monte carlo (EC), é possível ver que as cadeias simuladas fornecem boa aproximação para as estimativas dos parâmetros. Além disso, percebe-se também que a variável área não afeta de forma significativa a produção, com base no quantil 0,95, e que a produção nas microrregiões encontra-se com retornos decrescentes de escala, pois $\sum_{j=1}^3 \hat{\beta}_j = 0,6935$.

Como a forma funcional de Cobb-Douglas foi utilizada, pode-se dizer então que, para as microrregiões com melhores práticas de produção, o aumento de 1% no uso de máquinas e implementos agrícolas, acarreta em um aumento de 0,1617% na produção, mantendo área e mão-de-obra fixos. Por outro lado, se um aumento de 1% ocorresse no uso de mão-de-obra, a produção, em média, aumentaria em 0,5658% (Tabela 4.2).

Mais ainda, uma vez feito o uso do enfoque bayesiano, pode-se fazer afirmações probabilísticas sobre tais retornos. O aumento de 1% no uso de mão-de-obra acarretaria aumento entre 0,4448% e 0,6910% na produção com probabilidade de 0,95, enquanto, com mesma probabilidade, o aumento de 1% em máquinas e implementos acarretaria aumento entre 0,1273% e 0,2004% (Tabela 4.2).

A Figura 4.7 traz as densidades *a posteriori* dos parâmetros da fronteira de produção ajustada no quantil 0,95. Enquanto nenhum dos parâmetros apresenta distribuição *a posteriori* nitidamente simétrica, a variável máquinas (β_2) é aquela que cuja distribuição apresenta maior assimetria.

Tabela 4.2: Estimativas e Erros padrão dos coeficientes da regressão quantílica sob enfoque bayesiano ajustada no quantil 0,95.

	Coef.	EP	HPD		EC
			LI	LS	
Intercepto	5,7448 ⁺	0,3429	5,0578	6,4090	0,0008
área	-0,0340	0,0440	-0,1195	0,0559	0,0001
máquinas	0,1617 ⁺	0,0199	0,1273	0,2004	0,0001
mão-de-obra	0,5658 ⁺	0,0627	0,4448	0,6910	0,0001

+ : Significativo com base no intervalo HPD de 95%.

EP: erro-padrão; LI: limite inferior; LS: limite superior; EC: erro de monte carlo.

Nota: Dados modelados em escala logarítmica.

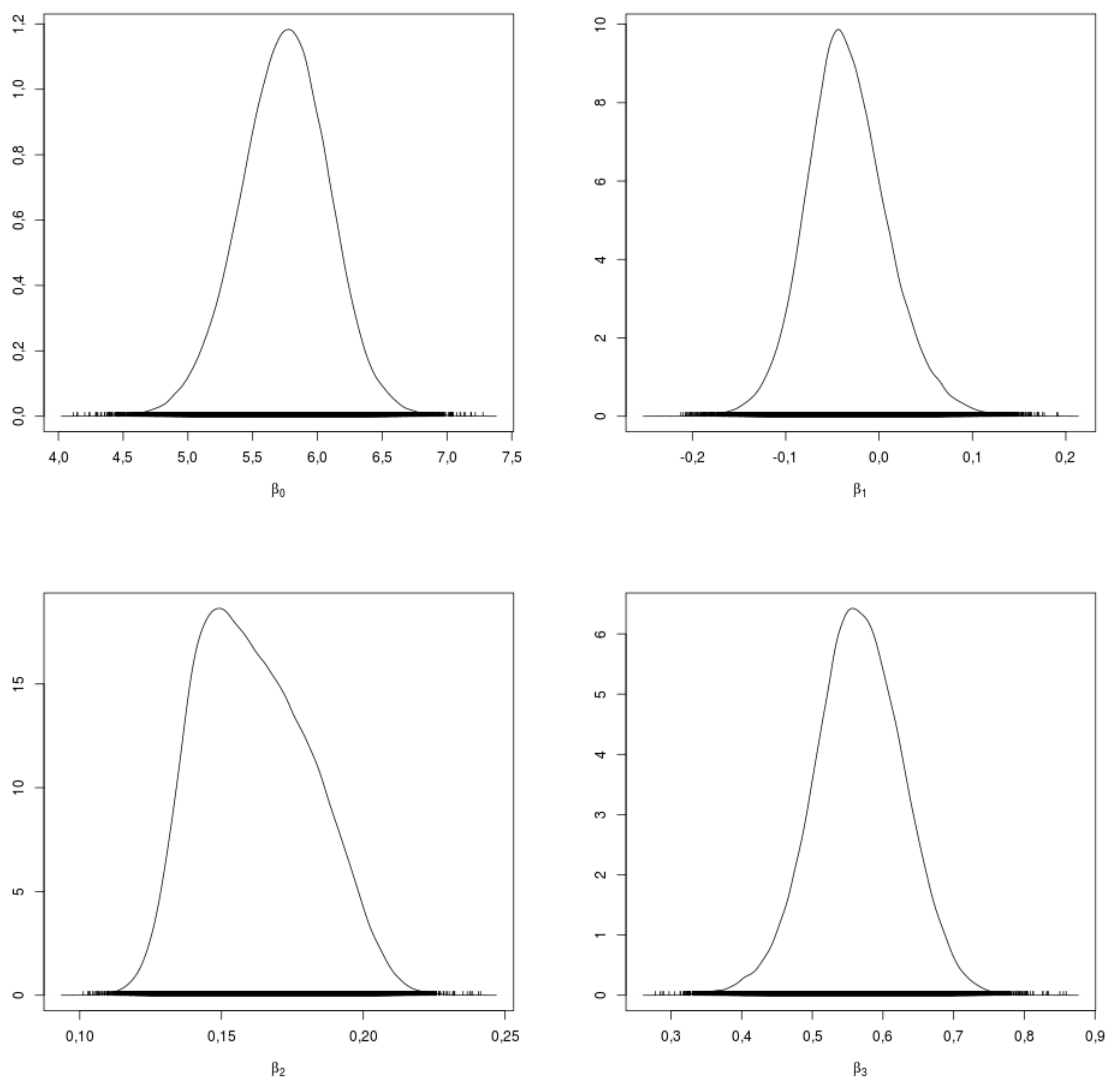


Figura 4.7: Densidades *a posteriori* dos parâmetros da fronteira de produção da agricultura familiar nas microrregiões brasileiras, em 2006, ajustada no quantil 0,95.

Com relação às eficiências técnicas ajustadas, pela Figura 4.8, percebe-se uma certa assimetria na distribuição das mesmas. Tal assimetria é esperada com base na assimetria verificada na distribuição dos dados e uma vez que o modelo ajusta um quantil alto da produção. A eficiência técnica mínima ajustada foi de 0,0619 e a mais alta 1,00. A média de eficiência técnica no país foi de 0,4109 com desvio-padrão de 0,2526. Ou seja, a agricultura familiar no Brasil, analisando-se a nível de microrregião, produzia, em 2006, apenas 41,09% do que poderia produzir com as mesmas quantidades de insumos utilizadas. Apenas 25% das microrregiões tem eficiência acima de 0,5614 (valor do terceiro quartil).

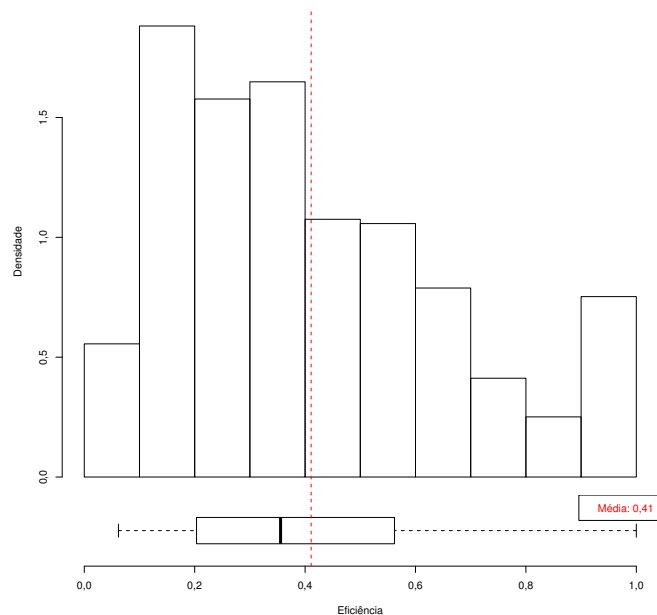
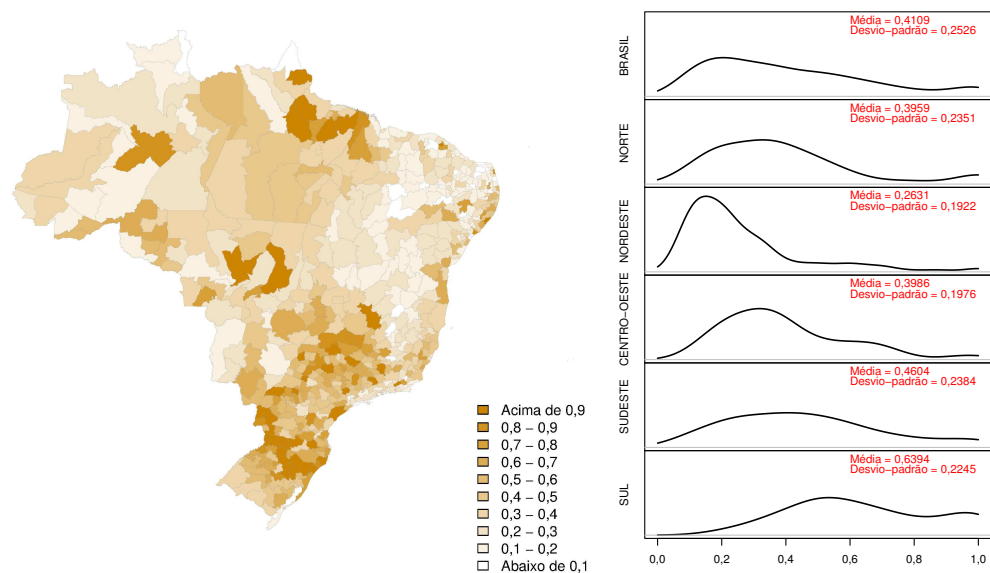


Figura 4.8: Histograma das eficiências técnicas ajustadas pela regressão quantílica no quantil 0,95.

4.2.3 Análise da eficiência técnica das microrregiões brasileiras

Ao verificar a distribuição geográfica da eficiência das microrregiões, Figura 4.9, pode-se perceber que há uma concentração de microrregiões com eficiência abaixo de 30% na região nordeste de Minas Gerais, no interior do estado da Bahia, no estado do Piauí, Paraíba, Rio Grande do Norte, partes de Pernambuco e Paraíba, isto é, no semiárido brasileiro. Enquanto isso, as áreas onde observa-se uma concentração de microrregiões com eficiência alta são ao norte do Rio Grande do Sul e sul de Santa Catarina, na região sudoeste do Paraná e na região nordeste do Pará. Das 28 microrregiões com eficiência igual a 1, 5 ficam Santa Catarina, 12 (distribuídas igualmente) em São Paulo, Rio Grande do Sul e Pará, 3 no Paraná, 2 em Minas Gerais e 6 (distribuídas igualmente) em Alagoas, Amapá, Ceará, Mato Grosso, Rio de Janeiro e Rio Grande do Norte.



(a) Distribuição geográfica da eficiência

(b) Distribuição da eficiência por região

Figura 4.9: Distribuição da eficiência técnica da agricultura familiar, em 2006, obtida por meio de regressão quantílica bayesiana ajustada no quantil 0,95.

A distribuição geográfica da eficiência técnica está de acordo com o esperado, com base na literatura especializada, de que a parte norte do país possui menores níveis de eficiência agrícola que o sul e sudeste. Ainda existe discrepância entre as regiões do Brasil, em termos de educação, infra-estrutura e etc, que impactam o desenvolvimento destas. Por exemplo, nas regiões sudeste e sul, que se destacam como as mais eficientes, foi onde o processo de modernização da agricultura ocorreu com mais afinco, (TEIXEIRA, 2005; SCHNEIDER, 2010), modernização esta, que alterou não apenas o processo de produção agrícola (MATOS; PESSOA, 2011; NAVARRO, 2001). O semiárido brasileiro, por outro lado, é historicamente conhecido por ser uma região com produção agrícola vulnerável devido à suas condições climáticas que, ao contrário de locais áridos em outros países, ainda carece de um sistema de produção adequado (COUTINHO *et al.*, 2013).

Capítulo 5

Considerações Finais

O uso da regressão quantílica sob o enfoque bayesiano foi apresentado como alternativa à regressão quantílica frequentista pelo fato de fornecer melhores propriedades, teóricas e práticas, mesmo em pequenas amostras. Trata-se de uma alternativa interessante para aqueles que não desejam justificar inferências sobre os parâmetros e incertezas sobre as estimativas pontuais da eficiência com a aplicação de *bootstrap*. O ajuste bayesiano permite ainda uma interpretação probabilística direta a respeito das estimativas intervalares e a possibilidade de incorporar conhecimentos adquiridos em outros estudos no processo de ajuste do modelo.

Os resultados apresentados no estudo de simulação corroboram com a literatura em favor da regressão quantílica. Nota-se que a regressão quantílica ajusta melhor a eficiência em todos cenários estudados, mostrando-se robusta à quebra de suposições aplicáveis a DEA e SFA, inclusive em hipóteses extremas de dominação do erro composto por erros aleatórios. A regressão quantílica sob enfoque bayesiano apresentou ajuste semelhante àquele sob enfoque frequentista, como esperado devido ao uso de distribuições *a priori* pouco informativas. Com relação as parametrizações da distribuição de Laplace assimétrica, a presença de um parâmetro de escala parece fornecer ajuste mais confiável para aplicações práticas, onde não se conhece a verdade sobre o cenário em que os dados são obtidos.

Ainda, a aplicação aos dados da agricultura familiar mostrou resultados coerentes com a literatura especializada, o que consiste em outro indício em favor da regressão quantílica bayesiana. Com relação à aplicação, entretanto, os dados são de 2006 e os níveis de eficiência técnica obtidos, bem como as informações sobre a fronteira de produção, certamente não são condizentes com o que existe hoje.

O uso de regressão quantílica para ajustar a eficiência é recente, de tal forma que ainda carecem desenvolvimentos. Por exemplo, calcular a eficiência

como $ET_i^{(\tau)} = \exp(\epsilon_i^{(\tau)})$, se $\exp(\epsilon_i^{(\tau)}) \leq 1$ e 1 caso contrário, não necessariamente é uma boa alternativa. Assim como observado na DEA, esse procedimento pode gerar muitos valores iguais a 1, podendo gerar aproximações pobres da distribuição (*a posteriori*) da eficiência quando se usa *bootstrap* (o enfoque bayesiano). Embora a quantidade de unidades produtivas com eficiência igual 1 depende de τ , isto é, o quantil considerado para ajustar a eficiência, valores de τ próximos a 1 fornecem menos unidades produtivas com eficiência igual a 1, considerar τ muito próximo de 1 não forneceria ajuste muito diferente do observado na DEA.

Como sugestões de trabalhos futuros, a investigação de formas de parametrizar o modelo de regressão quantílica para melhor adequar este modelo ao contexto de análise eficiência é extremamente válida, com respeito à maneira como a eficiência é obtida. Introduzir um erro composto, assim como acontece na SFA, em regressão quantílica parece ser algo que já está em desenvolvimento e pode fornecer resultados interessantes. Uma outra alternativa de desenvolvimento na regressão quantílica para melhorar a aplicação no contexto da análise eficiência é, como também acontece na SFA, introduzir uma forma de realizar os dois estágios da análise de uma única vez, isto é, permitir que os efeitos de variáveis que possam explicar a eficiência sejam estimados em paralelo à estimação dos parâmetros da fronteira.

Referências Bibliográficas

- AIGNER, D. *et al.* Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, Elsevier, v. 6, n. 1, p. 21–37, 1977.
- ALMANIDIS, P. *et al.* Stochastic frontier models with bounded inefficiency. In: *Festschrift in honor of Peter Schmidt*. [S.l.]: Springer, 2014. p. 47–81.
- ARROYO HINOSTROZA, A. A. *Regressão quantílica bayesiana em modelos de fronteira de produção estocástica*. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, 2017.
- BANKER, R. D. *et al.* Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, INFORMS, v. 30, n. 9, p. 1078–1092, 1984.
- BARRODALE, I. *et al.* Solution of an overdetermined system of equations in the 1 norm [f4]. *Communications of the ACM*, ACM, v. 17, n. 6, p. 319–320, 1974.
- BATTESE, G. E. *et al.* Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of econometrics*, Elsevier, v. 38, n. 3, p. 387–399, 1988.
- BATTESE, G. E. *et al.* Frontier production functions, technical efficiency and panel data: with application to paddy farmers in india. In: *International applications of productivity and efficiency analysis*. [S.l.]: Springer, 1992. p. 149–165.
- BATTESE, G. E. *et al.* A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical economics*, Springer, v. 20, n. 2, p. 325–332, 1995.
- BEHR, A. Quantile regression for robust bank efficiency score estimation. *European Journal of Operational Research*, Elsevier, v. 200, n. 2, p. 568–581, 2010.
- BERNINI, C. *et al.* Quantile estimation of frontier production function. *Empirical Economics*, Springer, v. 29, n. 2, p. 373–381, 2004.
- BOGETOFT, P. *et al.* *Benchmarking with Dea, Sfa, and R*. [S.l.]: Springer Science & Business Media, 2010. v. 157.
- BOLSTAD, W. M. *Introduction to Bayesian Statistics*. 2. ed. [S.l.]: Wiley, 2007.
- BROECK, J. Van den *et al.* Stochastic frontier models: A bayesian perspective. *Journal of Econometrics*, Elsevier, v. 61, n. 2, p. 273–303, 1994.
- CASELLA, G. *et al.* *Statistical inference*. [S.l.]: Duxbury Pacific Grove, CA, 2002. v. 2.
- CHARNES, A. *et al.* Measuring the efficiency of decision making units. *European journal of operational research*, Elsevier, v. 2, n. 6, p. 429–444, 1978.
- CHEN, C. *et al.* Computational issues for quantile regression. *Sankhyā: The Indian Journal of Statistics*, JSTOR, p. 399–417, 2005.

- CHERNICK, M. R. *et al.* An introduction to bootstrap methods with applications to r. Wiley Publishing, 2011.
- COELLI, T. *et al.* *frontier: Stochastic Frontier Analysis*. [S.l.], 2017. R package version 1.1-2. Disponível em: <<https://CRAN.R-Project.org/package=frontier>>.
- COELLI, T. J. *et al.* *An introduction to efficiency and productivity analysis*. [S.l.]: Springer Science & Business Media, 2005.
- COUTINHO, M. J. F. *et al.* A pecuária como atividade estabilizadora no semiárido brasileiro. *Veterinária e Zootecnia*, v. 20, n. 3, p. 434–441, 2013.
- DEBREU, G. The coefficient of resource utilization. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 273–292, 1951.
- FARAHMAND, M. *et al.* A combined data envelopment analysis and support vector regression for efficiency evaluation of large decision making units. *International Journal of Engineering and Technology (IJET)*, p. 2310–2321, 2014.
- FARRELL, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, JSTOR, v. 120, n. 3, p. 253–290, 1957.
- FENG, X. *et al.* Wild bootstrap for quantile regression. *Biometrika*, Oxford University Press, v. 98, n. 4, p. 995–999, 2011.
- GALVAO, A. F. *et al.* On bootstrap inference for quantile regression panel data: A monte carlo study. *Econometrics*, Multidisciplinary Digital Publishing Institute, v. 3, n. 3, p. 654–666, 2015.
- GELMAN, A. *et al.* *Bayesian data analysis*. [S.l.]: CRC press Boca Raton, FL, 2014. v. 3.
- GELMAN, A. *et al.* Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, International Society for Bayesian Analysis, v. 1, n. 3, p. 515–534, 2006.
- GELMAN, A. *et al.* Inference from iterative simulation using multiple sequences. *Statistical science*, JSTOR, p. 457–472, 1992.
- GREENE, W. H. A gamma-distributed stochastic frontier model. *Journal of Econometrics*, Elsevier, v. 46, n. 1-2, p. 141–163, 1990.
- GRIFFIN, J. E. *et al.* Flexible mixture modelling of stochastic frontiers. *Journal of Productivity Analysis*, Springer, v. 29, n. 1, p. 33–50, 2008.
- HAO, L. *et al.* *Quantile regression*. [S.l.]: Sage, 2007.
- HE, X. *et al.* Markov chain marginal bootstrap. *Journal of the American Statistical Association*, Taylor & Francis, v. 97, n. 459, p. 783–795, 2002.
- HORRACE, W. C. *et al.* A laplace stochastic frontier model. *Econometric Reviews*, Taylor & Francis, p. 1–27, 2015.
- HSU, A. *et al.* *Efficiency estimation with panel quantile regression: An application using longitudinal data from nursing homes in Ontario, Canada*. [S.l.], 2017.

- JONDROW, J. *et al.* On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, Elsevier, v. 19, n. 2-3, p. 233–238, 1982.
- KADITI, E. A. *et al.* Applying regression quantiles to farm efficiency estimation. In: *2010 Annual Meeting, July*. [S.l.: s.n.], 2010. p. 25–27.
- KNOX, K. J. *et al.* Technical efficiency in texas nursing facilities: a stochastic production frontier approach. *Journal of economics and finance*, Springer, v. 31, n. 1, p. 75–86, 2007.
- KOENKER, R. *quantreg: Quantile Regression*. [S.l.], 2017. R package version 5.34. Disponível em: <<https://CRAN.R-project.org/package=quantreg>>.
- KOENKER, R. *et al.* Regression quantiles. *Econometrica: journal of the Econometric Society*, JSTOR, p. 33–50, 1978.
- KOENKER, R. *et al.* Robust tests for heteroscedasticity based on regression quantiles. *Econometrica: Journal of the Econometric Society*, JSTOR, p. 43–61, 1982.
- KOENKER, R. *et al.* Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, Taylor & Francis Group, v. 94, n. 448, p. 1296–1310, 1999.
- KOENKER, R. W. *et al.* Algorithm as 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, JSTOR, v. 36, n. 3, p. 383–393, 1987.
- KOOPMANS, T. C. An analysis of production as efficient combination of activities. In: *Activity Analysis of Production and Allocation, volume 13 Cowles Commission for Research in Economics Monographs*. [S.l.]: John Wiley & Sons, 1951.
- LEI, D. *et al.* *A robust algorithm for least absolute deviations curve fitting*. [S.l.], 2001.
- LINK, W. A. *et al.* On thinning of chains in mcmc. *Methods in ecology and evolution*, Wiley Online Library, v. 3, n. 1, p. 112–115, 2012.
- LIU, C. *et al.* The quantile regression approach to efficiency measurement: insights from monte carlo simulations. *Health economics*, Wiley Online Library, v. 17, n. 9, p. 1073–1087, 2008.
- LIU, R. Y. *et al.* Bootstrap procedures under some non-iid models. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 16, n. 4, p. 1696–1708, 1988.
- LUO, Y. *et al.* Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, Taylor & Francis, v. 82, n. 11, p. 1635–1649, 2012.
- MATOS, P. F. *et al.* A modernização da agricultura no brasil e os novos usos do território. *Geo Uerj*, v. 2, n. 22, p. 290–322, 2011.
- MEDRANO, L. *et al.* Critérios baseados na “deviance” para comparação de modelos bayesianos de fronteira de produção estocástica. *Publicação interna*, v. 176, 2004.

- MEEUSEN, W. *et al.* Efficiency estimation from cobb-douglas production functions with composed error. *International Economic Review*, JSTOR, p. 435–444, 1977.
- NAVARRO, Z. Desenvolvimento rural no brasil: os limites do passado e os caminhos do futuro. *Estudos avançados*, SciELO Brasil, v. 15, n. 43, p. 83–100, 2001.
- NOGUEIRA, D. A. *et al.* Avaliação de critérios de convergência univariados para o método de monte carlo via cadeias de markov. *Revista Brasileira de Estatística*, IBGE, v. 65, n. 224, p. 59–88, 2004.
- NTZOUFRAS, I. *Bayesian modeling using WinBUGS*. [S.l.]: John Wiley & Sons, 2009. 506 p.
- PARMETER, C. F. *et al.* Efficiency analysis: a primer on recent advances. *Foundations and Trends® in Econometrics*, Now Publishers, Inc., v. 7, n. 3–4, p. 191–385, 2014.
- PARZEN, M. *et al.* A resampling method based on pivotal estimating functions. *Biometrika*, Oxford University Press, v. 81, n. 2, p. 341–350, 1994.
- PLUMMER, M. *rjags: Bayesian Graphical Models using MCMC*. [S.l.], 2016. R package version 4-6. Disponível em: <<https://CRAN.R-project.org/package=rjags>>.
- PLUMMER, M. *et al.* *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*. [S.l.], 2003. v. 124, 125 p.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>.
- RAFTERY, A. E. *et al.* How many iterations in the gibbs sampler. *Bayesian Statistics*, Oxford University Press, Oxford, Reino Unido, v. 4, p. 763–773, 1992.
- RASMUSSEN, S. *Production economics: the basic theory of production optimisation*. [S.l.]: Springer Science & Business Media, 2012.
- REICH, B. J. *et al.* Bayesian spatial quantile regression. *Journal of the American Statistical Association*, Taylor & Francis, v. 106, n. 493, p. 6–20, 2011.
- SAMUELSON, P. A. *et al.* *Economics*. 19. ed. [S.l.]: McGraw-Hill/Irwin, 2010.
- SCHNEIDER, S. Situando o desenvolvimento rural no brasil: o contexto e as questões em debate. *Revista de economia política*, SciELO Brasil, v. 30, n. 3, p. 511–531, 2010.
- SIMAR, L. *et al.* Sensitivity analysis of efficiency scores: How to bootstrap in non-parametric frontier models. *Management science*, INFORMS, v. 44, n. 1, p. 49–61, 1998.
- SIMM, J. *et al.* *rDEA: Robust Data Envelopment Analysis (DEA) for R*. [S.l.], 2016. R package version 1.2-5. Disponível em: <<https://CRAN.R-project.org/package=rDEA>>.
- STEVENSON, R. E. Likelihood functions for generalized stochastic frontier estimation. *Journal of econometrics*, Elsevier, v. 13, n. 1, p. 57–66, 1980.

TEIXEIRA, J. C. Modernização da agricultura no brasil: impactos econômicos, sociais e ambientais. *Revista Eletrônica AGB-TL*, v. 1, n. 2, p. 21–42, 2005.

TSIONAS, E. G. Efficiency measurement with the weibull stochastic frontier. *Oxford Bulletin of Economics and Statistics*, Wiley Online Library, v. 69, n. 5, p. 693–706, 2007.

WANG, J. Bayesian quantile regression for parametric nonlinear mixed effects models. *Statistical Methods & Applications*, Springer, v. 21, n. 3, p. 279–295, 2012.

WEN, M. *Uncertain Data Envelopment Analysis*. [S.l.]: Springer, 2015.

YU, K. *et al.* Bayesian quantile regression. *Statistics & Probability Letters*, Elsevier, v. 54, n. 4, p. 437–447, 2001.

YUAN, Y. *et al.* Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, Wiley Online Library, v. 66, n. 1, p. 105–114, 2010.

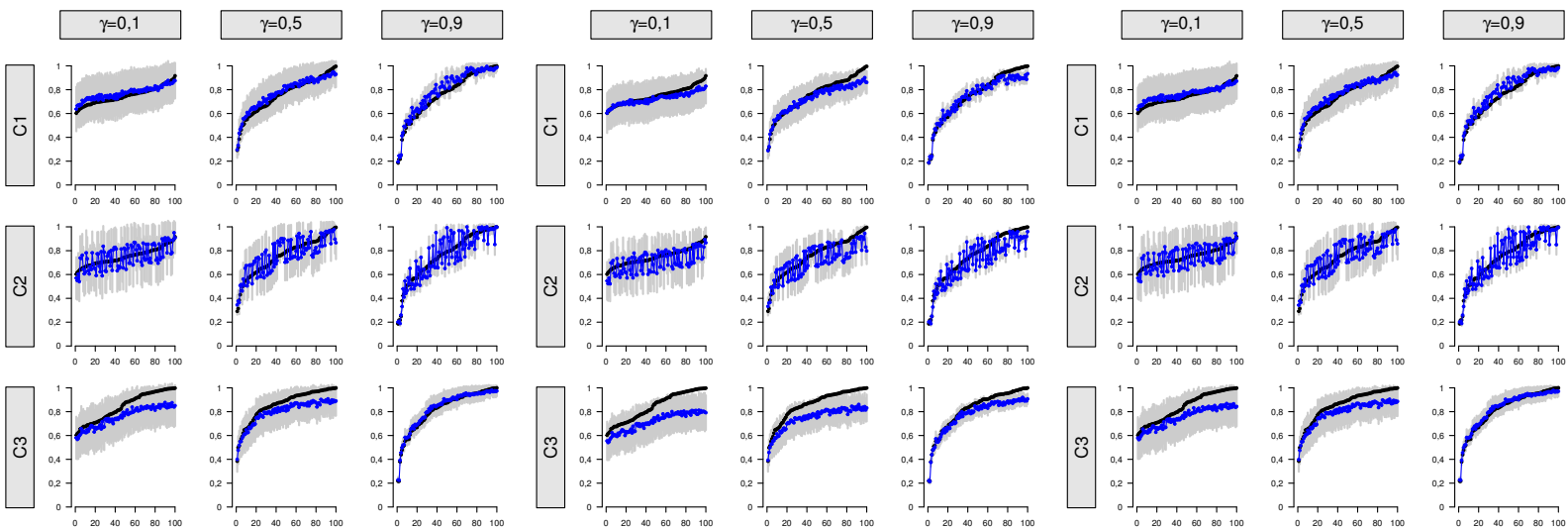


Figura A.1: Eficiência técnica ajustada por regressão quantílica, no quantil 0,8, (em azul, média das 300 repetições) e simulada (preto). Em cinza, um desvio padrão acima e abaixo da média de eficiência ajustada. No centro, Regressão quantílica bayesiana sem parâmetro de escala, à esquerda regressão quantílica frequentista e à direita regressão quantílica bayesiana com parâmetro de escala.

Nota:

C1: DMUs homogêneas; C2: Heterogeneidade da tecnologia de produção; C3: heterogeneidade na distribuição da ineficiência; γ = razão da variância da ineficiência pela variância do erro composto.

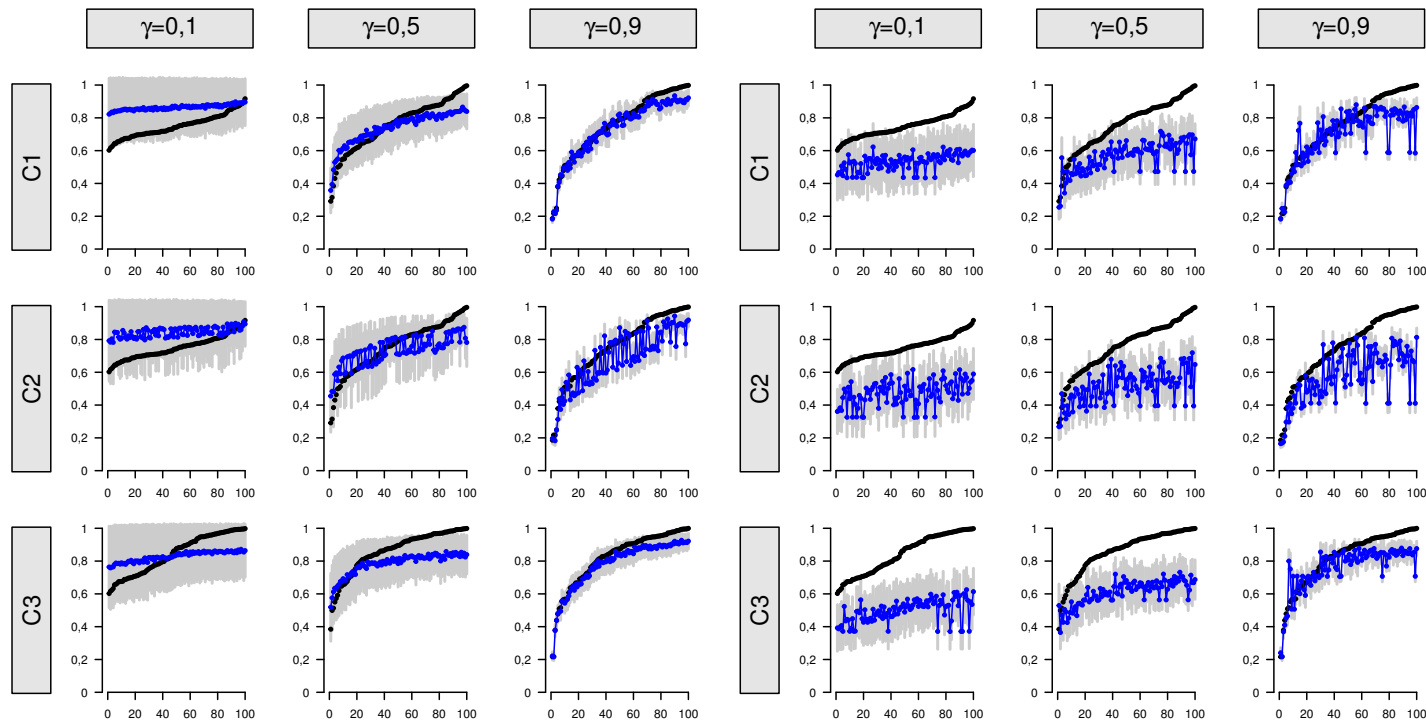


Figura A.2: Eficiência técnica ajustada por Fronteira Estocástica (à esquerda) e DEA (à direita), em azul, e eficiência simulada em preto. Em cinza, um desvio padrão acima e abaixo da média de eficiência ajustada.

Nota:

C1: DMUs homogêneas; C2: heterogeneidade da tecnologia de produção; C3: heterogeneidade na distribuição da ineficiência; γ = razão da variância da ineficiência pela variância do erro composto.

Apêndice B: Traceplots dos parâmetros da fronteira ajustada na aplicação aos dados da agricultura familiar brasileira.

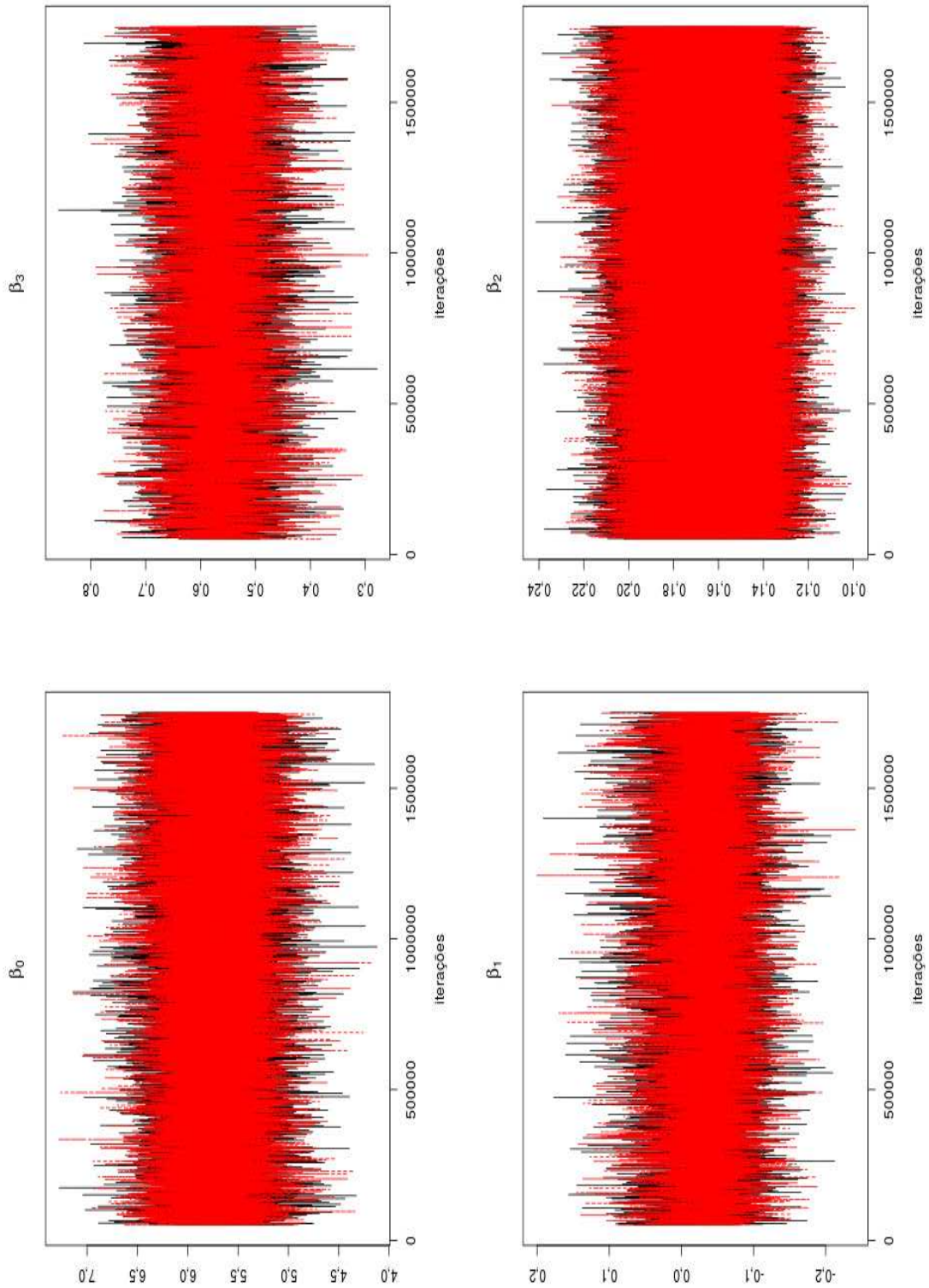


Figura B.1: Traceplots das cadeias para os coeficientes da regressão quantílica no quantil 0,95 ajustada na aplicação.

Apêndice C: Eficiências técnicas ajustadas para as microrregiões

Tabela C.1: Eficiências técnicas ajustadas para cada microrregião, em ordem alfabética de estado e microrregião no estado, pelo modelo de Regressão Quantílica Bayesiana com parâmetro de escala e respectivos Limites Inferiores (LI) e Superiores (LS) dos intervalos de credibilidade HPD de 95%.

Microrregião	Eficiência	LI	LS	Microrregião	Eficiência	LI	LS
AC - Brasília	0,505	0,423	0,594	PB - Catolé do Rocha	0,134	0,122	0,146
AC - Cruzeiro do Sul	0,265	0,229	0,303	PB - Curimataú Ocidental	0,301	0,273	0,329
AC - Rio Branco	0,533	0,470	0,595	PB - Curimataú Oriental	0,072	0,064	0,079
AC - Sena Madureira	0,162	0,146	0,178	PB - Esperança	0,063	0,054	0,073
AC - Tarauacá	0,312	0,257	0,373	PB - Guarabira	0,124	0,108	0,141
AL - Alagoano do Sertão do São Francisco	0,120	0,111	0,130	PB - Itabaiana	0,097	0,084	0,110
AL - Arapiraca	0,301	0,235	0,369	PB - Itaporanga	0,125	0,114	0,136
AL - Batalha	0,210	0,193	0,228	PB - João Pessoa	0,582	0,486	0,685
AL - Litoral Norte Alagoano	0,467	0,398	0,548	PB - Litoral Norte	0,153	0,132	0,174
AL - Maceió	1,000	1,000	1,000	PB - Litoral Sul	0,246	0,207	0,288
AL - Mata Alagoana	0,469	0,423	0,514	PB - Patos	0,121	0,110	0,132
AL - Palmeira dos Índios	0,192	0,164	0,220	PB - Piancó	0,155	0,143	0,167
AL - Penedo	0,305	0,251	0,358	PB - Sapé	0,232	0,200	0,265
AL - Santana do Ipanema	0,194	0,172	0,218	PB - Seridó Ocidental Paraibano	0,075	0,067	0,084
AL - São Miguel dos Campos	0,614	0,553	0,673	PB - Seridó Oriental Paraibano	0,111	0,103	0,119
AL - Serrana do Sertão Alagoano	0,079	0,070	0,087	PB - Serra do Teixeira	0,101	0,089	0,113
AL - Serrana dos Quilombos	0,717	0,638	0,799	PB - Sousa	0,304	0,281	0,327
AL - Traipu	0,066	0,059	0,073	PB - Umbuzeiro	0,335	0,288	0,383
AM - Alto Solimões	0,354	0,269	0,445	PE - Alto Capibaribe	0,118	0,100	0,135
AM - Boca do Acre	0,132	0,111	0,157	PE - Araripina	0,150	0,126	0,174
AM - Coari	0,816	0,679	0,974	PE - Brejo Pernambucano	0,335	0,290	0,379
AM - Itacoatiara	0,202	0,173	0,235	PE - Fernando de Noronha	0,092	0,057	0,129
AM - Japurá	0,169	0,133	0,214	PE - Garanhuns	0,227	0,189	0,267
AM - Juruá	0,380	0,308	0,457	PE - Itamaracá	0,458	0,370	0,556
AM - Madeira	0,217	0,187	0,248	PE - Itaparica	0,275	0,246	0,305
AM - Manaus	0,357	0,305	0,410	PE - Mata Meridional Pernambucana	0,630	0,537	0,723
AM - Parintins	0,438	0,355	0,526	PE - Mata Setentrional Pernambucana	0,511	0,435	0,585
AM - Purus	0,193	0,165	0,226	PE - Médio Capibaribe	0,587	0,470	0,704
AM - Rio Negro	0,268	0,177	0,375	PE - Pajeú	0,266	0,228	0,306
AM - Rio Preto da Eva	0,181	0,160	0,202	PE - Petrolina	0,522	0,452	0,595
AM - Tefé	0,153	0,134	0,174	PE - Recife	0,332	0,273	0,390
AP - Amapá	1,000	0,812	1,000	PE - Salgueiro	0,098	0,088	0,108
AP - Macapá	0,376	0,316	0,441	PE - Sertão do Moxotó	0,714	0,634	0,795
AP - Mazagão	0,285	0,233	0,347	PE - Suape	0,495	0,423	0,575
AP - Oiapoque	0,072	0,060	0,085	PE - Vale do Ipanema	0,198	0,179	0,218
BA - Alagoinhas	0,565	0,464	0,665	PE - Vale do Ipojuca	0,325	0,276	0,373
BA - Barra	0,287	0,243	0,331	PE - Vitória de Santo Antão	0,303	0,256	0,348
BA - Barreiras	0,260	0,229	0,292	PI - Alto Médio Canindé	0,123	0,103	0,144
BA - Bom Jesus da Lapa	0,127	0,111	0,143	PI - Alto Médio Gurguéia	0,129	0,113	0,146
BA - Boquira	0,100	0,087	0,115	PI - Alto Parnaíba Piauiense	0,214	0,186	0,247
BA - Brumado	0,105	0,091	0,119	PI - Baixo Parnaíba Piauiense	0,154	0,115	0,191
BA - Catu	0,240	0,215	0,265	PI - Bertolínia	0,183	0,156	0,215
BA - Cotegipe	0,126	0,107	0,144	PI - Campo Maior	0,148	0,115	0,183
BA - Entre Rios	0,189	0,167	0,212	PI - Chapadas do Extremo Sul Piauiense	0,132	0,116	0,148
BA - Euclides da Cunha	0,218	0,181	0,256	PI - Floriano	0,074	0,066	0,081
BA - Feira de Santana	0,206	0,158	0,255	PI - Litoral Piauiense	0,209	0,165	0,253
BA - Guanambi	0,227	0,192	0,265	PI - Médio Parnaíba Piauiense	0,145	0,122	0,168
BA - Ilhéus-Itabuna	0,685	0,554	0,821	PI - Picos	0,180	0,149	0,211
BA - Irecê	0,188	0,156	0,222	PI - Pio IX	0,091	0,081	0,101
BA - Itaberaba	0,240	0,207	0,273	PI - São Raimundo Nonato	0,095	0,083	0,108
BA - Itapetinga	0,259	0,226	0,297	PI - Teresina	0,105	0,084	0,125
BA - Jacobina	0,289	0,244	0,333	PI - Valença do Piauí	0,123	0,104	0,142
BA - Jequié	0,589	0,478	0,702	PR - Apucarana	0,627	0,585	0,666
BA - Jeremoabo	0,282	0,253	0,311	PR - Assaí	0,518	0,467	0,572
BA - Juazeiro	0,372	0,316	0,430	PR - Astorga	0,743	0,689	0,794

BA - Livramento do Brumado	0,177	0,160	0,194	PR - Campo Mourão	0,550	0,510	0,589
BA - Paulo Afonso	0,611	0,557	0,663	PR - Capanema	0,553	0,514	0,595
BA - Porto Seguro	0,357	0,308	0,406	PR - Cascavel	0,906	0,830	0,988
BA - Ribeira do Pombal	0,354	0,288	0,421	PR - Cerro Azul	0,236	0,207	0,271
BA - Salvador	0,103	0,083	0,123	PR - Cianorte	0,765	0,703	0,823
BA - Santa Maria da Vitória	0,194	0,168	0,221	PR - Cornélio Procópio	0,661	0,615	0,705
BA - Santo Antônio de Jesus	0,675	0,524	0,825	PR - Curitiba	0,647	0,591	0,703
BA - Seabra	0,197	0,166	0,228	PR - Faxinal	0,465	0,430	0,499
BA - Senhor do Bonfim	0,230	0,194	0,267	PR - Floráí	0,548	0,477	0,622
BA - Serrinha	0,209	0,171	0,246	PR - Foz do Iguaçu	0,810	0,754	0,864
BA - Valença	0,345	0,284	0,409	PR - Francisco Beltrão	1,000	0,962	1,000
BA - Vitória da Conquista	0,200	0,166	0,234	PR - Goioerê	0,617	0,569	0,661
CE - Baixo Curu	0,172	0,144	0,200	PR - Guarapuava	0,471	0,422	0,524
CE - Baixo Jaguaribe	0,272	0,240	0,306	PR - Ibaiti	0,370	0,340	0,399
CE - Barro	0,220	0,191	0,248	PR - Irati	0,669	0,623	0,715
CE - Baturité	0,321	0,270	0,375	PR - Ivaiporã	0,513	0,468	0,561
CE - Brejo Santo	0,157	0,140	0,173	PR - Jacarezinho	0,428	0,395	0,460
CE - Canindé	0,174	0,154	0,194	PR - Jaguaraiá	0,422	0,387	0,457
CE - Caririáçu	0,095	0,080	0,110	PR - Lapa	0,374	0,342	0,411
CE - Cariri	0,143	0,120	0,167	PR - Londrina	0,733	0,684	0,781
CE - Cascavel	0,118	0,105	0,130	PR - Maringá	0,560	0,513	0,608
CE - Chapada do Araripe	0,155	0,138	0,172	PR - Palmas	0,311	0,289	0,332
CE - Chorozinho	0,112	0,101	0,123	PR - Paranaguá	0,307	0,269	0,346
CE - Coreaú	0,104	0,090	0,119	PR - Paranaíba	0,922	0,846	0,999
CE - Fortaleza	0,232	0,188	0,276	PR - Pato Branco	0,676	0,627	0,724
CE - Ibiapaba	0,420	0,339	0,500	PR - Pitanga	0,562	0,520	0,605
CE - Iguatu	0,312	0,271	0,352	PR - Ponta Grossa	0,851	0,797	0,907
CE - Ipu	0,382	0,321	0,441	PR - Porecatu	1,000	1,000	1,000
CE - Itapipoca	0,260	0,212	0,308	PR - Prudentópolis	0,549	0,503	0,598
CE - Lavras da Mangabeira	0,171	0,157	0,184	PR - Rio Negro	0,422	0,390	0,452
CE - Litoral de Aracati	0,141	0,126	0,157	PR - São Mateus do Sul	0,470	0,436	0,505
CE - Litoral de Camocim e Acaraú	0,135	0,108	0,161	PR - Telêmaco Borba	0,227	0,210	0,243
CE - Médio Curu	0,783	0,698	0,872	PR - Toledo	1,000	1,000	1,000
CE - Médio Jaguaribe	0,307	0,279	0,334	PR - Umuarama	0,511	0,460	0,564
CE - Meruoca	0,088	0,073	0,103	PR - União da Vitória	0,502	0,469	0,534
CE - Pacajus	0,122	0,100	0,146	PR - Wenceslau Braz	0,465	0,433	0,494
CE - Santa Quitéria	0,185	0,168	0,203	RJ - Bacia de São João	0,183	0,149	0,222
CE - Serra do Pereiro	0,099	0,090	0,108	RJ - Baía da Ilha Grande	0,246	0,197	0,299
CE - Sertão de Cratêus	0,340	0,283	0,399	RJ - Barra do Pirai	0,194	0,162	0,231
CE - Sertão de Inhamuns	0,227	0,193	0,261	RJ - Campos dos Goytacazes	0,434	0,393	0,475
CE - Sertão de Quixeramobim	0,299	0,254	0,343	RJ - Cantagalo-Cordeiro	0,082	0,074	0,090
CE - Sertão de Senador Pompeu	0,234	0,199	0,271	RJ - Itaguaí	0,104	0,086	0,124
CE - Sobral	0,097	0,082	0,112	RJ - Itaperuna	0,318	0,289	0,346
CE - Uruburetama	1,000	1,000	1,000	RJ - Lagos	0,108	0,092	0,125
CE - Várzea Alegre	0,177	0,154	0,199	RJ - Macacu-Caceribu	0,165	0,143	0,188
DF - Brasília	0,384	0,336	0,435	RJ - Macaé	0,180	0,151	0,211
ES - Afonso Cláudio	0,714	0,650	0,778	RJ - Nova Friburgo	0,728	0,669	0,787
ES - Alegre	0,395	0,357	0,432	RJ - Rio de Janeiro	0,237	0,206	0,268
ES - Barra de São Francisco	0,322	0,286	0,361	RJ - Santa Maria Madalena	0,363	0,312	0,413
ES - Cachoeiro de Itapemirim	0,431	0,391	0,471	RJ - Santo Antônio de Pádua	0,577	0,517	0,634
ES - Colatina	0,521	0,470	0,575	RJ - Serrana	1,000	0,881	1,000
ES - Guarapari	0,547	0,497	0,599	RJ - Três Rios	0,242	0,200	0,289
ES - Itapemirim	0,686	0,613	0,764	RJ - Vale do Paraíba Fluminense	0,371	0,318	0,425
ES - Linhares	0,479	0,433	0,523	RJ - Vassouras	0,262	0,223	0,303
ES - Montanha	0,272	0,241	0,303	RN - Agreste Potiguar	0,127	0,113	0,141
ES - Nova Venécia	0,427	0,389	0,465	RN - Angicos	0,097	0,086	0,108
ES - Santa Teresa	0,463	0,426	0,502	RN - Baixa Verde	0,068	0,062	0,075
ES - São Mateus	0,401	0,367	0,436	RN - Borborema Potiguar	0,148	0,137	0,159
ES - Vitória	0,246	0,209	0,286	RN - Chapada do Apodi	0,197	0,181	0,212
GO - Anápolis	0,689	0,619	0,758	RN - Litoral Nordeste	0,196	0,177	0,216
GO - Anicuns	0,411	0,358	0,465	RN - Litoral Sul	0,090	0,076	0,104
GO - Aragarças	0,352	0,289	0,415	RN - Macaíba	0,264	0,234	0,294
GO - Catalão	0,384	0,327	0,437	RN - Macau	0,081	0,067	0,095
GO - Ceres	0,483	0,427	0,535	RN - Médio Oeste	0,102	0,091	0,112
GO - Chapada dos Veadeiros	0,164	0,139	0,190	RN - Mossoró	0,223	0,206	0,240
GO - Entorno de Brasília	0,325	0,288	0,361	RN - Natal	1,000	1,000	1,000
GO - Goiânia	0,656	0,593	0,714	RN - Pau dos Ferros	0,152	0,141	0,162
GO - Iporá	0,275	0,236	0,315	RN - Seridó Ocidental	0,155	0,132	0,177
GO - Meia Ponte	0,611	0,525	0,688	RN - Seridó Oriental	0,190	0,171	0,209

GO - Pires do Rio	0,546	0,481	0,605	RN - Serra de Santana	0,097	0,088	0,105
GO - Porangatu	0,357	0,302	0,411	RN - Serra de São Miguel	0,131	0,114	0,147
GO - Quirinópolis	0,388	0,321	0,454	RN - Umarizal	0,165	0,151	0,179
GO - Rio Vermelho	0,338	0,289	0,385	RN - Vale do Açu	0,203	0,184	0,223
GO - São Miguel do Araguaia	0,253	0,213	0,294	RO - Alvorada D'Oeste	0,524	0,445	0,606
GO - Sudoeste de Goiás	0,668	0,569	0,757	RO - Ariquemes	0,353	0,301	0,407
GO - Vale do Rio dos Bois	0,576	0,501	0,649	RO - Cacoal	0,517	0,441	0,594
GO - Vão do Paranã	0,150	0,132	0,168	RO - Colorado do Oeste	0,419	0,360	0,474
MA - Aglomeração Urbana de São Luís	0,258	0,190	0,327	RO - Guajará-Mirim	0,175	0,150	0,202
MA - Alto Mearim e Grajaú	0,384	0,313	0,458	RO - Ji-Paraná	0,605	0,499	0,712
MA - Baixada Maranhense	0,334	0,253	0,414	RO - Porto Velho	0,619	0,523	0,717
MA - Baixo Parnaíba Maranhense	0,301	0,234	0,372	RO - Vilhena	0,262	0,229	0,295
MA - Caxias	0,216	0,178	0,254	RR - Boa Vista	0,109	0,090	0,130
MA - Chapadas das Mangabeiras	0,229	0,195	0,267	RR - Caracará	0,159	0,135	0,185
MA - Chapadas do Alto Itapeturu	0,301	0,256	0,344	RR - Nordeste de Roraima	0,095	0,079	0,113
MA - Chapadinha	0,234	0,183	0,286	RR - Sudeste de Roraima	0,235	0,198	0,278
MA - Codó	0,430	0,337	0,527	RS - Cachoeira do Sul	0,471	0,436	0,507
MA - Coelho Neto	0,187	0,144	0,232	RS - Camaquã	0,550	0,509	0,591
MA - Gerais de Balsas	0,233	0,200	0,267	RS - Campanha Central	0,306	0,267	0,342
MA - Gurupi	0,851	0,682	1,000	RS - Campanha Meridional	0,380	0,333	0,425
MA - Imperatriz	0,326	0,285	0,365	RS - Campanha Ocidental	0,361	0,315	0,401
MA - Itapeturu Mirim	0,175	0,134	0,216	RS - Carazinho	0,488	0,454	0,525
MA - Lençóis Maranhenses	0,171	0,122	0,223	RS - Caxias do Sul	1,000	0,935	1,000
MA - Litoral Ocidental Maranhense	0,117	0,094	0,140	RS - Cerro Largo	0,652	0,605	0,701
MA - Médio Mearim	0,503	0,396	0,616	RS - Cruz Alta	0,671	0,613	0,730
MA - Pindaré	0,703	0,573	0,834	RS - Erechim	0,922	0,834	1,000
MA - Porto Franco	0,402	0,348	0,462	RS - Frederico Westphalen	0,722	0,646	0,803
MA - Presidente Dutra	0,217	0,182	0,257	RS - Gramado-Canela	0,551	0,515	0,588
MA - Rosário	0,123	0,103	0,142	RS - Guaporé	1,000	1,000	1,000
MG - Aimorés	0,393	0,350	0,435	RS - Ijuí	0,695	0,635	0,757
MG - Alfenas	0,667	0,619	0,715	RS - Jaguarão	0,351	0,287	0,412
MG - Almenara	0,324	0,282	0,365	RS - Lajeado-Estrela	1,000	1,000	1,000
MG - Andrelândia	0,447	0,397	0,495	RS - Litoral Lagunar	0,189	0,172	0,207
MG - Araçuaí	0,159	0,136	0,182	RS - Montenegro	0,955	0,871	1,000
MG - Araxá	0,719	0,623	0,806	RS - Não-Me-Toque	0,752	0,664	0,843
MG - Barbacena	0,528	0,493	0,562	RS - Osório	0,715	0,666	0,766
MG - Belo Horizonte	0,799	0,721	0,881	RS - Passo Fundo	1,000	0,957	1,000
MG - Bocaiúva	0,268	0,240	0,296	RS - Pelotas	0,602	0,541	0,666
MG - Bom Despacho	0,747	0,630	0,862	RS - Porto Alegre	0,410	0,384	0,436
MG - Campo Belo	0,380	0,349	0,409	RS - Restinga Seca	0,607	0,562	0,653
MG - Capelinha	0,239	0,199	0,279	RS - Sananduva	0,621	0,575	0,667
MG - Caratinga	0,360	0,325	0,394	RS - Santa Cruz do Sul	0,878	0,773	0,991
MG - Cataguases	0,236	0,214	0,256	RS - Santa Maria	0,518	0,469	0,567
MG - Conceição do Mato Dentro	0,196	0,173	0,220	RS - Santa Rosa	0,632	0,576	0,692
MG - Conselheiro Lafaiete	0,315	0,291	0,337	RS - Santiago	0,504	0,448	0,556
MG - Curvelo	0,342	0,304	0,377	RS - Santo Ângelo	0,524	0,478	0,572
MG - Diamantina	0,076	0,067	0,086	RS - São Jerônimo	0,496	0,463	0,529
MG - Divinópolis	0,633	0,567	0,695	RS - Serras de Sudeste	0,421	0,374	0,468
MG - Formiga	0,400	0,364	0,433	RS - Soledade	0,396	0,369	0,423
MG - Frutal	0,692	0,598	0,777	RS - Três Passos	0,803	0,730	0,884
MG - Governador Valadares	0,285	0,255	0,314	RS - Vacaria	0,978	0,881	1,000
MG - Grão Mogol	0,090	0,081	0,100	SC - Araranguá	1,000	1,000	1,000
MG - Guanhães	0,201	0,182	0,220	SC - Blumenau	0,518	0,481	0,555
MG - Ipatinga	0,124	0,112	0,137	SC - Campos de Lages	0,613	0,563	0,664
MG - Itabira	0,298	0,272	0,324	SC - Canoinhas	0,785	0,720	0,853
MG - Itaguara	0,326	0,303	0,349	SC - Chapecó	1,000	1,000	1,000
MG - Itajubá	0,427	0,393	0,458	SC - Concórdia	1,000	1,000	1,000
MG - Ituiutaba	0,363	0,310	0,414	SC - Criciúma	0,812	0,759	0,866
MG - Janaúba	0,165	0,147	0,184	SC - Curitibaanos	0,397	0,366	0,425
MG - Januária	0,236	0,202	0,269	SC - Florianópolis	0,706	0,640	0,775
MG - Juiz de Fora	0,395	0,351	0,434	SC - Itajaí	0,529	0,470	0,591
MG - Lavras	0,510	0,455	0,562	SC - Ituporanga	0,663	0,616	0,710
MG - Manhuaçu	0,588	0,499	0,675	SC - Joaçaba	1,000	1,000	1,000
MG - Mantena	0,215	0,191	0,239	SC - Joinville	0,794	0,739	0,849
MG - Montes Claros	0,236	0,206	0,266	SC - Rio do Sul	0,940	0,867	1,000
MG - Muriaé	0,382	0,346	0,419	SC - São Bento do Sul	0,287	0,250	0,327
MG - Nanuque	0,301	0,251	0,354	SC - São Miguel do Oeste	1,000	1,000	1,000
MG - Oliveira	0,508	0,465	0,548	SC - Tabuleiro	0,510	0,474	0,547
MG - Ouro Preto	0,084	0,074	0,096	SC - Tijucas	0,623	0,572	0,676

MG - Paracatu	0,498	0,431	0,560	SC - Tubarão	0,903	0,840	0,968
MG - Pará de Minas	0,478	0,396	0,567	SC - Xanxerê	0,938	0,873	1,000
MG - Passos	0,905	0,838	0,968	SE - Agreste de Itabaiana	0,338	0,282	0,391
MG - Patos de Minas	0,801	0,711	0,890	SE - Agreste de Lagarto	0,244	0,211	0,275
MG - Patrocínio	0,822	0,736	0,901	SE - Aracaju	0,076	0,063	0,089
MG - Peçanha	0,266	0,238	0,296	SE - Baixo Cotinguiba	0,062	0,052	0,072
MG - Pedra Azul	0,202	0,173	0,238	SE - Boquim	0,331	0,284	0,378
MG - Pirapora	0,249	0,216	0,281	SE - Carira	0,229	0,208	0,249
MG - Piuí	0,531	0,466	0,590	SE - Cotinguiba	0,183	0,157	0,212
MG - Poços de Caldas	0,874	0,810	0,939	SE - Estância	0,524	0,453	0,599
MG - Ponte Nova	0,238	0,217	0,258	SE - Japarutuba	0,897	0,764	1,000
MG - Pouso Alegre	0,531	0,490	0,573	SE - Nossa Senhora das Dores	0,159	0,140	0,180
MG - Salinas	0,164	0,143	0,185	SE - Propriá	0,411	0,358	0,466
MG - Santa Rita do Sapucaí	0,440	0,408	0,470	SE - Sergipana do Sertão do São Francisco	0,169	0,150	0,188
MG - São João Del Rei	1,000	1,000	1,000	SE - Tobias Barreto	0,169	0,153	0,185
MG - São Lourenço	0,935	0,851	1,000	SP - Adamantina	0,598	0,536	0,658
MG - São Sebastião do Paraíso	0,614	0,564	0,667	SP - Amparo	0,452	0,418	0,489
MG - Sete Lagoas	0,379	0,353	0,405	SP - Andradina	0,559	0,504	0,609
MG - Teófilo Otoni	0,298	0,259	0,337	SP - Araçatuba	0,405	0,338	0,477
MG - Três Marias	0,522	0,433	0,604	SP - Araraquara	0,654	0,592	0,718
MG - Ubá	0,607	0,563	0,649	SP - Assis	0,668	0,580	0,753
MG - Uberaba	1,000	1,000	1,000	SP - Auriflama	0,526	0,446	0,607
MG - Uberlândia	0,870	0,785	0,947	SP - Avaré	0,314	0,272	0,357
MG - Unai	0,389	0,339	0,434	SP - Bananal	0,140	0,113	0,168
MG - Varginha	0,606	0,556	0,659	SP - Barretos	0,430	0,343	0,527
MG - Viçosa	0,323	0,282	0,364	SP - Batatais	0,758	0,669	0,855
MS - Alto Taquari	0,466	0,386	0,542	SP - Bauru	0,591	0,527	0,654
MS - Aquidauana	0,211	0,179	0,242	SP - Birigui	0,804	0,704	0,898
MS - Baixo Pantanal	0,142	0,125	0,162	SP - Botucatu	0,507	0,434	0,583
MS - Bodoquena	0,279	0,244	0,313	SP - Bragança Paulista	0,414	0,384	0,447
MS - Campo Grande	0,571	0,515	0,623	SP - Campinas	1,000	0,918	1,000
MS - Cassilândia	0,286	0,217	0,356	SP - Campos do Jordão	0,336	0,276	0,402
MS - Dourados	0,646	0,584	0,706	SP - Capão Bonito	0,481	0,443	0,516
MS - Iguatemi	0,503	0,453	0,552	SP - Caraguatatuba	0,158	0,121	0,199
MS - Nova Andradina	0,234	0,207	0,260	SP - Catanduva	1,000	0,882	1,000
MS - Paranaíba	0,349	0,277	0,420	SP - Dracena	0,372	0,331	0,414
MS - Três Lagoas	0,186	0,156	0,216	SP - Fernandópolis	0,498	0,429	0,569
MT - Alta Floresta	0,354	0,300	0,408	SP - Franca	0,498	0,444	0,553
MT - Alto Araguaia	0,153	0,116	0,191	SP - Franco da Rocha	0,094	0,061	0,132
MT - Alto Guaporé	0,372	0,310	0,429	SP - Guaratinguetá	0,306	0,255	0,363
MT - Alto Pantanal	0,232	0,203	0,261	SP - Guarulhos	0,415	0,314	0,524
MT - Alto Paraguai	0,180	0,153	0,207	SP - Itanhaém	0,571	0,463	0,689
MT - Alto Teles Pires	0,943	0,786	1,000	SP - Itapecerica da Serra	0,265	0,193	0,342
MT - Arinos	0,239	0,199	0,277	SP - Itapetininga	0,486	0,435	0,540
MT - Aripuanã	0,309	0,257	0,363	SP - Itapeva	0,475	0,434	0,513
MT - Canarana	1,000	1,000	1,000	SP - Ituverava	0,316	0,274	0,362
MT - Colíder	0,394	0,333	0,454	SP - Jaboticabal	0,917	0,821	1,000
MT - Cuiabá	0,321	0,280	0,361	SP - Jales	0,917	0,827	1,000
MT - Jauru	0,725	0,624	0,819	SP - Jaú	0,700	0,614	0,789
MT - Médio Araguaia	0,224	0,170	0,281	SP - Jundiá	0,387	0,321	0,458
MT - Norte Araguaia	0,363	0,292	0,434	SP - Limeira	0,694	0,619	0,778
MT - Paranatinga	0,253	0,199	0,304	SP - Lins	0,469	0,402	0,539
MT - Parecis	0,185	0,156	0,213	SP - Marília	0,239	0,213	0,265
MT - Primavera do Leste	0,733	0,598	0,882	SP - Mogi das Cruzes	0,989	0,859	1,000
MT - Rondonópolis	0,398	0,348	0,443	SP - Mogi Mirim	0,645	0,570	0,728
MT - Rosário Oeste	0,273	0,240	0,306	SP - Nhandeara	0,762	0,682	0,836
MT - Sinop	0,414	0,355	0,467	SP - Novo Horizonte	0,593	0,514	0,678
MT - Tangará da Serra	0,486	0,421	0,551	SP - Osasco	0,090	0,055	0,129
MT - Tesouro	0,293	0,234	0,349	SP - Ourinhos	0,553	0,504	0,600
PA - Almeirim	0,172	0,147	0,199	SP - Paraibuna/Paraitinga	0,217	0,194	0,238
PA - Altamira	0,466	0,388	0,547	SP - Piedade	0,615	0,564	0,665
PA - Arari	0,343	0,279	0,417	SP - Piracicaba	0,470	0,408	0,536
PA - Belém	0,468	0,401	0,545	SP - Pirassununga	0,579	0,477	0,695
PA - Bragantina	0,363	0,296	0,430	SP - Presidente Prudente	0,410	0,370	0,450
PA - Cametá	0,743	0,569	0,921	SP - Registro	1,000	1,000	1,000
PA - Castanhal	0,604	0,528	0,682	SP - Ribeirão Preto	0,621	0,558	0,691
PA - Conceição do Araguaia	0,447	0,366	0,531	SP - Rio Claro	0,351	0,301	0,406
PA - Furos de Breves	1,000	1,000	1,000	SP - Santos	0,092	0,055	0,134
PA - Guamá	1,000	1,000	1,000	SP - São Carlos	0,597	0,480	0,733

PA - Itaituba	0,347	0,305	0,391	SP - São João da Boa Vista	0,562	0,522	0,601
PA - Marabá	0,507	0,431	0,590	SP - São Joaquim da Barra	0,503	0,409	0,608
PA - Óbidos	0,547	0,456	0,648	SP - São José do Rio Preto	0,807	0,724	0,883
PA - Paragominas	0,336	0,278	0,401	SP - São José dos Campos	0,175	0,159	0,192
PA - Parauapebas	0,426	0,357	0,501	SP - São Paulo	0,234	0,173	0,302
PA - Portel	1,000	1,000	1,000	SP - Sorocaba	0,487	0,438	0,540
PA - Redenção	0,308	0,263	0,357	SP - Tatuí	0,590	0,521	0,663
PA - Salgado	0,286	0,242	0,329	SP - Tupã	1,000	1,000	1,000
PA - Santarém	0,461	0,385	0,538	SP - Votuporanga	0,424	0,368	0,480
PA - São Félix do Xingu	0,432	0,359	0,510	TO - Araguaína	0,313	0,269	0,358
PA - Tomé-Açu	1,000	1,000	1,000	TO - Bico do Papagaio	0,327	0,278	0,380
PA - Tucuruí	0,389	0,329	0,454	TO - Dianópolis	0,252	0,211	0,296
PB - Brejo Paraibano	0,637	0,525	0,745	TO - Gurupi	0,220	0,180	0,255
PB - Cajazeiras	0,153	0,137	0,169	TO - Jalapão	0,225	0,183	0,268
PB - Campina Grande	0,188	0,161	0,214	TO - Miracema do Tocantins	0,347	0,286	0,407
PB - Cariri Ocidental	0,104	0,095	0,114	TO - Porto Nacional	0,163	0,141	0,187
PB - Cariri Oriental	0,150	0,138	0,162	TO - Rio Formoso	0,399	0,332	0,461

Apêndice D: Códigos R e JAGS

Programa para gerar os dados

```
# Pacote necessário para gerar dados da normal truncada
require(truncnorm)

set.seed(301)

# "grupos" é um vetor de zeros, do qual, 50 serão marcados como 1
# Isso é para facilitar a diferenciação nos cenários C2 e C3
grupos=numeric(100)
grupos[sample(1:100,50)]=1

#elasticidadeas
beta=c(2,0.45,0.3,0.4)
beta1=c(2,0.54,0.32,0.42)

#gerando os inputs
inputs=cbind(1,matrix(runif(300,1,3),ncol=3))

# 'output_clean' é produção que será usada nos cenários C1 e C3,
# enquanto 'output_clean1' será usada em C2.
# Ainda será adicionado o erro composto
output_clean=inputs%*%beta
output_clean1=output_clean
output_clean1[grupos==1] = inputs[grupos==1, ]%*%beta1

xbar_u=0.3 # média de ineficiência, para C1 e C2
s2 = 0.1 # variância do erro composto

inefs=matrix(NA,ncol=6,nrow=100)

gama = c(0.1,0.5,0.9)
s2v = (1-gama)*s2 # variância do erro aleatório

#-----
# Coluna 1 e 2 de tnpair, na linha i, são a média e desvio-padrão,
# respectivamente, da normal que após ser truncada, gera
# ineficiências com média xbar e variância gama[i]*s2.
```

```

#
# Coluna 3 e 4 de tnpair, na linha i, são a média e desvio-padrão,
# respectivamente, da normal que após ser truncada, gera
# ineficiências com média  $\bar{x}/6$  e variância  $\text{gama}[i]*s^2$ .
#
# Lembrando que s2 é a variância total do erro composto.
#-----
tnpair=matrix(c(0.29959103,0.1002072,-3.175941,0.4142047,
               0.02975919,0.3622150,-3.736891,0.4896798,
               -14.61796372,2.1365790,-4.198351,0.5499569),
              ncol=4,nrow=3,byrow=TRUE)

for(i in 1:3){
  # gerandos as ineficiências com média 'xbar_u' e variância s2u=gama[i]*s2
  # estas ineficiências serão usadas em C1 e C2
  u=rtruncnorm(100,a=0,b=Inf,mean=tnpair[i,1],sd=tnpair[i,2])

  # gerandos as ineficiências que serão usadas em C3
  # estas ineficiências tem média 'xbar_u/6' e variância s2u=gama[i]*s2
  u1=u
  u1[grupos==1]=rtruncnorm(50,a=0,b=Inf,mean=tnpair[i,3],sd=tnpair[i,4])

  #-----
  # Observe que o grupo que teve a tecnologia de produção
  # alterada em C2 é o mesmo que recebe ineficiência com média
  # menor em C3
  #-----

  if(i==1){
    # ineficiências geradas no caso de gamma=0,1
    efs[,c(1,4)]=cbind(u,u1)
  }else{
    if(i==2){
      # ineficiências geradas no caso de gamma=0,5
      efs[,c(2,5)]=cbind(u,u1)
    }else{
      # ineficiências geradas no caso de gamma=0,9
      efs[,c(3,6)]=cbind(u,u1)
    }
  }
}
}

```

```

# gerando o erro aleatório, que é o mesmo para C1, C2 e C3
v=rnorm(300*100,0,sqrt(s2v[i]))
dados=data.frame(grupos,inputs[,-1])

# C1
dados[,5:304]=output_clean-u
dados[,5:304]=dados[,5:304]+v
write.table(dados,paste("dados/C1","_",gsub("\\.", "", gama), ".txt", sep=""),
            col.names=TRUE,row.names=FALSE,sep=",")

# C2
dados[,5:304]=output_clean1-u
dados[,5:304]=dados[,5:304]+v
write.table(dados,paste("dados/C2","_",gsub("\\.", "", gama), ".txt", sep=""),
            col.names=TRUE,row.names=FALSE,sep=",")

# C3
dados[,5:304]=output_clean-u1
dados[,5:304]=dados[,5:304]+v
write.table(dados,paste("dados/C3","_",gsub("\\.", "", gama), ".txt", sep=""),
            col.names=TRUE,row.names=FALSE,sep=",")
}

write.table(exp(-inefs), "dados/efs.txt", col.names=FALSE, row.names=FALSE, sep=",")

```

Este programa ajusta os modelos regressão quantílica frequentista, SFA e DEA, obtém as eficiências ajustadas e salva as mesmas em arquivos para análises posteriores.

```

# pacotes necessários para regressão quantílica, SFA e DEA,
# respectivamente
require(quantreg)
require(frontier)
require(rDEA)

files = c("C1_01", "C1_05", "C1_09",
          "C2_01", "C2_05", "C2_09",
          "C3_01", "C3_05", "C3_09")

#-----

```

```

# Em cada conjunto de dados (files), o for mais externo
# ler os dados, ajusta o modelo para cada uma das 300
# repetições, colunas 5 a 304, obtém a eficiência e salva
# um arquivos com as eficiências ajustas para aquele
# conjunto de dados, naquele modelo
#-----

for(file in files){
  dados=read.csv(paste("dados/",file,".txt",sep=""))
  efs_rq=efs_sfa=efs_dea = matrix(NA,ncol=300,nrow=100)

  for(i in 5:304){
    m=rq(dados[,i]~dados[,2]+dados[,3]+dados[,4],tau=0.8)
    efs_rq[,i-4]=ifelse(exp(residuals(m))>1,1,exp(residuals(m)))

    m=sfa(dados[,i]~dados[,2]+dados[,3]+dados[,4],truncNorm=TRUE)
    efs_sfa[,i-4]=efficiencies(m,asInData=TRUE)

    m=dea.robust(X=dados[,2:4],Y=dados[,i],model="output",RTS="variable")
    efs_dea[,i-4]=m$theta_hat_hat
  }
  write.table(efs_rq,paste("efs/rq_",file,".txt",sep=""),
             col.names=FALSE,row.names=FALSE,sep=",")

  write.table(efs_sfa,paste("efs/sfa_",file,".txt",sep=""),
             col.names=FALSE,row.names=FALSE,sep=",")

  write.table(efs_dea,paste("efs/dea_",file,".txt",sep=""),
             col.names=FALSE,row.names=FALSE,sep=",")
}

```

Este programa ajusta a regressão quantílica bayesiana, com parâmetro de escala, obtém as eficiências ajustadas e salva as mesmas em arquivos para análises posteriores.

```

# pacote necessário para usar o JAGS
require(rjags)

#-----

# O modelo jags é armazenado numa string, que é usada na função

```

```

# posteriormente pela função jags.model(). Assim como no R, na
# linguagem do JAGS comentários podem ser escritos começando a
# linha com '#'. Eles podem ser feitos na string que será passada
# a jags.model().
#-----

model.jags="model{
  mu = x%%beta
  for (i in 1:n){
    # resíduo do modelo e cálculo da eficiência
    res[i] = y[i]-mu[i]
    te[i]=ifelse(exp(res[i])>1,1,exp(res[i]))

    # valor da função check
    check[i]=(abs(res[i])+(2*tau-1)*res[i]) * prec * 0.5

    # valor da densidade
    dens[i]=tau*(1-tau)*exp(-check[i]) * prec

    # ones-trick
    ones[i]~dbern(0.0001*dens[i])
  }

  # prioris para beta. normal é modelada em termos da
  # precisão, que é 1/sigma^2
  for (i in 1:(p+1)){
    beta[i] ~ dnorm(0,0.01)
  }
  prec~dgamma(0.001,0.001)
}"

inits1=list(beta=c(1,2,3,4),prec=100)
inits2=list(beta=c(100,200,300,400),prec=10)
iter=200000

files = c("C1_01","C1_05","C1_09",
          "C2_01","C2_05","C2_09",
          "C3_01","C3_05","C3_09")

for(file in files){

```

```

dados=read.csv(paste("dados/",file,".txt",sep=""))
efs=matrix(NA,ncol=300,nrow=100)
for(i in 5:304){
  m=jags.model(file=textConnection(model.jags),
               data=list(x=cbind(1,dados[,c(2,3,4)]),
                          y=dados[,i],
                          ones=rep(1,100),
                          n=100,p=3,tau=0.8),
               n.chains=2,inits=list(inits1,inits2),
               n.adapt=1000,quiet=TRUE)

  update(m,n.iter=iter/2)
  samples=jags.samples(m,variable.names="te",n.iter=iter/2,thin=5)
  efs[,i-4]=rowMeans(samples$te)
}
write.table(efs,
            paste("efs/brq1_",file,".txt",sep=""),
            col.names=FALSE,row.names=FALSE,sep=",")
}

```

Para ajustar o modelo sem parâmetro de escala, o programa acima foi utilizado apenas alterando `model.jags` para a forma abaixo:

```

model.jags="model {
  mu = x%*%beta
  for (i in 1:n){
    res[i] = y[i]-mu[i]
    te[i]=ifelse(exp(res[i])>1,1,exp(res[i]))
    check[i]=(abs(res[i])+(2*tau-1)*res[i]) / 2
    dens[i]=tau*(1-tau)*exp(-check[i])
    ones[i]~dbern(0.0001*dens[i])
  }
  for (i in 1:(p+1)){
    beta[i] ~ dnorm(0,0.01)
  }
}"

```

O programa abaixo ajusta o modelo de regressão quantílica aos dados da agricultura familiar e salva os dados para análise posterior.

```

dados=read.csv("AF/dados.csv")
model.jags="model{

```

```

mu = x%%beta
for (i in 1:n){
  res[i] = y[i]-mu[i]
  te[i]=ifelse(exp(res[i])>1,1,exp(res[i]))
  check[i]=(abs(res[i])+(2*tau-1)*res[i]) * prec * 0.5
  dens[i]=tau*(1-tau)*exp(-check[i]) * prec
  ones[i]~dbern(0.0001*dens[i])
}

for (i in 1:(p+1)){
  beta[i] ~ dnorm(0,0.01)
}
prec~dgamma(0.001,0.001)
}"

inits1=list(beta=c(1,0.3,3,0.7),prec=0.1)
inits2=list(beta=c(147,100,38,70),prec=10)

m=jags.model(file=textConnection(model.jags),
  data=list(x=cbind(1,log(dados[,c(3,4,5)])),
    # dados[,1] são os nomes das microrregiões
    y=log(dados[,2]),
    ones=rep(1,nrow(dados)),
    n=nrow(dados),
    p=3,
    tau=0.95),
  n.chains=2,inits=list(inits1,inits2),
  n.adapt=1000,quiet=TRUE)
update(m,n.iter=5*10^4)
samples=jags.samples(m,variable.names=c("te","beta","prec"),
  n.iter=1.7*10^6,thin=17)
# salvar para analisar posteriormente
saveRDS(samples,"AF/samples.rds")

```