JOSIMAR MENDES DE VASCONCELOS

EQUAÇÕES SIMULTÂNEAS NO CONTEXTO CLÁSSICO E BAYESIANO: UMA ABORDAGEM À PRODUÇÃO DE SOJA

RECIFE-PE

AGOSTO/2011

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

EQUAÇÕES SIMULTÂNEAS NO CONTEXTO CLÁSSICO E BAYESIANO: UMA ABORDAGEM À PRODUÇÃO DE SOJA

Dissertação apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Mestre.

Área de Concentração: Modelagem Estatística e Computacional

Orientador: Prof. Dr. Eufrázio de Souza Santos

RECIFE-PE

AGOSTO/2011

Ficha Catalográfica

V331e Vasconcelos, Josimar Mendes de
Equações simultâneas no contexto clássico e bayesiano: uma abordagem à produção de soja / Josimar Mendes de Vasconcelos. -- 2011.
91 f.: il.

Orientador (a): Eufrázio Santos.
Dissertação (Mestrado em Biometria e Estatística Aplicada) – Universidade Federal Rural de Pernambuco, Departamento de Informática, Recife, 2011.
Inclui apêndice e referências.

1. Soja – Produção 2. Modelos de equações simultâneas 3. Monte Carlo via cadeia de Markov 4. Algoritmo de Gibbs 5. Algoritmo de Metropolis-Hastings 6. Inferência clássica e Bayesiana I. Santos, Eufrázio, Orientador II. Título

CDD 574.018

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

EQUAÇÕES SIMULTÂNEAS NO CONTEXTO CLÁSSICO E BAYESIANO: UMA ABORDAGEM À PRODUÇÃO DE SOJA

JOSIMAR MENDES DE VASCONCELOS

Dissertação julgada adequada para obtenção do título de mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 08/08/2011 pela Comissão Examinadora.

Orientador:

Prof. Dr. Orientador Eufrázio de Souza Santos
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Prof. Dr. Moacyr Cunha Filho
Universidade Federal Rural de Pernambuco

Prof. Dr. Cláudio Tadeu Cristino
Universidade Federal Rural de Pernambuco

Prof. Dra. Maria Cristina Falcão Raposo
Universidade Federal de Pernambuco

Dedico este trabalho a minha mãe, Marli, meu pai, Kerginaldo, minhas irmãs, Mikarla e Mirlândia e a minha amada esposa, Marília.

Agradecimentos

Aos meus pais, Marli e Kerginaldo, pelo grande apoio que me deram em todas as decisões da minha vida. Eles sabem muito bem que as palavras não são suficientes para descrever o amor e agradecimento que eu tenho em meu coração. Lembro até hoje que minha infância, adolescência e juventude não foi simples como de muitas pessoas, mais sempre acreditaram em mim, a todo momento estiveram em oração profunda e nunca esquecerei disso. Muito obrigado por este amor Ágape!

Aos meus irmãos, agradeço por todo o carinho, atenção e confiança que me deram até o momento. Não poderia deixar de destacar Mirlândia que ficou quatro dias e quatro noites passando o artigo[1] para o programa "Microsoft Word 2003" e sua profunda oração para o meu sucesso, e a Mikarla pela sua silenciosa oração para a minha nova jornada de vida.

Agradeço a minha amiga/companheira/namorada/noiva/esposa, Marília, pela compreensão, carinho, dedicação, firmeza e apoio psicológico que teve o tempo todo comigo "desde o ensino médio", e sem dúvida, hoje tenho certeza que ela é a mulher da minha vida. Além disso, me ajudou na correção ortográfica e gramatical (por pura pressão e chantagem emocional).

Agradeço ao professor Eufrázio Santos, pela orientação, paciência, incentivo e cobranças para a realização dessa brilhante dissertação. Também, pela sua dedicação de muito tempo ao programa da pós−graduação da biometria e estatística aplicada.

Agradeço aos professores do programa da pós−graduação da biometria e estatística aplica, ao secretário Marcos e com exclusividade a Zuleide. Zuleide, que sempre exigiu da gente um bom dia, boa tarde, tudo bem? E ai aprendemos com ela o quanto a saudação é importante para termos um bom dia. Muito obrigado!

Aos professores da UFPE, meus agradecimentos, pela maturidade na minha vida acadêmica que os senhores me proporcionaram durante a minha passagem no programa da pós−graduação em estatística. Deles destaco os professores Cristiano Ferraz, Cribari-Neto e Sylvio Santos pelas boas aulas, orientação e incentivo a pós−graduação.

Agradeço a secretária da pós−graduação em estatística da UFPE, Valéria Bittencourt, pela grande competência, carinho e atenção com os alunos do programa.

[1] O artigo transposto se refere ao filtro desta dissertação, veja Filho *et all.* (2011).

Aos professores da Universidade Federal do Rio Grande do Norte, André, Carla e Formiga, pelo incentivo, apoio e confiança na minha vinda a Recife−PE para cursar o mestrado e consequentemente a aprovação do concurso. Tenho um imenso carinho pelos senhores!

Dificilmente passaria por esse tópico sem citar os meus sinceros agradecimentos pela minha primeira equipe de estudo no mestrado da Universidade Federal de Pernambuco; onde aprendi a estudar em grupo, contar as minhas piadas sem graças “todos riam” e sempre lembrar que o conhecimento é para ser divido em grupo. Sendo-os: Helton (Josi, olha aquela guri), Jeremias (que putaria é essa macho), Lutemberg (zarai), Manoel (tem que ter foco), Marcelo (vou me lascar) e Priscila (sempre sorrindo). Também, ressalvo os companheiros Fábio Bayer, Ivan, Diego, Tarciana, Tatiene e Francisco que de alguma forma contribuíram para o epílogo desse mestrado.

Depois de citar os companheiros da UFPE não poderia deixar de agradecer aos colegas do programa da pós−graduação da biometria e estatística aplicada pelo incentivo, grupos de estudos e motivação pelo curso. Em especial quero citar os jovens Carlão e Samuel (velho Samuca) pela boa guarita que me proporcionou nessa passagem da biometria, me suportando constantemente no 0800, noites e noites acordados estudando para as disciplinas e descobrindo o porque que somos bons! Em seguida me carregaram para uma nova guarita do grande Rodrigo *versus* Filipe (meus cordiais agradecimentos) que também me acolheram de braços abertos para eu finalizar a grande obra da vida acadêmica. Da mesma forma, agradeço aos demais colegas da pós−graduação (turmas anteriores/posteriores do mestrado e doutorado na biometria e estatística aplicada) que contribuíram de alguma forma para a conclusão deste mestrado: Anderson, Augusto, Cíntia, Dennis, Diego, Djalma, Milton, Rivelino, Silvio (O grande matemático e estatístico de “ôlinda”), Rodrigo, Rogério, Fábio, Lázaro, Darlon, David, Rosilda, Jáder, Fabia, Ana Célia e a Mércia Cardoso. Meus sinceros agradecimentos!

Agradeço a minha afilhada, Rita, pelo apoio, carinho fraterno e grande incetivo nos momentos bons e difíceis do mestrado/graduação. Sem contar do grande carisma que sempre teve comigo, independente dos meus erros e acertos.

Agradeço a companheira de estudo Danielle, pela força, carinho e paciência que teve no período do mestrado acadêmico.

Agradeço também a família da minha esposa que de alguma forma contribuíram para minha formação com suas orações e palavras de conforto.

Agradeço à Dona Maria José e Dona Severina pela acolhida na pensão, algumas dormidas no 0800 e por estarem sempre dispostas a ajudar no que fosse possível.

Aos companheiros de pensão João Batista, Jadson, Djekson, Josivandro, Júnior, Agosso e Paulo, pelos diversos momentos de brincadeiras, discussões religiosas e filosóficas, meus cordiais agradecimentos.

Agradeço ao companheiro Raphael, que foi a primeira pessoa a me receber em Recife−PE, no qual se disponibilizou-se em procurar uma moradia e dando uma boa assistência no período do mestrado.

Agradeço a todos e todas da Unidade Acadêmica de Garanhuns−UFRPE pelo apoio que me proporcionaram na minha vida profissional. Pessoas que me receberam excelentemente bem, do(a) zelador(a) ao diretor do centro, no qual se eu for citar nomes dará infinitas páginas.

Não posso esquecer de agradecer aos meus amigos (Natal/RN−Recife/PE−Garanhuns/PE) que, de uma forma ou de outra, contribuíram com sua amizade, sugestões e conversas "paralelas" para a realização desta obra.

Agradeço antecipadamente pelos comentários, sugestões e críticas dos participantes da banca examinadora, professores Maria Cristina Falcão Raposo, Moacyr Cunha Filho e Cláudio Tadeu Cristino.

Agradeço a existência do editor de texto LaTeX e os softwares estatísticos `R` e `WinBUGS`.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico−CNPq pelo apoio financeiro de um ano e meio no programa de mestrado da pós−graduação de estatística da UFPE.

Por fim, peço desculpas pelos nomes não citados nesse tópico, lembrem que o espaço é pouco para uma gama de pessoas que estão no meu coração, paciência!

Às vezes, nos vemos distantes do nosso possível, mas, na verdade, estamos bem mais perto do que esperávamos... O tempo nos mostra isso, ou melhor, o Deus que nos rodeia mostra-nos claramente. Assim, teremos sempre que caminhar com a paciência ao nosso lado e deixarmos que o rio nos leve com sua suavidade, gentileza, humildade e, mais do que tudo, sua sapiência divina no nosso meio interno. Com isso, sem dúvida, seremos fortes para vencer nossos obstáculos, como o rio que nunca perde para os seus opositores. **Josimar Vasconcelos.**

"*Eis que vos envio vocês como ovelhas no meio de lobos. Portanto, sejam prudentes como as serpentes e simples como as pombas.*" **Mateus 10:16.**

Resumo

Nos últimos anos tem aumentado a quantidade de pesquisadores e pesquisas científicas na plantação, produção e valor de soja no Brasil, em grão. Diante disso, a presente dissertação busca analisar os dados e ajustar modelos que expliquem, de forma satisfatória, a variabilidade observada da quantidade produzida e valor da produção de soja em grão no Brasil, no campo do estudo. Para o desenvolvimento dessas análises é utilizada a inferência clássica e bayesiana, no contexto de equações simultâneas através da ferramenta de mínimos quadrados em dois estágios. Na inferência clássica utiliza-se o estimador de mínimos quadrados em dois estágios. Na inferência bayesiana trabalhou-se o método de Monte Carlo via Cadeia de Markov com os algoritmos de *Gibbs* e *Metropolis-Hastings* por meio da técnica de equações simultâneas. No estudo, consideram-se as variáveis área colhida, quantidade produzida, valor da produção e produto interno bruto, no qual ajustou-se o modelo com a variável resposta quantidade produzida e depois a variável resposta valor da produção para finalmente fazer as correções e obter o resultado final, no método clássico e bayesiano. Através, dos desvios padrão, estatística do teste-t, critérios de informação Akaike e Schwarz normalizados destaca-se a boa aplicação do método de Monte Carlo via Cadeia de Markov pelo algoritmo de *Gibbs*, também é um método eficiente na modelagem e de fácil implementação nos softwares estatísticos `R` & `WinBUGS`, pois já existem bibliotecas prontas para compilar o método. Portanto, sugere-se trabalhar o método de Monte Carlo via cadeia de Markov através do método de *Gibbs* para estimar a produção de soja em grão, no Brasil.

Palavras-chave: Produção de soja em grão; modelos de equações simultâneas; Monte Carlo via Cadeia de Markov; algoritmo de *Gibbs*; algoritmo de *Metropolis-Hastings*.

Abstract

The last years has increased the quantity of researchers and search scientific in the plantation, production and value of the soybeans in the Brazil, in grain. In front of this, the present dissertation looks for to analyze the data and estimate models that explain, of satisfactory form, the variability observed of the quantity produced and value of the production of soya in grain in the Brazil, in the field of the study. For the development of these analyses is used the classical and Bayesian inference, in the context of simultaneous equations by the tools of indirect square minimum in two practices. In the classical inference uses the estimator of square minima in two practices. In the Bayesian inference worked the method of Mountain Carlo via Chain of Markov with the algorithms of Gibbs and Metropolis-Hastings by means of the technician of simultaneous equations. In the study, consider the variable area harvested, quantity produced, value of the production and gross inner product, in which it adjusted the model with the variable answer quantity produced and afterwards the another variable answer value of the production for finally do the corrections and obtain the final result, in the classical and Bayesian method. Through of the detours normalized, statistics of the proof-t, criteria of information Akaike and Schwarz normalized stands out the good application of the method of Mountain Carlo via Chain of Markov by the algorithm of Gibbs, also is an efficient method in the modelado and of easy implementation in the statistical softwares R & WinBUGS, as they already exist smart libraries to compile the method. Therefore, it suggests work the method of Mountain Carlo via chain of Markov through the method of Gibbs to estimate the production of soya in grain.

Key words: Production of soybean in grain; models of simultaneous equations; Mountain Carlo via Chain of Markov; algorithm of Gibbs; algorithm of Metropolis-Hastings.

Resumen

Los últimos años ha aumentado la cantidad de investigadores y pesquisas científicas en la plantación, producción y valor de la soja en el Brasil, en grano. Delante de eso, la presente dissertação busca analizar los datos y estimar modelos que expliquen, de forma satisfactoria, la variabilidade observada de la cantidad producida y valor de la producción de soja en grano en el Brasil, en el campo del estudio. Para el desarrollo de esos análisis es utilizada la inferencia clásica y Bayesiana, en el contexto de ecuaciones simultáneas por las herramientas de mínimos cuadrados en dos prácticas. En la inferencia clásica se utiliza el estimador de mínimos cuadrados en dos prácticas. En la inferencia Bayesiana se trabajó el método de Monte Carlo vía Cadena de Markov con los algoritmos de Gibbs y Metropolis-Hastings por medio de la técnica de ecuaciones simultáneas. En el estudio, se consideran las variables área cosechada, cantidad producida, valor de la producción y producto interior bruto, en el cual se ajustó el modelo con la variable respuesta cantidad producida y después la otra variable respuesta valor de la producción para finalmente hacer las correcciones y obtener el resultado final, en el método clásico y bayesiano. A través, de los desvíos normalizados, estadística de la prueba-t, criterios de información Akaike y Schwarz normalizados se destaca la buena aplicación del método de Monte Carlo vía Cadena de Markov por el algoritmo de Gibbs, también es un método eficiente en el modelado y de fácil implementación en los softwares estadísticos R & WinBUGS, pues ya existen bibliotecas listas para compilar el método. Por lo tanto, se sugiere trabajar el método de Monte Carlo vía cadena de Markov a través del método de Gibbs para estimar la producción de soja en grano.

Palabras clave: Producción de soja en grano; modelos de ecuaciones simultáneas; Monte Carlo vía Cadena de Markov; algoritmo de Gibbs; algoritmo de Metropolis-Hastings.

Lista de Figuras

Lista de Tabelas

Abreviaturas e Siglas Utilizadas

AC	Área colhida
a.C.	antes de Cristo
AIC	Critério de informação de Akaike
AIC$_n$	Critério de informação de Akaike normalizado
AP	Área plantada
BIC	Critério de informação de Bayes
BIC$_n$	Critério de informação de Bayes normalizado
BN	Distribuição binomial negativa
d.C.	depois de Cristo
DP	Desvio padrão
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
ENIAC	Electronic Numerical Integrator And Computer
FAC	Função de autocorrelação
fdp	Função de densidade de probabilidade
F(I)	Fator de dependência
IBGE	Instituo brasileiro de geografia e estatística
IC a 95%	Intervalo de Credibilidade a 95% de confiança
IM	Primeiro modelo
IIM	Segundo modelo
IPEA	Instituto de pesquisas econômica e aplicada
MCMC	Monte Carlo via Cadeia de Markov
MH	*Metropolis-Hastings*
MQI	Mínimos Quadrados Indireto
MQO	Mínimos Quadrados Ordinários
MQ2E	Mínimos Quadrados em 2 Estágios
N(μ, σ^2)	Distribuição normal com média μ e variância σ^2
PIB	Produto interno bruto
QMreg	Quadrado médio da regressão
QMres	Quadrado médio do resíduo
QP	Quantidade produzida
SQE	Soma de quadrado do erro (resíduo)
SQR	Soma de quadrado da regressão
SQT	Soma de quadrado total
VP	Valor da produção

Sumário

1 Introdução

1.1 Preâmbulo

Com o surgimento da informática no final do século XX a estatística se desenvolveu vastamente na programação e em especial na simulação de Monte Carlo via cadeia de Markov através de vários algoritmos. Entre esses algoritmos destacam-se os métodos de *Gibbs sampling* e *Metropolis-Hastings*, em que trabalha com distribuições condicionais complicadas na inferência bayesiana.

Na literatura estatística, se relata sobre a regressão clássica por meio do estimador de mínimos quadrados ordinários, no qual se emprega o modelo de uma variável resposta e uma ou mais variáveis explicativas. Contudo, nem sempre isso acontece, em algumas situações existem mais de uma variável resposta no modelo a ser estimado, e essa situação denomina-se de equações simultâneas. Os conceitos básicos do modelo de equações simultâneas podem ser introduzidos de um modo didático através de um exemplo real ou um caso particular e em seguida generalizar. Portanto, como no mundo acadêmico das ciências exatas e agrárias não foi explorado o método de Monte Carlo via cadeia de Markov no contexto de equações simultâneas, acredita-se que a exploração desse método trará um ganho na pesquisa científica e uma outra alternativa para modelagem de equações simultâneas utilizando o MCMC via *Gibbs* e *Metropolis-Hastings*, com referência no estimador de mínimos quadrados ordinários.

1.2 Método analítico

O estudo foi realizado em três métodos distintos, obtendo duas modelagens em cada condição (veja Fig. 1). Na primeira situação utilizou-se a aplicação de modelos de equações simultâneas pelo método de mínimos quadrados em dois estágios. Sendo que na primeira modelagem a variável dependente foi quantidade produzida e as variáveis independentes foram área colhida e produto interno bruto. Na segunda modelagem estimou-se o primeiro modelo para compor a variável independente e valor da produção entrou como variável dependente.

Dentro do contexto de MQ2E aplicaram-se os métodos clássico e bayesiano. Na metodologia clássica utilizaram-se duas sucessivas modelagens de MQO (vide Baltagi, 2005 & 2008; Magnus & Neudecker, 2007; Gujarati, 2006; Madalla,1992). Na abordagem bayesiana foram aplicados dois tipos de modelagens: o Monte Carlo via Cadeia de Markov através do método de *Gibbs sample* e o MCMC por meio do método de *Metropolis-Hastings*, ambos com aplicação do modelo de regressão linear. Na modelagem bayesiana utilizaram-se as *prioris* Gama e Normal, como os resultados foram semelhantes permaneceu-se com a *priori* Normal de média 0 (zero) e desvio padrão 0,01.

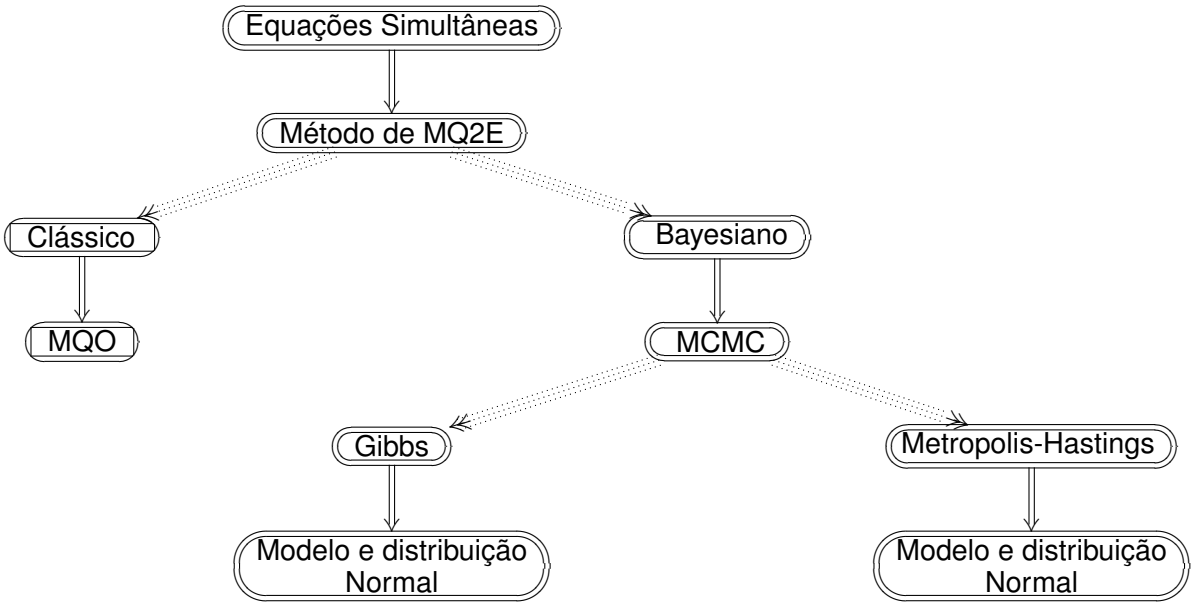

Figura 1: Estrutura do método analítico

1.3 Critérios para dados simulados

Após ter ajustado o modelo de regressão clássico, por meio do método de equações simultâneas usando a técnica de MQ2E, fez-se a aplicação do método de Monte Carlo via Cadeia de Markov através do método de *Gibbs* e *Metropolis-Hastings*. Nas duas modelagens dos métodos, utilizaram-se a *priori* Normal com média 0 (zero) e desvio padrão 0,01 e na precisão aplicou-se a distribuição Gama com o parâmetro de forma e escala $0,10$. O algoritmo de *Gibbs* foi implementado no `WinBUGS` (Bayesian inference Using Gibbs Sampling of windows) dentro da plataforma computacional `R` utilizando o pacote `R2WinBUGS` para gerar as cadeias de Markov dos parâmetros, veja Gelman *et al.* (2005). Para o método de *Metropolis-Hastings* construí-se o algoritmo no software estatístico `R`. Nesse estudo também aplicou-se a *priori* Gama com parâmetro de forma e escala 0,01, contudo o resultado ficou semelhante a *priori* Normal com os parâmetros acima. Portanto, optou-se a permanecer com a *priori* Normal.

Inicialmente realizou-se uma amostra piloto de tamanho 10.000. Depois, pelo diagnóstico de convergência de Geweke (1992), Heidelberger & Welch (1983) e Raftery & Lewis (1992B), sugeriu-se fixar 5.000 iterações e descartar "Burn-in" as 100 (cem) primeiras observações totalizando um tamanho de 4.900 observações para se ter um bom ajuste e retirar o efeito dos primeiros valores. Chama-se atenção do segundo método, o *Metropolis-Hastings*, em que no diagnóstico de Geweke os resultados não ficaram no padrão recomendado, causando assim um processo razoável. Essas aplicações do diagnóstico de convergência e os gráficos realizado ao longo da dissertação foram através do pacote `coda` dentro do `MCMCpack` do software `R` (veja Martin & Quinn, 2010 e Plummer *et al.*, 2010).

1.4 Suporte computacional

Utilizaram-se dois softwares estatísticos: o primeiro `WinBUGS` versão 1.4.3 e o segundo o `R` com pacotes `MCMCpack`, `coda` e `R2WinBUGS`. O software `WinBUGS` foi utilizado para simular os dados do método de *Gibbs sample* para plataforma `Windows`. Para o modelo clássico, o algoritmo de *Metropolis-Hastings*, construções gráficas, análise de regressão (estimação de parâmetros, testes de hipóteses, intervalos de confiança, entre outras análises) foram produzidas no ambiente de programação `R`, utilizando a versão 2.12.2 para a plataforma `Windows`.

Um dos maiores obstáculos na inferência bayesiana tem sido a complexidade da implementação de simulações ou em situações práticas. Assim, tem-se como exemplo a especificação da distribuição a *priori* e a convergência da *posteriori*. Então, para trabalhar esses aspectos goza-se do sistema computacional dotado de boa capacidade para simular ou aplicar a ferramenta de MCMC via *Gibbs* ou MH. Essa ferramenta é o software `WinBUGS`, versão em ambiente `Windows` do pacote `Bugs` "Bayesian Using Gibbs Sampling", implementado por Thomas *et al.* (1992) no qual produz um conjunto de procedimentos que permite a especificação de diferentes tipos de modelos e estima seus parâmetros via MCMC com algoritmos de *Gibbs* ou de *Metropolis-Hastings*, segundo Gamerman (1997).[1]

A plataforma  foi criada por Ross Ihaka & Robert Gentleman (veja Venables *et al.*, 2009) na Universidade de Auckland na Austria com o objetivo de produzir um ambiente de programação semelhante ao `S`, uma linguagem desenvolvida no AT & T Bell Laboratories, cuja versão comercial é o `S-Plus`, tendo as vantagens de ser de livre distribuição e de possuir código fonte aberto. O `R` é um ambiente integrado que possui grandes facilidades para manipulação de dados, geração de gráficos e modelagem estatística em geral. A linguagem e seus pacotes podem ser obtidos gratuitamente no endereço `http://www.r-project.org`.

[1] Para conceitos, materiais ou *download* do software consulte Lunn *et al.* (2000).

Dentro da plataforma `R` utilizou-se o pacote `MCMCpack` (Markov Chain Monte Carlo package) e `coda` que contém funções para utilização da inferência bayesiana (construções de gráficos e diagnósticos de convergência) usando simulação a *posteriori* para um número de modelos estatísticos (veja Martin & Quinn, 2010 e Plummer *et al.*, 2010). Outro pacote utilizado foi o `R2WinBUGS`, que fez a ligação do `R` com o `WinBUGS`, isto é, simulou os dados no `WinBUGS` e imprimiu os resultados no `R`.

Na parte final do suporte computacional cita-se o sistema tipográfico LaTeX, em que foi utilizado como editor de texto da presente dissertação. A tipografia LaTeX foi desenvolvida por Leslie Lamport na década de 1980, no qual consiste em uma série de macros ou rotinas do sistema TeX, criado por Donald Knuth na Universidade de Stanford, que facilita o desenvolvimento da edição do texto. Uma implementação LaTeX para a plataforma Windows (MikTeX) encontra-se disponível em `http://www.miktex.org`. Informações sobre o sistema de tipografia LaTeX podem ser encontrados em Lamport (1994), Mittelbach *et al.* (2004) e também através do sitio `http://www.tex.ac.uk/CTAN/latex`.

1.5 Objetivos do estudo

Nesse novo século XXI a soja tem se mostrado popular no mundo e em particular no Brasil, por meio de seus derivados como a torta, bolo, brigadeiros e outros. Então, os pesquisadores, com ênfase na empresa EMBRAPA, trabalham fortemente nas pesquisas sobre o produto, a produção, o valor da produção, o ganho com a exportação, os efeitos do produto no mercado nacional e internacional. Com isso, os cientistas procuram a melhor técnica para responder de forma precisa e simplória esses aspectos, e em particular a precisão do ajuste do modelo, através das ferramentas de mínimos quadrados em dois estágios. Por isso, essa dissertação pretende alcançar dois objetivos: o primeiro corresponde a detalhes metodológicos, onde pretende descrever o método de Monte Carlo via cadeia de Markov por meio dos algoritmos de *Gibss* e *Metropolis-Hastings*, através da técnica de equações simultâneas, obtendo como destaque os aspectos de inferência, diagnóstico de convergência e gráficos de convergência com respeito a regressão; o segundo é de forma empírica, em outras palavras, trata-se da aplicação da técnica de MCMC via *Gibbs* e MH para modelagem da equação do valor da produção de soja em grão no Brasil e ao mesmo tempo verificar quais desses dois métodos são mais eficientes para modelagem bayesiana. Os resultados obtidos via MCMC serão comparados com os ajustes realizados pela metodologia clássica. No geral, o que se busca nessa tarefa é melhorar a precisão da estimação do valor da produção de soja de forma simples e de fácil aplicação, também explorar a ferramenta de equações simultâneas na inferência bayesiana.

1.6 O que há nesta dissertação?

Este estudo é composto de 7 (sete) capítulos. O primeiro capítulo cita-se a breve parte histórica do método de MCMC, depois relata-se a equação simultânea com conceitos básicos e a importância do estudo dessa ferramenta. Em seguida, a metodologia trabalhada, os critérios para dados simulados, o suporte computacional e os objetivos do estudo.

O Capítulo 2, relata uma breve revisão de literatura enfatizando a técnica de equações simultâneas e o método de Monte Carlo via cadeia de Markov. Na equação simultânea busca-se uma das primeiras e últimas publicações, em livros e artigos, utilizando a ferramenta de mínimos quadrados em dois estágios. No método de MCMC destaca-se um dos primeiros e últimos artigos e livros, salientando os algoritmos de *Gibbs* e *Metropolis-Hastings*.

O Capítulo 3 e 4, descrevem-se a inferência clássica, bayesiana, o modelo de equações simultâneas, especificação do modelo e o teste de simultaneidade ou teste de Hausman. No modelo de equações simultâneas destaca-se a técnica de MQ2E, discorre-se analiticamente esse método.

O Capítulo 5, apresenta-se o método de Monte Carlo via cadeia de Markov por meio dos algoritmos de *Gibbs* e *Metropolis-Hastings*. Nos algoritmos, foi citado o histórico, discorrido o passo-a-passo e exemplos ilustrativos. Depois, descrevem-se os diagnósticos de convergências e os gráficos de convergências.

O Capítulo 6, encontram-se a história e produção da soja no Brasil e exterior. A aplicação em dados reais da quantidade produzida e valor da soja, no Brasil, entre os anos de 1994 e 2009 (dados em anos). Nisso, visa a estimação das equações simultâneas pelo estimador de MQ2E no contexto clássico e bayesiano. Também, verificar qual o método que melhor ajusta a técnica de equações simultâneas.

Para finalizar a dissertação, no Capítulo 7 são apresentadas as considerações finais, comentários e sugestões para futuras pesquisas. Da mesma forma, se o pesquisador quiser um estudo resumido sobre esta dissertação consulte o artigo do Filho *et al.* (2011).

2 Revisão de Literatura

Será mostrado nesse capítulo uma breve revisão de literatura da aplicação de modelos de equações simultâneas por meio da técnica de mínimos quadrados em dois estágios. Em seguida, o método de Monte Carlo via Cadeia de Markov através dos algoritmos de *Gibbs* e *Metropolis-Hastings*.

2.1 Modelos de equações simultâneas

A técnica de equações simultâneas é frequentemente utilizada na área da economia, por exemplo, o estudo de elasticidade da demanda e choques na oferta. Também, nas ciências agrárias, como a produção de soja, mandioca, madeira e desmatamento da Amazônia.

Liu (1983), estimou a elasticidade das variáveis preço e preço cruzado da demanda por gás natural, dos setores residencial, comercial e industrial nos Estados Unidos e suas regiões geográficas (definidas pelo Departamento de Energia - DoE). Esse estudo foi realizado com dados no período de 1967 a 1978, com o uso da variável *dummy* para a crise do petróleo de 1973. Aplicou-se nesse estudo os métodos de mínimos quadrados ordinários e mínimos quadrados em dois estágios, considerando as variáveis independentes a renda e o preço do petróleo.

Através de Epple & Mccallum (2006), foi apresentado a primeira aplicação, com dados reais sobre oferta e demanda da ferramenta de modelos de equações simultâneas, com o sugestivo título “Simultaneous Equation Econometrics: The Missing Example”. Realizou-se nesse trabalho as expressões de oferta e demanda separadamente por MQO, obtendo como objetivo principal a melhor especificação para as equações do Modelo de Equações Simultâneas. Logo após, estimou-se o sistema de equações em MQ2E realizando-se análises gráficas e plotando as curvas de oferta-demanda de curto-longo prazos.

Como consequência, os autores desenvolveram um exemplo, segundo eles “trivial, mas satisfatório”, sobre a oferta e demanda de frango nos Estados Unidos da América. Nessa apli-

cação utilizaram-se dados anuais de 1960 a 1999 e obtiveram resultados, a partir da estimação das equações simultâneas, melhores do que o estimador de mínimos quadrados ordinários.

Também, Soares *et al.* (2004) estudaram o mercado de carvão vegetal no Brasil no período de 1974 a 2000. Depois da análise realizada, concluíram que o método MQ2E é melhor do que o MQO, e portanto, chega-se a considerar que a taxa de câmbio e a taxa de juros interferem no equilíbrio entre oferta e demanda por carvão.

No artigo de Vicente (1994), aplica-se o sistema de equações simultâneas nas equações de oferta e demanda de carnes bovina, suína, de aves e ovos no Brasil para o período 1970-90. O estudo foi realizado por meio dos estimadores mínimos quadrados ordinários, variáveis instrumentais, mínimos quadrados em dois estágios e mínimos quadrados em três estágios; obtendo como melhor resultado o último processo de estimação.

No artigo do Braga & Gomes (2008), estimou-se os principais determinantes da Produtividade Total dos Fatores (PTF) da Amazônia legal, no período de 1990 a 2004. Utilizaram-se o estimador de MQ2E e o método dos momentos generalizados, nos quais os fatores escolaridade, infra-estrutura e o número de novas cooperativas são determinantes para elevar o nível do progresso tecnológico na Amazônia legal.

Por último, tem-se os livros textos de econometria do Gujarati (2006), Madalla (1992) e Baltagi (2008), que descreve a técnica de equações simultâneas na teoria e prática. Em que se teve como base para escrita da dissertação, tanto na parte teórica como na parte prática do contexto de equações simultâneas.

2.2 Monte Carlo via Cadeia de Markov

A ferramenta MCMC é sempre encontrada em livros textos, artigos e trabalhos científicos. Nos livros textos tem-se o volume do professor Paulino *et al.* (2003); Dimove (2008); Kalos & Whitlock (2004); Ntzoufras (2009); Andrade & Kinas (2010). Alguns dos principais artigos e trabalhos científicos serão descritos a seguir.

Um dos primeiros artigos com o desenvolvimento da técnica de *Metropolis-Hastings* foi em 1970[1]. Também, tem-se o artigo detalhado por Chib & Greenberg (1995), que demostraram o algoritmo e fizeram uma aplicação com a normal bivariada. Depois, foram desenvolvidos vários artigos fazendo aplicações da ferramenta MH, e um deles, é o artigo do Brooks & Roberts (1997), que utilizaram-se das técnicas de diagnóstico de convergência MCMC e revisaram vários méto-

[1] Para mais detalhes sobre a técnica consulte as referências, Hastings (1970).

dos de MCMC propostos na literatura. Da mesma forma, compararam os métodos em termos de interpretação e aplicabilidade.

Flury & Shephard (2008) mostraram a aplicação de quatro modelos: o primeiro modelo é um padrão *probit* transversal; o segundo modelo é um padrão de modelo linear de Gauss; o terceiro é um modelo de volatilidade estocástica simples e o quarto é um modelo de equilíbrio geral estocástico. Dos quatro modelos, o último é o mais interessante, visto que mostra o poder único de combinar o filtro de partículas com o algoritmo de *Metroplis-Hastings*.

Com o desenvolvimento computacional aliado ao uso do método de Monte Carlo via Cadeia de Markov, tornaram-se as ferramentas bayesianas mais acessíveis. Um dos tópicos mais ativos em estatística computacional é a inferência bayesiana, por meio de simulação iterativa, aplicando, o amostrador de *Gibbs* e o algoritmo de *Metroplis-Hastings*.[2] Devido a disponibilidade de computadores de alta velocidade, na segunda metade do século XX, as ferramentas tiveram um grande avanço na sociedade facilitando a aplicação dos algoritmos de *Gibbs* e *Metroplis-Hastings*, em dados reais. A essência da simulação iterativa é gerar valores de uma variável aleatória, utilizando uma sequência de distribuições que convergem iterativamente para a distribuição de origem (Leandro, 2001).

Ajustou-se o modelo Gama, da família exponencial com função de ligação potência, a dados de medidas de comprimento e peso de peixes da espécie *Trachydoras paraguayensis*, no artigo do Guedes *et al.* (2007). Assim, para obter as estimativas do modelo utilizou-se o método clássico de máxima verossimilhança e o método bayesiano MCMC por meio dos algoritmos *Gibbs* e MH. Os resultados clássicos e bayesianos foram semelhantes, mas o parâmetro de dispersão foi inferior ao obtido pela a estimativa clássica, além de fornecer um erro padrão muito baixo (30% menor).

Por fim, o artigo dos autores Carlin & Cowles (1996) que descreve 13 tipos de algoritmos de convergência e em seguida verifica se podem falhar em alguma determinada situação.[3] Segundo esse artigo quaisquer desses métodos pode falhar em algum determinado momento. Portanto, recomenda-se uma combinação de estratégias para avaliar e acelerar o MCMC, incluindo a aplicação de procedimentos de diagnóstico para um pequeno número de cadeias paralelas, autocorrelações e correlações cruzadas.

A dissertação de Mayrink (2006), tem-se como objetivo principal avaliar a influência de vizi-

[2]Caso queiram se aprofundar no assunto recomendam-se os artigos Hastings (1970), Gilks *et al.* (1996), Gamerman (1997), Casella & Robert (1999), Chen & Ibrahim (2000).

[3]Os 13 tipos de algoritmos são: Gelman & Rubin (1992); Raftery & Lewis (1992); Garren & Smith (1993); Geweke (1992); Johnson (1994); Liu, Liu & Rubin (1992); Mykland, Tiermey & Yu (1995); Ritter & Tanner (1992); Roberts (1992-1994); Yu (1994); Yu & Mykland (1994); Zellner & Min (1995); Heidelberg & Welch (1983).

nhança através dos resultados adquirido pela simulação de Monte Carlo via Cadeia de Markov por meio dos algoritmos *Gibbs* e *Metropolis-Hastings*.

Constantemente, utiliza-se na área médica, estudos com modelos de respostas dicotômicas ou transformadas em binárias. Rossi & Zellner (1984), analisaram modelos de resposta quanto ao ponto de vista bayesiano, usando distribuição *prioris* informativa. Também, o Ming-Hui & Qi-Man (2000) investigaram a propriedade da distribuição a *posteriori* para modelo de resposta dicotômica, usando uma distribuição *priori* para os parâmetros da regressão.

3 Inferência clássica e bayesiana

3.1 Introdução

A inferência clássica e bayesiana usa aspectos do método científico, isto é, o método científico envolve a coleta de evidências para confrontar com hipóteses formuladas anteriormente. Essa comparação possibilitará averiguar se há consistência ou inconsistência com as hipóteses inicialmente estabelecidas. Nesse contexto, a inferência bayesiana pode ser utilizada para diferenciar hipóteses objetivando estudar uma certa população através de evidências fornecidas por uma amostra. Assim, pode-se perceber que o próprio entendimento sobre o método científico tem uma lógica e estrutura bayesiana, em outras palavras, é essencialmente bayesiano quando relata e destaca hipóteses para verificar experimentalmente e reassentar as crenças iniciais com base na ênfase empírica. Portanto, os resultados experimentais devem ser vistos como algo que mude a sua credibilidade em uma determinada hipótese e não como uma maneira de chegar a uma verdade absoluta propriamente dita. Entretanto, os críticos da abordagem bayesiana dizem que este método de inferência pode ser tendencioso devido à inclusão da crença do pesquisador (probabilidades a *priori*) nos cálculos subsequentes, mesmo antes que qualquer evidência tenha sido adquirida (veja Vasconcelos, 2007).

O bayesianismo tem duas importantes bases no estudo de conhecimento e em vários ramos da sociedade. O primeiro é a versão do universo com base em graus de crença ou credibilidades, em vez do tudo ou nada. O segundo é uma regra matemática que explicita como se devem mudar suas crenças à luz de novos dados empíricos. A partir desses dois pilares pode-se deduzir uma série de implicações filosóficas do bayesianismo.

A primeira abordagem é a crença de cada um na experiência empírica, em que, comparando com a experiência do aprendizado produz algo denominado de “Inteligência Artificial”. Em Pena (2006), há um relato sobre uma teoria de aprendizado em inteligência artificial denominado “Aprendizado Bayesiano”. Por exemplo, os softwares livres e privados de grande porte baseiam-se em princípios bayesianos, programas que bloqueiam mensagens inadequadas nas caixas de correios eletrônicos e também a detecção de câncer de mama.

A segunda abordagem foi elaborada há aproximadamente dois mil anos pelo filósofo grego Aristóteles (384-322, a.C.), que elaborou uma estrutura lógica para verificar a racionalidade humana. Nesse relato chegou-se à seguinte conclusão: a lógica baseia-se em ideias que incluem duas suposições e nada mais. As tais suposições são *falso ou verdade*. No artigo do Alder (2005), conclui-se que Thomas Bayes leva a uma generalização da lógica Aristoteliana, em outras palavras, sair do raciocínio discreto, do tudo ou nada, para o raciocínio contínuo, no qual fatos novos relevantes alteram o valor *verdade ou credibilidade* sobre um determinado fato.

Em geral, a inferência clássica é uma área que apresenta os resultados através de uma amostra para retirar conclusões sobre uma determinada população. No método bayesiano pode-se representar a ferramenta através de uma distribuição de probabilidade, e essa distribuição de probabilidade é obtida como um agregado de parâmetros, pois dependerá desses parâmetros para caracterizá-la por completo.

3.2 Inferência clássica

Na inferência clássica estima-se através da amostra, ou melhor, função de verossimilhança. A inferência clássica é composta por duas suposições: a primeira é que toda informação necessária encontra-se nos dados amostrais e a segunda é o processo de estimação, a qual garante a informação sobre a distribuição limite dos estimadores. No esquema da inferência clássica, mostrado na Figura 2, percebe-se que a inferência estatística inicia com o modelo experimental, depois retira uma amostra para começar o processo de avaliação, posteriormente passa por um raciocínio indutivo para chegar à inferência estatística, veja Bickel & Doksum (1977) e Mood *et al.* (1983).

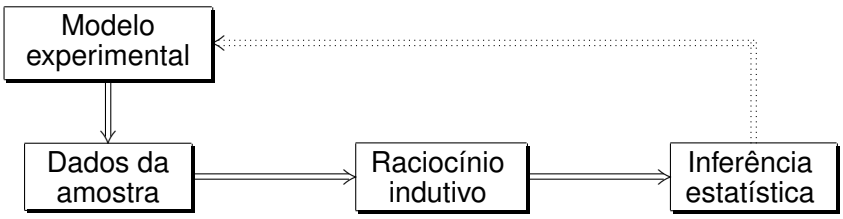

Figura 2: Esquema da inferência clássica.

3.3 Inferência bayesiana

Quando se estuda a inferência bayesiana verifica-se que existe uma diferença fundamental da inferência clássica. Essa diferença está no fato de que na abordagem bayesiana nem toda

informação sobre o parâmetro vem da amostra, mais especificamente, uma parte vem do conhecimento a *priori* sobre o parâmetro, e a outra parte da função de verossimilhança. A crença prévia do pesquisador é expressa por uma distribuição de probabilidade que é conhecida como distribuição a *priori* do parâmetro. Na Figura 3 verifica-se o esquema da inferência bayesiana, na qual inicia com o modelo experimental para depois tomar uma amostra conjuntamente com a distribuição a *priori*, logo em seguida aplica o teorema de Bayes para obter a distribuição a *posteriori* e finalmente com o raciocínio dedutivo chegar à inferência estatística ou a estimativa (Box & Tião, 1973 e Paulino *et al.* 2003).

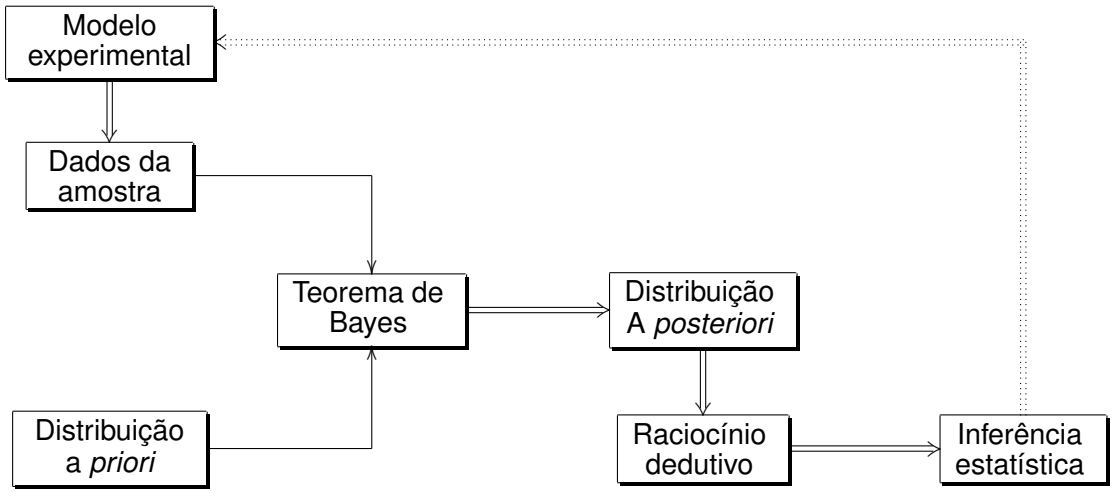

Figura 3: Esquema da inferência bayesiana.

3.4 Distribuição *priori* conjugada utilizando o modelo normal

Nesta seção será apresentada a distribuição a *posteriori* com a verossimilhança e *priori* Normal para em seguida fazer a implementação do modelo de regressão normal. Agora, supondo que as amostras são tomadas de uma distribuição normal para a qual o valor da média θ é desconhecida, mas o valor da variância σ^2 é conhecido, a família de distribuição normal é, ela própria; isto é, uma família conjugada de distribuições a *priori*, como é mostrado abaixo.

Suponha que $x_1, x_2, \ldots, x_n$ ou $\underset{\sim}{x}$ formam uma amostra aleatória de uma distribuição normal para a qual o valor da média θ ($\theta \in \mathbb{R}$) é desconhecido e o valor da variância σ^2 ($\sigma^2 \in \mathbb{R}^+$) é conhecido. Suponha também que a distribuição a *priori* de θ é uma distribuição normal com valores dados da média μ e variância ν^2. Com isso, tem-se que a função de verossimilhança é dada por

$$
\begin{aligned}
L_{\underset{\sim}{x}}(\underline{x}|\,\theta) &= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right\}. \qquad (3.1)
\end{aligned}
$$

Logo, tem-se que

$$
\begin{aligned}
L_{\underset{\sim}{x}}(\underline{x}|\,\theta) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\theta)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\theta-\bar{x})^2\right]\right\}.
\end{aligned}
$$

Também, pode-se descrever da seguinte maneira:

$$
L_{\underset{\sim}{x}}(\underline{x}|\,\theta) \propto \exp\left\{-\frac{1}{2\sigma^2}[n(\theta-\bar{x})^2]\right\}.
$$

A função de densidade de probabilidade a *priori* tem a forma

$$
\xi(\theta) \propto \exp\left\{-\frac{1}{2\nu^2}(\theta-\mu)^2\right\},
$$

por isso a fdp a *posteriori* segue abaixo,

$$
\begin{aligned}
\xi(\theta|\,\underline{x}) &\propto \exp\left\{-\frac{1}{2\sigma^2}\left[n(\theta-\bar{x})^2\right]\right\}\exp\left\{-\frac{1}{2\nu^2}\left[(\theta-\mu)^2\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\frac{n}{\sigma^2}\left(\theta^2-2\theta\bar{x}+\bar{x}^2\right)+\frac{1}{\nu^2}\left(\theta^2-2\theta\mu+\mu^2\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{n}{\sigma^2}+\frac{1}{\nu^2}\right)-2\theta\left(\frac{n\bar{x}}{\sigma^2}+\frac{\mu}{\nu^2}\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{n\nu^2+\sigma^2}{\sigma^2\nu^2}\right)\left[\theta^2-2\theta\left(\frac{n\bar{x}\nu^2+\mu\sigma^2}{n\nu^2+\sigma^2}\right)\right]\right\} \\
&\propto \exp\left[-\frac{1}{2\varphi^2}\left(\theta^2-2\theta\eta\right)\right],
\end{aligned}
$$

em que, $\varphi^2=\left(\frac{\sigma^2\nu^2}{n\nu^2+\sigma^2}\right)$ e $\eta=\left(\frac{n\bar{x}\nu^2+\mu\sigma^2}{n\nu^2+\sigma^2}\right)$. Como η não depende do parâmetro θ pode-se completar o quadrado da função e depois descartar o $\left(-\eta^2\right)$ para enfim, encontrar a *posteriori* da distribuição conjugada. Assim,

$$\xi(\theta|\underset{\sim}{x}) \quad \propto \quad \exp\left[-\frac{1}{2\varphi^2}(\theta-\eta)^2\right]. \tag{3.2}$$

Portanto, $\xi \sim \mathsf{N}(\eta, \varphi^2)$, isto é, $\mathsf{E}(\theta|\underset{\sim}{x}) = \eta$ e $\mathsf{Var}(\theta|\underset{\sim}{x}) = \varphi^2$.

Pode-se ver que η é uma média ponderada da média μ da distribuição a *priori* e da média amostral $\bar{x}$. Além disso, o peso relativo dado a $\bar{x}$ satisfaz as três propriedades:

I. Para valores fixos de σ^2 e n, obtém-se maior variância ν^2 da distribuição a *priori* e consequentemente será dado a $\bar{x}$ um maior peso relativo.

II. Fixando valores de σ^2 e ν^2, percebe-se que aumentando o tamanho n, também aumentará o peso relativo que é dado a $\bar{x}$;

III. Agora, para valores fixos de ν^2 e n, verifica-se que aumentando a variância σ^2 de cada observação da amostra, será menor o peso relativo que é dado a $\bar{x}$.

Exemplo 3.4.1 (**variância da distribuição normal a *posteriori*).** Dada uma distribuição normal com média θ, desconhecida, e a variância igual a uma unidade com a distribuição a *priori* de θ Gaussiana, no qual a variância é igual a 6. Imagina-se que as observações são capturadas até que a variância da distribuição a *posteriori* de θ tenha reduzido a valor menor ou igual a 1%. Logo, será determinado o número de observações que devem ser tomadas até que o processo de amostragem seja interrompido. Portanto, tem-se que a variância é dada por

$$\varphi^2 = \frac{\sigma^2\nu^2}{n\nu^2+\sigma^2} = \frac{6}{6n+1}.$$

Consequentemente, a relação $\varphi^2 \leq 0,01$ será satisfeita se e somente se $n \geq 99,83$, isto é, depois que 100 observações forem tomadas e não antes disso.

No caso da distribuição normal com média conhecida e variância desconhecida o desenvolvimento será igual ao processo que se discorreu nesta seção, na qual a distribuição a *posteriori* é semelhante a Equação 3.2. Para a distribuição normal com ambos os parâmetros desconhecidos, μ e σ^2, terá que fazer dois estágios analíticos: o primeiro para o parâmetro de escala (μ) e o segundo para o parâmetro de forma (σ^2), as principais distribuições conjugadas a *priori* utilizando as distribuições clássicas encontram-se na Tabela 1.

Tabela 1: Principais distribuições conjugadas a *priori* utilizando as distribuições clássicas.

Distribuição	Verossimilhança	Distribuição a *priori*	Distribuição a *posteriori*
Poisson	$Y_i \sim \text{Poisson}(\theta)$	$\theta \sim \text{Gama}(a,b)$	$\widehat{a} = n\bar{y} + a,\ \widehat{b} = n + b$
Binomial	$Y_i \sim \text{Binomial}(\theta, N_i)$	$\theta \sim \text{Beta}(a,b)$	$\widehat{a} = n\bar{y} + a,\ \widehat{b} = n\overline{N} + b$
Multinomial	$Y_i \sim \text{Multinomial}(\theta, n_i)$	$\theta \sim \text{Dirichlet}(\theta_i)$	$\widehat{a} = n\bar{y} + \theta_i$
Binomial Negativa$-$BN	$Y_i \sim \text{BN}(\theta, m_i)$	$\theta \sim \text{Beta}(a,b)$	$\widehat{a} = n\overline{m} + a,\ \widehat{b} = n\bar{y} + b$
Exponencial	$Y_i \sim \text{Exponencial}(\theta)$	$\theta \,\vert\, \alpha \sim \text{Gama}(a,b)$	$\widehat{a} = n + a,\ \widehat{b} = n\bar{y} + b$
Gama (α conhecido)	$Y_i \sim \text{Gama}(\alpha, \beta)$	$\theta \,\vert\, \alpha \sim \text{Gama}(a,b)$	$\widehat{a} = n\alpha + a,\ \widehat{b} = n\bar{y} + b$
Normal (σ^2 conhecido)	$Y_i \sim \text{Normal}(\eta, \sigma^2)$	$\eta \,\vert\, \varphi^2 \sim \text{N}(\mu, \nu)$	$\eta = \left(\frac{n\bar{y}\nu^2 + \mu\sigma^2}{n\nu^2 + \sigma^2}\right)$, $\varphi^2 = \left(\frac{\sigma^2\nu^2}{n\nu^2 + \sigma^2}\right)$

3.5 A regressão normal multivariada com a *priori* normal

A densidade normal multivariada é uma generalização da densidade da normal univariada, cuja a função foi citada na Equação 3.1. Iniciando com o expoente da densidade da normal univariada:

$$\frac{(x-\mu)^2}{\sigma^2} = (x-\mu)^t(\sigma^2)^{-1}(x-\mu).$$

O qual, mede a distância quadrada entre a média e o ponto observado. Agora, faz-se a generalização para o caso multivariado, com o vetor $(x_1, x_2, \ldots, x_n)$ ou $\underset{\sim}{x}$ dada por $\left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)^t (\Sigma)^{-1} \left(\underset{\sim}{x} - \underset{\sim}{\mu}\right)$. Suponha que se tem uma amostra de tamanho n com o vetor $\underset{\sim}{y}$ correspondendo a variável aleatória $\boldsymbol{Y}$, então tem-se a seguinte expressão:

$$\boldsymbol{Y}|\mu, \Sigma \sim N_n\left(\boldsymbol{X\beta}, \Sigma\right), \text{ e o modelo é } \boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon},$$

em que, $\boldsymbol{X}$ é a matriz de dados $(n \times p)$ conjuntamente com os valores das covariaveis, $\boldsymbol{\beta}$ é o vetor específico de parâmetros $(p \times 1)$, o Σ é a matriz $(n \times n)$ positiva definida que representa a matriz de covariâncias das variáveis e o $\boldsymbol{\varepsilon}$ é um vetor de variáveis aleatórias não observáveis $(n \times 1)$. Em seguida, é dado o modelo da função de verossimilhança.[1]

[1] O desenvolvimento, detalhado, da função de verossimilhança da distribuição normal multivariada encontra-se nos livros textos do Souza (1998), Montgomery & Runger (2003) ou Bussab & Morettin (2010).

$$\begin{aligned} f(y_1, y_2, \ldots, y_n \mid \boldsymbol{\beta}, \Sigma, x_{11}, x_{12}, \ldots, x_{np}) &= f(\boldsymbol{Y} \mid \boldsymbol{\beta}, \Sigma, \boldsymbol{X}) \\ &= (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(Y - X\beta)^t \Sigma^{-1} (Y - X\beta)\right]. \end{aligned}$$

É comum, se utilizar a distribuição normal como *priori* nos modelos de regressão logística e modelo de regressão linear. Logo, será considerado como a *priori* a distribuição normal. Então,

$$\beta_j \sim N(\mu_{\beta_j};\, \sigma^2_{\beta_j}) \quad \text{para} \quad j = 0, 1, 2, \ldots, p. \text{ Com } p \text{ parâmetros.}$$

Assim, aplica-se o Teorema de Bayes[2] para encontrar a distribuição a *posteriori* através da função de verossimilhança e *priori* normal para os parâmetros da distribuição de interesse.

$$\begin{aligned} f(\boldsymbol{\beta} \mid \boldsymbol{y}) &\propto f(\boldsymbol{y} \mid \beta_0, \beta_1, \beta_2, \ldots, \beta_p) f(\beta_0, \beta_1, \beta_2, \ldots, \beta_p) \\ &\propto \exp\left\{ -\frac{1}{2}(Y - X\beta)^t \Sigma^{-1} (Y - X\beta) - \frac{1}{2}\left[\left(\frac{\beta_0 - \mu_{\beta_0}}{\sigma_{\beta_0}} \right)^2 + \ldots + \left(\frac{\beta_p - \mu_{\beta_p}}{\sigma_{\beta_p}} \right)^2 \right] \right\} \\ &\propto \exp\left\{ -\frac{1}{2}\left[(Y - X\beta)^t \Sigma^{-1} (Y - X\beta) + \sum_{j=1}^{p} \left(\frac{\beta_j - \mu_{\beta_j}}{\sigma_{\beta_j}} \right)^2 \right] \right\}. \end{aligned} \tag{3.3}$$

Em linhas gerais, observa-se que na seção anterior, onde não havia envolvimento do modelo de regressão linear, obteve facilmente o resultado analítico e consequentemente chegando a um resultado satisfatório. Já nessa seção de modelo de regressão normal com *priori* normal, é diferente, verifica-se no Modelo 3.3 que a distribuição condicional marginal a *posteriori* é complicada de se resolver analiticamente. Por isso, sugere-se utilizar o método de Monte Carlo via cadeia de Markov por meio dos algoritmos de *Gibbs* ou *Metropolis-Hastings* para encontrar a distribuição a *posteriori*.

[2] O Teorema de Bayes encontra-se nos livros do Meyer (1983), James (2006) ou Magalhães (2006).

4 Equações simultâneas

4.1 Introdução

Em geral a grande preocupação é apenas com modelos de uma única equação, ou seja, com modelos em que exista uma única variável dependente Y e uma ou mais variáveis independentes X.[1] Nesses modelos, destaca-se a estimação e/ou previsão do valor médio de Y condicionado a valores de X. Porém, em alguns momentos essa relação de mão única não tem lógica e isso ocorre quando Y é determinado pelas variáveis independentes, então existe uma relação de mão dupla, ou simultânea, entre Y e alguns X's, em que se torna a distinção entre variáveis independentes e dependentes de valor muito duvidoso. Logo, é interessante fazer um agrupamento de variáveis que possa ser determinada simultaneamente pelo conjunto oposto dessas variáveis, onde é denominado de modelos de equações simultâneas, veja Madalla (1992), Greene (1997) e Baltagi (2005,2008).

No contexto de regressão, constantemente se utiliza uma especificação linear, mas devido ao problema de simultaneidade que existe entre as duas variáveis dependentes aplica-se esta técnica de equações simultâneas pelo método de MQ2E, método aplicado por Nieswiadomy & Molina (1988, 1991), Barbosa (1983), Braga & Gomes (2008). Enfim, tem-se que a interligação das variáveis dependentes é pelo fato da correlação entre o erro aleatório e a primeira variável dependente, causando um grande viés.

Exemplo 4.1.1. *Considerando o estudo do mercado de madeira em tora para produção de celulose no Brasil, os seguintes modelos conceituais são:*

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}.$$

[1] No contexto dos modelos de equações simultâneas, as variáveis conjuntamente endógenas são denominadas variáveis dependentes e as variáveis exógenas são chamadas de não estocásticas ou independentes.

Onde, $\boldsymbol{Y} = \begin{bmatrix} Q_t^D \\ Q_t^O \end{bmatrix}$, $\boldsymbol{X} = \begin{bmatrix} 1 & P_t & C_t & VE_t \\ 1 & P_t & IBNDES_t & PRO_t \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \alpha_0 & \beta_0 \\ \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \alpha_3 & \beta_3 \end{bmatrix}$ e $\boldsymbol{e} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$.

Em que, Q_t^D = quantidade demandada;

Q_t^O = quantidade ofertada;

P_t = preço de madeira em tora para produção de celulose;

C_t = capacidade instalada;

VE_t = volume de exportação;

$IBNDES_t$ = investimento do banco nacional de desenvolvimento;

PRO_t = produtividade;

t = tempo;

u_{1t} e u_{2t} = termos de erro aleatórios.

Essas variáveis respostas estocásticas deverão estar fortemente correlacionadas com os erros aleatórios estocásticos, onde torna o método clássico de MQO inaplicável à estimativa dos parâmetros das duas equações individualmente.

4.2 Identificação para o modelo estrutural

Antes de saber qual a técnica de equações simultâneas que irá aplicar; de mínimos quadrados em dois estágios, mínimos quadrados indireto ou método estrutural, é necessário testar o sistema de equações simultâneas a fim de observar se ele é subidentificado, superidentificado ou exatamente identificado. Então, utilizar-se-á a notação abaixo para fazer o teste do método.

I. DM : número de variáveis dependentes do modelo;

II. DE : número de variáveis dependentes da equação;

III. IM : número de variáveis predeterminadas do modelo, incluindo o intercepto;

IV. IE : número de variáveis predeterminadas da equação.

Para identificar a equação, por meio do modelo de equações simultâneas, o número de variáveis dependentes incluídas na equação menos um não deve ser maior que o número de

variáveis predeterminadas excluídas da equação, ou seja,

$$DE - 1 \leq IM - IE.$$

Caso aconteça de $DE - 1 < IM - IE$, ela será superidentificada; se $DE - 1 = IM - IE$, a equação está exatamente identificada e caso $DE - 1 > IM - IE$ a equação será subidentificado.

4.3 Teste de simultaneidade

Inicialmente aplica-se o teste Hausman para se determinar se existem problema de simultaneidade. O teste verifica se um regressor (dependente) se correlaciona com o termo de erro. Segundo Hausman (1976), se existir, pode-se utilizar um dos métodos de equações simultâneas, caso contrário, recorre-se a MQO. Com a presença de simultaneidade, os métodos de mínimos quadrados em dois estágios, mínimos quadrados indireto e método estrutural gerarão estimadores consistentes e eficientes.

Como foi citado anteriormente, esse problema surge devido a alguns dos regressores serem dependentes e, logo, se correlaciona com o termo do erro aleatório. Por isso, é essencial o teste de simultaneidade, onde se verifica a variável dependente está correlacionada com o erro. Para verificar este caso pode-se recorrer ao teste de especificação de Hausman, como é discorrido abaixo.

4.3.1 O teste de especificação de Hausman

Para explicar o teste de especificação de Hausman toma-se um caso particular com uma variável independente e duas variáveis dependentes. Agora, suponha que X é variável independente. Consequentemente, P e Q são variáveis dependentes.

Novamente, considere o caso particular. Caso não haja problema de simultaneidade, ou seja, as variáveis P e Q serão mutuamente independentes, P_t e u_t não deverão ter correlação. Caso haja simultaneidade, P_t e u_t serão correlacionadas. Então, para verificar se isto realmente ocorre, o teste de Hausman segue o seguinte processo:

Primeiramente, obtém-se as equações na forma reduzida:

$$\begin{aligned} \mathrm{P}_t &= \Pi_0 + \Pi_1 X_t + v_t, \\ \mathrm{Q}_t &= \Pi_2 + \Pi_3 X_t + w_t, \end{aligned} \tag{4.1}$$

em que, v e w são os termos do erro aleatório na forma reduzida. Agora, estimando a Equação (4.1) por MQO, obtém-se:

$$\widehat{\mathsf{P}}_t = \widehat{\Pi}_0 + \widehat{\Pi}_1 X_t + \widehat{v}_t. \tag{4.2}$$

Portanto,

$$\mathsf{P}_t = \widehat{P}_t + \widehat{v}_t, \tag{4.3}$$

No qual, $\widehat{P_t}$ são os P_t estimados e $\widehat{v}_t$ são os resíduos estimados. Fazendo a substituição da Expressão (4.3) na equação oferta, obtém-se:

$$\mathsf{Q}_t = \beta_0 + \beta_1 \widehat{P}_t + \beta_1 \widehat{v}_t + \widehat{u}_{2t}, \tag{4.4}$$

Lembrando que os parâmetros P_t e v_t terão os mesmos coeficientes.

- **As hipóteses estatísticas são:**

  $\rightsquigarrow$ H_0 : As equações não são simultâneas;

  $\rightsquigarrow$ H_1 : As equações são simultâneas.

Fazendo o teste sob a hipótese nula de que não existe simultaneidade, a correlação entre $\widehat{v}_t$ e $\widehat{u}_{2t}$ deveria ser, assintoticamente, igual a zero. Portanto, se estimar a regressão (4.4) e observar que esse coeficiente de v_t é estatisticamente igual a zero, deve-se afirmar que não existe problema de simultaneidade. Caso contrário, existirá problema de simultaneidade e tem-se que recorrer aos métodos de equações simultâneas. Por último, descreve-se de forma geral os dois passos do teste de Hausman.

I. **Primeiro Passo:** Faz-se a regressão da primeira variável dependente P_t contra a variável independente X_t para obter os resíduos $\widehat{v}_t$.

II. **Segundo Passo:** Gera a regressão com a segunda variável dependente Q_t contra as estimativas do primeiro modelo $\widehat{P}_t$ e os resíduos $\widehat{v}_t$, para em seguida aplicar o teste$-t$ ao coeficiente do resíduo $\widehat{v}_t$.[2] Caso seja significativo, não se rejeita a hipótese nula de simultaneidade, caso contrário, rejeita-se a hipótese nula.[3]

[2]Caso o modelo tenha mais de uma variável dependente envolvida e queira testar a significância conjunta em dois ou mais coeficientes, será aplicado o teste $\mathcal{F}$.

[3]Segundo Pindyck & Rubinfeld (2004), pode-se obter um melhor resultado se fizer a regressão de Q_t contra P_t e $\widehat{v}_t$.

4.4 Método de mínimos quadrados em dois estágios

No sistema de equações simultâneas superidentificadas ou exatamente identificadas, o método mais adequado é o MQ2E para estimar seus parâmetros. Segundo Martin & Perez (1975), tem grandes vantagens em utilizar este método, pela facilidade de implementação nos softwares como a obtenção de estimadores eficientes para pequenas amostras.

O método de mínimos quadrados em dois estágios consiste em duas modelagens de MQO: a primeira modelagem faz-se com todas as variáveis independentes e a variável dependente é a variável que está sobreidentificada, em seguida estima o primeiro modelo para gerar a segunda modelagem com a outra variável dependente, abaixo será discorrido de formal funcional.

I. **Primeiro Estágio:** Na primeira modelagem, faz-se a regressão Y_{1t} sobre todas as variáveis independentes, em todo o sistema. Por exemplo, supondo que se tem duas variáveis independentes e duas variáveis dependentes no modelo oferta, assim obtém-se o seguinte modelo,

$$Y_{1t} = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t, \tag{4.5}$$

onde u_t são os resíduos de MQO. Pelo modelo (4.5), obtém-se:

$$\widehat{Y}_{1t} = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1t} + \widehat{\beta}_2 X_{2t},$$

em que $\widehat{Y}_{1t}$ é as estimativas do valor esperado de Y_{1t} condicionado as variáveis independentes. Logo, a expressão (4.5) pode ser reescrita da seguinte forma:

$$Y_{1t} = \widehat{Y}_{1t} + \widehat{u}_t.$$

II. **Segundo Estágio:** Agora, pode-se escrever o segundo modelo da equação superidentificada de oferta do seguinte modo:

$$\begin{aligned} Y_{2t} &= \beta_{20} + \beta_{21}(\widehat{Y}_{1t} + \widehat{u}_t) + u_{2t} \\ &= \beta_{20} + \beta_{21}\widehat{Y}_{1t} + u_t^*. \quad \text{No qual,} \quad u_t^* = u_{2t} + \beta_{21}\widehat{u}_t. \end{aligned}$$

4.4.1 Correções dos desvios padrão dos estimadores de MQ2E

Os desvios padrão ou erros padrão que se obtém no segundo estágio do procedimento de MQ2E precisa fazer uma correção. Pois, se observar o modelo do segundo estágio verifica que o $\widehat{\sigma}^2_{u^*}$ é diferente do $\widehat{\sigma}^2_{u_2}$. Isto é, a primeira variância depende das estimativas da variável

resposta, enquanto o outro termo depende do verdadeiro valor real da resposta, veja Gujarati (2006) e Madalla (1992).

Através do exemplo de oferta, visto no método de MQ2E, obtém-se as seguintes expressões abaixo,

$$\widehat{\sigma}^2_{u_2} = \frac{\sum_{t=1}^{n}(\widehat{u}_{2t})^2}{n-2} = \frac{\sum_{t=1}^{n}(Y_{2t} - \widehat{\beta}_{20} - \widehat{\beta}_{21}Y_{1t})^2}{n-2}. \tag{4.6}$$

E,

$$\widehat{\sigma}^2_{u^*} = \frac{\sum_{t=1}^{n}(\widehat{u}^*_t)^2}{n-2} = \frac{\sum_{t=1}^{n}(Y_{2t} - \widehat{\beta}_{20} - \widehat{\beta}_{21}\widehat{Y}_{1t})^2}{n-2}. \tag{4.7}$$

Depois de calculado esses valores o modo de corrigir os erros padrão dos coeficiente estimados na regressão de mínimos quadrados em dois estágios é multiplicar cada um desses coeficientes pela divisão do resultado de (4.6) por (4.7). Caso o R^2 seja muito alto (mais ou menos acima de 0,80) na regressão do primeiro estágio, ou seja, o valor estimado esteja muito próximo do verdadeiro valor real, o fator de correção será aproximadamente 1 (um), onde o pesquisador poderá permanecer com os desvios padrão do segundo estágio, sem precisar atualizar os valores.

Depois de descrever as ferramentas chaves da técnica de equações simultâneas e o estimador de mínimos quadrados em dois estágios será discorrido a regressão linear com o método de mínimos quadrados ordinários, estatística do teste-t, coeficiente de determinação e os critérios de informações.

4.5 Regressão linear

Uma das utilidades do modelo de regressão linear é analisar a relação entre uma variável dependente e uma ou mais variáveis explicativas.[4] Sendo assim, o objetivo principal da análise de regressão é encontrar uma função linear que permita descrever e compreender a relação entre uma variável dependente e uma ou mais variáveis independentes. Na forma matricial, o modelo é dado por:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{4.8}$$

em que $\boldsymbol{Y}$ é o vetor de variáveis resposta (ou variável resposta) de dimensão $T \times 1$; $\boldsymbol{X}$ é uma matriz com as variáveis explicativas de dimensão $T \times k$ e tem pelo menos dois valores distin-

[4] Também pode-se relatar variável explicativa ao invés de descrever variável não estocástica ou independente.

tos na amostra; $\boldsymbol{\beta}$ é o vetor de parâmetros desconhecidos com dimensão $k \times 1$; as variáveis aleatórias $\boldsymbol{\varepsilon}$ de dimensão $T \times 1$ com média zero, são não correlacionadas e possuem variância comum σ^2.

4.5.1 Método de mínimos quadrados ordinários

O método de mínimos quadrados ordinários é útil para estimar o vetor de parâmetro $\boldsymbol{\beta}$ do modelo (4.8). O enfoque principal é a minimização da forma quadrática por meio da Soma de Quadrada dos Resíduos (SQR). Então, tem-se que:

$$\widehat{\boldsymbol{\beta}} = (X^t X)^{-1} X^t y. \tag{4.9}$$

O estimador (4.9) corresponde ao ponto de mínimo de (4.8), pelo fato da matriz $X^t X$ ser definida positiva. Por último, tem-se que o estimador (4.9) é estatisticamente suficiente e completo.[5]

4.5.2 Estatística do teste-t

As hipóteses para testar se as variáveis x_i, sendo $i = 1, 2, \ldots, k$ é significativo para o modelo são:

$$\begin{cases} H_0: & \beta_i = 0; \\ H_1: & \beta_i \neq 0. \end{cases}$$

Usa-se a estatística do teste (4.10) para testar se a hipótese é ou não significante, sob a hipótese nula. No qual, se segue uma distribuição t-Student com $(T-k)$ graus de liberdade, em que $|t| > t_{(\frac{\alpha}{2}, T-k)}$ ou o p-valor menor que α, em que, p-valor $= P(t_{(T-K)} > t)$ e α é o nível de significância.

$$t = \frac{b_i}{\text{DP}(b_i)}, \text{em que DP}(b_i) \text{ é o desvio padrão de } b_i \text{ e } b_i \text{ os valores estimados.} \tag{4.10}$$

Em algumas situações o pesquisador está interessado em saber, ou melhor, testar se há alguma relação linear entre a variável resposta e os regressores. Neste caso usa-se as seguintes hipóteses:

[5] Para maiores informações sobre a regressão linear, derivação de MQO, estatística do teste-t, coeficiente de determinação e critérios de informação, veja Montgomery & Runger (2003), Bussab & Morettin (2010) ou Casella & Berger (2011).

$$\begin{cases} H_0: & \beta_i = 0 \quad \forall\, i = 1,2,\ldots,k; \\ H_1: & \beta_i \neq 0 \quad \text{para algum } i. \end{cases}$$

Sendo assim, pode-se utilizar a estatística do teste $\mathcal{F}$ dada pela expressão 4.11, que sob a hipótese nula, tem distribuição $\mathcal{F}$ com $(k-1)$, $(T-k)$ graus de liberdade. Então, $\mathcal{F} > \mathcal{F}_{(1-\alpha,k-1,T-k)} = \mathcal{F}$, onde, $\mathcal{F}$ é o percentil $(1-\alpha)$ da distribuição $\mathcal{F}_{[(k-1),(T-k)]}$.

$$\mathcal{F} = \frac{\text{QMreg}}{\text{QMres}}, \tag{4.11}$$

Em que,

$$\text{QMreg} = \frac{\text{SQR}}{(k-1)} = \frac{\sum_{t=1}^{k}(\hat{y}_t - \bar{y})^2}{(k-1)}, \quad \text{e} \quad \text{QMres} = \frac{\text{SQE}}{(T-k)} = \frac{\sum_{t=1}^{k}(y_t - \hat{y}_t)^2}{(T-k)}.$$

No qual, T é igual ao número de observações e k igual ao número de variáveis.

4.5.3 Coeficiente de determinação

O coeficiente de determinação é uma medida da qualidade do ajuste através da quantidade da variabilidade da variável resposta y que é explicada pelo modelo de regressão linear ajustado. Então, o R^2 é dado por:

$$R^2 = \frac{SQR}{SQT}.$$

Porém, quando aumenta-se o número de variáveis no modelo, o R^2 tende a aumentar, dificultando a comparação de dois ou mais modelos. Entretanto, pode-se usar no lugar do R^2 o $\bar{R}^2$ ajustado que penaliza o R^2 pela adição de muitas variáveis ao modelo. O $\overline{R}^2$ ajustado é dado por:

$$\overline{R}^2 = 1 - \frac{SQE/(T-k)}{SQT/(T-1)} = 1 - \frac{(T-1)}{(T-k)}(1-R^2), \qquad \text{em que,} \quad \text{SQT} = \sum_{t=1}^{k}(y_t - \bar{y}_t)^2.$$

4.5.4 Critérios de informação

Em algumas situações, na modelagem estatística, tem-se o interesse em escolher o modelo através da minimização de algum critério de informação, no qual penalize a verossimilhança. Observa-se também que a função de verossimilhança é uma das propriedades mais sensível a

pequenos desvios dos parâmetros do verdadeiro valor do modelo. Portanto, Akaike (1970,1974) e Schwarz (1978) apresentaram os métodos abaixo

$$\text{AIC} = -2\ln(\text{verossimilhança}) + 2p \qquad \text{e} \qquad \text{BIC} = -2\ln(\text{verossimilhança}) + p\ln(n),$$

No qual, p é o número de parâmetro mais 1 (um) e o n é o tamanho amostral. Também, pode-se utilizar os critérios normalizados, isto é, dividi-os pelo tamanho da amostra. Então, tem-se que o $\text{AIC}_n = \text{AIC}/n$ e $\text{BIC}_n = \text{BIC}/n$. A expressão com o menor AIC_n e/ou BIC_n, entre todos os modelos ajustados, deve ser considerado como o que melhor explica as observações.

Por último, informa-se que os critérios de informação de Akaike (AIC) e Schwarz (BIC) são estatísticas bem conhecidas e de fácil implementação e interpretação para selecionar os modelos de regressão clássico e bayesiano (Veja Maddala, 1992; Gujarati, 2006 e Lopes & Salazar, 2011).

5 Monte Carlo via cadeia de Markov

5.1 Introdução

A simulação de Monte Carlo surgiu em 1964 quando os matemáticos Stanislaw Ulam, Richard Feynman e John Neuman (veja Fig. 4) jogavam paciência e no determinado momento surgiu a seguinte pergunta: `quais são as chances de que em um solitaire Canfield estabelecido com 52 cartões saia um sucesso?` Eles tentaram calcular as probabilidades utilizando a ferramenta de análise combinatória, porém perceberam que uma alternativa mais dinâmica seria simplesmente realizar várias jogadas — por exemplo, dez, cem ou mil jogadas — e fazer a contagem dos resultados que ocorresse.

Figura 4: Stanislaw Ulam, Richard Feynman e John Neuman.

Ulam era matemático e sabia que as técnicas de amostragem não se encaixava no problema, pelo fato de envolver cálculos extremamente demorados, fadigado e sujeito a muitos erros. Entretanto, nessa época ficou pronto o primeiro computador eletrônico, desenvolvido durante a segunda guerra mundial, o ENIAC, por J.P. Eckert e J.W. Mauchly na Universidade da Pensilvânia, Estados Unidos da América. Então, surgiu a simulação de Monte Carlo e o nome de Monte Carlo deve-se a Nicholas Metropolis que sugeriu aos pesquisadores, devido ao tio de Ulam que sempre visitava o Monte Carlo. A contribuição de Ulam foi o reconhecimento do potencial dos computadores eletrônicos para automatizar as amostragens. Assim, em 1949 eles publicaram o primeiro artigo com o algoritmo de Monte Carlo.

Por meio do livro do Hammersley & Handscomb (1964), o método de Monte Carlo é um método estatístico, em que utiliza simulações estocásticas, em vários tipos de aplicações como

por exemplo a matemática, física, biologia e ciências agrárias. A partir do final do século XX começaram a utilizar, com frequência, a simulação de Monte Carlo para obter aproximações numéricas de funções complicadas. O método envolve a geração de observações, através de vários experimentos de alguma distribuição de probabilidade e o uso da amostra obtida para aproximar a função de interesse.

O método MCMC surgiu no final do século XX tendo como ideia básica a construção de uma cadeia de Markov com uma distribuição de equilíbrio semelhante a de interesse, ou seja, simular um passeio aleatório no espaço do parâmetro que convirja para uma distribuição estacionária, no qual seja a distribuição de interesse no problema; cada situação, nessa cadeia, pode-se atingir a partir de qualquer outra distribuição com um número finito de iterações. Depois de um grande número de iterações, a cadeia converge para a distribuição de interesse e em seguida pode ser utilizado para inferências estatísticas.

Existem diversas vantagens nesse método e duas delas são: não se precisa ter o conhecimento sobre o tipo de distribuição em que se pretende simular, o que acontece em problemas práticos na inferência bayesiana; e, existem vários algoritmos para construir as cadeias de Markov que são necessários para a simulação, onde todos esses algoritmos tem como objetivo principal gerar observações para distribuição de interesse.

Na estatística computacional, um dos tópicos mais ativos é a inferência através de simulação iterativa, ou seja, aplicando o método de Monte Carlo via Cadeia de Markov por meio do amostrador de *Gibbs* e o algoritmo *Metropolis-Hastings*. Esses algoritmos, demanda um extensivo uso de recursos computacionais utilizando a teoria de MCMC para representar a dependência entre os parâmetros, por isso o amostrador de *Gibbs* e o MH faz uso do método conhecido como MCMC, veja Sorensen (1996).

Na segunda metade do século XX, onde já se tinha um grande avanço computacional, foram desenvolvidos os algoritmos de *Gibbs sample* e *Metropolis-Hastings*. O método de *Gibbs* foi o primeiro algoritmo a ser exposto para simulação estocástica usando cadeias de Markov. Geman & Geman (1984) foram os que desenvolveram o método para processamento de imagens sendo apresentado pela primeira vez na área da estatística. Contudo, o Gelfand & Smith (1990) tiveram sucesso em mostrar para comunidade estatística que o esquema desenvolvido poderia ser usado, em diversos casos, para mostrar a distribuição a *posteriori*. O algoritmo de *Metropolis-Hastings* é semelhante ao método de aceitação e rejeição, no qual é uma ferramenta do método de Monte Carlo via cadeia de Markov para simular distribuições de probabilidades complicadas. O método foi desenvolvido por Metropolis *et al.* (1953) e posteriormente generalizado por Hastings (1970). Quando o amostrador de *Gibbs* não se mostra eficiente, isto é,

para tipos de parâmetros cuja a marginal não se caracteriza como uma distribuição conhecida é interessante utilizar o método de *Metropolis-Hastings* para obtenção dos resultados.[1]

De forma geral, a diferença do amostrador de *Gibbs* para o *Metropolis-Hastings* é que o método *Gibbs* aceita todas as observações calculadas, enquanto o algoritmo *Metropolis-Hastings* gera-se observação a partir de uma distribuição proposta e esse valor é aceito ou não com uma certa probabilidade de aceitação, ou seja, é igual ao método de aceitação e rejeição. Agora será discorrido o algoritmos *Gibbs*, o algoritmo *Metropolis-Hastings* e os testes de convergências conjuntamente com os gráficos de estacionariedade e convergência, de acordo com Gilks *et al.* (1996) e Gamerman (1997).

5.2 *Gibbs*

No algoritmo *Gibbs* a cadeia sempre irá a um novo valor, ou seja, não existe o processo de aceitação ou rejeição. As passagens de um estado para o outro do método é feito de acordo com as distribuições condicionais completas representadas por $\pi(\theta_i|\theta_{-i})$, em que $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d)$'.

Os componentes do parâmetro θ_i, em geral, podem ser uni ou multidimensional. Logo, a distribuição condicional completa será a distribuição da $i-$ésima componente do parâmetro θ condicionado nas outras componentes. Sendo assim, pode-se obter por meio da seguinte distribuição conjunta:

$$\pi(\theta_i|\theta_{-i}) = \frac{\pi(\theta)}{\int \pi(\theta) d\theta_i}.$$

Em algumas determinadas circunstâncias a simulação de uma amostra de $\pi(\theta)$ pode levar muito tempo tornando-o insatisfatório, complicado ou impossível de encontrar. Porém, se forem conhecidas as distribuições condicionais completas a *posteriori*, pode-se utilizar o algoritmo de *Gibbs* pelos passos abaixo:

I. Comece o contador de iterações da cadeia $t = 0$;

II. Especifique os valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_n^{(0)})$';

[1] Referências complementares sobre os métodos de MCMC, veja Dey & Gelfand (1994), Chib & Greenberg (1993), Leotti (2007), Mayrink (2006), Duarte (2011) e Farias *et al.* (2011).

III. Obter um novo valor de $\theta^{(t)}$ a partir de $\theta^{(t-1)}$ através da geração sucessiva dos seguintes valores

$$\theta_1^{(t)} \sim \pi\left(\theta_1|\theta_2^{(t-1)},\theta_3^{(t-1)},\ldots,\theta_n^{(t-1)}\right)$$
$$\theta_2^{(t)} \sim \pi\left(\theta_2|\theta_1^{(t)},\theta_3^{(t-1)},\ldots,\theta_n^{(t-1)}\right)$$
$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$
$$\theta_n^{(t)} \sim \pi\left(\theta_n|\theta_1^{(t)},\theta_2^{(t)},\ldots,\theta_{n-1}^{(t)}\right);$$

IV. Atualize o contador de (t) para $(t+1)$ e volte ao segundo passo até obter a convergência.

Logo, só acontecerá uma iteração quando completar n movimentos ao longo dos eixos das coordenadas do parâmetro θ. Caso queira se aprofundar no amostrador de *Gibbs* pode recorrer a Casella & Robert (1999) ou Gamerman (1997,2006).

Exemplo 5.2.1.

Dado o exemplo do livro *Bayesian modeling using WinBUGS* para ilustrar o assunto do algoritmo de *Gibbs* tem-se o exemplo sobre a temperatura corporal (dados normal). Agora, considera-se a distribuição abaixo,

$$\mu \sim N(\mu_0,\sigma_0^2) \qquad \text{e} \qquad \sigma^2 \sim IG(a_0,b_0).$$

Para esse modelo deve-se construir o amostrador de *Gibbs* com a distribuição condicional dada por $f(\mu|\sigma^2,y)$ e $f(\sigma^2|\mu,y)$, ou seja,

$$\mu|\sigma^2,y \sim N\left(w\bar{y}+(1-w)\mu_0, w\frac{\sigma^2}{n}\right), \qquad \text{onde,} \qquad w=\frac{\sigma_0^2}{\sigma^2/n+\sigma_0^2},$$

e,

$$\sigma^2|\mu,y \sim IG\left(\alpha_0+\frac{n}{2}, b_0+\frac{1}{2}\sum_{i=1}^{n}(y_i-\mu)^2\right). \quad \text{Maiores detalhes, veja Ntzoufras (2009).}$$

Na Figura 5, estão os gráficos de iterações e densidade do exemplo acima, onde apresenta uma boa convergência sem apresentar irregularidades no processo. Em seguida, na Figura 6 se encontra o histograma dos dados que apresenta visualmente a forma de uma distribuição normal centrada nas médias $98,25$ e $0,74$.

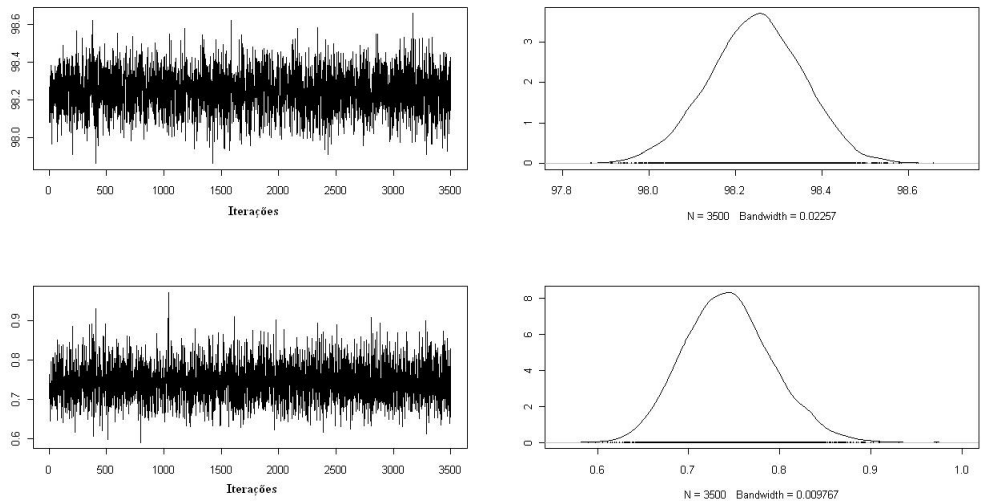

Figura 5: Gráfico de iterações e de densidade.

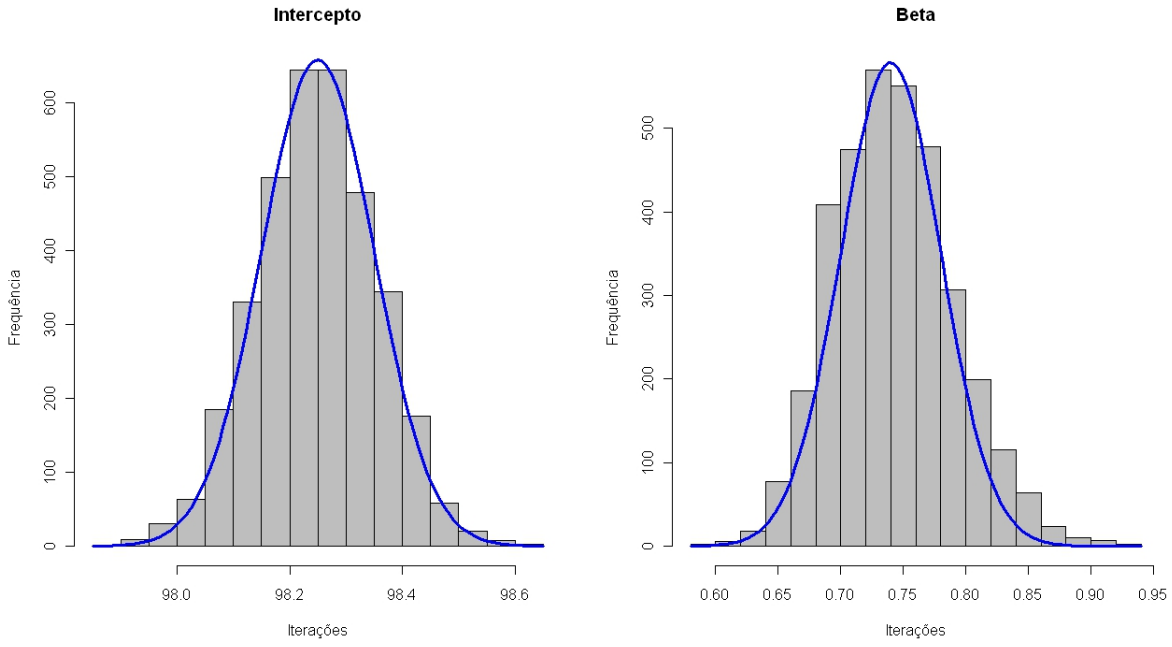

Figura 6: Histograma das observações simuladas.

5.3 *Metropolis-Hastings*

Uma das ferramentas mais poderosa do método MCMC é o algoritmo *Metropolis-hastings* para simulação de distribuições de probabilidade complicada. Ultimamente a comunidade estatística bayesiana tem se interessado em utilizar esse algoritmo, o qual foi desenvolvido inicialmente por *Metropolis* (1953) e em seguida *Hastings* (1970) fez a generalização, originando o algoritmo em *Metropolis-Hastings* (veja Müller, 1993; Phillips & Smith, 1994; Casella & Robert, (1999,2010) e Casella & Berger, 2011). A técnica foi publicada para a área de física e química,

neste meio tempo, mostrou-se bastante vantajoso para a simulação tornando-se um método aplicável na área da estatística.

Presumi-se que a cadeia esteja no estado θ e um novo valor θ^* é gerado de uma distribuição sugerida $Q(\cdot|\theta)$. Observa-se que a distribuição sugerida pode depender do estado atual da cadeia, por exemplo suponha que a distribuição sugerida seja uma normal centrada em θ.

$$\alpha(\theta,\theta^*) = \begin{cases} \min\left[\dfrac{\pi(\theta^*)\,Q(\theta|\theta^*)}{\pi(\theta)\,Q(\theta^*|\theta)},1\right], & \text{se} \quad \pi(\theta)\,Q(\theta^*|\theta) > 0; \\ 1, & \text{caso contrário.} \end{cases} \tag{5.1}$$

Em que $\pi(\cdot)$ é a distribuição de interesse.

O interessante é que basta conhecer o $\pi(\cdot)$ parcialmente, em outras palavras, tem-se que a probabilidade (5.1) não se altera, a menos de uma constante. Isso é essencial em aplicações bayesianas onde não é conhecido completamente a distribuição de interesse na simulação. Mais, a cadeia também pode ficar estável por muitas iterações e na prática costuma-se monitorar por meio da média, testes e gráficos.

Será descrito abaixo o algoritmo de *Metropolis-Hastings*, em termos práticos, através dos seguintes passos:

I. Comece o contador de iterações, $t = 0$, e especifique o valor inicial $\theta^{(0)}$;

II. Gere um novo valor θ^* da distribuição $Q(\cdot|\theta)$;

III. Calcule a probabilidade de aceitação da expressão (5.1) e gere uma $u \sim U(0,1)$;

IV. Caso $u \leq \alpha$ então aceite o novo valor e faça $\theta^{(t+1)} = \theta^*$, caso contrário não aceite e faça $\theta^{(t+1)} = \theta$;

V. Atualize o contador de (t) para $(t+1)$ e volte ao segundo passo.

A distribuição sugerida pode ser escolhida arbitrariamente, porém na prática deve-se ter alguns cuidados básicos para garantir uma boa eficiência do algoritmo. Nas aplicações a distribuição de interesse será a própria *posteriori*, ou seja, a expressão (5.1) e a $p(\theta|x)$ assumirá uma forma particular, como mostra em (5.2).

$$\alpha(\theta,\theta^*) = \begin{cases} \min\left[\dfrac{p(x|\theta^*)\,p(\theta^*)\,Q(\theta|\theta^*)}{p(x|\theta)\,p(\theta)\,Q(\theta^*|\theta)},1\right], & \text{se} \quad p(x|\theta)\,p(\theta)\,Q(\theta^*|\theta) > 0; \\ 1, & \text{caso contrário.} \end{cases} \tag{5.2}$$

Será mostrado agora, o exemplo ilustrativo da aplicação da ferramenta *Metropolis-Hastings*, fundamentado nas notas de aula de Ehlers (2003-2005) e Rossi (2011).

Exemplo 5.3.1.

Dada uma população de 209 plantas dos 4 (quatro) maiores tipos de genética do mundo $(y = 24, 40, 27, 118)$ com as seguintes probabilidades,

$$p_1 = \left(\frac{1}{2} + \frac{\theta}{4}\right), \qquad p_2 = \left(\frac{1-\theta}{4}\right), \qquad p_3 = \left(\frac{2-\theta}{16}\right) \qquad \text{e} \qquad p_4 = \left(\frac{1}{8} + \frac{\theta}{16}\right).$$

Em que θ é um parâmetro desconhecido e pertence ao intervalo 0 (zero) e 1 (um). Para todo o $\theta \in (0,1)$ tem-se que $p_i > 0, i = 1,2,3,4$ e $\sum_{i=1}^{4} p_i = 1$. Depois de observada as 209 plantas dentre os quais y_i pertencem à *i-ésima* genética e o vetor Y tem distribuição multinomial com os parâmetros p_i's. Logo,

$$\begin{aligned} p(y|\theta) &= \frac{n!}{y_1!y_2!y_3!y_4!} p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4} \\ &\propto (2+\theta)^{y_1+y_4}(1-\theta)^{y_2}(2-\theta)^{y_3}. \end{aligned} \tag{5.3}$$

Gerando uma *priori* $\theta \sim U(0,1)$ segue que a *posteriori* é proporcional á equação (5.3). Novamente simulando a *priori* acima como proposta inicial tem-se o $Q(\theta) = 1$, para qualquer valor do parâmetro θ e a probabilidade (5.1) será,

$$\alpha(\theta, \theta^*) = \min\left[\frac{p(y|\theta^*)}{p(y|\theta)}, 1\right] = \min\left[\left(\frac{2+\theta^*}{2+\theta}\right)^{(y_1+y_4)} \left(\frac{1-\theta^*}{1-\theta}\right)^{y_2} \left(\frac{2-\theta^*}{2-\theta}\right)^{y_3}, 1\right].$$

Em geral, foram verificadas 209 plantas nas categorias dadas por $y_i = (24, 40, 27, 118)$, com suas respectivas probabilidades. Também, foi gerado a cadeia de Markov com 3.500 iterações do parâmetro θ. Os gráficos das observações simuladas do parâmetro θ (depois do descarte das 100 primeiras observações) encontram-se nas Figuras 7 e 8. Em geral, observa-se que a cadeia é correlacionada ao longo da iteração e isto é devido a baixa taxa de aceitação pela escolha de $Q(\theta)$. Na aplicação obteve $29,18\%$ de aceitação das observações do processo.

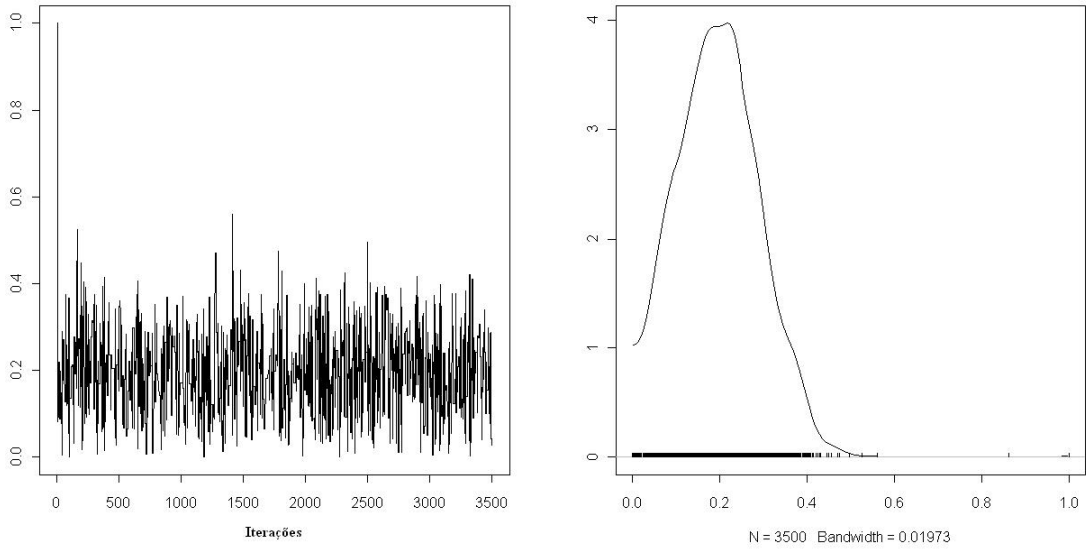

Figura 7: Gráfico de iterações e densidade.

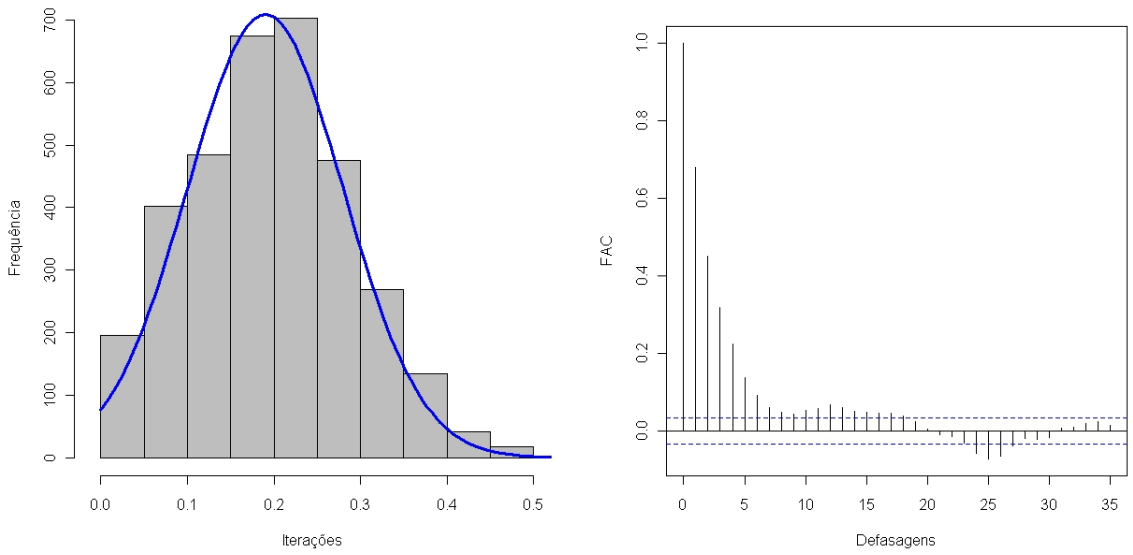

Figura 8: Histograma e correlações das observações simuladas após a queimação.

5.4 Análise de Convergência

No contexto bayesiano, um dos métodos de boa qualidade é o MCMC para resolução de muitos problemas práticos, mesmo assim, ainda requer uma vasta pesquisa no método MCMC e o diagnóstico de convergência estocástica. Logo, surgirá o seguinte problema: quantas iterações deve ter o processo de simulação MCMC para garantir que, de fato, a cadeia convergirá para a distribuição de interesse? Dificilmente responde-se a essa questão, pois a distribuição estacionária será na prática desconhecida. Contudo, pode-se avaliar a convergência

das cadeias detectando problemas que estejam fora do período de aquecimento.

inicialmente, faz-se uma análise de convergência por meio de gráficos ou medidas descritivas dos valores simulados da distribuição ou quantidade de interesse. Então, os gráficos a serem feitos são: o do parâmetro θ ao longo do processo de iterações, o histograma, a densidade de Kernel e o gráfico da estimativa da distribuição a posteriori de θ. As estatísticas utilizadas são as descritivas como a média, mediana, desvio padrão e os quartis $(25\%; 50\%; 75\%)$.

Os algoritmos desenvolvidos nesse artigo são o *Gibbs* e *Metropolis-Hastings* que são ferramentas iterativas, isto é, precisam ser verificados a convergência para fazer inferência sobre os valores das distribuições marginais dos parâmetros que estão sendo estudado. Sendo assim, tem-se os testes de diagnóstico de Raftery & Lewis (1992), Geweke (1992), Heidelberg & Welch (1983) e o de Gelman & Rubin (1992), dos 13 (treze) tipos de diagnóstico que existem atualmente para fazer a avaliação de convergência (Veja a Tabela 2). Também, será descrito e mostrado os gráficos de autocorrelações, histogramas, densidades de Kernel e os de diagnóstico de convergência.

Tabela 2: Sumário dos métodos de diagnóstico de convergência.

Método	Impressão do método	Quantidade de cadeias	Tipo da distribuição	Aplicação	
				Algoritmo	R
Gelman & Rubin (1992)	Quantitativo	Múltiplas	Univariada	MCMC	Sim
Raftery & Lewis (1992)	Quantitativo	Única	Univariada	MCMC	Sim
Geweke (1992)	Quantitativo	Única	Univariada	MCMC	Sim
Heidelberger & Welch (1983)	Quantitativo	Única	Univariada	MCMC	Sim
Roberts (1992,1994)	Gráfico	Múltiplas	Conjunta	Outros tipos	Não
Ritter & Tanner (1992)	Gráfico	Os dois tipos	Conjunta	MCMC-Gibbs	Não
Zellner & Min (1995)	Quantitativo	Única	Conjunta	Outros tipos	Não
Liu, Liu & Rubin (1992)	Os dois tipos	Múltiplas	Conjunta	MCMC-Gibbs	Não
Garren & Smith (1993)	Quantitativo	Múltiplas	Univariada	MCMC-Gibbs	Não
Johnson (1994)	Quantitativo	Múltiplas	Conjunta	Outros tipos	Não
Mykland, Tierney & Yu (1995)	Gráfico	Única	Conjunta	Outros tipos	Não
Yu (1994)	Gráfico	Única	Conjunta	Outros tipos	Não
Yu & Mykland (1994)	Gráfico	Única	Univariada	MCMC	Não

Fonte: resultados da pesquisa (veja, Carlin & Cowles 1996).

5.4.1 Diagnóstico de Raftery & Lewis

Caso, tenha interesse em obter o número de iterações necessárias para conseguir a convergência, da menor distância de uma iteração à outra, para conseguir uma amostra independente e também o número de iterações iniciais que deve ser descartadas é interessante o diagnóstico de Raftery & Lewis. O processo do algoritmo para obter os resultados segue abaixo:

I. Primeiro seleciona um quantil *posteriori* de interesse q (0,025 quantil);

II. Selecione uma tolerância r aceitável para este quantil, por exemplo se $r = 0,005$ significa que irá medir a 0,025 quantil tendo uma acurácia de $\pm 0,005$;

III. Agora será selecionado uma probabilidade s, em que deve está no intervalo de (q-r, q+r);

IV. Por fim, executa uma amostra piloto para gerar a cadeia de Markov de comprimento mínimo dado por,

$$\mathsf{n}_{\mathsf{min}} = \left[\psi^{-1} \frac{\sqrt{q(1-q)(s+1)^2}}{2r} \right]^2,$$

onde, ψ^{-1} é a inversa da distribuição normal. Caso aconteça do fator de dependência ser maior que 5 (cinco), pode-se afirmar que existe uma alta correlação entre coeficientes e portanto é interessante aumentar o tamanho da iteração.

5.4.2 Diagnóstico de Geweke

O diagnóstico de Geweke é um critério para a ausência de convergência. Igualmente, o método de Geweke propõe o diagnóstico de convergência para o método de Monte Carlo via cadeia de Markov baseado no teste de igualdade de médias da primeira (10%) e última parte da cadeia (50%), veja Geweke (1992) para obter uma descrição completa do método. A estatística de teste segue o Z-escore padrão seguindo dos erros padrão ajustados para autocorrelação.

Agora, suponha que o pesquisador está interessado em estimar o parâmetro θ por meio do MCMC, para fazer a análise da convergência pode-se utilizar esse diagnóstico, no qual calcula-se as devidas médias,

$$\mu_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \theta^i \qquad \text{e} \qquad \mu_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \theta^i, \qquad (n_1 + n_2) < n \quad (\text{iterações}),$$

em que, o i indica a posição na cadeia, μ_1 será o valor da média que foi calculada utilizando as n_1 primeiras posições na cadeia e o μ_2 será o valor da média calculada aplicando as n_2 últimas posições na cadeia. Caso os valores de μ_1 e μ_2 forem iguais descreve-se que a cadeia convergiu, em que será fixo o n_1 sobre a população e dividir o n_2 pela população, e também a população tenderá ao infinito. Portanto,

$$\frac{(\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \longrightarrow N(0,1),$$

em que, os valores de s_1^2 e s_2^2 são estimativas das variâncias, assintóticamente, das médias de μ_1 e μ_2. Logo, será assumido convergência se o valor desta estatística ficar entre $\pm 1,96$.

5.4.3 Diagnóstico de Heidelberg & Welch

O terceiro tipo de diagnóstico é o Heidelberg & Welch que testa a hipótese nula de estacionariedade da sequência gerada, com base na estatística de teste *Cramer-von Mises*, para aceitar ou rejeitar a hipótese nula de que a cadeia de Markov é estacionária. Nesse diagnóstico, o teste será repetido se a hipótese nula for rejeitado. Caso aconteça do teste ser rejeitado terá que descartar 10% das observaçõe e esse processo será parado quando for descartadas 50% dos valores iniciais. Caso a hipótese nula seja rejeitada, depois do primeiro processo, indicará um aumento no número de iterações, evento contrário, o número inicial de iterações descartadas será indicado como o tamanho da queimação. Também, o critério utiliza o teste de HalfWidth para fazer a análise da estimativa da média, isto é, verificar se a média calculada obteve uma boa acurácia pré-especificada. Portanto, se o resultado for positivo, a média estimada estará com um erro aceitável e podendo assim ser a média da distribuição de interesse. Esse diagnóstico consiste de duas partes, sendo-as:

Primeira parte do processo:

I. Será gerado uma cadeia de N iterações e defini-se um nível de significância α;

II. Calcular a estatística de teste na cadeia. Em seguida, rejeita ou aceita a hipótese nula, no qual a cadeia é ou não estacionária;

III. Caso a hipótese nula seja rejeitada, os primeiros 10% da cadeia será descartado e calculado de novo a estatística do teste;

IV. Caso a hipótese nula seja rejeitada, rejeita-se os próximos 10% e calcula a estatística do teste;

V. O processo será repetido até forem descartadas 50% das observações ou quando a hipótese nula for aceita. Mais, se o teste continuar rejeitando a hipótese nula, tem-se que aumentar o tamanho da iteração.

Segunda parte do processo:

I. Depois de executado a primeira etapa do diagnóstico será tomado a parte que não foi descartada para iniciar o teste da segunda parte;

II. Agora o teste de Halfwidth calcula a metade da largura de $(1-\alpha)\%$, que é o intervalo de credibilidade em torno do valor médio;

III. Tomando um ε arbitrário, verifica se a relação entre o teste de Halfwidth e a média calculada é inferior ao ε. Caso isso aconteça a cadeia passa no teste, caso contrário, deve-se aumentar o tamanho da iteração.

5.4.4 Diagnóstico Gelman & Rubin

Segundo os fundadores, o diagnóstico de Gelman & Rubin pressupõe que m cadeias serão geradas em paralelo, iniciando diferentes valores, tendo $2n$ iterações, no qual n iterações serão descartadas chamando assim de aquecimento ou "burni-in". Por isso, as m sequências renderá m possíveis inferências, e logo, em seguida tem-se a convergência estocástica. Abaixo será mostrado os 5 (cinco) passos do algoritmo para cada parâmetro.

I. Serão executados $m > 1$ cadeias de $2n$ de comprimento;

II. Rejeitam-se os primeiros n empates em cada cadeia;

III. Serão calculados as variâncias de dentro e entre as cadeias;

IV. Calculará a variância das estimativas do parâmetro de dentro e entre as cadeias;

V. Por último, calculará o fator de redução da potência escalar.

Então, depois de realizado o processo do algoritmo pode-se estimar as variâncias de dentro das cadeias. Logo, tem-se a média representada por V e a variância por S_i^2.

$$V = \frac{1}{m}\sum_{i=1}^{m} S_i^2, \quad \text{no qual,} \quad S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n-1}\left(\theta_{ij} - \bar{\theta}_i\right)^2.$$

Para o caso do cálculo das estimativas do parâmetro de entre as cadeias tem-se:

$$B = \frac{n}{m-1}\sum_{i=1}^{m-1}\left(\bar{\theta}_i - \bar{\bar{\theta}}\right)^2, \qquad \text{em que,} \qquad \bar{\bar{\theta}} = \frac{1}{m}\sum_{i=1}^{m}\bar{\theta}_i.$$

Acima citou-se a variância da cadeia e cada cadeia é baseado em n empates. Depois de estimado a variância entre e dentro das cadeias pode-se estimar a variância da distribuição de interesse da média ponderada de V e B.

$$\widehat{\text{Var}}(\theta) = n^{-1}[(n-1)V + B]. \tag{5.4}$$

Uma observação a fazer é que devido a superdispersão dos valores inicias esse resultado (5.4) superestima a variância populacional, porém é não tendencioso se a distribuição a *priori* for igual a distribuição estacionária. Também, pode-se encontrar o fator potencial de redução da escala por meio da seguinte expressão,

$$\widehat{\text{R}} = \sqrt{\frac{\widehat{\text{Var}}(\theta)}{V}}.$$

Algumas observações a fazer:

I. Caso o $\widehat{\text{R}}$ seja alto, maior que 1,1 ou 1,2, então, deve-se aumentar o número de execução das cadeias para melhor obter a convergência;

II. Caso tenha mais de um parâmetro, recomenda-se calcular o fator de redução potencial de escala para cada parâmetro;

III. O processo será realizado por tempo suficiente para que todos os $\widehat{\text{R}}$ sejam satisfeito, isto é, menor que 1,1 ou 1,2.

5.4.5 Gráfico da densidade de Kernel

Além do diagnóstico de convergência é interessante fazer o gráfico da densidade a *posteriori* utilizando o estimador de Kernel[2] para saber se existe ou não convergência, visualmente. Da mesma forma, através desse gráfico pode-se verificar quantas observações devem ser descartadas para trabalhar-se com os valores que estão em torno da média, isto é, tomar a fatia da parte que é convergente e fazer inferências estatísticas. Deste modo, na Figura 9 estão os gráficos ilustrativo de iteração e densidade utilizando o estimador de Kernel.

[2]O estimador de Kernel é dada pela seguinte função $EK(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$; no qual o n é o tamanho amostral; o h é o parâmetro de escala e o K é uma função de probabilidade simétrica, como por exemplo a distribuição normal (veja Silverman,1984-185; Härdle, 1990 e Souza, 2008).

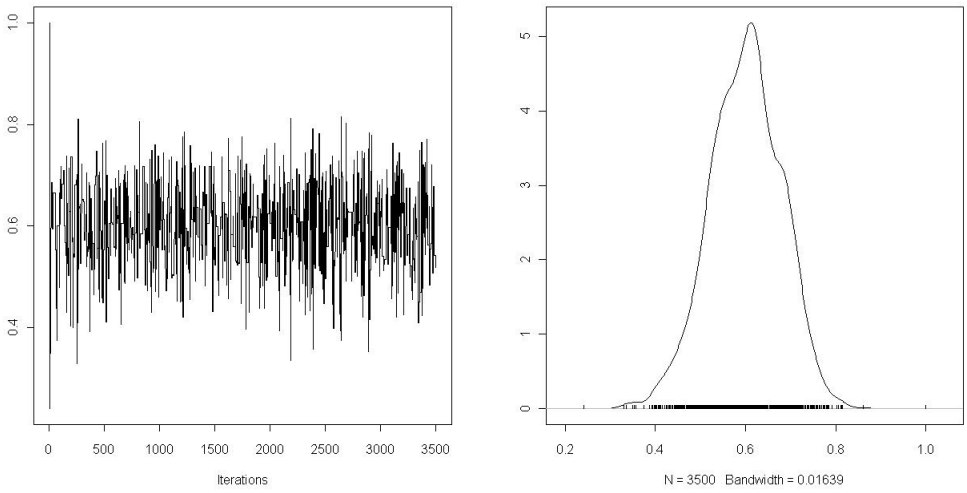

Figura 9: Gráfico ilustrativo da iteração e densidade de convergência.

5.4.6 Representação da autocorrelação

Na autocorrelação pode-se avaliar graficamente ou por meio do teste que calcula a função de autocorrelação da cadeia de Markov com defasagens “lags” de 0, 1, 5, 10, 20 e 50. Em geral, a FAC proporciona informação sobre a estrutura de correlação ao longo do tempo. Nessa dissertação também será focado o estudo de autocorrelação graficamente, pois é possível tirar boa parte da conclusão a respeito da convergência, como mostra a Figura 10.

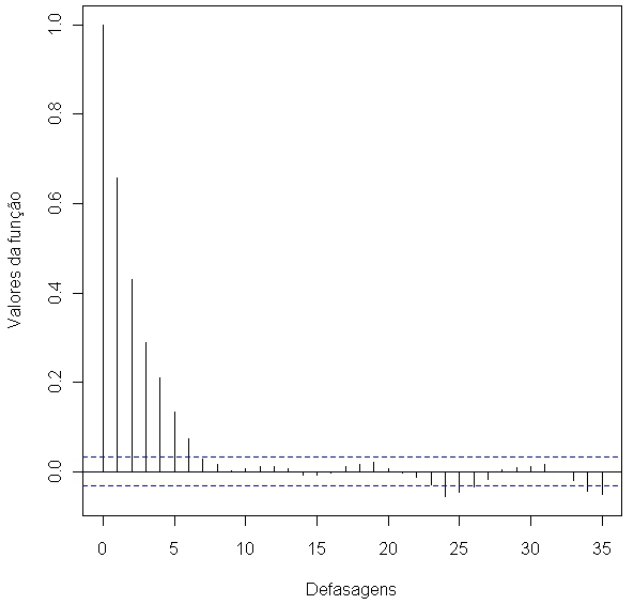

Figura 10: Gráfico ilustrativo de autocorrelação.

5.4.7 Descrição do Histograma

O histograma é a representação gráfica de uma distribuição de frequências através de retângulos justapostos, cujas as áreas são proporcionais ás frequências das classes, em outra linguagem, tem-se que dividir o intervalo de variação dos dados em subintervalos de comprimento h (na linguagem inglesa, denominado de “bins”), veja a Fig. 11. Também, é o método não-paramétrico utilizado para estimar as densidades, contudo não é rigoroso e a aplicabilidade é complicada quando está no caso multivariado, Silverman (1986).

Então, considera-se uma variável aleatória discreta X, no qual x assumirá alguns valores dessa variável, e tem-se que estimar o $\mathrm{HI}_h(x)$ a partir das observações $x_i,\ i = 1,2,\ldots,n$. Portanto, o histograma é definido por

$$\mathrm{HI}_h(x) = \frac{1}{nh} \times (\text{os valores de } x_1, x_2, \ldots, x_n \text{ iguais a } x).$$

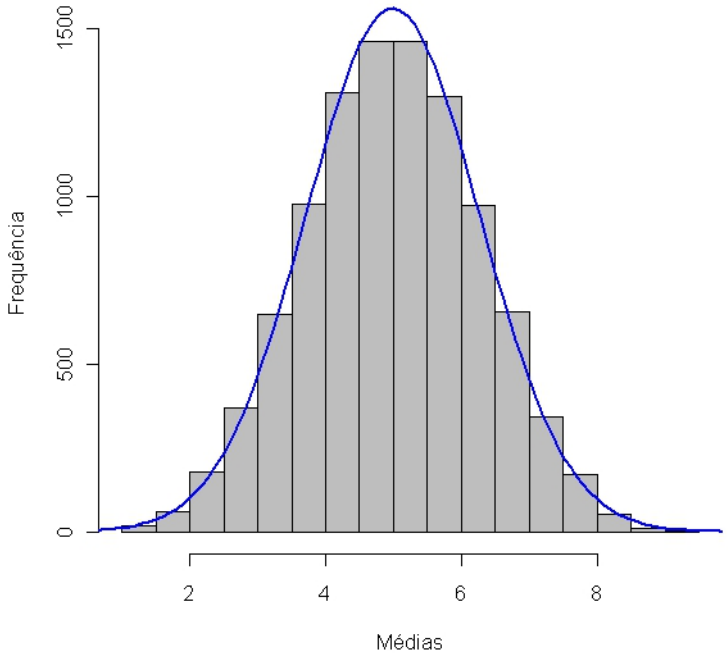

Figura 11: Gráfico ilustrativo do histograma.

6 Modelagem da produção de soja

6.1 Soja

Há cerca de cinco mil anos os chineses tem a soja como seu principal legume doméstico. A soja selvagem é denominada como a espécie mais antiga, que crescia nas terras baixas e úmidas conjuntamente com os juncos entre as proximidades dos rios e lagos da China Central. Segundo algumas evidências históricas e geográficas a soja surgiu no século XI a.C. no norte da china, vale do rio Amarelo, que é a origem da civilização chinesa. A primeira citação sobre a soja foi através do imperador Sheng Nung por meio do seu livro sobre a medicina, publicado nos anos dois mil a.C.


Figura 12: Foto ilustrativa da soja.

A produção de soja iniciou-se no norte da China e depois passou para o sul da China, Coréia, Japão e sudeste da Ásia. Contudo, devido a fraqueza da China na agricultura, a soja chegou a Coréia e Japão entre os anos 200 a.C. e século III d.C. Com a chegada dos navios europeus no ocidente, entre os século XV e XVI, surgiu a soja nas Américas. Mas, ficou apenas como curiosidade botânica até o início do século XIX. No final do século XX a soja passou a ser cultivada comercialmente nos Estados Unidos da América. Logo em seguida, houve um rápido crescimento na produção, com o desenvolvimento das primeiras cultivas comerciais.


Figura 13: Foto ilustrativa da soja.

A soja surgiu no Brasil em 1882 no estado da Bahia. Em seguida, foi para São Paulo por meio dos imigrantes japoneses e consequentemente no Rio Grande do Sul por volta de 1914, nesse estado a soja obteve um grande crescimento na produtividade. Em 1930, já considerava essa região como produtora de soja no Brasil (veja Trucom, 2009 e Santos *et al.*, 1997).

Nos primórdios da história da soja no Brasil, relata-se que esse alimento era direcionado aos suínos, como complemento alimentar da dieta à base de milho, abóbora e mandioca. Ou seja, foi como fonte de nitrogênio em adubação verde. Por volta de 1950, o presidente investiu no cultivo de soja e trigo, iniciando-se a dobradinha trigo-soja, quando teve uma vasta produção no cultivo da soja fazendo rodízio com o trigo.

A soja é um alimento que pertence a família das leguminosas, mesma família do feijão, da lentilha, da ervilha e do grão-de-bico. Além disso, o seu grande destaque na natureza é a riqueza em proteínas, lipídios, fibras, sais minerais e rico em vitaminas do complexo B. A soja é do gênero *Glycine* da família *Papilionoideae*, medindo em torno de 0,6 a 1,5 metro de altura, sendo herbácea e anual. O legume é uma vagem achatada, curta, de cor amarelo-palha, preta ou cinzenta, que fornece de duas a cinco sementes (vide Fig. 12 e 13).

Depois da criação da empresa EMBRAPA,[1] que realiza frequentemente estudos científicos, aliados a muitas pesquisas internacionais e à insistência da população brasileira, a soja está voltando ao menu do brasileiro, de forma diferente, pois a EMBRAPA desenvolveu, desde de 1987, um programa de incentivo ao consumo de soja através da elaboração e divulgação de novos conceitos e receitas em livro, site, cursos e palestras, visando a disseminação de técnicas adequadas de preparo do alimento.

O segundo maior produtor mundial de soja é o Brasil seguido apenas dos Estados Unidos. Entre os anos de 2009 e 2010, a plantação ocupou cerca de 23,6 milhões de hectares, o que totalizou uma produção de 68,7 milhões de toneladas e os Estados Unidos correspondeu a 91,4 milhões de toneladas de soja de grão. O maior estado da produção de soja no Brasil é o Mato Grosso chegando a produzir 3.036 Kg/ha, enquanto o Brasil tem em média 2.941 kg/ha.[1]

6.2 Material do estudo

Para ilustrar a aplicação da técnica de equações simultâneas por meio do método de mínimos quadrados em dois estágios no contexto clássico e bayesiano, utilizou-se o banco de dados da produção de soja em grão, no Brasil, dados em anos de 1994 a 2009, como mostra as Figuras ilustrativas da produção de soja. As variáveis estudadas foram classificadas como: resposta (quantidade produzida e valor da produção) e explicativas (área plantada, área colhida e produto interno bruto). As variáveis área plantada e área colhida foram medidas por hectares, quantidade produzida por tonelada e as variáveis valor da produção em reais $(1.000\,\mathbf{R\$})$ e pro-

[1] Para mais detalhes sobre a produção de soja, veja o sitio da Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA, vide sitio <http://www.cnpso.embrapa.br>. Também, pode consultar os artigos do Hungria *et all.*, 2006 e Gonçalvez *et all.*, 2006.

duto interno bruto em milhões.[2] As observações do banco de dados foram dividido por 10^6 para facilitar a visualização dos resultados. Para comparação dos testes estatísticos e análise de diagnóstico de convergência será adotado em todos os resultados 90% de confiança.

6.3 Análise descritiva dos dados

Nesta seção, será analisado o comportamento das variáveis área plantada, área colhida, quantidade produzida, valor da produção e produto interno bruto, através de algumas medidas descritivas e gráficos. A seguir, segue um sumário das variáveis envolvidas no estudo.

Tabela 3: Principais medidas descritivas das variáveis de interesse.

Variável	Sigla	Mínimo	Máximo	Amplitude	Média	Mediana	Moda	DP	CV(%)
Área plantada	AP	10,3561	23,4268	13,0706	16,5490	15,1822	12,4486	4,6361	28,0147
Área colhida	AC	10,2994	22,9489	12,6494	16,4984	15,1722	12,5198	4,5939	27,8442
Quantidade produzida	QP	23,1669	59,8331	36,6662	40,9660	40,0074	38,0902	13,2402	32,3199
Valor da produção	VP	3,5388	39,0772	35,5384	17,0997	14,1061	8,1189	12,4710	72,2931
Produto interno bruto	PIB	0,3092	2,7407	2,4315	1,3955	1,1959	0,7966	0,7284	52,1982

Na Tabela 3 obteve-se as principais estatísticas descritivas das variáveis em estudo e percebe-se que as variáveis AP e AC tem valores semelhantes, consequentemente as dispersões (Desvio Padrão$-$DP) são parecidas. Também, a média é maior do que a mediana e a moda de Pearson em todas as variáveis, então, os dados estão concentrados a esquerda da média amostral. Do mesmo modo, tem-se o CV (Coeficiente de Variação de Pearson), em percentual, onde AP, AC e QP são homogêneas. Portanto, a maior concentração das observações das variáveis AP, AC e QP estão a esquerda e próxima da média amostral.

A seguir, apresenta-se na Tabela 4 a matriz de correlação, dois a dois, em que todas as variáveis são correlacionadas positivamente. Aplicando-se o teste de correlação, com 90% de confiança, verifica-se que, de fato, essas variáveis são correlacionadas.

Tabela 4: Correlações entre as variáveis

	AP	AC	QP	VP	PIB
AP	1,0000	0,9997	0,9512	0,8739	0,8989
AC	·	1,0000	0,9559	0,8802	0,9040
QP	·	·	1,0000	0,9317	0,9486
VP	·	·	·	1,0000	0,9097
VE	·	·	·	·	1,0000

No diagrama de dispersão, Figura 14, verifica-se que há uma relação linear entre área plantada e área colhida, com uma inclinação positiva forte; a variável área colhida com QP, VP

[2] O banco de dados utilizado, encontra-se disponível no sitio do IBGE <http://www.ibge.gov.br>, IPEA <http://www.ipea.gov.br> e no Apêndice B.

e PIB, sendo dois a dois, exibe um padrão aproximadamente linear, não tão intenso quanto apresentado entre AP e AC.

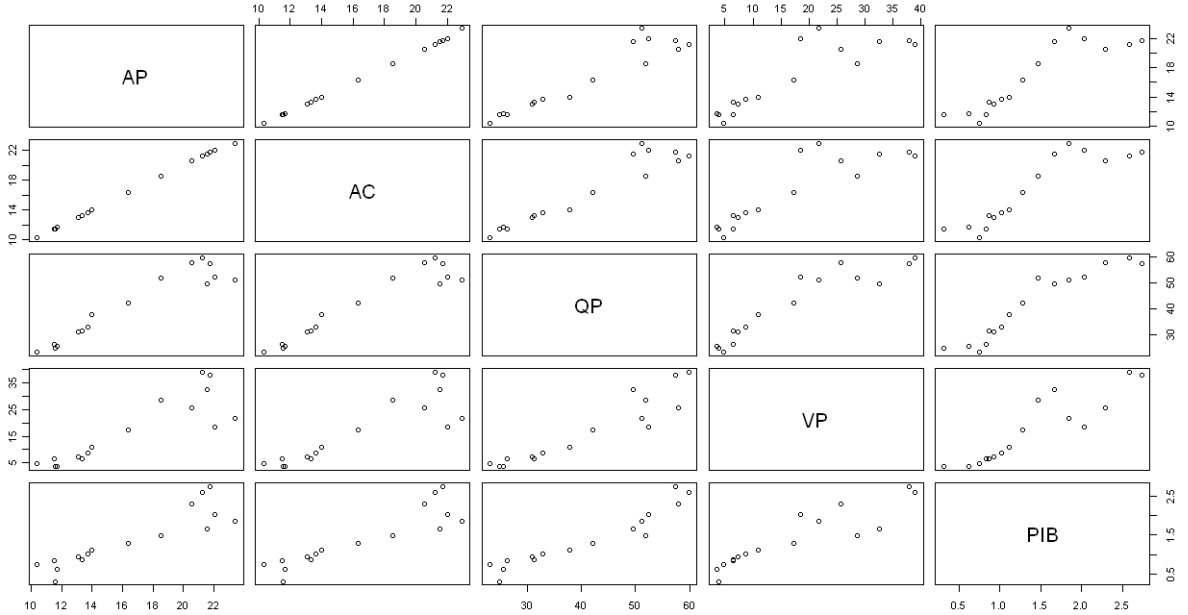

Figura 14: Diagrama de dispersão.

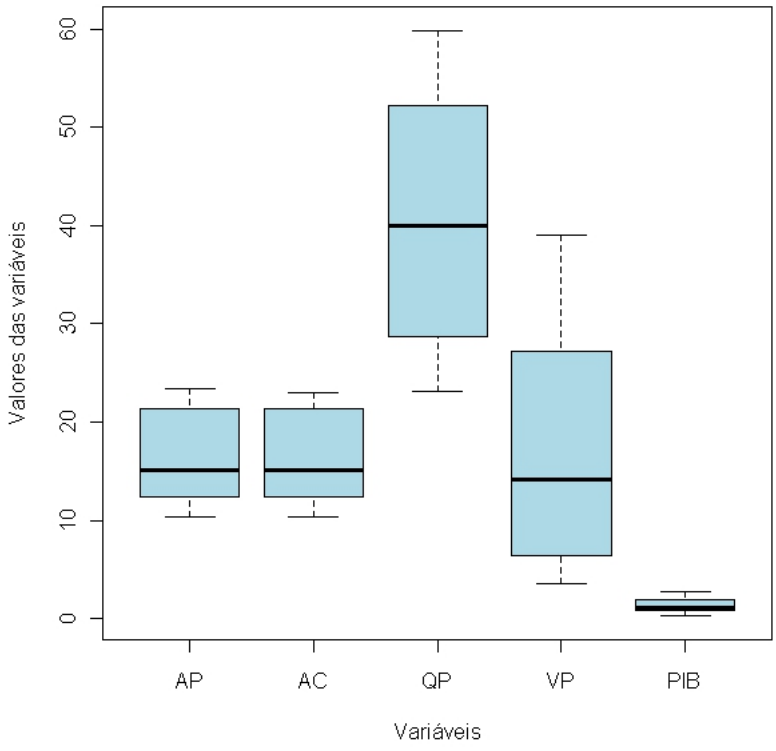

Figura 15: *box-plot* das variáveis AP, AC, QP, VP e PIB.

Depois, de realizado as estatísticas descritivas, a correlação das variáveis, o teste de correlação, o gráfico de dispersão (na Fig. 14) e o gráfico *box-plot* (na Fig. 15), sugere-se trabalhar na primeira modelagem as variáveis explanatórias AC e PIB e para segunda modelagem as estimativas do primeiro modelo como variável independente. Em seguida, fez-se o critério de escolha da variável resposta QP para compor o primeiro modelo e a variável VP para o segundo modelo, pois a variável VP está sobreidentificada a variável quantidade produzida.

6.4 Ajuste e análise do modelo: clássico

Nesta seção, será escolhido o método para compor o modelo final utilizando a técnica de equações simultâneas. Inicialmente, aplicou-se o método de mínimos quadrados em dois estágios. Em seguida, foi gerado o primeiro modelo com a variável independente área colhida & produto interno bruto e a variável dependente quantidade da produção. Depois, que se estimou o primeiro modelo gerou-se o segundo modelo com a variável dependente valor da produção e a variável independente as estimativas do primeiro modelo, lembrando que os dois modelos foram gerados pela técnica de mínimos quadrados ordinários.

- Modelo I: $\mathrm{QP} = \beta_0 + \beta_1 \mathrm{AC} + \beta_2 \mathrm{PIB} + \varepsilon$.

Tabela 5: Estimativas clássicas dos parâmetros do primeiro modelo.

	Estimativa	DP	Estatística do teste-t	p-valor
Intercepto	3,6500	3,8714	0,9430	0,3630
AC	1,5517	0,4064	3,8180	0,0021
PIB	8,3953	2,5629	3,2760	0,0060

No primeiro modelo o p-valor das variáveis AC e PIB são significantes (Tab. 5) e o coeficiente de determinação de Pearson ajustado é $\overline{R}^2 = 0,9455$. Para a segunda modelagem estimou-se esse modelo e ajustou-se com a variável resposta valor da produção.

- Modelo II: $\mathrm{VP} = \beta_0 + \beta_1 \widehat{\mathrm{QP}} + \varepsilon$.

Já no segundo modelo o p-valor da variável independente é significante (Tab. 6) e o coeficiente de determinação de Pearson ajustado é $\overline{R}^2 = 0,8274$. Depois, que foram realizadas as estimativas do primeiro e segundo modelo de regressão de MQ2E verifica-se que realmente as variáveis VP e QP são simultâneas, em outras palavras, as variáveis VP e QP são mutuamente dependentes. Para observar esse resultado utilizou-se o teste de Hausman. No qual, fez-se a

Tabela 6: Estimativas clássicas dos parâmetros do segundo modelo.

	Estimativa	DP	Estatística do teste-t	p-valor
Intercepto	$-19{,}1079$	4,4337	$-4{,}3100$	0,0007
$\widehat{\text{QP}}$	0,8838	0,1035	8,5390	$6,3603 \cdot 10^{-7}$

regressão do primeiro modelo, por essa regressão, obteve-se as estimativas e os valores residuais. Logo, em seguida, formou-se a regressão da QP sobre o VP estimado e os resíduos (u) para obter os seguintes resultados:

Tabela 7: Estimativas clássicas dos parâmetros para verificação do teste de Hausman.

	Estimativa	DP	Estatística do teste-t	p-valor
Intercepto	$-19{,}1079$	4,1494	$-4{,}6050$	0,0005
$\widehat{\text{QP}}$	0,8839	0,0969	9,1240	$5,1565 \cdot 10^{-7}$
$\widehat{\text{u}}$	0,7515	0,4350	1,7270	$0,1077$

Tendo em vista que a estatística do teste-t do resíduo é estatisticamente significante (vide Tabela 7), não se pode rejeitar a hipótese nula do teste de simultaneidade entre as variáveis dependentes. Em vista disso, as variáveis quantidade produzida e valor da produção são simultâneas, isto é, mutuamente dependentes e consequentemente pode-se utilizar o MQ2E ao invés de empregar o MQO.

6.5 Ajuste e análise de modelo: bayesiano

Após ter ajustado o modelo de regressão clássico, fez-se a aplicação do método de Monte Carlo via Cadeia de Markov por meio dos algoritmos de *Gibbs* e *Metropolis-Hastings* utilizando a técnica de MQ2E de equações simultâneas. O estudo foi realizado com as *prioris* Normal e Gama, como os resultado foram semelhantes, permaneceu-se com a *priori* Normal de média 0,0 e desvio padrão 0,01; obtendo como precisão uma distribuição gama de parâmetros de escala e de forma 0,1. Neste tópico será mostrado os resultados do método de *Gibbs*, e do método de *Metropolis-Hastings* encontra-se no Apêndice A.

Pode-se observar na Tabela 8 o resumo da *posteriori* do primeiro modelo utilizando o algoritmo de *Gibbs* com suas respectivas estimativas, desvios padrão, estatística do teste-t e p-valor. Diante disso, destacam-se os p-valores que são significantes.

Tabela 8: Estimativas bayesianas dos parâmetros do primeiro modelo.

	Estimativa	DP	Estatística do teste-t	p-valor
Intercepto	2,7663	3,7340	0,7409	0,4710
AC	1,6624	0,3940	4,2191	0,0008
PIB	7,7053	2,5631	3,0062	0,0094

Após o aquecimento (descartes dos dados iniciais que influência a divergência) foram construídos os gráficos para observar a existência da convergência[3]. Então, analisando o histograma verifica-se que os dados seguem aproximadamente uma distribuição normal (ver Figura 16); na Figura 17 os dados estão em torno da média demonstrando estacionariedade; na Figura 18 a densidade segue aproximadamente uma distribuição normal e pelo gráfico das correlações percebe-se que existe um caimento nos primeiros *lag's*, isto é, os dados estão estacionários (Fig. 36). Para complementar os resultados foram construídos os gráficos de diagnóstico de convergência de Geweke (Fig. 19), da acumulada (Fig. 20) e Gelman & Rubin (Fig. 22), no qual verifica-se que esses gráficos permaneceram no padrão recomendado, em outras palavras, o algoritmo obteve convergência. Diante desses gráficos chama-se atenção para o de Gelman & Rubin (Fig. 22), no qual esse gráfico mostrou-se que do início ao fim da cadeia possuem valores próximos da média mostrando que, de fato, convergiu.

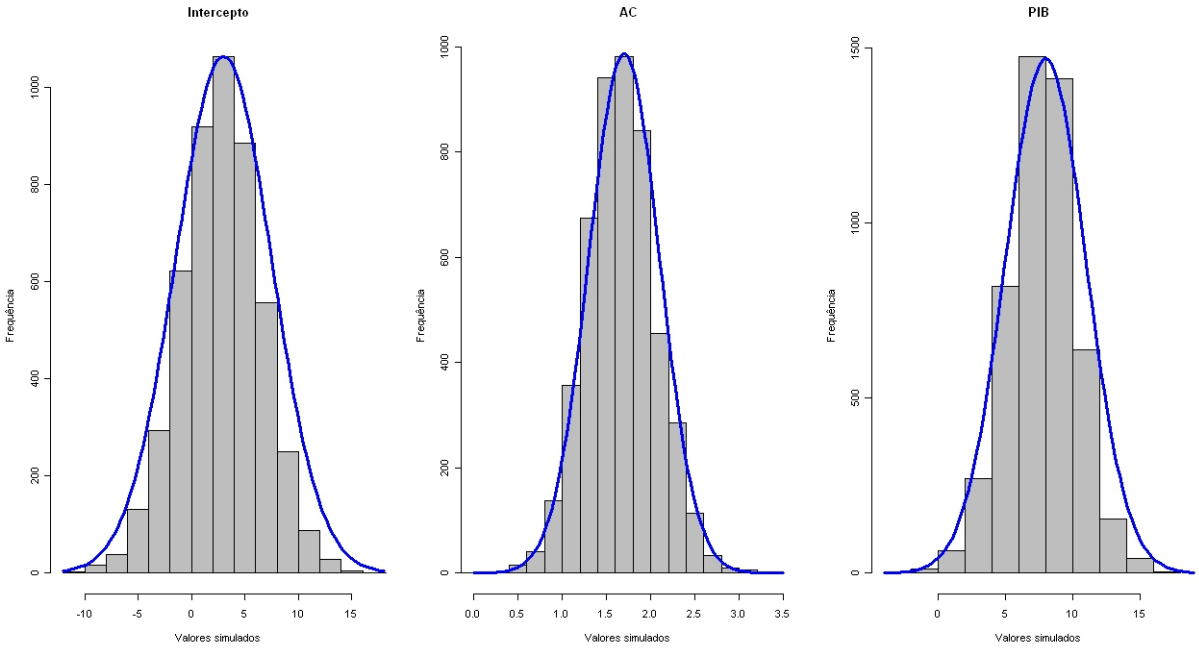

Figura 16: Histograma dos dados após o aquecimento do método de *Gibbs* do IM.

[3]Tanto no algoritmo de *Gibbs como no algoritmo de Metropolis-Hastings*.

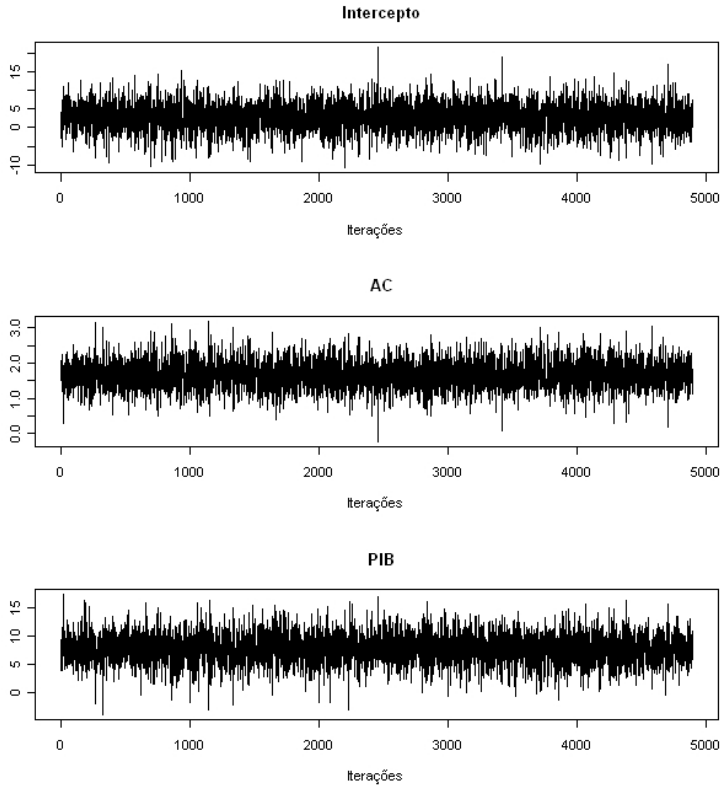

Figura 17: Visualização gráfica da convergência das iterações do método de *Gibbs* do IM.

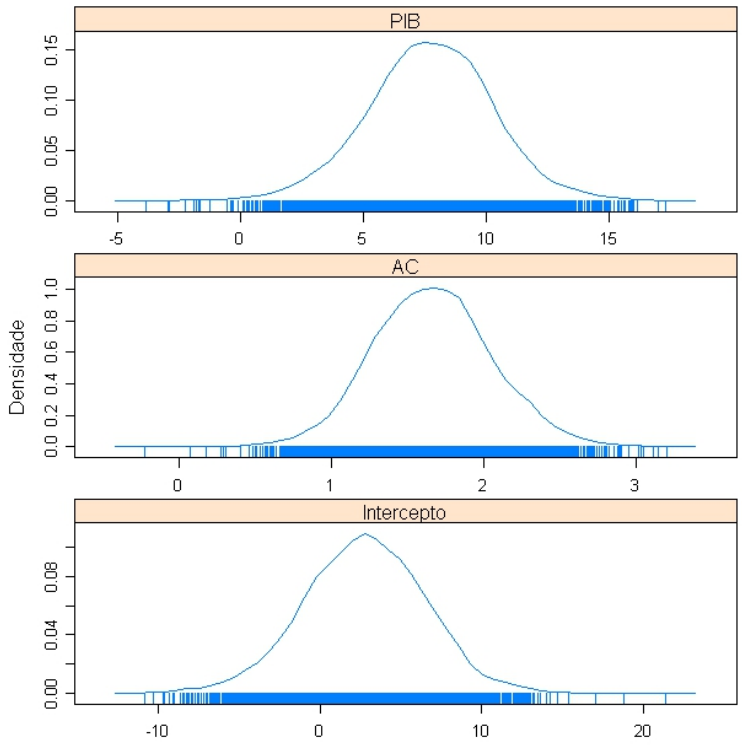

Figura 18: Gráfico da densidade de convergência do método de *Gibbs* do IM.

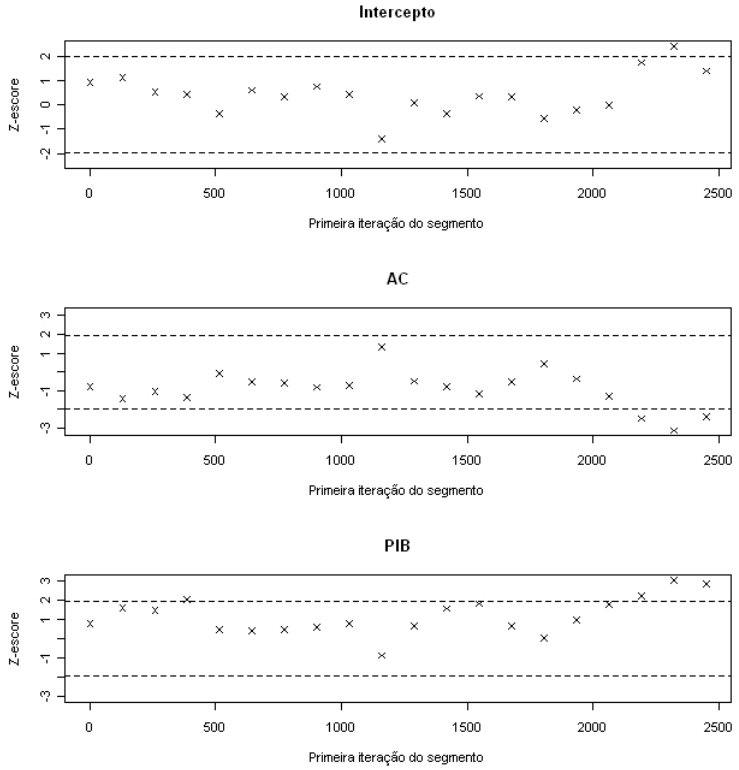

Figura 19: Diagnóstico de convergência de Geweke do algoritmo de *Gibss* do IM.

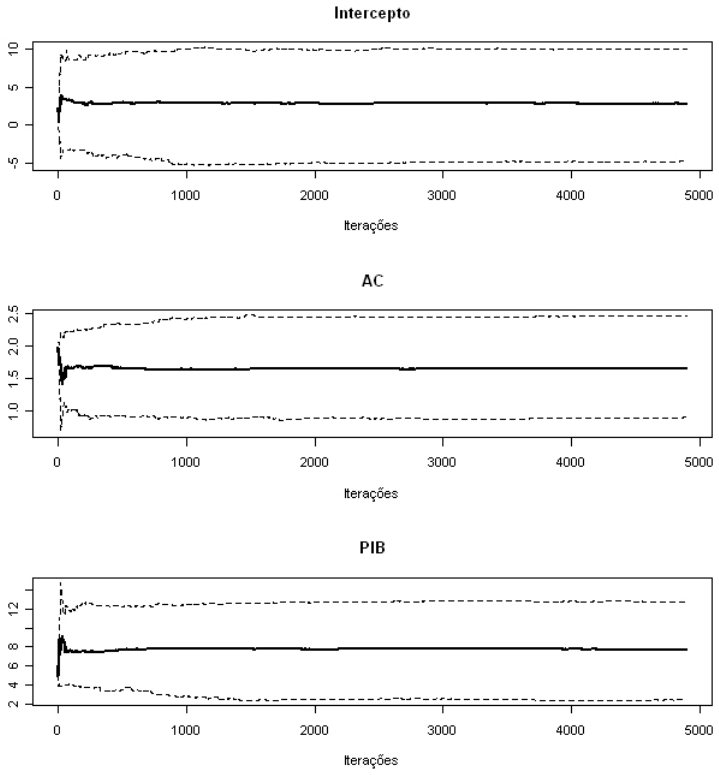

Figura 20: Gráfico da convergência acumulada do algoritmo de *Gibbs* do IM.

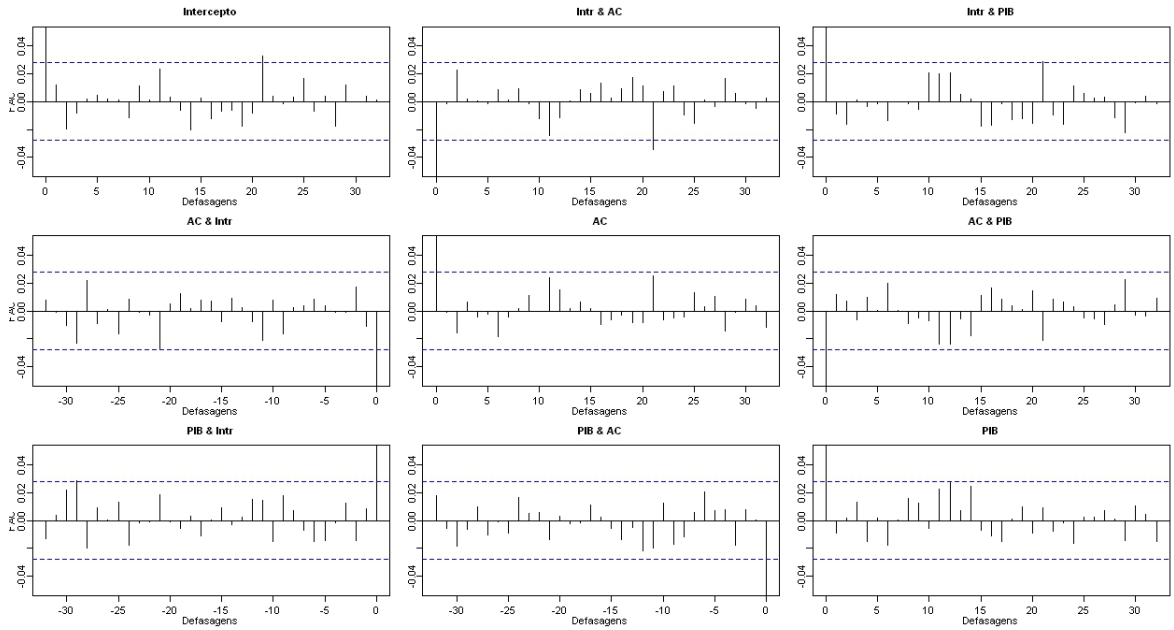

Figura 21: Gráfico das correlações do algoritmo de *Gibbs* do IM.

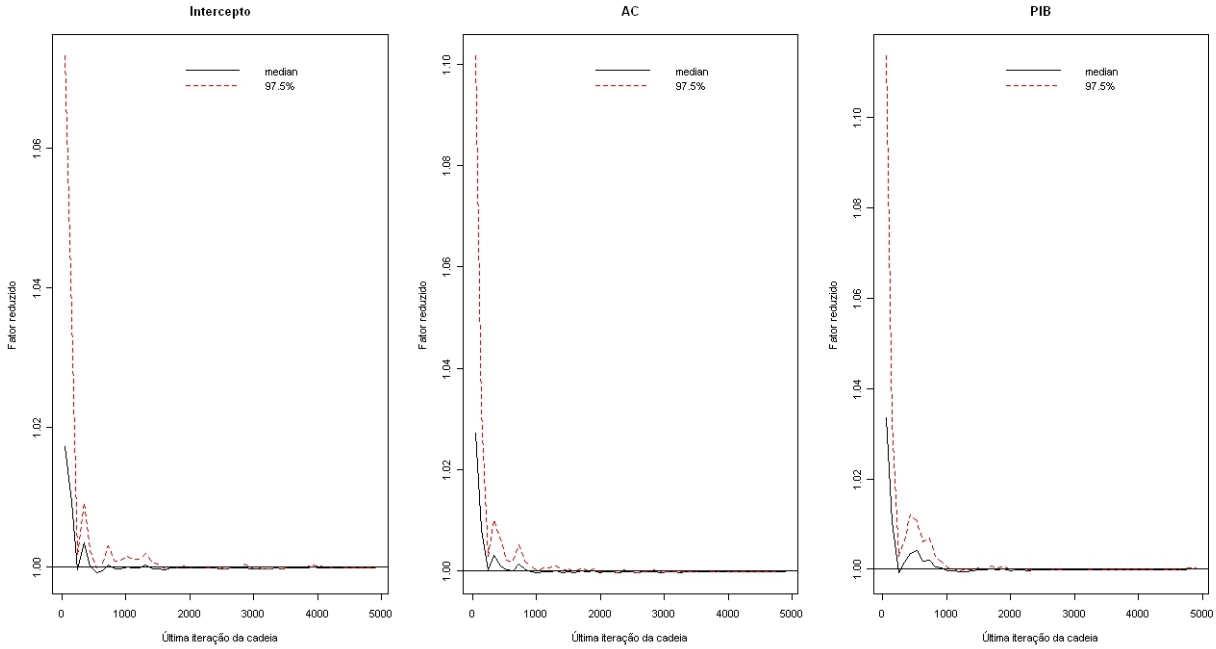

Figura 22: Visualização do diagnóstico de Gelman & Rubin do algoritmo de *Gibbs* do IM.

Depois de encontrar e avaliar as estimativas do primeiro modelo, fez-se o segundo modelo com a variável dependente VP e a variável explanatória *as estimativas do primeiro modelo*. Assim, tem-se abaixo a Tabela 9 que mostra o resumo da *posteriori* do segundo modelo, com média, desvio padrão, estatística do teste-t e o p-valor. Dentre esses resultados, destacam-se os desvios padrão que foram semelhantes ao clássico e os p-valores que são significantes.

Tabela 9: Estimativas bayesianas dos parâmetros do segundo modelo.

	Estimativa	DP	Estatística do teste-t	p-valor
Intercepto	$-15{,}4159$	4,4369	$-3{,}4745$	0,0037
$\widehat{\text{QP}}$	0,8007	0,1036	7,7305	$2{,}0360 \cdot 10^{-6}$

Então, após o aquecimento da segunda modelagem foram construídos os gráficos para verificar a existência da convergência. Logo, pelo histograma observa-se que os dados seguem aproximadamente uma distribuição normal (ver Figura 23). Na Figura 24 observa-se que os dados estão em torno da média demonstrando estacionariedade e consequentemente na Figura 25 observa-se que os dados seguem aproximadamente uma distribuição normal.

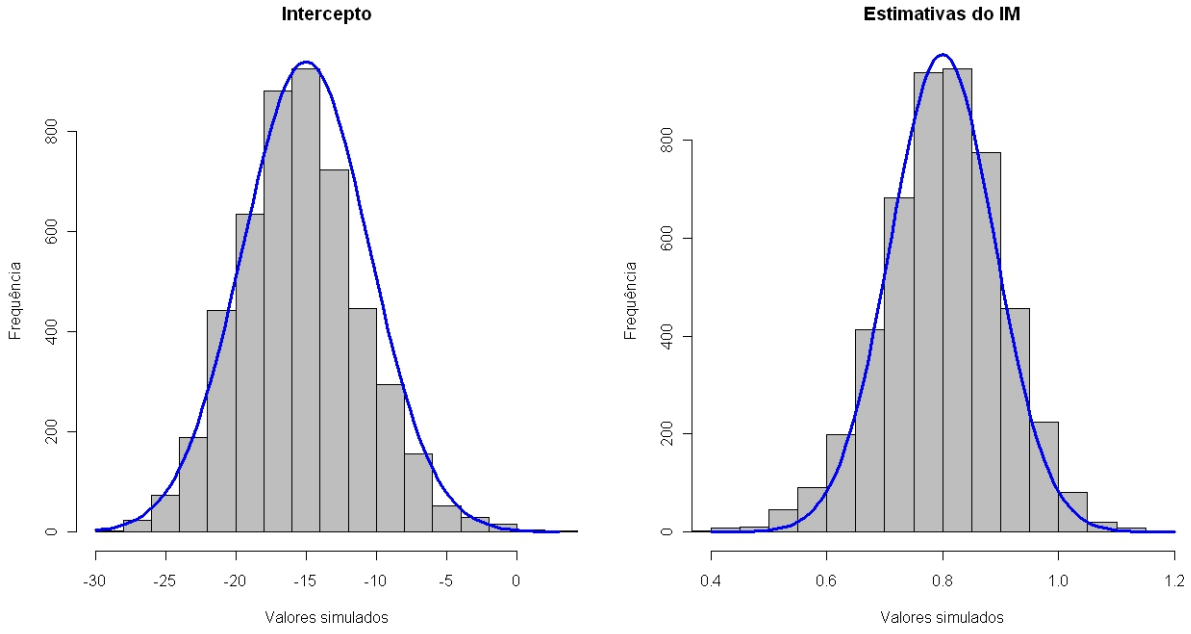

Figura 23: Histograma dos dados após o aquecimento do método de *Gibbs* do IIM.

Após os gráficos "básicos" de convergência (Histograma e Kernel) construíram-se os gráficos de escore (Geweke), acumulada, correlações e do diagnóstico de Gelman & Rubin, que se encontram nas Figuras abaixo. O algoritmo de *Gibbs* mostrou-se convergente em todos os casos. Isto é, as observações ficaram dentro do intervalo de convergência do diagnóstico de Geweke, no primeiro e segundo modelo; nas correlações, tanto o primeiro como o segundo modelo, houve um caimento exponencial na primeira defasagem; por último, os gráficos da acumulada e de Gelman & Ruben, no qual permaneceram no padrão adequado (a parte gráfica e numérica do algoritmo de *Metropolis-Hastings* se encontram no Apêndice A).

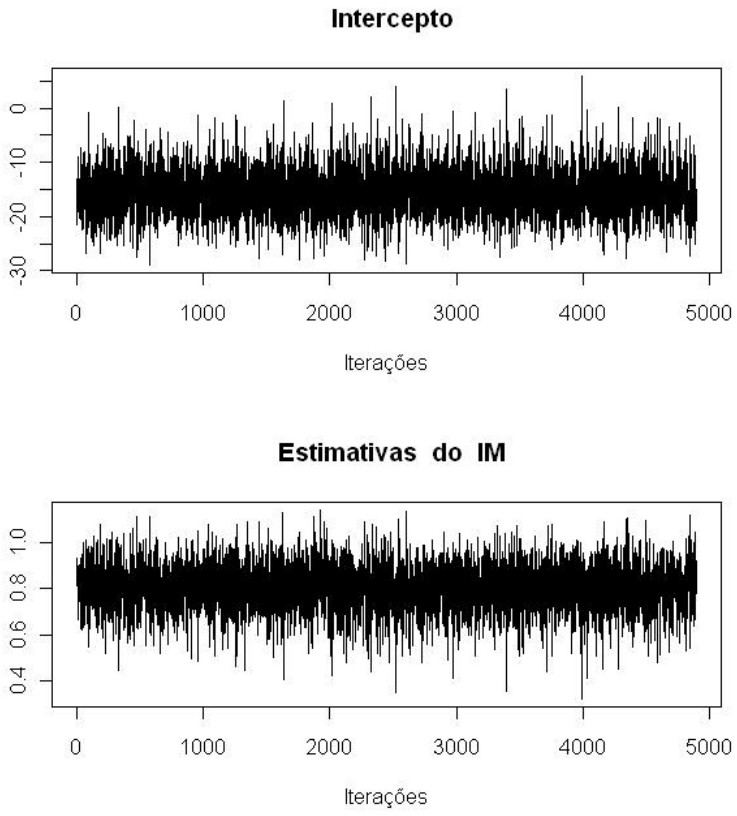

Figura 24: Visualização gráfica da convergência das iterações do método de *Gibbs* do IIM.

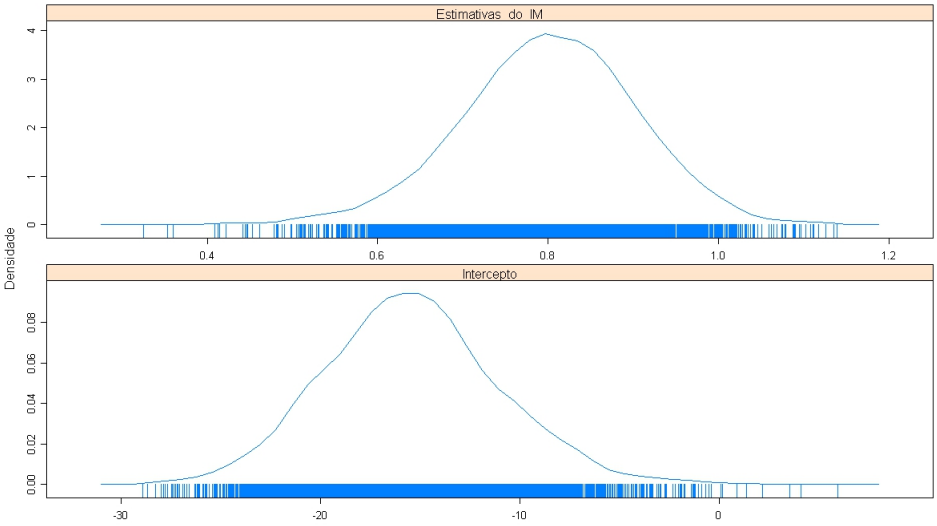

Figura 25: Gráfico da densidade de convergência do método de *Gibbs* do IIM.

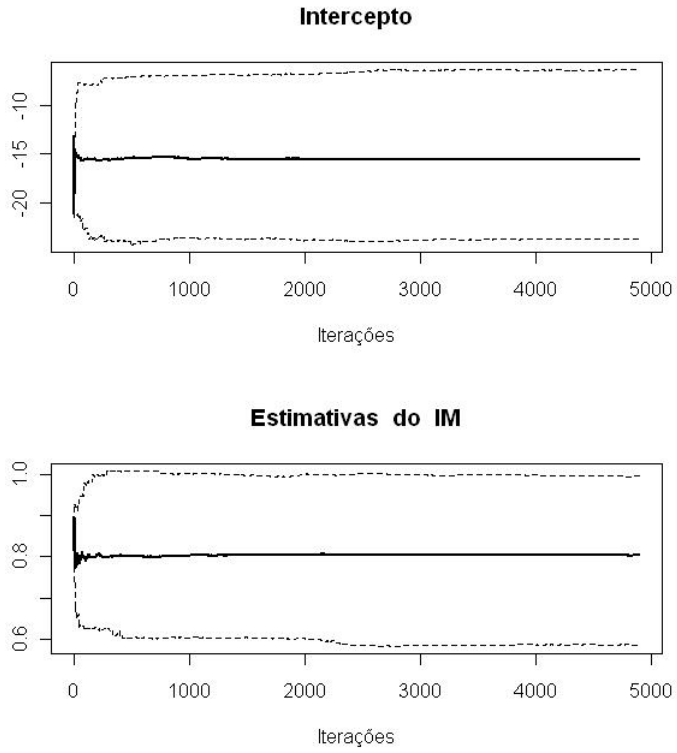

Figura 26: Gráfico da convergência acumulada do algoritmo de *Gibbs* do IIM.

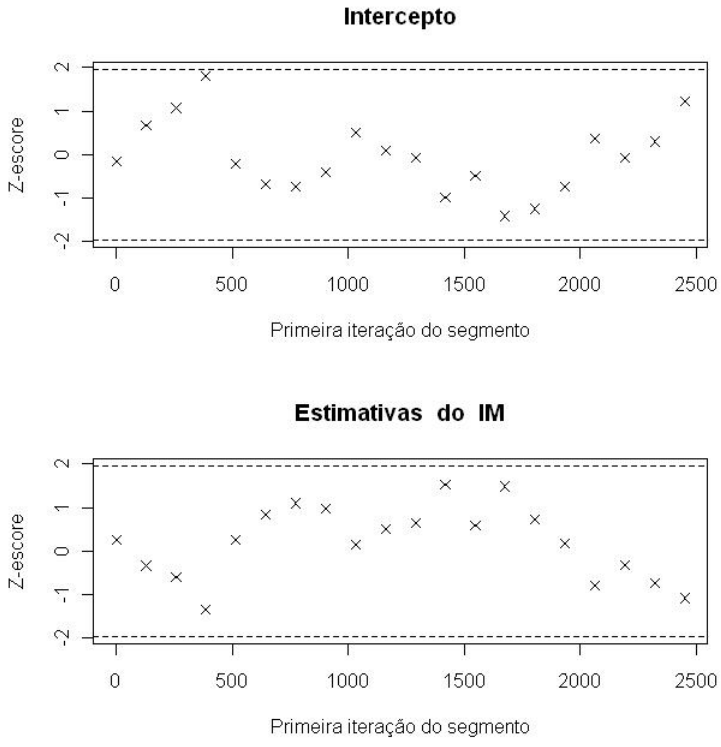

Figura 27: Diagnóstico de convergência de Geweke do algoritmo de *Gibss* do IIM.

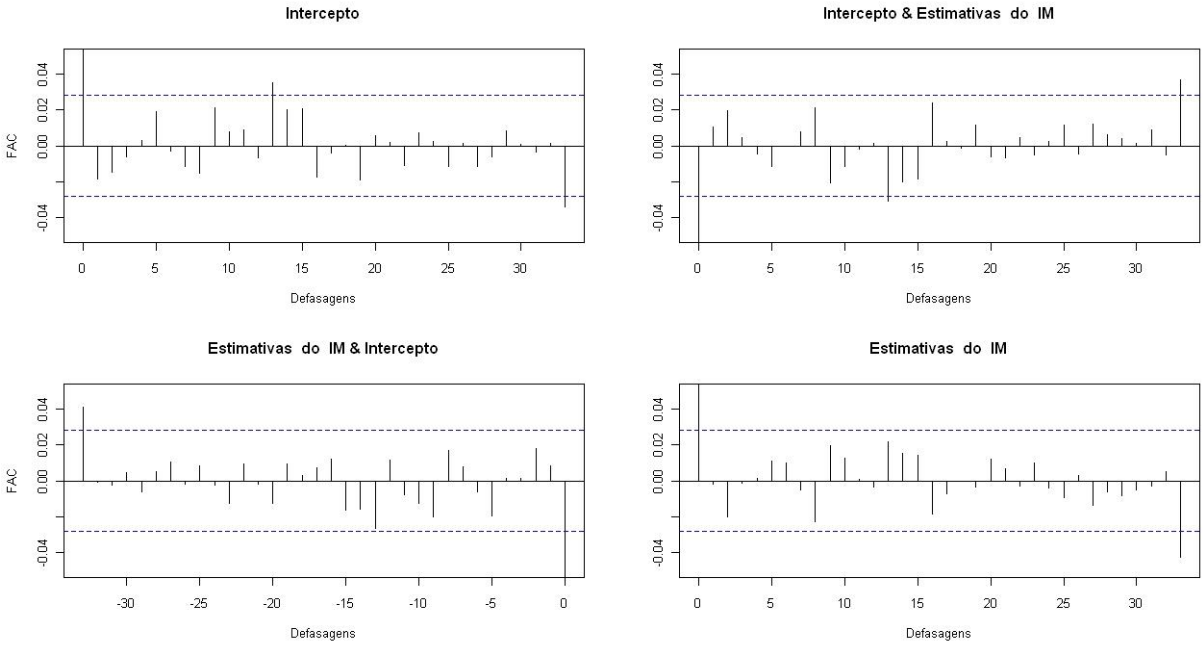

Figura 28: Gráfico das correlações do algoritmo de *Gibbs* do IIM.

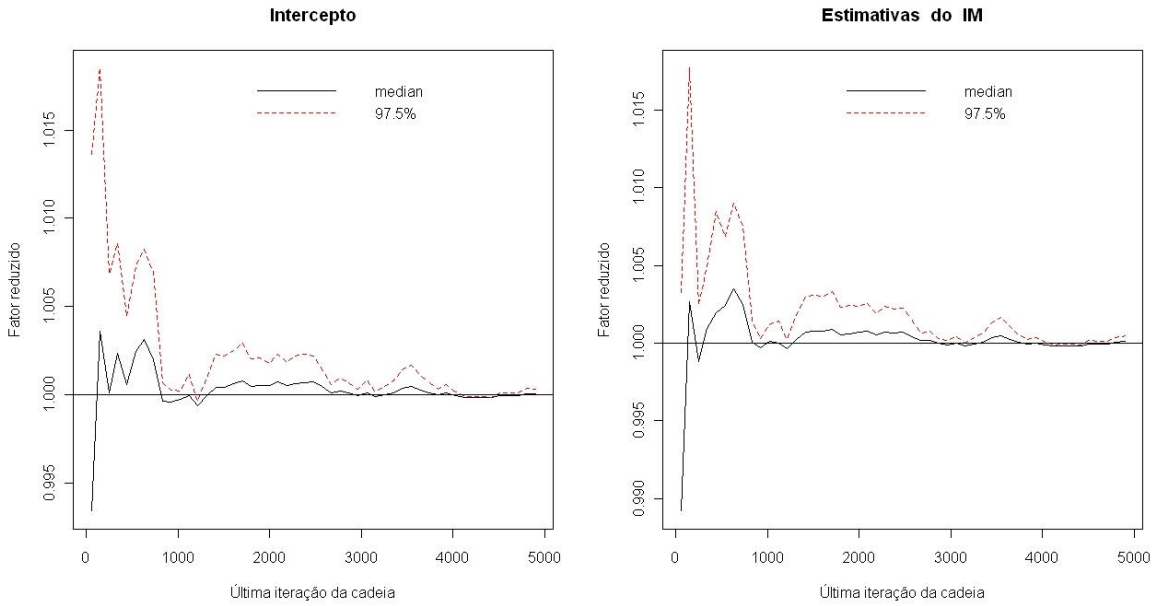

Figura 29: Visualização do diagnóstico de Gelman & Rubin do algoritmo de *Gibbs* do IIM.

Na Tabela 10, encontram-se os resultados do diagnóstico de convergência do método aplicado. Primeiramente, fez-se uma amostra piloto com 10.000 (dez mil) observações e por meio do diagnóstico de convergência de Raftery & Lewis foi sugerido uma amostra de no mínimo 2.638 para obter uma convergência, logo optou-se por trabalhar com a amostra de tamanho 5.000 (cinco mil) com uma queimação de tamanho 100, ou seja, descartando as cem primeiras

observações, e consequentemente obteve-se o fator de dependência−FI menor que 5 (cinco) nas duas modelagens. No teste de Heidelberger & Welch e Halfwidth, o algoritmo de *Gibbs* passou no teste de estacionariedade. Na aplicação do teste de Geweke observou-se que o algoritmo, nas duas modelagens, satisfez os níveis de convergência, pois a estatística do teste ficou dentro do intervalo $\pm 1,96$. Por último, aplicou-se o teste de Gelman & Rubin e calculou o fator de redução potencial de escala para cada parâmetro, com 5 cadeias, e todos os rendimentos foram satisfeitos, ou seja, o resultado dos parâmetros ficaram abaixo do recomendado. Da mesma forma, quando é observado os gráficos na Figura 29 de Gelman & Rubin verifica-se que a linha tracejada está próxima da linha contínua, portanto o algoritmo satisfaz o diagnóstico de convergência numericamente e graficamente.

Tabela 10: Sumário dos resultados do diagnóstico de convergência do modelo I e II.

			Raftery & Lewis	Heidelberger & Welch		Geweke	Gelman & Rubin		
Algoritmo		Modelo	(FI)	valor-p	Halfwidth		Ponto 95%	Quantil	Entre
Gibbs	I	Intercepto	1,0000	0,4687	0,1076	0,9119	1,0000	1,0000	1,00
		AC	1,0400	0,1423	0,0115	−0,8046	1,0000	1,0000	
		PIB	0,9710	0,0594	0,0743	0,7795	1,0000	1,0000	
	II	Intercepto	1,0200	0,9270	0,1171	−0,1510	1,0000	1,0000	1,00
		Estimativas do IM	1,0200	0,6920	0,0029	0,2615	1,0000	1,0000	

6.6 Correções dos desvios padrão dos modelos clássico e bayesiano

Para finalizar a parte prática da dissertação fez-se a correção dos desvios padrão, estatística do teste-t, p-valor e calcularam-se os critérios de informações de AIC_n e BIC_n normalizados dos modelos clássico e bayesiano, conforme exige a técnica de MQ2E de equações simultâneas. A seguir tem-se a tabela dos valores corrigidos e o gráfico dos valores observados *versus* valores preditos.

Tabela 11: Estimativas clássicas e bayesianas dos parâmetros dos modelos corrigido.

Método		Estimativa	DP	Estatística do teste-t	p-valor	AIC_n	BIC_n
Clássico	Intercepto	−19,1079	4,0127	−4,7618	0,0003	6,3692	6,5141
	$\widehat{\text{QP}}$	0,8838	0,0936	9,4348	$1{,}91 \cdot 10^{-7}$		
Bayesiano	Intercepto	−15,4159	3,9966	−3,8572	0,0017	6,6153	6,7602
(*Gibbs*)	$\widehat{\text{QP}}$	0,8007	0,0933	8,5821	$5{,}99 \cdot 10^{-7}$		

No modelo corrigido clássico observou-se que à medida que a quantidade produzida aumenta, o valor da produção da soja no Brasil também aumenta, isto é, há uma relação direta

entre a quantidade produzida e o valor da produção de soja. Portanto, uma ampliação da estimativa da média da quantidade produzida em 1 (um), origina uma elevação de 0,8838 no valor da produção de soja em grão. No modelo bayesiano segue a mesma análise, em que no algoritmo de *Gibbs* há um aumento médio de 0,8007. Na Tabela 11 encontram-se os modelos finais estimados com os desvios padrão corrigido e os critérios de informações normalizado AIC_n e BIC_n, no qual os desvios padrão e os critérios de informações do clássico e bayesiano ficaram semelhantes, visto que quanto menor os erros padrão e os critérios de informações melhor é o modelo. Como as estatísticas da clássica e bayesiana ficaram próximas sugere-se trabalhar com o método de *Gibbs*, pois é de fácil implementação nos softwares livres (`R` e `WinBUGS`) e da mesma forma cria uma outra alternativa de aplicação da técnica de equações simultâneas no contexto do método de mínimos quadrados em dois estágios. Por fim, tem-se abaixo o gráfico dos valores estimados *versus* valores preditos dos dois métodos aplicados.

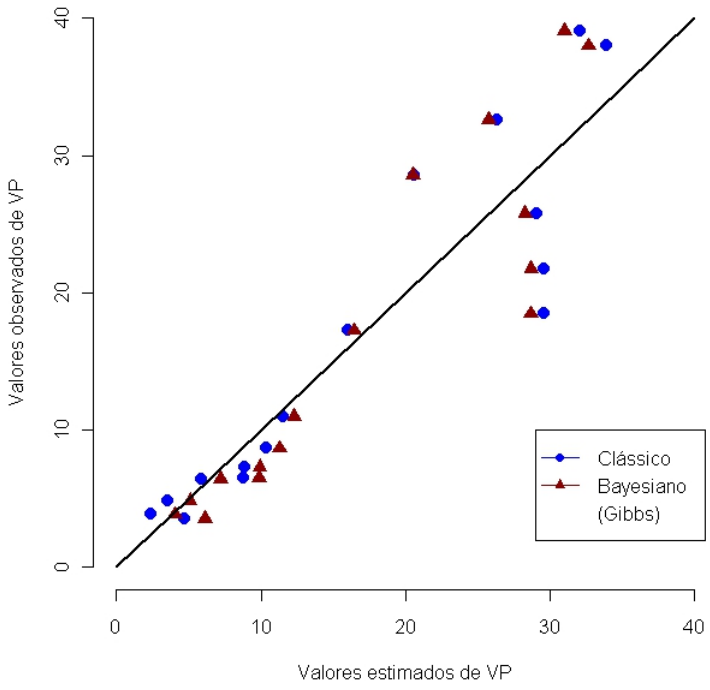

Figura 30: Gráficos dos valores observados *versus* valores preditos de VP.

7 Considerações finais

7.1 Conclusões

No decorrer da dissertação apresentaram-se as características e propriedades do método clássico de equações simultâneas através da estimação de mínimos quadrados em dois estágios; em seguida o método bayesiano por meio da simulação de Monte Carlo via Cadeia de Markov com os algoritmos de *Gibbs* e *Metropolis-Hastings*[1], dentro do contexto de equações simultâneas. Então, calcularam-se as estatísticas descritivas de posição e dispersão conjuntamente com as correlações, também aplicou-se o teste de correlações, com 90% de confiança, para verificar se de fato são correlacionadas; depois os gráficos de dispersão (Fig. 14) e box-plot (Fig. 15) para observar o comportamento das variáveis em estudo com relação a soja em grão, no Brasil, entre os anos de 1994 e 2009 (dados anuais). As variáveis em estudo foram área plantada, área colhida, quantidade produzida, valor da produção e produto interno bruto. Depois que foi aplicado as ferramentas acima, descartou-se inicialmente a variável área plantada, pelo fato dela ser semelhante a variável área colhida (vide Fig. 15) e do mesmo modo a variável área colhida apresenta uma melhor consistência e representativade para modelar a equação oferta-demanda nos métodos citados.

Na primeira aplicação clássica teve-se um bom ajuste na utilização do método de mínimos quadrados em dois estágios com os respectivos modelos, em seguida aplicou-se o teste de Hausman e concluiu-se, com 90% de confiança, que as variáveis quantidade produzida e valor da produção são mutuamente dependentes. Portanto, tem-se uma influência em utilizar o método de equações simultâneas ao contrário de se aplicar o método de mínimos quadrados ordinários.

Logo após a aplicação do método clássico realizou-se a modelação bayesiana através dos métodos de Monte Carlo via cadeia de Markov, em que o primeiro método (*Gibbs*) obteve o melhor ajuste com a *priori* Normal de média 0 (zero) e desvio padrão 0,01, tendo como precisão uma distribuição Gama com parâmetro de forma e escala 0,1.

[1] Os resultados desse algoritmo encontram-se no Apêndice A.

Em linhas gerais, registra-se que o método de Monte Carlo via cadeia de Markov com o algoritmo de *Gibbs* conseguiu-se bons resultados em "todos" os aspectos descritos nesta dissertação. Por exemplo, obtiveram-se boas estimativas no modelo de equações simultâneas (veja Figura 30); visualização gráfica das iterações, densidade, correlações, histogramas, gráficos de diagnósticos e os testes de diagnóstico de Raftery & Lewis, Geweke, Heidelberg & Welch e de Gelman & Rubin, foram significantes; também, os critérios de informação normalizado AIC_n e BIC_n ficaram semelhantes com a ferramenta clássica. Portanto, sugere-se trabalhar o método de Monte Carlo via cadeia de Markov através do método de *Gibbs* para estimar a produção de soja em grão. Da mesma forma, enfatiza-se que o uso desse novo método não deve ser confundido com o método sofisticado ou até depuração da inferência estatística, entretanto um método eficiente na modelagem e de fácil implementação nos softwares estatísticos `R` & `WinBUGS`, pois já existem bibliotecas prontas para compilar o método. Vale destacar que o uso da ferramenta MCMC é de forte uso dos estatísticos, matemáticos, economistas e outros estudiosos nessa linha de pesquisa, na qual está se tornando um método "popular" na literatura científica.

7.2 Sugestões para futuras pesquisas

Sem dúvida este trabalho não terminou na teoria e nem na prática. Existe uma gama de ideias a serem aplicados na literatura, motivo pelo qual se sugere para trabalhos futuros:

- ★ Repetir o processo utilizado nesta dissertação com um grupo maior de variáveis explicativas e respostas de oferta e demanda; por exemplo, os artigos do Gonçalvez *et al.* (2006), Guedes *et al.* (2007) ou Braga & Gomes (2008).
- ★ Nesse trabalho aplicou-se o método de mínimos quadrados em dois estágios, então poderá adaptar os métodos de mínimos quadrados indireto, instrumental e de mínimos quadrados em três estágios, no contexto bayesiano;
- ★ Utilizar o método de Monte Carlo via cadeia de Markov, na situação de equações simultâneas, com as *prioris* não informativa de Jeffreys, Haldane e Bayes-Laplace;
- ★ Trabalhar o modelo *logit* e *probit*, ou seja, categorizar a variável resposta para utilizar o modelo não linear e verificar, se de fato, o algoritmo de *Metropolis-Hastings* é melhor que o método de *Gibbs*;
- ★ Na dissertação aplicaram-se os diagnósticos de convergência para a análise da inferência bayesiana via MCMC, então pode-se fazer uma avaliação desses diagnóstico, isto é, veri-

ficar se eles são eficientes e quais os mais apropriados para cada situação dos algoritmos de *Gibbs* e *Metropolis-Hastings*;

★ No fim do século XX para o início deste século tem-se trabalhado exaustivamente o software estatístico `R`, então, foi desenvolvido diversos pacotes dentre eles o `R2WinBUGS`, no qual se realiza a simulação no software `WinBUGS` e imprime os resultados no `R` aplicando o algoritmo de *Gibbs* dentro da simulação de Monte Carlo via cadeia de Markov. Portanto, seria interessante desenvolver um pacote no `R`, nessa situação, com o algoritmo de *Metropolis-Hastings*.

Apêndice

A – Gráficos do primeiro e segundo modelo do algoritmo de *Metropolis-Hastings*

No corpo da dissertação foi discorrido a aplicação do método de Monte Carlo via cadeia de Markov por meio do algoritmo de *Gibbs*, pois obteve melhores resultados tanto na parte gráfica quanto na numérica. Já o *Metropolis-Hastings* não obteve sucesso com a regressão e *priori* normal na estimação da produção de soja em grão, no Brasil. Logo, será mostrado abaixo os gráficos de convergência e diagnóstico de convergência da primeira e segunda modelagem do método de Monte Carlo via cadeia de Markov através do algoritmo de *Metropolis-Hastings*.

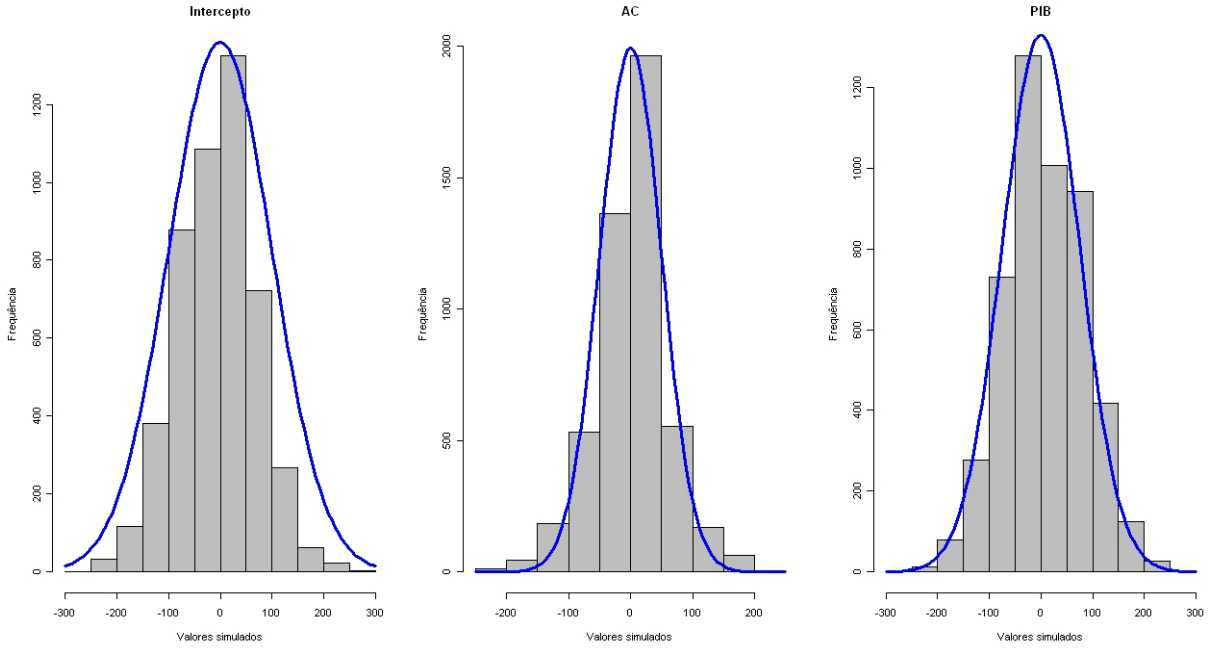

Figura 31: Histograma dos dados após o aquecimento do método de *MH* do IM.

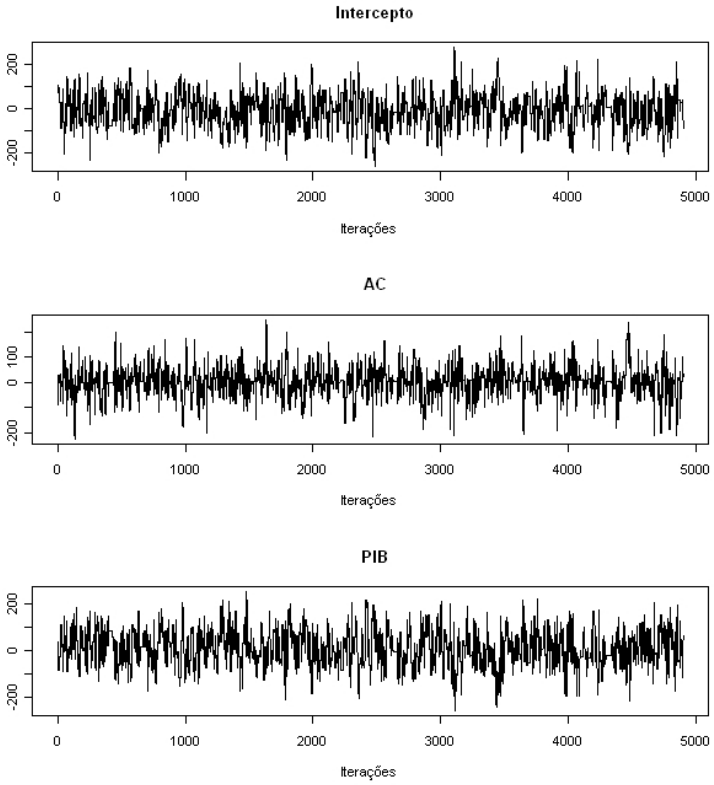

Figura 32: Visualização gráfica da convergência das iterações do método de *MH* do IM.

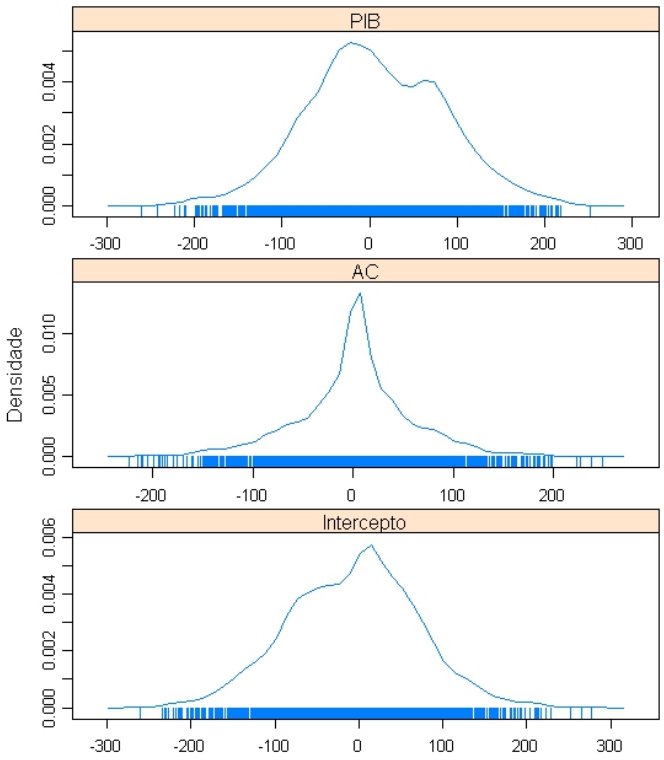

Figura 33: Gráfico da densidade de convergência do método de *MH* do IM.

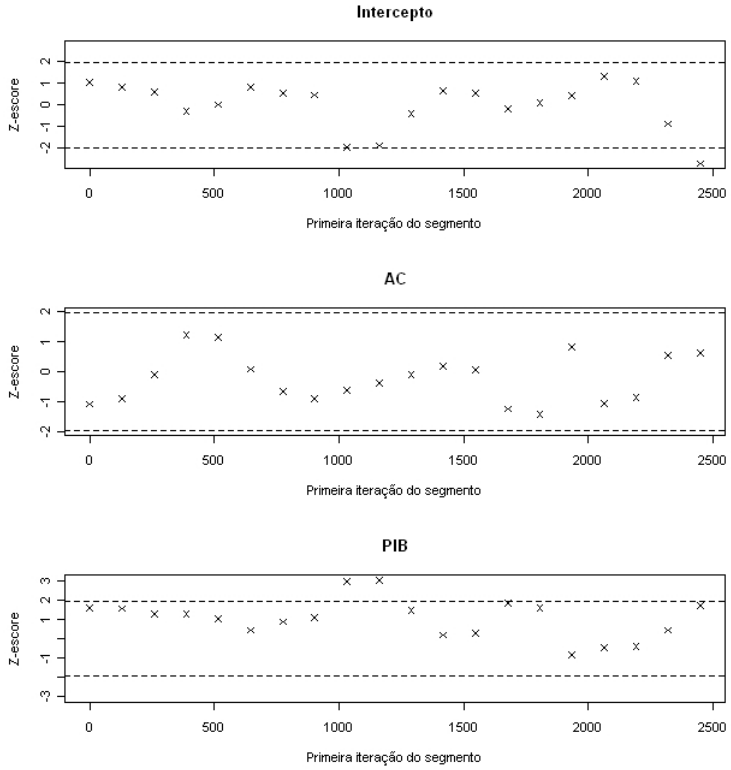

Figura 34: Diagnóstico de convergência de Geweke do algoritmo de *MH* do IM.

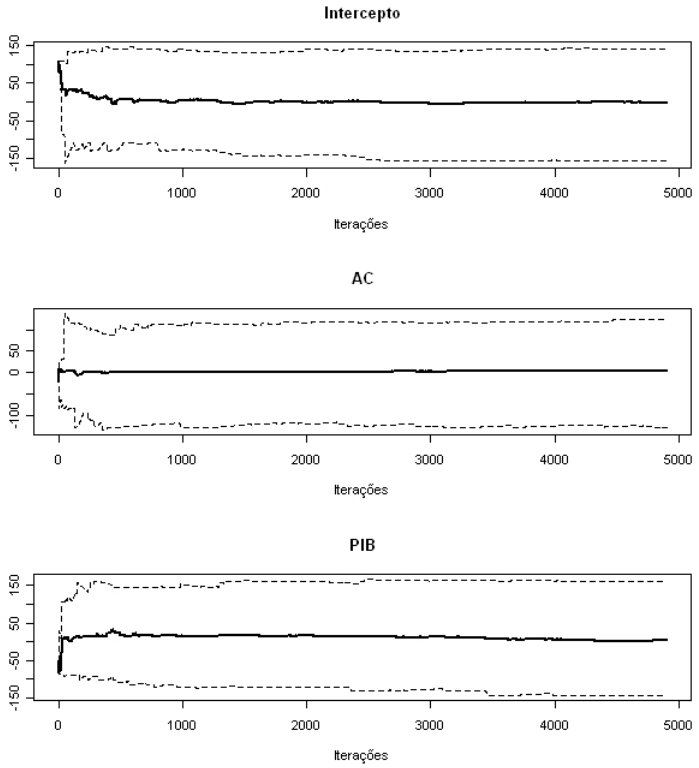

Figura 35: Gráfico da convergência acumulada do algoritmo de *MH* do IM.

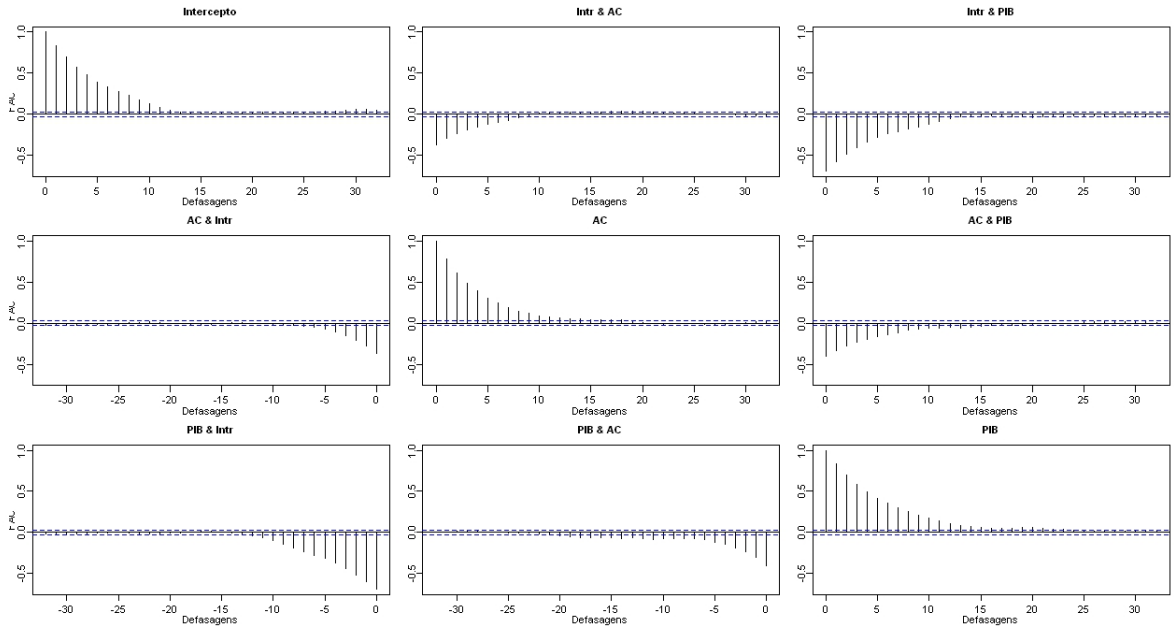

Figura 36: Gráfico das correlações do algoritmo de *MH* do IM.

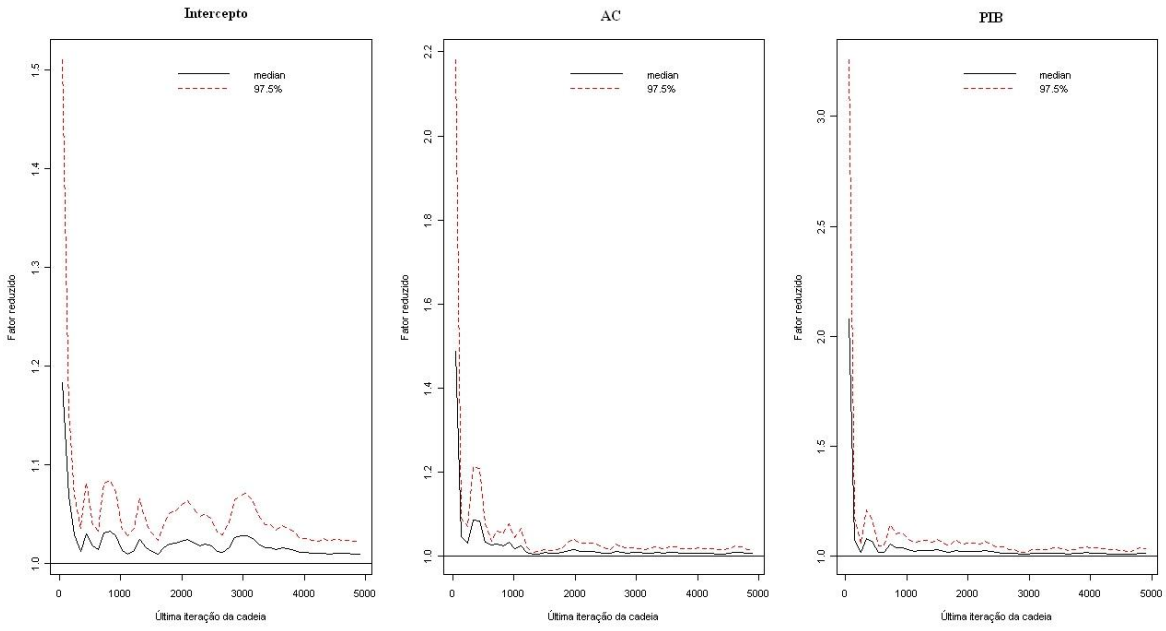

Figura 37: Visualização do diagnóstico de Gelman & Rubin do algoritmo de *MH* do IM.

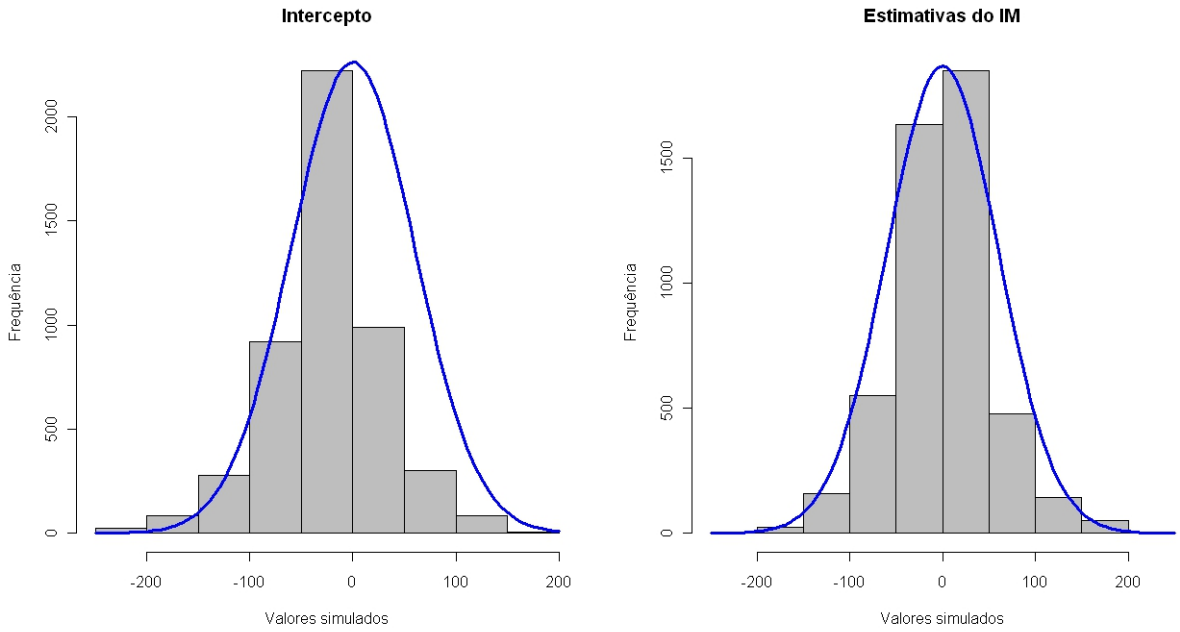

Figura 38: Histograma dos dados após o aquecimento do método de *MH* do IIM.

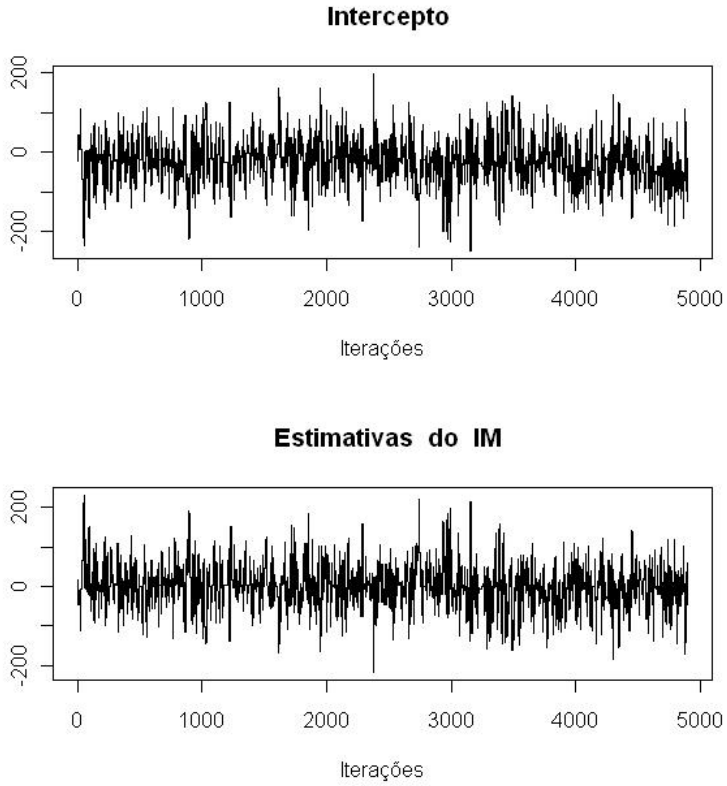

Figura 39: Visualização gráfica da convergência das iterações do método de *MH* do IIM.

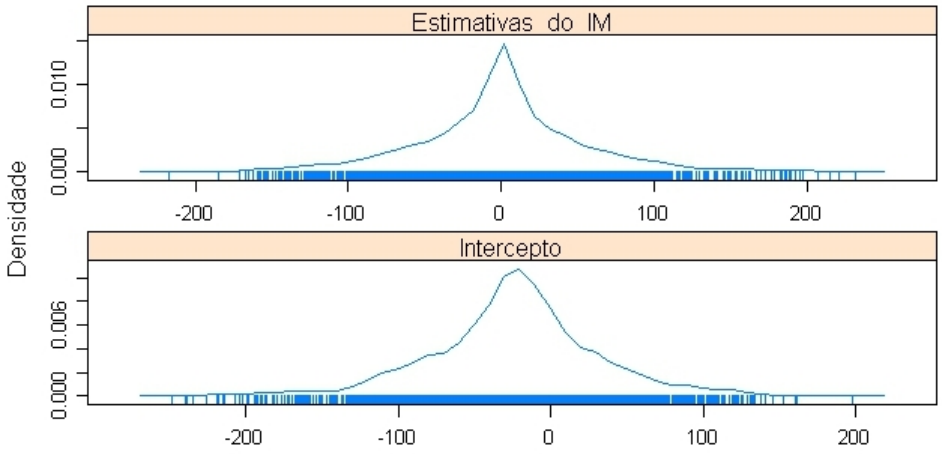

Figura 40: Gráfico da densidade de convergência do método de *MH* do IIM.

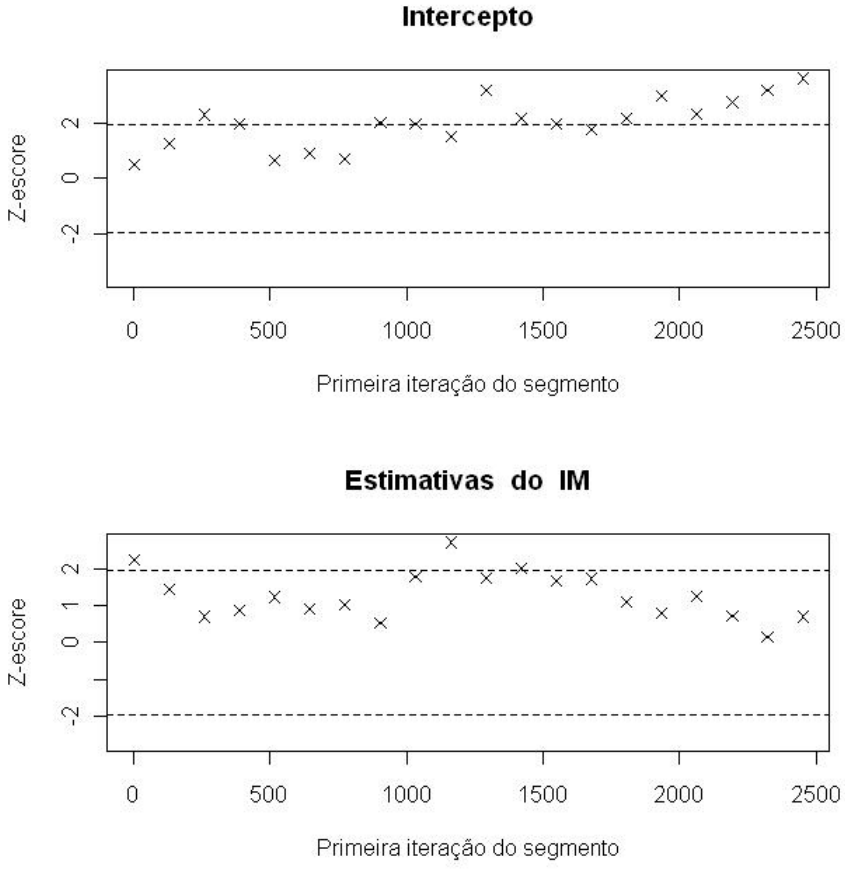

Figura 41: Diagnóstico de convergência de Geweke do algoritmo de *MH* do IIM.

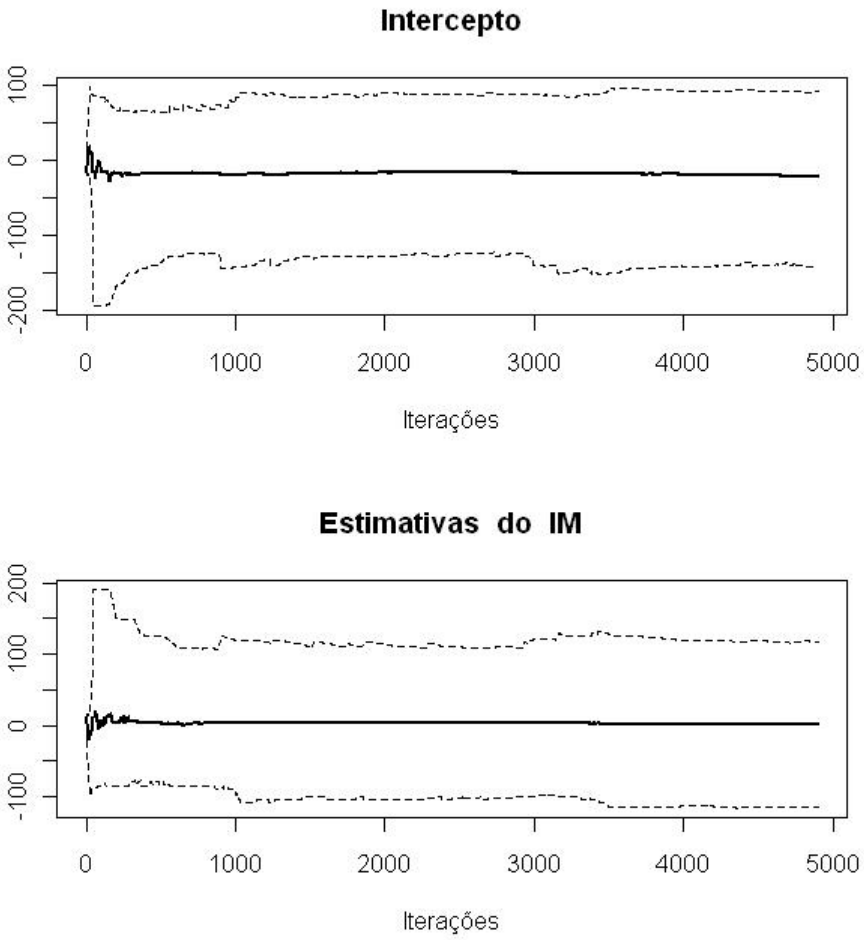

Figura 42: Gráfico da convergência acumulada do algoritmo de *MH* do IIM.

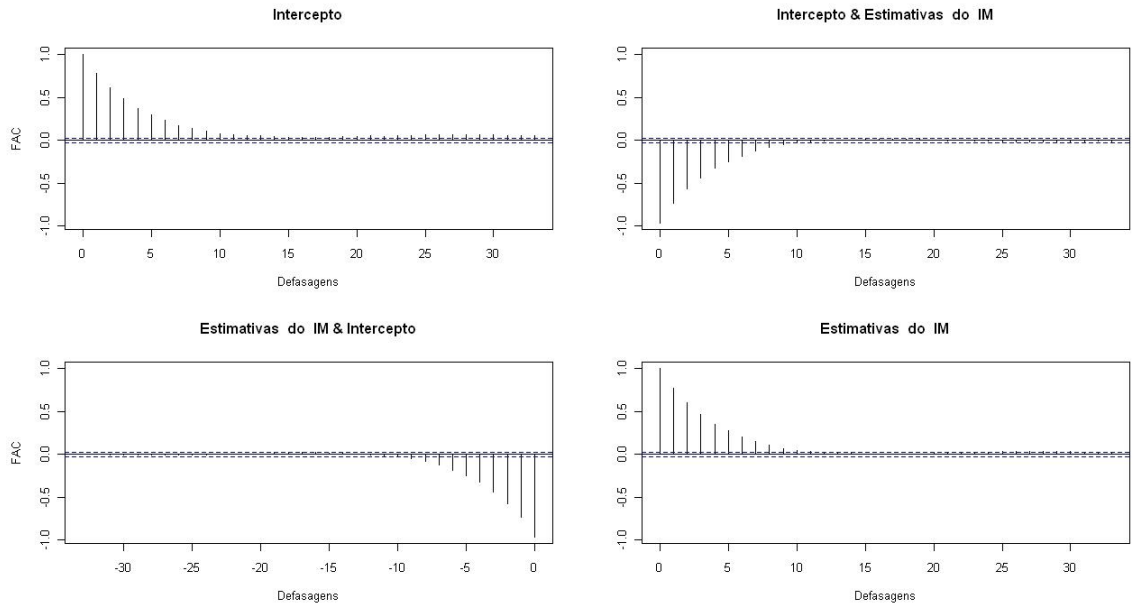

Figura 43: Gráfico das correlações do algoritmo de *MH* do IIM.

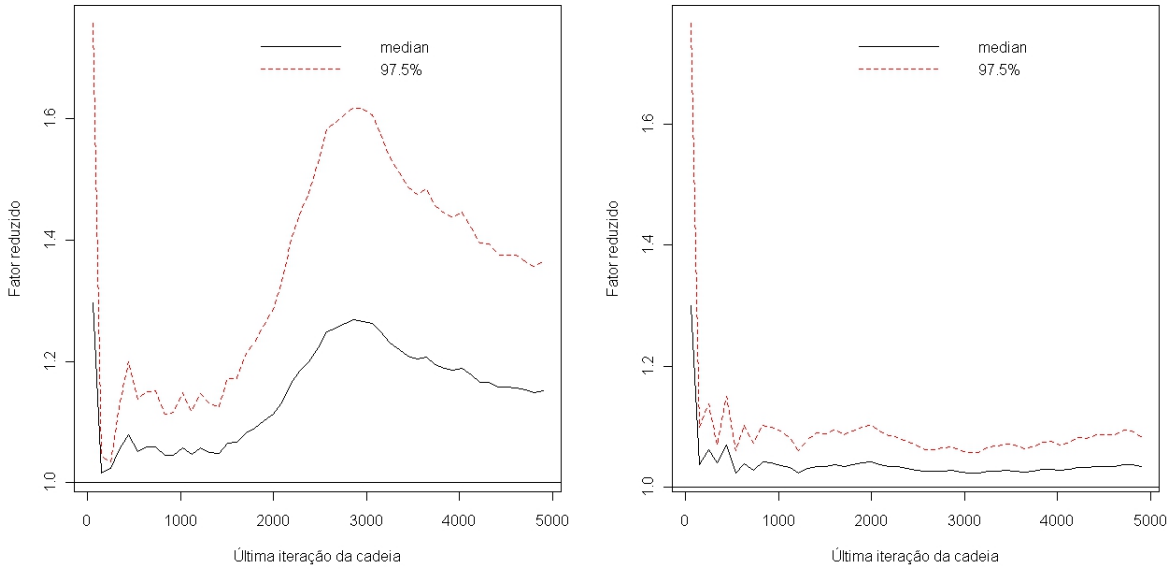

Figura 44: Visualização do diagnóstico de Gelman & Rubin do algoritmo de *MH* do IIM.

B – Conjunto de dados

Tabela 12: Dados sobre a área plantada(hectares), área colhida (hectares), quantidade produzida (toneladas), valor da produção (mil) e produto interno bruto (milhões) de soja em grão no Brasil, entre 1994 e 2009.

AP	AC	QP	VP	PIB
11.544.577	11.525.410	24.931.832	3.839.461	309.206,65
11.702.919	11.675.005	25.682.637	3.538.796	616.070,87
10.356.156	10.299.470	23.166.874	4.824.329	742.860,00
11.508.120	11.486.478	26.392.636	6.438.004	830.628,00
13.319.749	13.303.656	31.307.440	6.495.000	865.997,00
13.069.793	13.061.410	30.987.476	7.294.961	927.838,00
13.693.677	13.656.771	32.820.826	8.658.735	1.021.648,00
13.988.351	13.985.099	37.907.259	10.978.839	1.118.613,00
16.376.035	16.359.441	42.107.618	17.233.232	1.273.129,00
18.527.544	18.524.769	51.919.440	28.584.866	1.470.614,00
21.601.340	21.538.990	49.549.941	32.627.677	1.666.258,00
23.426.756	22.948.874	51.182.074	21.750.332	1.842.253,00
22.082.666	22.047.349	52.464.640	18.470.711	2.034.421,00
20.571.393	20.565.279	57.857.172	25.794.985	2.287.858,00
21.252.721	21.246.302	59.833.105	39.077.161	2.580.110,00
21.761.782	21.750.468	57.345.382	37.988.045	2.740.733,12

Fonte: IPEA e IBGE.

Referências

[1] AKAIKE, Hirotsugu (1970). Statistical prediction identification. *Annals of the Institute of Statistical Mathematics*. v. 22, p. 203-217.

[2] _____ (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, New York, v. 19, n. 6, p. 716-723.

[3] ALDER, Mike (2005). *Non-Aristotelian Logic in Practice, or How to be much cleverer than all your friends (so they really hate you)*. Disponível em: <http://www.indiana.edu/∼pietsch/callosum.html#mri>. Acesso em: 20 de fevereiro de 2011.

[4] ANDRADE, Humber A. & KINAS, Paul G. (2010). *Introdução à análise Bayesiana (com R)*. Porto Alegre: MaisQnada. 258p.

[5] BALTAGI, Badi H. (2005). *Econometric analysis of panel data*, 3th ed. Inglaterra: John Wiley & Sons, Ltd.

[6] _____ (2008). *Econometrics*, 4th ed. Nova York: Springer. 403p.

[7] BARBOSA, Fernando de Holanda (1983). Modelos de Equações Simultâneas. *Fundação Getúlio Vargas*. Disponível em: <http://bibliotecadigital.fgv.br/dspace/handle/10438/604>. Acesso em: 31 março 2011.

[8] BICKEL, P.J. & DOKSUM, K.A. (1977). Mathematical Statistical. Basic Ideas and Selected Topics. Holden-Day. 493p.

[9] BOX, G. E. P. & TIÃO, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Wiley Classics. 588p.

[10] BRAGA, Marcelo J. & GOMES, Sérgio Castro (2008). Determinantes da produtividade total dos fatores na Amazônia legal: uma aplicação de dados em painel. *Amazônia: Ci. & Desenv.*. Belém, v. 3, n. 6, janeiro/julho de 2008.

[11] BROOKS, Stefhen P. & ROBERTS, Gareth O. (1997) *Assessing convergence of Markov Chain Monte Carlo algorithms*. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.40.2642&rep=rep1&ty pe=pdf>. Acesso em: 31 março 2011.

[12] BUSSAB, Wilton de O. & MORETTIN, Pedro A. (2010). *Estatística Básica*. 6^a Edição. São Paulo: Editora Saraiva. 526p.

[13] CARLIN, Bradley P. & COWLES, Mary Kathryn (1996). Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association*, v. 91, n. 434. Jun, p. 883-904.

[14] CASELLA, G. & ROBERT, C. P. (1999). Monte Carlo Statistical Methods. *Springer-Verlag*, New York.

[15] _____ (2010). Introducing Monte Carlo methods with R. *Springer*, XX, p. 168-197.

[16] CASELLA, George & BERGER, Roger L. (2011). *Inferência estatística*. 2^a Edição. São Paulo: Editora Cengage Leaenig. 588p.

[17] CHEN, M. H. , SHAO, A. & IBRAHIM, J. (2000). *Monte Carlo Methos in Bayesian Computation*. Springer-Verlag, NY.

[18] CHIB, Siddhartha & GREENBERG, Edward (1993). Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory*, in press.

[19] _____ (1994). Bayes inference for regression models with ARMA(p,q) errors. *Journal of econometric*, v. 64, p. 183-206.

[20] _____ (1995). Understanding the Metropolis-Hastings algorrithm. *The American Statistician*, v. 49, n. 4. Novembro 9p.

[21] DEGROOT, M. H. (1986). *Probability and Statistics*, 2^a Edição, Addison-Wesley, Reading, MA, 1986, Cap. 6, apud RAMOS, Paulo C. Formiga.

[22] DEY, D. K & GELFAND, A. E. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society*, B, v. 56, n. 56, p. 501-514.

[23] DIMOVE, Ivan T. (2008). *Monte Carlo methods for applied scientists*. Bulgaria: World Scientific, 308p.

[24] DUARTE , José Antônio M. S. (2011). *O método de Monte Carlos*. Universidade Estadual de Maringá. Centro de Ciências Exatas. Departamento de Estatística.

[25] EHLERS, Ricardo Sanders (2003). *Inferência estatística*. Departamento de Estatística. UFPR, Paraná. Disponível em:<http://leg.ufpr.br/∼ehlers/notas/mci.pdf>. Acesso em: 01 agosto 2007.

[26] _____ (2004). *Métodos computacionais intensivos em estatística*. Departamento de Estatística. UFPR, Paraná. Disponível em: <http://www.est.ufpr.br/∼ehlers/notas>. Acesso em: 01 agosto 2007.

[27] _____ (2005). *Introdução a Inferência Bayesiana*. Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná.

[28] EPPLE, D. & MCCALLUM, B. T. (2006). *Simultaneous equation econometrics: the missing example*, Economic Inquiry, p. 374-384.

[29] FARIAS, Agnaldo Monteiro; ISHIBASHI ,Regina; MENDONÇA, Flávia Costa Gomes de; TASSO, João Paulo (2011). *Determinação da Integral de uma Função Utilizando o Método de Monte Carlo*. UNESP - Universidade Estadual Paulista - Presidente Prudente Sessão de Estatística.

[30] FILHO, Moacyr Cunha; SANTOS, Eufrázio de Souza; VASCONCELOS, Josimar Mendes de (2011). *Equações simultâneas no contexto clássico e Bayesiano: uma aplicação à produção de soja*. Recife. submetido para revista Pesquisa Agropecuária Brasileira.

[31] FLURY, Thomas & SHEPHARD, Neil (2008). *Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models*. Disponível em: <http://ideas.repec.org/p/oxf/wpaper/413.html>. Acesso em: 31 fevereiro 2011.

[32] GAMERMAN, D. (1997). *Markov Chain Monte Carlo: stochastic simulation for bayesian inference*. Chapman & Hall.

[33] GAMERMAN, D. & H. Lopes (2006). *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Texts in Statistical Science Series. CRC Press.

[34] GELFAND, A. E. & SMITH, A. F. M. (1990). Sampling-based approaches to calculating marginal distributions. *Journal of the American Statistical Association*, v. 85, n. 410, p. 398-409.

[35] GELMAN, Andrew; LIGGES, Uwe; STURTZ, Sibylle (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Jounal of Statistical Software*, v. 12, p. 1-17.

[36] GELMAN, A. & RUBIN, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, v. 7, 457-72.

[37] GEMAN, S. & GEMAN, D. (1984). Stochatisc relaxation, Gibbs distribution and the bayesian restoration of images. *IEEE Tramsaction on Patterm Analysis and Machine Intelligence*, v. 6, p. 721-741.

[38] GEWEKE, J. (1992). Evaluating the accuracy of sampling-based approaches to calcualting posterior moments. *Bayesian Statistics*, v. 4.

[39] GILKS, W. R.; RICHARDSON, S. & SPIEGELHALTER, D. J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

[40] GONÇALVEZ, Manoel Carlos; KURIHARA, Carlos Hissao; MARCHETTI, Marlene Estevão; URANO, Eliane O. Mocelin; VITORINO, A. C. Tadeu (2006). Avaliação do estado nutricional da soja. *Pesquisa agropecuária brasileira*. Brasília, v.41, n.9, p.1421-1428, setembro de 2006.

[41] GREENE, William H (1997). *Econometric Analysis*, 3rd ed. Upper Saddle River, New Jersey: Prentice Hall.

[42] GUEDES, Terezinha A.; JANEIRO, Vanderly; MARTINS, Elias Nunes & ROSSI, Robson Marcelo (2007). Estimação do fator de condição de peixes da espécie tracydoras paraguayensis: uma perspectiva bayesiana. *Acta Sci*. Anim. Sci. v. 29, n. 1, p. 85-92.

[43] GUJARATI, D. N. (2006). *Basic Econometrics*. 4th ed. Nova York: McGraw-Hill.

[44] HAMMERSLEY, J. M. & HANDSCOMB, D. C. (1964). Monte Carlo Methods. *Methuen & Co.*, London, e John Wiley & Sons, Inc. New York, 178p.

[45] HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.

[46] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, p. 97 a 109.

[47] HAUSMAN, J. A. Specification Tests in Econometrics. *Econometrica*, vol 46, novembro de 1976, p. 1251-1271.

[48] HEIDELBERG, P. & WELCH, P. (1983). Simulation run lenght control in the presence of an initial transient. *Operations Research*, v. 31, p. 1109-44.

[49] HUNGRIA, Mariangela; NICOLÁS, Marisa Fabiana; SANTOS, Maria Aparecida dos (2006). Identificação de QTL associados à simbiose entre Bradyrhizobium japonicum, B. elkanii e soja. *Pesquisa agropecuária brasileira*. Brasília, v.41, n.1, p. 67-75, janeiro de 2006.

[50] JAMES, Barry R. (2006). *Probabilidade: um Curso em nível intermediário*. 3^a Edição. Rio de Janeiro: Instituto de Matemática Pura e Aplicada. 304p.

[51] KALOS, Malvin H. & WHITLOCK, Paula A. (2004). *Monte Carlo methods*, New York: Wiley-VCH, 195p.

[52] LAMPORT, L. (1994). *A Document Preparation System LATEX, User's Guide and Reference Manual*, 2^a Edição. Massachusett: Addison-Wesley.

[53] LEANDRO R. A. (2001). *Introdução à Estatística Bayesiana*. Departamento de Ciências Exatas, ESALQ/USP Piracicaba, SP$-$7 a 13 julho.

[54] LEOTTI, Vanessa Bielefeldt (2007). *Comparação via simulação dos estimadores clássicos e Bayesianos no modelo de coeficientes aleatórios para dados longitudinais*. Dissertação. Universidade Federal do Rio Grande do Sul. Rio Grande do Sul, 99p.

[55] LIU, B. (1983). *Natural gas price elasticities - Variations by region and by sector in the USA*. Energy Economics.

[56] LOPES, Hedibert Freitas & SALAZAR, Esther (2011). *Bayesian Inference and Selection in Smooth Transition Autoregressive Models*. Rio de Janeiro: Universidade Federal do Rio de Janeiro. 102 slides, color. Disponível em: <http://acd.ufrj.br/$\sim$hedibert/lavras2003.pdf>. Arquivo obtido em 7 de Abril de 2011.

[57] LUNN, DJ; THOMAS, A.; BEST, N. & SPIEGELHALTER, D. (2000). *WinBUGS - um quadro de modelagem Bayesiana: conceitos, estrutura e capacidade de extensão*. Estatística e Computação, 10:325-337. Disponível em: <http://translate.google.com.br/translate?hl=pt-BR&langpair=en|pt&u=http://www.mrc-bsu.cam.ac.uk/bugs/>. Arquivo obtido em 7 de Abril de 2011.

[58] MADALLA, G. S. (1992). *Introduction to econometrics*, 2^a Edição. New York: Macmillan. 637p.

[59] MAGALHÃES, Marcos Nascimento (2006). *Probabilidade e variáveis aleatórias*. 2^a Edição. São Paulo: Edusp. 428p.

[60] MAGNUS, Jan R. & NEUDECKER, Heinz (2007). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley. 470p.

[61] MARTIN, A. D. & QUINN, K. M. (2010). *MCMCpack: Markov chain Monte Carlo (MCMC) Package*. R package version 0.7-1. Disponível em: <http://mcmcpack.wustl.edu>. Acesso em: 01 de maio de 2011.

[62] MARTIN, M. A. & PEREZ, M. C. R. C (1975). *O Método dos Mínimos Quadrados de Dois estágios: seus fundamentos e aplicação na estimação da demanda e oferta de ovos no estado de São Paulo*. Piracicaba: Universidade de São Paulo, 39p.

[63] MAYRINK, Vinícius Diniz (2006). *Aspectos computacionais de modelos espaciais*. Dissertação. Universidade Federal do Rio de Janeiro. Rio de Janeiro, 144p.

[64] METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H. & TELLER E. (1953). Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*.

[65] MEYER, Paul L. (1983). *Propabilidade: Aplicações à Estatística*. 2^a Edição. Rio de Janeiro. Livros Técnicos e Científicos Editora. 426p.

[66] MING-HUI, Chen & QI-MAN, Shao (2000). Propriety of posterior distribution for dichotomous quantal response models. *Proceedings of the American Mathematical Society*. v.129, n.1, p.293-302.

[67] MITTELBACH, F.; GOOSSENS, M.; BRAAMS, J.; CARLISL, D. & Rowley, C. (2004). *The Latex Companion. Tools and Techniques for Computer Typesetting*. Addison Wesley, Boston.

[68] MONTGOMERY, Douglas C. & RUNGER, George C. (2003). *Applied statistics and probability for engineers*. 3a. ed. United States of America: John Wiley & Sons, Inc. 822p.

[69] MOOD, Alexander M.; GRAYBILL, Franklin A. & BOES, Duane C. (1983). *Introduction to the Theory of Statistics*. 3^{nd}. ed. California: International Student. 564p.

[70] MÜLLER, P. (1993). *A generic Approach to posterior integration and Gibbs sampling*, manuscrito.

[71] NIESWIADOMY, M. L. & MOLINA, D. J. (1988). Urban water demand estimates under increasing block rates. *Growth and Change*, v.19, p. 1-12.

[72] _____ (1991). A note on price perception in water demand models. *Land Economics*, v.67. p. 352-592.

[73] NTZOUFRAS, Ioannis (2009). *Bayesian modeling using WinBUGS*. New Jersey: Wiley Series in Computacional Statistics. 506p.

[74] PAULINO, Carlos Daniel; TURKMAN, M. Antónia Amaral & MURTEIRA, Bento (2003). *Estatística Bayesiana*. Fundação Caloute Gulbenkian, Av. de Berna, Lisboa, 446p.

[75] PENA, Sérgio Danilo (2006). FILOSOFIA da ciência. *Revista de divulgação científica da SBPC*. v. 38, n. 228, p. 22-29, julho.

[76] PHILLIPS, D. B. & SMITH, A. F. M. (1994). Bayesian faces via hierarchical template modeling. *Journal of the american statistical association*, v. 89, p. 1151-1163.

[77] PINDYCK, Robert S. & RUBINFELD, Daniel L (2004). Econometria: Modelos e Previsões. Tradução da 4^a Edição. *Americana: Econometric Models and Economic Forecasts*. Rio de Janeiro: Elsevier.

[78] PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. (2010). *CODA: Output analysis and diagnostics for MCMC*, R package version 0.14-2.

[79] RAFTERY, A.L. & LEWIS, S. (1992). *How many iterations in the Gibbs sampler? Bayesian Statistics*. v. 4, 763-74.

[80] ROSSI, Robson Marcelo (2011). *Introdução aos métodos Bayesianos na análise de dados zootécnicos com uso do WinBUGS e R*. Departamento de Estatística. Fundação Universidade Estadual de Maringá, Paraná.

[81] ROSSI, P. & ZELLNER, A. (1984). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*. v. 25, n. 22, p. 365-393.

[82] SANTOS, Henrique Pereira dos; LHAMBY, Cesar Barreneche & SANDINI, Itacir (1997). Efeitos de culturas de inverno e de sistema de rotação de culturas sobre algumas características da soja. *Pesquisa Agropecuária Brasileira*.

[83] SCHWARZ, Gideon (1978). Estimating the dimension of a model. *Annals of Statistics*, v. 6, n. 2, p. 461-464.

[84] SILVERMAN, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, v. 12, 898-916.

[85] _____ (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *Journal of the Royal Statistical Society*, B 47, 1-52.

[86] SILVERMAN, B.W & GREEN, P.J. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

[87] SOARES, N.; SILVA, M. & FONTES, A. (2004). *Análise econométrica do mercado brasileiro de carvão vegetal no período de 1974 a 2000*, Scientia Forestalis (66), p. 84-93.

[88] SORENSEN, D. (1996). *Gibbs Sampling in quantitative genetic*. Copenhagen: Foulun, 186p.

[89] SOUZA, C. P. E. (2008). *Testes de Hipóteses para Dados Funcionais Baseados em Distâncias: Um Estudo Usando Splines*. Dissertação (Mestrado em Estatística) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica (UNICAMP/IMECC), Campinas-SP.

[90] SOUZA, Geraldo da Silva (1998). *Introdução aos modelos de regressão linear e não-linear*. Brasilia: EMBRAPA-SPI. 505p.

[91] THOMAS, A.; SPIEGELHALTER, D. J.; GILKS, W. R. (1992). BUGS: A program to perform Bayesian inference using Gibbs Sampling. Em *Bayesian Statistics*. V. 4, 837-842 (Editors Bernardo, J. M., Berger, J. O., Dawid, A.P. e Simth, A. F. M.).

[92] TRUCOM, Conceição (2009). *Soja: nutrição e saúde*. São Paulo: Alaúde. 151p.

[93] VASCONCELOS, Josimar Mendes de (2007). *Um curso introdutório de Inferência Bayesiana*. Monografia. Universidade Federal do Rio Grande do Norte. Rio Grande do Norte, 70p.

[94] VENABLES, W.N; Smith, D.M. & R Development Core Team. (2009). An introduction to R. Disponível em: *http://cran.r-project.org/doc/manuals/R-intro.pdf*. Arquivo obtido em 7 de Abril de 2011.

[95] VICENTE, José Roberto. (1994). Análise comparativa de métodos de estimação da oferta e demanda de carnes e ovos. *Agricultura em São Paulo*. São Paulo, v. 41, n. 1, p. 1-20.