

DENISMAR ALVES NOGUEIRA

**PROPOSTA E AVALIAÇÃO DE CRITÉRIOS DE
CONVERGÊNCIA PARA O MÉTODO DE MONTE
CARLO VIA CADEIAS DE MARKOV: CASOS UNI E
MULTIVARIADO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de Mestre”.

Orientadora:

Prof^a Dr^a Thelma Sáfadi

Co-orientador:

Prof^o Dr^o Daniel Furtado Ferreira

LAVRAS

MINAS GERAIS – BRASIL

2004

DENISMAR ALVES NOGUEIRA

**PROPOSTA E AVALIAÇÃO DE CRITÉRIOS DE
CONVERGÊNCIA PARA O MÉTODO DE MONTE
CARLO VIA CADEIAS DE MARKOV: CASOS UNI E
MULTIVARIADO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de Mestre”.

APROVADA em 27 de fevereiro de 2004

Prof.^a Dr.^a Rosângela Helena Loschi

UFMG

Prof.^o Dr.^o Augusto Ramalho de Moraes

UFLA

Prof.^o Dr.^o Daniel Furtado Ferreira

UFLA

Prof.^a Dr.^a Thelma Sáfyadi
(Orientadora)

LAVRAS
MINAS GERAIS – BRASIL

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Nogueira, Denismar Alves

Proposta e avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov: casos uni e multivariados / Denismar Alves Nogueira. -- Lavras : UFLA, 2004.

121 p. : il.

Orientadora: Thelma Sáfyadi.

Co-Orientador: Daniel Furtado Ferreira.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Amostrador de Gibbs. 2. Correlação. 3. Séries temporais. I. Universidade Federal de Lavras. II. Título.

CDD-519.282

*À mulher que eu amo, **Thaísa**. Companheira dedicada soube suportar com abnegação e carinho os caminhos espinhosos por que passei nesta jornada. Obrigado!*

*Ao meu motivo de inspiração e alegria de viver, **Mateus**!*

*Ao meu pai, **Emar**. Em sua humildade e seu jeito de ser simples eu me inspiro e a eles anseio chegar. Sua jornada não foi em vão. Obrigado!*

*À minha mãe, **Denaide**, pelo carinho que sempre me dedicou.*

AGRADECIMENTOS

A Deus por mais esta etapa vencida. A minha família que esteve sempre ao meu lado, e que acreditaram em mim.

A **Thelma Sáfadi** pela sua orientação, dedicação, apoio, ensinamentos, confiança, amizade e carinho. Mesmo a distância não deixou de estar presente.

Ao **Daniel Furtado Ferreira**, pela sua orientação, amizade, apoio e companheirismo. Foi fundamental para os resultados deste trabalho.

Aos professores **Eduardo Bearzoti, Augusto Ramalho de Moraes, Daniel Furtado Ferreira, Luís Henrique de Aquino, Júlio Silvio de Souza Bueno Filho, Ruben Delly, Joel Augusto Muniz, Thelma Sáfadi e Paulo Lima da Silva** pelos ensinamentos, apoio e dedicação.

A **Maria das Dores Corrêa Santos** por sua dedicação e carinho.

Ao professor **Joel Augusto Muniz** pela grande dedicação e esforço na melhora constante do curso.

Aos colegas **José Leite, Fabyano (Bia), Janser, José Waldemar (Zé), Francisco (Xodó), Vânia, Renata, Maria (Néia), Rafaela e Eliana**. pela convivência que tivemos, pela amizade, carinho e apoio nas horas difíceis .

A equipe Master Stat agradeço a amizade e desejo grande sucesso.

Aos professores e amigos não citados mas que estão presentes no coração.

Obrigado a CAPES pela bolsa, permitindo dedicação neste trabalho.

SUMÁRIO

	Página
RESUMO.....	i
ABSTRACT	ii
LISTA DE TABELAS	iii
LISTA DE FIGURAS	iv
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	4
2.1 Distribuição Normal Multivariada.....	4
2.2 Maximização de formas quadráticas	5
2.3 Maximização de pares de formas quadráticas	6
2.4 Séries Temporais	8
2.4.1 Modelo de intervenção com erro ARMA(p,q)	8
2.5 Inferência Bayesiana.....	9
2.5.1 Considerações Iniciais	9
2.5.2 Inferência Bayesiana.....	10
2.5.3 Inferência Bayesiana para o modelo de intervenção com erro ARMA(p,q)	13
2.6 Métodos Aproximados de Inferência.....	18
2.7 Simulação Estocástica via Cadeias de Markov	19
2.8 Amostrador de Gibbs (“Gibbs Sampler”).....	20
2.9 Autocorrelação e Correlação Cruzada.....	21
2.9.1 Autocorrelação.....	21
2.9.2 Correlação Cruzada	23

2.10 Sequências Múltiplas	24
2.11 Valores Iniciais.....	25
2.12 Verificação Informal da Convergência.....	26
2.12.1 Ergodicidade.....	26
2.13 Verificação Formal da Convergência.....	27
2.13.1 Critério de Gelman & Rubin (1992a).....	27
2.13.2 Critério de Raftery & Lewis (1992a).....	31
2.13.3 Análise Espectral	34
2.13.3.1 Critério de Geweke (1992)	34
2.13.3.2 Critério de Heidelberg & Welch (1983).....	35
2.13.4 Critério multivariado de Brooks & Gelman (1998).....	37
2.14 Justificativa.....	40
3 METODOLOGIA.....	42
3.1 Análise do modelo com duas intervenções e erro ARMA(2,2).....	43
3.2 Geração de números aleatórios para a normal trivariada utilizando o Amostrador de Gibbs.....	45
3.3 Critérios Univariados.....	47
3.3.1 Critério de Gelman & Rubin (1992a).....	47
3.3.2 Critério de Raftery & Lewis (1992a).....	48
3.3.3 Critério de Geweke (1992)	48
3.3.4 Critério de Heidelberg & Welch (1983).....	49
3.4 Critérios Multivariados.....	49
4 RESULTADOS E DISCUSSÃO.....	51
4.1 Critérios Univariados.....	51
4.1.1 Modelo de intervenção com erro ARMA(2,2)	51
4.1.1.1 Critério de Gelman & Rubin (1992a).....	51
4.1.1.2 Critério de Geweke (1992).....	54
4.1.1.3 Critério de Heidelberg & Welch (1983).....	57

4.1.1.4 Critério de Raftery & Lewis (1992a).....	62
4.1.2 Normal trivariada.....	67
4.1.2.1 Critério de Gelman & Rubin (1992a).....	67
4.1.2.2 Critério de Geweke (1992)	75
4.1.2.3 Critério de Heidelberger & Welch (1983).....	78
4.1.2.4 Critério de Raftery & Lewis (1992a).....	80
4.2 Critérios Multivariados.....	85
4.2.1 Cálculo alternativo para maximização de formas quadráticas	85
4.2.2 Dois critérios de convergência propostos.....	87
4.2.3 Aplicação dos critérios multivariados	88
4.2.3.1 Modelo de intervenção com erro ARMA(2,2)	88
4.2.3.2 Normal trivariada.....	91
4.3 Considerações Finais	94
5 CONCLUSÃO.....	98
REFERÊNCIAS BIBLIOGRÁFICA	99
ANEXO A	104
ANEXO B.....	111

RESUMO

NOGUEIRA, Denismar Alves. **Proposta e avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov: casos uni e multivariados**. 2004. 121p. Dissertação (Mestrado em Agronomia /Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG*.

Métodos de Monte Carlo via Cadeias de Markov têm sido estudados com aplicações em diversas áreas, mas uma das maiores dificuldades é determinar o tamanho adequado da amostra, ou seja, a convergência do processo para assim inferir sobre os parâmetros da distribuição de interesse. Dos critérios existentes na literatura, optou-se por estudar aqueles mais utilizados e mais simples de se implementar. Por isso, os objetivos foram: propor uma forma alternativa de aplicação combinada de quatro critérios univariados de convergência das seqüências de Markov (Gelman & Rubin, 1992a, Raftery & Lewis, 1992a, Geweke, 1992, Heidelberger & Welch, 1983), otimizando suas propriedades; apresentar uma forma alternativa de cálculo do critério de Brooks & Gelman (1998); e propor dois novos critérios multivariados de monitoramento da convergência. Para isso, dados referentes a dois modelos foram utilizados, sendo um de séries temporais com duas intervenções e erro ARMA(2,2) e o outro uma normal trivariada, considerando três situações distintas para a matriz de covariâncias. Em ambos os casos foram utilizados o Amostrador de Gibbs e aplicados os critérios para monitorar a convergência. Os resultados obtidos permitiram que se concluísse que: sugere-se a utilização dos critérios univariados do monitoramento da convergência de forma combinada; a forma alternativa de cálculo para o critério de Brooks & Gelman foi viável de ser aplicada e mostrou-se robusta numericamente; os dois novos critérios multivariados para monitorar a convergência das seqüências de Monte Carlo foram propostos com sucesso; o critério multivariado baseado na razão dos traços das matrizes de covariâncias $\hat{V}$ e D foi considerado melhor.

Palavras Chaves: Amostrador de Gibbs; Correlação; Séries Temporais.

* Comitê Orientador: Prof^a Dr^a. Thelma Sáfaci - UFLA(orientadora) e Prof. Dr. Daniel Furtado Ferreira – UFLA(co-orientador).

ABSTRACT

NOGUEIRA, D. A. **Proposal and evaluation of convergence diagnostics criterion for Markov Chain Monte Carlo methods: univariate and multivariate cases.** 2004. 121p. Dissertation (Masters in Statistics and Agricultural Experimentation) – Universidade Federal de Lavras, Lavras, MG*.

Markov Chain Monte Carlo Methods have been studied in several areas, but one of the largest difficulties is to determine the appropriate sample size, i.e., to determine the convergence of the process and then make inference over parameters of the target distribution. In spite of several accessible convergence diagnostic procedures in literature we have opted to study the most used and simplest to implement diagnostic procedures. The objectives of this work were: to propose an alternative form of combined application of 4 univariate Markov chain convergence diagnostics (Gelman & Rubin, 1992a, Raftery & Lewis, 1992a, Geweke, 1992, Heidelberger & Welch, 1983) and optimize their properties; to present an alternative numerical method for calculating the Brooks & Gelman's (1998) convergence diagnostic procedure; and to propose two new multivariate convergence diagnostic procedures. Simulated data from two models was used. The first a time series with two intervention parameters with ARMA {2,2} error and the second a normal trivariate, utilizing three different structures for the covariance matrix. In both cases the Gibbs sampler was used. The following conclusions were obtained: it was suggest that univariate diagnostic procedures should be applied in a combined way; the alternative form for calculating Brooks & Gelman's convergence diagnostic procedure was easy to be applied and numerically robust; the two new multivariate convergence diagnostics were proposed with success; the multivariate convergence diagnostic based on ratio of the $\hat{V}$ and D covariance matrix traces was considered the best.

Keywords: Correlation; Gibbs sampler; time series.

*Guidance Committee: Prof. DSc. Thelma Sáfiadi – UFLA(Adviser), Prof. DSc. Daniel Furtado Ferreira – UFLA(Co-Adviser).

LISTA DE TABELAS

TABELA 1. Resumo de alguns critérios de convergência encontrados na literatura.	41
TABELA 2. Valores arbitrários iniciais utilizados para iniciar cada sequência para o modelo de intervenção com erro ARMA	45
TABELA 3. Valores arbitrários iniciais utilizados para iniciar cada sequência para o modelo normal trivariado.....	47
TABELA 4. Análise descritiva para os 7 parâmetros em 3 diferentes níveis de convergência para o PRFS.	53
TABELA 5. Resultados do teste da estacionariedade para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações.....	59
TABELA 6. Resultados do teste de “half-width” para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações.....	60
TABELA 7. Resultados do “burn-in” para o parâmetro τ realizado a cada 100 iterações até 30000 iterações.....	62
TABELA 8. Resultados do Critério de Raftery & Lewis para uma amostra mínima de 3746 dados, aplicado em diferentes tamanhos de amostra piloto.....	63
TABELA 9. Resultados do Critério de Geweke para as duas novas amostras considerando 5% de significância para o teste Z.	64
TABELA 10. Resultados do Critério de Heidelberger & Welch para as duas novas amostras.	65

TABELA 11. Análise descritiva para as duas novas amostras, Erros de Monte Carlo, intervalo HPD com 95% de probabilidade e Comprimento do Intervalo (C.I.).	66
TABELA 12. Análise descritiva para os 3 parâmetros em 2 diferentes estágios de convergência considerando ausência de correlação.....	69
TABELA 13. Análises descritivas para os 3 parâmetros em 2 diferentes estágios de convergência considerando correlação média.	72
TABELA 14. Análise descritiva para os 3 parâmetros em 2 diferentes estágios de convergência considerando correlação média.	74
TABELA 15. Análise descritiva para os 3 parâmetros em 2 diferentes estágios de convergência considerando ausência de correlação.	78
TABELA 16. Resultados do Critério de Raftery & Lewis para uma amostra mínima de 3746 iterações aplicado para as 3 situações de correlação.	80
TABELA 17. Resultados do Critério de Geweke para as 3 novas amostras, considerando 5% de significância para o teste Z.	81
TABELA 18. Resultado do Critério de Heidelberger & Welch para as 3 novas amostras.	81
TABELA 19. Análise descritiva e respectivos Erros de Monte Carlo para as 3 novas amostras.	84

LISTA DE FIGURAS

FIGURA 1. Diagrama das possíveis técnicas de inferência Bayesiana.	12
---	----

FIGURA 2. Gráfico de exemplos de curvas de nível para exemplificar uma distribuição a posteriori. Em (a) têm-se parâmetros muito correlacionados e em (b) têm-se parâmetros pouco correlacionados, com uma ilustração do passeio do Amostrador de Gibbs.	24
FIGURA 3. Representação gráfica da série simulada para o modelo com duas intervenções e erro $ARMA(2,2)$	44
FIGURA 4. Representação gráfica do Fator de Redução de Escala Potencial (R) para o parâmetro θ_1 com 30000 iterações realizadas.	52
FIGURA 5. Representação gráfica da estatística do teste Z para o critério de Geweke para as 7 parâmetros até 30000 iterações.	56
FIGURA 6. Representações gráficas das médias ergóticas calculadas a cada iteração para os dois parâmetros mais perturbadores em função do número de iterações.....	57
FIGURA 7. Representação gráfica do resultado do teste da estacionariedade para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações.	58
FIGURA 8. Representação gráfica do resultado do teste de half-width para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações.	60
FIGURA 9. Representação gráfica do “burn-in” proposto para o parâmetro τ realizado a cada 100 iterações até 30000 iterações.	61
FIGURA 10. Representações gráficas do Fator de Redução de Escala Potencial (R) para as 3 variáveis com 5000 iterações realizadas considerando ausência de correlação.	68
FIGURA 11. Representações gráficas das médias ergóticas calculadas a cada iteração para as 3 variáveis com 5000 iterações realizadas para ausência de correlação.....	70

FIGURA 12. Representações gráficas do Fator de Redução de Escala Potencial (R) para as 3 variáveis com 5000 iterações realizadas considerando correlação média	71
FIGURA 13. Representações gráficas do Fator de Redução de Escala Potencial (R) para a variável 1 com um exemplar com escala ampliada do R com 5000 iterações realizadas considerando correlação alta....	73
FIGURA 14. Representações gráficas das médias ergóticas calculadas a cada iteração para as 3 variáveis para amostra de 2501-5000 iterações para alta correlação.	75
FIGURA 15. Representações gráficas da estatística do teste Z para o critério de Geweke para a variável 3 nos 3 níveis de correlação para 5000 iterações.	76
FIGURA 16. Representações gráficas das médias ergóticas das 3 variáveis para a situação de correlação alta.	83
FIGURA 17. Representações gráficas dos critérios multivariados (MPSRF) considerando 2 seqüências. Figura (b) da ênfase na escala do Fator R.	89
FIGURA 18. Representação gráfica dos critérios multivariados (MPSRF) considerando 7 seqüências	90
FIGURA 19. Representação gráfica dos critérios multivariados (MPSRF) considerando 7 seqüências na situação de ausência de correlação.....	92
FIGURA 20. Representação gráfica dos critérios multivariados (MPSRF) considerando 7 seqüências na situação de correlação média.....	93
FIGURA 21. Representações gráficas dos critérios multivariados (MPSRF) considerando 2 (a) e 7 (b) seqüências na situação de correlação alta.....	94

1 INTRODUÇÃO

Um assunto de grande interesse na atualidade é a inferência Bayesiana, referindo-se a um modo de se fazer estatística criado muito antes da análise que é utilizada atualmente. A inferência Bayesiana ficou durante um grande período esquecida devido à necessidade de se calcularem analiticamente integrais complicadas. Muitas vezes estas eram inviáveis de serem feitas de forma algébrica. Alguns métodos aproximados de inferências surgiram, mas apresentavam algumas limitações quanto à complexidade dos modelos. Por volta da década de 60 a análise Bayesiana ressurgiu, mas somente em 1990, com o trabalho de Gelfand & Smith (1990) utilizando Cadeias de Markov, foi utilizada para solucionar este problema de uma maneira alternativa. Seus resultados foram excelentes, atraindo, assim, outros pesquisadores. A primeira grande classe de esquemas de simulação estocástica utilizando Monte Carlo via Cadeias de Markov, genericamente conhecida, foi a do Amostrador de Gibbs. Os métodos de Monte Carlo via Cadeias de Markov têm sido estudados com aplicações em diversas áreas, mas uma das maiores dificuldades é a verificação empírica da convergência da cadeia (seqüência) à distribuição de equilíbrio. Existem várias técnicas que auxiliam na identificação e monitoração da convergência da seqüência.

A inferência Bayesiana tem ganhado cada vez mais espaço, sendo utilizada em inúmeros trabalhos, mas os resultados são muitas vezes questionados, pois muitos deles não utilizam, ou mesmo não abordam com clareza, os critérios implementados para verificação da convergência. Outro questionamento feito também seria o de que, em casos de modelos complicados,

esta exige um esforço computacional grandioso, necessitando, assim, de máquinas poderosas para computar os processos iterativos das Cadeias de Markov. Esses esforços podem ser minimizados com a monitoração da convergência, evitando que as iterações, aquém e além das necessárias, sejam executadas.

Vários critérios para monitorar a convergência existem. Cowles & Carlin (1996) classificaram vários testes quanto à facilidade de serem implementados, aos problemas específicos e à facilidade de interpretar os resultados. A partir destes diagnósticos escolheu-se trabalhar com os testes univariados de Gelman & Rubin (1992a), Raftery & Lewis (1992a), Geweke (1992), Heidelberger & Welch (1983) e o multivariado de Brooks & Gelman (1998) por apresentarem facilidade de implementação, interpretação e enorme utilização em vários procedimentos de amostragem.

Dentre os critérios tem-se aquele para o caso multivariado apresentado por Brooks & Gelman, sendo uma generalização do critério univariado. (seção 2.13.4). Brooks & Gelman desenvolveram um critério que averiguava a maior variabilidade do sistema p-dimensional através do maior autovalor. A questão é que o sistema pode apresentar outros parâmetros que são importantes, não somente o de maior variabilidade, e que este critério não considera. A partir das idéias e problemas apresentados por Brooks & Gelman (1998), estar-se-á discutindo sobre uma forma de cálculo para evitar os problemas e apresentar-se-ão dois critérios alternativos ao original que consideram toda a variabilidade do sistema.

Na seção 2.1 será descrito a distribuição normal multivariada cuja propriedade será utilizada posteriormente para caracterização do segundo modelo. A seção 2.2 trata da apresentação da maximização de formas quadráticas e a seção seguinte (seção 2.3) trata da maximização de pares de

formas quadráticas, ambas necessárias para melhor entendimento do critério de Brooks & Gelman (1998) e dos dois critérios propostos.

Na seção 2.4 é apresentado um breve resumo sobre séries temporais, dando ênfase a um modelo em especial, o modelo de intervenção com erros autorregressivos médias móveis (*ARMA*).

Na seção 2.5 são apresentados conceitos básicos sobre a inferência Bayesiana e as distribuições condicionais completas do modelo de intervenção com erro *ARMA*(p,q). Nas seções seguintes são apresentados os tópicos relacionados ao entendimento dos processos estocásticos via Amostrador de Gibbs e nas seções posteriores, suas características.

A verificação informal de convergência é apresentada na seção 2.12. Os programas utilizados são apresentados no Anexo B.

Os objetivos deste trabalho foram: propor uma forma alternativa de aplicação combinada de quatro critérios univariados de convergência das seqüências de Markov, otimizando suas propriedades; apresentar uma forma alternativa de cálculo do critério de Brooks & Gelman (1998); e propor dois novos critérios multivariados de monitoramento da convergência.

2 REFERENCIAL TEÓRICO

2.1 Distribuição Normal Multivariada

Devido aos fenômenos serem estudados a partir de dados coletados ou mensurados em muitas variáveis, os métodos estatísticos delineados para obter informações a partir destes conjuntos de variáveis são denominados de métodos de análise multivariados. A partir deste conceito pode-se ter a generalização de densidades multivariadas; como exemplo tem-se a distribuição normal, que desempenha um papel fundamental. Segundo Johnson & Wichern (1998), pode-se escrever a distribuição normal multivariada da seguinte forma:

$$f(\underline{x}; \underline{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu}) \right\},$$

sendo p o número de variáveis, $\underline{x}$ o vetor de variáveis observadas ($1 \times p$), $\underline{\mu}$ um vetor ($p \times 1$) dos valores esperados do vetor $\underline{x}$; Σ representa a matriz de covariâncias ($p \times p$) do vetor $\underline{x}$ e T , o símbolo indicador de transposição.

Segundo Bock (1975), uma propriedade da distribuição normal multivariada é que todos os subconjuntos de $\underline{x}$ têm distribuição também normal

multivariada. Tomando-se uma partição ${}_p \underline{X}_1 = \begin{bmatrix} {}_q \underline{X}_1 \\ (p-q) \underline{X}_1 \end{bmatrix} = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}$ e suas

correspondentes partições no vetor de média e covariância, tem-se:

$${}_p\mu_1 = \begin{bmatrix} {}_q\mu_1 \\ {}_{(p-q)}\mu_1 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \text{ e } \Sigma = \begin{bmatrix} \Sigma_{11q} & \Sigma_{12(p-q)} \\ \Sigma_{21q} & \Sigma_{22(p-q)} \end{bmatrix}, \text{ sendo } X_1 \sim N_p(\mu_1, \Sigma_{11}),$$

$X_2 \sim N_p(\mu_2, \Sigma_{22})$ para $q < p$, sendo Σ_{11q} e $\Sigma_{22(p-q)}$ matrizes de covariâncias de X_1 e X_2 , respectivamente.

Com esta mesma partição a distribuição condicional de $X_1 | X_2$ pode ser dada por:

$$X_1 | X_2 \sim N_q(\mu_c, \Sigma_c), \quad (2.1)$$

sendo o vetor de média $\mu_c = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ e a matriz de covariâncias dada por $\Sigma_c = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

2.2 Maximização de formas quadráticas

Seja $L = \underline{a}^T M \underline{a}$ uma forma quadrática, sendo M uma matriz simétrica real e $\underline{a}$ um vetor do tamanho de M . Pode-se estar interessado em maximizar a forma quadrática L com respeito a $\underline{a}$ condicionada ao comprimento de $\underline{a}$, ou seja, que $\underline{a}$ tenha comprimento unitário. Desta maneira, pode-se transformar L em:

$$\lambda = \frac{\underline{a}^T M \underline{a}}{\underline{a}^T \underline{a}},$$

para toda matriz M simétrica.

O máximo pode ser obtido derivando a função em relação a $\tilde{a}$ e igualando a zero.

$$\left(M - \frac{\tilde{a}^T M \tilde{a}}{\tilde{a}^T \tilde{a}} I \right) \tilde{a} = 0,$$

substituindo o valor de λ tem-se, para qualquer i :

$$(M - \lambda_i I) \tilde{a}_i = 0. \quad (2.2)$$

O sistema homogêneo de equações deverá ter posto incompleto para não possuir somente a solução trivial $\tilde{a} = 0$ e por isso o determinante de $M - \lambda_i I$ deve ser zero.

A equação polinomial de λ é chamada de equação característica de M , sendo λ_i denominado valor característico ou autovalor da i -ésima raiz da equação, sendo $\tilde{a}_i$ denominado de vetor característico de M associado a λ_i ou de autovetor.

Existem alguns métodos propostos na literatura para obter as raízes e vetores característicos de matrizes simétricas. Alguns métodos que merecem destaque são: “The Power Method”, “The Jacobi Method”, “The Tridia Gonalization Method”, “Sturm-sequence Bisection” e “QR Method”. Para maiores detalhes, ver Bock (1975).

2.3 Maximização de pares de formas quadráticas

É comum maximizar a razão entre duas formas quadráticas:

$$\lambda = \frac{\underline{\underline{a}}^T \underline{\underline{M}} \underline{\underline{a}}}{\underline{\underline{a}}^T \underline{\underline{N}} \underline{\underline{a}}}, \quad (2.3)$$

em que $\underline{\underline{N}}$ é uma matriz simétrica positiva definida e não singular.

O máximo é dado a partir da derivada de λ em relação a $\underline{\underline{a}}$, igualando-a a zero, como a seguir:

$$\frac{\partial \lambda}{\partial \underline{\underline{a}}} = 0,$$

então,

$$\begin{aligned} \frac{\partial \lambda}{\partial \underline{\underline{a}}} &= \frac{2\underline{\underline{M}}\underline{\underline{a}}\underline{\underline{a}}^T\underline{\underline{N}}\underline{\underline{a}} - 2\underline{\underline{N}}\underline{\underline{a}}\underline{\underline{a}}^T\underline{\underline{M}}\underline{\underline{a}}}{\left(\underline{\underline{a}}^T\underline{\underline{N}}\underline{\underline{a}}\right)^2} = \\ &= \frac{2}{\underline{\underline{a}}^T\underline{\underline{N}}\underline{\underline{a}}} [\underline{\underline{M}}\underline{\underline{a}} - \lambda\underline{\underline{N}}\underline{\underline{a}}] = \\ &= (\underline{\underline{M}} - \lambda_i \underline{\underline{N}})\underline{\underline{a}}_i = \underline{\underline{0}}. \end{aligned}$$

O sistema homogêneo de equações terá solução não trivial se e somente se $|\underline{\underline{M}} - \lambda \underline{\underline{N}}| = 0$. A solução pode ser obtida considerando $\underline{\underline{N}}^{-1}\underline{\underline{M}}$, mas é complicado obter, em alguns casos, a inversa de $\underline{\underline{N}}$ e obter autovalores da matriz $\underline{\underline{N}}^{-1}\underline{\underline{M}}$, por esta nova matriz não ser simétrica.

A alternativa para isso é obter o fator de Cholesky de $\underline{\underline{N}}$, que é positiva definida e, por meio de transformações lineares, transformar o sistema em um sistema similar ao de uma única forma quadrática do tipo $(\underline{\underline{H}} - \lambda_i \underline{\underline{I}})\underline{\underline{z}}_i = \underline{\underline{0}}$. Os métodos numéricos aplicados para uma forma quadrática são os mesmos que seriam aplicados nesse sistema transformado. O autovalor é invariante à transformação; no entanto, os autovetores $\underline{\underline{a}}_i$ devem ser recuperados do autovetor $\underline{\underline{z}}_i$.

2.4 Séries Temporais

2.4.1 Modelo de intervenção com erro $ARMA(p,q)$

Em alguns casos pode-se deparar com eventos que influenciam diretamente a série causando, por exemplo, mudança de nível em um dado tempo t . Este tipo de evento pode ser considerado na análise e é chamado intervenção. O objetivo da análise de intervenção é avaliar o impacto do evento no comportamento da série. Para exemplificar, seja uma série temporal para a qual identificou-se e estimou-se um modelo de séries temporais. Em um dado instante ocorre um evento independente do fenômeno que originou a série temporal, mas cujos efeitos podem se manifestar sobre a série. A intervenção é um evento externo, cujo efeito deve ser incorporado ao modelo (Pino, 1980).

Segundo Milani (2000), Diaz (1988) descreve o modelo de intervenção com erros autorregressivos médias móveis de ordem p e q , denotado por $ARMA(p,q)$, dado por

$$y_t = X_t^T \beta + e_t, \quad (2.4)$$

sendo, e_t erros correlacionados, tratados como uma série temporal e ajustado um modelo $ARMA(p,q)$, então têm-se:

$$y_t = X_t^T \beta + \frac{1 + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}}{1 - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p}}, \quad t = 1, 2, \dots, n,$$

sendo a_t o resíduo considerado ruído branco, que é uma seqüência de variáveis aleatórias $i.i.d. \sim N(0, \tau^{-1})$ sendo τ a precisão e $\tau^{-1} = \sigma^2$ a variância, e os

resíduos no instante t podem ser dados em termos das observações $y_1, y_2, \dots, y_t$.

$X_t^T = [\xi_{t,1} \ \xi_{t,2} \ \dots \ \xi_{t,w}]$ é uma matriz ($n \times m$) de variáveis binárias, na qual cada elemento é um vetor e w o número de intervenções, e $\beta^T = [\beta_1 \ \beta_2 \ \dots \ \beta_w]$ um vetor de parâmetros da intervenção.

2.5 Inferência Bayesiana

2.5.1 Considerações Iniciais

O problema fundamental da estatística é fazer inferência. Alguns dados são observados e deseja-se fazer inferência sobre uma ou mais características desconhecidas do sistema que teria dado origem a estes dados.

A inferência estatística trabalha na presença de observações y , cujos valores inicialmente incertos são descritos através de uma distribuição de probabilidade $f(y | \theta)$. O parâmetro θ é o que interessa ao pesquisador, e tem um sentido exato dentro da inferência Bayesiana. O pesquisador deve ter alguma informação prévia sobre θ , e esta informação deve ser incorporada à análise.

Para Gelman et al. (1997), a inferência Bayesiana é o processo de ajustar um modelo de probabilidade para um grupo de dados e resumir o resultado por uma distribuição de probabilidade nos parâmetros do modelo e em quantidades não observáveis como predição para novas observações.

A inferência Bayesiana é baseada no conceito de probabilidade subjetiva, que mede o grau de desconhecimento que alguém tem da ocorrência de um determinado evento do espaço amostral. Se existe uma incerteza, a descrição desta incerteza deve ser feita usando a probabilidade subjetiva. Então, a análise Bayesiana descreve toda quantidade desconhecida através da

probabilidade. Ao descrever uma incerteza usando a interpretação subjetivista de probabilidade, deve-se levar em conta o axioma de coerência estabelecido por De Finetti (1964).

2.5.2 Inferência Bayesiana

Ao incorporar uma opinião sobre θ à análise através de uma densidade de probabilidade $P(\theta)$, o Bayesiano está denominando a densidade a priori. A densidade a priori possui este nome por ser a distribuição de probabilidade de θ antes que se observem os dados. O parâmetro θ pode ser um escalar ou um vetor de parâmetros.

A densidade conjunta de um grupo de observações $y_1, \dots, y_n$, examinada como uma função do parâmetro, é denominada função de verossimilhança e é representada por $L(y_1, \dots, y_n | \theta)$, sendo n o número de observações.

Se $P(\theta)$ é a densidade a priori para θ , então a densidade a posteriori de θ é dada pelo Teorema de Bayes, ou seja

$$P(\theta | Y) = \frac{L(Y | \theta)P(\theta)}{\int L(Y | \theta)P(\theta)d\theta}, \quad (2.5)$$

sendo $Y = \{y_1, y_2, \dots, y_n\}$. Como o denominador não depende de θ , pode-se escrever

$$P(\theta | Y) \propto L(Y | \theta)P(\theta),$$

em que $\propto$ representa proporcionalidade.

Pode-se pensar no Teorema de Bayes como um mecanismo de atualização da opinião do estatístico sobre θ . O Teorema de Bayes é um elemento essencial para a análise Bayesiana, pois toda inferência é feita a partir da posteriori.

Por muitos anos, o uso de priori subjetiva gerou uma grande controvérsia para a inferência estatística baseada no Teorema de Bayes. A escolha da distribuição a priori é um problema pertinente na abordagem Bayesiana. Existem dois tipos de priori, a priori informativa e a priori não-informativa. Quando o pesquisador tem alguma informação prévia sobre o que está estudando, ele pode usar uma priori informativa. De todos os aspectos da inferência Bayesiana, a priori informativa é a mais difícil e a que traz maior controvérsia. Pode acontecer que em determinado estudo o pesquisador tenha pouca ou nenhuma informação para se incorporar à priori; quando isto acontece, a distribuição considerada é a não-informativa. Existe igualmente controvérsia quanto à priori não-informativa, pois muitas vezes esta priori faz com que se tenha uma distribuição a posteriori imprópria. A priori não-informativa mais usada é a priori de Jeffreys (Jeffreys, 1939).

Uma grande parte da literatura Bayesiana é dedicada a descobrir distribuições a priori nas quais $P(\theta | Y)$ pode ser facilmente trabalhada.

A representação probabilística de todo o conhecimento incerto é a essência da inferência Bayesiana, seja tal conhecimento relacionado ao futuro, quantidades observáveis ou parâmetros desconhecidos. Os conceitos de priori e posteriori são sempre relativos à observação considerada no momento.

Broemeling (1985) apresentou um diagrama para ilustrar a técnica usada na inferência estatística, que está apresentado na Figura 1.

Os componentes de inferência estatística consistem da informação a priori dos dados amostrais e do cálculo da densidade a posteriori dos parâmetros. A informação a priori é expressa pela densidade de probabilidade $P(\theta)$. A

informação nos dados $y_1, \dots, y_n$, sendo $y_1, \dots, y_n$ uma amostra aleatória de uma população com densidade f , está contida na função de verossimilhança $L(y_1, \dots, y_n | \theta)$, que é a densidade conjunta dos dados amostrais; então a função de verossimilhança é combinada com a densidade a priori de θ , pelo Teorema de Bayes, obtendo-se a densidade a posteriori de θ .

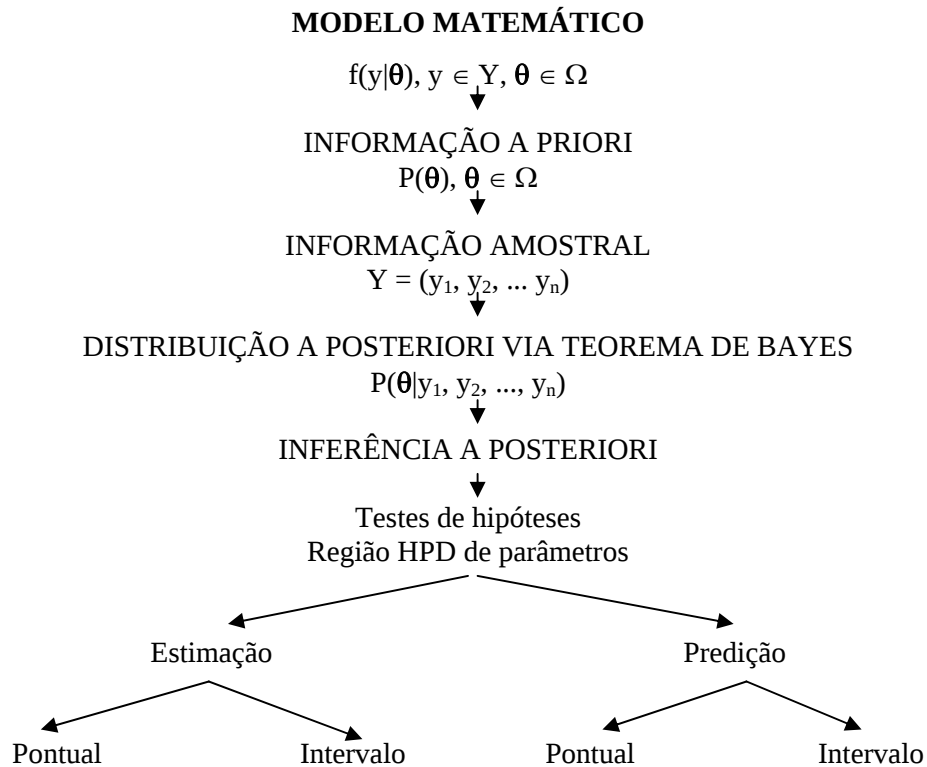


FIGURA 1. Diagrama das possíveis técnicas de inferência Bayesiana.

Broemeling (1985) afirma que a base da inferência Bayesiana é a distribuição a posteriori de θ , pois qualquer conclusão é feita a partir desta distribuição.

A diferença formal entre a inferência Bayesiana e a frequentista é que, para a inferência Bayesiana, o parâmetro θ é variável aleatória, possuindo, então, uma distribuição de probabilidade. E para a inferência frequentista, os parâmetros são valores fixos, não sendo possível atribuir a eles uma distribuição de probabilidade.

O'Hagan (1994) comenta o porquê de usar a inferência Bayesiana ao invés da inferência clássica. Ele diz que, em termos gerais, alguns dos argumentos a favor do enfoque Bayesiano são os de que a análise Bayesiana é fundamentalmente completa, muito flexível, produz conclusões claras e diretas e faz uso de toda informação disponível. Em contraste, o enfoque clássico ignora a informação a priori.

2.5.3 Inferência Bayesiana para o modelo de intervenção com erro ARMA(p,q)

A análise Bayesiana do modelo (2.4) é dada em Milani (2000) e aqui serão usados estes resultados.

Para realizar a análise a posteriori, consideraram-se as distribuições a priori de Jeffreys para os parâmetros $\phi^T = (\phi_1, \phi_2, \dots, \phi_p)$, $\theta^T = (\theta_1, \theta_2, \dots, \theta_q)$, $\beta^T = (\beta_1, \beta_2, \dots, \beta_w)$ e τ , que é a precisão do resíduo a_t , sendo p a ordem da parte referente ao autorregressivo (AR), q a ordem da parte referente ao médias móveis (MA) e w o número de intervenções.

A distribuição a posteriori conjunta para os parâmetros é obtida através do Teorema de Bayes e, segundo Milani (2000), é dada por:

$$P(\phi, \theta, \beta, \tau | Y) \propto \tau^{\frac{n-p}{2}-1} \exp \left\{ -\frac{\tau}{2} \sum_{t=p+1}^n (Z - U^T \beta)^T (Z - U^T \beta) \right\}, \quad (2.6)$$

sendo n o número de observações, p o número de parâmetros autoregressivos, β o vetor de parâmetros da intervenção e os componentes do vetor Z e as linhas da matriz U dadas, respectivamente, por:

$$Z_t = y_t - y^T \phi - \hat{a}^T \theta, \quad U_t^T = X_H^T - \phi^T X_p,$$

em que $y^T = [y_{t-1} \ y_{t-2} \ \dots \ y_{t-p}]$ é o vetor de dados, ϕ o vetor de parâmetros do médias móveis, θ o vetor de parâmetros autorregressivos, $\hat{a}^T = [\hat{a}_{t-1} \ \hat{a}_{t-2} \ \dots \ \hat{a}_{t-q}]$ o vetor de resíduos estimados, $X_H^T = [\xi_{t,1} \ \xi_{t,2} \ \dots \ \xi_{t,w}]$ a matriz de incidência das intervenções e X_p uma matriz dada por:

$$X_p = \begin{bmatrix} \xi_{t-1,1} & \xi_{t-1,2} & \dots & \xi_{t-1,m} \\ \xi_{t-2,1} & \xi_{t-2,2} & \dots & \xi_{t-2,m} \\ \vdots & \vdots & \dots & \vdots \\ \xi_{t-p,1} & \xi_{t-p,2} & \dots & \xi_{t-p,m} \end{bmatrix}.$$

Com o objetivo de utilizar o Amostrador de Gibbs, a ser definido na seção 2.8, determinou-se a distribuição a posteriori condicional completa para cada parâmetro (Milani, 2000).

Distribuição condicional completa para ϕ :

$$\phi | \theta, \beta, \tau, Y \sim \text{Normal}(N^{-1}M, \tau^{-1}N^{-1}) \quad (2.7)$$

sendo $N = N_1 - N_2 - N_3 + N_4$, em que:

$$N_1(i, j) = \sum_{t=p+1}^n y_{t-i} y_{t-j}, \quad N_2(i, j) = \sum_{t=p+1}^n \sum_{k=1}^w \beta_k y_{t-i} \xi_{t-j,k}, \quad N_3 = N_2^T \quad e$$

$$N_4(i, j) = \sum_{t=p+1}^n \sum_{k=1}^w \sum_{l=1}^w \beta_k \beta_l \xi_{t-i,k} \xi_{t-j,l}.$$

Para $M = M_1 - M_2 - M_3 - M_4 + M_5 + M_6$, tem-se:

$$M_1 = \begin{bmatrix} \sum_t y_{t-1} \xi_{t-1} \\ \sum_t y_{t-2} \xi_{t-2} \\ \vdots \\ \sum_t y_{t-p} \xi_{t-p} \end{bmatrix}, \quad M_2 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^q \theta_i y_{t-1} \hat{a}_{t-i} \\ \sum_{t=p+1}^n \sum_{i=1}^q \theta_i y_{t-2} \hat{a}_{t-i} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^q \theta_i y_{t-p} \hat{a}_{t-i} \end{bmatrix}$$

$$M_3 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^w \beta_i y_t \xi_{t-1,i} \\ \sum_{t=p+1}^n \sum_{i=1}^w \beta_i y_t \xi_{t-2,i} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^w \beta_i y_t \xi_{t-p,i} \end{bmatrix}, \quad M_4 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^w \beta_i y_{t-1} \xi_{t,i} \\ \sum_{t=p+1}^n \sum_{i=1}^w \beta_i y_{t-2} \xi_{t,i} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^w \beta_i y_{t-p} \xi_{t,i} \end{bmatrix}$$

$$M_5 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^w \sum_{j=1}^q \beta_i \theta_j \xi_{t-1,i} \hat{a}_{t-j} \\ \sum_{t=p+1}^n \sum_{i=1}^w \sum_{j=1}^q \beta_i \theta_j \xi_{t-2,i} \hat{a}_{t-j} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^w \sum_{j=1}^q \beta_i \theta_j \xi_{t-p,i} \hat{a}_{t-j} \end{bmatrix} \text{ e } M_6 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^w \sum_{j=1}^w \beta_i \beta_j \xi_{t,i} \xi_{t-1,j} \\ \sum_{t=p+1}^n \sum_{i=1}^w \sum_{j=1}^w \beta_i \beta_j \xi_{t,i} \xi_{t-2,j} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^w \sum_{j=1}^w \beta_i \beta_j \xi_{t,i} \xi_{t-p,j} \end{bmatrix}.$$

Distribuição condicional completa para θ :

$$\theta | \phi, \beta, \tau, Y \sim \text{Normal}(C^{-1}P, \tau^{-1}C^{-1}), \quad (2.8)$$

sendo $C(i, j) = \sum_{t=p+1}^n \hat{a}_{t-i} \hat{a}_{t-j}$ e $P = P_1 - P_2 - P_3 + P_4$, em que:

$$P_1 = \begin{bmatrix} \sum_{t=p+1}^n y_t \hat{a}_{t-1} \\ \sum_{t=p+1}^n y_t \hat{a}_{t-2} \\ \vdots \\ \sum_{t=p+1}^n y_t \hat{a}_{t-q} \end{bmatrix}, \quad P_2 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^p \phi_i \hat{a}_{t-1} y_{t-i} \\ \sum_{t=p+1}^n \sum_{i=1}^p \phi_i \hat{a}_{t-2} y_{t-i} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^p \phi_i \hat{a}_{t-p} y_{t-i} \end{bmatrix},$$

$$P_3 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^w \beta_i \hat{a}_{t-1} \xi_{t,i} \\ \sum_{t=p+1}^n \sum_{i=1}^w \beta_i \hat{a}_{t-2} \xi_{t,i} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^w \beta_i \hat{a}_{t-p} \xi_{t,i} \end{bmatrix} \quad \text{e} \quad P_4 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^w \sum_{j=1}^p \beta_i \phi_j \hat{a}_{t-1} \xi_{t-j,i} \\ \sum_{t=p+1}^n \sum_{i=1}^w \sum_{j=1}^p \beta_i \phi_j \hat{a}_{t-2} \xi_{t-j,i} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^w \sum_{j=1}^p \beta_i \phi_j \hat{a}_{t-q} \xi_{t-j,i} \end{bmatrix}$$

Distribuição condicional completa para β :

$$\beta \mid \phi, \theta, \tau, Y \sim \text{Normal}(B^{-1}A, \tau^{-1}B^{-1}), \quad (2.9)$$

sendo $A = A_1 - A_2 - A_3 + A_4 - A_5 + A_6$, em que:

$$A_1 = \begin{bmatrix} \sum_{t=p+1}^n y_t \xi_{t,1} \\ \sum_{t=p+1}^n y_t \xi_{t,2} \\ \vdots \\ \sum_{t=p+1}^n y_t \xi_{t,w} \end{bmatrix}, \quad A_2 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^p \phi_i y_t \xi_{t-i,1} \\ \sum_{t=p+1}^n \sum_{i=1}^p \phi_i y_t \xi_{t-i,2} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^p \phi_i y_t \xi_{t-i,w} \end{bmatrix},$$

$$A_3 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^p \phi_i y_{t-i} \xi_{t,1} \\ \sum_{t=p+1}^n \sum_{i=1}^p \phi_i y_{t-i} \xi_{t,2} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^p \phi_i y_{t-i} \xi_{t,w} \end{bmatrix}, \quad A_4 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^p \sum_{j=1}^p \phi_i \phi_j y_{t-i} \xi_{t-j,1} \\ \sum_{t=p+1}^n \sum_{i=1}^p \sum_{j=1}^p \phi_i \phi_j y_{t-i} \xi_{t-j,2} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^p \sum_{j=1}^p \phi_i \phi_j y_{t-i} \xi_{t-j,w} \end{bmatrix},$$

$$A_5 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^q \theta_i \hat{a}_{t-i} \xi_{t,1} \\ \sum_{t=p+1}^n \sum_{i=1}^q \theta_i \hat{a}_{t-i} \xi_{t,2} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^q \theta_i \hat{a}_{t-i} \xi_{t,w} \end{bmatrix} \quad e \quad A_6 = \begin{bmatrix} \sum_{t=p+1}^n \sum_{i=1}^q \sum_{j=1}^p \theta_i \phi_j \hat{a}_{t-i} \xi_{t-j,1} \\ \sum_{t=p+1}^n \sum_{i=1}^q \sum_{j=1}^p \theta_i \phi_j \hat{a}_{t-i} \xi_{t-j,2} \\ \vdots \\ \sum_{t=p+1}^n \sum_{i=1}^q \sum_{j=1}^p \theta_i \phi_j \hat{a}_{t-i} \xi_{t-j,w} \end{bmatrix}.$$

Para $B = B_1 - B_2 - B_3 + B_4$ tem-se $B_1(i, j) = \sum_{t=p+1}^n \xi_{t,i} \xi_{t,j},$

$$B_2(i, j) = \sum_{t=p+1}^n \sum_{k=1}^w \phi_k \xi_{t,i} \xi_{t-k,j}, \quad B_3 = B_2^T \quad e \quad B_4(i, j) = \sum_{t=p+1}^n \sum_{k=1}^p \sum_{l=1}^p \phi_k \phi_l \xi_{t-k,i} \xi_{t-l,j}.$$

Distribuição condicional completa para τ :

$$\tau | \phi, \theta, \beta, Y \sim \text{Gama} \left(\frac{n-p}{2}, \frac{\sum_{t=p+1}^n (Z_t - U_t^T \beta)' (Z_t - U_t^T \beta)}{2} \right). \quad (2.10)$$

2.6 Métodos Aproximados de Inferência

Para fazer a inferência sobre os parâmetros é necessário determinar as marginais a posteriori. Cada parâmetro terá a sua distribuição marginal, que é obtida integrando a distribuição a posteriori conjunta em relação aos outros parâmetros do modelo. Quando não é possível a obtenção das distribuições

marginais correspondentes analiticamente, podem ser utilizadas técnicas que se baseiam em preceitos analíticos e determinísticos, tais como linearização e aproximação pela normal, aproximação de Laplace e aproximação via quadratura Gaussiana, entre outras. Estas técnicas não iterativas estão limitadas ao número de parâmetros devido à complexidade. Estas formas de cálculos foram muito utilizadas antes do desenvolvimento das técnicas de simulação extensiva computacional (Gamerman, 1996).

Os métodos analíticos citados anteriormente têm erros que diminuem com o aumento do tamanho da amostra de observações. Os métodos baseados em simulação não dependem de quantas observações foram coletadas, mas fornecem aproximações que serão melhores quanto maior for o número de valores gerados.

2.7 Simulação Estocástica via Cadeias de Markov

O avanço da informática propiciou métodos cada vez melhores para obter soluções aproximadas com menores erros que os processos analíticos anteriormente comentados.

O termo simulação refere-se ao tratamento de problemas reais a partir de reproduções em ambientes controlados pelo pesquisador, sendo que, muitas vezes, esses ambientes são os equipamentos computacionais. Alguns problemas apresentam componentes aleatórios, os quais não podem ser descritos de forma exata e são baseados em informações probabilísticas. O processo de simulação é estocástico, ou seja, baseado em distribuições de probabilidades.

Dos métodos de simulação que utilizam cadeias de Markov podem ser citados o Amostrador de Gibbs e o Metropolis-Hasting (Hastings, 1970).

2.8 Amostrador de Gibbs (“Gibbs Sampler”)

O Amostrador de Gibbs é essencialmente um esquema iterativo de amostragem de uma cadeia de Markov cujo núcleo de transição é formado pelas distribuições condicionais completas. É uma técnica para gerar variáveis aleatórias de uma distribuição (marginal) sem que se conheça a sua densidade. A idéia do método de Monte Carlo via Cadeia de Markov (MCMC) é simular um passeio aleatório no espaço do parâmetro θ , o qual converge para uma distribuição estacionária que, em estatística Bayesiana, é a distribuição a *posteriori* $P(\theta|Y)$.

O Amostrador de Gibbs tem sido extremamente útil na resolução de problemas multidimensionais e é definido em termos de subvetores de θ . Seja o vetor de parâmetros θ dividido em p subvetores, $(\theta_1, \theta_2, \dots, \theta_p)$, e suponha-se que as distribuições condicionais de cada subvetor (que pode ser um parâmetro ou um vetor destes) θ_i , dado todos os outros, sejam conhecidas. Essas distribuições são denotadas por $f(\theta_1|\theta_2, \dots, \theta_p, Y)$, $f(\theta_2|\theta_1, \theta_3, \dots, \theta_p, Y)$, ..., $f(\theta_p|\theta_1, \theta_2, \dots, \theta_{p-1}, Y)$, sendo Y o vetor de n observações. Estas distribuições são denominadas distribuições *condicionais completas*.

Segundo Gamerman (1996), o Amostrador de Gibbs pode ser descrito da seguinte forma:

1. dados os valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$ para os parâmetros;
2. gerando $M+N$ grupos de números aleatórios, retira-se iterativamente

$$\theta_1^{(1)} \text{ de } f_1(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, Y),$$

$$\theta_2^{(1)} \text{ de } f_2(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, Y),$$

↓

$$\theta_p^{(1)} \text{ de } f_p(\theta_p|\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{p-1}^{(1)}, Y),$$

obtendo-se na primeira iteração $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_p^{(1)})'$.

3. descartando as primeiras M realizações no passo 2, as N restantes realizações são utilizadas para formar uma amostra aleatória $\{(\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_p^{(i)})_{i=M+1}^{M+N}\}$ e estima-se a posteriori usando esta amostra aleatória.

À medida que o número de iterações aumenta $N \rightarrow \infty$, a seqüência se aproxima de sua condição de equilíbrio, ou seja, ela atinge a característica de estacionariedade com um comportamento assintótico. Assim, assume-se que a convergência é atingida em uma iteração cuja distribuição esteja arbitrariamente próxima da distribuição de equilíbrio, ou seja, a marginal desejada, e não no sentido formal e inatingível do número de iterações tendendo para o infinito. Gamerman (1996) comenta que apesar de os resultados teóricos garantirem a convergência do Amostrador de Gibbs, sua utilização na prática pode ser bastante complicada quando se têm modelos muito complexos. Essa complexidade faz com que a convergência do Amostrador de Gibbs seja de difícil caracterização.

2.9 Autocorrelação e Correlação Cruzada

2.9.1 Autocorrelação

O Amostrador de Gibbs é um processo Markoviano por sustentar uma memória curta, sendo a variável gerada só dependente da variável gerada anteriormente. Desta forma, a série apresenta uma autocorrelação. Para formar uma amostra de uma condicional completa qualquer deve-se esperar o processo alcançar a estacionariedade, que ocorre após B iterações, e assim coletar a partir do b -ésimo valor para formar uma amostra de tamanho $M = N - B$. Como já comentado, esta amostra retida não é independente. O maior problema em se utilizar em amostras com presença de autocorrelação entre os valores é que a

estimativa do erro padrão da média será viesada, pois a variância da média é viesada. Estimadores para esta variância são dados em Carlin & Louis (2000).

Carlin et al. (1998) comentam que as estimativas provindas de um processo MCMC apresentam erros de Monte Carlo que, apesar de pequenos, não devem ser ignorados. O estimador da média dos parâmetros a posteriori é

$$\hat{E}(\theta_p | y) = \bar{\theta}_p = \frac{1}{M} \sum_{i=1}^M \theta_p^{(i)}, \text{ sendo } p \text{ índice do parâmetro correspondente, mas a}$$

$$\text{variância não pode simplesmente ser } S_p^2 = \frac{1}{M-1} \sum_{i=1}^M (\theta_p^{(i)} - \bar{\theta}_p)^2, \text{ pois certamente}$$

será uma estimativa negativamente viesada.

Este problema pode ser amenizado fazendo amostras aproximadamente independentes combinando as seqüências amostrais, ou seja, agrupar as seqüências paralelas iniciadas com valores diferentes, formando uma única amostra (Sorensen, 1996; Carlin & Louis, 2000). Segundo Carlin et al. (1998), esta deve proporcionar uma estimativa para o erro padrão da média razoavelmente boa sempre que a estimativa da média amostral for boa. Outra maneira seria a de calcular as autocorrelações e, assim, testar a partir de qual ordem pode-se assumir independência, com isso retira-se um valor a cada k iterações; este fato é conhecido como pulo ou “thin”. MacEachern & Berliner (1994) dão uma prova simples de que esta aplicação é sempre viesada. Uma outra aplicação simples para se produzirem estimativas melhores de erros padrão seria a estimativa “batching”, ou seja, fazendo grupos. Este procedimento é simples, entretanto também “naive”. Divida uma seqüência longa de comprimento N em m grupos sucessivos de tamanho k cada, sendo as médias dos grupos $B_1, B_2, \dots, B_m$. Assim, tem-se $\bar{B} = 1/m \sum_{i=1}^m B_i$ e a estimativa da variância é dada por:

$$\hat{\sigma}_{\text{batch}}^2 = \frac{1}{m(m-1)} \sum_{i=1}^m (B_i - \bar{B})^2,$$

sendo k grande o suficiente para que a correlação entre grupos seja desprezível e m , grande o bastante para calcular a estimativa da variância com confiança. É importante que as médias dos grupos sejam independentes e isto pode ser verificado se a autocorrelação no “lag” 1 for menor que 0,1. Caso isso não for verdade, deve-se aumentar k e, conseqüentemente, N . Para maiores detalhes sobre o assunto, ver Carlin & Louis (2000). Métodos de séries temporais, apesar de complicados, apresentam melhores resultados pois consideram a correlação existente (Ripley, 1987).

2.9.2 Correlação Cruzada

A correlação cruzada refere-se à correlação entre as condicionais completas. Se a correlação for alta, o passo do amostrador pode ser muito curto, acarretando lentidão no passeio pela distribuição. Com isto, a convergência fica lenta. Este fato está relacionado diretamente com a aceleração do Amostrador de Gibbs. Segundo Gamerman (1996), a reparametrização é uma forma de resolver este tipo de situação, reduzindo o tempo de convergência. Outra maneira seria construir blocos com as variáveis altamente correlacionadas, formando uma única condicional completa para um bloco. Liu et al. (1994) mostraram que estimativas obtidas para componentes de blocos são sempre mais precisas que quando as componentes são tratadas isoladamente, pois pode-se considerar as covariâncias entre as variáveis.

O caso extremo seria de todos os parâmetros em um único bloco, coincidindo com a posteriori conjunta; se fosse possível amostrar da posteriori, então não seria necessário o Método de Monte Carlo. Assim, os blocos devem

ser feitos de maneira a serem obtidas condicionais possíveis de amostrar e, assim, acelerar a convergência (Carlin & Louis, 2000). Na Figura 2 apresenta-se um exemplo de duas condicionais correlacionadas, θ_1 e θ_2 . Pode-se verificar que o passeio fica lento devido à dependência das condicionais.

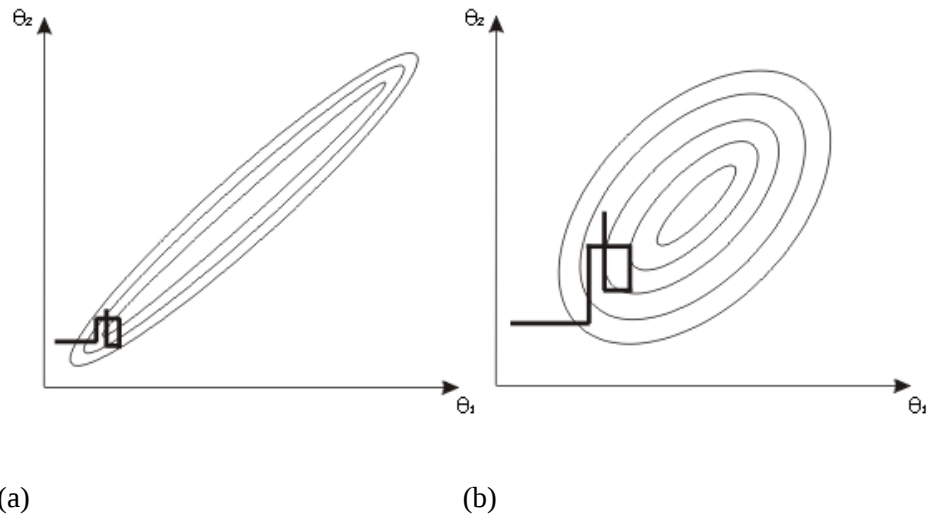


FIGURA 2. Gráfico de exemplos de curvas de nível para exemplificar uma distribuição a posteriori. Em (a) têm-se parâmetros muito correlacionados e em (b) têm-se parâmetros pouco correlacionados, com uma ilustração do passeio do Amostrador de Gibbs.

2.10 Seqüências Múltiplas

Em alguns casos pode-se ter a presença de posteriores multimodais; este fato possibilita ao processo de amostragem ou passeio do amostrador ficar preso em alguma destas modas, o que causa uma certa estabilidade no processo de amostragem, dando a falsa impressão de que este convergiu quando na verdade não conseguiu passear por toda a distribuição. Para contornar esse fato

utilizam-se várias seqüências, partindo de pontos distintos da distribuição a posteriori ou partindo dos máximos locais (Gelman em Carlin et al., 1998). Quando o processo convergir, significará que uma amostra de cada seqüência representa amostras da mesma distribuição. Estas amostras, quando unidas, podem representar uma amostra independente da distribuição de interesse, como comentado na seção anterior.

Alguns critérios utilizam seqüências múltiplas para verificar a convergência. Muitos autores não as utilizam porque elas exigirem muito esforço computacional, apresentando grande desperdício pois grande parte das iterações serão descartadas.

2.11 Valores Iniciais

É importante que os valores iniciais do processo MCMC pertençam ao espaço paramétrico da distribuição; caso contrário, pode-se comprometer a convergência. Gelman & Rubin (1992a) sugerem, ao se utilizarem seqüências múltiplas, iniciar cada uma com valor distinto, facilitando o processo de amostragem. Eles ainda sugerem que possivelmente poderia-se estudar os máximos locais e iniciar uma seqüência de cada um destes, sendo ótimos os valores iniciais. A máxima verossimilhança é um método que poderia ser aplicado neste caso para se acharem estes máximos. Outra maneira seria a de adaptar o Amostrador de Gibbs para buscar o máximo da posteriori mudando passos aleatórios por determinísticos (Gelman, 1995).

2.12 Verificação Informal da Convergência

Gelfand & Smith (1990) sugeriram, como proposta inicial de verificação da convergência, uma forma informal utilizando técnicas gráficas. Eles recomendam processar um número de iterações que se acredita ser suficiente (N); com estes resultados, um histograma é gerado. Em seguida, mais Z iterações são necessárias e em seguida cria-se um novo histograma ($N+Z$). A comparação do gráfico anterior com o último é realizada. Visualmente é verificado se há diferenças; caso não existam, significa que a sequência já estabilizou e o processo convergiu. Similarmente, um outro critério gráfico seria o do acompanhamento das médias ergóticas, doravante comentado (seção 2.12.1), que demonstra um comportamento assintótico à medida que o número de iterações aumenta.

Problemas podem ser verificados com estas técnicas pois a utilização de gráficos como única ferramenta pode não ser adequada. A escala dos gráficos pode fornecer uma falsa impressão de igualdade. Outro ponto discutido é sobre qual quantidade Z seria suficiente. Acredita-se que um número não muito grande seria necessário, pois se existe a suspeita de convergência, esta pode estar próxima, não necessitando, assim, de muitas iterações a mais (Gamerman, 1996).

2.12.1 Ergodicidade

A distribuição da cadeia gerada pelo Amostrador de Gibbs converge para a distribuição estacionária a uma taxa geométrica na iteração i . Esta propriedade é conhecida como ergodicidade (Tierney, 1994). Uma propriedade da ergodicidade da sequência é que as médias ergóticas (2.11) convergem para $E[f(\theta)]$ quando $i \rightarrow \infty$. Se, claro, a esperança existir, em que $f(\theta)$ é uma função

de θ . Embora possa parecer um critério interessante, isto não garante a convergência do Amostrador de Gibbs. Situações de alta correlação entre os parâmetros podem indicar uma falsa convergência, dando ao sistema uma idéia de estabilidade para a média que será falsa. Esta média pode ser calculada da seguinte maneira:

$$\bar{f}_i = \frac{1}{i} \sum_{t=1}^i f(\theta^{(t)}), \quad i = 1, 2, \dots \quad (2.11)$$

2.13 Verificação Formal de Convergência

2.13.1 Critério de Gelman & Rubin (1992a)

O algoritmo MCMC tem produzido um significativo impacto no alcance de problemas para os quais a análise Bayesiana pode ser aplicada (Gilks, Richardson & Spiegelhalter; 1996). O método envolve simulações provindas de uma complexa e, geralmente, multivariada distribuição de interesse, $P(\theta)$, indiretamente gerado de uma Cadeia de Markov com a densidade de interesse, uma densidade estacionária. Geralmente são simuladas $m \geq 1$ seqüências de simulações, cada uma de tamanho n , $(\theta_{j1}, \theta_{j2}, \dots, \theta_{jn})$, para $j = 1 \dots m$ (sendo cada θ_{jt} um vetor). Se $m > 1$, as m seqüências são usualmente, mas nem sempre, simuladas independentemente.

Gelman & Rubin (1992a,b) mostram que em muitos casos a convergência pode ser facilmente determinada provinda de seqüências múltiplas independentes em paralelo, mas não pode ser diagnosticada usando o resultado da simulação provinda de qualquer seqüência simples. Eles propuseram um método que utiliza múltiplas repetições de cadeias para decidir se a

estacionariedade foi alcançada dentro da segunda metade de cada amostra. A popularidade deste método pode ser, em grande parte, devida à simplicidade de sua implementação e ao fato de sua metodologia ter sido extensamente avaliada.

O método pressupõe que m seqüências tenham sido simuladas em paralelo, cada uma partindo de diferentes pontos iniciais. Tendo obtido os pontos iniciais satisfatórios, ou seja, pertencentes ao espaço paramétrico da posteriori, as seqüências são geradas para $2n$ iterações, das quais as primeiras n são descartadas para evitar o período de aquecimento (“burn-in”), ou seja, para evitar a influência dos valores iniciais.

As m seqüências rendem m possíveis inferências. Se estas inferências são bastante similares, é um indicativo de que a convergência foi alcançada ou está próxima. Gelman & Rubin (1992a) sugerem comparar estas inferências produzidas com as obtidas com a união das seqüências, ou seja, com mn valores.

Considere uma variável aleatória θ que tem média μ e variância σ^2 , sob a distribuição de interesse; considere também que se tem um estimador $\hat{\mu}$ não viesado para μ . Dado θ_{ji} , denota o i -ésimo valor de um total de n de θ na seqüência j , toma-se $\hat{\mu} = \bar{\theta}_{..}$, e calcula-se a variação entre seqüência E/n e a variação dentro de seqüência D , definidas por:

$$E/n = \frac{1}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\theta}_{..})^2 \quad (2.12)$$

$$D = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=1}^n (\theta_{ji} - \bar{\theta}_j)^2. \quad (2.13)$$

Note que a análise de variância (ANOVA) assume um “pool” de variâncias total que pode ser produzido porque, sobre convergência, a variação

total das seqüências é igual, de fato, à variação dentro de seqüência, pois a variação entre seqüências se anula.

Após calculadas as estimativas anteriores, pode-se estimar $\hat{\sigma}^2$ pela média de E e D:

$$\hat{\sigma}^2 = \frac{n-1}{n}D + \frac{E}{n}, \quad (2.14)$$

a qual pode ser uma estimativa não viesada da verdadeira variância se os valores iniciais das seqüências provirem da distribuição de interesse. O $\hat{\sigma}^2$ é uma ferramenta que pode levar a falsos diagnósticos de convergência.

Considerando a variância do estimador um “pool” de variâncias:

$$\hat{V} = \hat{\sigma}^2 + \frac{E}{mn}. \quad (2.15)$$

A comparação entre as variâncias é expressa por:

$$R = \frac{\hat{V}}{\sigma^2}, \quad (2.16)$$

o qual chama-se fator de redução de escala (da sigla inglesa SRF). O denominador de R não é conhecido, devendo ser estimado dos dados; pode-se, por isso, ganhar uma superestimativa de R, pois σ^2 será subestimada pelo D. Por isso, utiliza-se uma superestimativa de R, utilizando:

$$\hat{R} = \frac{\hat{V}}{D} = \frac{m+1}{m} \left(\frac{\hat{\sigma}^2}{D} \right) - \frac{n-1}{mn}, \quad (2.17)$$

o qual é conhecido como fator de redução de escala potencial (da sigla inglesa PSRF) e pode ser interpretado como um fator de diagnóstico da convergência. Quando o $\hat{R}$ for grande, sugere-se que qualquer das estimativas das variâncias pode ser reduzida com a utilização de mais iterações, desde que as seqüências simuladas ainda não tenham feito um passeio por toda a distribuição de interesse. Se PSRF aproxima-se de 1, pode-se concluir que cada m conjunto de n iterações se aproxima da distribuição de interesse. A estimativa para D parece ser razoável desde que não estimada com n pequeno, mas será mais acurada à medida que n aumenta. O uso do E também não será muito satisfatório quando se utiliza $m - 1$ graus de liberdade, o qual é em geral muito pequeno. O critério é de fácil interpretação e geralmente fidedigno quando aplicado em situações de aproximada simetria e unimodal, mas pode ser muito falho se aplicado a problemas multimodais (Brooks et al., 1995).

A limitação principal para o método é a suposição de normalidade da distribuição marginal de cada parâmetro. Gelman & Rubin (1992a) sugerem uma transformação apropriada, mas esta não é muito atrativa pois métodos MCMC utilizam freqüentemente distribuições não normais.

Brooks & Gelman (1998) comentam que Gelman & Rubin (1992a) adotaram incorretamente o fator de correção aplicado à variância para a variabilidade amostral, i.e. $d/(d-2)$, em que d representa os números de graus de liberdade. O uso deste fator incorreto conduziu a um número de problemas, em que o corrigido pode ser infinito ou até mesmo negativo. Então, seguindo o método de Fisher (1935), baseado em amostras independentes provindas da distribuição normal, o parâmetro chave para este ajustamento são os graus de liberdade d . Para informar corretamente a variabilidade amostral o fator correto é:

$$\hat{R} = \left(\frac{d+3}{d+1} \right) \frac{\hat{V}}{D}, \quad (2.18)$$

esta correção será importante para pequenas amostras.

Caso o processo indique convergência, deve-se lembrar que os 50% finais da seqüência são valores amostrados da marginal desejada. Gelman (1995) sugere aceitar como garantia de convergência valores de $\hat{R}$ menores que 1,1 ou 1,2.

Em alguns casos pode-se encontrar uma infinidade de parâmetros de interesse para serem monitorados. No entanto, para acompanhamento individual, acaba sendo inviável. Por isso, Brooks & Gelman (1998) estenderam o método para o caso de interesse não somente em funções escalares, mas algo mais geral, como o caso de função multivariada das seqüências.

2.13.2 Critério de Raftery & Lewis (1992a)

O critério sugerido por Raftery & Lewis (1992a) é um método que estima quantas iterações são necessárias para que o Amostrador de Gibbs apresente convergência à distribuição estacionária, propondo especificações de quantidades necessárias para isto. Deve-se ter uma seqüência do Amostrador de Gibbs com um N mínimo, que é o número mínimo de iterações requeridas para se obter a precisão necessária de estimação. O método fornece as estimativas do “burn-in” (B), que é o número de iterações que devem ser descartadas, o número de iterações que devem ser computadas N (total) e o k, que é a distância mínima de uma iteração à outra para se obter a subamostra aproximadamente independente (“thin”). Esses valores serão calculados mediante especificações para garantir que um quantil u de uma determinada função $f(\theta)$ seja estimada com precisão pré definida. Seja um exemplo no qual se deseja obter

$q = P(f(\theta) \leq u | y) = r = 0,025$, sendo u o limite inferior do intervalo de confiança com probabilidade $s = 95\%$. Pode-se desejar que o valor dessa probabilidade seja calculado com $0,01$ e probabilidade $0,99$. Desta forma, o intervalo de confiança de 95% está sendo calculado realmente a posteriori entre 93% e 97% com 99% de confiança.

Uma cadeia binária (0 ou 1) é formada segundo $Z_i = I(f(\theta)_{(i)} \leq u)$, sendo $f(\theta)_{(i)}$ o valor gerado da i -ésima iteração. Pode-se construir novas seqüências tomando-se valores de Z_i a cada k iterações, sendo $Z_i^{(k)} = Z_{1+(i-1)k}$, supondo que as dependências decaem rapidamente com a distância. A aproximação é baseada no modelo de Markov de segunda ordem, ou seja, a dependência é em dois estágios passados. Toma-se k como o menor valor para o qual a dependência em apenas um estágio é preferida.

Seja também

$$\begin{bmatrix} 1-\alpha & \beta \\ \beta & 1-\alpha \end{bmatrix}$$

a matriz de transição de $Z_i^{(k)}$. Com a definição do k , determina-se $B = mk$, sendo

$$m = \frac{\log \left(\frac{\varepsilon(\alpha + \beta)}{\max(\alpha, \beta)} \right)}{\log(1 - \alpha - \beta)},$$

em que ε é a distância requerida entre $p(Z_B^{(k)} = t | Z_0^{(k)} = j)$ e a probabilidade de equilíbrio. Esse valor é escolhido de maneira a garantir a probabilidade de que a seqüência esteja próxima da probabilidade de equilíbrio em um determinado estágio. Então escolhe-se $N = nk$ de maneira a garantir que $\hat{q}$ esteja próximo de

q de acordo com a probabilidade de cobertura definida. O estimador de q é

$$\bar{Z}_{k,n} = \frac{1}{n} \sum_{i=1}^n Z_k^{(m+i)}. \text{ Então}$$

$$n = \frac{\alpha\beta(2-\alpha-\beta)}{(\alpha+\beta)^3} \left[\frac{\Phi\left(\frac{1}{2}(1+s)\right)}{r} \right]^2,$$

sendo $\Phi(\cdot)$ a normal padrão. Para se utilizar o critério de Raftery & Lewis (1992a), basta especificar o quantil q a ser calculado (ex. 0,025), a precisão desejada r (ex. $\pm 0,005$), a probabilidade exigida s de atingir a precisão especificada e uma tolerância em sua estimação. Para se ter uma idéia das ordens de grandezas envolvidas para o caso mais simples de $k = 1$, e desejando-se estimar o quantil 0,025 com erro máximo de 0,0125 a um nível de 95% de confiança, obtém-se $N = n = 600$ (Gamerman, 1996).

A presença de inúmeras variáveis possibilita resultados dessa análise para cada uma delas; assim, deve-se tomar os maiores valores de B , N e k , pois a convergência só se dará quando todas as variáveis convergirem.

Outra saída importante é o Fator de Dependência que é responsável pelo acréscimo multiplicativo ao número de iterações necessárias para se alcançar a convergência, devido à autocorrelação dentro de seqüência. Segundo Raftery & Lewis (1992b), se este fator for maior que 5,0, pode-se dizer que a convergência ainda não foi obtida, necessitando reparametrizar o modelo em estudo.

2.13.3 Análise Espectral

MCMC são especiais casos de séries temporais estacionárias, então é natural usar métodos de séries temporais para testar a convergência. Dois métodos existentes na literatura utilizam técnicas padrão de análise espectral para estimar a verdadeira variância via densidade espectral, $S(w)$.

2.13.3.1 Critério de Geweke (1992)

A ausência de independência entre os dados amostrais é considerada um problema. Na literatura tem-se tratado este problema de modo informal, denotando-o a um problema computacional. Usando técnicas padrão de análise espectral, promove-se um erro padrão para a aproximação $E[g(\theta)]$ das médias amostrais correspondentes; desta maneira, leva-se em conta qualquer autocorrelação em que $g(\theta)$ é uma função real. Com base nestes resultados, um diagnóstico para a ausência de convergência do Amostrador de Gibbs é construído (Geweke, 1992).

Considerando a função real $g(\theta)$ e sua trajetória $g^{(1)}, g^{(2)}, \dots$ construída a partir de $g^{(i)} = g(\theta^{(i)})$, $i = 1, 2, \dots, n$, essa trajetória define uma série temporal. Pode-se, portanto, estimar a média desta série utilizando o seguinte estimador:

$$\bar{g}_i = \left(\frac{1}{n} \right) \sum_{i=1}^n g^{(i)}.$$

Ela fornece um estimador não viesado de $E[g(\theta)]$ cuja variância assintótica é dada por $S_g(0)/n$, sendo $S_g(w)$ a densidade espectral da série g de frequência 0. Em geral, S_g é desconhecido sendo estimado por $\hat{S}_g$ com base em análise

espectral. Após um número suficiente grande, n , de iterações, determina-se as médias $\bar{g}_A$ e $\bar{g}_B$, sendo $\bar{g}_A$ a média das primeiras n_A e $\bar{g}_B$, a média das n_B últimas iterações. Considerando, também, os estimadores independentes $\hat{S}_g^A(0)$ e $\hat{S}_g^B(0)$ das variâncias assintóticas de $\{g^{(i)} : i = 1, \dots, n_A\}$ e $\{g^{(i)} : i = n^*, \dots, n_B\}$, respectivamente, sendo $n^* = n - n_B + 1$, pode-se mostrar que se n_A/n e n_B/n são fixos, com $(n_A + n_B)/n < 1$, quando $n \rightarrow \infty$, tem-se a diferença padronizada entre as médias,

$$\frac{\bar{g}_A - \bar{g}_B}{\sqrt{(\hat{S}_g^A(0)/n_A) + (\hat{S}_g^B(0)/n_B)}} \sim N(0,1), \quad (2.19)$$

se a sequência $\{g^{(i)}\}$ for estacionária. Assim, um teste para uma sequência pode ser construído. Se a diferença padronizada entre as médias for grande, existe indicação de ausência de convergência. As médias devem ser construídas após algumas iterações iniciais terem sido descartadas, e que sejam usados os valores de $n_A = 0,1n$ e $n_B = 0,5n$ (Geweke, 1992).

Na prática, deve-se notar que esta tentativa de diagnóstico para verificar uma necessária, mas não suficiente condição de convergência, informa ao pesquisador somente se a convergência não foi alcançada, e não se realmente convergiu. Para outras discussões do problema associando análise espectral, veja Ripley (1987).

2.13.3.2 Critério de Heidelberg & Welch (1983)

Heidelberg & Welch (1983) apresentaram um método para determinar o valor inicial de uma sequência discreta de eventos, que é também apropriada para ser utilizada para o diagnóstico de convergência do Amostrador de Gibbs.

Suas idéias são baseadas na teoria Browniana. Inicialmente assume-se que no estado estável tem-se uma covariância ou um processo estacionário fraco. É razoável assumir isto desde que se esteja interessado em seqüências provindas de cadeias de Markov, as quais, conseqüentemente, terão estacionariedade completa. Para testar a hipótese nula de estacionariedade, a partir de uma seqüência de dados $\{g^{(i)} : i = 1, \dots, n\}$ oriundos de um processo não estacionário de desconhecida densidade espectral $S(w)$ utiliza-se a teoria Browniana para fazer esta ligação (ver maiores detalhes em Brooks & Roberts; 1998), e assim avaliar possível superdispersão na presença dos valores iniciais. Pode-se, portanto, utilizar vários testes estatísticos para testar a hipótese de nulidade, como Cramer-von Mises (Von Mises, 1931), a estatística de Kolmogorov-Smirnov (Kolmogorov, 1933) ou a estatística de Schruben's (Schruben et al., 1983), que testam se os valores amostrados para cada variável provêm de uma distribuição estacionária. Se a hipótese nula é rejeitada para um dado valor, o teste é repetido depois de descartadas os primeiros 10% das iterações. Se a hipótese é novamente rejeitada outros 10% são descartados após o descarte dos 10% primeiros. Este processo é repetido até se ter uma proporção de 50% (10%, 20%, ..., 50%). Se a hipótese for novamente rejeitada ou o teste não conseguir ser realizado, isto indica falha da estacionariedade, implicando que é necessário um número maior de iterações. Se o teste for satisfatório, o número inicial de iterações descartadas é indicado como o tamanho do “burn-in”.

O teste de “half-width” é utilizado para verificar se a média estimada está sendo calculada com uma acurácia pré-especificada, sendo utilizada na porção da seqüência que passou no teste de estacionariedade para cada variável em estudo. Calcula-se um intervalo de confiança com 95% para a média, utilizando o desvio padrão assintótico via métodos de séries temporais, ou seja, utilizando densidade espectral. Pelo teste pode-se verificar que a razão entre o $(\text{limite superior} - \text{limite inferior})/2$ e a estimativa da média é menor que a fração

específica pré definida (acurácia), o qual, segundo Heidelberg & Welch (1983), deverá ser de 0,1. O resultado passará no teste se isto for verdadeiro, indicando que a média amostral é estimada com um erro relativo aceitável, portanto julgada ser a média a posteriori, sendo a amostra realizada da distribuição de interesse. Se o teste falhou, é necessário um número maior de iterações. Para evitar problemas de dados seqüenciais, não se deve repetir este teste muitas vezes. Uma maneira de amenizar o problema é acrescentar à nova seqüência no mínimo 50% referente a seu tamanho, ou seja, a seqüência de dados deverá ser multiplicada por um fator $I > 1,5$ (Heidelberg & Welch, 1983).

Heidelberg & Welch (1983) sugerem este processo iterativo como estimativa do tamanho do “burn-in”.

O diagnóstico tem muito pouco poder para detectar a falta de convergência se a amostra é de tamanho muito pequeno, podendo ser problema também em obter estimativas fidedignas da densidade espectral, o que ocorre com o método de Geweke.

2.13.4 Critério multivariado de Brooks & Gelman (1998)

O critério de Brooks & Gelman (1998) é a extensão multivariada para o critério de Gelman & Rubin (1992a)(equação 2.17). Por analogia, quando se está interessado na estimação de um vetor de parâmetros $\underline{\theta}$ com base nas observações $\theta_{ji}^{(p)}$, sendo $\theta_{ji}^{(p)}$ o p-ésimo elemento do vetor de parâmetros da seqüência j na iteração i . Para grandes dimensões deve-se estimar a matriz de covariâncias a posteriori:

$$\hat{V} = \frac{n-1}{n} D + \left(1 + \frac{1}{m}\right) \frac{E}{n}, \quad (2.20)$$

em que:

$$D = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=1}^n (\theta_{ji} - \bar{\theta}_{j.})(\theta_{ji} - \bar{\theta}_{j.})^T \quad (2.21)$$

e

$$\frac{E}{n} = \frac{1}{m-1} \sum_{j=1}^m (\bar{\theta}_{j.} - \bar{\theta}_{..})(\bar{\theta}_{j.} - \bar{\theta}_{..})^T \quad (2.22)$$

são matrizes p -dimensional estimadas de p -variáveis. Pode-se, desta maneira, monitorar ambas as variâncias, $\hat{V}$ e D , em que E e D são as matrizes de covariâncias entre e dentro de seqüências, respectivamente.

Resume-se a distância entre $\hat{V}$ e D com uma medida escalar que deve se aproximar a 1 para verificar se a convergência foi alcançada. Uma maneira de se fazer isto é trabalhar com o máximo da raiz característica, que é o máximo do SRF de qualquer projeção linear de θ .

Seguindo o seguinte lema:

Lema. Para duas matrizes simétricas, não singular e positiva definidas M e N , tem-se:

$$\max_{\tilde{a}} \frac{\tilde{a}^T M \tilde{a}}{\tilde{a}^T N \tilde{a}} = \lambda,$$

sendo λ o maior autovalor da matriz positiva definida $N^{-1}M$. Veja a prova em Mardia, Kent & Bibby (1979).

O máximo é dado a partir da derivada em relação a $\underline{a}$, igualando-a a zero, como comentado na seção 2.3. Substituindo as matrizes fictícias pelas matrizes das covariâncias, tem-se:

$$\hat{R}^p = \max_{\underline{a}} \frac{\underline{a}^T \hat{V} \underline{a}}{\underline{a}^T D \underline{a}}, \quad (2.23)$$

que pode ser o multivariado PSRF (MPSRF). O Lema pode ser rescrito por:

$$\hat{R}^p = \frac{n-1}{n} + \left(\frac{m+1}{m} \right) \lambda_1,$$

sendo λ_1 o maior autovalor da matriz positiva definida simétrica $D^{-1}E/n$.

Demonstração:

$$\begin{aligned} \hat{R}^p &= \max_{\underline{a}} \frac{\underline{a}^T \hat{V} \underline{a}}{\underline{a}^T D \underline{a}} \\ &= \max_{\underline{a}} \frac{\underline{a}^T \left[\frac{n-1}{n} D + E \left(\frac{m+1}{mn} \right) \right] \underline{a}}{\underline{a}^T D \underline{a}} \\ &= \frac{n-1}{n} + \left(\frac{m+1}{m} \right) \max_{\underline{a}} \frac{(\underline{a}^T E \underline{a})/n}{\underline{a}^T D \underline{a}} \\ &= \frac{n-1}{n} + \left(\frac{m+1}{m} \right) \lambda_1. \end{aligned}$$

Sobre igualdade de médias entre seqüências, $\lambda_1 \rightarrow 0$ e $\hat{R}^p \rightarrow 1$ quando n for grande o suficiente.

Caso D seja uma matriz singular, não será possível calcular o MPSRF. Se ambas, D e E , são singulares, isto sugere que a razão é indeterminada. Isto pode ocorrer se dois ou mais parâmetros forem correlacionados. Se somente D for o problema, este pode ser contornado à medida que a amostra vai sendo alterada com o aumento das iterações. Outro problema que pode ocorrer é a impossibilidade de se inverter a matriz por esta ser muito grande, sendo um problema essencialmente devido ao conjunto de “hardware” utilizado (Brooks & Gelman, 1998).

2.14 Justificativa

Existe um número grande de métodos na literatura para verificar a convergência. Alguns critérios podem ser classificados como sendo puramente computacionais, outros necessitam de dispendioso gasto computacional e outros são facilmente implementáveis. O número de seqüências exigidas dependerá também do tipo de critério implementado. Outro problema discutido na literatura é qual critério utilizar para aquela devida ferramenta utilizada para se amostrar.

Na Tabela 1 resumiram-se alguns métodos de diagnósticos de acordo com as suas características.

TABELA 1. Resumo de alguns critérios de convergência encontrados na literatura.

Critério	Densidade Conjunta?	Seqüências múltiplas?	Aplicável para quais amostradores?	Problema específico?	Computação extensiva?	Interpretação	Fácil de usar
Raftery & Lewis (1992)	Não	Não	Qualquer	Não	Rápida	Fácil	1
Geweke (1992)	Não*	Não*	Qualquer	Não	Rápida	Fácil	1
Heidelberger & Welch (1983)	Não*	Não*	Qualquer	Não	Rápida	Fácil	1
Ritter & Tanner (1992)	Sim	Sim	GS, MH e híbridos	Sim	Pode ser extensiva	Subjetiva	3
Zellner & Min (1995)	Sim	Não	Somente GS	Sim	Deve ser extensiva	Fácil	4
Garren & Smith (1993)	Sim	Sim**	Qualquer, sob certas condições	Não	Extensiva	Subjetiva, requer experiência	2
Gelman & Rubin (1992)	Não	Sim	Qualquer	Não	Rápida	Subjetiva	1
Johnson (1994)	Sim	Sim	Somente GS	Não	Rápida se as condicionais forem padrão	Escolha subjetiva do parâmetro, mas geralmente fácil	1 ou 3
Liu, Liu & Rubin (1992)	Sim	Sim	Somente GS	Sim	Pode ser extensiva	Subjetiva	3
Roberts (1992, 1994)	Sim	Sim	GS Reversível	Sim	Pode ser extensiva	Subjetiva, requer experiência	3
Brooks, Dellaportas & Roberts (1995)	Sim	Sim	Qualquer	Sim	Pode ser extensiva	Fácil	3

* Pode ser generalizado; ** Valores iniciais iguais; Facilidade de usar: 1 – Código genérico simples está disponível para implementação; 2 - código genérico pode ser escrito e aplicado a saída do MCMC para qualquer problema; 3 - Deve ser escrito o código do específico problema; 4 - É preciso um trabalho analítico e o código para o específico problema. Gibbs Samples (GS) e Metropolis-Hastings (MH). Fonte: Cowles & Carlin (1996) e Brooks & Roberts (1998).

3 METODOLOGIA

Para avaliar os critérios de convergência de Cadeias de Markov foram adotados vários procedimentos, conforme descrição apresentada a seguir. Adotaram-se os critérios de Gelman & Rubin (1992a), Raftery & Lewis (1992a), Geweke (1992), Heidelberger & Welch (1983) e Brooks & Gelman (1998). Justificou-se a escolha desses critérios pelas razões apresentadas na seção 2.8. De maneira geral e sem perda de generalidades, dois modelos foram adotados para avaliação das características dos critérios estudados. Para isso, foram simuladas situações específicas para os valores dos parâmetros. O Amostrador de Gibbs foi utilizado para gerar amostras da distribuição a posteriori conjunta para obtenção das distribuições marginais dos parâmetros. O primeiro modelo considerado envolveu sete parâmetros e o segundo modelo, três variáveis aleatórias. Assim, o primeiro modelo considerado foi um modelo de intervenção com erros $ARMA(2,2)$ e com dois parâmetros para a intervenção. O segundo modelo adotado foi o da normal trivariada. Utilizou-se o Amostrador de Gibbs para gerar amostras de Monte Carlo das 3 variáveis. Procedimentos de geração de amostras de Monte Carlo sem utilizar o Amostrador de Gibbs existem na literatura; para maiores detalhes, ver Johnson & Wichern (1998). Justificou-se o uso do Amostrador de Gibbs devido ao fato de se pretender avaliar e comparar os critérios de convergência em uma amostra com correlação temporal (Cadeias de Markov).

Foram propostos dois novos critérios de convergência multivariados, sendo estes aplicados às amostras dos dois modelos. Comparações entre esses novos critérios e o critério de Brooks & Gelman (1998) foram realizadas.

Os dois modelos considerados e as amostras obtidas pelo amostrador de Gibbs, bem como a sua implementação, são descritos a seguir. Posteriormente, são descritos os critérios de convergência univariados e os multivariados. A simulação foi feita no software SAS[®] v.8.2 (SAS, 2000), no software R v.1.8.0., e também utilizou-se o pacote Bayesian Output Analysis (B.O.A.) (Smith, 2003).

3.1 Análise do modelo com duas intervenções e erro *ARMA(2, 2)*

Foram geradas 500 observações para o modelo, com duas intervenções e erro *ARMA(2, 2)* dado pelo modelo (2.4).

Os valores escolhidos para os parâmetros foram $\phi_1 = 0,5$; $\phi_2 = -0,3$; $\theta_1 = -0,3$; $\theta_2 = 0,5$; $\beta_1 = -30$; $\beta_2 = 20$ e $\tau = 1$. Os valores para os parâmetros foram escolhidos de forma que a série seja estacionária e invertível.

A matriz de variáveis binárias $X_t' = \begin{bmatrix} \xi_{t,1} & \xi_{t,2} \end{bmatrix}$ é dada por

$$\xi_{t,1} = \begin{cases} 0, & \text{se } t \leq 120 \\ 1, & \text{se } t > 120 \end{cases} \quad \xi_{t,2} = \begin{cases} 0, & \text{se } t \leq 300 \\ 1, & \text{se } t > 300 \end{cases}.$$

As intervenções caracterizam mudança de nível na série.

A Figura 3 apresenta a série simulada, na qual fica demonstrado que os parâmetros para a intervenção são $\beta_1 = -30$ e $\beta_2 = 20$.

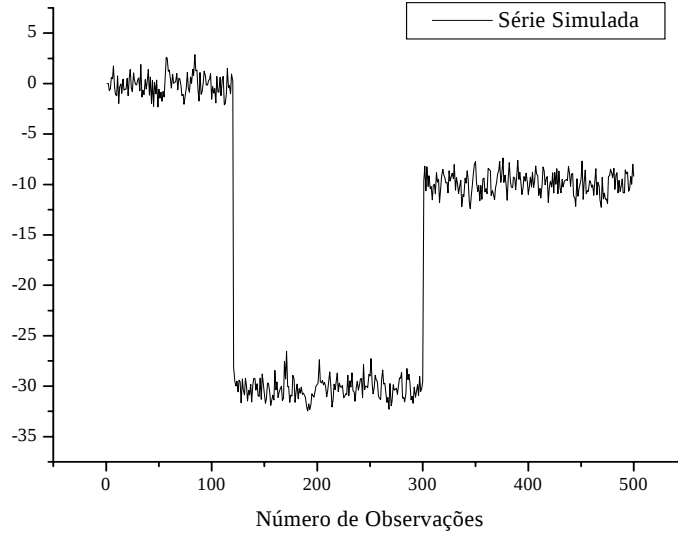


FIGURA 3. Representação gráfica da série simulada para o modelo com duas intervenções e erro $ARMA(2,2)$.

Utilizando as distribuições condicionais completas envolvidas, dadas por (2.7), (2.8), (2.9) e (2.10), pode-se, através do Amostrador de Gibbs, obter as distribuições marginais para os parâmetros ϕ , θ , β e τ , quando o processo alcançar a convergência. Os valores iniciais utilizados para cada seqüência se encontram na Tabela 2.

Para a situação univariada foram utilizadas somente duas seqüências em paralelo, mas para a situação multivariada foram utilizados 2, 3, 5 e 7 seqüências em paralelo.

TABELA 2. Valores arbitrários iniciais utilizados para iniciar cada sequência para o modelo de intervenção com erro ARMA.

Par.	Seqüências						
	primeira	segunda	terceira	quarta	quinta	sexta	sétima
ϕ_1	-0,2	0.6	0.1	0.8	-0.8	0.8	-0.3
ϕ_2	0,6	-0.1	-0.8	1.8	0.1	0.8	0.01
θ_1	-0.4	0.3	0.8	-0.8	1.8	1.6	8.0
θ_2	0.3	0.4	0.2	-0.2	0.8	0.6	-0.8
β_1	-2.5	-1.0	-3.0	-5.0	5.0	1.0	0.2
β_2	2.5	1.5	0.5	-0.5	-1.8	4.8	-0.8
τ	1.5	2.0	0.1	0.8	3.0	0.01	8.0

3.2 Geração de números aleatórios para a normal trivariada utilizando o Amostrador de Gibbs

A normal trivariada foi gerada considerando a propriedade da distribuição normal dada em (2.1). Assim, pode-se generalizar para o caso trivariado utilizando partições alternadas de ${}_3\mathbf{X}_1 = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \mathbf{X}_3 \end{bmatrix}$. As distribuições condicionais são:

$$\mathbf{X}_1 | \mathbf{X}_2, \mathbf{X}_3 \sim N(\mu_1 + \mathbf{A}_1^T \mathbf{B}_1^{-1} \mathbf{C}_1, \sigma_{11} - \mathbf{A}_1^T \mathbf{B}_1^{-1} \mathbf{A}_1),$$

sendo

$$\mathbf{A}_1^T = [\sigma_{12} \quad \sigma_{13}]; \mathbf{B}_1^{-1} = \begin{bmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{bmatrix}^{-1}; \mathbf{C}_1 = \begin{bmatrix} \begin{bmatrix} \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} - \begin{bmatrix} \mu_2 \\ \mu_3 \end{bmatrix} \end{bmatrix}.$$

$$\mathbf{X}_2 | \mathbf{X}_1, \mathbf{X}_3 \sim N(\mu_2 + \mathbf{A}_2^T \mathbf{B}_2^{-1} \mathbf{C}_2, \sigma_{22} - \mathbf{A}_2^T \mathbf{B}_2^{-1} \mathbf{A}_2),$$

sendo

$$A_2^T = [\sigma_{21} \quad \sigma_{23}]; B_2^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{bmatrix}^{-1}; C_2 = \begin{bmatrix} x_1 \\ x_3 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix}.$$

$$X_3 | X_1, X_2 \sim N(\mu_3 + A_3^T B_3^{-1} C_3, \sigma_{33} - A_3^T B_3^{-1} A_3),$$

sendo

$$A_3^T = [\sigma_{31} \quad \sigma_{32}]; B_3^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}^{-1}; C_3 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}.$$

O vetor $\underline{\mu} = [2, 0 \quad 200 \quad 0, 2]$ foi considerado o mesmo para todos os casos. A matriz de covariâncias variou de acordo com as três categorias consideradas: sem correlação (S), correlação média (M) e correlação alta (A) entre as variáveis. As matrizes de correlações e covariâncias, respectivamente, são:

$$\rho_S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ e } \Sigma_S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 100 \end{bmatrix},$$

$$\rho_M = \begin{bmatrix} 1 & 0,50 & 0,52 \\ 0,50 & 1 & 0,48 \\ 0,52 & 0,48 & 1 \end{bmatrix} \text{ e } \Sigma_M = \begin{bmatrix} 1 & 1,58 & 5,20 \\ 1,58 & 10 & 15,18 \\ 5,20 & 15,18 & 100 \end{bmatrix},$$

$$\rho_A = \begin{bmatrix} 1 & 0,98 & 0,98 \\ 0,98 & 1 & 0,99 \\ 0,97 & 0,99 & 1 \end{bmatrix} \text{ e } \Sigma_A = \begin{bmatrix} 1 & 3,10 & 9,70 \\ 3,10 & 10 & 31,31 \\ 9,70 & 31,31 & 100 \end{bmatrix},$$

sendo $\Sigma_i = V_i^{\gamma/2} \rho_i V_i^{\gamma/2}$, em que $V_i = \text{Diagonal}(\Sigma_i)$ e $i = S, M \text{ e } A$.

A justificativa para utilizar 3 categorias é de simular situações de ausência de correlação, correlação média e alta entre as condicionais completas e, assim, acompanhar como os critérios se comportam. Juntamente, foram considerados 3 tipos diferentes de variâncias para estudar como os critérios levariam esta diferença de variação em consideração, visto que em situações normais estas não são conhecidas. Os valores iniciais utilizados estão dispostos na Tabela 3.

TABELA 3. Valores arbitrários iniciais utilizados para iniciar cada seqüência para o modelo normal trivariado.

Var.	Seqüências						
	Primeira	segunda	terceira	quarta	quinta	sexta	sétima
	a						
X1	900	10	-800	100	500	-100	-500
X2	-10000	1000	-0,01	-700	70	-340	70
X3	-3	-50	100	36	6	60	-76

3.3 Critérios Univariados

3.3.1 Critério de Gelman & Rubin (1992a)

Para o critério de Gelman & Rubin utilizaram-se duas seqüências em paralelo. Considerou-se um “burn-in” de 50% da seqüência. O critério foi calculado juntamente com o processo iterativo, sendo realizado pela primeira vez com 20 iterações e repetido a cada 2 até as 30000 iterações para o modelo de séries temporais ($ARMA(2,2)$) e até as 5000 para os três casos do modelo normal trivariado. Os resultados foram dispostos de maneira gráfica, permitindo verificar em qual iteração o processo convergiu.

3.3.2 Critério de Raftery & Lewis (1992a)

O critério de Raftery & Lewis foi calculado com os parâmetros pré-especificados: Acurácia = $\pm 0,02$, Probabilidade = 0,90 e Quantil = 0,025. Para o modelo de séries temporais ($ARMA(2,2)$), inicialmente o critério foi realizado com 5000 iterações; em seguida, com 10000, 15000, 20000, 25000 e 30000. Para o cálculo da amostra piloto foi utilizado a saída do Amostrador de Gibbs na íntegra. Para o modelo normal trivariado foi realizada a amostra piloto para todos os casos com 5000 iterações.

A partir dos resultados do critério de Raftery & Lewis, estes foram utilizados para calcular a amostra ideal. Esta amostra foi exposta a outros critérios, como Geweke, Heidelberger & Welch e ao próprio Raftery & Lewis para confirmação de suas qualidades.

3.3.3 Critério de Geweke (1992)

O critério de Geweke foi realizado utilizando uma seqüência que foi dividida em duas partes, a inicial e a final. Estas partes da seqüência representam cada amostra. A primeira contém 10% das seqüências e a Segunda, os 50% finais das seqüências em estudo. Considerou-se um “burn-in” de 50%. O critério foi calculado juntamente com o processo iterativo, sendo realizado pela primeira vez com 20 iterações e repetido a cada 2 até as 30000 iterações para o modelo de séries temporais ($ARMA(2,2)$) e até as 5000 para os três casos do modelo normal trivariado. Para o cálculo das variâncias oriundas da análise espectral foi utilizado o algoritmo do software R, pacote BOA (Bayesian Output Analysis).

Os resultados foram dispostos de maneira gráfica, juntamente com os limites de significância para a normal padrão. Os limites são 1,96 e -1,96 para 5% e 2,58 e -2,58 para 1%. Estes permitiram verificar em qual iteração o

processo estabilizou, caracterizando aceitação da hipótese de igualdade das médias.

3.3.4 Critério de Heidelberg & Welch (1983)

O critério de Heidelberg & Welch foi calculado também de maneira iterativa, sendo realizado pela primeira vez com 100 iterações e repetido a cada 100 até as 30000 iterações para o modelo de séries temporais e até as 5000 para os três casos do modelo normal trivariado. Para o cálculo das iterações a serem descartadas utilizaram duas partes da seqüência, como utilizado pelo critério de Geweke, e as variâncias oriundas da análise espectral também foram calculadas pelo software R, pacote BOA (Bayesian Output Analysis). Os resultados também foram dispostos de maneira gráfica.

3.4 Critérios Multivariados

Foi utilizado o critério de Brooks & Gelman (1998) para avaliar simultaneamente a convergência dos sete parâmetros no modelo de intervenção com erro $ARMA(2,2)$ e das três variáveis no modelo trivariado normal. O critério foi realizado iterativamente e em conjunto com o Amostrador de Gibbs. A partir de 20 iterações o critério foi calculado pela primeira vez; depois, a cada 2 iterações o critério foi recalculado com o n igual a 22, considerando um “burn-in” de 50%. Para o modelo de séries temporais foram realizadas 3671 iterações com 2 seqüências em paralelo. Para 3, 5 e 7 seqüências em paralelo foram realizadas 5000 iterações. Para o modelo normal trivariado foram realizadas 5000 iterações para 2, 3, 5 e 7 seqüências em paralelo.

Adicionalmente, dois novos critérios foram propostos e aplicados conforme descrição anterior para o critério de Brooks & Gelman (1998).

O desenvolvimento teórico destes critérios, bem como as suas aplicações, serão apresentadas em Resultados e Discussão. Foi também proposto uma nova forma de se aplicar o critério multivariado de Brooks & Gelman para contornar o principal problema de obtenção de autovalores em matrizes não simétricas.

4 RESULTADOS E DISCUSSÃO

Estudaram-se os cinco critérios individualmente para cada modelo; na seção multivariada (4.2), estudaram-se as duas novas propostas em relação ao critério original de Brooks & Gelman (1998). Para concluir foram feitas as considerações finais (seção 4.3).

4.1 Critérios Univariados

4.1.1 Modelo de intervenção com erro *ARMA(2,2)*

4.1.1.1 Critério de Gelman e Rubin (1992a)

O critério de Gelman & Rubin apresentou uma característica para este modelo de sempre tender a 1 com a proximidade da convergência. Isto permite ter uma idéia sobre quão perto ele está da convergência.

Alguns parâmetros se mostraram bem estáveis com relação à escala dos valores de R , apresentando características de convergência próximo das iterações iniciais.

Na Figura 4 tem-se o gráfico do critério de Gelman & Rubin construído para realizações do critério ocorridas a cada 2 iterações para o parâmetro mais problemático (θ_1), no sentido de este estar mais distante da convergência. Os gráficos para os demais parâmetros são apresentados no Anexo A.

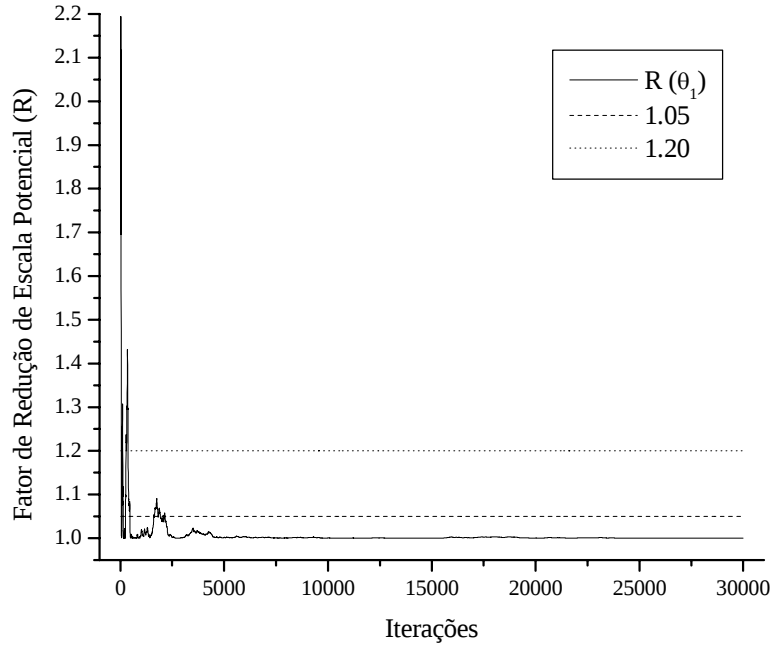


FIGURA 4. Representação gráfica do Fator de Redução de Escala Potencial (R) para o parâmetro θ_1 com 30000 iterações realizadas.

Os parâmetros considerados de maior perturbação para o sistema foram os parâmetros θ_1 e ϕ_1 . O parâmetro τ (Figura A1 (e)) se mostrou estável no processo de amostragem, ou seja, apresentou menor variância, o que foi verificado pelo PRFS, pois apresentou convergência logo de início, com aproximadamente 10 iterações, considerando $R = 1,2$ e $R = 1,05$. Observando a Figura 4, pode-se perceber que o processo realmente só estabilizou com um pouco mais de 5000 iterações, mas estas variações do critério não são significativas, sendo a escala importante para determinar isto. A convergência só será alcançada quando o processo estabilizar para os sete parâmetros, sendo isso verificado com 380 iterações para $R = 1,2$ e 2156 iterações para $R = 1,05$.

Na Tabela 4 estão dispostas as análises descritivas para todos os parâmetros, considerando diferentes pontos de convergência.

TABELA 4. Análise descritiva para os 7 parâmetros em 3 diferentes níveis de convergência para o PRFS.

Iterações	Par.	Média	Erro Padrão da Média ("Naive")	Erro de Monte Carlo	HPD*(95%)		C.I.**
					L.I.	L.S.	
380 iter. (R = 1,2)	θ_1	-0,16599	0,009623	0,01727	-0,4279	0,0574	0,485
	θ_2	0,56533	0,008155	0,01261	0,3680	0,7905	0,423
	β_1	-29,8548	0,003734	0,00438	-29,945	-29,743	0,202
	β_2	19,74254	0,003536	0,00401	19,660	19,839	0,179
	ϕ_1	0,33354	0,009216	0,01625	0,1022	0,5566	0,454
	ϕ_2	-0,25443	0,007474	0,01197	-0,4364	-0,0435	0,393
	τ	0,97746	0,004805	0,00577	0,8731	1,1215	0,248
2156 iter. (R = 1,05)	θ_1	-0,12314	0,004590	0,01210	-0,4211	0,1569	0,578
	θ_2	0,52020	0,003886	0,00941	0,2626	0,7612	0,499
	β_1	-29,8585	0,001446	0,00166	-29,953	-29,768	0,185
	β_2	19,74804	0,001445	0,00166	19,662	19,842	0,180
	ϕ_1	0,28859	0,004447	0,01173	-0,0107	0,5541	0,565
	ϕ_2	-0,20407	0,003937	0,00994	-0,4247	0,0840	0,509
	τ	0,97436	0,001905	0,00169	0,8561	1,1011	0,245
30000 iter.	θ_1	-0,10441	0,000836	0,00300	-0,3852	0,1812	0,566
	θ_2	0,50796	0,000750	0,00252	0,2445	0,7469	0,502
	β_1	-29,8601	0,000283	0,00035	-29,956	-29,763	0,193
	β_2	19,74731	0,000282	0,00034	19,654	19,845	0,191
	ϕ_1	0,27146	0,000802	0,00288	-0,0105	0,5349	0,545
	ϕ_2	-0,18854	0,000757	0,00270	-0,4337	0,0770	0,511
	τ	0,97429	0,000360	0,00039	0,8543	1,0989	0,245

*"Highest Probability Density Intervals"; ** Comprimento do Intervalo

O critério de Gelman & Rubin considera os resultados fornecidos pelo processo de amostragem e por isso estes apresentam autocorrelação. Os usuários interessados em estimar o erro padrão da média não podem deixar de considerar

esta autocorrelação, pois o estimador clássico é viesado. Este viés implicará em intervalos com maior comprimento.

Observa-se que os resultados melhoraram a precisão com o aumento da amostra, o que era esperado. A distribuição de equilíbrio pode ter sido alcançada considerando um $R = 1,2$, mas a região de maior probabilidade (HPD) considerando 95% pode variar um pouco com o aumento das iterações. Um fato importante que merece consideração é o comprimento do intervalo de HPD, este apresentou variação com aumento das iterações, mas o C.I. para 2156 iterações ($R = 1,05$), em comparação ao C.I. com 30000 iterações, apresentou mudanças insignificantes. Para $R = 1,2$, as diferenças foram grandes. Isto ocorreu mesmo com valores do erro de Monte Carlo e erro padrão da média parecidos.

4.1.1.2 Critério de Geweke (1992)

O critério de Geweke é uma condição necessária mas não suficiente para garantir a convergência, ou seja, a não rejeição da hipótese de igualdade entre médias das duas partes da cadeia não é garantia de convergência, embora a rejeição seja um indicativo de não convergência. No presente estudo verificou-se que o critério de Geweke é um critério que se mostrou extremamente rigoroso, caracterizando a convergência para todos os parâmetros simultaneamente somente com um grande número de iterações.

Devido ao critério de Geweke comparar médias das seqüências, pode-se afirmar que sua essência é a estabilização das médias. Por este motivo existe a necessidade das partes da seqüência conterem um grande número de valores para se ter uma alta precisão.

Na Figura 5 está apresentado o teste normal padrão (teste Z), que foi realizado a cada 2 iterações considerando 50% de “burn-in” para os 7 parâmetros. Concomitantemente foram apresentados os limites para caracterizar

a significância do teste. Os limites foram 1,96 e $-1,96$ ($\alpha = 5\%$), 2,58 e $-2,58$ ($\alpha = 1\%$).

O emprego da convergência do critério de Geweke para avaliar a convergência, acompanhada graficamente, foi interessante e se mostrou extremamente importante. Para o caso de se estudar a convergência de maneira pontual, ou seja, fazendo, por exemplo, 5000 iterações e depois aplicando-se o teste, verifica-se que ele fornecerá um resultado que caracteriza a convergência. No entanto, caso as iterações sejam aumentadas para 12500 e se aplique novamente o teste, ter-se-á um resultado que caracteriza a ausência de convergência. Para ilustrar este caso, veja a Figura 5; para este exemplo o critério caracterizou convergência e ao aumentar as iterações, esta não mais foi verificada. Este é um critério com o qual se deve ter muito cuidado ao utilizar para tomar alguma decisão caso se caracterize a convergência. Outro ponto que merece destaque é que este método caracteriza a convergência somente quando ocorre a estabilização da média, sendo, portanto, diferente do critério de Gelman & Rubin.

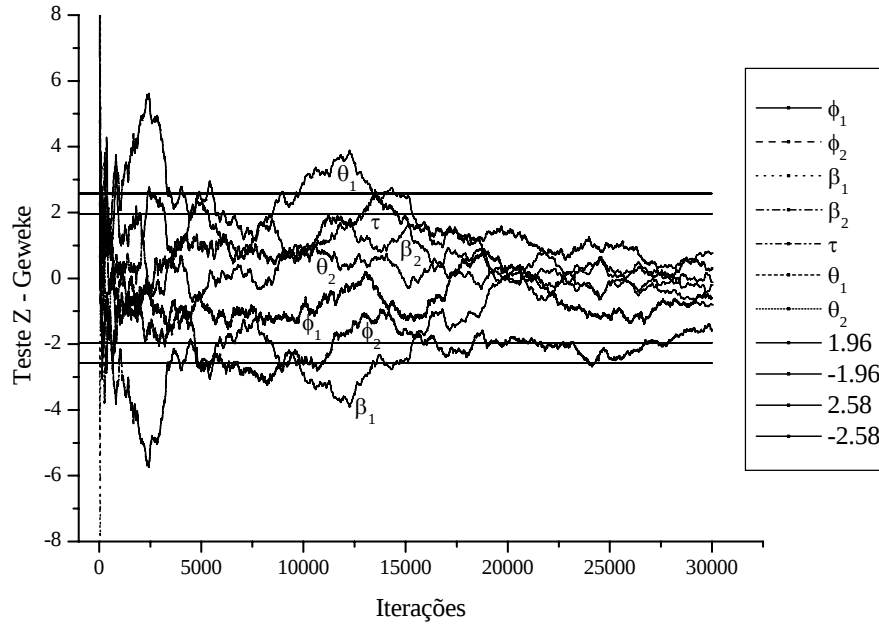


FIGURA 5. Representação gráfica das estatísticas do teste Z do critério de Geweke para os sete parâmetros até 30000 iterações.

Ao analisar mais detalhadamente a Figura 5 observa-se que as linhas retas, para o intervalo de confiança para a diferença de duas médias, foram ultrapassadas várias vezes no decorrer das iterações. As quatro linhas retas caracterizam os limites para aceitação da hipótese para confirmação da convergência. Para afirmar que a convergência ocorreu, as seqüências dos sete parâmetros teriam que estar dentro destes limites. Verifica-se, pelo gráfico, que esta só ocorreu próximo das 27500 iterações. Se for utilizado um rigor maior para rejeição (1%) haverá as duas linhas externas que caracterizam a convergência muito antes, próximo das 15000 iterações, as quais coincidirão com a estabilização das médias ergóticas, segundo as Figura 6a e 6b, que

mostram o comportamento das médias dos parâmetros mais perturbadores do sistema; as demais se encontram na Figura A2.

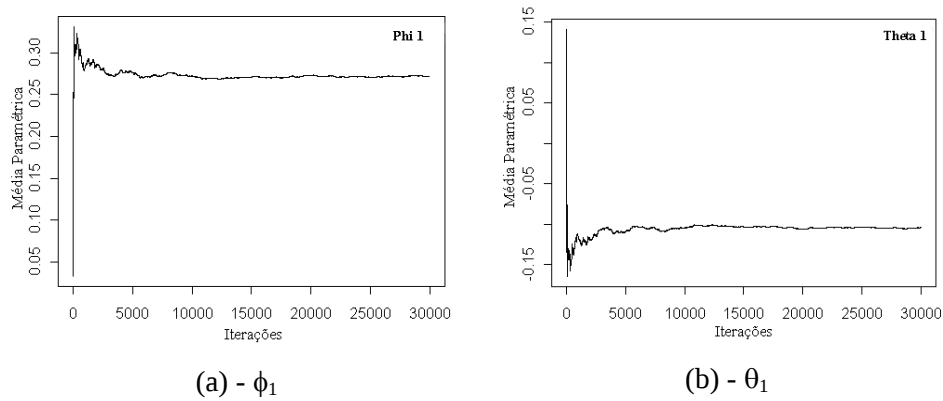


FIGURA 6. Representações gráficas das médias ergóticas calculadas a cada iteração para os dois parâmetros mais perturbadores em função do número de iterações.

4.1.1.3 Critério de Heidelberg & Welch (1983)

O critério de Heidelberg & Welch apresentou resultados que servem para chamar a atenção sobre riscos que se corre ao tirar conclusões precipitadas sobre a convergência. O problema também ocorreu quando se utilizou o critério de Geweke. O critério é mais indicado para informar sobre o tamanho do “burn-in”. O resultado do teste de “half-width” é importante para se ter uma noção sobre a precisão com que a média está sendo calculada.

Na Figura 7 observa-se o teste da estacionariedade disposto de maneira gráfica. Os resultados avaliaram a hipótese da estacionariedade da sequência até uma respectiva iteração. Pode-se verificar que para o parâmetro θ_1 ocorreu rejeição no início das iterações; depois, até a iteração 2900 a hipótese estava sendo aceita como estacionária. Entre 2900 e 3000 o teste rejeitou-a novamente,

sendo, na seqüência, aceita. Este tipo de comportamento permite inferir em resultados pouco confiáveis. Um ponto que se deve considerar é que a partir da iteração 3000 o sistema convergiu para este parâmetro. Uma maneira de saber o ponto em que a hipótese não mais será rejeitada seria a utilização do gráfico para facilitar a visualização, mas que exige que o processo seja repetido com o aumento das iterações. Os parâmetros apresentados na Figura 7, 8 e 9 são considerados suficientes para descrever a perturbação do sistema, não necessitando apresentar os outros parâmetros.

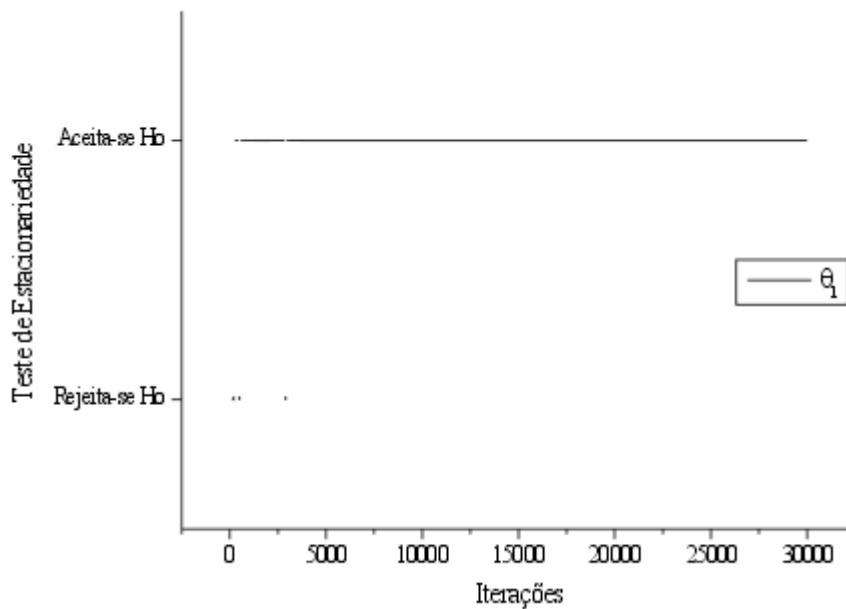


FIGURA 7. Representação gráfica do resultado do teste da estacionariedade para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações.

Na Tabela 5 estão dispostos os resultados do teste da estacionariedade para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações, os quais foram representados de maneira gráfica na Figura 7.

TABELA 5. Resultados do teste da estacionariedade para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações.

Iterações		Resultado do Teste
Início	Fim	
100	200	Rejeita-se H_0
300	400	Aceita-se H_0
500		Rejeita-se H_0
600	2800	Aceita-se H_0
2900		Rejeita-se H_0
3000	30000	Aceita-se H_0

Como descrito na seção 2.7.4.2, juntamente com o teste da estacionariedade é realizado o teste de “half-width”, que testa a acurácia com que o erro padrão da média está sendo calculado. Na Figura 8 está disposto o resultado do teste de “half-width” de maneira gráfica para facilitar a interpretação.

Como comentado, o critério de Heidelberger & Welch indica o valor para o “burn-in” a partir dos resultados do teste da estacionariedade. Os resultados serão aceitos de acordo com o teste da precisão (“half-width”). O critério fornece a estimativa do “burn-in” juntamente com seu complementar que será retido.

Na Tabela 6 estão dispostos os resultados do teste de “half-width” para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações, os quais foram representados de maneira gráfica na Figura 8.

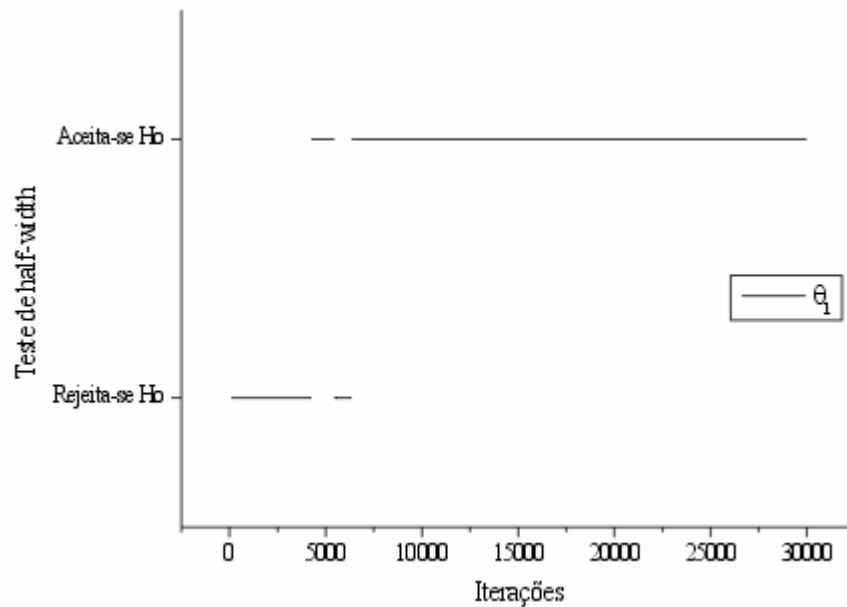


FIGURA 8. Representação gráfica do resultado do teste de “half-width” para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações.

TABELA 6. Resultados do teste de “half-width” para o parâmetro θ_1 realizado a cada 100 iterações até 30000 iterações.

Iterações		Resultado do Teste
Início	Fim	
100	4200	Rejeita-se H_0
4300	5400	Aceita-se H_0
5500	6300	Rejeita-se H_0
6400	30000	Aceita-se H_0

Observando a Figura 8, verifica-se que um dos parâmetros de maior perturbação apresenta resultados de rejeição da hipótese da precisão com que a média deste está sendo calculada e que este pode ser considerado como um indicativo de qualidade da amostra. Este comportamento também implica em resultados pouco confiáveis para um número pequeno de iterações. Na Figura 9

apresentam-se os valores estimados para o “burn-in” à medida que as iterações foram calculadas. O critério apresentou algumas contradições, pois hora indicava a ausência de “burn-in”, para o parâmetro τ , e logo em seguida indicava presença deste. Se se executassem, por exemplo, 13000 iterações para o parâmetro τ , este indicaria ausência de “burn-in”, mas se fossem executadas 13100, o resultado seria de um “burn-in” de 3930; se fossem 13800, não seria mais necessário o “burn-in”.

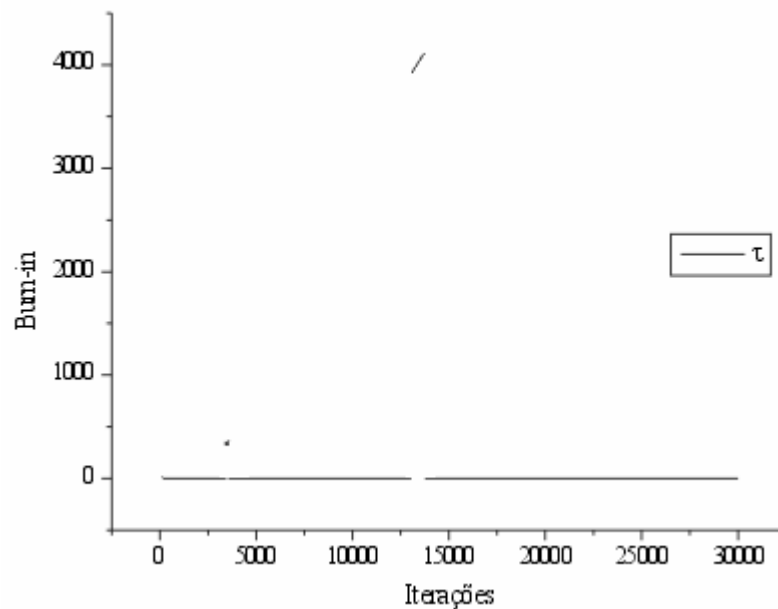


FIGURA 9. Representação gráfica do “burn-in” proposto para o parâmetro τ realizado a cada 100 iterações até 30000 iterações.

Na Tabela 7 estão dispostos os “burn-in” propostos para o parâmetro τ realizado a cada 100 iterações até 30000 iterações, os quais foram representados de maneira gráfica na Figura 9.

TABELA 7. Resultados do “burn-in” proposto para o parâmetro τ realizado a cada 100 iterações até 30000 iterações.

Iterações		Resultados do Critério	Iterações		Resultados do Critério
Início	Fim		Início	Fim	
100		10	13300		3990
200	3400	0	13400		4020
3500		350	13500		4050
3600	13000	0	13600		4080
13100		3930	13700		4110
13200		3960	13800	30000	0

Este tipo de situação exige que o pesquisador tenha um conhecimento mais detalhado sobre o comportamento dos parâmetros.

O critério apresentou variações nos resultados com o aumento das iterações, não sendo indicado para ser realizado de maneira pontual. À medida que o número de iterações aumentou, este critério não mais rejeitou nenhuma das três hipóteses. Segundo este critério um número grande de iterações, implica que não é necessário retirar o “burn-in”.

4.1.1.4 Critério de Raftery & Lewis (1992a)

O critério de Raftery & Lewis para estimação de uma amostra apresentou algumas contradições. Foi realizado o critério em seis amostras piloto diferentes; os resultados para cada uma delas e para cada parâmetro se encontram na Tabela 8.

Os resultados da amostra de tamanho 5000 foram diferentes dos resultados obtidos com a amostra 30000. Cada amostra se comporta de maneira diferente, apesar de provindas do mesmo modelo. O fator de dependência para a amostra de tamanho 30000 apresentou resultados maiores que os da amostra 5000, sendo ambos superiores a 5. Este fato, segundo Raftery & Lewis, é

indicativo de ausência de convergência no processo ou necessidade de se utilizarem reparametrizações.

TABELA 8. Resultados do Critério de Raftery & Lewis para uma amostra mínima de 3746 dados, aplicado em diferentes tamanhos de amostra piloto.

Iterações	Par.	“Thin”	“Burn-in”	Total	Fator de Dependência
5000	β_1	2	4	8990	2,40
	β_2	1	3	4165	1,11
	ϕ_1	6	24	28038	7,48
	ϕ_2	3	9	12408	3,31
	τ	1	2	3777	1,01
	θ_1	6	18	18996	5,07
	θ_2	5	20	21230	5,67
10000 ¹	θ_2	7	28	32466	8,67
15000 ¹	θ_2	8	32	39528	10,55
20000 ¹	ϕ_1	11	33	50281	13,42
25000 ¹	ϕ_1	12	36	61188	16,33
30000	β_1	2	2	7426	1,98
	β_2	2	2	7286	1,95
	ϕ_1	12	36	60612	16,18
	ϕ_2	5	10	19595	5,23
	τ	1	2	3787	1,01
	θ_1	7	21	28693	7,66
	θ_2	10	30	44960	12,00

¹ Para estas amostras o critério de Raftery & Lewis está sendo representado somente pelo parâmetro de maior perturbação do sistema.

Seguindo os resultados apresentados para a amostra de 5000 e de 30000, que estão em *itálico* na Tabela 8, realizou-se uma nova amostra para cada tamanho. Uma com pulo (“thin”) de 6, “burn-in” de 24 e 28038 iterações na totalidade e outra de pulo (“thin”) 12, “burn-in” de 36 e 60612 iterações na totalidade, respectivamente. Os resultados das novas amostras foram testados

com os critérios de Geweke e de Heidelberg & Welch. Juntamente, foi calculada a estatística descritiva e o erro de Monte Carlo.

Na Tabela 9 encontram-se os resultados do critério de Geweke e na Tabela 10, os resultados do critério de Heidelberg & Welch.

TABELA 9. Resultados do Critério de Geweke para as duas novas amostras considerando 5% de significância para o teste Z.

		β_1	β_2	ϕ_1	ϕ_2	τ	θ_1	θ_2
30000 ¹	Teste Z	-0,521	-1,294	-0,084	0,383	-0,608	0,295	-0,835
	p-valor	0,602	0,196	0,933	0,702	0,543	0,768	0,404
5000 ²	Teste Z	1,415	0,134	-0,592	-0,050	2,234	0,656	0,445
	p-valor	0,157	0,893	0,554	0,960	0,026	0,512	0,657

¹ 30000 apresenta amostra real de 5048. ² 5000 apresenta amostra real de 4669.

Pode-se dizer que os resultados de ambas as amostras foram satisfatórios para a amostra vinda de 5000; o parâmetro τ apresentou resultado significativo se for considerado um nível de 5% de significância, sendo um indicativo de ausência de convergência; porém, se for considerado um nível de 1% de significância, este não mais o será. Os demais parâmetros foram não significativos. Para a amostra vinda de 30000 iterações, os resultados confirmaram a convergência do processo pelo critério de Geweke. Pela Tabela 10 verifica-se a não rejeição da estacionariedade para todos os parâmetros em ambos os critérios e também da precisão pelo teste de “half-width”. A única diferença está nas estimativas das médias dos parâmetros, sendo os resultados da amostra 30000 de menor Erro Quadrático Médio 0,008934 em relação a 0,009775 da 5000.

Na Tabela 11 está a estatística descritiva de todas as variáveis, os erros de Monte Carlo e os respectivos intervalos HPD para todos os parâmetros e o Comprimento do Intervalo.

TABELA 10. Resultados do Critério de Heidelberger & Welch para as duas novas amostras.

Iterações	Parâmetro	Teste da Estacionariedade	“Burn-in”	Teste de “Half-width”	Média
30000 ¹	β_1	Aceita-se Ho	0	Aceita-se Ho	-30,01
	β_2	Aceita-se Ho	0	Aceita-se Ho	19,90
	ϕ_1	Aceita-se Ho	0	Aceita-se Ho	0,49
	ϕ_2	Aceita-se Ho	0	Aceita-se Ho	-0,17
	τ	Aceita-se Ho	0	Aceita-se Ho	1,08
	θ_1	Aceita-se Ho	0	Aceita-se Ho	-0,21
	θ_2	Aceita-se Ho	0	Aceita-se Ho	0,36
5000 ²	β_1	Aceita-se Ho	0	Aceita-se Ho	-30,04
	β_2	Aceita-se Ho	0	Aceita-se Ho	19,96
	ϕ_1	Aceita-se Ho	0	Aceita-se Ho	0,34
	ϕ_2	Aceita-se Ho	0	Aceita-se Ho	-0,32
	τ	Aceita-se Ho	0	Aceita-se Ho	0,93
	θ_1	Aceita-se Ho	0	Aceita-se Ho	-0,12
	θ_2	Aceita-se Ho	0	Aceita-se Ho	0,52

¹ 30000 apresenta amostra real de 5048. ² 5000 apresenta amostra real de 4669.

Ambas as amostras apresentaram erros de Monte Carlo pequenos e muito parecidos. Esta não é a única estimativa parecida; todas as outras, média, teste de “half-width”, “burn-in” e teste da estacionariedade, também o foram. Os comprimentos dos intervalos apresentaram valores muito próximos, apresentando a mesma precisão no formato da distribuição paramétrica. Pode-se afirmar que a amostra piloto não necessita ser grande; neste caso, a amostra de 5000 apresentou resultados tão bons quanto a de 30000 e necessitou de menor número de iterações.

TABELA 11. Análise descritiva para as duas novas amostras, Erros de Monte Carlo, intervalo HPD com 95% de probabilidade e Comprimento do Intervalo (C.I.).

Iterações	Par.	Média	SE "Naive"	Erro de Monte Carlo	HPD (0,95)		
					L.I.	L.S.	C.I.
30000 ¹	β_1	-30,011	0,001	0,001	-30,099	-29,918	0,182
	β_2	19,904	0,001	0,001	19,813	19,995	0,182
	ϕ_1	0,488	0,003	0,004	0,118	0,844	0,726
	ϕ_2	-0,167	0,002	0,003	-0,446	0,121	0,567
	τ	1,077	0,001	0,001	0,939	1,209	0,270
	θ_1	-0,206	0,003	0,004	-0,547	0,181	0,728
	θ_2	0,356	0,002	0,002	0,098	0,597	0,498
5000 ²	β_1	-30,038	0,001	0,001	-30,130	-29,939	0,191
	β_2	19,956	0,001	0,001	19,858	20,050	0,192
	ϕ_1	0,337	0,002	0,004	0,020	0,651	0,630
	ϕ_2	-0,321	0,002	0,003	-0,598	-0,040	0,558
	τ	0,934	0,001	0,001	0,820	1,054	0,234
	θ_1	-0,117	0,002	0,004	-0,430	0,210	0,640
	θ_2	0,518	0,002	0,003	0,221	0,776	0,555

¹ 30000 apresenta amostra real de 5048. ² 5000 apresenta amostra real de 4669.

As duas novas amostras foram submetidas novamente ao critério de Raftery & Lewis. Os resultados não foram muito coerentes com as amostras, pois era esperado que o critério afirmasse que ambas as amostras estivessem estacionárias, ou seja, apresentassem resultados insignificantes para “burn-in” e “thin”. Um ponto interessante é que o Fator de Dependência foi menor que 5 para todos os parâmetros, indicando que realmente a convergência ocorreu. Os resultados apresentados para os parâmetros perturbadores foram: pulo de 3, “burn-in” de 12 e 13572 de iterações totais e pulo de 3, “burn-in” de 9 e 12990 de iterações totais para as amostras de 5000 e 30000, respectivamente.

Estes resultados para o número de iterações totais não se mostraram muito coerentes com os obtidos, mas são calculados a partir do fator de dependência.

Para o nosso modelo pode-se dizer que as amostras apresentaram resultados próximos dos valores simulados, sendo um indicativo de convergência para as novas amostras. Utilizar uma amostra piloto grande ou pequena não influenciou os resultados, sendo estes muito parecidos.

4.1.2 Normal trivariada

4.1.2.1 Critério de Gelman & Rubin (1992a)

Para o modelo da normal trivariada na condição de nenhuma correlação entre as variáveis, tem-se o caso extremo de se obterem as próprias marginais e, assim, o processo de amostragem do Amostrador de Gibbs se comportar como um processo de Monte Carlo. Na Figura 10 encontram-se os gráficos do critério de Gelman & Rubin para as 3 variáveis.

Como era de se esperar os gráficos mostram que a convergência é imediata, sendo influenciada somente pelo valor inicial escolhido arbitrariamente do processo. O valor de $R = 1,05$ mostrou-se adequado para verificar a convergência neste caso de independência entre as variáveis. Considerando $R = 1,2$, o processo já convergiria com 20 iterações, pois observando a Figura 10 verifica-se que os valores foram sempre inferiores a 1,2.

De acordo com a Tabela 12 observa-se que com o aumento da amostra as estimativas apresentaram maior precisão, aproximando-se mais dos valores reais. Os Erros Padrão da Média ($S_{\bar{x}}$) diminuíram com o aumento da amostra, o

que era esperado. O que também é esperado é a diminuição do erro de Monte Carlo, que é um bom indicativo de precisão.

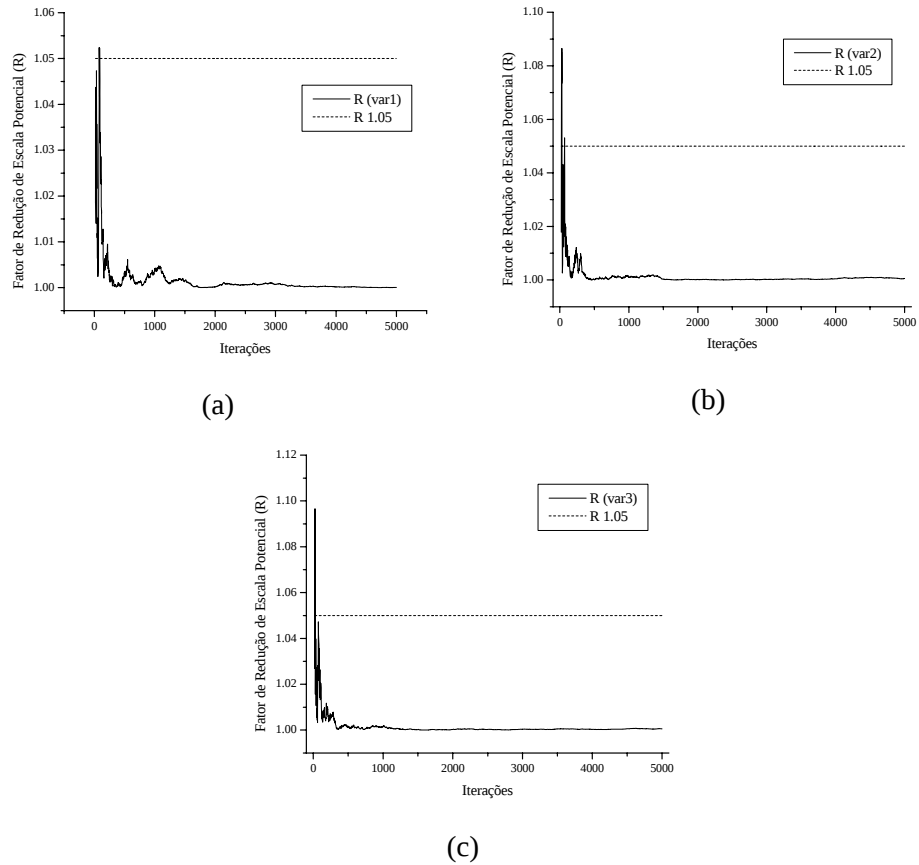


FIGURA 10. Representações gráficas do Fator de Redução de Escala Potencial (R) para as 3 variáveis com 5000 iterações realizadas considerando ausência de correlação.

O processo de convergência é independente, ou seja, pode ou não Ter boas estimativas, pois a convergência pode ter ocorrido, mas se a amostra for pequena, o erro de Monte Carlo será grande. O tamanho da amostra é importante

para reduzir este erro e, assim, obterem-se estimativas mais precisas. Isto fica claro quando se visualizam os gráficos na Figura 11.

TABELA 12. Análise descritiva para os 3 parâmetros em 2 diferentes estágios de convergência considerando ausência de correlação.

Iterações	Variáveis	$\bar{X}$	$S_{\bar{x}}^1$	Quantil 2,5	Quantil 97,5	S^2	Erro de Monte Carlo
90* iter. (R = 1,05)	X1	2,40	0,178	0,000	4,87	1,43	0,087
	X2	199,69	0,465	191,12	204,00	9,72	0,319
	X3	2,18	1,156	-13,62	14,00	60,19	0,858
5000** iter.	X1	1,98	0,021	0,00	4,00	1,13	0,013
	X2	200,03	0,063	194,00	206,00	9,86	0,042
	X3	-0,17	0,199	-20,00	20,00	98,67	0,128

* Contempla uma amostra da iteração 46 a 90. ** Amostra de 2501 a 5000.

¹ Estimador viesado. As médias reais foram 2,0; 200,0 e 0,2 e as variâncias reais foram 1; 10 e 100, respectivamente.

A amostra converge com 90 iterações, mas as estimativas das médias só estabilizaram com aproximadamente 2000 iterações para as variáveis 1 e 2 (Figuras 11a e 11b) e não estabiliza para a variável 3 (Figura 11c), necessitando maior amostragem. É evidente que quando diagnosticada a convergência utilizando o critério de Gelman & Rubin, não necessariamente ter-se-á maior precisão. É uma condição necessária, mas não suficiente.

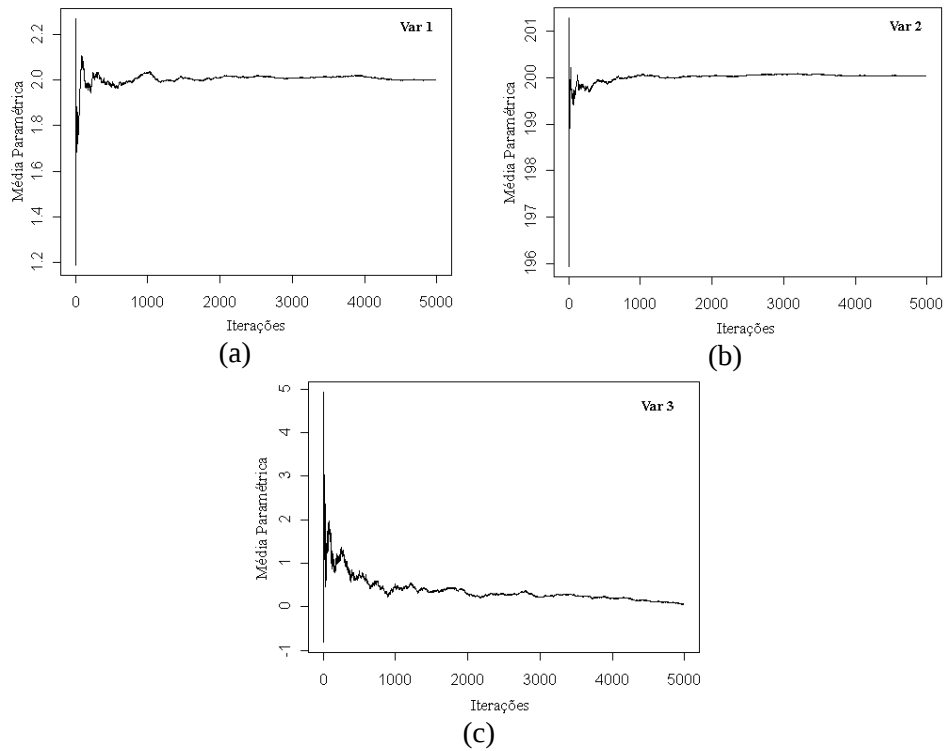
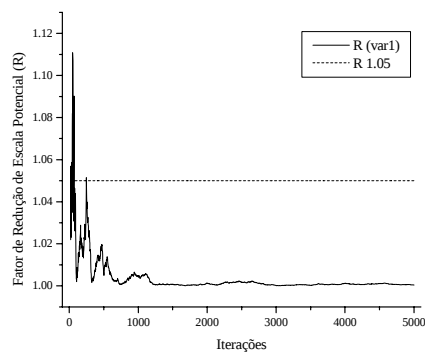


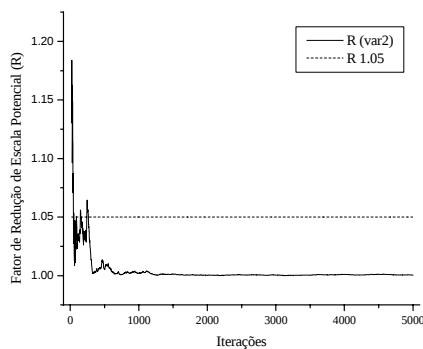
FIGURA 11. Representações gráficas das médias ergóticas calculadas a cada iteração para as 3 variáveis com 5000 iterações realizadas para baixa correlação.

Considerando agora uma correlação média entre as variáveis, pode-se perceber que o critério de Gelman & Rubin convergiu com um número maior de iterações.

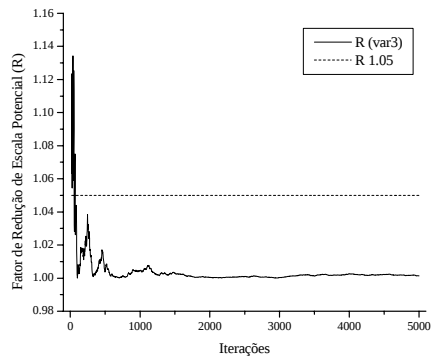
Na Figura 12 estão dispostos os gráficos do critério de convergência para as 3 variáveis considerando correlações médias.



(a)



(b)



(c)

FIGURA 12. Representações gráficas do Fator de Redução de Escala Potencial (R) para as 3 variáveis com 5000 iterações realizadas considerando correlação média.

Como era de se esperar, os gráficos mostram que a convergência é rápida, mas não imediata. O valor do $R = 1,05$ mostrou-se um nível adequado como resultado para a indicação de convergência. Na Tabela 13 estão dispostas as análises descritivas.

TABELA 13. Análises descritivas para os 3 parâmetros em 2 diferentes estágios de convergência considerando correlação média.

Iterações	Variáveis	$\bar{X}$	$S_{\bar{X}}$ ¹	Quantil 2,5	Quantil 97,5	S^2	Erro de Monte Carlo
262* iter. (R = 1,05)	X1	2,4	0,103	0,0	4,75	1,39	0,791
	X2	200,98	0,259	196,0	206,0	8,71	1,281
	X3	2,86	0,916	-17,0	22,0	109,16	4,155
5000** iter.	X1	1,96	0,020	0,0	4,0	1,05	0,027
	X2	199,87	0,063	194,0	206,0	10,05	0,068
	X3	-0,39	0,200	-20,0	19,0	100,50	0,217

* Contempla uma amostra da iteração 132 a 262. ** Amostra de 2501 a 5000.

¹ Estimador viesado.

Observa-se, na Tabela 13, o mesmo comportamento ocorrido para o caso anterior. As estimativas apresentaram maior precisão, aproximando-se mais dos valores simulados à medida que aumenta a amostra. Isto também ocorre para todas as estimativas, o que era esperado. O que também era esperado é a diminuição do erro de Monte Carlo, o qual é um bom indicativo de precisão, mas o que chama a atenção neste caso é a magnitude do erro de Monte Carlo para convergência com $R = 1,05$. A amostra apresentou maior número de observações, mas a dependência é maior, portanto é de suma importância acompanhar o erro de Monte Carlo para se obterem melhores estimativas.

A convergência para este caso foi mais lenta devido à correlação entre as condicionais. A amostra converge com 262 iterações para $R = 1,05$, enquanto, na ausência de correlação, a convergência ocorre com 90 iterações.

Não se pode deixar de lembrar que o critério de Gelman & Rubin é calculado com 50% das iterações finais; por isso, à medida que o processo iterativo caminha, o critério se afasta dos valores influenciados pelo valor inicial arbitrário.

Considerando finalmente uma correlação alta, foi verificado o que era esperado. A convergência se apresentou de maneira lenta, pois o processo de amostragem percorre muito lentamente o espaço paramétrico.

Na Figura 13a está disposto o gráfico do critério de convergência para a variável 1. Na Figura 13b os dados também pertencem à variável 1, mas a escala agora se apresenta diminuída para o R; desta maneira ilustra-se o problema de utilizar a forma gráfica para tirar conclusões. Na Figura 13a poder-se-ia afirmar sem dificuldade que o processo convergiria com menos de 1000 iterações; mas, ao observar a Figura 13b, pode-se perceber que existem oscilações entre aceitar e não a convergência, tanto para $R = 1,2$ quanto para $R = 1,05$.

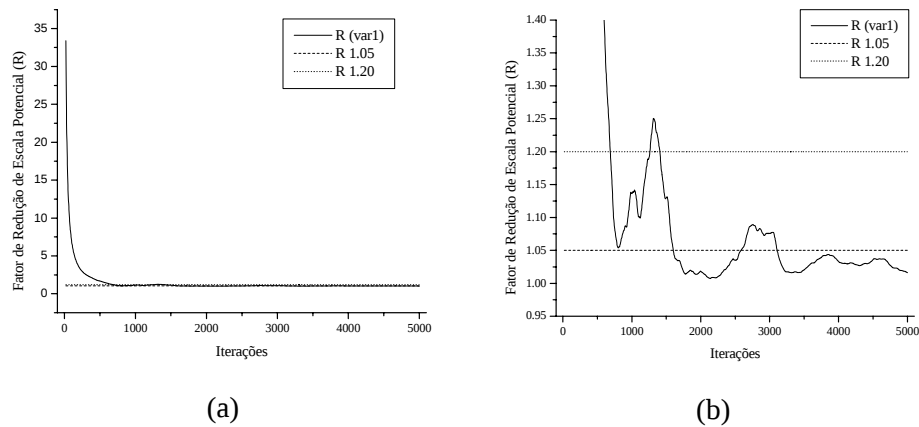


FIGURA 13. Representações gráficas do Fator de Redução de Escala Potencial (R) para a variável 1 com um exemplar com escala ampliada do R com 5000 iterações realizadas considerando correlação alta.

De acordo com a Figura 13 e a Figura A3, pode-se perceber a grande equivalência dos gráficos devido à correlação alta. Referindo-se à convergência, pode-se afirmar que quando $R = 1,2$, tem-se a convergência ocorrendo próxima

das 1500 iterações, mas se for levado em conta um $R = 1,05$, a convergência é caracterizada por volta da iteração 3000.

Na Tabela 14 estão dispostos os resultados para os 3 níveis de indicação da convergência e a análise descritiva.

Apesar de a caracterização da convergência para este caso ter sido verificada, observa-se que a média estimada da variável 3 está muito distante da verdadeira média ($\mu_3 = 0,2$), o que pode ser devido à alta variabilidade que esta variável apresenta.

TABELA 14. Análise descritiva para os 3 parâmetros em 2 diferentes estágios de convergência considerando correlação média.

Iterações	Variáveis	$\bar{X}$	$S_{\bar{x}}^1$	Quantil 2,5	Quantil 97,5	S^2	Erro de Monte Carlo
1408 iter.* ($R = 1,2$)	X1	2,529	0,034	0,2809	3,9826	0,825	0,07732
	X2	201,7	0,108	194,88	206,49	8,269	0,24853
	X3	5,607	0,343	-15,721	20,124	83,133	0,78351
3108 iter.** ($R = 1,05$)	X1	2,193	0,021	0,601	3,758	0,716	0,06336
	X2	200,6	0,067	195,58	205,61	7,052	0,20195
	X3	2,160	0,213	-13,767	18,745	70,791	0,63718
5000*** iter.	X1	2,353	0,018	0,5115	4,3297	0,8334	0,06027
	X2	201,1	0,058	195,43	207,40	8,2945	0,19236
	X3	3,820	0,182	-14,562	23,896	83,292	0,60509

* Contempla uma amostra da iteração 705 a 1408. ** Contempla uma amostra da iteração 1555 a 3108. *** Amostra de 2501 a 5000. ¹ Estimador viesado.

Observando os gráficos das médias ergóticas na Figura 14, as outras duas variáveis também não apresentaram uma caracterização de estabilidade (Figuras 14a e 14b), apesar de seus valores estarem próximos do exato. Esta proximidade talvez se deva à variância que foi atribuída às variáveis 1 e 2, $\sigma_1^2 = 1$ e $\sigma_2^2 = 10$, respectivamente. Era esperado que os valores estimados para

as variâncias estivessem sempre distantes dos valores exatos, mas isso só foi verificado para a situação de correlação alta entre as condicionais.

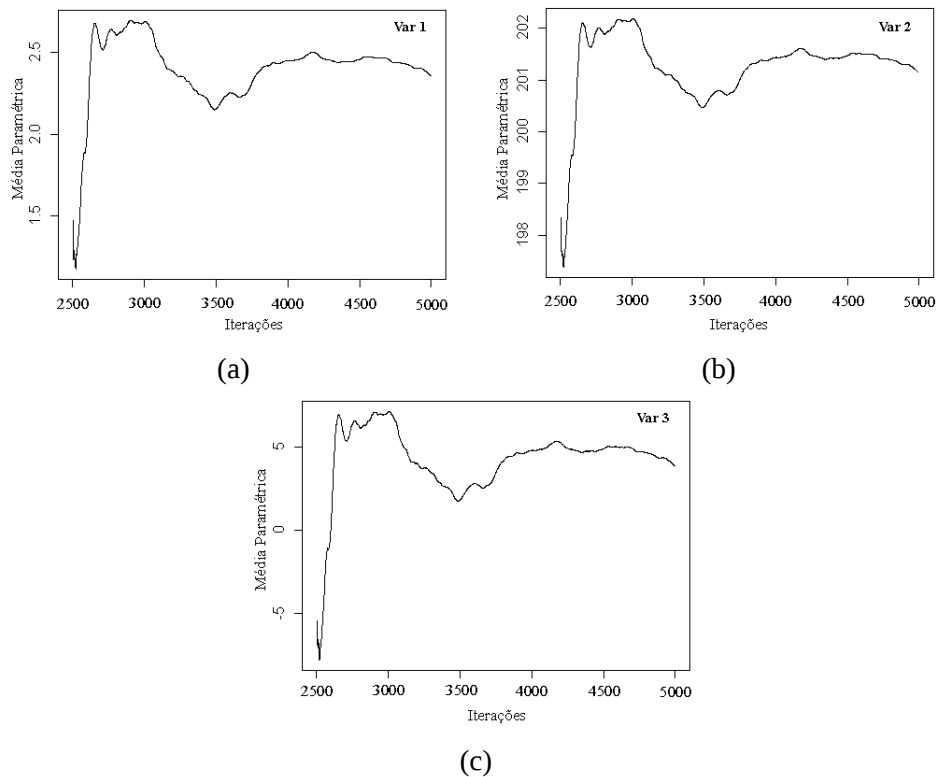


FIGURA 14. Representações gráficas das médias ergóticas calculadas a cada iteração para as 3 variáveis para amostra de 2501-5000 iterações para alta correlação.

4.1.2.2 Critério de Geweke (1992)

O critério de Geweke também se mostrou rigoroso para o modelo normal trivariado, mesmo para a situação de ausência de correlação. Pela Figura 15 verifica-se a necessidade de um número maior de iterações em comparação

ao critério de Gelman & Rubin para a situação de ausência de correlação. Se forem considerados 5% de significância para o teste Z, a convergência será alcançada com aproximadamente 5000 iterações, mas se considerado 1%, esta será verificada com aproximadamente 200 iterações.

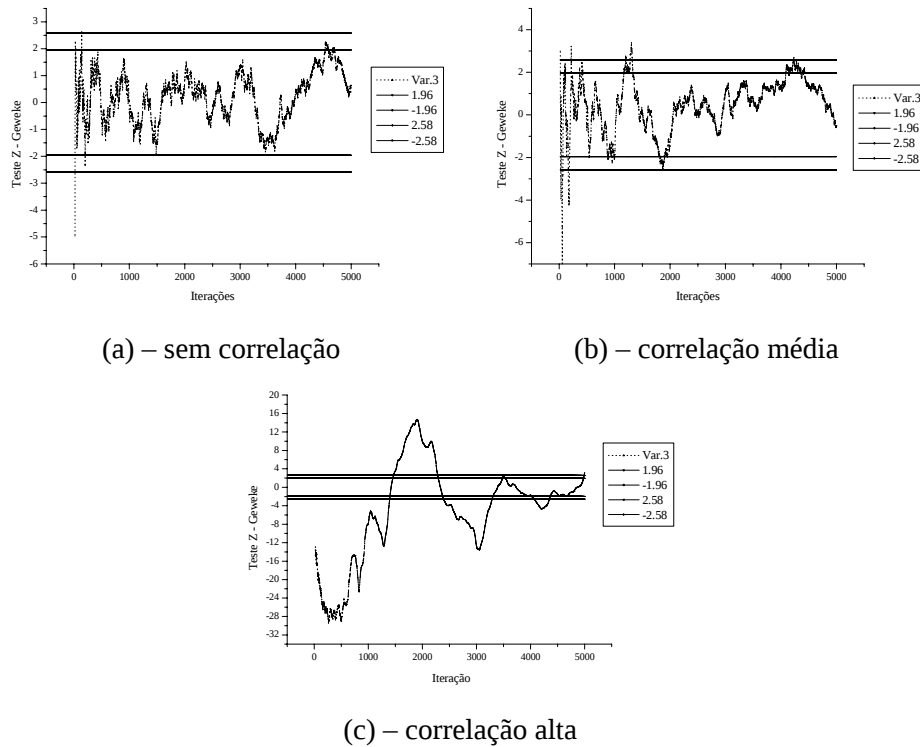


FIGURA 15. Representações gráficas da estatística do teste Z para o critério de Geweke para a variável 3 nos 3 níveis de correlação para 5000 iterações.

Para as situações de ausência de correlação e correlação média, observa-se que as variáveis apresentaram comportamento similar, apresentando convergência para 5% próxima das 4600 iterações. Para a situação de correlação alta pode-se verificar a tendência do critério permanecer dentro das linhas, ou

seja, convergir com o aumento do número de iterações. Para o modelo de séries temporais comentou-se a respeito do comportamento do critério com a igualdade das médias ergóticas dos parâmetros (neste caso das variáveis), podendo ser verificado este fato para a situação de correlação alta (Figura 15c e Figura 14). Esta comparação não é tão clara, pois as médias ergóticas são calculadas desde as primeiras iterações e o critério de Geweke é realizado descartando o “burn-in”. As seqüência das variáveis para a situação de correlação alta apresentaram um comportamento de igualdade entre si, devido a esta correlação, pois o processo de amostragem fica muito dependente. Apesar disto, a Figura 12c apresentou uma característica de tendência em aproximar-se dos limites de significância quando aumenta o número de iterações.

O critério de Geweke é interessante, mas deve ser realizado sempre ao longo das iterações e de maneira gráfica para facilitar a verificação. Outro ponto a ser discutido é sobre o valor a ser adotado para a significância nominal, pois o valor de 5% de significância se mostrou pouco poderoso, não conseguindo confirmar a convergência com poucas iterações. Na realidade o nível de significância global não é 5%, e sim maior. Portanto, aumentar a confiança para 99% se mostrou uma alternativa interessante, pois diminui o nível de significância global e, por isso, aumenta o poder do teste, visto que com 200 iterações para a situação de ausência de correlação, a convergência já está assegurada. As outras variáveis se encontram na Figura A7. Na Tabela 15 as estimativas para 200 iterações e para 5000 iterações são expostas. Pode-se verificar que as diferenças são pequenas.

TABELA 15. Análise descritiva para os 3 parâmetros em 2 diferentes estágios de convergência considerando ausência de correlação.

Iterações	Variáveis	$\bar{X}$	$S_{\bar{X}}^1$	Quantil 2,5	Quantil 97,5	S^2	Erro de Monte Carlo
200 iter.** ($\alpha = 0,01$)	X1	1,990	0,032	0,011	3,985	1,038	0,0313
	X2	200,0	0,099	193,98	206,11	9,939	0,0925
	X3	0,129	0,316	-20,487	19,507	99,615	0,2855
5000*** iter.	X1	1,986	0,02	-0,051	3,932	1,024	0,0191
	X2	200,0	0,062	194,16	206,10	9,780	0,0634
	X3	-0,169	0,199	-19,79	19,72	98,568	0,1849

* Contempla uma amostra da iteração 101 a 200. ** Amostra de 2501 a 5000. ¹ Estimador viesado.

O critério se mostrou extremamente rigoroso, mostrando que a convergência só ocorrerá com um número muito grande de iterações, se comparado ao critério de Gelman & Rubin. O único fato que justificaria este rigor seria de a amostra apresentar um número grande de valores, diminuindo, assim, o erro de Monte Carlo.

4.1.2.3 Critério de Heidelberg & Welch (1983)

Para o modelo da normal trivariada na ausência de correlação, o teste de estacionariedade não rejeitou a hipótese para as 3 variáveis em todas as iterações testadas. O teste de “half-width” não rejeitou a acurácia para as variáveis 1 e 2. Para a variável 3, este rejeitou-a em todo processo iterativo. A variável 3 foi simulada com maior variância, ou seja, 100 vezes maior que a variável 1 e 10 vezes maior que a variável 2. O teste não foi sensível a esta diferença de variabilidade, rejeitando a hipótese de que a média foi estimada com acurácia pré-especificada. Talvez essa hipótese deixe de ser rejeitada para esta situação de alta variabilidade somente com amostras de tamanho muito grande, pois a acurácia é pré-especificada para todos os parâmetros (variáveis)

simultaneamente, quando utilizam-se softwares como BOA e CODA (ambos pacotes do software R). O teste do “burn-in” apresentou total contradição, indicando necessidade de “burn-in” em certos momentos e em outros não.

Para a situação de ausência de correlação o Amostrador de Gibbs, como já comentado, é na verdade um algoritmo de geração de valores de Monte Carlo, ou seja, já está gerando valores da marginal desejada, só influenciada pelo único valor inicial arbitrário, ou seja, deveria apresentar um “burn-in” de tamanho 1 apenas.

Considerando agora a situação de correlação média, os resultados foram ainda mais confusos e inconsistentes. Para o teste da estacionariedade as variáveis foram consideradas estabilizadas até um determinado ponto; assim, a hipótese foi rejeitada por um número de iterações específico; a partir desse ponto, foi não rejeitada. Este tipo de situação compromete a interpretação, pois, como já comentado, se um usuário tivesse simulado um número de 3500 iterações, este poderia tomar uma decisão totalmente diferente da que tomaria se tivesse considerado 4000 iterações. Para o teste de “half-width” o resultado foi parecido com a situação de ausência de correlação. O “burn-in” apresentou a característica de aumentar com o aumento do número de iterações com algumas oscilações.

Considerando agora a situação de correlação alta tem-se, para o teste da estacionariedade, o resultado de rejeição da hipótese, ou seja, o teste descartou 50% da seqüência como “burn-in” e, mesmo assim, a hipótese não foi satisfeita. O teste de “half-width” não rejeitou a hipótese em todas as iterações. Este resultado era esperado, visto que a média para este caso não se apresentava estável e possuía alto erro de Monte Carlo. O critério de Heidelberger & Welch é conhecido por apresentar baixo poder e isso pode ser verificado, pois apresenta valores de “burn-in” muito variados. Aceita-se muito a hipótese falsa de presença de “burn-in” na tentativa de justificar a falta de estabilidade.

4.1.2.4 Critério de Raftery & Lewis (1992a)

Para o modelo normal trivariado, o critério de Raftery & Lewis apresentou bons resultados na situação de ausência de correlação e de correlação média, sendo os resultados coerentes. Para a situação de correlação alta, os resultados foram inconsistentes. Na Tabela 16 estão dispostos os resultados do teste de Raftery & Lewis para as 3 situações de correlação.

TABELA 16. Resultados do Critério de Raftery & Lewis para uma amostra mínima de 3746 iterações aplicado para as 3 situações de correlação.

Iterações ¹	Variáveis	“Thin”	“Burn-in”	Total	Fator de Dependência
Sem correlação					
5000	X1	1	2	3866	1,03
	X2	1	2	3741	1,00
	X3	1	2	3620	0,97
Correlação Média					
5000	X1	2	6	7320	1,95
	X2	2	6	9958	2,66
	X3	2	6	8522	2,27
Correlação Alta					
5000	X1	1	861	1722	0,23
	X2	1	861	1722	0,23
	X3	1	861	1722	0,23

¹ Tamanho efetivo da amostra piloto.

Como sugerido pelo critério, novas amostras foram realizadas para cada situação e submetidas aos testes para verificar se ocorreu a convergência. Os gráficos das médias ergóticas foram feitos para verificar a estabilidade das médias.

Os resultados do critério de Geweke, para as 3 situações de correlação, estão na Tabela 17.

TABELA 17. Resultados do Critério de Geweke para as 3 novas amostras, considerando 5% de significância para o teste Z.

Correlação		X1	X2	X3
Ausência	Teste Z	-0,406	-0,696	1,221
	p-valor	0,685	0,486	0,222
Média	Teste Z	0,591	0,940	0,118
	p-valor	0,555	0,347	0,906
Alta	Teste Z	9,741	9,819	9,801
	p-valor	0,000	0,000	0,000

Pode-se observar que o critério de Geweke não rejeitou a hipótese da estacionariedade do sistema para ausência de correlação e correlação média e rejeitou a hipótese da estacionariedade para a situação de correlação alta, ou seja, o sistema não convergiu para esta última situação.

Para as 3 situações de correlação os resultados do critério de Heidelberg e Welch estão na Tabela 18.

TABELA 18. Resultado do Critério de Heidelberg & Welch para as 3 novas amostras.

Correlação	Variáveis	Teste da Estacionariedade	“Burn-in”	Teste de “Half-width”
Ausência	X1	Aceita-se Ho	0	Aceita-se Ho
	X2	Aceita-se Ho	0	Aceita-se Ho
	X3	Aceita-se Ho	0	Rejeita-se Ho
Média	X1	Aceita-se Ho	0	Aceita-se Ho
	X2	Aceita-se Ho	0	Aceita-se Ho
	X3	Aceita-se Ho	0	Rejeita-se Ho
Alta	X1	Rejeita-se Ho	516	Aceita-se Ho
	X2	Rejeita-se Ho	516	Aceita-se Ho
	X3	Rejeita-se Ho	516	Rejeita-se Ho

Pelo teste de estacionariedade tem-se aceitação da hipótese para todas as variáveis para ausência de correlação e correlação média. O teste rejeitou a hipótese de estacionariedade para todas as variáveis na presença de correlação

alta, confirmando o resultado do critério de Geweke. O teste de “half-width” apresentou um resultado surpreendente; para correlação alta, rejeitou as 3 variáveis, o que era esperado se realmente o processo não convergiu, mas em todos os casos a convergência da variável 3 foi rejeitada pela precisão da sua média. A variável 3 foi gerada com maior variância (100) em relação às outras e, por isso, o teste apresenta este resultado. As médias apresentadas para ausência de correlação e correlação média foram estimadas muito próximas dos valores simulados (2; 200; 0,2). As médias estimadas para a situação de correlação alta apresentaram maiores diferenças em relação às simuladas. Para o teste do “burn-in” foram sugeridas, além dos indicados pelo critério de Raftery & Lewis, 516 iterações, mas o critério não conseguiu prosseguir pois atingiu 50% da sequência. Portanto, verifica-se que a sequência que originou a amostra não estabilizou-se. Na Figura 16 apresentaram-se as médias ergóticas para a situação de correlação alta. Percebe-se que estas realmente não caracterizam estabilidade, mesmo levando em consideração a escala.

O erro de Monte Carlo foi maior para a variável 3 em todas as situações, mas para a situação de correlação alta este foi superior, sendo maior na variável 3. Apesar de os resultados apresentados pelos critérios de Geweke e de Heidelberger & Welch rejeitarem a possibilidade de estabilidade do sistema, obtiveram-se boas estimativas para as médias. Somente a estimativa da média da variável 3 apresentou desvio do seu real valor. A variância e o erro padrão da média da variável 3 foram superestimados. Este resultado reafirma que a convergência do processo deve ser buscada, mas que as estimativas acuradas são outra preocupação que se deve ter.

Todas as 3 novas amostras foram submetidas novamente ao critério de Raftery & Lewis. Os resultados foram coerentes com as amostras que já foram obtidas, mas somente para as situações de ausência de correlação e correlação média. Para a situação de correlação alta, os resultados não foram possíveis de

serem calculados, pois o número mínimo de iterações mantendo-se as especificações de acurácia e probabilidade pré-especificados são de 3746.

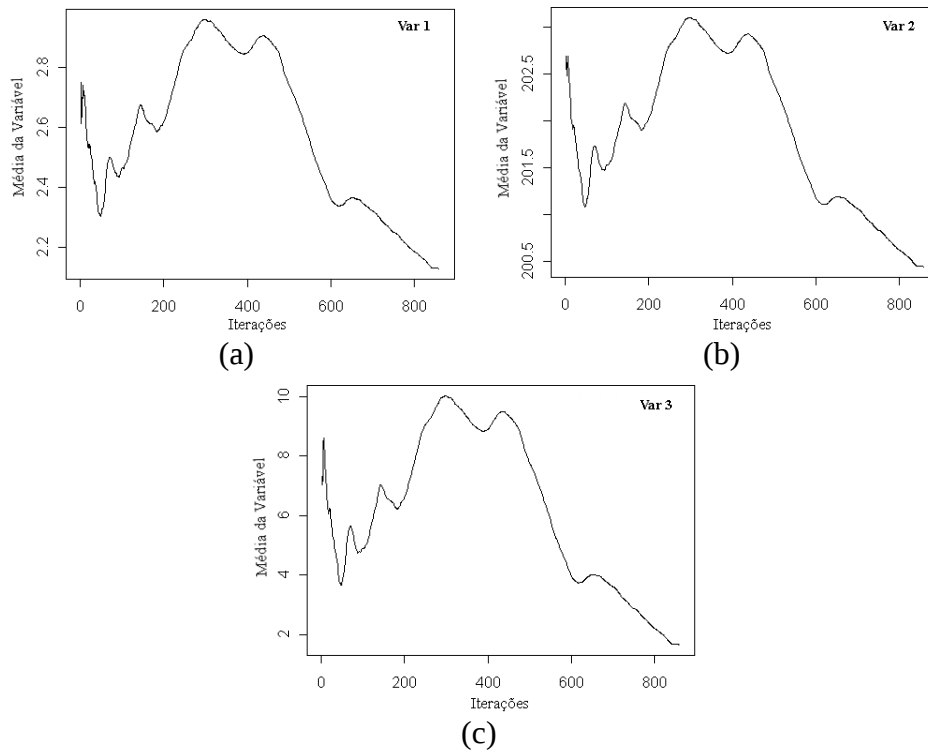


FIGURA 16. Representações gráficas das médias ergóticas das 3 variáveis para a situação de correlação alta.

O critério de Raftery e Lewis não leva em consideração as correlações entre condicionais, somente a correlação dentro da sequência, e isto permite inferir que em correlações altas o processo não consegue estimar com confiança a amostra necessária.

Segundo Raftery & Lewis (1992a) se o Fator de Dependência for maior que 5, indicará ausência de convergência, indicando a necessidade de

reparametrizar, mas em algumas situações como esta da normal trivariada, não se tem interesse algum em utilizar uma transformação linear das variáveis.

Na Tabela 19 estão dispostas as estatísticas descritivas para as 3 variáveis nas 3 situações para as novas amostras. As estimativas das médias nas situações de ausência de correlação e correlação média apresentaram resultados muito próximo do valor exato, com baixo valor de erro de Monte Carlo. Para a situação de correlação alta têm-se boas estimativas para a variável 1 e 2, mas há uma variação um pouco maior para a variável 3 em relação às estimativas das outras situações apesar de o intervalo de credibilidade ser muito próximo dos observados pelos demais, o erro de Monte Carlo apresentou-se muito elevado, o que enfatiza a necessidade de uma amostra melhor, ou seja, um maior número de iterações é necessário.

TABELA 19. Análise descritiva e respectivos Erros de Monte Carlo para as 3 novas amostras.

Correlação	Var.	Média	SE "Naive"	Erro de Monte Carlo	Quantil 0,025	Quantil 0,5	Quantil 0,975
Ausência	X1	2,02	0,016	0,015	0,02	2,00	4,03
	X2	200,05	0,050	0,046	193,99	200,11	206,04
	X3	0,23	0,164	0,156	-20,02	0,39	20,34
Média	X1	2,00	0,010	0,014	0,01	2,00	3,95
	X2	199,96	0,032	0,042	193,71	199,99	206,09
	X3	0,16	0,100	0,139	-19,88	0,30	19,09
Alta	X1	2,13	0,037	0,104	0,10	2,19	3,94
	X2	200,44	0,116	0,331	193,88	200,74	206,30
	X3	1,64	0,367	1,043	-19,20	2,58	20,02

4.2 Critérios Multivariados

4.2.1 Cálculo alternativo para maximização de formas quadráticas

A maximização de formas quadráticas é muito utilizada para situações em que se deseja obter um valor que represente a maior variabilidade do sistema. Quando se tem a maximização de formas quadráticas há um sistema dado em (2.2). Para maximização de um par de formas quadráticas tem-se um sistema (seção 2.1.2.) parecido, mas relacionado à razão entre duas formas quadráticas. Este fato dificulta a obtenção das raízes características e do vetor característico. Então, este trabalho propõe uma rotação nos eixos cartesianos do sistema buscando novos eixos no sentido de maior variabilidade, tornando-o, assim, um sistema de duas formas quadrática, em uma única forma quadrática.

Seja o par de formas quadráticas do critério de Brooks & Gelman dados na equação (2.23) que se deseja maximizar. Na literatura específica de análise Bayesiana, é normal encontrar a maximização dada pela obtenção dos autovalores e autovetores de $D^{-1}\hat{V}$. No presente trabalho propõe-se a forma alternativa descrita doravante. Para isso, a matriz D é fatorada (fator de Cholesky) na seguinte forma:

$$D = SS^T.$$

Definindo-se $\underline{z}$ como a transformação linear do vetor $\underline{a}$ por $\underline{z} = S^T \underline{a}$, tem-se $\underline{a} = (S^{-1})^T \underline{z}$, pois de acordo com as propriedades do fator de Cholesky,

$SS^{-1} = S^T (S^{-1})^T = I$. Se a equação $(\hat{V} - \lambda_i D) \hat{a}_i = 0$ for pré multiplicada por S^{-1} e $\tilde{a} = (S^{-1})^T \hat{a}$ for substituído na expressão, têm-se:

$$\left[S^{-1} \hat{V} - \lambda S^{-1} D \right] (S^{-1})^T \tilde{z} = 0$$

$$\left[S^{-1} \hat{V} (S^{-1})^T - \lambda S^{-1} D (S^{-1})^T \right] \tilde{z} = 0$$

$$\left[S^{-1} \hat{V} (S^{-1})^T - \lambda S^{-1} S S^T (S^{-1})^T \right] \tilde{z} = 0$$

$$\left[S^{-1} \hat{V} (S^{-1})^T - \lambda I \right] \tilde{z} = 0,$$

desde que $S^{-1} D (S^{-1})^T = I$. Chamando $H = S^{-1} \hat{V} (S^{-1})^T$, tem-se $(H - \lambda_i I) \tilde{z}_i = 0$, sendo, portanto, a mesma solução para o caso de maximização de uma forma quadrática, exceto que $\tilde{a} = (S^{-1})^T \hat{a}$ deve ser recuperado, pois os autovetores encontrados $\tilde{z}$ são alterados. Os autovalores são invariantes mediante a transformação não-singular efetuada.

De acordo com Brooks & Gelman (1998), o autovalor máximo λ_i é o próprio R^p , sendo R^p o fator de redução de escala potencial p dimensional, dado pela equação (2.23); desta maneira, a transformação nos fornece o sistema:

$$(H - R_i^p I) \tilde{z}_i = 0. \quad (4.1)$$

A partir dessa equação pode-se verificar que os problemas mencionados por Brooks & Gelman (1998) não ocorrem, pois não é necessário inverter a matriz D para se obter a solução da maximização de duas formas quadráticas. A matriz D precisa ser positiva definida para aceitar o fator de Cholesky.

4.2.2 Dois critérios de convergência propostos

O critério multivariado de Brooks & Gelman (1998) é teoricamente considerado eficiente se existirem elevadas correlações entre os parâmetros. O caso extremo é dado pela situação de perfeita correlação entre os parâmetros. Nesse caso, esse critério multivariado seria equivalente ao critério univariado de algum dos p parâmetros, uma vez que monitorar a correlação de um parâmetro refletirá o que acontece nas seqüências dos outros. Para o caso de baixas correlações entre os parâmetros, ou subgrupos deles, esse critério pode falhar na monitoração da convergência, uma vez que considera apenas o sentido de maior variabilidade no hiperespaço p -dimensional. No caso extremo de ausência de correlação, esse critério vai monitorar apenas o parâmetro de maior perturbação (variância) para o sistema.

Essas limitações proporcionam uma maior reflexão sobre esse método. Essa reflexão permitiu que duas novas alternativas multivariadas fossem propostas. Ambas consideram as variabilidades em todos os eixos do hiperespaço, que são transformações lineares dos eixos dos parâmetros no sentido de maior variabilidade. A primeira alternativa é baseada na substituição dos escalares $\hat{\mathbf{a}}^T \hat{\mathbf{V}} \hat{\mathbf{a}}$ e $\hat{\mathbf{a}}^T \mathbf{D} \hat{\mathbf{a}}$ do critério original por novos escalares $\text{traço}(\hat{\mathbf{V}})$ e $\text{traço}(\mathbf{D})$, respectivamente; sendo $\mathbf{D}$ a matriz de covariâncias dentro de seqüência dada pela equação (2.21) e $\hat{\mathbf{V}}$ a matriz de covariância estimada total dada pela equação (2.20). O segundo critério é relativo ao produto dos autovalores não-

nulos da matriz $D^{-1}\hat{V}$, $\prod_{i=1}^r R_i^p = \det(\hat{V})/\det(D)$. Ambas as medidas consideram de forma sumariadora univariada a variabilidade de todo sistema. Assim, definem-se os seguintes critérios:

$$\hat{R}_{\text{det.}}^p = \prod_{i=1}^r R_i^p \quad \text{e} \quad \hat{R}_{\text{traço}}^p = \text{Traço}(\hat{V})/\text{Traço}(D), \quad (4.2)$$

sendo r o número de autovalores não nulos de (4.1).

4.2.3 Aplicação dos critérios multivariados

4.2.3.1 Modelo de intervenção com erro *ARMA(2,2)*

O número de autovalores do sistema é r , como comentado anteriormente. O valor r é dado pelo número $\min(m-1, p)$, sendo m o número de seqüências e p o número de parâmetros. Então o número de autovalores é diretamente relacionado com o número de seqüências utilizadas. Por este motivo, estudaram-se os casos com 3, 5 e 7 seqüências em paralelo.

Para este modelo, os critérios estão representados de maneira gráfica monitorados a cada 2 iterações. Os critérios se comportaram de maneira muito parecida, ou seja, todos os critérios fizeram o mesmo trajeto. O critério do Traço apresentou um comportamento um pouco mais suave e com valores menores em relação aos outros. Observa-se que, na Figura 17, o critério do Determinante concorda com o critério original de Brooks & Gelman, indicando a convergência para $R = 1,2$ com 396 e 458, mas apresenta uma pequena flutuação perto de $R = 1,05$ em 1760 e 1770 iterações, respectivamente. O critério do Traço, como

comentado, caracteriza a convergência para $R = 1,2$ com 384 iterações. Pode-se perceber que o critério do traço caracterizou a convergência com um número de iterações muito parecido com o do critério univariado de Gelman & Rubin (380 iterações).

Poder-se-ia pensar em adotar $R = 1,05$. Com essa escolha, haveria um excelente ponto de referência. O critério do Traço sempre apresentou valores menores.

Considerando agora 3, 5 e 7 seqüências, têm-se os resultados de maneira gráfica na Figura 18. Quando se considera maior número de seqüências, tem-se maior precisão nas estimativas da matriz de covariâncias e, com isso, diminuem as grandes oscilações.

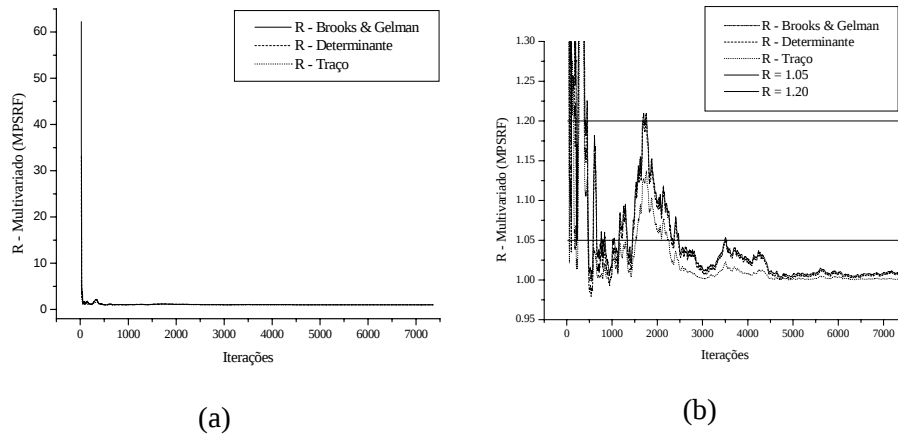


FIGURA 17. Representações gráficas dos critérios multivariados (MPSRF) considerando 2 seqüências. Figura (b) dá ênfase na escala do Fator R.

Para o aumento no número de seqüência o Amostrador de Gibbs foi realizado, como eventos independentes, ou seja, sendo amostragens diferentes. Considerando os resultados apresentados para 3 seqüências, tem-se a

convergência caracterizada para $R = 1,2$ com 486 iterações para o Traço, 526 iterações para o Determinante e 524 iterações para o critério original de Brooks & Gelman. Quando foram utilizadas somente 5 seqüências em paralelo, houve a convergência dos critérios caracterizada com 352, 370 e 364 iterações, respectivamente. Para 7 seqüências, tem-se a convergência caracterizada com 422, 880 e 850 iterações, respectivamente (Figura 18). Observa-se que o critério do Traço caracterizou a convergência sempre antes dos demais. Outro ponto interessante é que as iterações não oscilaram muito com o aumento do número de seqüências para o critério do Traço. O critério do Determinante acompanhou os resultados do critério original de Brooks & Gelman em todas as situações.

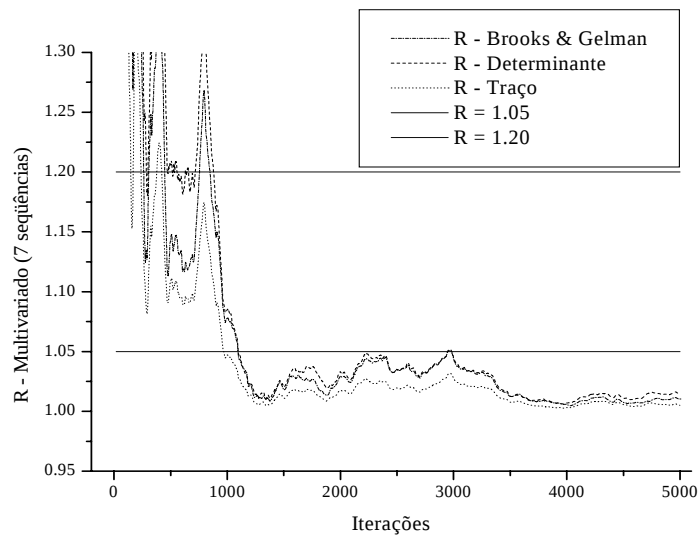


FIGURA 18. Representação gráfica dos critérios multivariados (MPSRF) considerando 7 seqüências em paralelo para o modelo de intervenção.

O aumento no número de seqüências aumentou o número de autovalores. Avaliando os resultados destes autovalores pode-se perceber que existia sempre um autovalor que representava mais de 70% da variação do sistema devido aos parâmetros do modelo de intervenção serem muito correlacionados.

4.2.3.2 Normal trivariada

Utilizando um modelo normal trivariado e considerando a correlação inexistente, tem-se, como já comentado anteriormente, a situação de uma simulação de Monte Carlo; por isso é esperado que o processo só sofra influência do valor arbitrário inicial. Para esta situação, têm-se os resultados considerando 2, 3, 5 e 7 seqüências em paralelo. Na Figura 19 encontra-se o gráfico dos critérios para 7 seqüências.

Observa-se, nas Figuras 21 e A10, que o aumento das seqüências causou uma pequena diminuição da variabilidade em torno do valor 1. Como comentado, era de se esperar que o sistema só fosse influenciado pelo valor inicial. A caracterização da convergência quando se utilizaram 2 seqüências para $R = 1,2$ ocorreu com 28 iterações para o critério original de Brooks & Gelman, 26 iterações para o critério do Determinante e menos que 20 iterações para o critério do Traço. A caracterização da convergência quando se utilizaram somente 7 seqüências ocorreu muito próximo dos valores anteriores, 30, 30, e menos que 20 iterações, respectivamente.

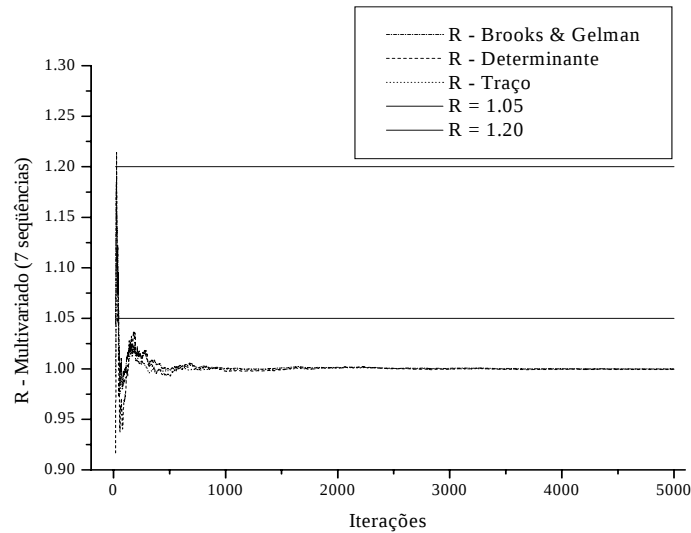


FIGURA 19. Representação gráfica dos critérios multivariados (MPSRF) considerando 7 seqüências na situação de ausência de correlação.

Pode-se perceber que existem poucas diferenças quando se utiliza um número de seqüências diferentes. Somente para o critério do Traço estas possíveis diferenças não foram perceptíveis, pois este é o critério que apresentou resultados próximos do esperado para a situação de ausência de correlação. Para $R = 1,05$, os resultados estão próximos dos obtidos para o $R = 1,2$.

Considerando agora uma correlação média, observa-se na Figura 20 uma certa igualdade dos critérios com 7 seqüências; entretanto, pode-se verificado que o mesmo ocorreu com o aumento no número das seqüências (Figura A11). A caracterização da convergência quando foram utilizadas 2 seqüências para $R = 1,2$ ocorreu com 54 iterações para o critério original de Brooks & Gelman, 40 iterações para o critério do Determinante e 40 iterações para o critério do Traço. A caracterização da convergência quando se utilizaram 7 seqüências apresentou

pouca diferença dos valores anteriores, ou seja, 40, 32 e menos que 20 iterações, respectivamente.

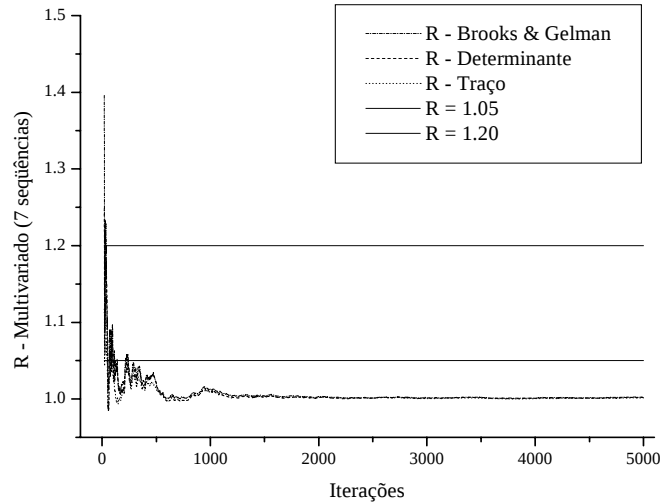


FIGURA 20. Representação gráfica dos critérios multivariados (MPSRF) considerando 7 seqüências em paralelo para o modelo normal trivariado na situação de correlação média.

Correlação alta entre as variáveis permite que uma única variável explique as outras duas, ou seja, o processo de amostragem se torna lento devido à dependência entre as condicionais completas, as quais apresentam o mesmo comportamento (Gamerman, 1996). Por este motivo tem-se um autovalor que explica praticamente 100% da variação. Na Figura 21 observa-se, como era esperado, a igualdade dos critérios mesmo com o aumento no número de seqüências. A caracterização da convergência quando se utilizaram 2 seqüências para $R = 1,2$ ocorreu com 1544 iterações para o critério original de Brooks & Gelman, 1544 iterações para o critério do Determinante e 1538 iterações para o critério do Traço. A caracterização da convergência quando se utilizaram 7

seqüências apresentou os seguintes valores, respectivamente: 1760, 1762 e 1754 iterações.

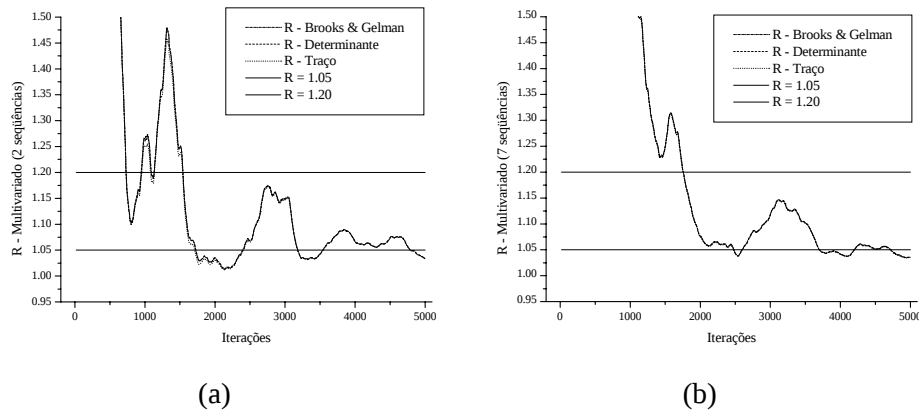


FIGURA 21. Representações gráficas dos critérios multivariados (MPSRF) considerando 2 (a) e 7 (b) seqüências na situação de correlação alta.

Fica claro que na presença de correlação alta entre as condicionais o processo se torna lento, necessitando de mais iterações. Uma maneira de acelerar a convergência seria o de reparametrizar, mas em algumas situações este procedimento não é desejado, necessitando de atenção maior a estas situações.

4.3 Considerações Finais

Para o caso univariado os critérios apresentaram resultados variados, entre os quais destacam-se: Gelman & Rubin (1992a) é um critério que se preocupa em caracterizar a convergência, mas que sozinho não garante a qualidade da inferência, pois é fundamental que a amostra seja representativa. Por isso, um número maior de iterações é necessário, além daquela quantidade atingida quando o fator $R = 1,2$ já for alcançado. Geweke (1992) é um critério

muito rigoroso e por isso a convergência é atingida com um número grande de iterações. A utilização do nível de significância de 1% se mostrou uma boa alternativa para o teste de hipótese realizado pelo critério de Geweke. Este método não deve ser aplicado sozinho, e sim combinado com outro método, pois seu resultado positivo é uma condição necessária mas não suficiente para garantir a convergência. Heidelberg & Welch (1983) é um método que se preocupa com o tamanho do “burn-in” para seqüências pequenas, pois o efeito deste “burn-in” sobre a média da parte inicial é absorvido. O teste de “half-width” presente neste método não considerou a diferença de variabilidade dos parâmetros, pois depende desta precisão que é pré-estabelecida.

O critério de Geweke e de Heidelberg & Welch apresentou resultados contraditórios quando a verificação foi feita de maneira pontual. Este fato também pode ocorrer com o critério de Gelman & Rubin. Uma alternativa que se mostrou extremamente viável foi a verificação da convergência de maneira gráfica. Esta deve ser construída concomitantemente com o processo de amostragem, sendo atualizada a cada x iterações. O uso de um número pequeno para x se mostrou desnecessário, aumentando o esforço computacional. A utilização de x em torno de 10 a 20 e a utilização de métodos de suavização seriam os métodos indicados para a construção do gráfico.

O critério de Raftery & Lewis (1992a) para a determinação do tamanho adequado da amostra, que se baseia em uma amostra piloto, se mostrou extremamente importante, fornecendo bons resultados. Chamaram a atenção os resultados desse critério quando as condicionais completas apresentavam correlação alta. Para este caso específico o critério de Raftery & Lewis apresentou resultados de convergência inconsistentes com a qualidade da amostra obtida.

Para se obter uma maior eficiência no monitoramento da convergência utilizando os critérios estudados nesse trabalho, alguns procedimentos são

recomendados. A princípio, é necessário gerar uma amostra piloto e aplicar o critério de Raftery & Lewis. Esse critério fornecerá um tamanho de sequência a ser obtida. O Amostrador de Gibbs é implementado e na proximidade do fim do tamanho da sequência recomendada pelo critério de Raftery & Lewis são aplicados os critérios de Gelman & Rubin e Geweke. O objetivo desse passo é monitorar a sequência a cada x iterações para verificar se a convergência foi alcançada. Caso esse objetivo tenha sido verificado, o processo é interrompido. Por outro lado, se a convergência não foi alcançada, o Amostrador de Gibbs deve ser acionado para se obterem novos x componentes da sequência e novamente o processo é repetido até que seja atingida a convergência. Espera-se que a convergência seja atingida próxima ao tamanho de sequência apontado pelo critério de Raftery & Lewis. Finalmente, pode-se obter o tamanho do “burn-in” utilizando o critério de Heidelberger & Welch.

É importante enfatizar que o procedimento de verificação da convergência dependerá muito da opinião do pesquisador, por isso nunca se deve realizar um único procedimento de amostragem. A qualidade da amostra é fundamental para realizar inferência; por isso, o erro de Monte Carlo deve ser averiguado.

Os métodos multivariados são indispensáveis mesmo em situações com poucos parâmetros, pois levam em consideração a variação e a correlação no hiperespaço. O critério de Brooks & Gelman (1998) apresentou resultados de convergência coniventes com os modelos simulados; todavia, por ser uma generalização do critério de Gelman & Rubin, apresenta também a característica de se preocupar somente com a monitoração da convergência, e não com a qualidade da amostra. A maneira alternativa de cálculo proposta para este critério foi implementada facilmente e foi robusta numericamente durante o processo de simulação. Dois critérios alternativos foram propostos para contemplar toda a variação dos parâmetros nas sequências. O critério do Traço

apresentou fácil implementação e mostrou bons resultados, em alguns casos mais coerentes que os dos outros.

Em algumas situações, como, por exemplo, quando se está interessado em combinações lineares entre parâmetros, a matriz de covariâncias dentro de seqüência (D) apresentará colunas linearmente dependentes. Seu determinante será nulo, não possibilitando a realização do critério do Determinante. Os algoritmos presentes em software como SAS e R permitem obter o fator de Cholesky de matrizes semi positivas definidas. Em situações iguais às comentadas, por problemas numéricos, estes algoritmos calculam autovalores negativos, não possibilitando o cálculo do critério de Brooks & Gelman (1998). O critério do Traço não apresentou qualquer problema neste tipo de situação.

A utilização de uma nova referência para o fator de redução de escala potencial para os critérios de Gelman & Rubin, Brooks & Gelman, Determinante e do Traço, que passou de $R = 1,2$ para $R = 1,05$, se mostrou interessante, pois a amostra apresentou resultados mais próximos dos obtidos pela amostra com PSRF próximos de 1 (Tabela 2). Assim, essa mudança, além de apontar que a convergência foi alcançada, também permitiu obter amostras representativas, possibilitando melhores inferências.

5 CONCLUSÃO

1. Sugere-se a utilização dos critérios univariados do monitoramento da convergência de forma combinada. O seguinte procedimento é recomendado:
 - a. aplicar Raftery & Lewis em uma amostra piloto e determinar o tamanho ideal da seqüência;
 - b. determinar o tamanho do “burn-in” pelo critério de Heidelberger & Welch;
 - c. monitorar a convergência das seqüências nas proximidades do tamanho ideal indicado pelo critério de Raftery & Lewis por meio dos critérios de Gelman & Rubin e Geweke.
2. O monitoramento das seqüências, no passo 1 (b), pode ser realizado por meio de um dos critérios multivariados estudados;
3. a forma alternativa de cálculo para o critério de Brooks & Gelman foi viável de ser aplicada e mostrou-se robusta numericamente;
4. os dois novos critérios multivariados para monitorar a convergência das seqüências de Monte Carlo foram propostas com sucesso;
5. o critério multivariado, baseado na razão dos traços das matrizes de covariâncias $\hat{V}$ e D , apresenta bons resultados e é de fácil implementação.

REFERÊNCIAS BIBLIOGRÁFICAS

BOCK, R. D. **Multivariate statistical methods in behavioral research**. New York: McGraw Hill, 1975. 623 p.

BROEMELING, L. D. **Bayesian analysis of linear models**. New York: M. Dekker, 1985. 454 p.

BROOKS, S. P.; DELLAPORTAS, P.; ROBERTS, G. O. An approach to diagnosing total variation convergence of MCMC algorithms. **Journal of Computational and Graphical Statistics**, Alexandria, v. 6, n. 3, p. 251-265, Sept. 1997.

BROOKS, S. P.; GELMAN, A.; General methods for monitoring convergence of iterative simulations, **Journal of Computational and Graphical Statistics**, Alexandria, v. 7, n. 4, p. 434-455, Dec. 1998.

BROOKS, S. P.; ROBERTS, G. O. Some L^2 results for diagnosing convergence of MCMC algorithms. **Tech. Rep.**, University of Cambridge, In Preparation. 1995.

BROOKS, S. P.; ROBERTS, G. O. Assessment convergence of markov chain monte carlo algorithms. **Statistics and Computing**, Dordrecht, v. 8, n. 4, p. 319-335, dec. 1998.

CARLIN, B. P.; LOUIS, T. A. **Bayes and empirical bayes methods for data analysis**. 2. ed. London: Chapman and Hall, 2000. 419 p.

CARLIN, B. P.; GELMAN, A.; NEAL, R. M.; KASS, R. E. Markov Chain Monte Carlo in Practice: A Roundtable Discussion, **Journal of the American Statistical Associations**, Statistical Practice, Alexandria, v. 52, n. 2, p. 93-99, May 1998.

COWLES, M. K.; CARLIN, B. P. Markov chain Monte Carlo convergence diagnostics: a comparative review. **Journal of the American Statistical Associations**, Alexandria, v. 91, n. 434, p. 883-904, June 1996.

DE FINETTI, B.; La prévision: ses lois logiques, ses sources subjectives. **Annales de l'Institut Henri Poincaré**, Paris, v. 7, p. 1-68, 1937. (Translated into English by Henry E. Kyburg Jr., Foresight: Its Logical Laws, its Subjective

Sources. In: KYBURG JR., H. E.; SMOKLER, H. E. (Ed.). **Studies in Subjective Probability**. New York: Wiley, 1964. p. 53-118.)

DIAZ, J. Bayesian analysis of regression models with time series erros. In SPALL, J. C. (Ed.). **Bayesian analysis of Time Series and Dynamic Models**. New York: M. Dekker, 1988. p. 347-358.

FISHER, R. A. **The desing of experiments**, Edinburgh: Oliver and Boyd, 1935.

GAMERMAN, D. **Simulação estocástica via cadeias de Markov**. São Paulo: ABE, 1996. 196 p.

GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches to calculating marginal densities. **Journal of the American Statistical Association**, Alexandria, v. 85, n. 410, p. 398-409, June 1990.

GELMAN, A. Inference and monitoring convergence. In: GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. (ED.). **Markov Chain Monte Carlo in Practice**. New York: Chapman and Hall, 1995. chap. 8.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-511, 1992a.

GELMAN, A.; RUBIN, D. B. A single series from the Gibbs sampler provides a false sense of security. In: BERNARDO, J. M.; BERGER, J. O.; DAWID, A. P.; SMITH, A. F. M. (Ed.). **Bayesian Statistics 4**. New York: Oxford University Press, 1992b. p. 625-631.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. London: Chapman and Hall, 1997. 526 p.

GEWEKE, J. Evaluating the accurary of sampling-based approaches to the calculation of posterior moments. N: BERNARDO, J. M.; BERGER, J. O.; DAWID, A. P.; SMITH, A. F. M. (Ed.). **Bayesian Statistics 4**. New York: Oxford University Press, 1992. p. 625-631.

GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. **Markov Chain Monte Carlo in Practice**, London: Chapman and Hall, 1996.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, London, v. 57, n. 1, p. 97-109, Apr. 1970.

HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations Research**, Landing, v. 31, n. 6, p. 1109-44, 1983.

JEFFREYS, H. **Theory of Probability**. Oxford: Claredon Press, 1939. 380 p.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4. ed. New Jersey: Prentice Hall, 1998. 816 p.

KOLMOGOROV, A. N. Sulla determinazione Empirica delle leggi di probabilita. **Giornal Istituto Italia Attuari**, v. 4, p. 1-11, 1933.

LIU, C.; WONG, W. H.; KONG, A. Correlation structure and convergence rate of the Gibbs sampler: applications to the comparison of estimators and augmentation schemes. **Biometrika**, London, v. 81, n. 1, p. 27-40, Apr. 1994.

MACEACHERN, S. N.; BERLINER, L. M. Subsampling the Gibbs sampler. **The American Statistician**, Alexandria, v. 48, n. 3, p. 188-190, Aug. 1994.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate Analysis**. London: Academic Press, 1979.

MILANI, L. L. **Análise bayesiana do modelo de intervenção com erro arma**. 2000. 142 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.

O'HAGAN, A. **Bayesian inference**: Kendall's advanced theory of statistic. London: Cambridge University Press, 1994. 330 p.

PINO, F. A. **Análise de intervenção em séries temporais**: aplicações em economia agrícola. 1980. Dissertação. (Mestrado em Estatística) – Universidade de São Paulo, São Paulo.

RAFTERY, A. E.; LEWIS, S. Comment: One long run with diagnostics: implementation strategies for markov chain monte carlo. **Statistical Science**, Hayward, v. 7, n. 4, p. 493-497, 1992a.

RAFTERY, A. E.; LEWIS, S. How many iterations in the Gibbs sampler? In: BERNARDO, J. M.; BERGER, J. O.; DAWID, A. P.; SMITH, A. F. M. (Ed.). **Bayesian Statistics 4**. Oxford: University Press, 1992. p. 763-773b.

RIPLEY, B. D. **Stochastic Simulation**. New York: John Wiley, 1987.

SAS INSTITUTE. **SAS for Windows, Release 8**. Cary, N. C, 2000.

SCHURUBEN, L.; SINGH H.; TIERNEY, L. Optimal tests for initialization bias in simulation output. **Operations Research**, Landing, v. 31, n. 6. p. 1167-1178. 1983.

SMITH, B. J. **Bayesian output analysis program (BOA)**. Disponível em: <http://www.public-health.uiowa.edu/boa/boa_user.pdf>. Version 1.0 User's Manual, 2003. Acesso em: 2003.

SORENSEN, D. **Gibbs sampling in quantitative genetics**. Foulun, 1996. 186 p.

TIERNEY, L. Markov chains for exploring posterior distributions. **Annals of Statistics**, Hayward, v. 22, n. 4, p. 1701-1762, Dec. 1994.

VON MISES, R. **Wahrscheinlichkeitsrechnung**. Vienna: Deuticke, 1931.

ANEXOS

ANEXO A

FIGURA A1. Representações gráficas do SRFP (R) para os 6 parâmetros do modelo $ARMA(2,2)$ com 30000 iterações realizadas.

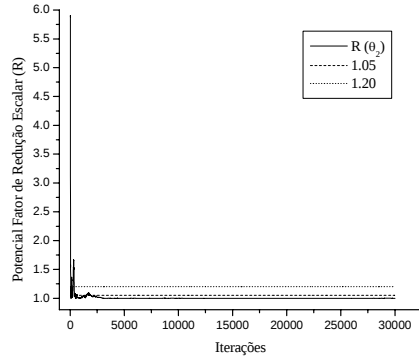
FIGURA A2. Representações gráficas das médias ergóticas calculadas a cada iteração para 5 parâmetros da amostra de 30000 iterações.

FIGURA A3. Representações gráficas do Potencial Fator de Redução Escalar (R) para as variáveis 2 e 3 com um exemplar de cada com a escala do R ampliada com 5000 iterações realizadas considerando correlação alta.

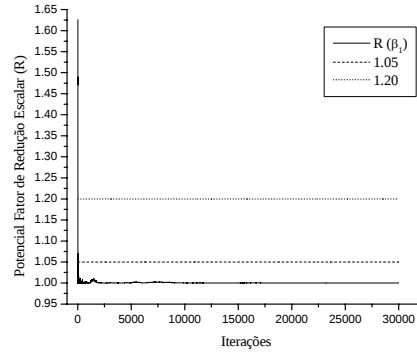
FIGURA A4. Representações gráficas da estatística do teste Z para o critério de Geweke para as variáveis 1 e 2 nos 3 níveis de correlação para 5000 iterações.

FIGURA A5. Representações gráficas dos critérios multivariados (MPSRF) considerando 2(a), 3(b) e 5(c) seqüências na situação de ausência de correlação.

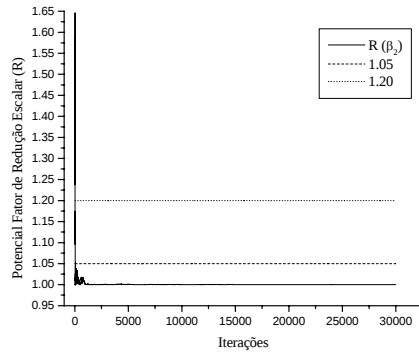
FIGURA A6. Representações gráficas dos critérios multivariados (MPSRF) considerando 2 (a), 3 (b) e 5 (c) seqüências na situação de correlação média.



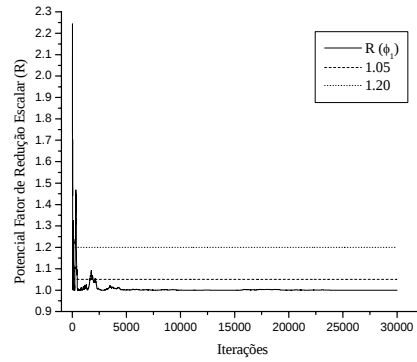
(a)



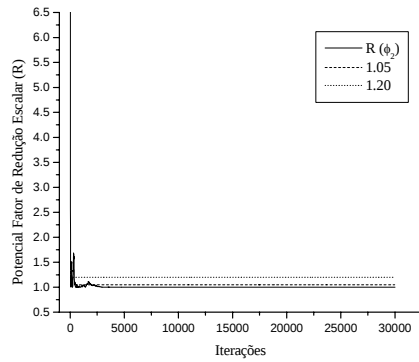
(b)



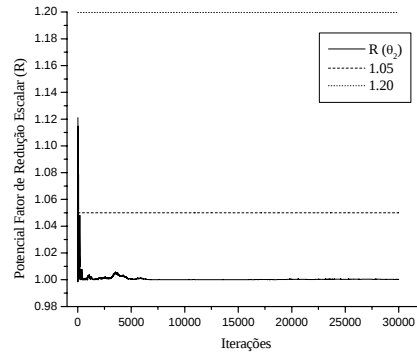
(c)



(d)

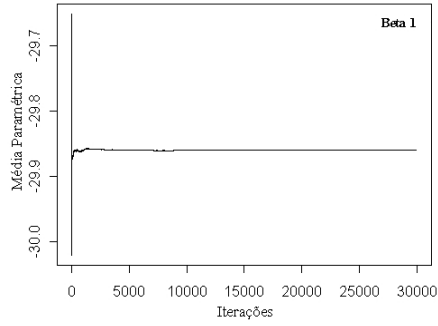


(e)

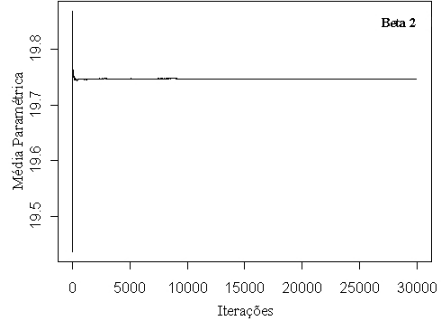


(f)

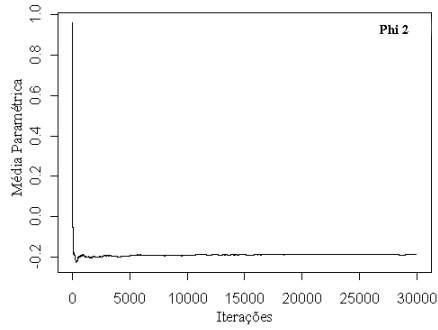
FIGURA A1. Representações gráficas do SRFP (R) para os 6 parâmetros do modelo $ARMA(2,2)$ com 30000 iterações realizadas.



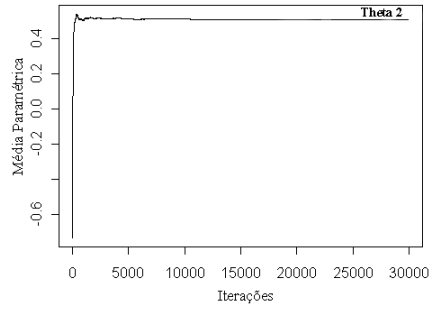
(a) - β_1



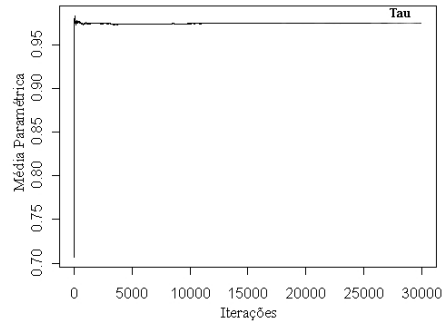
(b) - β_2



(c) - ϕ_2

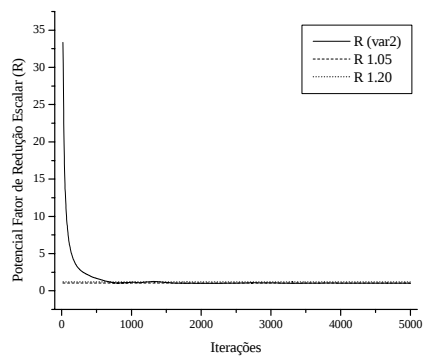


(d) - θ_2

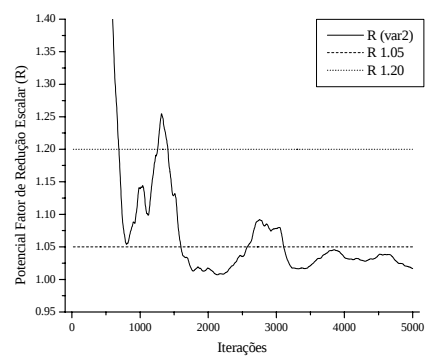


(e) - τ

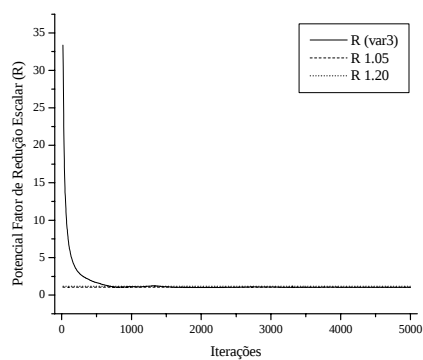
FIGURA A2. Representações gráficas das médias ergóticas calculadas a cada iteração para 5 parâmetros da amostra de 30000 iterações.



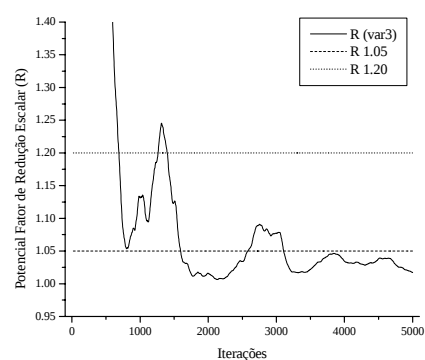
(a)



(b)



(c)



(d)

FIGURA A3. Representações gráficas do Potencial Fator de Redução Escalar (R) para as variáveis 2 e 3 com um exemplar de cada com a escala do R ampliada com 5000 iterações realizadas considerando correlação alta.

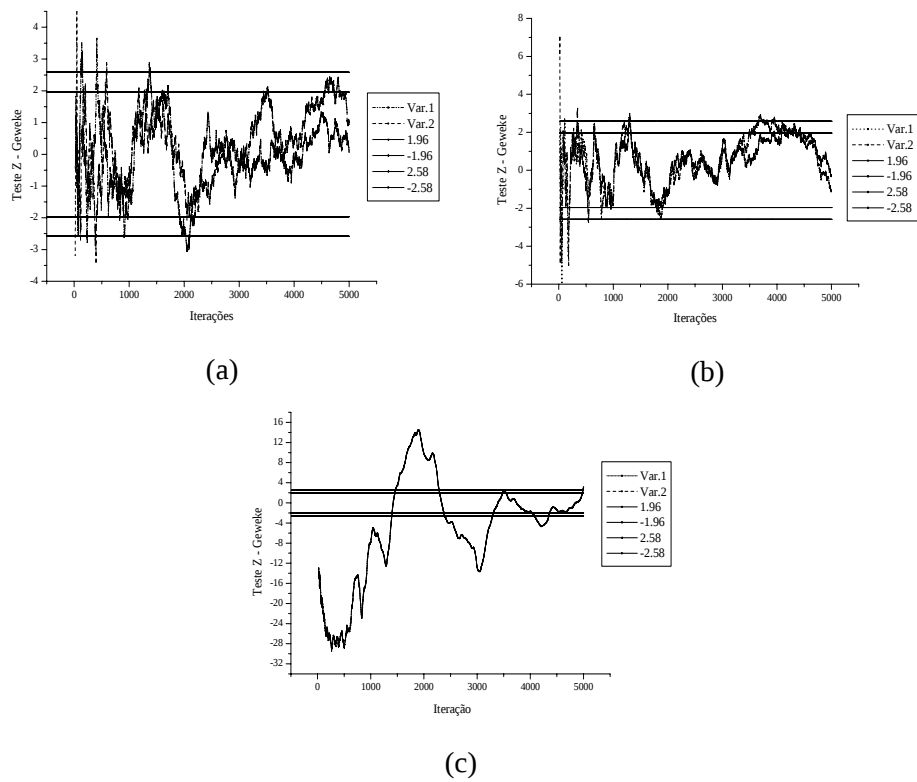
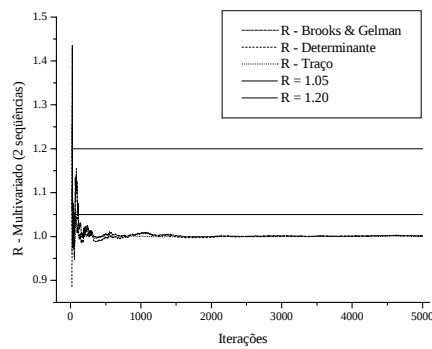
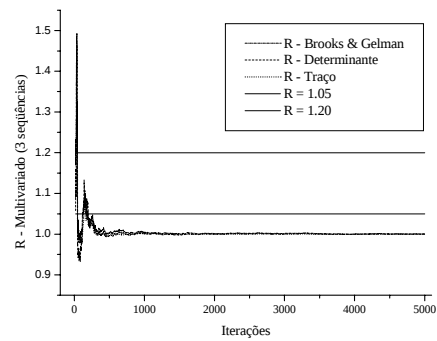


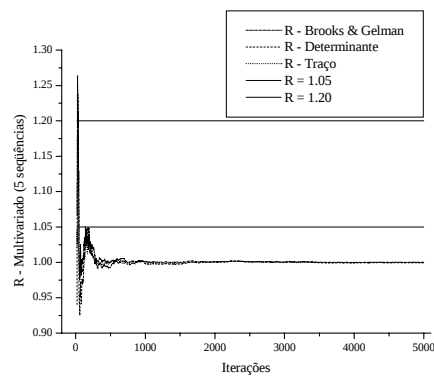
FIGURA A4. Representações gráficas da estatística do teste Z para o critério de Geweke para as variáveis 1 e 2 nos 3 níveis de correlação para 5000 iterações.



(a)

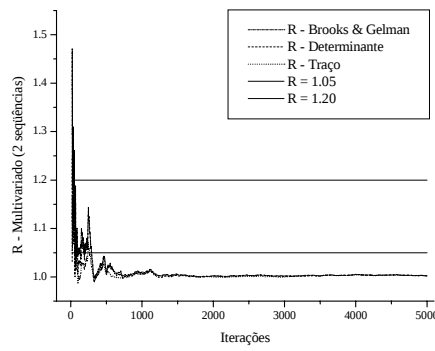


(b)

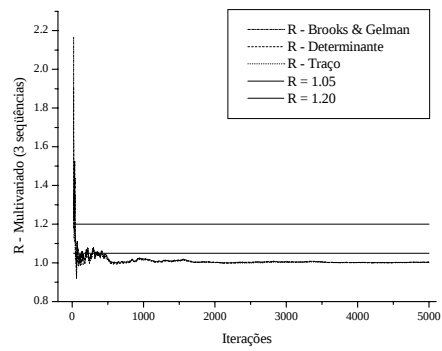


(c)

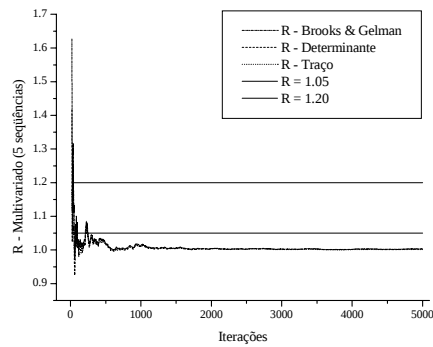
FIGURA A5. Representações gráficas dos critérios multivariados (MPSRF) considerando 2(a), 3(b) e 5(c) seqüências na situação de ausência de correlação.



(a)



(b)



(c)

FIGURA A6. Representações gráficas dos critérios multivariados (MPSRF) considerando 2 (a), 3 (b) e 5 (c) seqüências na situação de correlação média.

ANEXO B – PROGRAMAS UTILIZADOS NOS PROCEDIMENTOS DE SIMULAÇÃO E ANÁLISE.

Programa de simulação do modelo de séries temporais com erro
ARMA(2,2) com duas intervenções. Programação realizada no software SAS®

```
proc iml;
  theta1=-0.3; theta2=0.5; beta1=-30; beta2=20; phi1=0.5; phi2=-0.3; tau=1;
  n=500;
  y=j(n,1,0);
  at_1=0;at_2=0;
  x1t_1=0;x1t_2=0;
  x2t_1=0;x2t_2=0;
  y[1]=0;y[2]=0;
  vt=j(n+2,1,0);
  vt[1]=1;vt[2]=2;
  vat=j(n,1,0);
  do t=3 to n;
    at=rannor(0);
    if t<=120 then x1t=0;else x1t=1;
    if t<=300 then x2t=0;else x2t=1;
    yt=phi1*y[t-1]+phi2*y[t-2]+beta1*x1t+beta2*x2t+theta1*
at_1+theta2* at_2+at-phi1*x1t_1*beta1-phi2*x1t_2*beta1-phi1*x2t_1*beta2-
phi2*x2t_2* beta2;
    at_2=at_1;at_1=at;
    x1t_2=x1t_1;x1t_1=x1t;
    x2t_2=x2t_1;x2t_1=x2t;
    y[t]=yt;
    vat[t]=at;
    vt[t]=t;
  end;
  x=y||vat;
  create serie from x ;
  append from x;
quit;
```

Programa do Amostrador de Gibbs que utiliza a série simulada.

```
proc iml;
  use serie;
  read next 500 into y;
  at=y[1:500,2:2];
  y=y[1:500,1:1];
  nit=100;pbur_in=0.40;salto=10;
  n=500;seed=0;
  res=j(nit,14,0);
  do cad=1 to 2;/*inicio do loop de 2 cadeias*/
    /*criando vetores de retardo*/
    if cad=1 then atarm=at;
    if cad=2 then at=atarm;
    y1=j(n,1,0);y2=j(n,1,0);
    y1[1]=0;y2[1]=0;y2[2]=0;
    y1[2:n]=y[1:n-1];y2[3:n]=y[1:n-2];
    at1=j(n,1,0);at2=j(n,1,0);
    at1[1]=0;at2[1]=0;at2[2]=0;
    at1[2:n]=at[1:n-1];at2[3:n]=at[1:n-2];
    x=j(n,2,0);
    do t=1 to n;
      if t<=120 then x1t=0;else x1t=1;
      if t<=300 then x2t=0;else x2t=1;
      x[t,1]=x1t;x[t,2]=x2t;
    end;
    x1=j(n,2,0);x2=j(n,2,0);
    x1[1,1]=0;x1[1,2]=0;
    x2[1,1]=0;x2[1,2]=0;x2[2,1]=0;x2[2,2]=0;
    x1[2:n,1:2]=x[1:n-1,1:2];x2[3:n,1:2]=x[1:n-2,1:2];

    /** Iterações – Bayesiana **/
    C=j(2,2,0);
    P=j(2,1,0);P1=j(2,1,0);P2=j(2,1,0);P3=j(2,1,0);P4=j(2,1,0);
    B=j(2,2,0);B1=j(2,2,0);B2=j(2,2,0);B3=j(2,2,0);B4=j(2,2,0);
    vtheta=j(2,1,0);
    vbeta=j(2,1,0);
    vphi=j(2,1,0);
    A=j(2,1,0);
    A1=j(2,1,0);A2=j(2,1,0);A3=j(2,1,0);
    A4=j(2,1,0);A5=j(2,1,0);A6=j(2,1,0);
    NN=j(2,2,0);
```

```

NN1=j(2,2,0);NN2=j(2,2,0);NN3=j(2,2,0);NN4=j(2,2,0);
M=j(2,1,0);
M1=j(2,1,0);M2=j(2,1,0);M3=j(2,1,0);M4=j(2,1,0);M5=j(2,1,0);M6=j(2,1,0);
nit_toburn=int(nit*pbur_in);
vet_res=j(1,7,0);
n_efet=nit-nit_toburn;
it_val=1;
/*valores iniciais*/
if cad=1 then
do;
theta1=-0.4;theta2=0.3;beta1=-2.5;beta2=2.5;phi1=-0.2;phi2=0.6;tau=1.5;
end;
if cad=2 then
do;
theta1=0.3;theta2=0.4;beta1=-1.0;beta2=1.5;phi1=0.6;phi2=-0.1;tau=2.0;
end;
do it=1 to nit;
/* calculo das posteriores*/
/*condicional completa para theta*/
/*montando as matrizes necessárias*/
/*matriz C - posteriori de theta*/
C[1,1]=at1`*at1;C[1,2]=at1`*at2;
C[2,1]=at2`*at1;C[2,2]=at2`*at2;
/*vetor P- posteriori de theta*/
P1[1]=y`*at1;
P2[1]=phi1*at1`*y1+phi2*at1`*y2;
P3[1]=beta1*at1`*x[,1]+beta2*at1`*x[,2];
P4[1]=phi1*beta1*at1`*X1[,1]+phi1*beta2*at1`*X1[,2]
+phi2*beta1*at1`*X2[,1]+phi2*beta2*at1`*X2[,2];
P1[2]=y`*at2;
P2[2]=phi1*at2`*y1+phi2*at2`*y2;
P3[2]=beta1*at2`*x[,1]+beta2*at2`*x[,2];
P4[2]=phi1*beta1*at2`*X1[,1]+phi1*beta2*at2`*X1[,2]
+phi2*beta1*at2`*X2[,1]+phi2*beta2*at2`*X2[,2];
P=P1-P2-P3+P4;
vtheta[1]=rannor(seed);vtheta[2]=rannor(seed);
mu=inv(C)*p;sigma=(1/tau)*inv(C);
st=root(sigma);
vtheta=st`*vtheta+mu;
/*atualizando theta*/
theta1=vtheta[1];
theta2=vtheta[2];

```

```

/*condicional completa para beta */
/*vetor B - posteriori para beta*/
    B2[1,1]=phi1*x[,1]*x1[,1]+phi2*x[,1]*x2[,1];
    B2[1,2]=phi1*x[,1]*x1[,2]+phi2*x[,1]*x2[,2];
    B2[2,1]=phi1*x[,2]*x1[,1]+phi2*x[,2]*x2[,1];
    B2[2,2]=phi1*x[,2]*x1[,2]+phi2*x[,2]*x2[,2];
    B1[1,1]=x[,1]*x[,1];B1[1,2]=x[,1]*x[,2];
    B1[2,1]=x[,2]*x[,1];B1[2,2]=x[,2]*x[,2];
    B3=B2`;
    B4[1,1]=phi1*phi1*X1[,1]*X1[,1]+phi1*phi2*X1[,1]*X2[,1]+phi2*phi1*
X2[,1]*X1[,1]+phi2*phi2*X2[,1]*X2[,1];
    B4[1,2]=phi1*phi1*X1[,1]*X1[,2]+phi1*phi2*X1[,1]*X2[,2]+phi2*phi1*
X2[,1]*X1[,2]+phi2*phi2*X2[,1]*X2[,2];
    B4[2,1]=phi1*phi1*X1[,2]*X1[,1]+phi1*phi2*X1[,2]*X2[,1]+phi2*phi1*
X2[,2]*X1[,1]+phi2*phi2*X2[,2]*X2[,1];
    B4[2,2]=phi1*phi1*X1[,2]*X1[,2]+phi1*phi2*X1[,2]*X2[,2]+phi2*phi1*
X2[,2]*X1[,2]+phi2*phi2*X2[,2]*X2[,2];
    B=B1-B2-B3+B4;
/*vetor A - posteriori para beta*/
    A1[1]=y*X[,1];A1[2]=y*X[,2];
    A2[1]=phi1*y*X1[,1]+phi2*y*X2[,1];
    A2[2]=phi1*y*X1[,2]+phi2*y*X2[,2];
    A3[1]=phi1*y1*X[,1]+phi2*y2*X[,1];
    A3[2]=phi1*y1*X[,2]+phi2*y2*X[,2];
    A4[1]=phi1*phi1*y1*X1[,1]+phi1*phi2*y1*X2[,1]+phi2*phi1*y2*X1[,1]
+phi2*phi2*y2*X2[,1];
    A4[2]=phi1*phi1*y1*X1[,2]+phi1*phi2*y1*X2[,2]+phi2*phi1*y2*X1[,2]
+phi2*phi2*y2*X2[,2];
    A5[1]=theta1*at1*X[,1]+theta2*at2*X[,1];
    A5[2]=theta1*at1*X[,2]+theta2*at2*X[,2];
    A6[1]=theta1*phi1*at1*X1[,1]+theta1*phi2*at1*X2[,1]+theta2*phi1*at2*
X1[,1]+theta2*phi2*at2*X2[,1];
    A6[2]=theta1*phi1*at1*X1[,2]+theta1*phi2*at1*X2[,2]+theta2*phi1*at2*
X1[,2]+theta2*phi2*at2*X2[,2];
    A=A1-A2-A3+A4-A5+A6;
    mu=inv(B)*A;sig=(1/tau)*inv(B);
    st=root(sigma);
    vbeta[1]=rannor(seed);
    vbeta[2]=rannor(seed);
    vbeta=st*vbeta+mu;
/*atualizando beta*/
    beta1=vbeta[1];

```

```

    beta2=beta[2];
/*posteriori do phi- condicional completa*/
    NN1[1,1]=y1`*y1;NN1[1,2]=y1`*y2;NN1[2,1]=y2`*y1;NN1[2,2]=y2`*y2;
    NN2[1,1]=beta1*y1`*X1[,1]+beta2*y1`*X1[,2];NN2[1,2]=beta1*y1`*X2[,1]
+beta2*y1`*X2[,2];
    NN2[2,1]=beta1*y2`*X1[,1]+beta2*y2`*X1[,2];NN2[2,2]=beta1*y2`*X2[,1]
+beta2*y2`*X2[,2];
    NN3=NN2`;
    NN4[1,1]=beta1*beta1*X1[,1]*X1[,1]+beta1*beta2*X1[,1]*X1[,2]+beta2*
beta1*X1[,2]*X1[,1]+beta2*beta2*X1[,2]*X1[,2];
    NN4[1,2]=beta1*beta1*X1[,1]*X2[,1]+beta1*beta2*X1[,1]*X2[,2]+beta2*
beta1*X1[,2]*X2[,1]+beta2*beta2*X1[,2]*X2[,2];
    NN4[2,1]=beta1*beta1*X2[,1]*X1[,1]+beta1*beta2*X2[,1]*X1[,2]+beta2*
beta1*X2[,2]*X1[,1]+beta2*beta2*X2[,2]*X1[,2];
    NN4[2,2]=beta1*beta1*X2[,1]*X2[,1]+beta1*beta2*X2[,1]*X2[,2]+beta2*
beta1*X2[,2]*X2[,1]+beta2*beta2*X2[,2]*X2[,2];
    NN=NN1-NN2-NN3+NN4;
/*matriz M*/
    M1[1]=y`*y1;M1[2]=y`*y2;/*neste ponto tem um erro de impressao na tese*/
    M2[1]=theta1*y1`*at1+theta2*y1`*at2;
    M2[2]=theta1*y2`*at1+theta2*y2`*at2;
    M3[1]=beta1*y`*x1[,1]+beta2*y`*x1[,2];
    M3[2]=beta1*y`*x2[,1]+beta2*y`*x2[,2];
    M4[1]=beta1*y1`*x[,1]+beta2*y1`*x[,2];
    M4[2]=beta1*y2`*x[,1]+beta2*y2`*x[,2];
    M5[1]=beta1*theta1*x1[,1]*at1+beta1*theta2*x1[,1]*at2+beta2*theta1*
x1[,2]*at1+beta2*theta2*x1[,2]*at2;
    M5[2]=beta1*theta1*x2[,1]*at1+beta1*theta2*x2[,1]*at2+beta2*theta1*
x2[,2]*at1+beta2*theta2*x2[,2]*at2;
    M6[1]=beta1*beta1*x[,1]*x1[,1]+beta1*beta2*x[,1]*x1[,2]+beta2*beta1*
x[,2]*x1[,1]+beta2*beta2*x[,2]*x1[,2];
    M6[2]=beta1*beta1*x[,1]*x2[,1]+beta1*beta2*x[,1]*x2[,2]+beta2*beta1*
x[,2]*x2[,1]+beta2*beta2*x[,2]*x2[,2];
    M=M1-M2-M3-M4+M5+M6;
    mu=inv(NN)*M;sig=(1/tau)*inv(NN);
    st=root(sigma);
    vphi[1]=rannor(seed);
    vphi[2]=rannor(seed);
    vphi=st`*vphi+mu;
/*atualizando phi*/
    phi1=vphi[1];
    phi2=vphi[2];

```

```

/* posteriori para tau - condicional completa*/
z=j(n,1,0);
z=y-y1*phi1-y2*phi2-at1*theta1-at2*theta2;
UB=x[,1]*beta1+x[,2]*beta2-beta1*(phi1*x1[,1]+phi2*x2[,1])-beta2*
(phi1*x1[,2]+phi2*x2[,2]);
T=(z-UB)*(z-UB);
/*este é uma opção que eu fiz para ser avaliada*/
/*atualizando at*/
/*at=z-ub; pode ser obtido dessa forma também*/
at=y-(phi1*y1+phi2*y2+beta1*x[,1]+beta2*x[,2]+theta1*at1+theta2*at2
-phi1*x1[,1]*beta1-phi2*x2[,1]*beta1-phi1*x1[,2]*beta2-phi2*x2[,2]*beta2);
at1[1]=0;at1[2]=at[1];
at2[1]=0;at2[2]=0;
do ii=3 to n;
    at1[ii]=at[ii-1];
    at2[ii]=at[ii-2];
end;
/*fim da atualização*/
bb1=(n-2)/2;
c1=T/2;
tau1=rangam(seed,bb1)/c1;
/*atualizando tau*/
tau=tau1;
/*Burn in de pbun_in observações*/
if it>0 then (Burn-in = 0)
do;
    vet_res=vtheta`||vbeta`||vphi`||tau;
    if cad=1 then res[it_val,1:7]=vet_res;
    if cad=2 then res[it_val,8:14]=vet_res;
    it_val=it_val+1;
end;
end;/*do iterações*/
end;/*do cadeias*/
/*Res com cadeia 1-7 e cadeia 2 8-14*/
/* Inserir se necessário aqui o critério de Gelman & Rubin (1992) */
/* Inserir se necessário aqui o critério de Brooks & Gelman (1998), critério do
Determinante e critério do Traço */
create arma22 var {theta11, theta21, beta11, beta21, phi11, phi21, tau1,
theta12, theta22, beta12, beta22, phi12, phi22, tau2};
append from res;
quit;

```

PROGRAMA DOS CRITÉRIOS

Critério de Gelman & Rubin (1992)

Programa utilizado para o modelo com 7 parâmetros

```
k=20;k1=k/2;m=2;total=int((nit-k)/2)+1;
ngel=(int((nit-k)/2)+1);
Gel_Rub=j(ngel,9);cont=1;
do t=1 to total;
  Gel_rub[cont,1]=k1;Gel_rub[cont,2]=k;
  do par=1 to 7;
    p1=0;p2=0;Sx1=0;sx2=0;
    B=0;W=0;
    P1=sum(res[k1+1:k,par]);
    P2=sum(res[k1+1:k,par+7]);
    sx1=sum(res[k1+1:k,par]#res[k1+1:k,par]);
    sx2=sum(res[k1+1:k,par+7]#res[k1+1:k,par+7]);
    B=(p1*p1/k1+p2*p2/k1-(p1+p2)**2/(2*k1))/(m-1);
    W=((sx1-p1*p1/k1)/(k1-1)+((sx2-p2*p2/k1)/(k1-1)))/m;
    df = ((k1-1)*W/k1+(m+1)*B/(k1*m))**2/((k1-1)**2*W**2/
    (k1**2*(2*k1-2))+(m+1)**2*B**2/(k1**2*m**2*(m-1)));
    Rroot=((((k1-1)/k1+(m+1)/(m*k1)*B/W)*(df+3)/(df+1))**0.5;
    *print par k1 k B W Rroot df;
    Gel_rub[cont,par+2]=rroot;
  end; /*do par*/
  cont=cont+1;
  k=k+2;k1=k/2;
end; /*do total*/
create data_uni var{k1, k, R1, R2, R3, R4, R5, R6, R7};
append from Gel_rub;
```

Critério de Brooks & Gelman (1998) (cálculo alternativo), Critério do

Determinante e Critério do Traço.

Programa utilizado para o modelo com 7 parâmetros

```
k=20;k1=k/2;m=2;total=int((nit-k)/2)+1;Resp=j(total,4,0);
do t=1 to total;
  Q1=i(k1)-(1/k1)*J(k1,k1,1);
  X1=res[k1+1:k,1:7];
  X2=res[k1+1:k,8:14];
  W=((1/(k1-1))*X1`*Q1*X1+(1/(k1-1))*X2`*Q1*X2)/m;
```

```

T1=j(7,1,0);T2=j(7,1,0);
s1=sum(res[k1+1:k,1]);
s2=sum(res[k1+1:k,2]);
s3=sum(res[k1+1:k,3]);
s4=sum(res[k1+1:k,4]);
s5=sum(res[k1+1:k,5]);
s6=sum(res[k1+1:k,6]);
s7=sum(res[k1+1:k,7]);
T1[1]=s1;T1[2]=s2;T1[3]=s3;T1[4]=s4;T1[5]=s5;T1[6]=s6;T1[7]=s7;
s1=sum(res[k1+1:k,8]);
s2=sum(res[k1+1:k,9]);
s3=sum(res[k1+1:k,10]);
s4=sum(res[k1+1:k,11]);
s5=sum(res[k1+1:k,12]);
s6=sum(res[k1+1:k,13]);
s7=sum(res[k1+1:k,14]);
T2[1]=s1;T2[2]=s2;T2[3]=s3;T2[4]=s4;T2[5]=s5;T2[6]=s6;T2[7]=s7;
B=((T1*T1'+T2*T2')/k1-(T1+T2)*(T1+T2)'/(k1*m))/(m-1);
V=(k1-1)/k1*W+B/k1+B/(m*k1);
Sw=root(W);
Swinv=inv(Sw);
H=Swinv*B/k1*Swinv;
auh1=max(eigval(H));
Rp1=(k1-1)/k1+(m+1)/m*auh1;
Rp2=det(V)/Det(W);
Rp3=trace(V)/trace(W);
resp[t,1]=k1;resp[t,2]=Rp1;resp[t,3]=Rp2;resp[t,4]=Rp3;
k=k+2;k1=k/2;
end;
create data_mult var{k1, Rp1, Rp2, Rp3};
append from resp;

```

Os critérios a baixo foram programados para o software R[®] estes programas necessitam que o pacote BOA (Bayesian Output Analysis) esteja sendo executado para funcionarem.

CRITÉRIO DE GEWEKE (1992)

```

boa.start()
boa.par(path =./simulacao/Normal trivariada/Alta correlacao/"")
boa.chain.import("dados")
work <- boa.chain("work")
k=20
k1=k/2
total=2491 #(5000-20)/2+1
ngew=total
cont=1
gew=matrix(0,ngew,16)
for (ii in 1:total) {
  gew[ii,15]<-k
  gew[ii,16]<-k1
  iter <- unique(boa.iter(work[["dados"]]))
  # n <- length(iter)
  n <- k
  link.first <- boa.getiter(work[["dados"]], iter[k1:(k1+round(0.1 * n))])
  link.last <- boa.getiter(work[["dados"]], iter[(n - round(0.5 * k1) + 1):n])
  result <- (colMeans(link.first) - colMeans(link.last)) /
    sqrt(boa.gewekePwr(link.first) / nrow(link.first) +
          boa.gewekePwr(link.last) / nrow(link.last))
  result <- cbind(result, 2 * (1 - pnorm(abs(result))))
  result1<-result
  for (j in 1:7)
  {
    gew[ii,j]<-result1[j,1]
    gew[ii,j+7]<-result1[j,2]
  }
  k<-k+2
  k1<-k/2
}
write.table(gew,file="gew.txt")

```

CRITÉRIO DE HEIDELBERGER & WELCH (1983)

```

boa.start()
boa.par(path="C:\\caminho\\")
boa.chain.import("dados") #nome do arquivo de dados sem a extensão .txt
work <- boa.chain("work")
k=100
total=300
ngew=total
cont=1
gew1=matrix(0,ngew,7)
gew2=matrix(0,ngew,7)
gew3=matrix(0,ngew,7)
gew4=matrix(0,ngew,7)
gew5=matrix(0,ngew,7)
alpha=0.05
error=0.1
for(i in names(work)) {
  header <- paste("\nChain: ", i, "\n", sep = "")
  #cat(header, rep("-", nchar(header) - 2), "\n\n", sep = "")
  #print(boa.handw(work[[i]], error, alpha))
  link<-work[[i]]
  error<-error
  alpha<-alpha
}
#function(link, error, alpha)
for(ii in 1:total) {
  pnames <- boa.pnames(link)
  stest <- keep <- I <- htest <- xbar <- halfwidth <- structure(rep(NA,
length(pnames)), names = pnames)
  iter <- unique(boa.iter(link))
  n <- length(iter)
  n.min <- round(0.5 * k)
  idx.drop <- -1 * 1:max(round(0.10 * k), 1)
  S0 <- boa.gewekePwr(boa.getiter(link, iter[(k - n.min + 1):k]))
  q.upper <- qnorm(1 - alpha / 2)
  for(i in pnames) {
    parm1 <- boa.getparms(link, i)
    parm <- as.matrix(parm1[1:k,])
    piter1 <- iter
    piter <- iter[1:k]
    keep[i] <- k
    stest[i] <- "failed"
    htest[i] <- "failed"
  }
}

```

```

while((keep[i] >= n.min) && (stest[i] == "failed")) {
  n.parm <- nrow(parm)
  xbar[i] <- mean(parm)
  halfwidth[i] <- q.upper * sqrt(S0[i] / n.parm)
  if(abs(halfwidth[i] / xbar[i]) <= error) htest[i] <- "passed"
  B <- cumsum(parm) - xbar[i] * 1:n.parm
  Bsqr <- (B * B) / (n.parm * S0[i])
  I[i] <- (2 * sum(Bsqr[seq(2, n.parm - 2, by = 2)]) +
    4 * sum(Bsqr[seq(1, n.parm - 1, by = 2)]) + Bsqr[n.parm]) /
    (3 * n.parm)
  if(I[i] < 0.46) {
    stest[i] <- "passed"
  } else {
    piter <- piter[idx.drop]
    keep[i] <- length(piter)
    parm <- boa.getiter(parm, piter)
  }
}
}
result <- data.frame(stest, keep, k - keep, I, htest, xbar, halfwidth)
names(result) <- c("Stationarity Test", "Keep", "Discard", "C-von-M",
  "Halfwidth Test", "Mean", "Halfwidth")
htestm <- as.matrix(htest)
halfwidthm <- as.matrix(halfwidth)
stestm <- as.matrix(stest)
keepm <- as.matrix(keep)
nkeepm <- as.matrix(k - keep)
gew1[ii,]<-t(htestm)
gew2[ii,]<-t(halfwidthm)
gew3[ii,]<-t(stestm)
gew4[ii,]<-t(keepm)
gew5[ii,]<-t(nkeepm)
k<-k+100
}

```

Para o modelo normal trivariado basta adaptar os vetores para 3 parâmetros (variáveis) ao invés de 7 parâmetros.