**Análise bayesiana
do modelo fatorial dinâmico
para um vetor de séries temporais
usando distribuições elípticas**

LÍVIA COSTA BORGES

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO DE
DOUTOR EM CIÊNCIAS

Programa: **Estatística**

Orientadora: **Profa. Dra. Lúcia Pereira Barroso**

São Paulo, agosto de 2008.

**Análise bayesiana
do modelo fatorial dinâmico
para um vetor de séries temporais
usando distribuições elípticas**

Este exemplar corresponde à redação final
da tese devidamente corrigida e
defendida por Lívia Costa Borges
e aprovada pela comissão julgadora.

Banca Examinadora:

Profa. Dra. Lúcia Pereira Barroso (presidente) - IME/USP

Prof. Dr. Pedro A. Morettin - IME/USP

Prof. Dr. Daniel Furtado Ferreira - DEX/UFLA

Prof. Dr. Marinho Gomes de Andrade Filho - ICMC/USP

Profa. Dra. Thelma Sáfadi - DEX/UFLA

*Aos meus pais, Albertino e Sara
e às minhas irmãs Andreísa e Valéria.*

Agradecimentos

Desejo expressar meus mais sinceros agradecimentos à minha família e aos meus amigos queridos por me acompanharem e apoiarem, sempre. Obrigado por serem tão pacientes comigo e por suportarem a minha ausência várias vezes durante este período de dedicação quase exclusiva aos estudos. Aos meus pais e irmãs agradeço pelo estímulo e apoio incondicional, desde o primeiro instante. Aos meus sobrinhos Gabriel e Vinícius, agradeço pelo fato de existirem em minha vida. Mesmo sem entender suportaram minha ausência.

À Prof. Lúcia Pereira Barroso, pela orientação, paciência, críticas e sugestões feitas durante a orientação.

À minha amiga Tatiana Terabayashi Melhado, pela amizade e disponibilidade desde o nosso primeiro contato. O seu apoio foi fundamental durante esta jornada.

Minha gratidão a todo o pessoal do Departamento de Ciências Exatas da Universidade Federal de Lavras - DEX/UFLA, especialmente aos professores Thelma Sáfydi pela co-orientação e Daniel Furtado Ferreira, por apostar e acreditar em minha capacidade e pela disponibilidade em ajudar sempre. Ao Professor Daniel Furtado dedico esta vitória. Serei eternamente grata aos meus colegas de mestrado no DEX, Marcelo Cirillo e Denys Nogueira, pelo apoio.

A Juan Carlos Ruilova, pela atenção especial concedida na etapa final do meu trabalho.

Aos meus amigos do Citibank, Carlos Auricchio, Ricardo Almeida, Jack, Fer e Paulo Sirimarco pela força e pelos momentos legais proporcionados.

Obrigado, mil vezes obrigado, a meus mentores e anjos da guarda por me mostrarem a luz, me acompanharem e me orientarem em minha jornada. Um agradecimento especial ao meu terapeuta Dr. Mateus, que me ajudou a crescer muito.

À Capes, pela bolsa de doutorado concedida.

Resumo

A análise fatorial é uma importante ferramenta estatística que tem amplas aplicações práticas e explica a correlação entre um grande número de variáveis observáveis em termos de um pequeno número de variáveis não observáveis, conhecidas como variáveis latentes. A proposta deste trabalho é fazer a análise Bayesiana, que incorpora à análise o conhecimento que se tenha sobre os parâmetros antes da coleta dos dados, do modelo fatorial dinâmico na classe de modelos elípticos multivariados, assumindo que a um vetor de q séries temporais pode-se ajustar um modelo fatorial com $k < q$ fatores mais um ruído branco, e que a parte latente segue um modelo vetorial auto-regressivo. A classe de modelos elípticos citada acima é rica em distribuições simétricas com caudas mais pesadas que as da distribuição normal, característica importante na análise de séries financeiras. Essa classe inclui as distribuições t de Student, exponencial potência, normal contaminada, entre outras. A inferência sobre os parâmetros foi feita utilizando métodos de Monte Carlo via Cadeias de Markov, com os algoritmos Metropolis-Hastings e Griddy-Gibbs, através da obtenção das distribuições a posteriori dos parâmetros e dos fatores. A determinação da convergência do processo foi feita por técnicas gráficas e pelos métodos de Geweke (1992), de Heidelberger e Welch (1983) e Half-Width. O método foi ilustrado usando dados reais e simulados.

Abstract

The factor analysis is an important statistical tool that has wide practical applications and it explains the correlation among a large number of observable variables in terms of a small number of unobservable variables, known as latent variables. The proposal of this work is the Bayesian analysis, which incorporates the information we have concerning the parameters before collecting data into the analysis of a dynamical factor model in the class of multivariate elliptical models, where the factors follow a multivariate autoregressive model, assuming that a vector of q time series can be adjusted with $k < q$ factors and a white noise. The class of elliptical models is rich in symmetrical distributions with heavier tails than the normal distribution, which is an important characteristic in financial series analysis. This class includes t-Student, power exponential, contaminated normal and other distributions. The parameters inference was made through Monte Carlo Markov Chain methods, with Metropolis-Hastings and Griddy-Gibbs algorithms, by obtaining the parameters and factors posteriori distributions. The convergence process was made through graphical technics and by Geweke (1992) and by Heidelberger and Welch (1983) and Half-Width methods. The method was illustrated using simulated and real data.

Índice

1	Introdução	1
2	Distribuições Elípticas	5
2.1	Definições e Exemplos	7
2.2	Propriedades da Distribuição Elíptica	12
2.3	Distribuição t multivariada	14
3	Análise Bayesiana	16
3.1	Inferência Bayesiana	16
3.2	Métodos de Monte Carlo via Cadeias de Markov	18
3.2.1	O Algoritmo Metropolis-Hastings	20
3.2.2	O Amostrador Griddy-Gibbs	25
3.2.2.1	O Algoritmo Griddy-Gibbs	25
3.2.3	Procedimentos para Monitoração da Convergência	27
4	Estimação Bayesiana do Modelo Fatorial Proposto	32
4.1	Modelo Fatorial Dinâmico	33
4.2	Distribuições a Posteriori	35
4.3	Simulação Estocástica via Cadeias de Markov	38
4.3.1	Simulação via Algoritmo Metropolis-Hastings	38
4.3.2	Simulação via Amostrador Griddy-Gibbs	38
4.4	Monitoração da Convergência	39

4.4.1	Convergência do Algoritmo Metropolis-Hastings	39
4.4.2	Convergência do Amostrador Griddy-Gibbs	40
4.4.3	Simulação via Máxima Verossimilhança	44
5	Aplicação	48
6	Conclusões	58
A	Programa R - simulação e análise via Metropolis-Hastings.	60
B	Programa S-Plus - simulação e análise via Griddy-Gibbs.	68
C	Séries financeiras utilizadas na aplicação	80
	Bibliografia	84

Capítulo 1

Introdução

Inovações metodológicas e aplicações da análise fatorial, ou seja, modelos com estruturas latentes, têm se desenvolvido rapidamente nos últimos anos, devido ao incremento de ferramentas computacionais apropriadas.

A análise fatorial tem como objetivo principal descrever as relações entre um conjunto de variáveis observadas e outras não observáveis, conhecidas como variáveis latentes ou fatores.

O estudo de variáveis latentes teve início no começo do Século XX, com a criação da análise fatorial na área de Psicometria por Spearman (1904), para analisar escores individuais em testes mentais. Esse primeiro trabalho envolveu o uso de somente uma variável latente, a habilidade geral ou inteligência. Thurstone (1944) estendeu essa teoria para modelos com mais variáveis, introduzindo a análise fatorial múltipla, também na área de Psicologia. Mais detalhes a respeito podem ser encontrados em Thurstone (1947).

Somente após 1960 é que o assunto teve tratamento estatístico e um dos trabalhos dessa época é o livro de Lawley e Maxwell (1971).

Ainda nos anos 60 teve início o desenvolvimento de *covariance structure analysis* com o trabalho de Jöreskog (1966), sendo esta técnica conhecida como *modelagem de equações estruturais*. Esse modelo nasceu na área de Ciências Sociais na busca de relações de causalidade e seus resultados são baseados na comparação da matriz de covariância estimada

pelos dados livres de estrutura com a matriz de covariância estimada dado o modelo. Vários softwares como o AMOS, o PRELIS, o CALIS e o LISREL foram desenvolvidos para o ajuste desse modelo, sendo que o último tornou-se um sinônimo do mesmo.

Atualmente, análise fatorial tem sido aplicada em um grande número de situações, por exemplo, para analisar quantidades econômicas, testes de atitudes e comportamentos, medidas físicas, etc.

Peña e Box (1987) propuseram um método para identificar fatores comuns em um vetor de séries temporais. Eles assumem que um vetor de séries temporais observado de dimensão q pode ser escrito em função de uma matriz de parâmetros desconhecidos com dimensão $(q \times k)$, um vetor não observável k -dimensional e um vetor de erros q -dimensional (ruído branco), com matriz de covariância Σ_ϵ de posto completo. Esse modelo é relevante se e somente se $k < q$ ou se $k = q$ e $\Sigma_\epsilon = \mathbf{0}$. No primeiro caso ocorre redução da dimensionalidade sem perda de informação e no segundo caso, em que $k = q$, um problema interessante é encontrar uma transformação linear das séries que permitam uma representação simples do sistema. Os autores mostraram que uma representação simples usando um pequeno número de fatores foi adequada para o modelo de séries temporais e que, diferentemente da análise fatorial para variáveis observadas em unidades independentes, as restrições colocadas pela presença de fatores comuns na correlação, autocorrelação parcial e matrizes paramétricas tornam a identificação do número de fatores e a estimativa da matriz de cargas fatoriais relativamente fáceis.

Diebold e Nerlove (1989) utilizaram um modelo multivariado para um vetor de séries temporais em que a parte latente segue um modelo ARCH (autoregressive conditional heterocedastic) para estudar padrões de volatilidade temporal na taxa de câmbio. Primeiramente, estimaram o modelo univariado e em seguida usaram as informações obtidas para especificar o modelo multivariado. A principal finalidade da aproximação multivariada nesse trabalho foi explorar a estrutura fatorial, que facilita a estimação através da redução do número de parâmetros a serem estimados.

Uma estrutura latente nos modelos de séries temporais foi proposta por Harvey e Chung (2000), aplicada à série de desemprego no Reino Unido. Esse trabalho teve abor-

dagem clássica. Sob abordagem Bayesiana, em que qualquer informação obtida antes da coleta dos dados pode ser incorporada à análise, Lee e Shi (2000) obtiveram estimativas conjuntas dos escores fatoriais e dos parâmetros estruturais de covariância, através da obtenção conjunta das distribuições a posteriori. A idéia desses autores foi tratar os escores fatoriais latentes como *missing data* e aumentá-los com os dados observados na geração de uma seqüência de observações aleatórias da distribuição a posteriori, por amostrador de Gibbs. As estimativas Bayesianas foram tomadas como as médias amostrais destas observações aleatórias. Esses autores desenvolveram uma aproximação Bayesiana em um modelo de análise fatorial confirmatória, ou seja, o modelo no qual o pesquisador impõe restrições de acordo com o seu conhecimento sobre o que se está estudando, técnica essa conhecida como modelagem de equações estruturais. Mais detalhes sobre este tipo de modelagem podem ser encontrados em Latif (2000) e Melhado (2004).

Procedimentos Bayesianos menos gerais que os desenvolvidos por Lee e Shi (2000) foram desenvolvidos por Bartholomew (1981), que obteve estimativas dos escores fatoriais em um modelo de análise fatorial confirmatória, sendo que os elementos do vetor paramétrico não foram estimados. Baseado no mesmo modelo confirmatório, Lee (1981) usou uma aproximação Bayesiana hierárquica para estimar os parâmetros do modelo, mas os escores fatoriais não foram estimados.

Sáfadi e Peña (2007) consideraram q séries temporais nas quais se ajustou um modelo fatorial dinâmico, em que os fatores seguem um modelo vetorial auto-regressivo. Os autores utilizaram o procedimento Bayesiano para análise do modelo fatorial dinâmico, considerando erros independentes e normalmente distribuídos. A inferência sobre os parâmetros foi feita através do amostrador de Gibbs e do algoritmo Metropolis-Hastings e a convergência foi verificada utilizando-se o critério de Gelman e Rubin (1992) para cadeias paralelas. Nesse trabalho, foram obtidas distribuições a posteriori conjuntas para os parâmetros do modelo e para os fatores.

Métodos de simulação iterativos Monte Carlo via Cadeias de Markov têm aberto acesso ao tratamento Bayesiano completo de modelos fatoriais, como os desenvolvidos em, por exemplo, Geweke e Zhou (1996), Polasek (1997) e Arminger e Muthén (1998). Ao

considerar componentes fatoriais dinâmicos em modelagem de séries temporais financeiras, pode-se citar Aguiar e West (2000) e Pitt e Shephard (1998). Os estudos citados, assim como outros, exploram inferência Bayesiana em modelos fatoriais latentes.

Lopes e West (2004) introduziram no contexto de análise fatorial multivariada o algoritmo *Reversible Jump Markov Chain Monte Carlo*, que permite ajustar e comparar modelos com diferentes números de fatores, quando se considera modelos fatoriais com número de fatores desconhecido. Em problemas práticos, especialmente com grandes valores de q , o número de fatores k será freqüentemente menor em relação a q , de maneira que a estrutura de variância-covariância é explicada pelos fatores comuns.

O modelo apresentado por Sáfadi e Peña (2007), citado anteriormente, é estendido no presente trabalho, supondo que os erros seguem uma distribuição multivariada elíptica. São considerados os erros em uma classe mais ampla de distribuições, ou seja, a classe das distribuições elípticas multivariadas, sendo esta, uma classe geral de distribuições simétricas que são invariantes por transformações ortogonais, e que inclui a distribuição normal como um caso especial. Mais detalhes são apresentados no Capítulo 2, com definições, exemplos e propriedades desta classe de distribuições.

No Capítulo 3, apresentamos conceitos básicos sobre análise Bayesiana, dando ênfase a Métodos de Monte Carlo via Cadeias de Markov. São apresentados os tópicos relacionados ao entendimento dos processos estocásticos via o algoritmo Metropolis-Hastings e o amostrador Griddy-Gibbs. Alguns procedimentos disponíveis na literatura para monitoração da convergência da cadeia de Markov são considerados ainda nesse capítulo.

O modelo proposto é apresentado no Capítulo 4, onde pode-se também encontrar as distribuições a posteriori dos parâmetros, bem como os resultados dos programas para os ajustes do modelo t , caso particular da distribuição elíptica. No Capítulo 5 apresentamos uma aplicação em dados reais, que é seguido das conclusões.

Capítulo 2

Distribuições Elípticas

A maior parte da teoria de inferência estatística para variáveis contínuas está desenvolvida em torno do modelo normal. Contudo, em muitas situações este modelo é impróprio, como por exemplo, quando os dados provêm de uma distribuição com caudas mais pesadas que as da distribuição normal ou quando há influência de *outliers*. Assim, em muitas situações práticas o uso da suposição de normalidade pode levar a inferências pouco apropriadas para os parâmetros de interesse.

Segundo Arellano-Vale (1994), a inferência estatística com base na distribuição normal pode ser generalizada para fins de cobrir uma classe mais ampla de modelos paramétricos que preservam a simetria elíptica do modelo normal. Essa classe é conhecida como classe das distribuições elípticas, gerada por distribuições com caudas mais leves ou mais pesadas que as da distribuição normal e que superam a falta de robustez com respeito a observações aberrantes, sendo portanto uma ferramenta útil para a realização de estudos de robustez da teoria normal.

A partir da década de 70, as distribuições elípticas vêm ocupando um importante e crescente lugar na teoria estatística. As distribuições esféricas, que são as distribuições elípticas na sua forma padronizada, são caracterizadas por sua invariância com respeito a transformações ortogonais e formam uma classe geral de distribuições com a mesma simetria esférica da distribuição normal padrão. As distribuições elípticas podem ser gene-

realizadas mediante transformações lineares de posição e escala das distribuições esféricas, sendo que cada distribuição esférica é um elemento particular de alguma família de distribuições elípticas.

Pode-se dizer que as distribuições elípticas formam uma classe generalizada de famílias paramétricas de distribuições de probabilidades que preservam a estrutura simétrica da distribuição normal, sendo a distribuição normal um caso particular desta classe. Outros elementos típicos na classe das distribuições elípticas são, por exemplo, os componentes da família de distribuições normais compostas (veja Exemplo 2.1.2), uma subclasse que tem sido enfatizada especialmente na estatística Bayesiana, que é dada por distribuições que são misturas de distribuições normais e que contêm a família de distribuições normais contaminadas e a família de distribuições t .

Diversos autores têm estudado as propriedades das distribuições elípticas. O primeiro estudo teórico a respeito, bem organizado, foi feito por Kelker (1970). Alguns resultados em termos de propriedades teóricas, assim como áreas de aplicação são dados em Devlin et al. (1976), Kariya e Eaton (1977) e Muirhead e Waternaux (1980), os quais se encontram resumidos em Muirhead (1982). Destacam-se também os trabalhos de Cambanis et al. (1981) e o de Chmielewski (1981), sendo que neste último existe uma ampla discussão sobre aplicações. Estudos mais recentes sobre as propriedades das distribuições elípticas podem ser encontrados em Mitchell e Krzanowski (1985), Berkane e Bentler (1987), Anderson et al. (1986) e Mitchell (1988; 1989). Um desenvolvimento resumido e bem organizado das distribuições esféricas e elípticas encontra-se em Fang et al. (1990, cap.2).

A distribuição elíptica q -variada pode ser obtida misturando a distribuição uniforme sobre a superfície da esfera unitária q -dimensional com uma distribuição radial, definida sobre $(0, \infty)$. Assim, cada distribuição elíptica ou sua versão padronizada, denominada distribuição esférica, é completamente caracterizada pela distribuição radial. Mais detalhes podem ser vistos em Arellano-Vale (1994). Em particular, a distribuição elíptica (ou esférica) normal q -variada é caracterizada pela distribuição radial qui-quadrado com q graus de liberdade. Já a distribuição elíptica ou esférica t q -variada com ν graus de liberdade é caracterizada pela distribuição radial de Fisher com q e ν graus de liberdade

(mais formalmente, pela distribuição beta de tipo II com parâmetros $\frac{q}{2}$ e $\frac{\nu}{2}$).

A teoria estatística desenvolvida em torno da classe das distribuições elípticas é baseada em dois enfoques para especificar o modelo estatístico, descritos a seguir:

- ✓ **Primeiro enfoque:** assume-se que a distribuição conjunta das variáveis de interesse é elíptica, de modo que estas variáveis podem ser não correlacionadas sem serem independentes, a menos que se suponha normalidade. Este enfoque é seguido pelos autores Anderson et al. (1986), Ghosh e Sinha (1980) e Zellner (1976). Os resultados destes autores mostram que grande parte da inferência estatística baseada no modelo normal permanece válida para estes modelos elípticos conjuntos ou dependentes.
- ✓ **Segundo enfoque:** supõe-se que as variáveis de interesse são independentes e que têm distribuições marginais elípticas. Este enfoque é seguido por Kano et al. (1993) e Lange et al. (1989). Neste caso, a inferência estatística deve ser baseada na teoria assintótica. A vantagem destes modelos elípticos independentes é que muitos deles, tais como o modelo t independente, implicam em inferência robusta em relação a *outliers*.

Vale a pena ressaltar que os dois enfoques somente coincidem sob normalidade e têm conseqüências diferentes sobre a teoria de inferência estatística.

A classe das distribuições simétricas esféricas e elípticas são extensões das distribuições normais multivariadas. De uma maneira geral, pode-se dizer que a distribuição esférica multivariada é uma extensão da distribuição normal multivariada padrão $N(\mathbf{0}, \mathbf{I})$ e a distribuição elíptica uma extensão da distribuição normal multivariada $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Nos últimos anos tem-se observado um vigoroso desenvolvimento das distribuições elípticas multivariadas como generalizações diretas da distribuição normal multivariada, as quais mantêm a maioria das propriedades atrativas da normal multivariada.

2.1 Definições e Exemplos

Podemos definir a distribuição elíptica de diversas maneiras, uma delas restringindo-se

ao caso não-singular e baseando-se em sua função densidade (veja, por exemplo, Muirhead (1982), Mitchell (1989) e Berkane e Bentler (1987)). Já se o interesse for incorporar o caso singular, podemos definir a distribuição com base em sua função característica (veja, por exemplo, Kelker (1970), Cambanis et al. (1981) e Anderson et al. (1986)).

Dada uma definição para a distribuição elíptica multivariada, a distribuição esférica multivariada pode ser estabelecida como caso particular. Neste trabalho, optamos por definir primeiro a distribuição esférica, caracterizada por sua invariância com respeito a transformações ortogonais, e depois estender a definição para as distribuições elípticas q -dimensionais.

Definição 2.1.1. *Dizemos que um vetor aleatório $\mathbf{Y}_{(q \times 1)}$ tem distribuição esférica multivariada, se $\mathbf{Y}$ e $\mathbf{\Gamma Y}$ têm a mesma distribuição para todas as matrizes ortogonais $\mathbf{\Gamma}_{(q \times q)}$, o que pode ser denotado por*

$$\mathbf{Y} \stackrel{\mathcal{D}}{=} \mathbf{\Gamma Y}, \forall \mathbf{\Gamma} : \mathbf{\Gamma}^\top \mathbf{\Gamma} = \mathbf{I}_q, \quad (2.1)$$

ou seja, a distribuição esférica é invariante por transformações ortogonais, o que impõe simetria esférica.

A distribuição esférica também pode ser definida com base em sua função característica. A função característica de um vetor aleatório $\mathbf{Y}_{(q \times 1)}$ é definida como:

$$\begin{aligned} \phi(\mathbf{t}) &= E(e^{i\mathbf{t}^\top \mathbf{Y}}) \\ &= h(\mathbf{t}^\top \mathbf{t}), \end{aligned} \quad (2.2)$$

em que $h(\cdot)$ é uma função escalar, denominada função geradora da função característica, e $\mathbf{t} = (t_1, \dots, t_q)^\top$ é um vetor de números reais.

Definição 2.1.2. *Dizemos que um vetor aleatório $\mathbf{Y}_{(q \times 1)}$ tem distribuição esférica se, e somente se, sua função característica, denotada por $\phi(\mathbf{t})$ satisfizer a seguinte condição:*

$$\begin{aligned}\phi(\mathbf{\Gamma}^\top \mathbf{t}) &= \phi(\mathbf{t}) \\ &= h(\mathbf{t}^\top \mathbf{t}),\end{aligned}\tag{2.3}$$

para alguma função $h(\cdot)$.

Segundo Arellano-Vale (1994), um vetor aleatório com distribuição elíptica não possui necessariamente uma densidade. No caso de ela existir, deverá ser da forma $g(\mathbf{y}^\top \mathbf{y})$, para alguma função não negativa g , função esta que pode definir uma densidade para alguma distribuição esférica, sendo portanto conhecida como função geradora de densidade. Assim a função densidade da distribuição esférica é dada por:

$$f(\mathbf{y}) = g(\mathbf{y}^\top \mathbf{y}), \mathbf{y} \in \mathbb{R}^q.\tag{2.4}$$

A Tabela 2.1 é apresentada em Fang et al. (1990) e mostra algumas subclasses de distribuições esféricas q -dimensionais com função densidade $f(\mathbf{y})$, $\mathbf{y} \in \mathbb{R}^q$. As duas últimas subclasses nesta tabela são apresentadas pela função característica $\phi(\mathbf{t})$, $\mathbf{t} \in \mathbb{R}^q$.

Mais detalhes sobre a classe das distribuições esféricas podem ser encontrados em Muirhead (1980; 1982) e em Fang et al. (1990).

Como citado anteriormente, podemos estender a definição de distribuições esféricas para definir distribuições elípticas q -dimensionais. Segundo Cambanis et al. (1981), a distribuição elíptica pode ser definida da seguinte maneira:

Definição 2.1.3. Dizemos que um vetor aleatório $\mathbf{Y}_{(q \times 1)}$ tem distribuição elíptica q -variada com vetor de posição $\boldsymbol{\mu}_{(q \times 1)}$ e matriz de dispersão $\boldsymbol{\Sigma}_{(q \times q)} > 0$, se a função característica de $\mathbf{Y}$ é da forma

$$\phi(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} h(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}), \mathbf{t} \in \mathbb{R}^q\tag{2.5}$$

para alguma função h . Dizemos que $\mathbf{Y}$ tem distribuição elíptica q -variada com parâmetros $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$, e h é denominada função geradora da função característica. Escrevemos $\mathbf{Y} \sim El_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h)$.

De forma equivalente, podemos dizer que $\mathbf{Y} \sim El_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h)$ se

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{AZ} \quad (2.6)$$

onde $\mathbf{A}$ é uma matriz $(q \times k)$ tal que $posto(\mathbf{A}) = k$, $\mathbf{A}^\top \mathbf{A} = \boldsymbol{\Sigma}$ e $\mathbf{Z}_{(k \times 1)}$ é um vetor aleatório tendo distribuição esférica k -variada com função característica $\psi(\mathbf{z}), \mathbf{z} \in \mathbb{R}^k$.

A função densidade de probabilidade da distribuição elíptica é dada por

$$f(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} g[(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})], \mathbf{y} \in \mathbb{R}^q, \quad (2.7)$$

onde a função geradora de densidade g satisfaz a condição

$$\int_0^\infty \mathbf{u}^{q/2-1} g(\mathbf{u}) d\mathbf{u} = \frac{\Gamma(q/2)}{\pi^{q/2}}, \quad (2.8)$$

em que $\Gamma(q/2) = \int_0^\infty x^{\frac{q}{2}-1} e^{-x} dx, \frac{q}{2} > 0$.

Teorema 2.1.1. *(Transformações lineares nas distribuições elípticas). Se $\mathbf{Y}$ é um vetor aleatório tal que $\mathbf{Y} = (Y_1, Y_2, \dots, Y_q)^\top \in \mathbb{R}^q$, com distribuição elíptica simétrica com parâmetros dados pelo vetor de médias $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_q]^\top \in \mathbb{R}^q$ e matriz simétrica e positiva definida $\boldsymbol{\Sigma}_{(q \times q)}$, então $\mathbf{Z} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{Y} - \boldsymbol{\mu})$ tem distribuição elíptica simétrica, ou mais precisamente, esférica simétrica, em que $\boldsymbol{\Sigma}^{-\frac{1}{2}}$ é a inversa da matriz raiz quadrada de $\boldsymbol{\Sigma}$, tal que $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{\Sigma}^{\frac{1}{2}}$.*

É possível verificar que as variâncias em $\boldsymbol{\Sigma}_{\mathbf{Z}}$ são iguais a 1 e as covariâncias são nulas, ou seja, $\boldsymbol{\Sigma}_{\mathbf{Z}} = \mathbf{I}_q$ e, ainda, que os contornos da função densidade de mesma probabilidade são esféricos. Portanto $\mathbf{Z}$ tem distribuição elíptica simétrica com parâmetros $\mathbf{0}$ e $\mathbf{I}_q$, ou ainda, a chamada distribuição esférica simétrica.

Alguns exemplos de distribuições esféricas e elípticas são apresentados a seguir. Vale a pena ressaltar que mais detalhes sobre os exemplos apresentados, assim como outros

exemplos, podem ser encontrados em Muirhead (1980; 1982) e em Fang et al. (1990).

Exemplo 2.1.1. Distribuição Normal Multivariada. *Seja $\mathbf{Y} \sim N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, ou seja, $\mathbf{Y}$ tem distribuição Normal Multivariada com função característica dada por:*

$$\phi(\mathbf{t}) = \exp\{i\mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t}\}, \mathbf{t} \in \mathbb{R}^q. \quad (2.9)$$

Se $\boldsymbol{\Sigma} > 0$, então $\mathbf{Y}$ tem densidade

$$f(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-\frac{q}{2}} \exp\left\{-\frac{1}{2} [(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})]\right\}, \mathbf{y} \in \mathbb{R}^q, \quad (2.10)$$

sendo $g(\mathbf{y}) = (2\pi)^{-\frac{q}{2}} \exp\left\{-\frac{1}{2} [(\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})]\right\}$.

Exemplo 2.1.2. Distribuições Normais q -variadas Compostas. *Sejam $Y_1, Y_2, \dots, Y_q$ e $V, V > 0$, variáveis aleatórias tais que, dado $V = v$, $Y_1, Y_2, \dots, Y_q$ são independentes e identicamente distribuídas (iid) com distribuição $N(0, v)$. Se V tem função de distribuição F_V , então a função densidade de probabilidade marginal de $\mathbf{Y} = (Y_1, Y_2, \dots, Y_q)^\top$ é*

$$f(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-1/2} \int_0^\infty (2\pi v)^{-\frac{q}{2}} \exp\left\{-\frac{1}{2v} (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right\} dF_V(v), \mathbf{y} \in \mathbb{R}^q, \quad (2.11)$$

sendo $g(\mathbf{y}) = \int_0^\infty (2\pi v)^{-\frac{q}{2}} \exp\left\{-\frac{1}{2v} (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right\} dF_V(v)$.

A distribuição elíptica construída da forma acima é chamada de mistura de escala de distribuições normais e a classe de distribuições elípticas obtida variando F_V em (2.11) é denominada **classe das distribuições normais compostas**. Assim, um vetor aleatório com distribuição elíptica padrão pode ser representado por

$$\mathbf{Y} \stackrel{\mathcal{D}}{=} \sqrt{V} \mathbf{Z}, \quad (2.12)$$

onde $\mathbf{Z} \sim N_q(\mathbf{0}, \mathbf{I}_q)$ e é independente de V .

Esta representação é útil na geração de valores para $\mathbf{Y}$. Se V assume valores 1 e σ^2 com probabilidade $1 - \varepsilon$ e ε , respectivamente, então $\mathbf{Y}$ tem distribuição normal q -variada ε -contaminada. Se V tem distribuição gama inversa, com parâmetros $(\frac{\nu}{2}, \frac{\nu}{2})$, $\nu > 0$, isto é, se sua densidade é

$$f_V(v|\nu) = \frac{(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} v^{-\frac{(\nu+2)}{2}} \exp\{-\frac{\nu}{2v}\}, v > 0, \quad (2.13)$$

então a distribuição de $\mathbf{Y}$ é t de Student q -variada padrão com ν ($\nu > 0$) graus de liberdade e tem densidade dada por

$$f(\mathbf{y}) = g(\mathbf{y}^\top \mathbf{y}) = \frac{\Gamma(\frac{\nu+q}{2})\nu^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})\pi^{\frac{q}{2}}} \{\nu + \mathbf{y}^\top \mathbf{y}\}^{-\frac{(\nu+q)}{2}}, \mathbf{y} \in \mathbb{R}^q. \quad (2.14)$$

Vale a pena ressaltar que (2.14) se reduz à distribuição normal q -variada padrão quando $(\nu \rightarrow \infty)$, e à distribuição de Cauchy quando $\nu = 1$. Assim, por (2.12), temos que as distribuições t -multivariada padrão, normal q -variada padrão e Cauchy pertencem à classe das distribuições normais compostas e, conseqüentemente, são distribuições esféricas.

Exemplo 2.1.3. Distribuição t q -variada. Seja $\mathbf{Y} \sim t_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ com ν graus de liberdade e parâmetros de posição $\boldsymbol{\mu}$ e dispersão $\boldsymbol{\Sigma}$. Se $\boldsymbol{\Sigma} > 0$, então $\mathbf{Y}$ tem densidade dada por

$$f(\mathbf{y}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu+q}{2})\nu^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})\pi^{\frac{q}{2}}} \{\nu + (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})\}^{-\frac{(\nu+q)}{2}}, \mathbf{y} \in \mathbb{R}^q. \quad (2.15)$$

Comentários sobre esta distribuição são feitos na Seção 2.3.

2.2 Propriedades da Distribuição Elíptica

Como foi provado por diferentes autores, a maioria das propriedades da distribuição normal também são válidas na classe geral das distribuições elípticas. Contudo, algumas propriedades tais como equivalência entre independência e covariância nula de variáveis

aleatórias que têm distribuição conjunta normal, são próprias dessa distribuição e permitem caracterizá-la dentro da classe das distribuições elípticas.

Vamos enumerar algumas propriedades das distribuições esféricas e elípticas para um vetor aleatório $\mathbf{Y} = (Y_1, Y_2, \dots, Y_q)^\top \in \mathbb{R}^q$ com distribuição elíptica, citadas em Ferreira (2007). As propriedades dadas a seguir podem ser derivadas supondo que $\mathbf{Y} \sim El(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h)$ com $\text{posto}(\boldsymbol{\Sigma}) = k \leq q$, excluindo o caso em que $\boldsymbol{\Sigma} = \mathbf{0}$.

P1) $E(\mathbf{Y}) = \boldsymbol{\mu}$;

P2) $\text{Cov}(\mathbf{Y}) = E[(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})^\top] = -2h'(\mathbf{0})\boldsymbol{\Sigma}$, onde $h'(\mathbf{0})$ é a derivada de h no ponto $\mathbf{0}$;

P3) O vetor $\mathbf{X}$ resultante da combinação linear de $\mathbf{Y}$ dada por $\mathbf{X} = \mathbf{A}\mathbf{Y} + \mathbf{b}$, em que $\mathbf{A}_{(q \times q)}$ é não singular e $\mathbf{b}$ é um vetor de dimensão $(q \times 1)$, possui distribuição $El(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top; h)$;

P4) Sejam $\mathbf{Y}$, $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ particionados da seguinte maneira:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad (2.16)$$

sendo os sub-vetores $\mathbf{Y}_1$ e $\boldsymbol{\mu}_1$ de dimensões $(p \times 1)$ e $\mathbf{Y}_2$ e $\boldsymbol{\mu}_2$ de dimensões $((q-p) \times 1)$, para $p < q$. Podemos afirmar que:

i) a distribuição marginal de $\mathbf{Y}_1$ é elíptica, $El_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, h)$;

ii) a distribuição marginal de $\mathbf{Y}_2$ é elíptica, $El_{q-p}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}, h)$;

iii) a distribuição condicional de $\mathbf{Y}_2$ dado $\mathbf{Y}_1 = \mathbf{y}_1$ é elíptica, $El_{q-p}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22}^*; h_2)$, em que $\boldsymbol{\Sigma}_{22}^* = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ e h_2 uma função obtida a partir de h ;

P5) Se $\mathbf{Y} = (Y_1, \dots, Y_q)^\top \sim El_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h)$, então as variáveis aleatórias $Y_1, \dots, Y_q$ são independentes se, e somente se, $\mathbf{Y} \sim N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Mais detalhes sobre as propriedades das distribuições elípticas podem ser encontrados em Arellano-Vale (1994).

2.3 Distribuição t multivariada

Nas últimas décadas tem se observado crescente interesse pela distribuição $t_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$ na literatura estatística, tanto em teoria como em aplicações. O interesse por esta distribuição que tem caudas mais pesadas que as da distribuição normal tem sido bastante enfatizado em estudos de robustez.

Lange et al. (1989) ilustraram o uso da família t na modelagem de dados com diferentes aplicações. Outros estudos interessantes baseados na distribuição t multivariada foram desenvolvidos por Zellner (1976), Sutradhar e Ali (1989), Singh (1991), Cowles e Carlin (1992) e Sutradhar (1993). Como esta distribuição tem sido aceita como alternativa robusta à distribuição normal, há grande desenvolvimento teórico em torno dela.

Muitos dados nas áreas de economia e finanças exibem distribuições com caudas mais pesadas do que as da distribuição normal. Alguns autores como Prucha e Kelejian (1984) discutiram a inadequabilidade do modelo normal e sugeriram o modelo t -correlacionado como uma distribuição alternativa mais apropriada para muitos problemas reais. Estes mesmos autores (Kelejian e Prucha (1985)) provaram posteriormente que distribuições t não correlacionadas têm maior capacidade para capturar comportamentos de variáveis cujas distribuições têm caudas pesadas.

Vale a pena ressaltar que nos últimos anos a distribuição t -multivariada tem desempenhado papel importante na análise Bayesiana multivariada de dados. Atualmente pode ser considerada como a distribuição a priori mais popular, devido ao fato de vários fenômenos físicos, de engenharia ou financeiros serem modelados com esta distribuição.

Mais detalhes sobre esta distribuição podem ser encontrados em Cambanis et al. (1981), Nadarajah e Kotz (2005) e em Kibria e Joarder (2006), sendo que estes últimos apresentam uma revisão das mais importantes teorias desenvolvidas recentemente para análises estatísticas utilizando o modelo t .

Tabela 2.1. Algumas subclasses de distribuições esféricas q -variadas.

Tipo	f.d.p. $f(\mathbf{y})$ ou f.c. $\phi(\mathbf{t})$
Kotz	$f(\mathbf{y}) = c (\mathbf{y}^\top \mathbf{y})^{N-1} \exp\{-r(\mathbf{y}^\top \mathbf{y})^s\},$ $r, s > 0, 2N + q > 2$
Multinormal	$f(\mathbf{y}) = c \exp\left\{-\frac{\mathbf{y}^\top \mathbf{y}}{2}\right\}$
Pearson Tipo VII	$f(\mathbf{y}) = c \left(1 + \frac{\mathbf{y}^\top \mathbf{y}}{s}\right)^{-N}, N > q/2, s > 0$
t Multivariada	$f(\mathbf{y}) = c \left(1 + \frac{\mathbf{y}^\top \mathbf{y}}{m}\right)^{-(q+m)/2}, m > 0 \text{ inteiro}$
Cauchy Multivariada	$f(\mathbf{y}) = c \left(1 + \frac{\mathbf{y}^\top \mathbf{y}}{s}\right)^{-(q+1)/2}, s > 0$
Pearson Tipo II	$f(\mathbf{y}) = c (1 - \mathbf{y}^\top \mathbf{y})^m, m > 0$
Logística	$f(\mathbf{y}) = c \frac{\exp\{-\mathbf{y}^\top \mathbf{y}\}}{[1 + \exp\{-\mathbf{y}^\top \mathbf{y}\}]^2}$
Bessel Multivariada	$f(\mathbf{y}) = c \left(\frac{\ \mathbf{y}\ }{\beta}\right)^a K_a\left(\frac{\ \mathbf{y}\ }{\beta}\right), a > -q/2,$ $\beta > 0$, onde $K_a(\cdot)$ denota função de Bessel do terceiro tipo modificada
Mistura de Escala	$f(\mathbf{y}) = c \int_0^\infty v^{-d/2} \exp\left\{-\frac{\mathbf{y}^\top \mathbf{y}}{2v}\right\} dF_V(v),$ onde $F_V(v)$ é uma f.d.a.
Leis Estáveis	$\phi(\mathbf{t}) = \exp\{r(\mathbf{t}^\top \mathbf{t})^{\alpha/2}\}, 0 < \alpha \leq 2, r < 0$
Multiuniforme	$\phi(\mathbf{t}) = {}_0F_1\left(q/2; -\frac{1}{4}\ \mathbf{t}\ ^2\right)$, onde ${}_0F_1(\cdot; \cdot)$ é uma função hipergeométrica generalizada

c é uma constante de normalização

Capítulo 3

Análise Bayesiana

A análise estatística de séries temporais tem aplicações em diversas áreas, mas na sua maior parte envolve métodos não-Bayesianos. Algumas referências de tratamento Bayesiano de séries temporais são Zellner (1971), Broemeling (1985), West e Harrison (1989) e Pole et al. (1994).

3.1 Inferência Bayesiana

Na teoria Bayesiana admite-se que a descrição da incerteza deve ser feita por meio de probabilidades e por esta razão descreve probabilisticamente toda quantidade desconhecida, sendo os procedimentos de inferência Bayesiana baseados em simples propriedades da teoria de probabilidade.

Considere $\boldsymbol{\theta}$ um vetor de parâmetros a serem estimados. O enfoque Bayesiano incorpora o conhecimento sobre $\boldsymbol{\theta}$ à análise por meio de uma função de probabilidade $P(\boldsymbol{\theta})$ ou função densidade de probabilidade $f(\boldsymbol{\theta})$, denominada distribuição *a priori*, que recebe este nome por se tratar da distribuição de probabilidade de $\boldsymbol{\theta}$ antes que se observem os dados, representados pelo vetor $\mathbf{y}$. Neste texto, tratamos $\boldsymbol{\theta}$ como um vetor de variáveis contínuas, por isso sua distribuição a priori é denotada por $f(\boldsymbol{\theta})$.

A inferência Bayesiana trabalha na presença de observações $\mathbf{y}$ cujos valores inicial-

mente incertos são descritos através de uma distribuição de probabilidades com densidade ou função de probabilidade $f(\mathbf{y}|\boldsymbol{\theta})$, da qual deriva $\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})$, denominada função de verossimilhança, sendo função de $\boldsymbol{\theta}$ dada a amostra $\mathbf{y}$ efetivamente observada.

A inferência sobre $\boldsymbol{\theta}$ é baseada na sua função densidade de probabilidade depois que os dados são observados. Esta função é denominada função densidade de probabilidade a *posteriori*, representada por $f(\boldsymbol{\theta}|\mathbf{y})$ e obtida por meio do Teorema de Bayes dado por:

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{y})}, \quad (3.1)$$

sendo $f(\mathbf{y})$ a função densidade marginal de $\mathbf{Y}$, que pode ser obtida por $f(\mathbf{y}) = \int f(\mathbf{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$. Pode-se reescrever o teorema de Bayes como,

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})f(\boldsymbol{\theta})}{\int \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})f(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (3.2)$$

Uma vez que $\int \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})f(\boldsymbol{\theta})d\boldsymbol{\theta}$ não é função de $\boldsymbol{\theta}$, pode-se escrever

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})f(\boldsymbol{\theta}), \quad (3.3)$$

onde o símbolo $\propto$ indica proporcionalidade.

Pode-se dizer que a inferência Bayesiana implica em processo de aprendizagem sobre algo desconhecido (parâmetros de um modelo), dado o que é conhecido (os dados). O teorema de Bayes pode ser entendido como uma regra de atualização na qual os dados permitem atualizar o conhecimento a priori sobre $\boldsymbol{\theta}$, e o resultado do processo de atualização é a distribuição a posteriori, que combina a informação amostral e a informação não amostral.

Podem ocorrer situações em que exista muito pouca ou nenhuma informação disponível a priori sobre os parâmetros do modelo, ou mesmo situações em que se deseje omitir a opinião do pesquisador. Neste caso, a distribuição a priori considerada é do tipo não informativa.

Portanto, a inferência Bayesiana exige o estabelecimento da verossimilhança e de uma densidade a priori, informativa ou não informativa. Estas são, então, utilizadas para encontrar a densidade a posteriori dos parâmetros, que é base para todos os procedimentos de inferência Bayesiana sobre os parâmetros do modelo. Caso se considerem vários modelos, o pesquisador pode selecioná-los e ou compará-los utilizando as probabilidades a posteriori dos modelos, as razões de chances a posteriori ou o fator de Bayes, no caso de densidades a priori informativas. No caso de priori não informativa, pode-se utilizar o pseudo-fator de Bayes.

3.2 Métodos de Monte Carlo via Cadeias de Markov

A teoria dos métodos de Monte Carlo via Cadeias de Markov (MCMC) é extremamente extensa e, portanto, apresentamos apenas os principais pontos que fundamentam sua metodologia. São apresentados de forma mais detalhada os algoritmos Metropolis-Hastings, Griddy-Gibbs e os procedimentos utilizados para verificar a convergência dos mesmos.

Esta seção foi baseada nos trabalhos de Chib e Greenberg (1995), Bawens et al. (1999), Geweke e Zhou (1996) e Geweke (1997).

Os métodos MCMC permitem obter uma amostra da distribuição a posteriori de interesse, que não poderia ser simulada diretamente. Os valores são gerados de forma iterativa, baseados em cadeias de Markov. A idéia geral dos métodos MCMC é simular um passeio aleatório no espaço do parâmetro $\boldsymbol{\theta}$, que converge para uma distribuição estacionária, que é a distribuição a posteriori $f(\boldsymbol{\theta}|\mathbf{y})$, onde $\mathbf{y}$ é o vetor de observações.

Na análise Bayesiana, a inferência sobre os parâmetros é baseada em sua distribuição a posteriori, como vimos anteriormente. As distribuições marginais de interesse podem ser obtidas da distribuição a posteriori conjunta dos parâmetros, mas, na maioria das vezes, não conseguimos obter analiticamente estas distribuições marginais e o pesquisador tem que recorrer a métodos numéricos. Os mais utilizados são os métodos de Monte Carlo via Cadeias de Markov, cuja teoria básica é apresentada nesta seção. São considerados métodos de simulação através dos quais geram-se amostras aleatórias de uma distribuição

marginal sem que se conheça sua densidade.

Segundo Chib e Greenberg (1995), a abordagem tradicional da teoria das cadeias de Markov inicia-se com um núcleo de transição, que é uma função de distribuição condicional que representa a probabilidade da cadeia de Markov movimentar-se de $\mathbf{x}$ para outro ponto em um conjunto A , representada por $P(\mathbf{x}, A)$, para $\mathbf{x} \in \mathbb{R}^q$ e A um conjunto em $\mathbb{R}^q$. Como se trata de uma função de distribuição, tem-se que $P(\mathbf{x}, \mathbb{R}^q) = 1$, permitindo que a cadeia possa fazer transição de $\mathbf{x}$ para $\mathbf{x}$, ou seja, $P(\mathbf{x}, \mathbf{x})$ não é necessariamente zero.

Os autores citados acima afirmam que uma questão central na simulação via cadeias de Markov é determinar condições sob as quais existe uma distribuição invariante $\pi(\cdot)$ e as iterações de um núcleo de transição convergem para a distribuição invariante na medida em que $n \rightarrow \infty$, sendo n o número de iterações da cadeia de Markov.

Nos métodos MCMC, a densidade invariante $\pi(\cdot)$ é conhecida (a menos de uma constante multiplicadora), e é a densidade objetivo, da qual se deseja amostrar, enquanto o núcleo de transição $P(\mathbf{x}, d\mathbf{y})$ é desconhecido. Para gerar amostras de $\pi(\cdot)$, a metodologia busca encontrar um núcleo de transição $P(\mathbf{x}, d\mathbf{y})$ que convirja para $\pi(\cdot)$ na n -ésima iteração. O processo é iniciado em um ponto arbitrário $\mathbf{x}$, e a partir desse ponto pode-se iterar a cadeia um grande número de vezes. Após muitas iterações, a distribuição das amostras geradas por simulação é, aproximadamente, a distribuição objetivo.

Portanto, o problema é encontrar um núcleo de transição $P(\mathbf{x}, d\mathbf{y})$ adequado. A estratégia para encontrar tal núcleo é descrita a seguir. Suponha que o núcleo de transição, para alguma função $p(\mathbf{x}, \mathbf{y})$ possa ser representado por:

$$P(\mathbf{x}, d\mathbf{y}) = p(\mathbf{x}, \mathbf{y})d\mathbf{y} + r(\mathbf{x})\delta_{\mathbf{x}}(d\mathbf{y}), \quad (3.4)$$

sendo $p(\mathbf{x}, \mathbf{x}) = 0$, $\delta_{\mathbf{x}}$ é a medida de Dirac relativa ao ponto $\mathbf{x}$, ou seja, $\delta_{\mathbf{x}}(d\mathbf{y}) = 1$ se $\mathbf{x} \in d\mathbf{y}$ e 0 caso contrário, e $r(\mathbf{x}) = 1 - \int_{\mathbb{R}^q} p(\mathbf{x}, \mathbf{y})d\mathbf{y}$ é a probabilidade de que a cadeia permaneça em $\mathbf{x}$ na próxima iteração. Conforme Chib e Greenberg (1995), da possibilidade que $r(\mathbf{x})$ seja diferente de 0, torna-se claro que a integral de $p(\mathbf{x}, \mathbf{y})$ em relação a $\mathbf{y}$ não é necessariamente 1.

Pode-se dizer que π é a distribuição invariante de $P(\mathbf{x}, \cdot)$ se a função $p(\mathbf{x}, \mathbf{y})$ em (3.4) satisfizer a condição de reversibilidade dada por:

$$\pi(\mathbf{x})p(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y})p(\mathbf{y}, \mathbf{x}). \quad (3.5)$$

Segundo Silva (2005), pode-se dizer que os lados esquerdo e direito da condição de reversibilidade são, respectivamente, as probabilidades incondicionais da cadeia de Markov mover-se de $\mathbf{x}$ para $\mathbf{y}$ e de $\mathbf{y}$ para $\mathbf{x}$, sendo $\mathbf{x}$ e $\mathbf{y}$ gerados a partir de $\pi(\cdot)$.

Tierney (1994) afirma que a reversibilidade, que deve ser satisfeita pelo núcleo de transição, é uma condição suficiente para que $\pi(\cdot)$ seja estacionária.

3.2.1 O Algoritmo Metropolis-Hastings

O algoritmo Metropolis-Hastings foi proposto por Metropolis et al. (1953) e generalizado por Hastings (1970) e é um poderoso método MCMC que pode ser utilizado para amostrar uma distribuição desconhecida $\pi(\cdot)$.

Esse algoritmo é uma técnica para gerar amostras aleatórias de uma distribuição sem que se conheça sua densidade. É um caso particular dos métodos MCMC, sendo extremamente útil na resolução de problemas multidimensionais e é definido em termos de sub-vetores de $\boldsymbol{\theta}$.

Suponhamos que o vetor de parâmetros a serem estimados $\boldsymbol{\theta}$ seja dividido em k sub-vetores, $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k)^\top$ e que as distribuições condicionais de cada sub-vetor de parâmetros $\boldsymbol{\theta}_i$, dados todos os outros, tenham sido obtidas. Estas distribuições podem ser denotadas por $f_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k, \mathbf{y})$, $f_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1, \boldsymbol{\theta}_3, \dots, \boldsymbol{\theta}_k, \mathbf{y})$, $\dots$, $f_k(\boldsymbol{\theta}_k|\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_{k-1}, \mathbf{y})$.

Dados os valores iniciais $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)})^\top$, selecionamos aleatoriamente

$\boldsymbol{\theta}_1^{(1)}$ de $f_1(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2^{(0)}, \boldsymbol{\theta}_3^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)}, \mathbf{y})$,

$\boldsymbol{\theta}_2^{(1)}$ de $f_2(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_3^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)}, \mathbf{y})$,

$\vdots$

$\boldsymbol{\theta}_k^{(1)}$ de $f_k(\boldsymbol{\theta}_k|\boldsymbol{\theta}_1^{(1)}, \boldsymbol{\theta}_2^{(1)}, \dots, \boldsymbol{\theta}_{k-1}^{(1)}, \mathbf{y})$,

obtendo-se na primeira iteração $\boldsymbol{\theta}^{(1)} = (\boldsymbol{\theta}_1^{(1)}, \dots, \boldsymbol{\theta}_k^{(1)})^\top$. Após um número grande de iterações, digamos B , obtém-se $\boldsymbol{\theta}^{(B)} = (\boldsymbol{\theta}_1^{(B)}, \dots, \boldsymbol{\theta}_k^{(B)})^\top$. As distribuições a posteriori desejadas podem ser aproximadas pelas distribuições marginais dos N valores seguintes $(\boldsymbol{\theta}_1^{(i)}, \dots, \boldsymbol{\theta}_k^{(i)})_{i=B+1}^M$, em que $N = M - B$. Os valores apropriados de M e N podem ser escolhidos monitorando a convergência da cadeia gerada. Os N valores são vistos como N realizações do parâmetro $\boldsymbol{\theta}$.

Cowles e Carlin (1996) apresentam um estudo comparativo entre vários métodos para monitoração da convergência através de métodos gráficos, quantitativos ou qualitativos.

Pode-se pensar no algoritmo Metropolis-Hastings como uma implementação prática do fato que a obtenção das distribuições condicionais completas, mesmo sendo não conhecidas, é suficiente para obtermos as distribuições marginais de interesse. No caso de se ter distribuições condicionais conhecidas, é mais adequado o uso do Amostrador de Gibbs, o qual é também um particular algoritmo MCMC, útil na resolução de problemas multidimensionais e também definido em termos de sub-vetores de $\boldsymbol{\theta}$.

O algoritmo Metropolis-Hastings fornece amostras da densidade a posteriori de interesse, uma vez atingida a convergência, através de uma aproximação de tal densidade e utiliza um mecanismo de rejeição para decidir se a amostra gerada é uma amostra da distribuição a posteriori de interesse.

Suponha que amostras candidatas de uma determinada densidade possam ser geradas. Denota-se por $q(\mathbf{x}, \mathbf{y})$ a densidade geradora de candidatos, sendo que $\int q(\mathbf{x}, \mathbf{y}) d\mathbf{y} = 1$. Se o processo está em determinado ponto $\mathbf{x}$, pode-se gerar uma amostra $\mathbf{y}$ da densidade candidata $q(\mathbf{x}, \mathbf{y})$. Caso $q(\mathbf{x}, \mathbf{y})$ satisfaça a condição de reversibilidade dada em (3.5), para todo $\mathbf{x}$ e $\mathbf{y}$, pode-se terminar a busca pela densidade invariante.

Entretanto, segundo Chib e Greenberg (1995), muito provavelmente isto não ocorrerá. Pode acontecer que a cadeia de Markov se movimente com muita frequência de $\mathbf{x}$ para $\mathbf{y}$ e muito raramente de $\mathbf{y}$ para $\mathbf{x}$, ou seja,

$$\pi(\mathbf{x})p(\mathbf{x}, \mathbf{y}) > \pi(\mathbf{y})p(\mathbf{y}, \mathbf{x}). \quad (3.6)$$

Uma maneira de corrigir esta situação é reduzir o número de movimentos de $\mathbf{x}$ para $\mathbf{y}$ pela introdução de uma probabilidade $\alpha(\mathbf{x}, \mathbf{y}) < 1$ do movimento ocorrer, sendo $\alpha(\mathbf{x}, \mathbf{y})$ a probabilidade de mudança de $\mathbf{x}$ para $\mathbf{y}$. Se o movimento não é realizado, a cadeia de Markov retorna para $\mathbf{x}$, considerando-o um valor da densidade objetivo. O núcleo de transição da cadeia, denotado por $P_{MH}(\mathbf{x}, d\mathbf{y})$, ($\mathbf{y} \neq \mathbf{x}$) será então dado por:

$$P_{MH}(\mathbf{x}, d\mathbf{y}) = q(\mathbf{x}, \mathbf{y})\alpha(\mathbf{x}, \mathbf{y}) \quad (3.7)$$

e $\alpha(\mathbf{x}, \mathbf{y})$ deve ser calculada.

Vale a pena ressaltar que se os movimentos de $\mathbf{y}$ para $\mathbf{x}$ não acontecem com frequência, deve-se definir $\alpha(\mathbf{x}, \mathbf{y})$ tão grande quanto possível, não esquecendo que seu limite superior é 1 por tratar-se de uma probabilidade. Assim a condição de reversibilidade fica dada por:

$$\begin{aligned} \pi(\mathbf{x})q(\mathbf{x}, \mathbf{y})\alpha(\mathbf{x}, \mathbf{y}) &= \pi(\mathbf{y})q(\mathbf{y}, \mathbf{x})\alpha(\mathbf{y}, \mathbf{x}) \\ &= \pi(\mathbf{y})q(\mathbf{y}, \mathbf{x}), \end{aligned} \quad (3.8)$$

pois $\alpha(\mathbf{y}, \mathbf{x})$ assume seu valor máximo que é 1.

Portanto,

$$\alpha(\mathbf{x}, \mathbf{y}) = \frac{\pi(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x}, \mathbf{y})}. \quad (3.9)$$

As probabilidades $\alpha(\mathbf{x}, \mathbf{y})$ e $\alpha(\mathbf{y}, \mathbf{x})$ asseguram que P_{MH} satisfaça a condição de reversibilidade. Pode-se concluir então que, para P_{MH} ser reversível, a probabilidade de movimento da cadeia de Markov deve ser calculada por:

$$\alpha(\mathbf{x}, \mathbf{y}) = \begin{cases} \min \left\{ \frac{\pi(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})q(\mathbf{x}, \mathbf{y})}, 1 \right\}, & \text{se } \pi(\mathbf{x})q(\mathbf{x}, \mathbf{y}) > 0 \\ 1, & \text{se } \pi(\mathbf{x})q(\mathbf{x}, \mathbf{y}) = 0. \end{cases} \quad (3.10)$$

Para completar a definição do núcleo de transição da cadeia de Metropolis-Hastings, deve-se considerar a probabilidade, que pode ser não-nula, da cadeia permanecer no ponto $\mathbf{x}$. Pela definição dada anteriormente, $r(\mathbf{x})$ pode ser reescrita como:

$$r(\mathbf{x}) = 1 - \int_{\mathbb{R}^q} q(\mathbf{x}, \mathbf{y}) \alpha(\mathbf{x}, \mathbf{y}) d\mathbf{y}. \quad (3.11)$$

Conseqüentemente, o núcleo de transição da cadeia de Metropolis-Hastings, denotado por $P_{MH}(\mathbf{x}, d\mathbf{y})$ será dado por:

$$P_{MH}(\mathbf{x}, d\mathbf{y}) = q(\mathbf{x}, \mathbf{y}) \alpha(\mathbf{x}, \mathbf{y}) d\mathbf{y} + [1 - \int_{\mathbb{R}^q} q(\mathbf{x}, \mathbf{y}) \alpha(\mathbf{x}, \mathbf{y}) d\mathbf{y}] \delta_{\mathbf{x}}(d\mathbf{y}), \quad (3.12)$$

que é um caso particular de (3.4). Como $P_{MH}(\mathbf{x}, d\mathbf{y})$ é reversível por construção, Chib e Greenberg (1995) afirmam que o núcleo de transição da cadeia de Markov tem distribuição invariante dada por $\pi(\mathbf{x})$.

O algoritmo Metropolis-Hastings fica determinado após ser definida a distribuição geradora de candidatos $q(\mathbf{x}, \mathbf{y})$. Duas escolhas comuns na literatura são as distribuições normal e t -Student multivariadas. No nosso caso optamos por usar a distribuição t -Student multivariada, pois esta apresenta caudas mais pesadas que as da distribuição normal e é mais compatível com o modelo considerado no Capítulo 4, sendo que os erros do nosso modelo seguem também uma distribuição t -multivariada. No caso de distribuições simétricas, como as duas citadas acima, tem-se que $q(\mathbf{x}, \mathbf{y}) = q(\mathbf{y}, \mathbf{x})$ e a probabilidade de movimento reduz-se a $\frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}$. Portanto, se $\pi(\mathbf{y}) \geq \pi(\mathbf{x})$, a cadeia move-se para $\mathbf{y}$, caso contrário, a cadeia permanece em $\mathbf{x}$ com probabilidade não nula.

Suponha uma densidade geradora de candidatos $q(\mathbf{x}, \mathbf{y})$ simétrica, como exemplificado na Figura 3.1. Verifica-se que do ponto corrente $\mathbf{x}$, uma mudança para a candidata $\mathbf{y}_1$ é feita certamente, enquanto que uma mudança para a candidata $\mathbf{y}_2$ é feita com probabilidade $\frac{\pi(\mathbf{y}_2)}{\pi(\mathbf{x})}$. Em outras palavras, se o valor da função densidade no ponto $\mathbf{y}_1$ é maior que o valor da densidade no ponto $\mathbf{x}$, o salto é crescente e a mudança para $\mathbf{y}_1$ é feita certamente, enquanto que se o valor da densidade em $\mathbf{y}_2$ for menor que o valor da densidade em $\mathbf{x}$, o

salto é decrescente e a mudança poderá ocorrer com probabilidade positiva, mas inferior a 1.

Figura 3.1: Probabilidade de mudança de $\mathbf{x}$ para $\mathbf{y}_i$

Seja $\boldsymbol{\theta}$ um vetor de parâmetros a ser estimado. O algoritmo Metropolis-Hastings pode ser resumido por:

- ✓ Inicialize com um *valor inicial arbitrário* $\boldsymbol{\theta}^{(0)}$ para a cadeia de Markov. Considere $s = 1$;
- ✓ Gere um candidato, $\boldsymbol{\theta}^*$ da densidade geradora de candidatos $q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$;
- ✓ Calcule a probabilidade de movimento,

$$\alpha\left(\boldsymbol{\theta}^{(s-1)}; \boldsymbol{\theta}^*\right) = \min \left\{ \frac{\pi(\boldsymbol{\theta}^*)q\left(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(s-1)}\right)}{\pi(\boldsymbol{\theta}^{(s-1)})q\left(\boldsymbol{\theta}^{(s-1)}; \boldsymbol{\theta}^*\right)}, 1 \right\}$$

- ✓ Gere u de uma distribuição uniforme (0,1);

✓ Faz-se

$$\boldsymbol{\theta}^{(s)} = \begin{cases} \boldsymbol{\theta}^*, & \text{se } u \leq \alpha(\boldsymbol{\theta}^{(s-1)}; \boldsymbol{\theta}^*) \\ \boldsymbol{\theta}^{(s-1)}, & \text{caso contrário} \end{cases}$$

✓ Repete-se o processo acima até que a distribuição estacionária tenha sido obtida.

Mais detalhes sobre o algoritmo Metropolis-Hastings podem ser encontrados em Casella e George (1992), Chib e Greenberg (1995) e Paulino et al. (2003).

3.2.2 O Amostrador Griddy-Gibbs

Quando a densidade condicional a posteriori de algum dos parâmetros não é nenhuma densidade conhecida, mas temos uma expressão analítica possível de ser avaliada em uma rede de pontos, podemos usar o amostrador Griddy-Gibbs para calcular a correspondente função de distribuição usando algum método de integração numérica para gerar variáveis aleatórias dessa função de distribuição (Ritter e Tanner (1992)). O algoritmo Griddy-Gibbs foi proposto por Ritter e Tanner (1992) e é um método MCMC que pode ser utilizado para amostrar uma distribuição desconhecida $\pi(\cdot)$. Este método é conhecido como *método de inversão aproximado* (Paulino et al. (2003)).

Considere $f(\theta)$ uma função densidade de probabilidades univariada. O amostrador Griddy-Gibbs permite obter uma aproximação para a inversa da função de distribuição baseada no cálculo de $f(\theta)$ numa grade de pontos $\theta_1, \dots, \theta_n$.

3.2.2.1 O Algoritmo Griddy-Gibbs

De acordo com Paulino et al. (2003), o algoritmo pode ser resumido da seguinte maneira:

- ✓ Calcula-se $f(\theta)$ em $\theta_1, \dots, \theta_n$, obtendo-se $w_i = f(\theta_i), i = 1, 2, \dots, n$;
- ✓ Usa-se $w_1, \dots, w_n$ para obter uma aproximação da inversa da função de distribuição correspondente a $f(\theta)$, denominada $\hat{F}^{-1}(\cdot)$;
- ✓ Gera-se u de uma distribuição uniforme em $(0,1)$;

✓ Obtém-se $\theta = \hat{F}^{-1}(u)$.

Dois métodos para obter aproximações para a inversa da função de distribuição sugeridos por Ritter e Tanner (1992) são:

- Constante por intervalos: corresponde a uma distribuição discreta para $\theta_1, \dots, \theta_n$ com probabilidades $P(\theta_i) = w_i / \sum_{j=1}^n w_j$;
- Linear por intervalos: corresponde a distribuições uniformes nos intervalos $[a_i, a_{i+1}]$, $i = 1, \dots, n$, escolhidos de modo que $\theta_i \in [a_i, a_{i+1}]$ e densidade $g_i = \frac{w_i}{\sum_{j=1}^n w_j (a_{j+1} - a_j)}$.

Portanto, para aplicar o algoritmo é necessário apenas conhecer $f(\theta)$ a menos de uma constante de proporcionalidade. No caso multivariado, como o aqui tratado, o algoritmo é aplicado para cada parâmetro individualmente. A estimação das distribuições a posteriori é feita em cada iteração, sendo um parâmetro estimado de cada vez, fixando-se os outros parâmetros.

Ritter e Tanner (1992) sugerem usar uma grade de pontos mais apertada à medida que o método se aproxima da convergência. Os pontos θ_i também não devem ser igualmente espaçados, sendo mais eficiente usar para a grade mais pontos no centro da distribuição do que nas caudas. Para este efeito, os autores sugerem a construção de uma grade adaptativa, usando a função de distribuição empírica ($\hat{F}$) construída em uma dada iteração, para obter novos pontos $\theta_1, \dots, \theta_n$ com $\theta_i = \hat{F}^{-1}(\xi_i)$ em valores equidistantes, pré-selecionados, $0 < \xi_1 < \dots < \xi_n < 1$, sendo que a atualização da grade não precisa ser feita em cada iteração.

Este algoritmo foi utilizado por Teran (2007) e segundo o autor, uma das grandes dificuldades que se tem quando se usa o amostrador Griddy-Gibbs a determinao da quantidade de pontos onde as funes condicionais so avaliadas. A determinao errada desse intervalo pode levar a erros nas estimativas, sendo que um pequeno nmero de pontos na grade pode levar a resultados insatisfatrios ou um nmero muito grande pode tornar o algoritmo lento. Para tentar solucionar esse problema, pode-se estabelecer uma janela observando o intervalo de variao da funo densidade condicional na primeira iterao

do amostrador Griddy-Gibbs e estabelecer esse intervalo como janela de variação para cada parâmetro a ser estimado. Este algoritmo apresenta a desvantagem de necessitar de grande esforço computacional sendo, portanto, lento computacionalmente.

Mais detalhes sobre o algoritmo Griddy-Gibbs podem ser encontrados em Tanner (1996).

3.2.3 Procedimentos para Monitoração da Convergência

No processo MCMC, a variável gerada numa etapa depende da variável gerada na etapa anterior, de modo que a série tende a apresentar uma autocorrelação. Para se formar uma amostra para o parâmetro de interesse é necessário que o processo tenha alcançado estacionariedade, que ocorre após B iterações. Então a coleta deve ser feita a partir do B -ésimo valor (*burn-in*) para formar uma amostra de tamanho $N = M - B$. A utilização de amostras com presença de autocorrelação entre os valores leva a um estimador viesado do erro padrão da média.

Segundo Carlin et al. (1998) as estimativas provenientes 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 $\hat{E}(\theta_p|\mathbf{y}) = \bar{\theta}_p = \frac{1}{N} \sum_{i=1}^N \theta_p^{(i)}$, sendo p o índice do parâmetro correspondente. Já o estimador da variância será viesado e não pode simplesmente ser dado por $S_p^2 = \frac{1}{N-1} \sum_{i=1}^N (\theta_p^{(i)} - \bar{\theta}_p)^2$. Este viés pode ser amenizado combinando seqüências amostrais, de maneira a proporcionar uma estimativa do erro padrão da média razoavelmente boa. Mais detalhes sobre este assunto são abordados em Carlin et al. (1998).

Uma outra maneira de solucionar o problema seria calcular as autocorrelações e testar a partir de qual ordem (K) pode-se assumir independência. Com isto, faz-se o *pulo* ou *thin*, que é a retirada de um valor a cada K iterações.

Para a análise da convergência podem ser consideradas técnicas gráficas e uma combinação de critérios univariados e multivariados de avaliação de convergência das seqüências de Markov, encontradas em Gelman e Rubin (1992), Raftery e Lewis (1992), Geweke

(1992), Heidelberger e Welch (1983) e Brooks e Gelman (1997). Mais detalhes sobre os critérios citados acima podem ser vistos em Nogueira (2004), onde o autor sugere uma forma alternativa de aplicação combinada de quatro critérios univariados de convergência das seqüências de Markov e apresenta uma forma alternativa de cálculo do critério multivariado de Brooks e Gelman (1997), além de propor dois novos critérios multivariados de monitoramento.

São consideradas a seguir as definições de alguns critérios de análise de convergência disponíveis no pacote B.O.A (Bayesian Output Analysis Program), desenvolvido por ?, para o aplicativo R, que será utilizado para análise de convergência das cadeias geradas.

Critério de Geweke

O procedimento sugerido por Geweke (1992) baseia-se em um teste de igualdade das médias da primeira e última parte da cadeia de Markov. Se as amostras são retiradas da distribuição estacionária da cadeia, as duas médias tendem a ser iguais, e sob a hipótese nula de convergência, a estatística do teste possui uma distribuição assintótica normal. Trata-se de um teste Z-score padrão, ou seja, a estatística é a diferença entre as duas médias amostrais dividida pelo erro-padrão estimado a partir da densidade espectral em zero, levando-se em consideração, portanto, qualquer autocorrelação presente. A estatística do teste é dada por

$$\frac{\bar{\theta}_1 - \bar{\theta}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}},$$

em que $\bar{\theta}_1$ é a média das n_1 primeiras observações da cadeia, $\bar{\theta}_2$ é a média das n_2 últimas observações da cadeia, S_1^2 e S_2^2 são os estimadores das variâncias de $\hat{\theta}$ nas partes inicial e final da cadeia, respectivamente. Admitindo-se que $\frac{n_1}{N}$ e $\frac{n_2}{N}$ são fixos, é possível mostrar que a distribuição da estatística converge para $N(0, 1)$ quando $N \rightarrow \infty$.

Critério de Raftery e Lewis

O critério sugerido por Raftery e Lewis (1992) é um método que estima quantas iterações são necessárias para que o algoritmo apresente convergência a uma distribuição estacionária. O método fornece alternativas do período de aquecimento ou *burn in* (B),

que se refere ao número de iterações que devem ser descartadas, do número de iterações que devem ser computadas (N) e o valor de K que é a distância mínima de uma iteração a outra (*thin*) para se obter uma sub-amostra aproximadamente independente. Esse critério consiste em apresentar resultados para todos os parâmetros estimados, portanto, deve-se tomar os maiores valores de B , N e K , pois a convergência só se dará quando as cadeias de todas as variáveis convergirem. Esse critério considera também o fator de dependência, que é responsável pelo acréscimo multiplicativo ao número de iterações necessárias para se alcançar a convergência. Segundo Raftery e Lewis (1992), se esse fator for maior que 5, pode-se dizer que a convergência não foi obtida, necessitando reparametrizar o modelo em estudo.

Suponha que se deseje estimar $P(\theta \leq a|\mathbf{y})$ com um erro ϵ e probabilidade γ e que o valor correto a ser estimado seja p . O método de Raftery e Lewis consiste dos seguintes passos:

- (i) para cada iteração s , calcula-se $\theta^{(s)}$ e o valor de $Z_s = I_{(-\infty, a]}(\theta^{(s)})$, onde $I(\cdot)$ é a função indicadora e Z_s é uma sequência de 0's e 1's, que não é uma cadeia de Markov;
- (ii) considera-se então um novo processo construído a partir da seleção de observações Z_s a saltos de K unidades, ou seja,

$$Z_s^{(K)} = Z_s + (s - 1)K.$$

Esse processo também é uma sequência de 0's e 1's e será aproximadamente uma cadeia de Markov se K for suficientemente grande;

- (iii) admita agora que P seja a matriz de transição de $Z_s^{(K)}$,

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}.$$

A distribuição de equilíbrio é dada por $\pi = (\pi_0, \pi_1)$, em que $\pi_0 = P(\theta \leq a|\mathbf{y}) = \frac{\beta}{\alpha + \beta}$ e $\pi_1 = 1 - \pi_0 = \frac{\alpha}{\alpha + \beta}$. Pode-se mostrar que o período de aquecimento é $B = Ks^*$,

onde

$$s^* = \frac{\ln(\frac{\epsilon(\alpha+\beta)}{\max(\alpha,\beta)})}{\ln(1-\alpha-\beta)},$$

ϵ é a distância entre $P(Z_{s^*}^{(K)} = i | Z_0^{(K)} = j)$ e π_i para $i, j = 0, 1$;

- (iv) usando o fato que a média amostral de $Z_s^{(K)}$ tem distribuição assintótica normal, pode-se obter o tamanho da amostra necessário para que se estime $P(\theta \leq a | \mathbf{y})$ com um erro máximo igual a ϵ com probabilidade γ . A distribuição assintótica de $\bar{Z}^{(K)} = \frac{1}{N} \sum_{s=1}^N Z_s^{(K)}$ é normal com média p e variância $\frac{\alpha\beta(2-\alpha-\beta)}{N(\alpha+\beta)^3}$ e daí tem-se que

$$N = \frac{\alpha\beta(2-\alpha-\beta)}{(\alpha+\beta)^3} \left\{ \frac{\Phi^{-1}\left(\frac{1+\gamma}{2}\right)}{\epsilon} \right\}^2,$$

onde Φ é a função distribuição acumulada da distribuição normal padrão. Com isso, $M = NK$.

Para implementar o método é necessário simular o processo e estimar os valores de α e β .

Critério de Heidelberger e Welch

Em outro método para diagnóstico da convergência, Heidelberger e Welch (1983) propõem testar a hipótese nula de estacionariedade a partir de uma seqüência de dados $\{\theta^{(i)}, i = 1, \dots, N\}$ oriundos de um processo não estacionário de densidade espectral $S(w)$ desconhecida, utilizando a teoria Browniana para fazer esta ligação. Mais detalhes podem ser vistos em Brooks e Roberts (1998). Se a hipótese nula for rejeitada para um dado valor, o teste é repetido depois de descartadas as primeiras 10% das iterações. Se a hipótese é novamente rejeitada, outras 10% iniciais são descartadas. Este processo é repetido por no máximo 5 vezes, podendo eliminar até 50% da cadeia. Se a hipótese for novamente rejeitada ou o teste não poder ser realizado, isto indica falha de estacionaridade, o que implica ser necessário um maior número de iterações. Se o teste for satisfatório, o número inicial de iterações descartadas é indicado como o tamanho do *burn in*.

Critério Half-Width

O teste de *half-width* é utilizado para verificar se a média estimada está sendo calculada com acurácia. Calcula-se um intervalo de confiança com coeficiente de confiança de 95% para a média, utilizando o desvio padrão assintótico por meio da densidade espectral. O resultado passará no teste se a razão entre $(\text{limite superior} - \text{limite inferior})/2$ e a estimativa da média for menor que 0,1.

Outros Critérios

Outros métodos, como por exemplo o de Gelman e Rubin (1992) são encontrados na literatura. Os autores mostram que em muitos casos a convergência pode ser facilmente determinada, quando se consideram seqüências múltiplas independentes em paralelo, mas não pode ser diagnosticado usando o resultado da simulação proveniente de qualquer seqüência simples. O método pressupõe que m seqüências tenham sido simuladas em paralelo, cada uma partindo de um diferente ponto inicial. São geradas $2N$ iterações, das quais N são descartadas como período de aquecimento (*burn in*), com isto evitando a influência de valores iniciais. As m seqüências rendem m possíveis inferências. As médias das m seqüências são comparadas por Análise de Variância (ANOVA), em que se compara as variâncias entre e dentro das cadeias. Se essas variâncias são similares é um indicativo de que a convergência foi alcançada ou está próxima.

No Capítulo 4, ao realizarmos a monitoração da convergência do modelo fatorial proposto, serão considerados alguns dos testes apresentados nesta seção.

Capítulo 4

Estimação Bayesiana do Modelo Fatorial Proposto

O processo de estimação Bayesiana do modelo proposto é baseado nas distribuições a posteriori conjuntas para todos os parâmetros desconhecidos. Neste trabalho, consideramos que a distribuição dos erros do modelo pertence à classe das distribuições elípticas.

Para a obtenção das distribuições a posteriori, consideramos o caso particular em que os erros seguem uma distribuição t -multivariada. Vale a pena ressaltar que essa distribuição foi escolhida, pois além de fazer parte da classe de distribuições elípticas, é bastante utilizada em modelos de séries temporais financeiras, muito estudadas nos dias de hoje. O uso dessa distribuição na análise clássica de séries temporais sob modelos ARCH (autoregressive conditional heteroscedasticity) é comum, mas na análise Bayesiana ainda é restrita, sendo a distribuição normal mais usada.

O modelo fatorial dinâmico é apresentado na Seção 4.1 e as distribuições a posteriori são obtidas para cada um dos parâmetros do modelo, considerando o caso particular em que os erros do modelo têm distribuição t -multivariada, na seção seguinte.

4.1 Modelo Fatorial Dinâmico

Considere uma amostra aleatória de tamanho T denotada por $\{\mathbf{y}_t, t = 1, \dots, T\}$, sendo $\mathbf{y}_t$ um vetor de q séries temporais. Para qualquer inteiro positivo $k < q$ o modelo k -fatorial relaciona cada $\mathbf{y}_t$ a um vetor de fatores comuns $\mathbf{f}_t$ via:

$$\mathbf{y}_t = \boldsymbol{\beta} + \mathbf{C}\mathbf{f}_t + \mathbf{e}_t \quad (4.1)$$

$$\mathbf{f}_t = \sum_{i=1}^p \boldsymbol{\rho}_i \mathbf{f}_{t-i} + \mathbf{w}_t, \quad (4.2)$$

em que

$\mathbf{y}_t$ é um vetor de q séries temporais observáveis;

$\mathbf{f}_t$ são vetores ($k \times 1$) de variáveis latentes, ou seja, não observáveis, denominadas fatores comuns, os quais seguem um modelo autoregressivo multivariado (AR) de ordem p , em que as matrizes $\boldsymbol{\rho}_i$ são diagonais com $\boldsymbol{\rho}_i = \text{diag}(\rho_{i1}, \dots, \rho_{ik})$, $i = 1, \dots, p$ e $\{\rho_{1j}, \rho_{2j}, \dots, \rho_{pj}\}$, $j = 1, 2, \dots, k$, satisfazem a condição de estacionariedade;

$\mathbf{C}$ é uma matriz ($q \times k$) de cargas fatoriais, que contém as covariâncias entre os fatores comuns e as variáveis observadas;

$\boldsymbol{\beta}$ é o vetor ($q \times 1$) de médias das observações $\mathbf{y}_t$;

$\mathbf{e}_t$ são vetores independentes ($q \times 1$) com distribuição t multivariada $t_q(\mathbf{0}, \boldsymbol{\Sigma}, \nu_e)$, sendo $\boldsymbol{\Sigma}$ uma matriz diagonal positiva definida, $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_q^2)$ e ν_e os graus de liberdade de $\mathbf{e}_t$;

$\mathbf{w}_t$ são vetores independentes ($k \times 1$) com $\mathbf{w}_t \sim t_k(\mathbf{0}, \boldsymbol{\Lambda}, \nu_w)$, sendo $\boldsymbol{\Lambda}$, uma matriz diagonal positiva definida, $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$ e ν_w os graus de liberdade de $\mathbf{w}_t$;

$\mathbf{e}_t$ e $\mathbf{w}_s$ são mutuamente independentes para todo t e s , $t, s = 1, 2, \dots, T$.

Freqüentemente é assumida normalidade para os vetores de erros aleatórios $\mathbf{e}_t$ e $\mathbf{w}_s$. Neste trabalho, supomos que as distribuições dos vetores de erros pertencem à classe das distribuições elípticas.

Em problemas práticos, especialmente com grandes valores de q , o número de fatores k será freqüentemente menor em relação a q , de maneira que a estrutura de variância

é explicada pelos fatores comuns. É usual impor restrições ao modelo k -fatorial, com a finalidade de se obter um modelo único, livre de problemas de identificação. A solução adotada neste texto foi restringir a matriz de cargas fatoriais $\mathbf{C}$, de maneira que seja uma matriz triangular inferior, de posto completo, ou seja, para o caso em que $q = 4$ e $k = 2$, que a matriz $\mathbf{C}$ tenha a forma

$$\begin{pmatrix} 1 & 0 \\ c_{21} & 1 \\ c_{31} & c_{32} \\ c_{41} & c_{42} \end{pmatrix}.$$

Como no modelo fatorial não são observados nem a matriz de cargas fatoriais e nem os fatores, os mesmos são estimados. A idéia essencial é determinar as distribuições a posteriori de todos os parâmetros desconhecidos, condicionais no fator latente, observações e outros parâmetros e a distribuição do fator latente condicionada aos dados e aos parâmetros do modelo.

Os parâmetros a serem estimados são representados pelo vetor $\boldsymbol{\theta}$, que contém os parâmetros $\boldsymbol{\beta}$, $\mathbf{C}$, $\boldsymbol{\rho}$, $\boldsymbol{\Sigma}$ e $\boldsymbol{\Lambda}$. As distribuições condicionais de $\boldsymbol{\theta}$ e dos fatores $\mathbf{f}$ são representadas por $f(\boldsymbol{\theta}|\mathbf{f}, \mathbf{y})$ e $f(\mathbf{f}|\boldsymbol{\theta}, \mathbf{y})$, onde $\mathbf{f} = (\mathbf{f}_{p+1}, \dots, \mathbf{f}_T)$ e $\mathbf{y} = (\mathbf{y}_{p+1}, \dots, \mathbf{y}_T)$. Os parâmetros do vetor $\boldsymbol{\theta}$ e dos escores fatoriais do modelo de análise fatorial são estimados conjuntamente através de uma aproximação Bayesiana.

Sáfadi e Peña (2007) utilizaram este tipo de procedimento para análise Bayesiana de um modelo fatorial dinâmico para avaliar a associação entre poluição do ar e mortalidade na cidade de São Paulo, de 1994 a 1997, assumindo que os vetores de erros no modelo considerado fossem independentes e suas distribuições fossem normais.

Neste trabalho, a inferência sobre os parâmetros foi feita via métodos MCMC, pelo algoritmo Metropolis-Hastings e pelo método Griddy-Gibbs. Para a implementação dos algoritmos, a distribuição a posteriori condicional completa de cada parâmetro foi obtida. A convergência dos resultados foi analisada por meio de métodos numéricos e pelos critérios de Geweke (1992) e de Heidelberger e Welch (1983).

4.2 Distribuições a Posteriori

A distribuição a posteriori conjunta dos parâmetros desconhecidos e fatores pode ser amostrada usando procedimentos MCMC no conjunto completo das distribuições condicionais.

A verossimilhança conjunta para o modelo fatorial dinâmico proposto é dada por:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{f}) &\propto \prod_{t=p+1}^T [\nu_e + (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t)]^{-\frac{(\nu_e+q)}{2}} \\ &\quad \times [\nu_w + (\mathbf{f}_t - \boldsymbol{\rho}\mathbf{f}_{t-1})^\top \boldsymbol{\Lambda}^{-1}(\mathbf{f}_t - \boldsymbol{\rho}\mathbf{f}_{t-1})]^{-\frac{(\nu_w+k)}{2}}. \end{aligned} \quad (4.3)$$

Nossas análises foram feitas considerando distribuições a priori constantes, independentes para os parâmetros: $\boldsymbol{\beta}$, $\mathbf{C}$, $\boldsymbol{\rho}$, $\boldsymbol{\Sigma}$ e $\boldsymbol{\Lambda}$. Com isso é possível obter as seguintes distribuições a posteriori condicionais completas dos parâmetros do modelo.

✓ distribuição a posteriori do vetor de médias $\boldsymbol{\beta}$:

$$\begin{aligned} f(\boldsymbol{\beta}|\mathbf{C}, \boldsymbol{\rho}, \mathbf{f}, \mathbf{y}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) &\propto f(\boldsymbol{\beta}) \prod_{t=p+1}^T f(\mathbf{y}_t|\boldsymbol{\beta}, \mathbf{C}, \boldsymbol{\rho}, \mathbf{f}_t, \boldsymbol{\Sigma}) \\ &\propto \prod_{t=p+1}^T \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \Gamma\left(\frac{\nu_e+q}{2}\right) \nu_e^{\frac{\nu_e}{2}}}{\Gamma\left(\frac{\nu_e}{2}\right) \pi^{\frac{q}{2}}} \\ &\quad \times \left\{ \nu_e + (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t) \right\}^{-\frac{(\nu_e+q)}{2}} \\ &\propto \prod_{t=p+1}^T \left\{ \nu_e + (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t)^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t) \right\}^{-\frac{(\nu_e+q)}{2}} \end{aligned}$$

✓ Distribuição a posteriori da matriz de cargas fatoriais $\mathbf{C}$:

$$\begin{aligned}
f(\mathbf{C}|\boldsymbol{\beta}, \boldsymbol{\rho}, \mathbf{y}, \mathbf{f}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) &\propto f(\mathbf{C}) \prod_{t=p+1}^T f(\mathbf{y}_t|\boldsymbol{\beta}, \mathbf{C}, \boldsymbol{\rho}, \mathbf{f}_t, \boldsymbol{\Sigma}) \\
&\propto \prod_{t=p+1}^T \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \Gamma(\frac{\nu_e+q}{2}) \nu_e^{\frac{\nu_e}{2}}}{\Gamma(\frac{\nu_e}{2}) \pi^{\frac{q}{2}}} \\
&\times \left\{ \nu_e + (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t) \right\}^{-\frac{(\nu_e+q)}{2}} \\
&\propto \prod_{t=p+1}^T \left\{ \nu_e + (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t) \right\}^{-\frac{(\nu_e+q)}{2}}
\end{aligned}$$

✓ Distribuição a posteriori para a matriz $\boldsymbol{\rho}$:

Considere $\boldsymbol{\rho}_i$ matriz diagonal com, $\boldsymbol{\rho}_i = \text{diag}(\rho_{i1}, \dots, \rho_{ik})$, $i = 1, \dots, p$.

Seja $\mathbf{B}_t$ uma matriz ($k \times kp$) tal que $\mathbf{B}_t = [\text{diag}(\mathbf{f}_{t-1}), \text{diag}(\mathbf{f}_{t-2}), \dots, \text{diag}(\mathbf{f}_{t-p})]$.

Seja $\boldsymbol{\rho} = (\boldsymbol{\rho}_{11}, \dots, \boldsymbol{\rho}_{1p}, \boldsymbol{\rho}_{21}, \dots, \boldsymbol{\rho}_{2p}, \dots, \boldsymbol{\rho}_{k1}, \dots, \boldsymbol{\rho}_{kp})^\top$ um vetor ($kp \times 1$).

Temos que,

$$\sum_{i=1}^p \boldsymbol{\rho}_i \mathbf{f}_{t-i} = \mathbf{B}_t \boldsymbol{\rho}.$$

Então,

$$\begin{aligned}
f(\boldsymbol{\rho}|\mathbf{C}, \boldsymbol{\beta}, \mathbf{f}, \mathbf{y}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) &\propto f(\boldsymbol{\rho}) \prod_{t=p+1}^T f(\mathbf{f}_t|\boldsymbol{\rho}, \mathbf{f}_{t-1}, \dots, \mathbf{f}_{t-p}, \boldsymbol{\Lambda}) \\
&\propto \prod_{t=p+1}^T \frac{|\boldsymbol{\Lambda}|^{-\frac{1}{2}} \Gamma(\frac{\nu_w+k}{2}) \nu_w^{\frac{\nu_w}{2}}}{\Gamma(\frac{\nu_w}{2}) \pi^{\frac{k}{2}}} \\
&\times \left\{ \nu_w + (\mathbf{f}_t - \sum_{i=1}^p \boldsymbol{\rho}_i \mathbf{f}_{t-i})^\top \boldsymbol{\Lambda}^{-1} (\mathbf{f}_t - \sum_{i=1}^p \boldsymbol{\rho}_i \mathbf{f}_{t-i}) \right\}^{-\frac{(\nu_w+k)}{2}} \\
&\propto \prod_{t=p+1}^T \left\{ \nu_w + (\mathbf{f}_t - \mathbf{B}_t \boldsymbol{\rho})^\top \boldsymbol{\Lambda}^{-1} (\mathbf{f}_t - \mathbf{B}_t \boldsymbol{\rho}) \right\}^{-\frac{(\nu_w+k)}{2}}
\end{aligned}$$

✓ Distribuição a posteriori dos fatores $\mathbf{f}$:

$$\begin{aligned}
f(\mathbf{f}|\boldsymbol{\beta}, \mathbf{C}, \boldsymbol{\rho}, \mathbf{y}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) &\propto f(\mathbf{f}) \prod_{t=p+1}^T f(\mathbf{y}_t|\boldsymbol{\beta}, \mathbf{C}, \boldsymbol{\rho}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) \prod_{t=p+1}^T f(\mathbf{f}_t|\boldsymbol{\rho}, \mathbf{f}_{t-1}, \dots, \mathbf{f}_{t-p}) \\
&\propto \prod_{t=p+1}^T \left\{ \nu_e + (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t) \right\}^{-\frac{(\nu_e+q)}{2}} \\
&\times \prod_{t=p+1}^T \left\{ \nu_w + (\mathbf{f}_t - \mathbf{B}_t\boldsymbol{\rho})^\top \boldsymbol{\Lambda}^{-1} (\mathbf{f}_t - \mathbf{B}_t\boldsymbol{\rho}) \right\}^{-\frac{(\nu_w+k)}{2}}
\end{aligned}$$

✓ Distribuição a posteriori de $\boldsymbol{\Sigma}$:

$$\begin{aligned}
f(\boldsymbol{\Sigma}|\mathbf{y}, \mathbf{f}, \boldsymbol{\beta}, \mathbf{C}, \boldsymbol{\rho}, \boldsymbol{\Lambda}) &\propto \prod_{t=p+1}^T f(\boldsymbol{\Sigma}) f(\mathbf{y}_t|\mathbf{f}_t, \boldsymbol{\beta}, \mathbf{C}, \boldsymbol{\rho}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}) \\
&\propto \prod_{t=p+1}^T \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \Gamma(\frac{\nu_e+q}{2}) \nu_e^{\frac{\nu_e}{2}}}{\Gamma(\frac{\nu_e}{2}) \pi^{\frac{q}{2}}} \left\{ \nu_e + (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t)^\top \boldsymbol{\Sigma}^{-1} \right. \\
&\quad \left. (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t) \right\}^{-\frac{(\nu_e+q)}{2}} \\
&\propto \prod_{t=p+1}^T |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left\{ \nu_e + (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t)^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \boldsymbol{\beta} - \mathbf{C}\mathbf{f}_t) \right\}^{-\frac{(\nu_e+q)}{2}}
\end{aligned}$$

✓ Distribuição a posteriori de $\boldsymbol{\Lambda}$:

$$\begin{aligned}
f(\boldsymbol{\Lambda}|\mathbf{y}, \mathbf{f}, \boldsymbol{\beta}, \mathbf{C}, \boldsymbol{\rho}, \boldsymbol{\Sigma}) &\propto \prod_{t=p+1}^T f(\boldsymbol{\Lambda}) f(\mathbf{f}_t|\mathbf{f}_{t-1}, \dots, \mathbf{f}_{t-p}, \boldsymbol{\rho}, \boldsymbol{\Lambda}) \\
&\propto \prod_{t=p+1}^T \frac{|\boldsymbol{\Lambda}|^{-\frac{1}{2}} \Gamma(\frac{\nu_w+k}{2}) \nu_w^{\frac{\nu_w}{2}}}{\Gamma(\frac{\nu_w}{2}) \pi^{\frac{k}{2}}} \left\{ \nu_w + (\mathbf{f}_t - \sum_{i=1}^p \boldsymbol{\rho}_i \mathbf{f}_{t-i})^\top \boldsymbol{\Lambda}^{-1} \right. \\
&\quad \left. (\mathbf{f}_t - \sum_{i=1}^p \boldsymbol{\rho}_i \mathbf{f}_{t-i}) \right\}^{-\frac{(\nu_w+k)}{2}} \\
&\propto \prod_{t=p+1}^T |\boldsymbol{\Lambda}|^{-\frac{1}{2}} \left\{ \nu_w + (\mathbf{f}_t - \mathbf{B}_t\boldsymbol{\rho})^\top \boldsymbol{\Lambda}^{-1} (\mathbf{f}_t - \mathbf{B}_t\boldsymbol{\rho}) \right\}^{-\frac{(\nu_w+k)}{2}}
\end{aligned}$$

4.3 Simulação Estocástica via Cadeias de Markov

Para a geração das cadeias dos parâmetros foi utilizado o algoritmo Metropolis-Hastings, que fica determinado após ter-se definido a distribuição geradora de candidatos, e o amostrador Griddy-Gibbs, que utiliza o método não paramétrico de Kernel, para o cálculo das densidades condicionais aproximadas.

4.3.1 Simulação via Algoritmo Metropolis-Hastings

Para implementar o algoritmo Metropolis-Hastings foi desenvolvido um programa para o *software* R, que está apresentado no Apêndice A. A distribuição candidata considerada foi a distribuição t de Student multivariada, devido ao fato da mesma ter sido eficiente em algumas situações encontradas na literatura.

Embora o programa tenha sido desenvolvido de forma geral, no processo de simulação foram consideradas quatro séries temporais obedecendo o modelo fatorial AR(1), com dois fatores latentes. As matrizes de covariâncias dos erros, Σ e Λ , foram consideradas como sendo a matriz identidade e os graus de liberdade foram fixados em 5 e 3 para ν_e e ν_w , respectivamente. Foram geradas 100.000 iterações de cada série temporal de tamanho 100.

4.3.2 Simulação via Amostrador Griddy-Gibbs

No processo de simulação foram consideradas quatro séries temporais obedecendo o modelo fatorial AR(1), com dois fatores latentes. As matrizes Σ e Λ foram consideradas identidade e os graus de liberdade fixados em 5 e 3 para ν_e e ν_w , respectivamente. Consideramos séries de tamanho 10.000.

O cálculo das densidades condicionais aproximadas foi efetuado considerando-se uma malha de 50 pontos. Foram utilizadas prioris não informativas. O intervalo de variação da densidade foi considerado móvel, isto é, a cada iteração o intervalo variou segundo a média e a variância calculadas na iteração anterior.

Utilizamos o pacote FinMetrics para desenvolver o programa para o S-Plus, que está

apresentado no Apêndice B.

4.4 Monitoração da Convergência

Um ponto importante da aplicação dos métodos MCMC é a verificação empírica da convergência da cadeia para a distribuição de equilíbrio. Em casos de modelos mais complicados, a inferência Bayesiana exige um grande esforço computacional que, segundo Nogueira (2004), pode ser minimizado com a monitoração da convergência, evitando que iterações além das necessárias sejam executadas.

O pacote B.O.A. (Bayesian Output Analysis), disponível no aplicativo *R* foi utilizado para se fazer o monitoramento da convergência das cadeias geradas pelo algoritmo Metropolis-Hastings e pelo Amostrador Griddy-Gibbs.

4.4.1 Convergência do Algoritmo Metropolis-Hastings

Após a inspeção visual da cadeia por meio de técnicas gráficas e de alguns testes disponíveis no B.O.A., foram desconsideradas 40% das observações da cadeia, para fins de eliminar o efeito inicial. Analisou-se em seguida, a função de autocorrelação amostral da cadeia de cada parâmetro, adotando-se intervalos (*thin*) de 150 observações, de forma a se obter uma amostra aproximadamente não correlacionada da cadeia de cada parâmetro.

A Tabela 4.1 mostra os valores fixados, as médias, os desvios padrões, os intervalos de maior densidade a posteriori (HPD - highest probability density) e os testes de Geweke para cada parâmetro do modelo. O valor verdadeiro de cada parâmetro está contido no intervalo de 95% de credibilidade, entretanto, os intervalos são tão grandes que essa informação passa a ser não relevante. Em especial, isso acontece com os intervalos para ρ_1 e ρ_2 , que são quase iguais aos espaços paramétricos correspondentes (intervalo $[-1, 1]$).

Os valores absolutos das médias são bem menores do que os respectivos valores dos desvios padrões. As médias dos β 's são próximas dos valores fixados para esses parâmetros, o que não ocorre com os parâmetros de $\mathbf{C}$ e $\boldsymbol{\rho}$, especialmente para ρ_2 . A pior performance

para ρ_2 também pode ser verificada pelos valores p do teste de Geweke. Esse parâmetro foi o único que apresentou valor p menor do que 0,05.

Outros critérios para avaliação da convergência foram aplicados e todos eles mostraram que a convergência não foi alcançada.

Tabela 4.1: Valores fixados para os parâmetros, médias e desvios padrões a posteriori e teste de Geweke das cadeias geradas via Metropolis-Hastings.

Parâmetro	Valor		Desvio padrão	Intervalo HPD		Teste de Geweke	
	fixado	Média		LI	LS	Z-escore	Valor p
β_1	0	0,0478	1,5799	-2,2236	2,3354	-1,6587	0,0758
β_2	0	-0,0436	1,3657	-2,5786	2,9985	-1,3548	0,1875
β_3	0	0,0758	1,5478	-2,5789	2,4875	-1,2357	0,2254
β_4	0	0,0358	1,6354	-2,5649	3,1542	1,8759	0,2069
C_{21}	0,7	0,1269	0,9457	-1,9876	1,9958	-0,1123	0,7458
C_{31}	0,2	0,0458	1,0657	-1,9987	2,1136	-0,7458	0,4256
C_{41}	0,1	-0,0587	0,9954	-1,8745	1,7845	1,3652	0,1221
C_{32}	0,2	0,0874	1,0614	-2,0045	2,0687	0,5478	0,7745
C_{42}	0,1	0,068	1,0657	-1,8745	2,5416	0,5874	0,7459
ρ_1	0,5	0,3458	0,5347	-0,8975	0,8547	0,9587	0,3874
ρ_2	0,5	0,0657	0,5526	-0,8745	0,9035	2,4587	0,0092

4.4.2 Convergência do Amostrador Griddy-Gibbs

As amostras resultantes do algoritmo Griddy-Gibbs foram obtidas de 1000 iterações, considerando o tamanho da série como 10.000. Foram desconsideradas 10% das observações da cadeia (*burn-in*), para fins de eliminar o efeito inicial. As Figuras 4.1, 4.2 e 4.3 mostram as cadeias originais, sem considerar o *burn-in*, para os parâmetros β , C e ρ , respectivamente. Aparentemente, as cadeias são convergentes para todos os parâmetros.

Tabela 4.2: Valores fixados para os parâmetros, médias e desvios padrões a posteriori das cadeias geradas via Griddy-Gibbs.

Parâmetro	Valor fixado	Média	Desvio padrão
β_1	2,0	1,9966	0,000874
β_2	-5,0	-5,0841	0,000951
β_3	4,0	3,9460	0,000307
β_4	-6,0	-5,9859	0,000214
C_{21}	0,0	0,0013	1,256E-05
C_{31}	0,3	0,2705	1,281E-04
C_{41}	0,5	0,4803	1,265E-04
C_{32}	0,5	0,4880	1,245E-04
C_{42}	0,3	0,2499	1,375E-04
ρ_1	0,5	0,4930	0,000145
ρ_2	0,5	0,5079	0,000143

Figura 4.1: Traço das cadeias geradas para o vetor de médias β .

Os valores fixados para os parâmetros no processo de simulação, bem como as médias a posteriori e os respectivos desvios padrões estão apresentados na Tabela 4.2. Pode-se notar que as estimativas são bem próximas dos verdadeiros valores dos parâmetros.

Figura 4.2: Traço das cadeias geradas para os parâmetros da matriz de cargas fatoriais $\mathbf{C}$.

Figura 4.3: Traço das cadeias geradas para os parâmetros da matriz ρ .

As densidades a posteriori dos parâmetros, obtidas via Griddy-Gibbs estão apresentadas nas Figuras 4.4, 4.5 e 4.6. Verifica-se que para todos os parâmetros considerados, a densidade a posteriori se concentra em torno do valor paramétrico adotado. Observa-se que as densidades são simétricas, indicando que distribuições simétricas poderiam ser usadas como prioris para os parâmetros.

Os resultados do teste de Geweke para as cadeias geradas pelo algoritmo Griddy-Gibbs estão apresentados na Tabela 4.3. Destaca-se que os valores p associados com os parâmetros C_{31} e C_{32} são um pouco menores do que 0,05, mas com base na Tabela 4.2 e nos gráficos da Figura 4.5, pode-se dizer que as cadeias convergiram.

A Tabela 4.4 mostra os resultados do critério de estacionariedade de Heidelberger e Welch e do teste Half-Width. Esses resultados confirmam que as cadeias de todos os

Figura 4.4: Distribuição a posteriori do vetor de médias β .

parâmetros são estacionárias e que a convergência foi alcançada.

Tabela 4.3: Valores do teste de Geweke para os parâmetros estimados pelo algoritmo Griddy-Gibbs.

Parâmetro	Z-Score	Valor p
β_1	0,4796	0,6483
β_2	0,9634	0,8323
β_3	0,6657	0,7472
β_4	0,8198	0,7938
C_{21}	-0,0934	0,4628
C_{31}	-1,6838	0,0461
C_{41}	-0,5653	0,2859
C_{32}	-1,6628	0,0482
C_{42}	-0,0397	0,4842
ρ_1	-0,4837	0,3142
ρ_2	-1,2042	0,1143

Figura 4.5: Distribuição a posteriori dos parâmetros da matriz de cargas fatoriais **C**.

Tabela 4.4: Resultados dos testes de Heidelberg e Welch e Half-Width.

Parâmetro	Heidelberg e Welch	Half-Width
β_1	passed	passed
β_2	passed	passed
β_3	passed	passed
β_4	passed	passed
C_{21}	passed	passed
C_{31}	passed	passed
C_{41}	passed	passed
C_{32}	passed	passed
C_{42}	passed	passed
ρ_1	passed	passed
ρ_2	passed	passed

4.4.3 Simulação via Máxima Verossimilhança

Quando a função de verossimilhança é complexa ou a quantidade de informações de uma série é muito grande, a estimação por máxima verossimilhança torna-se complicada. Por

Figura 4.6: Distribuição a posteriori dos parâmetros da matriz ρ .

este motivo a implementação de novos algoritmos de otimização ou de métodos Bayesianos para se inferir sobre os parâmetros do modelo são usados freqüentemente.

Para as estimações dos parâmetros do modelo fatorial dinâmico proposto foi considerado, a título de comparação, além dos métodos Bayesianos citados anteriormente, o método de *quase máxima-verossimilhança* e o método de *máxima-verossimilhança*, utilizando os otimizadores *BFGS* (Dennis et al. (1981)) e *SimulatedAnnealing – SANN* (Belisle (1992)). No método de quase máxima-verossimilhança foi feita uma aproximação da função de verossimilhança pela distribuição normal com variância equivalente à variância da distribuição *t*-Student. As rotinas foram desenvolvidas no *S – Plus* e encontram-se no Apêndice B.

As tabelas a seguir apresentam os valores paramétricos considerados e suas respectivas estimativas de médias e desvios padrões quando se utiliza os métodos de quase máxima-verossimilhança (Tabela 4.5), máxima-verossimilhança via otimizadores *BFGS* (Tabela 4.6) e *SANN* (Tabela 4.7). Essas análises foram feitas no *S-Plus*. Todos os resultados obtidos por esses três métodos são satisfatórios, as médias são próximas dos valores fixados para os parâmetros e os desvios padrões são pequenos. Em comparação com os resultados obtidos via Griddy-Gibbs, pode-se concluir que os resultados para os parâmetros de β são muito parecidos nos quatro métodos.

Com relação às matrizes $\mathbf{C}$ e $\boldsymbol{\rho}$, os resultados obtidos via Griddy-Gibbs são melhores do que os obtidos pelos outros métodos, especialmente se comparados às médias encontradas por quase máxima-verossimilhança para a matriz $\mathbf{C}$. Estas foram as mais distantes dos verdadeiros valores. Os desvios padrões relacionados à matriz $\mathbf{C}$ são menores quando obtidos pelo Griddy-Gibbs, particularmente aquele associado a C_{41} . Conclui-se que o amostrador Griddy-Gibbs apresentou resultados melhores do que os apresentados pelos outros métodos considerados.

Tabela 4.5: Estimativa dos parâmetros por quase máxima-verossimilhança.

Parâmetro	Valor fixado	Média	Desvio padrão
β_1	2,0	1,9989	0,000744
β_2	-5,0	-5,0715	0,000846
β_3	4,0	3,9527	0,000296
β_4	-6,0	-5,9772	0,000185
C_{21}	0,0	-0,3299	1,588E-05
C_{31}	0,3	0,2481	1,459E-04
C_{41}	0,5	0,7873	1,790E-04
C_{32}	0,5	0,6567	1,368E-04
C_{42}	0,3	0,4822	1,488E-04
ρ_1	0,5	0,5589	0,000136
ρ_2	0,5	0,5628	0,000158

Tabela 4.6: Estimativa dos parâmetros por máxima verossimilhança (*BFGS*).

Parâmetro	Valor fixado	Média	Desvio padrão
β_1	2,0	2,0054	0,000854
β_2	-5,0	-5,0248	0,000754
β_3	4,0	3,9958	0,000358
β_4	-6,0	-6,0014	0,000256
C_{21}	0,0	0,0458	1,536E-05
C_{31}	0,3	0,3588	1,463E-04
C_{41}	0,5	0,5674	1,854-04
C_{32}	0,5	0,5583	1,248E-04
C_{42}	0,3	0,3487	1,326E-04
ρ_1	0,5	0,5253	0,000257
ρ_2	0,5	0,5489	0,000326

Tabela 4.7: Estimativa dos parâmetros por máxima verossimilhança (*SANN*).

Parâmetro	Valor fixado	Média	Desvio padrão
β_1	2,0	2,0158	0,000768
β_2	-5,0	-5,0234	0,000976
β_3	4,0	3,9875	0,000186
β_4	-6,0	-6,0235	0,000193
C_{21}	0,0	0,0261	1,468E-05
C_{31}	0,3	0,3678	1,366E-04
C_{41}	0,5	0,5561	1,646-04
C_{32}	0,5	0,5574	1,287E-04
C_{42}	0,3	0,3281	1,366E-04
ρ_1	0,5	0,5254	0,000125
ρ_2	0,5	0,5499	0,000148

Capítulo 5

Aplicação

A estrutura fatorial em séries temporais foi explorada considerando índices diários - valores de fechamento - de bolsas de valores de quatro diferentes países: Down Jones (Estados Unidos), Ibovespa(Brasil), KLSE-Kuala Lumpur Stock Exchange (Malásia) e RTSI-Russian Trading System Index (Rússia). As séries consideradas têm 242 observações, os dados referem-se ao ano de 2004 e estão disponíveis no Apêndice C.

Optamos por trabalhar com variações de índices de ativos financeiros entre os instantes $t - 1$ e t , dadas pelos retornos.

Denotando $i_t = \log I_t$, sendo $\log$ o logaritmo neperiano e I_t o preço de um ativo (valor do fechamento do índice) no dia t , definimos retorno como:

$$\begin{aligned} r_t &= \log \frac{I_t}{I_{t-1}} \\ &= \log I_t - \log I_{t-1} = i_t - i_{t-1}. \end{aligned} \tag{5.1}$$

Segundo Morettin (2008), na prática é preferível trabalhar com retornos, que são livres de escala e têm propriedades estatísticas mais interessantes, como estacionariedade e ergodicidade.

Os índices diários de bolsas de valores estão apresentados na Figura 5.1. Os retornos destes índices estão apresentados na Figura 5.2 e os gráficos de autocorrelação, autocorrelação parcial, histogramas e QxQ-normal nas Figuras 5.3, 5.4, 5.5 e 5.6, respectivamente.

Figura 5.1: Índices diários de bolsas de valores.

Figura 5.2: Retornos diários dos índices de bolsas de valores.

Embora os gráficos de autocorrelação e autocorrelação parcial indiquem que as séries de retornos sejam ruído branco, ajustamos um modelo AR(1) pois, no modelo considerado, essa estrutura é suposta para os fatores latentes e não diretamente para os retornos.

Os histogramas e gráficos QxQ-normal indicam que a distribuição dos retornos dos índices apresentam certo distanciamento da distribuição normal, principalmente para as séries do Ibovespa, KLSE e RTSI. O retorno do índice Down Jones parece mostrar boa aderência à distribuição normal, entretanto a aplicação do teste de Jarque e Bera (1980) resultou em valores p menores do que 0,0001 para todas as séries de retornos.

Assim sendo, a título de ilustração, ajustamos aos dados dois modelos fatoriais, um sob distribuição normal, como ajustado por Sáfadi (2008), e outro sob distribuição t de Student. Os retornos foram multiplicados por 100.

Figura 5.3: Funções de autocorrelação dos retornos diários de bolsas de valores.

Figura 5.4: Funções de autocorrelação parcial dos retornos diários de bolsas de valores.

A análise Bayesiana do modelo fatorial dinâmico foi feita considerando o modelo fatorial com 2 fatores, os quais seguem um modelo auto-regressivo de ordem 1. As distribuições a priori são as mesmas consideradas no processo de simulação dos dados e a matriz de cargas fatoriais obedece às seguintes restrições: $C_{11} = 1, C_{12} = 0$ e $C_{22} = 1$. Os erros do

Figura 5.5: Histogramas dos retornos diários dos índices de bolsas de valores.

Figura 5.6: Gráficos QxQ-normal dos retornos diários dos índices de bolsas de valores.

modelo seguem uma distribuição normal ou t -Student multivariada, sendo que, as matrizes Σ e Λ foram consideradas como a matriz identidade. Essa restrição não implica em que os dados das séries sejam não correlacionados; a correlação entre eles é traduzida em função da correlação entre os fatores latentes de instantes diferentes.

No modelo t foram fixados os graus de liberdade em 5 e 3, para as distribuições de e_t e w_t , respectivamente. Os números de graus de liberdade poderiam ser considerados desconhecidos e serem também estimados. Segundo Lange et al. (1989), valores baixos fixados têm apresentado boa performance para pequenas amostras. Preferimos fixar valores baixos

para os graus de liberdade, pois valores altos levariam a distribuições aproximadamente normais e o que estamos procurando é uma alternativa a essa distribuição.

Modelo normal

O modelo normal foi ajustado por meio do algoritmo Metroplis-Hastings, com 64.000 iterações, sendo 40% delas descartadas como *burn-in* e tendo saltos de 40 unidades.

As médias e desvios padrões a posteriori dos parâmetros da matriz de cargas fatoriais $\mathbf{C}$ estão apresentadas na Tabela 5.1. O primeiro fator poderia ser interpretado como a própria série de retornos do índice Down Jones, uma vez que a carga fatorial associada a ele é bem maior do que as demais. O segundo fator pode ser visto como uma combinação linear entre os retornos dos outros três índices. O índice KLSE apresenta carga fatorial baixa nos dois fatores, sendo pouco explicado por eles.

Tabela 5.1: Médias (desvios padrões) a posteriori de $\mathbf{C}$, para o modelo normal.

	Fator 1	Fator 2
Down Jones	1	0
Ibovespa	-0,1994 (0,2368)	1
KLSE	-0,0231 (0,0964)	0,3796 (0,0914)
RTSI	-0,1341 (0,2385)	0,7598 (0,2353)

A Tabela 5.2 mostra as médias e desvios padrões a posteriori dos parâmetros de média β e de autocorrelação ρ . Podemos dizer que ρ_1 e ρ_2 são não significantes e isso poderia nos levar a concluir que o modelo mais indicado seria um modelo de análise fatorial para observações independentes. Entretanto, é importante lembrar que a suposição de normalidade das séries parece não ser adequada e isso poderia afetar as estimativas.

A convergência das cadeias foi verificada pelo critério de Gelman e Rubin.

Tabela 5.2: Médias e desvios padrões a posteriori dos parâmetros de β e ρ , para o modelo normal.

Parâmetro	Média	Desvio padrão
β_1	0,200	0,0414
β_2	2,8583	0,3416
β_3	0,5607	0,0750
β_4	3,8198	0,4421
ρ_1	0,0088	0,1031
ρ_2	0,1656	0,1005

Modelo t -Student

O modelo t foi ajustado por meio do algoritmo Griddy-Gibbs, considerando uma malha de 50 pontos.

A Tabela 5.3 contém as médias e os desvios padrões a posteriori das cargas fatoriais. O primeiro fator poderia ser interpretado como um contraste entre os índices Down Jones e RTSI, uma vez que esses índices apresentam cargas altas com sinais trocados. Já o segundo fator poderia ser interpretado como o próprio índice Ibovespa. No modelo t , a variabilidade da série de retornos de KLSE também é pouco explicada.

Tabela 5.3: Médias (desvios padrões) a posteriori de $\mathbf{C}$, para o modelo t .

	Fator 1	Fator 2
Down Jones	1	0
Ibovespa	-0,0117 (0,0244)	1
KLSE	-0,0648 (0,0090)	0,0797 (0,0058)
RTSI	-0,8554 (0,0210)	0,2318 (0,0173)

As médias e desvios padrões a posteriori dos parâmetros β e ρ são apresentados na Tabela 5.4. Pode-se notar que essas estimativas são bem diferentes das encontradas sob o modelo normal. No modelo t , os parâmetros ρ_1 e ρ_2 são significantes, embora sejam baixos.

A convergência das cadeias foi verificada pelo critério de Geweke.

Tabela 5.4: Médias e desvios padrões a posteriori de β e ρ , para o modelo t .

Parâmetro	Média	Desvio padrão
β_1	0,0510	0,0185
β_2	0,1222	0,0253
β_3	0,0725	0,0120
β_4	0,0394	0,0162
ρ_1	0,0507	0,0134
ρ_2	0,0911	0,0137

Análise de resíduos

A análise de resíduos poderia ser feita com base na comparação entre a matriz de covariância amostral dos dados e a matriz estruturada imposta pelo modelo. Optamos por calcular os resíduos pela diferença entre os valores reais das séries e os valores ajustados pelo modelo.

No instante 1, as estimativas dos fatores latentes f_1 e f_2 foram calculadas pelo método de mínimos quadrados, como é habitual fazer em análise fatorial, isto é, $\hat{\mathbf{f}} = (\hat{\mathbf{C}}^\top \hat{\mathbf{C}})^{-1} \hat{\mathbf{C}}^\top (\mathbf{y} - \hat{\beta})$. Para os demais instantes, as estimativas dos fatores foram calculadas pelo modelo, bem como os valores ajustados de $\mathbf{y}$. A fim de eliminar o efeito dessa estimativa no primeiro instante, desprezamos arbitrariamente os primeiros 50 resíduos.

As Figuras 5.7 e 5.8 mostram os gráficos QxQ dos resíduos de cada série de retorno do modelo normal e t , respectivamente. É claro que a observação univariada de cada distribuição não garante que a distribuição multivariada seja a suposta no modelo, mas auxilia na análise descritiva. O que se pode observar é que os resíduos dos retornos do índice Down Jones estão melhor ajustados no modelo normal, mas os resíduos dos índices KLSE e RTSI têm melhor ajuste no modelo t . Outro ponto que favorece o modelo t é que os resíduos desse modelo estão mais próximos do valor zero se comparados aos resíduos do

modelo normal. Isso pode ser visto pelos gráficos *dotplot* dos resíduos nas Figuras 5.9 a 5.12. A variabilidade dos resíduos nos dois modelos é a mesma, mas no modelo t eles estão mais centrados em torno do zero.

Figura 5.7: Gráficos QxQ-normal dos resíduos do modelo normal.

Figura 5.8: Gráficos QxQ- t -Student dos resíduos do modelo t .

Figura 5.9: *Dotplot* dos resíduos do índice Down Jones.

Figura 5.10: *Dotplot* dos resíduos do índice Ibovespa.

Figura 5.11: *Dotplot* dos resíduos do índice KLSE.

Figura 5.12: *Dotplot* dos resíduos do índice RTSI.

Capítulo 6

Conclusões

Aplicações da análise fatorial têm sido úteis em muitas situações práticas, sendo recentemente bastante usadas na área de séries temporais. Como nem sempre podemos assumir normalidade de certo conjunto de dados, optamos por considerar o estudo de um modelo fatorial dinâmico cujos erros têm distribuição na classe de distribuições simétricas mais geral, ou seja, a classe das distribuições elípticas. Uma revisão de suas principais propriedades e utilizações foi considerada no Capítulo 3.

No nosso estudo, consideramos uma estrutura latente em modelos de séries temporais, sob abordagem Bayesiana, em que os fatores seguem um modelo vetorial autoregressivo, supondo que os erros seguem uma distribuição multivariada elíptica e não normal. Assumimos que a um vetor de q séries temporais pode-se ajustar um modelo com k fatores ($k < q$) mais um ruído branco e que a parte latente segue um modelo vetorial autoregressivo.

Para fins de obtenção das distribuições a posteriori dos parâmetros de interesse, consideramos o caso particular em que os erros do modelo seguem uma distribuição t -multivariada. Para a estimação Bayesiana dos parâmetros utilizamos os algoritmos Metropolis-Hastings e Griddy-Gibbs.

Uma avaliação combinada de vários critérios de análise de convergência foi considerada. Ao analisar a convergência do modelo via utilização do algoritmo Metropolis-Hastings, considerando como distribuição candidata a t -Student multivariada e 100.000

iterações no processo de simulação, verificamos que os parâmetros da matriz $\boldsymbol{\rho}$ foram os mais problemáticos. É possível que com um maior número de iterações possa se obter um melhor resultado.

O método de estimação Bayesiana Griddy-Gibbs apresentou excelentes resultados, atingindo convergência após um número pequeno de iterações. Os valores estimados através deste método apresentaram-se bem próximos dos valores estimados quando se utiliza os métodos de máxima-verossimilhança ou quase máxima-verossimilhança.

Como continuidade deste trabalho poderíamos considerar os números de graus de liberdade desconhecidos, estudar critérios para determinar o número de fatores latentes a serem mantidos no modelo e desenvolver o modelo fatorial para outras distribuições da classe das distribuições elípticas.

Apêndice A

Programa R - simulação e análise via Metropolis-Hastings.

```
# PASSO 1: Geração dos valores de Y em função do modelo proposto tese
# Função para gerar uma amostra N da t multivariada: t(N,p,mu,sigma,nu)
tmult = function(N=1,Sigma=diag(4),nu=5)
{
  x=rmvnorm(N,sigma=Sigma)/sqrt(rchisq(N,nu)/nu)
  return(x)
}

##### GERANDO OS DADOS #####
## Início da Função para se gerar os dados de interesse ##

# Observações sobre os atributos da função abaixo
# n: tamanho da série temporal
# nu: graus de liberdade de et(epst na função abaixo)
# eta: graus de liberdade de wt (erro do modelo de ft)
# A função geradados abaixo vai ser usada depois de definidos os valores iniciais
# para os valores entre parenteses

geradados = function(n=200,beta,C,rho,F,nu=5,eta=3) # Default do Programa
{
  # PASSO 1 - Gerar o Vetor y1
  k = nrow(as.matrix(rho))
  q = nrow(as.matrix(C))
  f1 = F[1,]
  eps1 = t(tmult(1,diag(q),nu))
```

```

y1 = beta + C%%f1 + eps1
ft = f1
Y = t(y1)

# PASSO 2 - Gerar yt de 2 a n para t
for (t in 2:n)
{
  ft = F[t,]
  epst = t(tmult(1,diag(q),nu))
  yt = beta + C%%ft + epst
  Y = rbind(Y,t(yt)) }
return(Y) }

### Função para especificar PARÂMETROS da simulação ###

define_par = function(n,q,k,eta,nu)
{
  rho=diag(0.5,k,k)
  beta=matrix(0,q,1)
  C=matrix(c(0.3,0.7,0.2,0.1,0.2,0.6,0.1,0.05),q,k)
  f1 = t(tmult(1,diag(k),eta))
  F = t(f1)
  ft = f1
  for (t in 2:n)
  {
    wt = t(tmult(1,diag(k),eta))
    ft = rho*ft + wt
    F = rbind(F,t(ft))
  }
  return(list(rho=rho,beta=beta,C=C,F=F))
}

### Criando os VALORES INICIAIS para os parâmetros ###

valores_arb = function(q,k,n)
{
  beta = matrix(0,q,1)
  C = matrix(c(0.4,0.8,0.2,0.1,0.3,0.9,0.8,0.2),q,k)
  rho = diag(0.4,k,k)
  F = matrix(0,n,k)
  F[1,]= tmult(1,diag(k),1)
  for (ii in 2:n)
  {

```

```

    F[ii,] = t(rho*%F[ii-1,]) + tmult(1,diag(k),1)
  }
return(list(beta=beta,C=C,rho=rho,F=F))
}

##### FUNÇÃO CANDIDATA #####

# Utilizada para gerar novos betas (betastar), novos Cs (C Cstar), etc.
# Atualmente está usando como candidata uma t esférica com dfb graus de liberdade
candidata_par = function(q,dfb)
{
  return(t(tmult(1,diag(q),dfb)))
}

##### Função para calcular a DENSIDADE POSTERIORI PARA BETA #####

post_beta = function(Y,beta,C,F,n,nu,q,k)
{
  pbeta = 0 # pbeta: posteriori para beta
  for (ii in 2:n)
  {
    yt = as.matrix(Y[ii,],q,1)
    f = as.matrix(F[ii,],k,1)
    aux = nu + t(beta)*%beta-t(beta)*%(yt-C*%f)-(t(yt)-t(f)*%t(C))*%beta
    if (aux<=0) aux = 0 else aux = -0.5*(nu+q)*log(aux)
    pbeta = pbeta+aux
  }
  return(exp(pbeta))
}

##### Função para calcular DENSIDADE DA DISTRIBUIÇÃO CANDIDATA #####

#(Considerando a t multivariada esférica como candidata)#

qdensidade = function(x,nu)
{
  pp = length(x)
  # pp = dimensão do parâmetro
  pq = (1+t(x)*%x/nu)^(-(nu+pp)/2) #fdp da t esférica
}

## Função para calcular a probabilidade de transição de BETA para BETA STAR ##

```

```

alphaxy_beta = function(Y,beta,betastar,C,F,n,nu,q,k,nu_beta)
{
  pibetastar = post_beta(Y,betastar,C,F,n,nu,q,k)
  pibeta = post_beta(Y,beta,C,F,n,nu,q,k)
  qbetastar = qdensidade(betastar,nu_beta)
  qbeta = qdensidade(beta,nu_beta)
  alphaxy = min(1,pibetastar*qbeta/(pibeta*qbetastar))
  if (alphaxy=="NaN") alphaxy=0.25
  return (alphaxy)
}

# Função para calcular a DENSIDADE A POSTERIORI PARA C #

post_C = function(Y,C,beta,F,n,nu,q,k)
{
  pC = 0
  for (ii in 2:n)
  {
    yt = as.matrix(Y[ii,],q,1)
    f = as.matrix(F[ii,],k,1)
    aux = yt-beta-C%%f
    aux = nu + t(aux)%%aux
    if (aux<=0) aux = 0 else aux = -0.5*(nu+q)*log(aux)
    pC = pC + aux
  }
  return(exp(pC))
}

## Função para calcular a probabilidade de transição de C para CSTAR ##
alphaxy_C = function(Y,C,Cstar,beta,F,n,nu,q,k,nu_C)
{
  piCstar = post_C(Y,Cstar,beta,F,n,nu,q,k)
  piC = post_C(Y,C,beta,F,n,nu,q,k)
  qCstar = qdensidade(as.vector(Cstar),nu_C)
  qC = qdensidade(as.vector(C),nu_C)
  alphaxy = min(1,piCstar*qC/(piC*qCstar))
  if (alphaxy=="NaN") alphaxy=0.25
  return(alphaxy)
}

## Função para gerar candidata F (MATRIZ n X k) ##
candidata_F = function(k,nu_f,rho)
{

```

```

Fst = matrix(0,n,k)
Fst[1,] = tmult(1,diag(k),1) # matriz n x k de fatores iniciais
for (ii in 2:n)
{
  Fst[ii,] = t(rho%%Fst[ii-1,]) + tmult(1,diag(k),nu_f)
}
return(Fst)
}

```

Função para calcular a DENSIDADE A POSTERIORI PARA F

```

post_F = function(Y,F,C,beta,rho,n,nu,nu_f,q,k) {
  pf = 0
  for (ii in 2:n)
  {
    yt = as.matrix(Y[ii,],q,1)
    ft = as.matrix(F[ii,],k,1)
    ft_1 = as.matrix(F[ii-1,],k,1)
    aux = yt-beta-C%%ft
    aux = nu + t(aux)%%aux
    aux1 = ft - rho%%ft_1
    aux1 = nu_f + t(aux1)%%aux1
    if (aux<=0) aux = 0 else aux = -0.5*(nu+q)*log(aux)
    if (aux1<=0) aux1 = 0 else aux1 = -0.5*(nu_f+k)*log(aux1)
    pf = pf + aux + aux1
  }
  return(exp(pf))
}

```

Função para calcular a PROBABILIDADE DE TRANSIÇÃO DE F PARA FSTAR

```

alphaxy_F = function(Y,F,Fstar,C,beta,rho,n,nu,q,k,nu_f)
{
  piFstar = post_F(Y,Fstar,C,beta,rho,n,nu,nu_f,q,k)
  piF = post_F(Y,F,C,beta,rho,n,nu,nu_f,q,k)
  qFstar = qdensidade(as.vector(t(Fstar)),nu_f)
  qF = qdensidade(as.vector(t(F)),nu_f)
  alphxy = min(1,piFstar*qF/(piF*qFstar))
  if (alphxy=="NaN") alphxy=0.25
  return(alphxy)
}

```

Função para gerar a CANDIDATA rho (MATRIZ k X k)


```

candidata_rho = function(k)
{
  return(diag(runif(k)*2-1,k,k))
}

# Função para calcular a DENSIDADE POSTERIORI PARA rho #

post_rho = function(F,C,beta,rho,n,eta,k)
{
  prho = 0
  for (ii in 2:n)
  {
    ft = as.matrix(F[ii,],k,1)
    ft_1 = as.matrix(F[ii-1,],k,1)
    aux = ft - rho%%ft_1
    aux = eta + t(aux)%%aux
    if (aux<=0) aux = 0 else aux = -0.5*(eta+k)*log(aux)
    prho = prho + aux
  }
  return(exp(prho))
}

# Função para calcular a PROBABILIDADE DE TRANSIÇÃO DE rho para rhoSTAR #

alphaxy_rho = function(Y,rho,rhostar,beta,C,F,n,nu,q,k,nu_f)
{
  pirhostar = post_rho(F,C,beta,rhostar,n,eta,k)
  pirho = post_rho(F,C,beta,rho,n,eta,k)
  qrhostar = 1
  qrho = 1
  alphxy = min(1,pirhostar*qrho/(pirho*qrhostar))
  if (alphxy=="NaN") alphxy=0.25
  return(alphxy)
}

#### ALGORITMO DE METROPOLIS-HASTINGS ####

MH = function(Y,k,NS)
{
  q = ncol(Y)
  n = nrow(Y)
  aux = valores_arb(q,k,n)
  beta = aux$beta

```

```

C = aux$C
rho = aux$rho
F = aux$F
nu = n - 1
nu_beta = 3
nu_C = nu
nu_f = eta
Cadeia.beta = beta
Cadeia.C = as.vector(C)
Cadeia.F = as.vector(F)
Cadeia.rho = diag(rho)

### APLICAÇÃO DO ALGORITMO DE METROPOLIS-HASTINGS ###

for (it in 1:NS)
{
# Atualização de valores de BETA: utiliza como candidata uma t_q(0,I,df=3) #
  betastar = candidata_par(q,nu_beta)
  axybeta = alphaxy_beta(Y,beta,betastar,C,F,n,nu,q,k,nu_beta)
  ubeta = runif(1)
  if (ubeta<=axybeta) beta = betastar

# Atualização de valores de C: utiliza como candidata uma t_(kq)(0,I,df=nu) #
  Cstar = matrix(candidata_par(q*k,nu_C),q,k)
  axyC = alphaxy_C(Y,C,Cstar,beta,F,n,nu,q,k,nu_C)
  uC = runif(1)    if (uC <= axyC) C = Cstar

# Atualização de valores de F: utiliza como candidata uma t_(k)(0,I,df=eta) #
  Fstar = candidata_F(k,nu_f,rho)
  axyF = alphaxy_F(Y,F,Fstar,C,beta,rho,n,nu,q,k,nu_f)
  uF = runif(1)
  if (uF<=axyF) F = Fstar

# Atualização de valores de rho: utiliza como candidata uma uni_k(-1,1) #
  rhostar = candidata_rho(k)
  axyrho = alphaxy_rho(Y,rho,rhostar,beta,C,F,n,nu,q,k,nu_f)
  urho = runif(1)
  if (urho<=axyrho) rho = rhostar

### Atualização das CADEIAS ###
Cadeia.beta = cbind(Cadeia.beta,beta)
Cadeia.C = cbind(Cadeia.C,as.vector(C))
Cadeia.F = cbind(Cadeia.F,as.vector(F))

```

```

Cadeia.rho = cbind(Cadeia.rho,diag(rho))
}
return(list(beta=t(Cadeia.beta),C=t(Cadeia.C),F=t(Cadeia.F),rho=t(Cadeia.rho)))
}

```

```

# OBTENÇÃO DOS RESULTADOS DE INTERESSE #

```

```

library(mvtnorm)
eta=3
nu=5
n=100 #tamanho da série é 00
q=4
k=2
NS=10000 #numero de simulações
par=define_par(n,q,k,eta,nu)
Y=gerados(n,par$beta,par$C,par$rho,par$F,eta,nu)
result=MH(Y,k,NS)

```

```

## SOLTAR A CADEIA (MH) PARA CADA PARÂMETRO SIMULADO ##

```

```

result$beta
result$C
result$F
result$rho

```

Apêndice B

Programa S-Plus - simulação e análise via Griddy-Gibbs.

```
# Simulacao de uma série seguindo o seguinte modelo:
#  $f(t) = \text{Rho} \cdot f(t-1) + \text{Eps-f}(t)$ 
#  $y(t) = \text{Beta} + \text{Cf}(t) + \text{Eps-y}(t)$ 
# Onde  $f(t)$  é um vetor  $2 \times 1$ ,  $y(t)$  é um vetor  $4 \times 1$ ,  $t$  variando de 1 até  $n$ ,
#  $\text{Rho}$  é a matrix  $2 \times 2$ 
# |0.5 0.0|
# |0.0 0.5|
#  $\text{Beta}$  é um vetor  $4 \times 1$  com valores (2,-5,4,-6)
#  $\text{C}$  é uma matrix  $4 \times 2$ 
# |1.0 0.0|
# |0.0 1.0|
# |0.3 0.5|
# |0.5 0.3|
#  $\text{Eps-f}$  e  $\text{Eps-y}$  são resíduos de uma distribuição Normal com média zero e matrix de variâncias  $3 \times \text{Identidade}$ . (Isto para
# ser equivalente a uma  $t$  com 3 graus de liberdade.

module(finmetrics)
mPhi <- matrix(c(0.5, 0, 0, 0.5, c(1,0,0,1,0.3,0.5,0.5,0.3)), nrow=6, byrow=T)
mOmega <- 3diag(6)
mDelta <- c(0,0,2,-5,4,-6)
mSigma <- matrix(c(0,0,0,0,0,0), ncol=2)
ssf <- list(mPhi=mPhi, mOmega=mOmega, mDelta=mDelta, mSigma=mSigma)
ssf <- CheckSsf(ssf)
rm(mPhi, mOmega, mDelta, mSigma)
```

```

simulacoes <- SsfSim(ssf, n=10000)

# Estimação por Máxima Verossimilhança do processo simulado acima.
ssf <- function(parm)
{
  mOmega <- 3diag(6)
  mPhi <- matrix(c(parm[1],0,0,parm[2],1,0,parm[3],1,parm[4:7]), ncol=2, byrow=T)
  mDelta <- c(0,0,parm[8:11])
  mSigma <- matrix(c(0,0,0,0,0,0), ncol=2)
  ssf <- list(mPhi=mPhi, mOmega=mOmega, mDelta=mDelta, mSigma=mSigma)
  ssf <- CheckSsf(ssf)
}
y <- simulacoes[,3:6]
parm0 <- rnorm(11)
likelihood <- SsfFit(parm0, y, "ssf")
Estimações <- likelihood$par
# > Estimações (Considerando valores de C: c(1,0,0,1,0.3,0.5,0.5,0.3))
# [1] 0.596254396 0.558190827 -0.124354757 0.356859114 0.753921817 0.709402642
# [7] 0.479337993 -0.001675221 -0.001178974 -0.011695067 -0.015320497
# > Estimações (Considerando valores de betas: c(0,0,2,-5,4,-6))
# [1] 0.59203430 0.58140524 -0.04518784 0.41866669 0.72834082 0.73798107 0.43044833
# [8] 1.98300553 -4.98936987 3.98528554 -5.99617807

# Simulação de uma série seguindo o modelo acima mas com os resíduos sendo  $t$ 
tmult = function(N=1,Sigma=diag(4),nu=3) #Default do programa
{
  x=rmvnorm(N,cov=Sigma^2)/sqrt(rchisq(N,nu)/nu)
  return(x)
}

gerados = function(n=10000,beta,C,rho,Ft,nu=3,eta=3) # Default do Programa
{
  # PASSO 1 - Gerar o Vetor y1
  k = nrow(as.matrix(rho))
  q = nrow(as.matrix(C))
  f1 = Ft[1,]
  eps1 = t(tmult(1,diag(q),nu))
  y1 = beta + C%%f1 + eps1
  ft = f1
  Y = t(y1)
  # PASSO 2 - Gerar yt de 2 a n para t
  for (t in 2:n)

```

```

{
  epst = t(tmult(1,diag(q),nu))
  wt = t(tmult(1,diag(k),eta))
  ft = rho%%ft + wt
  Y = rbind(Y,t(yt))
  Ft = rbind(Ft,t(ft))
}
return(Y)
}

definepar = function(n=10000,q,k,eta=3,nu=3)
{
  rho=diag(0.5,k,k)
  beta=matrix(c(2,-5,4,-6),q,1)
  C=matrix(c(1,0.0,0.3,0.5,0.0,1,0.5,0.3),q,k)
  f1 = c(0,0)
  Ft = t(f1)
  ft = f1
  for (t in 2:n)
  {
    wt = t(tmult(1,diag(k),eta))
    ft = rho%%ft + wt
    Ft = rbind(Ft,t(ft))
  }
  return(list(rho=rho,beta=beta,C=C,F=Ft))
}

eta=3
nu=3
n=10000
q=4
k=2
par=definepar(n,q,k,eta,nu)
yt=geradados(n,par$beta,par$C,par$rho,par$F,eta,nu)

# Estimação por Máxima Verossimilhança do processo simulado acima usando normal
parm0 <- rnorm(11)
likelihood <- SsfFit(parm0, yt, "ssf")
Estimações <- likelihood$par
# > Estimações
# [1] 0.48644819 0.50246531 -0.09220364 0.23649234 0.49578197 0.46275100
# [7] 0.26568146 1.99931183 -5.07183685 3.95248605 -5.97706571

fLogLik <- function(parm)

```

```

{
value <- SsfLoglike(yt, ssf(parm))
return(-value$loglike)
}
fLogLik(estimacoes)

#####

# Estimação por Máxima Verossimilhança do processo simulado acima usando  $t$ 
SS.LogLik <- function(y, f, Beta, C, Rho, nu, eta)
{
q <- length(y[1,])
k <- length(f[1,])
n <- length(y[,1])
LogLik <- 0
for (i in 2:n)
{
vyt <- y[i,]
vft <- f[i,]
vft1 <- f[(i-1),]
passo1t <- vyt - Beta - C%%vft
passo2t <- vft - Rho%%vft1
passo3t <- nu + sum(passo1t^2)
passo4t <- eta + sum(passo2t^2)
passo5t <- -log(passo3t)
passo6t <- -log(passo4t)
passofinalt <- passo5t + passo6t
LogLik <- LogLik + passofinalt
}
return(LogLik)
}
fLogLik <- function(parm)
{
ss <- ssf(parm)
teste <- SsfMomentEst(yt, ss, task="STFIL")
f <- teste$state.moment
Beta <- ss$mDelta[3:6]
C <- ss$mPhi[3:6,]
Rho <- ss$mPhi[1:2,]
fLogLik <- SS.LogLik(yt, f, Beta, C, Rho, 3, 3)
return(-fLogLik)
}
parm0 <- estimacoes

```

```

#parm0 <- otimo.nlmin$x
#parm0 <- otimo.sann$par

otimo.nlmin <- nlmin(f=fLogLik, x=parm0 , max.iter=100)
otimo.sann <- optim(fn=fLogLik, par=parm0, method="SANN")

yt <- bolsamundo100
#> otimo.nlmin
#$x:
# [1] 0.031806799 0.091476242 0.001091947 -0.071857338 0.081490410
-0.877897902 0.236121137 0.020935015 0.119682397 0.055437584 0.044576792
# > otimo.sann
#$par:
# [1] 0.031790008 0.091586353 0.001314815 -0.071832948 0.081529709
-0.877773548 0.236393928 0.020918031 0.119753583 0.055459933 0.044607751

# Estimação usando Griddy-Gibbs
Dens2Dist <- function(x, y)
{
  Dist <- (x - tslag(x, pad=0)) (y + tslag(y, pad=0))/2
  Dist[1] <- 0
  cumsum(Dist)
}
# número de pontos na malha. grids.
n.grids <- 50
# Definição das malhas de pontos.
rho1.malha <- (0:n.grids)/n.grids
rho2.malha <- (0:n.grids)/n.grids
C1.malha <- (0:n.grids)/2/n.grids
C2.malha <- (0:n.grids)/n.grids
C3.malha <- (0:n.grids)/n.grids
C4.malha <- (0:n.grids)/n.grids
C5.malha <- (0:n.grids)/n.grids
Beta1.malha <- 4(0:n.grids)/n.grids
Beta2.malha <- -10(0:n.grids)/n.grids
Beta3.malha <- 8(0:n.grids)/n.grids
Beta4.malha <- -12(0:n.grids)/n.grids
# Valores iniciais.
rho1.ini <- 0.5
rho2.ini <- 0.5
C1.ini <- 0.5
C2.ini <- 0.5
C3.ini <- 0.5

```



```

C4.ini <- 0.5
C5.ini <- 0.5
Beta1.ini <- 5
Beta2.ini <- 5
Beta3.ini <- 5
Beta4.ini <- 5
# Distribuições a priori.
f.priori.rho1 <- function(par.rho1) 1
f.priori.rho2 <- function(par.rho2) 1
f.priori.C1 <- function(par.C1) 1
f.priori.C2 <- function(par.C2) 1
f.priori.C3 <- function(par.C3) 1
f.priori.C4 <- function(par.C4) 1
f.priori.C5 <- function(par.C5) 1
f.priori.Beta1 <- function(par.Beta1) 1
f.priori.Beta2 <- function(par.Beta2) 1
f.priori.Beta3 <- function(par.Beta3) 1
f.priori.Beta4 <- function(par.Beta4) 1
# Funcao de verossimilhanca
f <- fLogLik
# Calculo das Posteriores.
f.cond.rho1 <- function(par.rho1)
{
-f(c(par.rho1,rho2.ini,C1.ini,C2.ini,C3.ini,C4.ini,C5.ini,Beta1.ini,Beta2.ini,Beta3.ini,Beta4.ini))
+ log(f.priori.rho1(par.rho1))
}
f.cond.rho2 <- function(par.rho2)
{
-f(c(rho1.ini,par.rho2,C1.ini,C2.ini,C3.ini,C4.ini,C5.ini,Beta1.ini,Beta2.ini,Beta3.ini,Beta4.ini))
+ log(f.priori.rho2(par.rho2))
}
f.cond.C1 <- function(par.C1)
{
-f(c(rho1.ini,rho2.ini,par.C1,C2.ini,C3.ini,C4.ini,C5.ini,Beta1.ini,Beta2.ini,Beta3.ini,Beta4.ini))
+ log(f.priori.C1(par.C1))
}
f.cond.C2 <- function(par.C2)
{
-f(c(rho1.ini,rho2.ini,C1.ini,par.C2,C3.ini,C4.ini,C5.ini,Beta1.ini,Beta2.ini,Beta3.ini,Beta4.ini))
+ log(f.priori.C2(par.C2))
}
f.cond.C3 <- function(par.C3)
{

```

```

-f(c(rho1.ini,rho2.ini,C1.ini,C2.ini,par.C3,C4.ini,C5.ini,Beta1.ini,Beta2.ini,Beta3.ini,Beta4.ini))
+ log(f.priori.C3(par.C3))
}
f.cond.C4 <- function(par.C4)
{
-f(c(rho1.ini,rho2.ini,C1.ini,C2.ini,C3.ini,par.C4,C5.ini,Beta1.ini,Beta2.ini,Beta3.ini,Beta4.ini))
+ log(f.priori.C4(par.C4))
}
f.cond.C5 <- function(par.C5)
{
-f(c(rho1.ini,rho2.ini,C1.ini,C2.ini,C3.ini,C4.ini,par.C5,Beta1.ini,Beta2.ini,Beta3.ini,Beta4.ini))
+ log(f.priori.C5(par.C5))
}
f.cond.Beta1 <- function(par.Beta1)
{
-f(c(rho1.ini,rho2.ini,C1.ini,C2.ini,C3.ini,C4.ini,C5.ini,par.Beta1,Beta2.ini,Beta3.ini,Beta4.ini))
+ log(f.priori.Beta1(par.Beta1))
}
f.cond.Beta2 <- function(par.Beta2)
{
-f(c(rho1.ini,rho2.ini,C1.ini,C2.ini,C3.ini,C4.ini,C5.ini,Beta1.ini,par.Beta2,Beta3.ini,Beta4.ini))
+ log(f.priori.Beta2(par.Beta2))
}
f.cond.Beta3 <- function(par.Beta3)
{
-f(c(rho1.ini,rho2.ini,C1.ini,C2.ini,C3.ini,C4.ini,C5.ini,Beta1.ini,Beta2.ini,par.Beta3,Beta4.ini))
+ log(f.priori.Beta3(par.Beta3))
}
f.cond.Beta4 <- function(par.Beta4)
{
-f(c(rho1.ini,rho2.ini,C1.ini,C2.ini,C3.ini,C4.ini,C5.ini,Beta1.ini,Beta2.ini,Beta3.ini,par.Beta4))
+ log(f.priori.Beta4(par.Beta4))
}

# Definindo vetores para pontos das Densidades e Distribuições dos parâmetros.
Dens.rho1 <- rep(NA, length(rho1.malha))
Dens.rho2 <- rep(NA, length(rho2.malha))
Dens.C1 <- rep(NA, length(C1.malha))
Dens.C2 <- rep(NA, length(C2.malha))
Dens.C3 <- rep(NA, length(C3.malha))
Dens.C4 <- rep(NA, length(C4.malha))
Dens.C5 <- rep(NA, length(C5.malha))
Dens.Beta1 <- rep(NA, length(Beta1.malha))

```

```

Dens.Beta2 <- rep(NA, length(Beta2.malha))
Dens.Beta3 <- rep(NA, length(Beta3.malha))
Dens.Beta4 <- rep(NA, length(Beta4.malha))
num.iter <- 3000
Resultados.Esperados <- matrix(NA, ncol=33, nrow=num.iter)
# Processo Iterativo.
for (j in 1:num.iter)
{
# Calculando a Densidade e a Distribuição Condicional para rho1.
for (i in 1:length(rho1.malha)) {
Dens.rho1[i] <- f.cond.rho1(rho1.malha[i])
}
Dens.rho1 <- Dens.rho1 - max(Dens.rho1)
Dens.rho1 <- exp(Dens.rho1)
Dist.rho1 <- Dens2Dist(rho1.malha, Dens.rho1)
Dens.rho1 <- Dens.rho1 / Dist.rho1[length(rho1.malha)]
Dist.rho1 <- Dist.rho1 / Dist.rho1[length(rho1.malha)]
Esp.cond.rho1 <- Dens2Dist(rho1.malha, Dens.rho1 rho1.malha)[length(rho1.malha)]
Var.cond.rho1 <- Dens2Dist(rho1.malha, Dens.rho1 rho1.malha ^ 2)[length(rho1.malha)]
- Esp.cond.rho1^ 2
Aleatório <- runif(1)
rho1.ini <- approx(y=rho1.malha, x=Dist.rho1, xout=Aleatorio)$y
Resultados.Esperados[j,1:3] <- c(rho1.ini, Esp.cond.rho1, Var.cond.rho1)
rho1.malha <- (-25:25)/253sqrt(Var.cond.rho1)+Esp.cond.rho1

# Calculando a Densidade e a Distribuição Condicional para rho2.
for (i in 1:length(rho2.malha))
{
Dens.rho2[i] <- f.cond.rho2(rho2.malha[i])
}
Dens.rho2 <- Dens.rho2 - max(Dens.rho2)
Dens.rho2 <- exp(Dens.rho2)
Dist.rho2 <- Dens2Dist(rho2.malha, Dens.rho2)
Dens.rho2 <- Dens.rho2 / Dist.rho2[length(rho2.malha)]
Dist.rho2 <- Dist.rho2 / Dist.rho2[length(rho2.malha)]
Esp.cond.rho2 <- Dens2Dist(rho2.malha, Dens.rho2 rho2.malha)[length(rho2.malha)]
Var.cond.rho2 <- Dens2Dist(rho2.malha, Dens.rho2 rho2.malha^2)[length(rho2.malha)]
- Esp.cond.rho2^2 Aleatório <- runif(1)
rho2.ini <- approx(y=rho2.malha, x=Dist.rho2, xout=Aleatorio)$y
Resultados.Esperados[j,4:6] <- c(rho2.ini, Esp.cond.rho2, Var.cond.rho2)
rho2.malha <- (-25:25)/253sqrt(Var.cond.rho2)+Esp.cond.rho2
# Calculando a Densidade e a Distribuicao Condicional para C1.
for (i in 1:length(C1.malha))

```

```

{
Dens.C1[i] <- f.cond.C1(C1.malha[i])
}
Dens.C1 <- Dens.C1 - max(Dens.C1)
Dens.C1 <- exp(Dens.C1)
Dist.C1 <- Dens2Dist(C1.malha, Dens.C1)
Dens.C1 <- Dens.C1 / Dist.C1[length(C1.malha)]
Dist.C1 <- Dist.C1 / Dist.C1[length(C1.malha)]
Esp.cond.C1 <- Dens2Dist(C1.malha, Dens.C1 C1.malha)[length(C1.malha)]
Var.cond.C1 <- Dens2Dist(C1.malha, Dens.C1 C1.malha2)[length(C1.malha)]
- Esp.cond.C12
Aleatorio <- runif(1)
C1.ini <- approx(y=C1.malha, x=Dist.C1, xout=Aleatorio)$y
Resultados.Esperados[j,7:9] <- c(C1.ini, Esp.cond.C1, Var.cond.C1)
C1.malha <- (-25:25)/253sqrt(Var.cond.C1)+Esp.cond.C1
# Calculando a Densidade e a Distribuição Condicional para C2.
for (i in 1:length(C2.malha))
{
Dens.C2[i] <- f.cond.C2(C2.malha[i])
}
Dens.C2 <- Dens.C2 - max(Dens.C2)
Dens.C2 <- exp(Dens.C2)
Dist.C2 <- Dens2Dist(C2.malha, Dens.C2)
Dens.C2 <- Dens.C2 / Dist.C2[length(C2.malha)]
Dist.C2 <- Dist.C2 / Dist.C2[length(C2.malha)]
Esp.cond.C2 <- Dens2Dist(C2.malha, Dens.C2 C2.malha)[length(C2.malha)]
Var.cond.C2 <- Dens2Dist(C2.malha, Dens.C2 C2.malha2)[length(C2.malha)]
- Esp.cond.C22
Aleatorio <- runif(1)
C2.ini <- approx(y=C2.malha, x=Dist.C2, xout=Aleatorio)$y
Resultados.Esperados[j,10:12] <- c(C2.ini, Esp.cond.C2, Var.cond.C2)
C2.malha <- (-25:25)/253sqrt(Var.cond.C2)+Esp.cond.C2
# Calculando a Densidade e a Distribuicao Condicional para C3.
for (i in 1:length(C3.malha))
{
Dens.C3[i] <- f.cond.C3(C3.malha[i])
}
Dens.C3 <- Dens.C3 - max(Dens.C3)
Dens.C3 <- exp(Dens.C3)
Dist.C3 <- Dens2Dist(C3.malha, Dens.C3)
Dens.C3 <- Dens.C3 / Dist.C3[length(C3.malha)]
Dist.C3 <- Dist.C3 / Dist.C3[length(C3.malha)]
Esp.cond.C3 <- Dens2Dist(C3.malha, Dens.C3 C3.malha)[length(C3.malha)]

```

```

Var.cond.C3 <- Dens2Dist(C3.malha, Dens.C3 C3.malha2)[length(C3.malha)]
- Esp.cond.C32
Aleatorio <- runif(1)
C3.ini <- approx(y=C3.malha, x=Dist.C3, xout=Aleatorio)$y
Resultados.Esperados[j,13:15] <- c(C3.ini, Esp.cond.C3, Var.cond.C3)
C3.malha <- (-25:25)/253sqrt(Var.cond.C3)+Esp.cond.C3
# Calculando a Densidade e a Distribuicao Condicional para C4.
for (i in 1:length(C4.malha))
{
Dens.C4[i] <- f.cond.C4(C4.malha[i])
}
Dens.C4 <- Dens.C4 - max(Dens.C4)
Dens.C4 <- exp(Dens.C4)
Dist.C4 <- Dens2Dist(C4.malha, Dens.C4)
Dens.C4 <- Dens.C4 / Dist.C4[length(C4.malha)]
Dist.C4 <- Dist.C4 / Dist.C4[length(C4.malha)]
Esp.cond.C4 <- Dens2Dist(C4.malha, Dens.C4 C4.malha)[length(C4.malha)]
Var.cond.C4 <- Dens2Dist(C4.malha, Dens.C4 C4.malha2)[length(C4.malha)]
- Esp.cond.C42
Aleatorio <- runif(1)
C4.ini <- approx(y=C4.malha, x=Dist.C4, xout=Aleatorio)$y
Resultados.Esperados[j,16:18] <- c(C4.ini, Esp.cond.C4, Var.cond.C4)
C4.malha <- (-25:25)/253sqrt(Var.cond.C4)+Esp.cond.C4
# Calculando a Densidade e a Distribuicao Condicional para C5.
for (i in 1:length(C5.malha))
{
Dens.C5[i] <- f.cond.C5(C5.malha[i])
}
Dens.C5 <- Dens.C5 - max(Dens.C5)
Dens.C5 <- exp(Dens.C5)
Dist.C5 <- Dens2Dist(C5.malha, Dens.C5)
Dens.C5 <- Dens.C5 / Dist.C5[length(C5.malha)]
Dist.C5 <- Dist.C5 / Dist.C5[length(C5.malha)]
Esp.cond.C5 <- Dens2Dist(C5.malha, Dens.C5 C5.malha)[length(C5.malha)]
Var.cond.C5 <- Dens2Dist(C5.malha, Dens.C5 C5.malha2)[length(C5.malha)]
- Esp.cond.C52
Aleatorio <- runif(1)
C5.ini <- approx(y=C5.malha, x=Dist.C5, xout=Aleatorio)$y
Resultados.Esperados[j,19:21] <- c(C5.ini, Esp.cond.C5, Var.cond.C5)
C5.malha <- (-25:25)/253sqrt(Var.cond.C5)+Esp.cond.C5
# Calculando a Densidade e a Distribuicao Condicional para Beta1.
for (i in 1:length(Beta1.malha))
{

```

```

Dens.Beta1[i] <- f.cond.Beta1(Beta1.malha[i])
}
Dens.Beta1 <- Dens.Beta1 - max(Dens.Beta1)
Dens.Beta1 <- exp(Dens.Beta1)
Dist.Beta1 <- Dens2Dist(Beta1.malha, Dens.Beta1)
Dens.Beta1 <- Dens.Beta1 / Dist.Beta1[length(Beta1.malha)]
Dist.Beta1 <- Dist.Beta1 / Dist.Beta1[length(Beta1.malha)]
Esp.cond.Beta1 <- Dens2Dist(Beta1.malha, Dens.Beta1 Beta1.malha)[length(Beta1.malha)]
Var.cond.Beta1 <- Dens2Dist(Beta1.malha, Dens.Beta1 Beta1.malha2)[length(Beta1.malha)]
- Esp.cond.Beta12
Aleatorio <- runif(1)
Beta1.ini <- approx(y=Beta1.malha, x=Dist.Beta1, xout=Aleatorio)$y
Resultados.Esperados[j,22:24] <- c(Beta1.ini, Esp.cond.Beta1, Var.cond.Beta1)
Beta1.malha <- (-25:25)/253sqrt(Var.cond.Beta1)+Esp.cond.Beta1
# Calculando a Densidade e a Distribuicao Condicional para Beta2.
for (i in 1:length(Beta2.malha))
{
Dens.Beta2[i] <- f.cond.Beta2(Beta2.malha[i])
}
Dens.Beta2 <- Dens.Beta2 - max(Dens.Beta2)
Dens.Beta2 <- exp(Dens.Beta2)
Dist.Beta2 <- Dens2Dist(Beta2.malha, Dens.Beta2)
Dens.Beta2 <- Dens.Beta2 / Dist.Beta2[length(Beta2.malha)]
Dist.Beta2 <- Dist.Beta2 / Dist.Beta2[length(Beta2.malha)]
Esp.cond.Beta2 <- Dens2Dist(Beta2.malha, Dens.Beta2 Beta2.malha)[length(Beta2.malha)]
Var.cond.Beta2 <- Dens2Dist(Beta2.malha, Dens.Beta2 Beta2.malha2)[length(Beta2.malha)]
- Esp.cond.Beta22
Aleatorio <- runif(1)
Beta2.ini <- approx(y=Beta2.malha, x=Dist.Beta2, xout=Aleatorio)$y
Resultados.Esperados[j,25:27] <- c(Beta2.ini, Esp.cond.Beta2, Var.cond.Beta2)
Beta2.malha <- (-25:25)/253sqrt(Var.cond.Beta2)+Esp.cond.Beta2
# Calculando a Densidade e a Distribuicao Condicional para Beta3.
for (i in 1:length(Beta3.malha))
{
Dens.Beta3[i] <- f.cond.Beta3(Beta3.malha[i])
}
Dens.Beta3 <- Dens.Beta3 - max(Dens.Beta3)
Dens.Beta3 <- exp(Dens.Beta3)
Dist.Beta3 <- Dens2Dist(Beta3.malha, Dens.Beta3)
Dens.Beta3 <- Dens.Beta3 / Dist.Beta3[length(Beta3.malha)]
Dist.Beta3 <- Dist.Beta3 / Dist.Beta3[length(Beta3.malha)]
Esp.cond.Beta3 <- Dens2Dist(Beta3.malha, Dens.Beta3 Beta3.malha)[length(Beta3.malha)]
Var.cond.Beta3 <- Dens2Dist(Beta3.malha, Dens.Beta3 Beta3.malha2)[length(Beta3.malha)]

```

```

- Esp.cond.Beta32
Aleatorio <- runif(1)
Beta3.ini <- approx(y=Beta3.malha, x=Dist.Beta3, xout=Aleatorio)$y
Resultados.Esperados[j,28:30] <- c(Beta3.ini, Esp.cond.Beta3, Var.cond.Beta3)
Beta3.malha <- (-25:25)/253sqrt(Var.cond.Beta3)+Esp.cond.Beta3
# Calculando a Densidade e a Distribuicao Condicional para Beta4.
for (i in 1:length(Beta4.malha))
{
Dens.Beta4[i] <- f.cond.Beta4(Beta4.malha[i])
}
Dens.Beta4 <- Dens.Beta4 - max(Dens.Beta4)
Dens.Beta4 <- exp(Dens.Beta4)
Dist.Beta4 <- Dens2Dist(Beta4.malha, Dens.Beta4)
Dens.Beta4 <- Dens.Beta4 / Dist.Beta4[length(Beta4.malha)]
Dist.Beta4 <- Dist.Beta4 / Dist.Beta4[length(Beta4.malha)]
Esp.cond.Beta4 <- Dens2Dist(Beta4.malha, Dens.Beta4 - Beta4.malha)[length(Beta4.malha)]
Var.cond.Beta4 <- Dens2Dist(Beta4.malha, Dens.Beta4 - Beta4.malha2)[length(Beta4.malha)]
- Esp.cond.Beta42
Aleatório <- runif(1)
Beta4.ini <- approx(y=Beta4.malha, x=Dist.Beta4, xout=Aleatorio)$y
Resultados.Esperados[j,31:33] <- c(Beta4.ini, Esp.cond.Beta4, Var.cond.Beta4)
Beta4.malha <- (-25:25)/253sqrt(Var.cond.Beta4)+Esp.cond.Beta4
print(j)

```

Apêndice C

Séries financeiras utilizadas na aplicação

Down Jones	Down Jones	Down Jones	Down Jones	Down Jones	Down Jones
10544,07	10678,14	10472,84	10416,41	9938,32	10216,54
10538,66	10591,48	10444,73	10371,47	9814,59	10177,68
10529,03	10593,11	10478,16	10395,07	9825,35	10239,92
10592,44	10588,00	10342,60	10479,57	9954,55	10125,40
10458,89	10595,55	10272,27	10443,81	9972,83	10055,20
10485,18	10529,48	10225,57	10371,84	10083,15	10077,18
10427,18	10456,96	10314,00	10357,09	10040,82	10002,33
10538,37	10296,89	10317,20	10413,43	10110,14	9928,31
10553,85	10128,38	10310,95	10435,48	10073,05	9933,38
10600,51	10240,08	10241,26	10334,16	10098,63	9956,32
10600,75	10102,89	10117,34	10282,83	10181,74	9897,62
10528,66	10184,67	9990,02	10219,34	10173,41	9886,93
10623,62	10300,30	10019,47	10240,29	10195,01	9865,76
10623,18	10295,78	10045,16	10171,56	10122,52	9757,81
10568,29	10186,60	10010,74	10213,22	10173,92	9749,99
10702,51	10064,75	10012,87	10238,22	10168,46	9888,48
10609,92	10063,64	9906,91	10247,59	10290,28	10002,03
10468,37	10048,23	9968,51	10208,80	10260,20	10004,54
10510,29	10218,82	9937,71	10163,16	10342,79	10035,73
10488,07	10212,97	9937,64	10139,78	10313,36	10137,05
10499,18	10329,63	9966,74	10094,06	10289,10	10314,76
10505,18	10381,70	9958,43	10149,07	10313,07	10387,54
10470,74	10357,70	10117,62	10046,13	10314,76	10391,31
10495,55	10373,33	10109,89	10050,33	10318,16	10386,37
10593,03	10470,59	10205,20	9962,22	10231,36	10385,48
10579,03	10558,37	10188,45	9961,92	10244,49	10469,84
10613,85	10570,81	10202,65	10085,14	10284,46	10550,24
10737,70	10480,15	10262,97	10117,07	10204,89	10487,65
10694,07	10442,03	10195,91	10129,24	10244,93	10549,57
10627,85	10515,56	10242,82	10139,71	10109,18	10572,55
10714,88	10381,28	10391,08	10179,16	10038,90	10456,91
10671,99	10377,95	10432,52	10120,24	10047,24	10489,42
10664,73	10397,46	10410,10	10126,51	9988,54	10492,60
10566,37	10451,97	10334,73	9963,03	10077,40	
10601,62	10437,85	10380,43	9815,33	10136,24	
10580,14	10317,27	10379,58	9814,66	10080,27	
10583,92	10461,20	10377,52	9944,67	10192,65	

Ibovespa	Ibovespa	Ibovespa	Ibovespa	Ibovespa	Ibovespa
23531,70	20763,10	18325,60	21676,60	22501,50	23660,10
23320,10	21775,70	18401,70	21672,70	22534,30	23879,60
23716,80	21233,60	18611,20	22064,60	22286,40	23211,90
23916,70	21612,90	18122,30	22447,10	21968,00	23215,90
24237,00	21900,90	18556,20	22102,40	21809,20	23772,00
23939,60	22369,80	18688,20	22361,20	22309,40	24169,80
23398,60	22260,60	18240,00	21810,30	22875,70	24143,00
22961,70	21663,00	18285,40	21730,30	23073,50	24034,20
23154,60	21204,90	19069,40	21591,10	23078,20	24444,90
23385,20	20984,70	19734,30	21317,70	23105,70	24368,30
23471,00	21000,30	19667,30	21737,00	22748,90	24866,60
24349,80	21541,00	19544,60	22168,30	22943,50	24997,80
24226,00	21530,70	19545,80	22227,80	22972,00	24854,70
23851,60	22042,60	19716,60	22336,90	22834,00	25128,30
22386,20	22142,20	19399,90	22448,00	23230,60	25234,70
22280,90	22647,10	19818,70	22372,90	23208,30	25200,40
21091,60	23146,10	20446,70	22177,90	23245,20	25468,00
21968,80	23071,60	20283,70	21330,20	23777,00	25632,50
21961,50	22724,70	19864,70	21652,70	24150,40	24968,40
22171,40	22779,60	19487,60	21260,90	24205,20	24526,00
23197,10	22619,60	20047,40	21736,70	24027,10	25225,10
23063,40	22311,90	20459,70	21569,70	24104,20	25563,10
22529,80	21739,40	20334,60	21570,20	23953,90	25575,80
22188,80	21768,90	20293,00	21763,70	23289,30	25831,10
22426,20	21626,70	20199,20	22060,50	22959,50	25659,10
22000,10	21077,60	20836,10	22778,80	23367,80	25536,60
20950,90	21141,60	20708,30	22934,80	23411,10	25889,70
21609,20	21590,50	20750,20	23195,60	22893,00	25723,20
21449,60	21323,60	20350,90	22852,10	22873,50	25878,20
21755,00	21316,60	20800,80	22870,10	23058,90	25936,90
22498,50	20473,60	21148,90	23058,40	22735,50	26116,60
22442,40	19865,20	21348,70	22582,00	22602,60	26161,30
22550,40	19987,50	21568,30	22596,80	22848,90	26196,20
22392,90	20026,20	21188,00	22803,20	23170,80	
22993,30	19190,40	21171,20	22512,80	22928,50	
22673,70	17604,10	20887,40	22688,00	23052,20	
21670,20	18536,60	21532,20	22415,00	23273,00	

KLSE	KLSE	KLSE	KLSE	KLSE	KLSE
792,66	879,50	818,72	851,82	838,77	878,45
787,29	884,55	803,11	845,02	850,11	872,02
813,20	890,29	793,97	838,60	846,97	871,48
825,42	885,70	781,05	842,65	850,29	876,37
823,68	890,14	783,27	843,15	857,25	882,29
818,42	901,47	793,83	838,64	856,71	901,66
815,98	904,45	788,87	840,88	853,24	904,48
814,80	908,96	791,53	839,37	859,66	897,03
817,27	897,87	794,39	838,20	865,34	895,94
818,91	895,31	811,74	838,49	865,13	909,38
824,66	890,14	808,44	836,15	856,30	905,22
827,87	895,12	810,67	838,64	852,89	908,78
823,27	899,58	812,27	833,50	855,32	915,90
819,85	901,52	806,36	833,98	848,28	917,19
817,11	901,85	798,58	827,49	845,46	914,62
818,94	887,08	802,82	825,53	846,72	919,97
804,83	887,34	815,73	824,88	849,96	917,17
815,42	884,16	817,39	824,37	853,93	916,36
817,79	882,34	818,40	820,10	865,07	912,41
812,06	879,78	824,32	816,81	861,56	894,98
818,22	880,37	810,03	820,78	861,54	894,23
825,65	866,29	818,16	821,45	864,16	897,02
825,91	861,19	821,30	819,85	861,08	904,76
828,17	861,81	821,93	815,62	853,31	904,29
845,80	855,76	818,47	808,51	852,16	903,00
857,15	861,33	822,09	809,56	851,29	909,61
858,66	871,63	826,80	810,15	851,69	906,50
861,72	875,50	824,23	808,47	851,74	903,30
885,34	870,12	822,22	808,61	848,65	909,30
879,24	868,32	823,28	804,89	849,21	906,89
882,07	861,57	819,86	807,56	853,88	906,55
894,45	849,72	828,10	818,53	850,95	907,58
877,20	838,21	830,68	821,49	851,55	910,13
882,33	846,68	843,61	827,98	851,17	
886,15	842,77	855,08	828,38	852,99	
882,96	835,37	851,77	842,04	861,14	
875,83	810,16	851,13	841,90	864,04	

RTSI	RTSI	RTSI	RTSI	RTSI	RTSI
586,08	671,33	589,57	590,28	598,19	657,69
592,63	669,55	595,68	588,67	596,02	662,65
588,01	676,20	596,87	593,74	600,41	670,84
597,33	680,57	616,24	592,28	604,50	674,37
595,32	698,70	619,80	604,13	605,65	664,87
591,6	708,59	635,21	585,10	612,49	673,06
595,16	721,81	630,97	575,70	612,04	681,87
589,60	712,88	627,12	553,79	612,22	655,44
593,35	713,14	632,48	565,77	612,58	662,30
594,18	705,16	613,52	555,13	621,50	653,97
603,74	706,71	583,17	550,56	624,60	632,44
619,31	716,55	581,07	518,15	623,88	633,44
618,75	721,35	557,45	526,64	624,41	631,87
611,51	736,28	568,20	540,27	627,11	627,98
620,40	752,66	570,61	545,81	629,14	611,07
611,10	748,95	567,49	545,52	631,58	604,65
623,64	759,01	576,27	547,10	631,65	595,39
616,43	750,03	586,08	548,20	648,28	596,43
623,62	762,33	589,70	539,05	663,26	607,55
620,23	781,55	566,83	550,46	674,36	546,11
626,67	763,57	542,57	544,48	679,01	545,52
640,52	739,25	533,26	537,05	689,01	563,55
644,71	736,82	587,10	539,01	690,60	572,25
645,04	731,30	600,86	555,37	663,47	567,37
654,02	726,57	609,27	560,67	662,19	555,88
655,21	716,84	603,13	546,11	658,96	549,02
650,40	680,09	601,09	547,69	668,52	585,86
656,90	679,59	599,97	551,69	668,23	582,94
664,29	686,29	589,90	548,52	641,19	595,46
670,14	664,24	576,23	557,34	638,77	605,78
688,72	670,53	583,32	558,49	648,14	600,13
686,21	636,09	585,15	563,52	649,21	598,55
667,27	631,11	566,45	575,10	653,23	605,66
669,26	636,38	560,10	575,28	668,64	
673,25	633,05	572,34	583,67	657,66	
688,68	601,13	573,90	587,01	663,67	
680,86	576,63	574,32	589,24	665,71	

Bibliografia

- Aguilar, O. e West, M. (2000). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation, *Journal of Business and Economic Statistics* **18**: 338–357.
- Anderson, T., Fang, K. e Hsu, H. (1986). Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptical contoured distributions, *The Canadian Journal of Statistics* **44**: 55–59.
- Arellano-Vale, R. B. (1994). *Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão*, Tese de Doutorado, Instituto de Matemática e Estatística - USP, São Paulo, Brasil.
- Arminger, G. e Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm, *Psychometrika* **63**: 271–300.
- Bartholomew, D. J. (1981). Posterior analysis of the factor model, *British Journal of Mathematical and Statistical Psychology* **34**: 93–99.
- Bawens, L., Lubrano, M. e Richard, J. (1999). *Bayesian Inference in Dinamic Econometric Models*, Oxford University Press, Oxford.
- Belisle, C. (1992). Convergence theorems for a class of simulated annealing algorithms on r ., *Journal of Applied Probability* **29**: 885–895.

- Berkane, M. e Bentler, P. (1987). Characterizing parameters of multivariate elliptical distributions, *Communication in Statistics - Simulation and Computation* **16**: 193–198.
- Broemeling, L. (1985). *Bayesian Analysis of Linear Models*, Marcel Decker, New York.
- Brooks, S. e Gelman, A. (1997). General methods for monitoring convergence of iterative simulations, *Journal of Computational and Graphical Statistics* **6**: 251–265.
- Brooks, S. e Roberts, G. (1998). Assessment convergence of Markov Chain Monte Carlo algorithms, *Statistics and Computing* **8**: 319–335.
- Cambanis, S., Huang, S. e Simons, G. (1981). On the theory of elliptically contoured distributions, *Journal of Multivariate Analysis* **11**: 368–385.
- Carlin, B., Gelman, A., Neal, R. e Kass, R. (1998). Markov Chain Monte Carlo in practice: a roundtable discussion, *Journal of the American Statistical Association* **52**: 93–99.
- Casella, G. e George, E. (1992). Explaining the Gibbs Sampler, *The American Statistician* **46**: 327–335.
- Chib, S. e Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm, *The American Statistician* **49**: 1–21.
- Chmielewski, M. (1981). Elliptically symmetric distributions: a review and bibliography, *International Statistical Review* **49**: 67–74.
- Cowles, M. e Carlin, B. (1992). Bayesian robustness for elliptical distributions, *Revista Brasileira de Probabilidade e Estatística* **61**: 97–119.
- Cowles, M. e Carlin, B. (1996). Markov Chain Monte Carlo convergence diagnostics: a comparative review, *Journal of the American Statistical Association* **91**: 424–434.
- Dennis, J., Gay, D. e Welsch, R. (1981). An adaptive nonlinear least-squares algorithm., *ACM Transactions on Mathematical Software* **7**: 348–383.

- Devlin, S., Gnanadesikan, R. e Kettering, J. (1976). Some multivariate applications of elliptical distributions, *em Essays in Probability and Statistics*, S. K. et al. (editor), Shinko Tsusho, Tokyo, chapter 24, págs. 365–393.
- Diebold, F. e Nerlove, M. (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH, *Journal of Applied Econometrics* **4**: 1–21.
- Fang, K. T., Kotz, S. e Ng, K. (1990). *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.
- Ferreira, D. (2007). *Estatística Multivariada*, Editora UFLA, Lavras.
- Gelman, A. e Rubin, D. (1992). Inference from iterative simulation using multiple sequence, *Statistical Science* **7**: 457–511.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posteriori moments, *em Bayesian Statistics 4*, J. Bernardo, J. Berger, A. David e A. Smith (editores), Oxford University Press, New York.
- Geweke, J. (1997). Posterior simulators in econometrics, *em Advances in Economics and Econometrics: Theory and Applications*, D. Kreps e K. Wallis (editores), Cambridge University Press, Cambridge.
- Geweke, J. e Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory, *The Review of Financial Studies* **9**: 557–587.
- Ghosh, J. e Sinha, K. (1980). On the robustness of least squares procedures in regression models, *Journal of Multivariate Analysis*. **10**: 332–342.
- Harvey, A. e Chung, C. (2000). Estimating the underlying change in unemployment in the UK, *Journal of the Royal Statistical Society* **163**: 303–339.
- Hastings, W. (1970). Monte Carlo sampling methods using markov chains and their applications, *Biometrika* **57**: 97–109.

- Heidelberger, P. e Welch, P. (1983). Simulation run length control in the presence of an initial transient, *Operations Research* **31**: 1109–1144.
- Jarque, C. e Bera, A. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals., *Economics Letters*. **6**: 255–259.
- Jöreskog, K. (1966). Testing a simple structure hypothesis in factor analysis, *Psychometrika* **31**: 165–178.
- Kano, Y., Berkane, M. e Bentler, P. (1993). Statistical inference based on pseudo-maximum likelihood estimators in elliptical populations, *Journal of the American Statistical Association* **88**: 135–143.
- Kariya, T. e Eaton, M. (1977). Robust tests for spherical symmetry, *The Annals of Statistics* **5**: 206–215.
- Kelejian, H. e Prucha, I. (1985). Independent or uncorrelated distributions in linear regression, *Economics Letters* **19**: 35–38.
- Kelker, D. (1970). Distribution theory of special distributions and a location-scale parameter, *Sankhya* **32**: 319–430.
- Kibria, B. e Joarder, A. (2006). A short review of multivariate t -distribution, *Journal of Statistical Research* **40**: 59–72.
- Lange, K., Little, J. e Taylor, M. (1989). Robust statistical modeling using t -distribution, *Journal of the American Statistical Association* **84**: 881–896.
- Latif, S. A. (2000). *Modelagem de equações estruturais*, Dissertação de Mestrado, Instituto de Matemática e Estatística — USP, São Paulo.
- Lawley, D. N. e Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method*, American Elsevier Publishing Co, New York.

- Lee, S. (1981). A Bayesian approach to confirmatory factor analysis, *Psychometrika* **44**: 99–113.
- Lee, S. e Shi, J. (2000). Joint Bayesian analysis of factor scores and structural parameters in the factor analysis model, *Annals of the Institute of Statistical Mathematics*, Vol. 52, págs. 722–736.
- Lopes, H. e West, M. (2004). Bayesian model assessment in factor analysis, *Statistica Sinica* **14**: 41–67.
- Melhado, T. T. (2004). *Medidas de ajuste de modelos de equações estruturais*, Dissertação de Mestrado, Instituto de Matemática e Estatística — USP, São Paulo.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. e Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1091.
- Mitchell, A. (1988). Statistical manifolds of univariate elliptic distributions, *International Statistical Review* **56**: 1–16.
- Mitchell, A. (1989). The information matrix, skewness tensor and α -connections for the general multivariate elliptic distribution, *Annals of the Institute of Statistical Mathematics* **41**: 289–304.
- Mitchell, A. e Krzanowski, W. (1985). The Mahalanobis distance and elliptic distributions, *Biometrika* **72**: 464–467.
- Morettin, P. (2008). *Econometria Financeira*, Edgard Blücher, São Paulo.
- Muirhead, R. (1980). The effect of elliptical distributions on some standard procedures involving correlation coefficients: a review, *Multivariate Statistical Analysis* págs. 143–159.
- Muirhead, R. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley, New York.

- Muirhead, R. e Waternaux, C. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika* **67**: 31–43.
- Nadarajah, S. e Kotz, S. (2005). Mathematical properties of the multivariate t distribution, *Acta Applicandae Mathematica* **89**: 53–84.
- Nogueira, D. A. (2004). *Proposta e avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov: casos uni e multivariados*, Dissertação de Mestrado, Universidade Federal de Lavras — UFLA, Lavras.
- Paulino, C., Turkman, M. e Murteira, B. (2003). *Estatística Bayesiana*, Fundação Calouste Gulbenkian, Lisboa.
- Peña, D. e Box, G. (1987). Identifying a simplifying structure in time séries, *Journal of the American Statistical Association* **82**: 836–843.
- Pitt, M. e Shephard, N. (1998). Time varying covariances: a factor stochastic volatility approach, *em Bayesian Statistics*, J. B. et al. (editor), Oxford University Press, London.
- Polasek, W. (1997). Factor analysis and outliers: a Bayesian approach, *Discussion paper*.
- Pole, A., West, M. e Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*, Chapman and Hall, New York.
- Prucha, I. e Kelejian, H. (1984). The structure of simultaneous equation estimators: A generalization towards nonnormal disturbances, *Econometrica* **52**: 721–736.
- Raftery, A. e Lewis, S. (1992). Coment: One long run with diagnostics: implementation strategies for Markov Chain Monte Carlo, *Statistical Science* **7**: 493–497.
- Ritter, C. e Tanner, M. (1992). Facilitating the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sampler, *Journal of the American Statistical Association* **87**: 861–868.

- Sáfadi, T. (2008). Análise Bayesiana do modelo fatorial dinâmico aplicado às séries Down Jones, Ibovespa, KLSE e RTSI., *Submetido*.
- Sáfadi, T. e Peña, D. (2007). Bayesian analysis of dynamic factor models: an application to air pollution and mortality in São Paulo, Brazil, *Environmetrics* **18**: 1–20.
- Silva, W. S. (2005). *Inferência Bayesiana para os modelos da classe ARCH com potência assimétrica*, Tese de doutorado, Universidade Federal de Lavras - UFLA, Lavras.
- Singh, R. (1991). James-Stein rule estimators in linear regression models with multivariate t distributed error, *Australian Journal of Statistics* **33**: 145–158.
- Spearman, C. (1904). General intelligence objectively determined and measured, *American Journal of Psychology* **15**: 201–293.
- Sutradhar, B. (1993). Score test for the covariance matrix of the elliptical t distribution, *Journal of Multivariate Analysis* **46**: 1–12.
- Sutradhar, B. e Ali, M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate t model, *Journal of Multivariate Analysis* **29**: 155–162.
- Tanner, M. (1996). *Tools for Statistical Inference*, Springer Verlag, New York.
- Thurstone, L. (1944). A multiple group method of factoring the correlation matrix, *Psychometrika* **10**: 73–78.
- Thurstone, L. (1947). *Multiple Factor Analysis*, Chicago Press, Chicago.
- Tierney, L. (1994). Markov Chains for exploring posterior distributions, *Annals of Statistics* **22**: 1701–1762.
- Téran, J. (2007). *Modelos ARCH heterogêneos e aplicações à análise de dados de alta frequência*, Tese de Doutorado, Instituto de Matemática e Estatística — USP, São Paulo.

- West, M. e Harrison, J. (1989). *Bayesian Forecasting and Dynamic Models*, Springer Verlag, New York.
- Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York.
- Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms, *Journal of the American Statistical Association* **71**: 400–405.