

Universidade Federal do Rio de Janeiro



MODELO DE VOLATILIDADE ESTOCÁSTICA NA MÉDIA  
COM ERROS HIPERBÓLICOS GENERALIZADOS T-STUDENT  
ASSIMÉTRICOS

William Lima Leão

2012

# MODELO DE VOLATILIDADE ESTOCÁSTICA NA MÉDIA COM ERROS HIPERBÓLICOS GENERALIZADOS T-STUDENT ASSIMÉTRICOS

William Lima Leão

Dissertação submetida ao Corpo Docente do Instituto de Matemática - Departamento de Métodos Estatísticos da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do grau de Mestre em Estatística.

Orientador: Carlos Antonio Abanto-Valle

Rio de Janeiro

Fevereiro de 2012

# Modelo de Volatilidade Estocástica na Média com erros Hiperbólicos Generalizados t-Student Assimétricos

William Lima Leão

Dissertação submetida ao Corpo Docente do Instituto de Matemática - Departamento de Métodos Estatísticos da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do grau de Mestre em Estatística.

Aprovada por:

---

Prof<sup>o</sup>. Carlos A. Abanto-Valle  
DME - UFRJ - Orientador

---

Prof<sup>o</sup>. Fernando Antonio da Silva Moura  
DME - UFRJ

---

Prof<sup>o</sup>. Basilio de Bragança Pereira  
COPPE - UFRJ

Rio de Janeiro

Fevereiro de 2012

# Agradecimentos

Aos meus avós que me deram todo apoio que eu precisei não só no mestrado como em toda em minha vida.

A minha mãe, por sempre acreditar em mim e me dar força para continuar.

Ao Professor Carlos Abanto, pela confiança, amizade e apoio que me deu durante essa jornada.

À Professora Flávia Landim, que sempre me ajudou quando eu precisei não só mestrado como também na graduação.

À Professora Beatriz Mendez, que graças a ela consegui chegar onde estou.

Ao Giuseppe por me ajudar nas matérias mais difíceis e sempre me dar dicas sobre o R e o Latex, a Julia pela amizade e por sempre me dar força para trabalhar, a Estelina por ser a “rainha do drama” me ajudou bastante a ficar calmo e também a todas as pessoas que conheci no mestrado que me ajudaram a completar este trabalho.

Aos meus amigos do mestrado que sempre me ajudaram quando eu precisei.

Ao Departamento de Métodos Estatísticos do Instituto de Matemática, pela oportunidade e à CAPES e FAPERJ pelo apoio financeiro.

# Resumo

Esta dissertação de mestrado, considera o modelo de volatilidade estocástica na média na qual a distribuição condicional dos dados é a distribuição hiperbólica generalizada t-Student assimétrica. Esta classe de modelos exibe uma flexibilidade muito grande capturando os principais fatos estilizados presentes nas séries de retornos financeiros: conglomerados de volatilidade, caudas pesadas, assimetria negativa, os efeitos de alavancagem e “feedback”.

Fazendo uso de uma abordagem Bayesiana, o procedimento de inferência e estimação é baseado em métodos de simulação estocástica via cadeias de Markov (MCMC). Devido à natureza não linear da volatilidade, a presença de alavancagem e explorando a representação de mistura normal média-variância da distribuição hiperbólica generalizada t-Student assimétrica, um algoritmo é desenvolvido para estimar eficientemente por blocos as log-volatilidades.

Uma aplicação com dados artificiais é realizada com o intuito de mostrar a eficiência do método. Finalmente, os algoritmos desenvolvidos são aplicados na análise dos retornos do índice da Bolsa de São Paulo. Os resultados indicam que o ajuste do modelo de volatilidade estocástica na média com erros hiperbólicos generalizados t-Student assimétricos supera ao equivalente normal e t-Student.

**Palavras Chaves:** modelos dinâmicos não lineares/não gaussianos, MCMC, volatilidade estocástica, amostragem por blocos.

# Abstract

This thesis consider the stochastic volatility in mean model in which the conditional distribution of the data is the generalized hyperbolic skew Student-t density. This class of models exhibit a great deal of flexibility in capturing the most important features of high frequency financial time series: volatility clustering, excess kurtosis, negative skewness, the feedback and leverage effects.

Using the Bayesian paradigm, the estimation and inference procedure is based on Markov Chain Monte Carlo methods. Due to the non-linear form of the volatility, the leverage effect and based on the mean-variance normal mixture representation of the generalized hyperbolic skew Student-t , an efficient algorithm is developed to sampling the log-volatilities by blocks.

An application with an artificial data set is performed with the objective to show the efficiency of the method. Finally, we apply the algorithm to analyze daily stock return data from the São Paulo Stock,Mercantile & Futures Exchange index(IBOVESPA). The Bayesian predictive information criteria(BPIC) is used as model selection criteria. The results reveal that the stochastic volatility in mean model with correlated errors and generalized hyperbolic skew Student-t distribution provides a significant improvement in model fit for the IBOVESPA data over the normal and the Student-t models.

**Key words:** block sampling, dynamic non linear/non gaussian models, MCMC, stochastic volatility.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Preliminares</b>	<b>4</b>
2.1	Inferência Bayesiana . . . . .	4
2.2	Método Aceitação-Rejeição . . . . .	5
2.3	Teoria de Monte Carlo com Cadeia de Markov . . . . .	7
2.3.1	Cadeias de Markov . . . . .	7
2.3.2	O algoritmo de Metropolis-Hastings . . . . .	8
2.3.3	O algoritmo Aceitação-Rejeição Metropolis-Hastings . . . . .	9
2.3.4	O amostrador de Gibbs . . . . .	11
2.3.5	Avaliação de Convergência . . . . .	12
2.4	Seleção de Modelos . . . . .	14
2.4.1	Introdução . . . . .	14
2.4.2	Estrutura Geral . . . . .	14
2.4.3	Critério de Informação Preditiva Bayesiana . . . . .	15
2.4.4	Fator de Bayes . . . . .	16
<b>3</b>	<b>Modelos Dinâmicos</b>	<b>18</b>
3.1	Introdução . . . . .	18
3.2	Modelo Linear Dinâmico (MLD) . . . . .	18
3.2.1	Filtro de Kalman . . . . .	19
3.2.2	Suavizador de Kalman . . . . .	20
3.2.3	Filtro de Perturbações . . . . .	21

3.2.4	Pertubações Suavizadas . . . . .	22
3.2.5	Algoritmo de de Jong e Shephard . . . . .	23
<b>4</b>	<b>Distribuição Hiperbólica Generalizada t-Student Assimétrica</b>	<b>24</b>
4.1	Introdução . . . . .	24
4.2	Distribuição Hiperbólica Generalizada . . . . .	25
4.2.1	Parâmetros da distribuição HG . . . . .	26
4.3	Distribuição Hiperbólica Generalizada t-Student Assimétrica . . . . .	29
<b>5</b>	<b>Modelo VEM utilizando a distribuição HG t-Student Assimétrica</b>	<b>33</b>
5.1	Modelo VEM-HGt . . . . .	34
5.2	Estimação dos parâmetros . . . . .	35
5.2.1	Distribuições condicionais completas . . . . .	37
5.3	Amostragem por blocos . . . . .	40
<b>6</b>	<b>Aplicação</b>	<b>49</b>
6.1	Aplicação com dados simulados . . . . .	49
6.2	Aplicação com dados empíricos . . . . .	55
<b>7</b>	<b>Conclusões e Futuros Desenvolvimentos</b>	<b>63</b>



# Capítulo 1

## Introdução

O mercado financeiro sempre esteve vinculado à idéia de alto risco e grande possibilidade de perda. Para ajudar a minimizar os riscos das operações são utilizados diversos modelos estatísticos aplicados aos dados históricos das cotações, que, com isso, tentam prever seu comportamento futuro do mercado. No aspecto estatístico, esses dados podem ser vistos como uma série temporal financeira.

Uma forma flexível de descrever modelos para séries financeiras é fazer uso de modelos dinâmicos, muitos dos problemas em estatística podem ser tratados dentro dessa classe. Desenvolvimentos em métodos de simulação via Cadeias de Markov tem incrementado a utilização de inferência Bayesiana em modelos dinâmicos em diversas áreas como por exemplo a área financeira o que será o foco desta dissertação.

A volatilidade ocupa um papel muito importante em finanças, ela é uma medida de variação do preço de ativos financeiros ao longo do tempo. Existem duas classes de modelos de volatilidade: os modelos de volatilidade determinística representada pelos modelos ARCH, GARCH entre outros e os modelos de volatilidade estocástica.

O modelo de Volatilidade Estocástica (VE) foi introduzido por Tauchen e Pitts (1983) e Taylor (1982,1986) como um modo de descrever a variação no tempo da volatilidade de retornos de ativos, assumindo a distribuição normal nos retornos. Infelizmente a suposição de normalidade é muito restritiva podendo haver falta de robustez na presença de observações extremas, “outliers”, e a presença de assimetria o que pode causar um efeito significativo na inferência do modelo.

A relação entre retornos e volatilidades esperadas tem sido examinada amplamente nos últimos anos na literatura financeira. A teoria geralmente prediz a existência de uma relação positiva entre retornos esperados e a volatilidade, se os investidores são avessos ao risco. Em outras palavras, os investidores exigem um retorno esperado maior do título que é mais arriscado. Entretanto French et al. (1987) sugerem uma forte evidência de uma relação negativa entre retornos e inovações para o processo de volatilidade, o que interpreta como uma evidência indireta de uma correlação positiva entre o prêmio de risco esperado e volatilidade *ex ante*. Esta teoria, conhecida como volatilidade “feedback”, estabelece que as reações nos preços dos ativos induzidas por notícias ruins tendem a ser maiores que as reações a notícias boas. Em outras palavras, uma notícia ruim (boa) diminui (aumenta) os preços das ações e aumenta (diminui) a volatilidade, assim determinando uma redução (aumento) ainda maior dos preços. A consequência é uma certa quantidade de assimetria na distribuição dos retornos e uma correlação negativa entre retornos correntes e a volatilidade futura.

Uma explicação alternativa para a assimetria na volatilidade onde a causalidade corre na direção oposta é o efeito de alavancagem apresentado por Black (1976), que afirmou que um choque de retorno negativo (positivo) leva a um aumento (redução) na razão de alavancagem de uma empresa financeira, que tem um efeito (negativo) positivo na volatilidade dos retornos de suas ações. Campbell e Hentschel (1992) encontraram evidência do efeito “feedback” e de alavancagem, ao passo que, Bekaert e Wu (2000) apresentaram resultados sugerindo que a volatilidade “feedback” domina empiricamente o efeito de alavancagem.

Frequentemente, a volatilidade de retornos diários de ações tem sido estimada com modelos VE, mas os resultados tem contado com uma extensiva pre-modelagem destas séries para evitar o problema de estimação simultânea da média e variância. Para lidar com este problema, Koopman e Uspensky (2002) introduziram o modelo VE na Média (VEM), incorporando a não observada volatilidade como uma variável explicativa na equação da média dos retornos e supondo normalidade na distribuição dos retornos. Eles utilizaram o método de máxima verossimilhança simulada para estimação dos parâmetros e forneceram uma justificação empírica que o coeficiente de volatilidade na média está

relacionado com o efeito “feedback”.

Nesta dissertação será proposto um modelo que captura diversos fatos estilizados presentes nas séries de retornos financeiros como conglomerados de volatilidade, caudas pesadas, assimetria negativa, os efeitos de alavancagem e “feedback”, usando o modelo de volatilidade estocástica na média na qual a distribuição condicional dos dados é a distribuição generalizada t-Student assimétrica (HGt) como uma alternativa ao equivalente modelo de volatilidade estocástica normal. Nos referimos a esta generalização como modelo VEM-HGt.

Fazendo uso de uma abordagem Bayesiana, o procedimento de inferência e estimação é baseado em métodos de simulação estocástica via cadeias de Markov (MCMC). Devido à natureza não linear da volatilidade, a presença de alavancagem e explorando a representação de mistura normal média-variância da distribuição hiperbólica generalizada t-Student assimétrica, um algoritmo é desenvolvido para estimar eficientemente por blocos as log-volatilidades.

A dissertação está organizada da seguinte forma: No Capítulo 2, são descritos de uma forma geral os conceitos fundamentais de inferência Bayesiana, métodos de simulação de Monte Carlo com Cadeia de Markov (MCMC) e critérios para seleção de modelos.

No capítulo 3, apresenta-se o modelo linear dinâmico, o filtro de Kalman, o suavizador de Kalman, o filtro de perturbações, perturbações suavizadas e o algoritmo de de Jong e Shephard.

No capítulo 4, apresenta-se a distribuição hiperbólica generalizada t-Student assimétrica, que será utilizada no modelo proposto da dissertação.

No capítulo 5, apresenta-se o de volatilidade estocástica na média com erros hiperbólicos generalizados t-student assimétricos proposto nesta dissertação, além do método de estimação do modelo e o algoritmo de amostragem por blocos.

No capítulo 6, é realizada uma aplicação com dados simulados e uma aplicação empírica do modelo.

No capítulo 7, finalizando a dissertação são apresentadas conclusões e sugestões para pesquisas futuras.

# Capítulo 2

## Preliminares

Neste Capítulo são apresentados os conceitos básicos que serão utilizados nos próximos capítulos no desenvolvimento da dissertação. Na seção 2.2 são apresentadas noções sobre inferência Bayesiana. Na seção 2.3 é realizada uma breve introdução sobre métodos de Monte Carlo com Cadeias de Markov (MCMC). Finalmente na seção 2.4 são apresentados critérios de seleção de modelos.

### 2.1 Inferência Bayesiana

Considere que  $\mathbf{H}$  denota a informação inicial disponível sobre algum vetor de parâmetros de interesse  $\boldsymbol{\theta}$ . Além disso assume-se que esta informação inicial é expressa em termos probabilísticos por  $p(\boldsymbol{\theta}|\mathbf{H})$ .

Considere-se que o vetor de quantidades aleatórias  $\mathbf{Y}$  relacionado a  $\boldsymbol{\theta}$  é observado, assim provendo mais informação a respeito de  $\boldsymbol{\theta}$ . Antes de observar  $\mathbf{Y}$ , devemos conhecer a distribuição amostral de  $\mathbf{Y}$  dada por  $p(\mathbf{Y}|\boldsymbol{\theta}, \mathbf{H})$ , onde a dependência em  $\boldsymbol{\theta}$  é claramente indicada.

Depois de observar  $\{\mathbf{Y} = \mathbf{y}\}$  a quantidade de informação que tem-se sobre  $\boldsymbol{\theta}$  é modificada de  $\mathbf{H}$  para  $\mathbf{H} \cap \{\mathbf{Y} = \mathbf{y}\}$ .

Agora a informação sobre  $\boldsymbol{\theta}$  é dada por  $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{H})$  e só o que é preciso saber é como passar de  $p(\boldsymbol{\theta}|\mathbf{H})$  para  $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{H})$ . Para atualizar esta informação usa-se o teorema de Bayes, isto é:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{H}) &= \frac{p(\boldsymbol{\theta}, \mathbf{y}|\mathbf{H})}{p(\mathbf{y}|\mathbf{H})} \\ &= \frac{p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{H})p(\boldsymbol{\theta}|\mathbf{H})}{p(\mathbf{y}|\mathbf{H})}, \end{aligned}$$

onde

$$p(\mathbf{y}|\mathbf{H}) = \int_{\Theta} p(\mathbf{y}, \boldsymbol{\theta}|\mathbf{H})d\boldsymbol{\theta},$$

tal que  $\Theta$  é conjunto de valores assumidos por  $\boldsymbol{\theta}$ .

Em diversas aplicações, a integral acima não possui forma analítica fechada e sua avaliação usando métodos numéricos em dimensões maiores do que 20 torna-se inviável.

Como  $\mathbf{H}$  é um termo comum em todas as distribuições ele será omitido, por simplicidade de notação. Assim, o teorema de Bayes pode ser reescrito in termos mais usuais da forma:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (2.1)$$

Para aproximar a distribuição a posteriori (2.1) usa-se, frequentemente, integração por Monte Carlo ou MCMC.

O teorema resulta em uma atualização de informação sobre  $\boldsymbol{\theta}$ , começando de  $p(\boldsymbol{\theta})$  e indo para  $p(\boldsymbol{\theta}|\mathbf{y})$ . Por esta razão as distribuições  $p(\boldsymbol{\theta})$  e  $p(\boldsymbol{\theta}|\mathbf{y})$  são chamadas de distribuição a priori e distribuição a posteriori, respectivamente.

A seguir são introduzidos conceitos fundamentais sobre métodos de simulação que serão utilizados nos próximos capítulos.

## 2.2 Método Aceitação-Rejeição

O método (A-R) é uma técnica de amostragem utilizada para gerar amostras independentes e identicamente distribuídas de uma distribuição de interesse  $\pi(\boldsymbol{\theta})$ .

Esse método requer a especificação de uma densidade proposta  $q(\boldsymbol{\theta})$  da qual seja simples de se amostrar. A escolha de  $q(\boldsymbol{\theta})$  é feita de forma que  $\pi(\boldsymbol{\theta}) \leq cq(\boldsymbol{\theta})$  para todo  $\boldsymbol{\theta}$ ,

onde  $c > 0$  é uma constante conhecida.

### Algoritmo A-R

1. Gerar  $\boldsymbol{\theta}^*$  de  $q(\boldsymbol{\theta})$ ,
2. Gerar  $u \sim U(0, 1)$ ,
3. Calcular a probabilidade de aceitação  $w(\boldsymbol{\theta}^*) = \pi(\boldsymbol{\theta}^*)/cq(\boldsymbol{\theta}^*)$ ,
4. Se  $u < w(\boldsymbol{\theta}^*)$  aceita-se o valor  $\boldsymbol{\theta}^*$ , caso contrário rejeita-se

**Proposição 2.2.1.** *Se  $\boldsymbol{\theta}^*$  é um valor obtido pelo método A-R, a distribuição de  $\boldsymbol{\theta}^*$  é  $\pi$ .*

**Prova:** (Ripley, 1987) Seja  $\boldsymbol{\theta}^*$  uma variável gerada pelo método A-R, e  $u \sim U(0, 1)$ . Consideremos que  $\mathcal{A}$  é um subconjunto do espaço paramétrico. Logo, tem que ser provado que  $P(\boldsymbol{\theta}^* \in \mathcal{A}) = P(\boldsymbol{\theta} \in \mathcal{A} | \text{aceitar } \boldsymbol{\theta}^*)$ .

Por definição, a probabilidade condicional  $P(\boldsymbol{\theta} \in \mathcal{A} | \text{aceitar } \boldsymbol{\theta}^*)$  é dada por:

$$P(\boldsymbol{\theta}^* \in \mathcal{A} | \text{aceitar } \boldsymbol{\theta}^*) = \frac{P(\boldsymbol{\theta}^* \in \mathcal{A}, \text{aceitar } \boldsymbol{\theta}^*)}{P(\text{aceitar } \boldsymbol{\theta}^*)} \quad (2.2)$$

O numerador em (2.2) é equivalente a

$$P(\boldsymbol{\theta}^* \in \mathcal{A}, \text{aceitar } \boldsymbol{\theta}^*) = P(\boldsymbol{\theta}^* \in \mathcal{A}, u \leq w(\boldsymbol{\theta}^*))$$

isto é

$$\begin{aligned} P(\boldsymbol{\theta}^* \in \mathcal{A}, \text{aceitar } \boldsymbol{\theta}^*) &= \int_{\mathcal{A}} \int_0^{w(\boldsymbol{\theta})} du q(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \int_{\mathcal{A}} w(\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned}$$

O denominador de (2.2), a probabilidade de  $\boldsymbol{\theta}$  ser aceito, é dada por

$$\begin{aligned} P(\text{aceitar } \boldsymbol{\theta}^*) &= P(u \leq w(\boldsymbol{\theta}^*)) \\ &= \int \int_0^{w(\boldsymbol{\theta})} q(\boldsymbol{\theta}) du d\boldsymbol{\theta} \\ &= \int w(\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned}$$

Finalmente tem-se que

$$\begin{aligned} P(\boldsymbol{\theta}^* \in \mathcal{A} | \text{aceitar } \boldsymbol{\theta}^*) &= \frac{\int_{\mathcal{A}} w(\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int w(\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta}} \\ &= \frac{\int_{\mathcal{A}} \pi(\boldsymbol{\theta})}{\int \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \end{aligned}$$

Portanto  $\boldsymbol{\theta}^*$  se distribui com função de densidade  $\pi(\boldsymbol{\theta})$ .

Para que o método funcione com eficiência é preciso escolher cuidadosamente a constante  $c$ , tal que garanta o envelopamento total da densidade  $\pi$ , isto é,  $\pi(\boldsymbol{\theta}) \leq cq(\boldsymbol{\theta})$  para todo  $\boldsymbol{\theta}$ . A melhor escolha para  $c$  é dada por

$$c = \sup_{\boldsymbol{\theta}} \frac{\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}$$

## 2.3 Teoria de Monte Carlo com Cadeia de Markov

Esta seção introduz técnicas de simulação estocástica via Cadeia de Markov os quais serão usados para obter amostras de densidades complexas.

### 2.3.1 Cadeias de Markov

Uma cadeia de Markov é uma coleção de variáveis ou vetores aleatórios  $\{\mathbf{X}_i \in \Omega | i \in M\}$ , onde usualmente considera-se  $M = \mathbb{N}$ . A evolução da cadeia de Markov no espaço  $\Omega \subset \mathbb{R}^p$  é dada pelo núcleo de transição

$$\begin{aligned} P(x, A) &= Pr(X_{i+1} \in A | X_i = x, \{X_j, j < i\}) \\ &= Pr(X_{i+1} \in A | X_i = x) \end{aligned}$$

onde  $x \in \Omega$  e  $A \subset \Omega$ .

Isso quer dizer que uma cadeia de Markov é um processo estocástico, em que, a probabilidade de transição estado o estado  $i + 1$  depende somente do estado anterior  $i$ .

### 2.3.2 O algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings (M-H) foi inicialmente proposto por Metropolis et al. (1953) e depois estendido por Hastings (1970). O algoritmo M-H é um método MCMC para obtenção de uma sequência de amostras aleatórias de uma distribuição,  $\pi(\boldsymbol{\theta})$ , para a qual a amostragem direta é difícil.

A ideia geral do algoritmo é usar Cadeia de Markov que, em períodos suficientemente grandes, geram estados que obedecem a distribuição de interesse.

Suponhamos que  $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)})$  é a distribuição proposta, para a distribuição de interesse  $\pi(\boldsymbol{\theta})$ , a qual será utilizada para obter  $\boldsymbol{\theta}^*$  dado o valor atual  $\boldsymbol{\theta}^{(i-1)}$  e seja  $\alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}^*)$  a probabilidade de aceitação

$$\alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}^*) = \min \left\{ \frac{\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{(i-1)}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{(i-1)})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})}, 1 \right\} \quad (2.3)$$

Os valores simulados podem ser obtidos a partir do seguinte procedimento recursivo:

#### Algoritmo M-H

1. Fazer  $i = 0$ , inicializar com o valor inicial  $\boldsymbol{\theta}^{(0)}$  tal que  $\pi(\boldsymbol{\theta}^{(0)}) > 0$ ,
2. Gerar da proposta  $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)})$ ,
3. Gerar  $u \sim U(0, 1)$ ,
4. Fazer  $\boldsymbol{\theta}^{(i)} = \begin{cases} \boldsymbol{\theta}^*, & \text{se } u \leq \alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}^*) \\ \boldsymbol{\theta}^{(i-1)} & \text{caso contrário} \end{cases}$ ,
5. Incrementar  $i = i + 1$ , voltar ao passo 2 e continuar o procedimento até alcançar a convergência.

**Proposição 2.3.1** (Tierney,1994). *A cadeia de Markov obtida pelo algoritmo M-H é reversível e tem a  $\pi(\boldsymbol{\theta})$  como sua distribuição estacionária.*



## Casos Particulares

O algoritmo M-H é bastante geral, e pode, pelo menos a princípio, ser implementado com qualquer distribuição  $\pi(\boldsymbol{\theta})$  e para qualquer proposta. Entretanto, sob o ponto de vista prático, a escolha da proposta é crucial para sua convergência para a distribuição a posteriori.

Quando a distribuição a posteriori não depende do valor da iteração anterior,  $q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i-1)}) = q(\boldsymbol{\theta})$ , resulta em uma cadeia independente e a probabilidade é dada por

$$\begin{aligned}\alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}^*) &= \min \left\{ \frac{\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{(i-1)})}{\pi(\boldsymbol{\theta}^{(i-1)})q(\boldsymbol{\theta}^*)}, 1 \right\} \\ &= \min \left\{ \frac{w(\boldsymbol{\theta}^*)}{w(\boldsymbol{\theta}^{(i-1)})}, 1 \right\}\end{aligned}$$

onde  $w(\boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}$ .

Quando a distribuição proposta é simétrica em torno da iteração anterior, isto é,  $q(\boldsymbol{\theta}^a|\boldsymbol{\theta}^b) = q(\boldsymbol{\theta}^b|\boldsymbol{\theta}^a)$ , e

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

Na prática, com modelos altamente dimensionais é muito difícil encontrar uma distribuição proposta adequada para tais espaços que garantam alta probabilidade de aceitação.

Para mais informações sob o algoritmo M-H ver Tierney (1994), Chib e Greenberg (1995), Gilks et al. (1996) e Gamerman e Lopes (2006).

### 2.3.3 O algoritmo Aceitação-Rejeição Metropolis-Hastings

Considere  $\pi(\boldsymbol{\theta})$  nossa distribuição alvo e seja  $g(\boldsymbol{\theta})$  a densidade proposta. Como vimos na sessão 2.2 a condição chave é que exista uma constante  $c$  tal que  $\pi(\boldsymbol{\theta}) \leq cg(\boldsymbol{\theta})$  para todo valor  $\boldsymbol{\theta}$  pertencente ao suporte da densidade  $\pi$ .

Seja  $\mathcal{D} = \{\boldsymbol{\theta} : \pi(\boldsymbol{\theta}) \leq cg(\boldsymbol{\theta})\}$  o conjunto de valores de  $\boldsymbol{\theta}$  que seguem a condição do algoritmo A-R.

O algoritmo Aceitação-Rejeição Metropolis-Hastings (ARMH) é um procedimento de amostragem MCMC no qual a condição dominante  $\pi(\boldsymbol{\theta}) \leq cg(\boldsymbol{\theta})$  não é satisfeita para algum  $\boldsymbol{\theta}$ , e assim  $g(\boldsymbol{\theta})$  é chamada de densidade pseudo-dominante. Neste caso, considere  $\mathcal{D}^c$  seja o complemento de  $\mathcal{D}$ , e suponha que corrente estado da cadeia é  $\boldsymbol{\theta}$ . Então o algoritmo ARMH é definido a seguir:

### Algoritmo

1. passo A-R: Gere uma amostra  $\boldsymbol{\theta}^*$  de  $g(\boldsymbol{\theta})$ ; aceite  $\boldsymbol{\theta}^*$  com probabilidade:

$$\alpha_{AR}(\boldsymbol{\theta}^*|\mathbf{y}) = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^*)} \right\}$$

Continue o processo até que amostra  $\boldsymbol{\theta}^*$  tenha sido aceita.

2. passo M-H: Dado o valor corrente  $\boldsymbol{\theta}$  e o valor proposto  $\boldsymbol{\theta}^*$ :

- (a) se  $\boldsymbol{\theta} \in \mathcal{D}$ , então  $\alpha_{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}^*|\mathbf{y}) = 1$ ;
- (b) se  $\boldsymbol{\theta} \in \mathcal{D}^c$  e  $\boldsymbol{\theta}^* \in \mathcal{D}$  então  $\alpha_{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}^*|\mathbf{y}) = \frac{cg(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})}$ ;
- (c) se  $\boldsymbol{\theta} \in \mathcal{D}^c$  e  $\boldsymbol{\theta}^* \in \mathcal{D}^c$ , então  $\alpha_{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}^*|\mathbf{y}) = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}^*)g(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta})g(\boldsymbol{\theta})} \right\}$

Aceite  $\boldsymbol{\theta}^*$  com probabilidade  $\alpha_{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}^*|\mathbf{y})$ . Caso contrário continue com  $\boldsymbol{\theta}$ .

Como discutido por Chib e Greenberg (1995), o algoritmo ARMH é reversível e, sobre condições de regularidade apropriadas, produz amostras da correta densidade  $\pi(\boldsymbol{\theta})$ . Chib e Greenberg (1995) também mostra que a amostra produzida no passo A-R tem uma densidade dada por

$$q(\boldsymbol{\theta}) = d^{-1} \alpha_{AR}(\boldsymbol{\theta}) g(\boldsymbol{\theta}),$$

a qual serve como a densidade proposta para o passo M-H. A constante de normalização  $d = \int \alpha_{AR}(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}$  não está disponível analiticamente, ao contrário do algoritmo M-H padrão onde a densidade proposta é completamente conhecida.

Para mais informações sobre o ARMH ver Chib e I. Jeliazkov (2004).

### 2.3.4 O amostrador de Gibbs

O amostrador de Gibbs é um algoritmo para gerar uma sequência de amostras de uma distribuição a posteriori  $\pi(\boldsymbol{\theta})$ , de duas ou mais variáveis aleatórias, desde que as distribuições condicionais completas, dos elementos de interesse, estejam disponíveis para amostragem. O amostrador de Gibbs é caso especial do algoritmo M-H e é aplicado quando a distribuição a posteriori não é explícita ou difícil de ser amostrada diretamente, mas a distribuição condicional de cada variável é conhecida e fácil de ser amostrada. O amostrador de Gibbs gera uma amostra a partir da distribuição de cada variável condicional sobre os valores atuais das outras variáveis.

Suponha que a distribuição de interesse é a distribuição  $\pi(\boldsymbol{\theta})$  onde  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)'$ . Considere também que todas as condicionais completas  $\pi_i(\theta_i | \boldsymbol{\theta}_{-i})$  estejam disponíveis, onde  $\boldsymbol{\theta}_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$ , e que se sabe gerar amostras de cada uma delas. Portanto, o esquema de amostragem é dado por:

#### Algoritmo de Gibbs

1. Fazer  $i = 0$ , inicializar  $\boldsymbol{\theta}^{(i)}$ ,
2. Dado  $\boldsymbol{\theta}^{(i-1)}$ , o próximo valor é obtido por simulação da forma:

$$\begin{aligned}\theta_1^{(i)} &\sim \pi_1\left(\theta_1 | \theta_2^{(i-1)}, \dots, \theta_d^{(i-1)}\right) \\ \theta_2^{(i)} &\sim \pi_2\left(\theta_2 | \theta_1^{(i)}, \theta_3^{(i-1)}, \dots, \theta_d^{(i-1)}\right) \\ &\vdots \\ \theta_d^{(i)} &\sim \pi_d\left(\theta_d | \theta_1^{(i)}, \dots, \theta_{d-1}^{(i)}\right)\end{aligned}$$

Note que o processo de atualização segue uma ordem fixa. Isto não é necessário.

A ordem pode ser aleatoriamente permutada a cada passo.

3. Incrementar  $i = i + 1$ , voltar ao passo 2 e continuar o procedimento até alcançar convergência.

O amostrador de Gibbs é simples e amplamente utilizado mas não é necessariamente o mais eficiente na solução de um problema. Assim, em casos onde o amostrador de Gibbs

não é a única aproximação possível, a simplicidade da implementação é uma vantagem que compensa sua ineficiência.

Para mais informações sobre o amostrador de Gibbs ver Gamerman e Lopes (2006) e Gelman et al. (2003).

### 2.3.5 Avaliação de Convergência

Os métodos MCMC são ótimas ferramentas para resolução de muitos problemas práticos na análise Bayesiana. Porém, algumas questões relacionadas à convergência nestes métodos ainda merecem bastante atenção.

Uma questão que pode surgir é “Quantas iterações deve ter o processo de simulação para garantir que a cadeia convergiu para o estado de equilíbrio?”. Como a cadeia não é inicializada na distribuição estacionária, uma prática comum é usar um período de aquecimento (Gilks et al., 1996). A cadeia é rodada por  $L + M$  iterações, sendo as  $L$  iterações iniciais descartadas. Espera-se que depois deste período de aquecimento a cadeia tenha “esquecido” os valores iniciais e convergido para a distribuição de equilíbrio; a amostra resultante de tamanho  $M$ , será uma amostra da distribuição de equilíbrio. Para eliminar uma possível auto correlação das cadeias após o aquecimento as amostras são salvas a cada  $k$  iterações, esse valor  $k$  chamamos de *lag*.

O grau de correlação da amostra final afetará a acurácia do estimador. O conceito de tempo de autocorrelação (Madras, 1998) é usado para quantificar este efeito. Assumindo que a cadeia tenha alcançado o equilíbrio, seja  $\boldsymbol{\theta}^t$  o valor da cadeia no tempo  $t$ . A autocorrelação,  $\rho_g(k)$ , no *lag*  $k$  para alguma função  $g(\boldsymbol{\theta})$  é definida por:

$$\rho_g(k) = \frac{E [g(\boldsymbol{\theta}^t) g(\boldsymbol{\theta}^{t+k})] - E [g(\boldsymbol{\theta}^t)]^2}{Var [g(\boldsymbol{\theta}^t)]}$$

A esperança acima é com relação à densidade  $\pi(\boldsymbol{\theta}^t)$ . O tempo de autocorrelação  $\tau_g$ , para a função  $g$  é definido como

$$\tau_g = \sum_{-\infty}^{\infty} \rho_g(k)$$

Se  $M \gg \tau_g$ , então a aproximação para a variância de

$$\frac{1}{M} \sum_{t=1}^M g(\boldsymbol{\theta}_t),$$

o estimador da esperança de  $g(\boldsymbol{\theta})$ , é  $V[g(\boldsymbol{\theta})] \tau_g / M$  o qual é maior que o estimador baseado em uma amostra, independente e identicamente distribuídas, de tamanho  $M$ . Em outras palavras, o número efetivo de amostras independentes em uma cadeia de tamanho  $M$  é aproximadamente  $M/\tau_g$ . Métodos para estimação  $\tau_g$  podem ser encontrados em Geyer (1992).

Observe que o tempo de autocorrelação é uma estimativa da eficiência da cadeia de Markov uma vez que alcança o equilíbrio, e portanto a estacionaridade, e não uma estimativa de quantas iterações são necessárias para a cadeia alcançar a distribuição estacionária.

Para avaliar a convergência dos métodos MCMC faz-se uso de alguns critérios que existem na literatura. As técnicas mais populares são as Geweke (1992) e Heidelberger e Welch (1983) que usam resultados baseados em análise espectral, Raftery e Lewis (1992) que calculam quantas iterações são necessárias para uma cadeia atingir a distribuição estacionária através da estimação de quantis a posteriori com uma precisão previamente fixada e Gelman e Rubin (1992) que utilizam resultados baseados na análise de variância clássica para duas ou mais cadeias simuladas com valores iniciais diferentes. Estes métodos e outros foram comparados no trabalho de Cowles e Carlin (1996), onde se chegou a conclusão de que não se pode afirmar qual deles é o mais eficiente.

## 2.4 Seleção de Modelos

### 2.4.1 Introdução

A tarefa de selecionar um modelo estatístico de um conjunto de modelos possíveis é uma questão crucial em vários campos de estudo.

Comparação de modelos Bayesianos não vai informar qual é o modelo “verdadeiro”, mas sim sobre a preferência para o modelo com base nos dados e outras informações. O Critério de Informação Preditiva Bayesiana (BPIC) proposto por Ando (2007) é uma ferramenta útil para calcular a qualidade do modelo Bayesiano de um ponto de vista preditivo. O BPIC é derivado corrigindo o viés assintótico na media a posteriori da log-verossimilhança. Esta sessão explica a estrutura do BPIC, o qual iremos utilizar no capítulo 7.

### 2.4.2 Estrutura Geral

Suponhamos que nós estamos interessados em selecionar um modelo de um conjunto de modelos possíveis  $M_1, \dots, M_r$ . É suposto que cada modelo  $M_k$  é caracterizado pela densidade de probabilidade  $f_k(\mathbf{y}|\theta_k)$ , onde  $\theta_k$  é um vetor real de parâmetros desconhecidos de dimensão  $p_k$ . Considere  $\pi_k(\theta_k)$  seja a distribuição a priori para o vetor de parâmetros  $\theta_k$  sobre o modelo  $M_k$ . A probabilidade a posteriori do modelo  $M_k$  para um particular conjunto de dados  $\mathbf{Y}_n = \{y_1, \dots, y_n\}$  é dado por:

$$P(M_k|\mathbf{Y}_n) = \frac{P(M_k) \int f_k(\mathbf{Y}_n|\theta_k) \pi_k(\theta_k) d\theta_k}{\sum_{\alpha=1}^r P(M_\alpha) \int f_\alpha(\mathbf{Y}_n|\theta_\alpha) \pi_\alpha(\theta_\alpha) d\theta_\alpha}, \quad (2.4)$$

onde  $f_k(\mathbf{Y}_n|\theta_k)$  e  $P(M_k)$  são a verossimilhança e a probabilidade a priori para o modelo  $M_k$ . A probabilidade a priori  $P(M_k)$  e  $\pi_k(\theta_k|M_k)$  para o modelo  $M_k$  especificam uma visão inicial da incerteza do modelo. Tendo observado a informação  $\mathbf{Y}_n$ , nós então atualizamos o ponto de vista da incerteza modelo baseado na probabilidade a posteriori do modelo  $P(M_k|\theta_n)$ .

Em princípio, a abordagem Bayesiana para seleção de um modelo é escolher o modelo com a maior probabilidade a posteriori entre um conjunto de modelos possíveis. Portanto, a probabilidade a posteriori do modelo  $P(M_1), \dots, P(M_r)$  é de fundamental interesse para seleção do modelo. Isto é equivalente a escolher o modelo que maximiza

$$P(M_k) \int f_k(\mathbf{Y}_n | \boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k. \quad (2.5)$$

A quantidade

$$P(\mathbf{Y}_n | M_k) = \int f_k(\mathbf{Y}_n | \boldsymbol{\theta}_k) \pi_k(\boldsymbol{\theta}_k) d\boldsymbol{\theta}_k \quad (2.6)$$

obtida pela integração sobre o espaço paramétrico  $\boldsymbol{\Theta}_k$  é a probabilidade marginal dos dados  $\mathbf{Y}_n$  sobre o modelo  $M_K$ . Esta quantidade mede quão bem a especificação da distribuição a priori ajusta aos dados observados.

A escolha da probabilidade a priori para um modelo é muita vezes feita baseado nas características do modelo tal como a complexidade do modelo.

### 2.4.3 Critério de Informação Preditiva Bayesiana

O Critério de Informação Preditiva Bayesiana (BPIC) proposto por Ando (2007) é uma ferramenta útil para avaliar a qualidade de modelos Bayesianos de um ponto de vista preditivo. O BPIC é um estimador para a média a posteriori da esperada log-verossimilhança da distribuição preditiva Bayesiana e é obtido corrigindo o viés assintótico na média a posterior da log-verossimilhança.

O BPIC é definido como:

$$BPIC = -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:n}} [\log\{p(\mathbf{y}_{1:n}|\boldsymbol{\theta})\}] + 2n\hat{b} \quad (2.7)$$

onde o viés assintótico corrigido,  $\hat{b}$ , é dado por:

$$\hat{b} \approx \frac{1}{n} \left\{ E_{\boldsymbol{\theta}|\mathbf{y}_{1:n}} [\log\{p(\mathbf{y}_{1:n}|\boldsymbol{\theta})\}p(\boldsymbol{\theta})] - \log\{p(\mathbf{y}_{1:n}|\hat{\boldsymbol{\theta}})\}p(\hat{\boldsymbol{\theta}}) + tr \left\{ J_n^{-1}(\hat{\boldsymbol{\theta}}) I_n(\hat{\boldsymbol{\theta}}) \right\} + 0.5q \right\} \quad (2.8)$$

tal que  $q$  é a dimensão de  $\boldsymbol{\theta}$ ,  $E_{\boldsymbol{\theta}|\mathbf{y}_{1:n}}[\cdot]$  denota a esperança com respeito a distribuição a posteriori,  $\hat{\boldsymbol{\theta}}$  é a moda a posteriori e

$$I_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{t=1}^n \left( \frac{\partial \eta_n(y_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \eta_n(y_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (2.9)$$

$$J_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{t=1}^n \left( \frac{\partial^2 \eta_n(y_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad (2.10)$$

com  $\eta_n(y_t, \boldsymbol{\theta}) = \log \{p(y_t|\mathbf{y}_{1:n-1}, \boldsymbol{\theta})\} + \log \{p(\boldsymbol{\theta})/n\}$ .

O BPIC tem sido aplicado em diversos problemas de seleção de modelos estatísticos Bayesianos como, por exemplo, para prever o risco da estrutura a termo, Ando (2009c) utilizou o BPIC para avaliar inferências Bayesianas com preditores funcionais. Ando (2009b) utilizou o BPIC para estudar a seleção de um portfolio Bayesiano sobre um modelo multifatorial.

O melhor modelo será que tiver o menor BPIC.

#### 2.4.4 Fator de Bayes

Definição:

O Fator de Bayes, uma quantidade para comparação de modelos e teste de hipótese em estrutura Bayesiana, tem desempenhado um papel importante em bondade de ajuste de modelos concorrentes. Ele permite considerar uma comparação em pares de modelos, seja  $M_k$  e  $M_j$  baseados na probabilidade a posteriori  $P(M_k|\mathbf{Y}_n)$  visto em (2.4). O fator de Bayes é definido como:

$$\text{Fator de Bayes}(M_k, M_j) \equiv \frac{P(\mathbf{Y}_n|M_k)}{P(\mathbf{Y}_n|M_j)}, \quad (2.11)$$

o qual mede preferência entre os modelos  $M_k$  e  $M_j$  baseado na informação dos dados. O Fator de Bayes escolhe o modelo com maior valor de verossimilhança marginal em um conjunto de modelos propostos.

Note que

$$\frac{P(M_k|\mathbf{Y}_n)}{P(M_j|\mathbf{Y}_n)} = \frac{P(\mathbf{Y}_n|M_k)}{P(\mathbf{Y}_n|M_j)} \times \frac{P(M_k)}{P(M_j)}, \quad (2.12)$$



o fator de Bayes também é dado como a razão da razão da chance a posteriori e razão de chance a priori.

$$\text{Fator de Bayes}(M_k, M_j) = \frac{\frac{P(M_k|\mathbf{Y}_n)}{P(M_j|\mathbf{Y}_n)}}{\frac{P(M_k)}{P(M_j)}}.$$

Quando as probabilidades a priori  $P(M_k)$  e  $P(M_j)$  são iguais, o fator de Bayes se reduz a razão de chances a posteriori de  $M_k$  e  $M_j$ .

Jeffreys (1961) recomendou interpretar o Fator de Bayes como uma escala de evidência.

Tabela 2.1: Escala de Jeffreys para o Fator de Bayes

Fator de Bayes	Interpretação
$\beta_{kj} < 1$	Evidência negativa para $M_k$
$1 < \beta_{kj} < 3$	Baixa evidência para $M_k$
$3 < \beta < 10$	Substancial evidência para $M_k$
$10 < \beta_{kj} < 30$	Forte evidência para $M_k$
$30 < \beta_{kj} < 100$	Evidencia muito forte para $M_k$
$100 < B_{kj}$	Evidência decisiva para $M_k$

Kass and Raftery (1995) também deram orientações para a interpretação do Fator de Bayes.

Para mais informações sobre seleção de modelos ver Ando (2010)

# Capítulo 3

## Modelos Dinâmicos

### 3.1 Introdução

Neste capítulo introduziremos algumas noções sobre os Modelos Dinâmicos, uma ampla classe de modelos com parâmetros variando no tempo, que vêm sendo utilizados para a análise e previsão de séries temporais e processos espaço-temporais em diversas áreas. Avanços em computação juntamente com os métodos MCMC aumentarão muito a utilização dessa classe de modelos nos últimos anos.

### 3.2 Modelo Linear Dinâmico (MLD)

**Definição 3.1.** Para cada  $t$ , o MLD geral, representado pela quadrupla  $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$ , é definido por:

$$\text{Equação observacional:} \quad \mathbf{Y}_t = \mathbf{c}_t + \mathbf{F}_t \boldsymbol{\theta}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{\nu} \sim N_q(\mathbf{0}, \mathbf{V}_t) \quad (3.1)$$

$$\text{Equação do sistema:} \quad \boldsymbol{\theta}_t = \mathbf{d}_t + \mathbf{G}_{t-1} \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N_p(\mathbf{0}, \mathbf{W}_t) \quad (3.2)$$

$$\text{Informação inicial:} \quad (\boldsymbol{\theta}_1 | \mathbf{D}_0) \sim N_p(\mathbf{m}_0, \mathbf{C}_0) \quad (3.3)$$

onde  $\mathbf{y}_t$  é o vetor de observações e  $\boldsymbol{\theta}_t$  é o vetor de estados não observáveis.  $\mathbf{m}_0$  e  $\mathbf{C}_0$  são a media e a variância da distribuição inicial respectivamente e são supostos

conhecidos. Assume-se que os erros observacionais,  $\{\boldsymbol{\nu}_t\}$ , e os erros da evolução,  $\{\boldsymbol{\omega}_t\}$ , são mutuamente independentes entre si e da distribuição inicial.

Os vetores  $\mathbf{c}_t$  e  $\mathbf{d}_t$  e as matrizes  $\mathbf{F}_t$ ,  $\mathbf{G}_t$ ,  $\mathbf{V}_t$  e  $\mathbf{W}_t$  são consideradas conhecidas para cada  $t$ . A inclusão de  $\mathbf{c}_t$  e  $\mathbf{d}_t$  na definição 3.1 foi motivada pela classe de modelos de volatilidade estocástica e suas extensões que serão apresentadas nos capítulo 5.

Esta definição do MLD foi introduzida por Harrison e Stevens (1976) no contexto bayesiano. Um amplo tratamento usando a abordagem Bayesiana pode ser encontrada em West e Harrison (1997).

### 3.2.1 Filtro de Kalman

O Filtro de Kalman é um algoritmo que permite atualizar de forma recursiva a distribuição dos estados quando uma nova observação está disponível.

**Teorema 3.1.** *No MLD da definição 3.1, as distribuições um passo a frente e a posteriori, para cada  $t$ , são dadas por*

*Evolução:*

$$\boldsymbol{\theta}_t | \mathbf{D}_{t-1} \sim N_p(\mathbf{a}_t, \mathbf{R}_t)$$

onde

$$\mathbf{a}_t = \mathbf{d}_t + \mathbf{G}_t \mathbf{m}_{t-1} \tag{3.4}$$

$$\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}'_t + \mathbf{W}_t \tag{3.5}$$

*Previsão:*

$$\mathbf{y}_t | \mathbf{D}_{t-1} \sim N_q(\mathbf{f}_t, \mathbf{Q}_t)$$

onde

$$\mathbf{f}_t = \mathbf{c}_t + \mathbf{F}'_t \mathbf{a}_t \tag{3.6}$$

$$\mathbf{Q}_t = \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t \tag{3.7}$$

Atualização:

$$\boldsymbol{\theta}_t | \mathbf{D}_t \sim N_p(\mathbf{m}_t, \mathbf{C}_t)$$

onde

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t \text{ tal que } \mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}^{-1} \text{ e } \mathbf{e}_t = \mathbf{y}_t - \mathbf{f}_t \quad (3.8)$$

$$\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t' \quad (3.9)$$

A matriz adaptativa  $\mathbf{A}_t$ , também conhecida como ganho de Kalman,

**Prova:**(West e Harrison, 1997) Partindo da distribuição de  $(\boldsymbol{\theta}_{t-1} | \mathbf{D}_{t-1})$  realizamos a evolução a partir da equação 3.2 assim obtendo  $(\boldsymbol{\theta}_t | \mathbf{D}_{t-1})$ . Utilizando  $(\boldsymbol{\theta}_t | \mathbf{D}_{t-1})$  conjuntamente com a equação observacional (3.1) obtemos facilmente  $N_q(\mathbf{f}_t, \mathbf{Q}_t)$ . A distribuição de  $(\boldsymbol{\theta}_t | \mathbf{D}_t)$  é obtida através do condicionamento da distribuição conjunta normal de  $(\boldsymbol{\theta}_t, \mathbf{y}_t)$  em relação a  $\mathbf{y}_t$ .

### 3.2.2 Suavizador de Kalman

A distribuição do vetor de estados  $\boldsymbol{\theta}_t$ , utilizando toda a informação disponível,  $\mathbf{D}_t$ , é chamada de distribuição suavizada. O algoritmo que permite obter estas distribuições para todo  $t$  é chamado suavizador de Kalman.

**Teorema 3.2.** *No MLD da definição 3.1, defini-se*

$$\mathbf{B}_t = \mathbf{C}_t \mathbf{G}_{t+1} \mathbf{R}_{t+1}^{-1}$$

para todo  $t$ . Então a distribuição da definição suavizada de  $\boldsymbol{\theta}_t$  dado  $\mathbf{D}_T$  é dada por

$$\boldsymbol{\theta}_t | \mathbf{D}_t \sim N(\tilde{\mathbf{m}}_t, \tilde{\mathbf{C}}_t)$$

onde

$$\tilde{\mathbf{m}}_t = \mathbf{m}_t + \mathbf{B}_t(\tilde{\mathbf{m}}_{t+1} - \mathbf{a}_t)$$

e

$$\tilde{\mathbf{C}}_t = \mathbf{C}_t + \mathbf{B}_t(\tilde{\mathbf{C}}_{t+1} - \mathbf{R}_{t+1})\mathbf{B}'_t$$

onde  $\tilde{\mathbf{m}}_T = \mathbf{m}_T$  e  $\tilde{\mathbf{C}}_T = \mathbf{C}_T$

**Prova:**(West e Harrison, 1997) A dedução das formulas recursivas do algoritmo de suavização de Kalman é baseada no princípio de indução para trás em t. Logo no tempo t=T tem-se que

$$\boldsymbol{\theta}_T | \mathbf{D}_T \sim N(\tilde{\mathbf{m}}_T, \tilde{\mathbf{C}}_T) \quad (3.10)$$

tal que  $\tilde{\mathbf{m}}_T = \mathbf{m}_T$  e  $\tilde{\mathbf{C}}_T = \mathbf{C}_T$ . Pela hipótese indutiva tem-se que  $\boldsymbol{\theta}_{t+1} | \mathbf{D}_T \sim N(\tilde{\mathbf{m}}_{t+1}, \tilde{\mathbf{C}}_{t+1})$ .

Logo temos que provar que  $\boldsymbol{\theta}_t | \mathbf{D}_T \sim N(\tilde{\mathbf{m}}_t, \tilde{\mathbf{C}}_t)$ . Do Filtro de Kalman tem-se

$$\boldsymbol{\theta}_{t+1} | \mathbf{D}_t \sim N_p(\mathbf{a}_{t+1}, \mathbf{R}_{t+1})$$

$$\boldsymbol{\theta}_t | \mathbf{D}_t \sim N_p(\mathbf{m}_t, \mathbf{C}_t)$$

as quais determinam a distribuição condicional conjunta de  $\boldsymbol{\theta}_{t+1}$  e  $\boldsymbol{\theta}_t$  dado  $\mathbf{D}_t$ . A covariância  $\mathbb{C}(\boldsymbol{\theta}_{t+1}, \boldsymbol{\theta}_t | \mathbf{D}_t) = \mathbf{C}_t \mathbf{G}'_{t+1}$  é obtida a partir de (3.2). Logo, da distribuição normal multivariada tem-se que

$$(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, \mathbf{D}_t) \sim N_p(\mathbf{m}_t + \mathbf{B}_t(\boldsymbol{\theta}_{t+1} - \mathbf{a}_{t+1}), \mathbf{C}_t - \mathbf{B}_t \mathbf{R}_{t+1} \mathbf{B}'_t) \quad (3.11)$$

Finalmente, usando a hipótese de indução, obtém-se

$$\begin{aligned} \tilde{\mathbf{m}}_t &= E[E[\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, \mathbf{D}_T] | \mathbf{D}_T] \\ &= \mathbf{m}_t + \mathbf{B}_t(\tilde{\mathbf{m}}_{t+1} - \mathbf{a}_{t+1}) \end{aligned} \quad (3.12)$$

$$\begin{aligned} \tilde{\mathbf{C}}_t &= E[V[\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}] | \mathbf{D}_t] + V[E[\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, \mathbf{D}_T] | \mathbf{D}_T] \\ &= \mathbf{C}_t + \mathbf{B}_t(\tilde{\mathbf{C}}_{t+1} - \mathbf{R}_{t+1})\mathbf{B}'_t. \end{aligned} \quad (3.13)$$

Desta maneira, o algoritmo suavizador de Kalman é processado depois do Filtro de Kalman e inicializado pela equação (3.10). As demais recursões seguem de (3.12) e (3.13).

### 3.2.3 Filtro de Pertubações

O filtro de perturbações é matematicamente equivalente ao filtro de Kalman. A razão para esta terminologia é que a saída do filtro de perturbações fornece os erros de previsão

um passo a frente  $\mathbf{e}_t$ , a inversa das variâncias destes erros,  $\mathbf{Q}_t^{-1}$ , e as matrizes adaptativas escaladas  $\mathbf{K}_t$ .

As equações do filtro de perturbações são obtidas a partir do filtro de Kalman e o algoritmo é inicializado usando  $\mathbf{m}_0$  e  $\mathbf{C}_0$ , os parâmetros de  $[\boldsymbol{\theta}_0|\mathbf{D}_0]$ . No filtro de Kalman temos que  $\mathbf{a}_1 = \mathbf{d}_1 + \mathbf{G}_1\mathbf{m}_0$  e  $\mathbf{R}_1 = \mathbf{G}\mathbf{G}_0\mathbf{G}'_1 + \mathbf{W}_1$ . Então para  $t = 1, \dots, T$  temos a seguinte recursão:

$$\begin{aligned}\mathbf{e}_t &= \mathbf{y}_t - \mathbf{c}_t - \mathbf{F}'_t\mathbf{a}_t \\ \mathbf{Q}_t &= \mathbf{F}'_t\mathbf{R}_t\mathbf{F}_t + \mathbf{V}_t \\ \mathbf{K}_t &= \mathbf{G}_{t+1}\mathbf{F}_t\mathbf{R}_t\mathbf{Q}_t^{-1} \\ \mathbf{a}_{t+1} &= \mathbf{d}_{t+1} + \mathbf{G}_{t+1}\mathbf{a}_t + \mathbf{K}_t\mathbf{e}_t \\ \mathbf{R}_{t+1} &= \mathbf{G}_{t+1}\mathbf{R}_t(\mathbf{G}_{t+1} - \mathbf{K}_t\mathbf{F}'_t)' + \mathbf{W}_{t+1},\end{aligned}$$

onde  $\mathbf{K}_t = \mathbf{G}_{t+1}\mathbf{A}_t$  é a conexão com o coeficiente adaptativo definido em (3.8). Deste modo o filtro de perturbações é equivalente ao filtro de Kalman, exceto pelas saídas. Somente  $\mathbf{K}_t$ ,  $\mathbf{Q}_t^{-1}$  e  $\mathbf{e}_t$  são armazenados, enquanto que o Filtro de Kalman armazena adicionalmente em  $\mathbf{m}_t$  e  $\mathbf{C}_t$ .

### 3.2.4 Perturbações Suavizadas

Como no caso da filtragem, é possível obter um ganho computacional usando perturbações ao invés dos estados, isto é, determinam-se  $\hat{\boldsymbol{\nu}} = E[\boldsymbol{\nu}|\mathbf{D}_T]$  e  $\tilde{\boldsymbol{\omega}} = E[\boldsymbol{\omega}|\mathbf{D}_T]$  em um lugar de  $\tilde{\mathbf{m}} = E[\boldsymbol{\theta}|\mathbf{D}_T]$ . O algoritmo de perturbações suavizadas introduzido por Koopman (1993), fornece o estimador do erro quadrático médio do vetor de perturbações dadas todas as observações, a partir das saídas do filtro de perturbações,  $\mathbf{K}_t$ ,  $\mathbf{Q}_t^{-1}$  e  $\mathbf{e}_t$ .

Como na derivação de suavizador de Kalman, a obtenção das perturbações suavizadas é uma recursão para trás no tempo. Seja  $\mathbf{r}_t$  o vetor  $p$ -dimensional para as perturbações dos estados e  $\varepsilon_t$  o vetor  $q$ -dimensional de perturbações das observações. A recursão é

inicializada com  $r_T = 0$ . Então para  $t = T, \dots, 1$  tem-se a recursão para trás

$$\boldsymbol{\varepsilon}_t = \mathbf{Q}_t^{-1} \mathbf{e}_t - \mathbf{K}_t \mathbf{r}_t \quad (3.14)$$

$$\tilde{\boldsymbol{\nu}}_t = \mathbf{V}_t \boldsymbol{\varepsilon}_t \quad (3.15)$$

$$\tilde{\boldsymbol{\omega}}_t = \mathbf{W}_t \mathbf{r}_{t-1} \quad (3.16)$$

$$\mathbf{r}_t = \mathbf{F}_t \boldsymbol{\varepsilon}_t + \mathbf{G}'_{t+1} \mathbf{r}_t \quad (3.17)$$

observa-se que comparadas ao suavizador de Kalman, não é necessário inversão de matrizes, já que  $\mathbf{Q}_t$  é invertida no filtro de perturbações. Então os valores suavizados  $\tilde{\mathbf{m}}_t$  são obtidos pela recursão

$$\tilde{\mathbf{m}}_t = \mathbf{G}_t \tilde{\mathbf{m}}_{t-1} + \tilde{\boldsymbol{\omega}}_t, \quad t = 1, \dots, T$$

### 3.2.5 Algoritmo de de Jong e Shephard

Esta técnica é chamada de algoritmo de *simulação de perturbações suavizadas*.

A recursão é iniciada fazendo  $\mathbf{r}_T^* = \mathbf{0}$ ,  $\mathbf{N}_T = \mathbf{0}$ . Logo para  $t = T, \dots, 1$

$$\boldsymbol{\Omega} = \mathbf{W}_t - \mathbf{W}_t \mathbf{N}_t \mathbf{W}_t$$

$$\boldsymbol{\nu} \sim N_p(\mathbf{0}, \boldsymbol{\Omega}_t)$$

$$\boldsymbol{\Upsilon}_t = \mathbf{W}_t \mathbf{N}_t \overbrace{\{\mathbf{G}_{t+1} - \mathbf{K}_t \mathbf{F}'_t\}}^{\boldsymbol{\Lambda}_t}$$

$$\mathbf{r}_{t-1}^* = \mathbf{F}_t \boldsymbol{\varepsilon}_t \mathbf{G}'_{t+1} \mathbf{r}_t^* - \boldsymbol{\varepsilon}_t \boldsymbol{\Omega}_t^{-1} \mathbf{v}_t$$

$$\mathbf{N}_{t-1} = \mathbf{F}_t \mathbf{Q}_t \mathbf{F}'_t + \boldsymbol{\Lambda}_t \mathbf{N}_t \boldsymbol{\Lambda}_t + \boldsymbol{\Upsilon}_t \boldsymbol{\Omega}_t^{-1} \boldsymbol{\Upsilon}_t$$

onde  $\boldsymbol{\varepsilon}_t$  vem de (3.14). Note que se  $\mathbf{v}_t = \mathbf{0}$  e  $\boldsymbol{\Upsilon}_t = \mathbf{0}$ , então  $\mathbf{r}_t^* = \mathbf{r}_t$ , o algoritmo de de Jong e Shephard se reduz ao algoritmo de perturbações suavizadas.

Uma amostra  $\boldsymbol{\omega}_t^{(i)}$  de  $p(\boldsymbol{\omega}_t | \mathbf{y}_{1:T})$  é dada por

$$\boldsymbol{\omega}_t^{(i)} = \mathbf{W}_t \mathbf{r}_{t-1}^* + \mathbf{v}_t \quad (3.18)$$

e desta

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1}^{(i)} + \boldsymbol{\omega}_t^{(i)} \quad (3.19)$$

Para obter uma amostra *i.i.d* de  $p(\boldsymbol{\theta}_{0:T} | \boldsymbol{\psi}, \mathbf{y}_{1:T})$ , repete-se o procedimento.

# Capítulo 4

## Distribuição Hiperbólica Generalizada t-Student Assimétrica

### 4.1 Introdução

Este capítulo introduz conceitos sobre a classe distribuições Hiperbólicas Generalizadas (HG) dando uma atenção a distribuição Hiperbólica Generalizada t-Student assimétrica (HGt) que será utilizada no capítulo 5.

A distribuição HG foram introduzidas por Barndorff-Nielsen (1977), o nome das distribuições é derivado do fato que a log-densidade possui a forma de uma hipérbole, enquanto a log-densidade da distribuição normal é uma parábola. O que tornou as distribuições hiperbólicas tão populares foi o fato dos seus parâmetros serem suficientemente flexíveis para se adequarem aos mais variados conjuntos de dados e contextos.

As distribuições HG tem recentemente atraído a atenção em econometria financeira como uma promissora abordagem para uma modelagem flexível da assimetria e caudas pesadas. Essa família contempla uma ampla classe de distribuições paramétricas tais como a Gaussiana Normal-Inversa (NIG) e a t-Student assimétrica e adicionado a isso essa classe possui boas propriedades como, por exemplo, são fechadas em relação a condicionamento, marginalização e transformações afim. Podem ser ambas simétrica ou assimétrica, e as caudas geralmente são semi-pesadas.

O caso particular das distribuições HG, a distribuição HG t-Student assimétrica, que



será discutida neste capítulo tem a importante propriedade que uma das caudas possui comportamento polinomial e a outra exponencial. Isto faz ela ser diferente das outras distribuições t-Student assimétrica propostas na literatura que tem duas caudas pesadas. Sabe-se que retornos de variáveis de mercado financeiro tais como taxa de câmbio, preços de ações, e taxas de juros, medidas sobre um curto período de tempo, isto é, diariamente ou semanalmente, são caracterizadas pela falta de normalidade sendo que as distribuições empíricas de tais retornos possuem caudas mais pesadas que a distribuição normal e além disso, muitas vezes, há assimetria, tendo uma cauda pesada e a outra semi-pesada ou provavelmente gaussiana.

## 4.2 Distribuição Hiperbólica Generalizada

A distribuição HG pode ser parametrizada de diversas maneiras, aqui mostramos a representação feita por Prause (1999):

$$f(x) = \frac{(\alpha^2 - \delta^2)^{\lambda/2} K_{\lambda-1/2} \left( \alpha \sqrt{\vartheta^2 + (x - \mu)^2} \right) \exp(\delta(x - \mu))}{\sqrt{2\pi} \alpha^{\lambda-1/2} \vartheta^\lambda K_\lambda \left( \vartheta \sqrt{\alpha^2 - \delta^2} \right) \left( \sqrt{\vartheta^2 + (x - \mu)^2} \right)^{1/2-\lambda}} \quad (4.1)$$

onde  $K_j$  é a função de Bessel modificada do terceiro tipo de ordem  $j$  (Abramowitz e Stegun, 1972):

$$K_j(x) = \frac{1}{2} \int_0^\infty y^{j-1} \exp\left(-\frac{x}{2}(y + y^{-1})\right) dy.$$

onde  $K_j$  é estritamente positiva e decrescente em relação a  $x$ .

Além de  $\mu \in \mathbb{R}$ , os valores dos parâmetros devem satisfazer as condições:

$$\vartheta \geq 0, |\delta| < \alpha \text{ se } \lambda > 0 \quad (4.2)$$

$$\vartheta > 0, |\delta| < \alpha \text{ se } \lambda = 0 \quad (4.3)$$

$$\vartheta > 0, |\delta| \leq \alpha \text{ se } \lambda < 0 \quad (4.4)$$

A distribuição HG comporta-se na cauda como

$$f(x) \sim \text{const}|x|^{\lambda-1} \exp(-\alpha|x| + \delta x) \text{ com } x \rightarrow \pm\infty, \quad (4.5)$$

para todos os valores de  $\lambda$ . Por isso, desde que  $|\delta| \neq \alpha$ , a distribuição HG possui duas caudas semi-pesadas.

### 4.2.1 Parâmetros da distribuição HG

Nesta seção vamos estudar o efeito de cada parâmetro da distribuição HG baseado na função de densidade definida em (4.1). Boas referências para esse estudo são Wang (2005) e Vitor (2006).

Os parâmetros da distribuição HG, grosseiramente falando, possuem a seguinte interpretação:  $\mu$  indica a locação,  $\vartheta$  representa escala,  $\delta$  determina a assimetria e  $\alpha$  a forma.  $\lambda$  determina o peso das caudas assim como as subclasses da distribuições HG.

A seguir serão apresentados gráficos dos efeitos de cada parâmetro da distribuição HG individualmente dado demais parâmetros constantes.

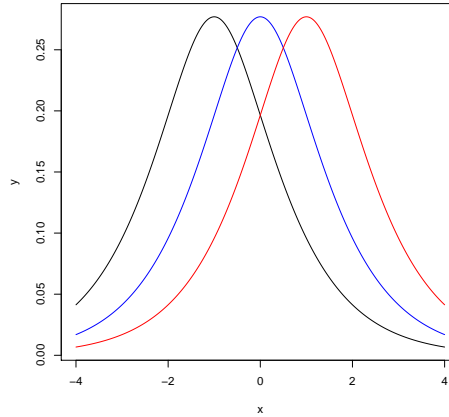


Figura 4.1: Três curvas da distribuição HG onde  $\lambda = 1.3$ ,  $\alpha = 1$ ,  $\delta = 0$ ,  $\vartheta = 1$  e  $\mu = -1$ (preto),  $\mu = 0$ (azul) e  $\mu = 1$ (vermelho).

A Figura 4.1 mostra claramente o efeito do parâmetro  $\mu$ , tal que um aumento no valor de  $\mu$  desloca as curvas da distribuição HG horizontalmente para a direita.

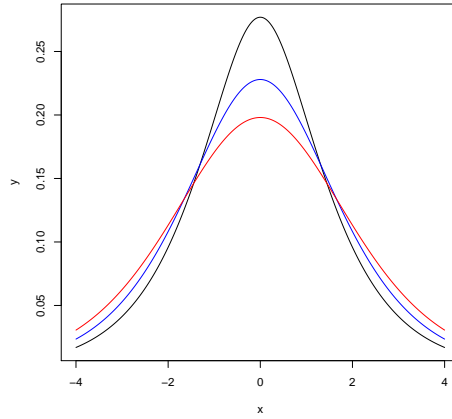


Figura 4.2: Três curvas da distribuição HG onde  $\lambda = 1.3$ ,  $\alpha = 1$ ,  $\delta = 0$ ,  $\mu = 0$  e  $\vartheta = 1$ (preto),  $\vartheta = 2$ (azul) e  $\vartheta = 3$ (vermelho).

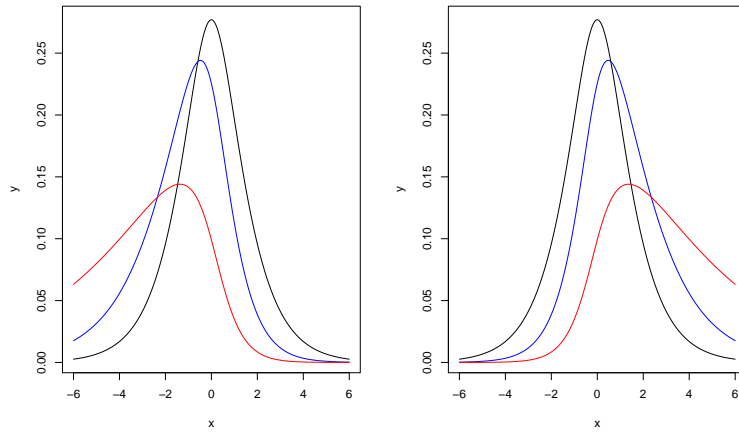


Figura 4.3: Três curvas da distribuição HG onde  $\lambda = 1.3$ ,  $\alpha = 1$ ,  $\mu = 0$ ,  $\vartheta = 1$  e à esquerda tem-se  $\delta = 0$ (preto),  $\delta = -0.35$ (azul) e  $\delta = -0.7$ (vermelho) e à direita tem-se  $\delta = 0$ (preto),  $\delta = 0.35$ (azul) e  $\delta = 0.7$ .

Na Figura 4.2 pode ser visto que um incremento no valor de  $\vartheta$  torna o gráfico mais achatado e Figura 4.3 observa-se que para  $\delta < 0$  tem-se assimetria à esquerda e para  $\delta > 0$  tem-se assimetria à direita e quando  $\delta = 0$  a densidade é simétrica.

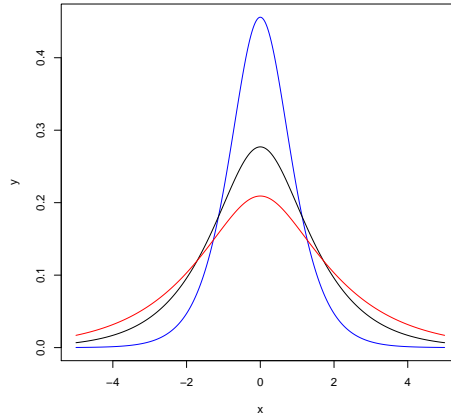


Figura 4.4: Três curvas da distribuição HG onde  $\lambda = 1.3$ ,  $\delta = 0$ ,  $\vartheta = 1$  e  $\alpha = 1$ (preto),  $\alpha = 2$ (azul) e  $\alpha = 0.7$ (vermelho).

A Figura 4.4 mostra que há um achatamento maior da função a medida que diminui-se o valor do parâmetro  $\alpha$ . Como foi visto anteriormente o valor do parâmetros  $\vartheta$  também influência no achatamento da função assim uma análise interessante é avaliar o efeito conjunto dos parâmetros  $\alpha$  e  $\vartheta$ .

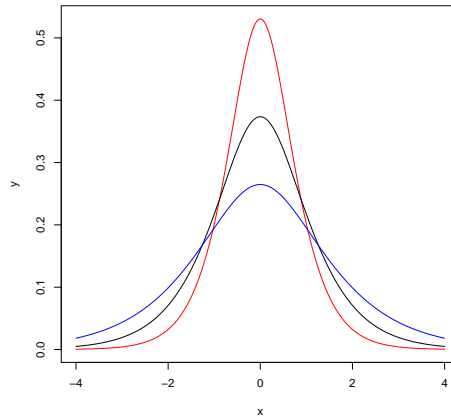


Figura 4.5: Três curvas da distribuição HG onde  $\lambda = 1.3$ ,  $\delta = 0$ , e  $\alpha = 1.5$ ,  $\vartheta = 1$ (preto),  $\alpha = 1$ ,  $\vartheta = 1.2$ (azul) e  $\alpha = 0.8$ ,  $\vartheta = 2.25$ (vermelho).

Os valores de  $\alpha\vartheta$  são 1.2, 1.5 e 1.8 e é fácil ver que um aumento no valor de  $\alpha\vartheta$  diminui o achatamento do gráfico.

### 4.3 Distribuição Hiperbólica Generalizada t-Student Assimétrica

A distribuição HG t-Student assimétrica é um caso particular das distribuições HG e ela pode ser representada como uma mistura variância-média normal com a distribuição gaussiana inversa generalizada (GIG), onde a distribuição GIG tem densidade dada por

$$f(z; \lambda, \vartheta, \gamma) = \left(\frac{\gamma}{\vartheta}\right)^\lambda \frac{z^{\lambda-1}}{2K_\lambda(\gamma\vartheta)} \exp\left\{-\frac{1}{2}(\vartheta^2 z^{-1} + \gamma^2 z)\right\} \quad (4.6)$$

Isso significa que uma variável HG  $X$ , pode ser representada estocasticamente por:

$$X = \mu + \delta Z + \sqrt{Z}Y, \quad (4.7)$$

onde  $Y \sim N(0, 1)$ ,  $Z \sim GIG(\lambda, \vartheta, \gamma)$ , com  $Y$  e  $Z$  independentes e  $\gamma = \sqrt{\alpha^2 - \delta^2}$ .

Segue da equação (4.7) que  $X|Z = z \sim N(\mu + \delta z, z)$ .

Considerando  $\lambda = -\frac{\nu}{2}$  e  $\alpha \rightarrow |\delta|$  na equação (4.1), obtemos a distribuição HG t-Student assimétrica cuja densidade é dada por

$$f(x) = \frac{2^{\frac{1-\nu}{2}} \vartheta^\nu |\delta|^{\frac{\nu+1}{2}} K_{\frac{\nu+1}{2}} \left( \sqrt{\delta^2(\vartheta^2 + (x - \mu)^2)} \right) \exp(\delta(x - \mu))}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi} \left( \sqrt{\vartheta^2 + (x - \mu)^2} \right)^{\frac{\nu+1}{2}}} \text{ se } \delta \neq 0 \quad (4.8)$$

e

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\vartheta\Gamma\left(\frac{\nu}{2}\right)} \left[ 1 + \frac{(x - \mu)^2}{\vartheta^2} \right]^{-(\nu+1)/2} \text{ se } \delta = 0 \quad (4.9)$$

A densidade  $f(x)$  na equação (4.8) é a distribuição t-Student não-central com  $\nu$  graus de liberdade, esperança  $\mu$  e variância  $\vartheta^2/(\nu - 2)$ .

A Figura 4.3 mostra a densidade da HG t-Student para diversas combinações dos valores de  $\delta$  e  $\nu$  para interpretação em relação a assimetria e as caudas pesadas. Na Figura 4.3(i), as densidades avaliadas para  $\delta = 0, -1$  e  $-2$  utilizando  $\nu$  fixo igual a 10. Como mencionado,  $\delta = 0$  corresponde a densidade a t-Student simétrica. Um pequeno valor de  $\delta$  implica em uma assimetria mais negativa ou assimetria a esquerda como também caudas pesadas. Na Figura 4.3(ii) mostra a densidades para os valores de  $\nu$

igual a 5, 10 e 15 com  $\delta$  fixo igual a  $-2$ . A medida que  $\nu$  se torna maior, a densidade se torna menos assimétrica e tem caudas leves. Então a assimetria e as caudas pesadas são determinadas conjuntamente por uma combinação dos valores dos parâmetros  $\delta$  e  $\nu$ .

A média e a variância da variável aleatória  $X$  de distribuição HG t-Student assimétrica é:

$$E[X] = \mu + \frac{\delta\vartheta^2}{\nu - 2}$$

e

$$Var(X) = \frac{2\delta^2\vartheta^4}{(\nu - 2)^2(\nu - 4)} + \frac{\vartheta^2}{\nu - 2}$$

A variância somente é finita quando  $\nu > 4$ , diferente da distribuição t-Student simétrica que ocorre somente quando  $\nu > 2$ . A obtenção da assimetria e curtose é relativamente direta (Aas e Haff, 2006), devido a estrutura de mistura normal da distribuição.

A assimetria e a curtose são dadas por:

$$s = \frac{2(\nu - 4)^{1/2}\delta\vartheta}{[2\delta^2\vartheta^2 + (\nu - 2)(\nu - 4)]^{3/2}} \left[ 3(\nu - 2) + \frac{8\delta^2\vartheta^2}{\nu - 6} \right]$$

e

$$k = \frac{6}{[2\delta^2\vartheta^2 + (\nu - 2)(\nu - 4)]^2} \left[ (\nu - 2)^2(\nu - 4) + \frac{16\delta^4\vartheta^2(\nu - 2)(\nu - 4)}{\nu - 6} + \frac{8\delta^4\vartheta^4(5\nu - 22)}{(\nu - 6)(\nu - 8)} \right]$$

A assimetria e a curtose não existem quando  $\nu \leq 6$ , e  $\nu \leq 8$ , respectivamente.

Segue da equação (4.5) que nas caudas, a densidade t-Student assimétrica é dada por:

$$f(x) \sim \text{const}|x|^{-\nu/2-1} \exp(-|\delta x| + \delta x) \text{ quando } x \rightarrow \pm\infty.$$

Além disso a cauda mais pesada decai como

$$f(x) \sim \text{const}|x|^{-\nu/2-1} \text{ quando } \begin{cases} \delta < 0 \text{ e } x \rightarrow -\infty, \\ \delta > 0 \text{ e } x \rightarrow +\infty, \end{cases}$$

e a mais leve como

$$f(x) \sim \text{const}|x|^{-\nu/2-1} \exp(-2|\delta x|) \text{ quando } \begin{cases} \delta < 0 \text{ e } x \rightarrow +\infty, \\ \delta > 0 \text{ e } x \rightarrow -\infty, \end{cases}$$

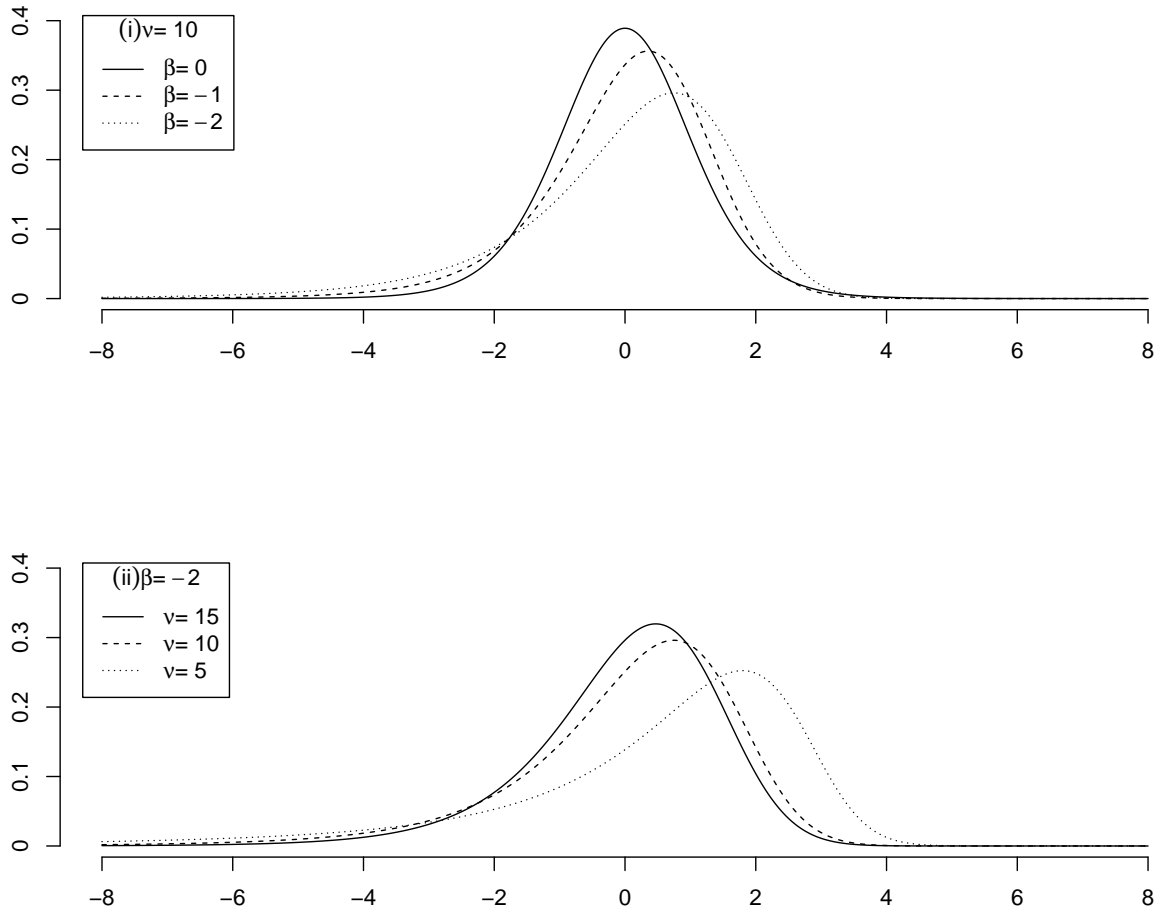


Figura 4.6: A Distribuição HG t-Student assimétrica. (i)  $\nu = 10$  fixo;  $\delta = 0$  (t-simétrica),  $-1$  e  $-2$ . (ii)  $\delta = -2$  fixo;  $\nu=5, 10$  e  $15$ .

Assim a HG t-Student assimétrica possui uma cauda pesada e outra semi-pesada. É o único membro da família de distribuições HG que possui essa propriedade. A equação (4.5) mostra que o único modo de obter uma cauda pesada e outra semi-pesada é considerar  $\alpha \rightarrow |\delta|$ . De acordo com as condições paramétricas dadas na equação (4.2), é necessário que  $\lambda < 0$ . Finalmente, se  $\lambda < 0$  e  $\alpha = |\delta|$ , nos podemos obter a HG t-Student assimétrica independente da magnitude de  $\lambda < 0$ .

O comportamento da distribuição HG t-Student assimétrica é diferente da distribuição t-Student assimétrica na qual as duas caudas decaem como polinomiais. Isto torna a única para modelagem substancial de assimetria e caudas pesadas nos dados.



# Capítulo 5

## Modelo VEM utilizando a distribuição HG t-Student Assimétrica

Os modelos de volatilidade estocástica (VE) têm sido considerados ferramentas úteis para modelagem da variância no tempo, principalmente em aplicações financeiras onde os agentes econômicos enfrentam problemas de decisão que geralmente dependem de medidas de volatilidade e risco. Uma característica atraente do modelo VE é a sua estreita relação com as teorias econômicas financeiras (Melino e Turnbull, 1990) e sua capacidade de captar os principais fatos estilizados presentes nos retornos financeiros de forma mais adequada (Carnero et al., 2004).

Frequentemente, a volatilidade de retornos diários de ações tem sido estimada com modelos VE, mas os resultados tem contado com uma extensiva pre-modelagem destas séries para evitar o problema de estimação simultânea da média e variância. Para lidar com este problema, Koopman e Uspensky (2002) introduziram o modelo VE na Média (VEM), incorporando a não observada volatilidade como uma variável explicativa na equação da média dos retornos. Eles utilizaram o método de máxima verossimilhança simulada para estimação dos parâmetros e forneceram uma justificação empírica que o coeficiente de volatilidade na média está relacionado com o efeito “feedback”. Nesta

dissertação nos propomos aumentar a robustez da especificação da inovações dos retornos em modelos VEM introduzido a distribuição HG t-Student assimétrica. Nos referimos a esta generalização como modelo VEM-HGt.

Há várias tipos de distribuições t-Student assimétricas na literatura (e.g. Hansen (1994); Fernandez e Steel (1998); Prause (1999); Azzaline e Capitanio (2003) e Aas e Haff (2006)). Entre esses, a distribuição dos erros HG t-Student assimétrico é simples, flexível e facilmente incorporada no modelo VEM porém a estimação desse modelo não é simples, já que agora a volatilidade aparece tanto na equação da variância quanto da média, logo, métodos computacionais intensivos fazem-se necessários. A inferência sobre o modelo VEM-HGt é desenvolvida sobre a abordagem Bayesiana através de métodos MCMC. Uma eficiente estratégia de amostragem por blocos é desenvolvida para simular as log-volatilidades (Abanto et al., 2010);(Shephard e Pitt, 1997);(Watanabe e Omori, 2004).

## 5.1 Modelo VEM-HGt

Com o objetivo de levar em conta as caudas pesadas, assimetria, efeito de alavancagem e “feedback” em retornos financeiros, é descrito aqui o modelo de volatilidade na média com efeito de alavancagem e erros com caudas assimetricamente pesadas utilizando a distribuição hiperbólica generalizada t-student assimétrica, o qual é definido como:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 e^{h_t} + e^{h_t/2} \omega_t \quad (5.1)$$

$$h_{t+1} = \alpha + \phi h_t + \sigma_\eta \eta_t \quad (5.2)$$

onde  $y_t$  e  $h_t$  são respectivamente, as componentes de retorno e a log-volatilidade no tempo  $t$ . A distribuição HG t-Student assimétrica é denotada no modelo por  $\omega_t$  a qual pode ser reescrita da forma:

$$\omega_t = \mu_\omega + \delta z_t + \sqrt{z_t} \epsilon_t \quad (5.3)$$

onde

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \text{ e } z_t \sim GI \left( \frac{\nu}{2}, \frac{\nu}{2} \right) \quad (5.4)$$

Assumi-se que  $\mu_\omega = -\delta\mu_z$  onde  $\mu_z = E[z_t] = \nu/(\nu - 2)$ , para que  $E[w_t] = 0$  e  $\nu > 4$  para que tenha variância finita. Consideramos  $|\phi| < 1$  para que o processo da log-volatilidade seja estacionário e o valor inicial  $h_1 \sim \mathcal{N}(\alpha/(1 - \phi), (1 - \rho^2)\sigma_\eta^2/(1 - \phi^2))$ . O parâmetro  $\rho$  mede a correlação entre  $\epsilon_t$  e  $\eta_t$ . Quando  $\rho < 0$  indica o efeito de alavancagem.

Então pode-se reescrever o modelo da seguinte maneira:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 e^{h_t} + e^{h_t/2} \{\delta(z_t - \mu_z) + \sqrt{z_t} \epsilon_t\} \quad (5.5)$$

$$h_{t+1} = \alpha + \phi h_t + \sigma_\eta \eta_t \quad (5.6)$$

onde  $(\epsilon_t, \eta_t)$  são como em (5.4). Quando  $\delta = 0$  o modelo se reduz ao modelo VEM t-Student simétrico (VEMt) estudado em Abanto-Valle et al. (2011)

## 5.2 Estimação dos parâmetros

Seja  $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \alpha, \phi, \tau^2, \varphi, \delta, \nu)$ , o vetor de parâmetros do modelo VEM-GHt, onde  $\tau^2 = \sqrt{1 - \rho^2}\sigma$  e  $\varphi = \rho\sigma_\eta$  pode-se reescrever as equações (5.5) e (5.6) como:

$$\begin{pmatrix} y_t \\ h_{t+1} \end{pmatrix} \Big| \boldsymbol{\theta}, h_t, z_t, y_{t-1} \sim \mathcal{N}_2 \left[ \begin{pmatrix} \beta_0 + \beta_1 y_{t-1} + \beta_2 e^{h_t} + e^{h_t/2} \delta(z_t - \mu_z) \\ \alpha + \phi h_t \end{pmatrix}, \begin{bmatrix} z_t e^{h_t} & \varphi \sqrt{z_t} e^{h_t/2} \\ \varphi \sqrt{z_t} e^{h_t/2} & \varphi^2 + \tau^2 \end{bmatrix} \right] \quad (5.7)$$

Com essa estrutura tem-se que a distribuição condicional  $y_t | \boldsymbol{\theta}, h_{t+1}, h_t, z_t, y_{t-1}$  segue uma distribuição normal com média e variância dadas por:

$$\mu_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 e^{h_t} + e^{h_t/2} \delta(z_t - \mu_z) + \frac{\varphi}{\varphi^2 + \tau^2} \sqrt{z_t} e^{h_t/2} (h_{t+1} - \alpha - \phi h_t) \quad (5.8)$$

$$V_t = \frac{\tau^2}{\varphi^2 + \tau^2} z_t e^{h_t} \quad (5.9)$$

Esta densidade vai ser útil no desenvolvimento do amostrador por blocos dado na seção (5.3).

Seja,  $\mathbf{y}_{0:n} = (y_0, \dots, y_n)'$  a informação disponível até o tempo  $n$ ,  $\mathbf{h}_{1:n} = (h_1, \dots, h_n)'$  o vetor das log-volatilidades e  $\mathbf{z}_{1:n} = (z_1, \dots, z_n)'$  as variáveis da mistura. A densidade conjunta a posteriori dos parâmetros e variáveis latentes não-observáveis pode ser escrita como:

$$p(\boldsymbol{\theta}, \mathbf{h}_{1:n}, \mathbf{z}_{1:n} | \mathbf{y}_{0:n}) \propto \left[ \prod_{t=1}^T p(y_t, h_{t+1} | z_t, h_t, y_{t-1}, \boldsymbol{\theta}) p(z_t | \nu) \right] p(h_1 | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \quad (5.10)$$

onde  $p(y_t, h_{t+1} | z_t, h_t, y_{t-1}, \boldsymbol{\theta})$  é dado pela equação (5.2) e  $p(\boldsymbol{\theta})$  é a distribuição a priori. Para realizar uma análise Bayesiana na estimação dos parâmetros do modelo VEM-HGt, geram-se amostras aleatórias da distribuição a posteriori de  $(\boldsymbol{\theta}, \mathbf{h}_{1:n}, \mathbf{z}_{1:n})$  dado  $\mathbf{y}_{1:n}$  utilizando métodos de simulação MCMC. O esquema de amostragem é descrito pelo algoritmo a seguir.

### Algoritmo

1. Fazer  $i = 0$ , inicializar os parâmetros  $\boldsymbol{\theta}^{(i)}$  e as variáveis latentes  $\mathbf{z}_{1:n}^{(i)}$  e  $\mathbf{h}_{1:n}^{(i)}$ ,
2. Gerar  $\boldsymbol{\theta}^{(i)}$  das distribuições condicionais completas, dados  $\mathbf{y}_{0:n}$ ,  $\mathbf{z}_{1:n}^{(i-1)}$  e  $\mathbf{h}_{1:n}^{(i-1)}$ ,
3. Gerar  $\mathbf{z}_{1:n}^{(i)}$  da distribuição  $p(\mathbf{z}_{1:n} | \boldsymbol{\theta}^{(i)}, \mathbf{h}_{1:n}^{(i-1)}, \mathbf{y}_{0:n})$ ,
4. Gerar  $\mathbf{h}_{1:n}$  por blocos como:
  - (i) Para  $l = 1, \dots, K$  as posições dos nós,  $k_l$ , são gerados como a parte inteira de  $[T\{(l+u_l)/(K+2)\}]$  onde os  $u_l$ 's são amostras independentes de variáveis aleatórias uniformes no intervalo  $(0, 1)$ .
  - (ii) Para  $l = 1, \dots, K$  gere  $\mathbf{h}_{k_{l-1}+1:k_l-1}$  conjuntamente condicionalmente em  $\mathbf{y}_{k_{l-1}+1:k_l-1}$ ,  $\boldsymbol{\theta}^{(i)}$ ,  $\mathbf{z}_{k_{l-1}+1:k_l-1}^{(i)}$  e  $\mathbf{h}_{k_{l-1}+1:k_l-1}^{(i-1)}$ .
  - (iii) Para  $l = 1, \dots, K$  gere  $\mathbf{h}_{k_l}$  condicional em  $\mathbf{y}_{0:n}$ ,  $\boldsymbol{\theta}^{(i)}$ ,  $\mathbf{h}_{k_l-1}^{(i)}$  e  $\mathbf{h}_{k_l+1}^{(i)}$ .
5. Fazer  $i = i + 1$  e retorne a 2 até a convergência ser atingida.

### 5.2.1 Distribuições condicionais completas

Aqui serão apresentadas as distribuições condicionais completas dos parâmetros e variáveis latentes do modelo VEM-HGt.

Para as distribuições a priori dos parâmetros assumi-se:

$$\begin{aligned} \beta_0 &\sim \mathcal{N}(\bar{\beta}_0, \sigma_{\beta_0}^2) & \alpha|\tau^2 &\sim \mathcal{N}(\alpha_0, \tau^2/q_0) & \phi|\tau^2 &\sim \mathcal{N}(\phi_0, s_\phi^2)\mathbb{I}_{\{|\phi|<1\}} \\ \beta_1 &\sim \mathcal{N}(\bar{\beta}_1, \sigma_{\beta_1}^2)\mathbb{I}_{\{|\beta_1|<1\}} & \delta &\sim \mathcal{N}(\bar{\delta}, \sigma_\delta^2) & \tau^2 &\sim \mathcal{IG}(A_\tau/2, S_\tau/2) \\ \beta_2 &\sim \mathcal{N}(\bar{\beta}_2, \sigma_{\beta_2}^2) & \varphi|\tau^2 &\sim \mathcal{N}(\varphi_0, \tau^2/p_0) & \nu &\sim \mathcal{G}(A_\nu, B_\nu)\mathbb{I}_{\{\nu>4\}}; \end{aligned}$$

**Distribuição condicional completa de  $\beta_0$ ,  $\beta_1$  e  $\beta_2$ :**

Definamos  $m_t$  e  $v_t$  por:

$$m_t = \begin{cases} \frac{\varphi}{\varphi^2 + \tau^2} \sqrt{z_t} e^{h_t/2} (h_{t+1} - \alpha - \phi h_t) & \text{se } t < T, \\ 0 & \text{se } t = T, \end{cases} \quad v_t^2 = \begin{cases} \frac{\tau^2}{\varphi^2 + \tau^2} z_t e^{h_t} & \text{se } t < T, \\ z_t e^{h_t} & \text{se } t = T. \end{cases}$$

$$\beta_0 | \mathbf{y}_{0:n}, \mathbf{h}_{1:n}, \mathbf{z}_{1:n}, \beta_1, \beta_2 \sim \mathcal{N}(a_{\beta_0} b_{\beta_0}, b_{\beta_0})$$

$$\beta_1 | \mathbf{y}_{0:n}, \mathbf{h}_{1:n}, \mathbf{z}_{1:n}, \beta_0, \beta_2 \sim \mathcal{N}(a_{\beta_1} b_{\beta_1}, b_{\beta_1}) \mathbb{I}_{|\beta_1|<1}$$

$$\beta_2 | \mathbf{y}_{0:n}, \mathbf{h}_{1:n}, \mathbf{z}_{1:n}, \beta_0, \beta_1 \sim \mathcal{N}(a_{\beta_2} b_{\beta_2}, b_{\beta_2})$$

onde

$$\begin{aligned} a_{\beta_0} &= \sum_{t=1}^n \frac{1}{v_t^2} \left( y_t - \beta_1 y_{t-1} - \beta_2 e^{h_t} - e^{h_t/2} \delta(z_t - \mu_z) - m_t \right) + \frac{\bar{\beta}_0}{\sigma_{\beta_0}^2} & b_{\beta_0} &= \left( \sum_{t=1}^n \frac{1}{v_t^2} + \frac{1}{\sigma_{\beta_0}^2} \right)^{-1} \\ a_{\beta_1} &= \sum_{t=1}^n \frac{y_{t-1}}{v_t^2} \left( y_t - \beta_0 - \beta_2 e^{h_t} - e^{h_t/2} \delta(z_t - \mu_z) - m_t \right) + \frac{\bar{\beta}_1}{\sigma_{\beta_1}^2} & b_{\beta_1} &= \left( \sum_{t=1}^n \frac{y_{t-1}^2}{v_t^2} + \frac{1}{\sigma_{\beta_1}^2} \right)^{-1} \\ a_{\beta_2} &= \sum_{t=1}^n \frac{e^{h_t}}{v_t^2} \left( y_t - \beta_0 - \beta_1 y_{t-1} - e^{h_t/2} \delta(z_t - \mu_z) - m_t \right) + \frac{\bar{\beta}_2}{\sigma_{\beta_2}^2} & b_{\beta_2} &= \left( \sum_{t=1}^n \frac{e^{2h_t}}{v_t^2} + \frac{1}{\sigma_{\beta_2}^2} \right)^{-1} \end{aligned}$$

**Distribuição condicional completa de  $\alpha$ ,  $\varphi$ ,  $\tau^2$  e  $\delta$ :**

Considere  $g_t = y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t} - e^{h_t/2} \delta(z_t - \mu_z)$ .

Logo as distribuições condicionais completas para os parâmetros são dadas por

$$\alpha | \cdot \sim N(a_\alpha b_\alpha, b_\alpha \tau^2)$$

$$\varphi | \cdot \sim \mathcal{N}(a_\varphi b_\varphi, b_\varphi \tau^2)$$

$$\delta | \cdot \sim \mathcal{N}(a_\delta b_\delta, b_\delta)$$

$$\tau^2 | \cdot \sim \mathcal{GI}(a_\tau/2, b_\tau/2)$$

onde

$$a_\alpha = q_0 \alpha_0 + (1 + \phi) h_1 + \sum_{t=1}^{n-1} \left[ h_{t+1} - \phi h_t - \varphi z_t^{-1/2} e^{-h_t/2} g_t \right]$$

$$b_\alpha = \left( q_0 + n - 1 + \frac{1 + \phi}{1 - \phi} \right)^{-1}$$

$$a_\varphi = \varphi_0 p_0 + \sum_{t=1}^{n-1} g_t z_t^{-1/2} e^{-h_t/2} (h_{t+1} - \alpha - \phi h_t)$$

$$b_\varphi = \left( p_0 + \sum_{t=1}^{n-1} z_t^{-1} e^{-h_t} g_t^2 \right)^{-1}$$

$$a_\tau = n + A_\tau + 2$$

$$b_\tau = \sum_{t=1}^{n-1} \left( h_{t+1} - \alpha - \phi h_t - \varphi z_t^{-1/2} e^{-h_t/2} g_t \right)^2 + \\ + (1 - \phi^2) \left( h_1 - \frac{\alpha}{1 - \phi} \right)^2 + p_0 (\varphi - \varphi_0)^2 + q_0 (\alpha - \alpha_0)^2 + S_\tau$$

$$\begin{aligned}
a_\delta &= -\frac{\varphi}{\tau^2} \sum_{t=1}^{n-1} \frac{1}{\sqrt{z_t}} (z_t - \mu_z) (h_{t+1} - \alpha - \phi h_t) \\
&+ \left( \frac{\varphi^2 + \tau^2}{\tau^2} \right) \sum_{t=1}^{n-1} \frac{1}{z_t e^{h_t/2}} (z_t - \mu_z) (y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t}) \\
&+ \frac{1}{z_n e^{h_n/2}} (z_n - \mu_z) (y_n - \beta' x_n) + \frac{\bar{\delta}}{\sigma_\delta^2} \\
b_\delta &= \left[ \left( \frac{\varphi^2 + \tau^2}{\tau^2} \right) \sum_{t=1}^{n-1} \left( \frac{1}{z_t} (z_t - \mu_z)^2 \right) + \frac{1}{z_n} (z_n - \mu_z)^2 + \frac{1}{\sigma_\delta^2} \right]^{-1}
\end{aligned}$$

As distribuições até aqui são conhecidas e com isso são diretamente amostradas. A partir daqui as distribuições não possuem forma fechada e para que sejam amostradas serão utilizadas os métodos vistos na sessões (2.2) e (2.3).

#### Distribuição condicional completa de $\phi$ :

$$p(\phi|\cdot) \propto \sqrt{1 - \phi^2} \exp \left[ -\frac{(1 - \phi^2)}{2\tau^2} \left( h_1 - \frac{\alpha}{1 - \phi} \right)^2 \right] \exp \left[ -\frac{1}{2b_\phi} (\phi - a_\phi b_\phi)^2 \right]$$

onde

$$a_\phi = \frac{\phi_0}{s_\phi^2} + \frac{1}{\tau^2} \sum_{t=1}^{n-1} h_t \left( h_{t+1} - \alpha - \varphi z_t^{-1/2} e^{-h_t/2} g_t \right) \text{ e } b_\phi = \left( \frac{1}{s_\phi^2} + \sum_{t=1}^{n-1} \frac{h_t^2}{\tau^2} \right)^{-1}$$

Amostra-se  $\phi$  utilizando o algoritmo de M-H sendo a densidade proposta  $N(a_\phi b_\phi, b_\phi) \mathbb{I}_{|\phi| < 1}$ .

#### Distribuição condicional completa de $\nu$ :

$$\begin{aligned}
p(\nu|\cdot) &\propto \frac{\nu^{\frac{n\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)^n} \exp \left\{ -\sum_{t=1}^n \frac{1}{2v_t^2} \left[ y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t} - e^{h_t/2} \delta \left( z_t - \frac{\nu}{\nu - 2} \right) - m_t \right]^2 - \right. \\
&\quad \left. -\frac{\nu}{2} \sum_{t=1}^n \left[ \frac{1}{z_t} + \log z_t \right] \right\} \nu^{A_\nu - 1} \exp\{-B_\nu \nu\} \mathbb{I}_{\nu > 4}
\end{aligned}$$

Amostra-se  $\nu$  utilizando o algoritmo ARMH. Considere  $\nu^*$  a moda de  $p(\nu|\cdot)$  e seja  $l(\nu) = \log\{p(\nu|\cdot)\}$ .  $l(\nu)$  é uma função concava, assim é utilizada a densidade proposta

$N(\mu_\nu, \sigma_\nu^2) \mathbb{I}_{4 \leq \nu \leq 40}$ , onde  $\mu_\nu = \nu^* - l'(\nu^*)/l''(\nu^*)$  e  $\sigma_\nu^2 = -1/l''(\nu^*)$  tal que  $l'(\nu^*)$  e  $l''(\nu^*)$  são, respectivamente, a primeira e a segunda derivadas de  $l(\nu)$  calculadas em  $\nu^*$ .

**Distribuição condicional completa de  $z_t$ :**

$$p(z_t | \cdot) \propto h(z) \exp \left\{ \frac{\varphi}{\tau^2} \left[ z_t^{-1/2} e^{-h_t/2} (h_{t+1} - \alpha - \phi h_t) (y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t} + \mu_z e^{h_t/2} \delta) - \right. \right. \\ \left. \left. - z_t^{1/2} \delta (h_{t+1} - \alpha - \phi h_t) \right] \mathbb{I}_{\{t < n\}} \right\}$$

onde  $h(z)$  é o núcleo da densidade da distribuição  $\text{GIG}(\lambda, \vartheta, \gamma)$  dada na equação (4.6) cujos parâmetros são dados abaixo:

$$\lambda = -\frac{\nu + 1}{2} \\ \gamma^2 = \delta^2 \frac{\varphi^2 + \tau^2}{\tau^2} \\ \vartheta^2 = \frac{\varphi^2 + \tau^2}{\tau^2} e^{-h_t} (y_t - \beta_0 - \beta_1 y_{t-1} - \beta_2 e^{h_t} + \mu_z e^{h_t/2} \delta)^2 + \frac{\nu}{2}$$

A amostragem de  $z_t$  é feita através do algoritmo M-H, onde proposta utilizada é a  $\text{GIG}(\lambda, \vartheta, \gamma)$ .

Devido a estrutura não linear da equação das observações a amostragem das log-volatilidades  $\mathbf{h}_{1:n}$  é uma tarefa difícil. Uma estratégia eficiente é amostrar da distribuição a posteriori condicional de  $\mathbf{h}_{1:n}$  dividindo em diversos blocos e amostrar cada bloco dados demais blocos. Essa estratégia é chamada de Amostragem por Blocos, o assunto que iremos tratar na próxima seção.

## 5.3 Amostragem por blocos

Nesta seção será apresentado o método de amostragem por blocos para as log-volatilidades  $\mathbf{h}_{1:n}$  do modelo VEM-HGt. O método é baseado em Omori e Watanabe (2008) e Abanto-Valle et al. (2011) pois os erros da equação observacional e das log-volatilidades são correlacionados. Eles mostraram que a amostragem por blocos pode produzir amostras eficientes da distribuição condicional a posteriori em comparação com a amostragem



componente a componente a qual primitivamente amostra um estado  $h_t$ , dado os outros  $h_s (s \neq t)$ .

Para amostrar do vetor de log-volatilidades  $\mathbf{h}_{1:n} = (h_1, \dots, h_n)'$ , dividimos em  $K+1$  blocos,  $\mathbf{h} = (h_{k_{l-1}+1}, \dots, h_{k_l-1})'$  para  $l = 1, \dots, K+1$ , com  $k_0 = 0$  e  $k_{K+1} = n$ , onde  $k_l - 1 - k_{l-1} \geq 2$  é o tamanho do  $l$ -ésimo bloco. Nos amostra-se do bloco de ruídos  $\boldsymbol{\eta}_{k_{l-1}:k_l-2} = (\eta_{k_{l-1}}, \dots, \eta_{k_l-2})'$  dado as condições  $h_{k_{l-1}}$  e  $h_{k_l}$  no lugar de  $\mathbf{h} = (h_{k_{l-1}+1}, \dots, h_{k_l-1})'$  como em Omori e Watanabe (2008).

A fim de facilitar a notação, omitimos a dependência em  $\theta$  e supomos que  $k_{l-1} = t$  e  $k_l = t+k+1$  para o  $l$ -ésimo bloco, tal que  $t+k < n$ . Então  $\boldsymbol{\eta}_{t:t+k-1} = (\eta_t, \dots, \eta_{t+k-1})'$  são amostrados de uma vez da sua distribuição condicional completa, a qual sem os termos constantes é expreso na escala logarítmica como:

$$\log f(\boldsymbol{\eta}_{t:t+k-1} | h_t, h_{t+k+1}, \mathbf{y}_{t:t+k}, \mathbf{z}_{t+1:t+k}) \doteq - \sum_{s=t}^{t+k-1} \frac{\eta_s^2}{2} + \sum_{s=t}^{s+k} l_s - \frac{1}{2\sigma_\eta^2} (h_{t+k+1} - \alpha - \phi h_{t+k})^2 \mathbb{I}(t+k < n) \quad (5.11)$$

onde, excluindo os termos constantes,  $l_s$  denota a distribuição condicional de  $y_s$  dado  $h_s$  e  $h_{s+1}$  para  $s < T$ , a qual é normal com média  $\mu_s$  e variância  $V_s$ , dadas pelas equações (5.8) e (5.9), respectivamente.

Define-se

$$L = \sum_{s=t}^{t+k} l_s - \frac{1}{2\sigma_\eta^2} (h_{t+k+1} - \alpha - \phi h_{t+k})^2 \mathbb{I}(t+k < n) \quad (5.12)$$

Como  $-\frac{1}{2} \sum_{s=t}^{t+k-1} \eta_s^2 + L$  em (5.11) não possui forma fechada, nos usamos o algoritmo ARMH, Chib e Greenberg (1995), para amostragem. Para obter a densidade proposta é desenvolvida um aproximação linear do modelo de espaço-estado que imita (5.11), de forma que seja fácil de amostra-la.

---

<sup>1</sup>Para o último bloco, tem-se que:

$$y_T | y_{T-1}, h_T \sim \mathcal{N}(\beta_0 + \beta_1 y_{T-1} + \beta_2 e^{h_T} + e^{h_T} \delta(z_T - \mu_z), z_T e^{h_T})$$

A seguir são definidas algumas equações que serão utilizadas no desenvolvimento da densidade proposta.

Considere  $\mathbf{d}_{t+1:t+k} = (d_{t+1}, \dots, d_{t+k})'$  e sejam  $d_s$  e  $\mathbf{Q}$  definidos como:

$$d_s = \frac{\partial L}{\partial h_s} = -\frac{1}{2} + \frac{(y_s - \mu_s)^2}{2V_s} + \frac{(y_s - \mu_s)}{V_s} \frac{\partial \mu_s}{\partial h_s} + \frac{(y_{s-1} - \mu_{s-1})}{V_{s-1}} \frac{\partial \mu_{s-1}}{\partial h_s} - \phi \frac{(h_{s+1} - \alpha - \phi h_s)}{\sigma_\eta^2} \mathbb{I}(s = t+k, s < n) \quad (5.13)$$

e

$$\mathbf{Q} = \begin{bmatrix} M_{t+1} & N_{t+2} & 0 & \dots & 0 \\ N_{t+2} & M_{t+2} & N_{t+3} & \dots & 0 \\ 0 & N_{t+3} & M_{t+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & N_{t+k} \\ 0 & \dots & 0 & N_{t+k} & M_{t+k} \end{bmatrix} \quad (5.14)$$

De maneira que:

$$\frac{\partial \mu_s}{\partial h_s} = \begin{cases} \beta_2 e^{h_s} + \frac{1}{2} e^{h_s/2} \delta(z_s - \mu_z) + \frac{\varphi}{\varphi^2 + \tau^2} \sqrt{z_s} e^{h_s/2} \left[ \frac{(h_{s+1} - \alpha - \phi h_s)}{2} - \phi \right], & s = 1, \dots, n-1 \\ \beta_2 e^{h_s} + \frac{1}{2} e^{h_s/2} \delta(z_s - \mu_z), & s = n \end{cases}$$

$$\frac{\partial \mu_{s-1}}{\partial h_s} = \begin{cases} 0, & s = 1 \\ \frac{\varphi}{\varphi^2 + \tau^2} \sqrt{z_{s-1}} e^{h_{s-1}/2}, & s = 2, \dots, n \end{cases}$$

e

$$M_s = -E \left[ \frac{\partial^2 L}{\partial h_s^2} \right] = \frac{1}{2} + \frac{1}{V_s} \left( \frac{\partial \mu_s}{\partial h_s} \right)^2 + \frac{1}{V_{s-1}} \left( \frac{\partial \mu_{s-1}}{\partial h_s} \right)^2 + \frac{\phi^2}{\sigma_\eta^2} \mathbb{I}(s = t+k < n) \quad (5.15)$$

$$N_s = -E \left[ \frac{\partial^2 L}{\partial h_s \partial h_{s-1}} \right] = \frac{1}{V_{s-1}} \frac{\partial \mu_{s-1}}{\partial h_{s-1}} \frac{\partial \mu_{s-1}}{\partial h_s} \quad (5.16)$$

onde

$$\begin{aligned} \frac{\partial^2 L}{\partial h_s^2} = & -\frac{(y_s - \mu_s)^2}{2V_s} - \frac{(y_s - \mu_s)}{V_s} \frac{\partial \mu_s}{\partial h_s} - \frac{1}{V_s} \left( \frac{\partial \mu_s}{\partial h_s} \right)^2 + \left( \frac{y_s - \mu_s}{V_s} \right) \left[ \frac{\partial^2 \mu_s}{\partial h_s^2} - \frac{\partial \mu_s}{\partial h_s} \right] - \\ & - \frac{1}{V_{s-1}} \left( \frac{\partial \mu_{s-1}}{\partial h_s} \right)^2 + \frac{\phi^2}{\sigma_\eta^2} \mathbb{I}(s = t+k, s < n) \end{aligned}$$

$$\frac{\partial^2 L}{\partial h_s \partial h_{s-1}} = -\frac{1}{V_{s-1}} \frac{\partial \mu_{s-1}}{\partial h_{s-1}} \frac{\partial \mu_{s-1}}{\partial h_s} + \frac{(y_{s-1} - \mu_{s-1})}{V_{s-1}} \left[ -\frac{\partial \mu_{s-1}}{\partial h_s} + \frac{\partial^2 \mu_{s-1}}{\partial h_s \partial h_{s-1}} \right]$$

Expandindo  $L$  em serie de Taylor até a segunda ordem em torno da moda  $\hat{\boldsymbol{\eta}}_{t:t+k-1}$  e realizando-se algumas contas tem-se que:

$$\begin{aligned} \log f(\boldsymbol{\eta}_{t:t+k-1} | h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \mathbf{z}_{t+1:t+k}) &\approx \\ &\approx \text{const} - \frac{1}{2} \sum_{r=t+1}^{t+k} \eta_r^2 + \hat{L} + \frac{\partial L}{\partial \boldsymbol{\eta}'_{t:t+k-1}} \Big|_{\boldsymbol{\eta}_{t:t+k-1} = \hat{\boldsymbol{\eta}}_{t:t+k-1}} (\boldsymbol{\eta}_{t:t+k-1} - \hat{\boldsymbol{\eta}}_{t:t+k-1}) + \\ &\quad + \frac{1}{2} (\boldsymbol{\eta}_{t:t+k-1} - \hat{\boldsymbol{\eta}}_{t:t+k-1})' E \left( \frac{\partial^2 L}{\partial \boldsymbol{\eta}_{t:t+k-1} \partial \boldsymbol{\eta}'_{t:t+k-1}} \right) \Big|_{\boldsymbol{\eta}_{t:t+k-1} = \hat{\boldsymbol{\eta}}_{t:t+k-1}} (\boldsymbol{\eta}_{t:t+k-1} - \hat{\boldsymbol{\eta}}_{t:t+k-1}) \\ &= \text{const} - \frac{1}{2} \sum_{r=t+1}^{t+k} \eta_r^2 + \hat{L} + \hat{\mathbf{d}}'_{t+1:t+k} (\mathbf{h}_{t+1:t+k} - \hat{\mathbf{h}}_{t+1:t+k}) - \\ &\quad - \frac{1}{2} (\mathbf{h}_{t+1:t+k} - \hat{\mathbf{h}}_{t+1:t+k})' \hat{\mathbf{Q}} (\mathbf{h}_{t+1:t+k} - \hat{\mathbf{h}}_{t+1:t+k}) \\ &= \text{const} + \log f(\boldsymbol{\eta}_{t:t+k-1} | h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \mathbf{z}_{t+1:t+k}) \end{aligned}$$

onde  $\hat{\mathbf{d}}_{t+1:t+k}$ ,  $\hat{L}$  e  $\hat{\mathbf{Q}}$  denotam  $\mathbf{d}_{t+1:t+k}$ ,  $L$  e  $\mathbf{Q}$  calculados em  $\mathbf{h}_{t+1:t+k} = \hat{\mathbf{h}}_{t+1:t+k}$ . As esperanças que são apresentadas acima são dos  $y'_s$ s condicionais aos  $h'_s$ s. A matriz de informação  $\mathbf{Q}$  é utilizada é necessário uma matriz positiva definida. A seguir mostra-se que a densidade proposta  $f^*(\boldsymbol{\eta}_{t:t+k-1} | h_t, h_{t+k+1}, \mathbf{y}_{t+1:t+k}, \mathbf{z}_{t+1:t+k})$  é a densidade a posteriori de  $\boldsymbol{\eta}_{t:t+k-1}$  para um modelo de espaço-estado Gaussiano dado as equações (5.17) e (5.18) abaixo (Omori e Watanabe, 2008).

Seja:

$$\frac{\partial L}{\partial \eta_j} = \sum_{s=j+1}^{t+k} \frac{\partial L}{\partial h_s} \frac{\partial h_s}{\partial \eta_j}$$

onde

$$\frac{\partial h_s}{\partial \eta_j} = \begin{cases} \frac{\partial h_s}{\partial h_{s-1}} \cdots \frac{\partial h_{j+2}}{\partial h_{j+1}} \frac{\partial h_{j+1}}{\partial \eta_j} = \phi^{s-1-j}, & s \geq j+1 \\ 0, & s \leq j. \end{cases}$$

Da equação (5.6), tem-se:

$$\begin{aligned} h_s &= \alpha + \phi h_{s-1} + \sigma_\eta \eta_{s-1} \\ &= \alpha \left( \frac{1 - \phi^{s-t}}{1 - \phi} \right) + \phi^{s-t} h_t + \sum_{j=t}^{s-1} \phi^{s-1-j} \sigma_\eta \eta_j \\ &= \alpha \left( \frac{1 - \phi^{s-t}}{1 - \phi} \right) + \phi^{s-t} h_t + \sum_{j=t}^{s-1} \frac{\partial h_s}{\partial \eta_j} \eta_j \end{aligned}$$

então

$$h_s - \hat{h}_s = \sum_{j=t}^{s-1} \frac{\partial h_s}{\partial \eta_j} (\eta_j - \hat{\eta}_j)$$

dado isso, segue que:

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{\eta}_{t:t+k-1}} \Big|_{\boldsymbol{\eta}_{t:t+k-1} = \hat{\boldsymbol{\eta}}_{t:t+k-1}} (\boldsymbol{\eta}_{t:t+k-1} - \hat{\boldsymbol{\eta}}_{t:t+k-1}) &= \sum_{j=t}^{t+k-1} \frac{\partial L}{\partial \eta_j} \Big|_{\eta_j = \hat{\eta}_j} (\eta_j - \hat{\eta}_j) \\ &= \sum_{j=t}^{t+k-1} \sum_{s=j+1}^{t+k} \frac{\partial L}{\partial h_s} \Big|_{h_j = \hat{h}_j} \frac{\partial h_s}{\partial \eta_j} \Big|_{\eta_j = \hat{\eta}_j} (\eta_j - \hat{\eta}_j) \\ &= \sum_{s=t+1}^{t+k} \hat{d}_s \sum_{j=t}^{s-1} \frac{\partial h_s}{\partial \eta_j} \Big|_{\eta_j = \hat{\eta}_j} (\eta_j - \hat{\eta}_j) \\ &= \sum_{s=t+1}^{t+k} \hat{d}_s (h_s - \hat{h}_s) \\ &= \hat{d}'_{t+1:t+k} (\boldsymbol{h}_{t+1:t+k} - \hat{\boldsymbol{h}}_{t+1:t+k}). \end{aligned}$$

Da mesma maneira:

$$\begin{aligned}
\frac{\partial^2 L}{\partial \eta_i \partial \eta_j} &= \frac{\partial}{\partial \eta_i} \left[ \frac{\partial L}{\partial \eta_j} \right] \\
&= \frac{\partial}{\partial \eta_j} \left[ \sum_{s_2=j+1}^{t+k} \frac{\partial L}{\partial h_{s_2}} \frac{\partial h_{s_2}}{\partial \eta_j} \right] \\
&= \sum_{s_2=j+1}^{t+k} \left[ \frac{\partial^2 L}{\partial \eta_i \partial h_{s_2}} \frac{\partial h_{s_2}}{\partial \eta_j} + \frac{\partial L}{\partial h_{s_2}} \frac{\partial^2 h_{s_2}}{\partial \eta_i \partial \eta_j} \right] \\
&= \sum_{s_2=j+1}^{t+k} \sum_{s_1=i+1}^{t+k} \frac{\partial^2 L}{\partial h_{s_1} \partial h_{s_2}} \frac{\partial h_{s_1}}{\partial \eta_i} \frac{\partial h_{s_2}}{\partial \eta_j} + \sum_{s_2=j+1}^{t+k} \frac{\partial L}{\partial h_{s_2}} \frac{\partial^2 h_{s_2}}{\partial \eta_i \partial \eta_j},
\end{aligned}$$

calculando-se seu valor esperado temos:

$$E \left( \frac{\partial^2 L}{\partial \eta_i \partial \eta_j} \right) = \sum_{s_2=j+1}^{t+k} \sum_{s_1=i+1}^{t+k} E \left[ \frac{\partial^2 L}{\partial h_{s_1} \partial h_{s_2}} \right] \frac{\partial h_{s_1}}{\partial \eta_i} \frac{\partial h_{s_2}}{\partial \eta_j}$$

então

$$\begin{aligned}
&(\boldsymbol{\eta}_{t:t+k-1} - \hat{\boldsymbol{\eta}}_{t:t+k-1})' E \left( \frac{\partial^2 L}{\partial \boldsymbol{\eta}_{t:t+k-1} \partial \boldsymbol{\eta}'_{t:t+k-1}} \right) \Big|_{\boldsymbol{\eta}_{t:t+k-1} = \hat{\boldsymbol{\eta}}_{t:t+k-1}} (\boldsymbol{\eta}_{t:t+k-1} - \hat{\boldsymbol{\eta}}_{t:t+k-1}) = \\
&= \sum_{s_1=t+1}^{t+k} \sum_{s_2=t+1}^{t+k} \sum_{i=t}^{s_1-1} \sum_{j=t}^{s_2-1} (\eta_i - \hat{\eta}_i) \frac{\partial h_{s_1}}{\partial \eta_i} \Big|_{\eta_i = \hat{\eta}_i} E \left( \frac{\partial^2 L}{\partial h_{s_1} \partial h_{s_2}} \right) \Big|_{h_i = \hat{h}_i, h_j = \hat{h}_j} \frac{\partial h_{s_2}}{\partial \eta_j} \Big|_{\eta_j = \hat{\eta}_j} (\eta_j - \hat{\eta}_j) \\
&= \sum_{s_1=t+1}^{t+k} \sum_{s_2=t+1}^{t+k} (h_{s_1} - \hat{h}_{s_1}) E \left( \frac{\partial^2 L}{\partial h_{s_1} \partial h_{s_2}} \right) \Big|_{h_i = \hat{h}_i, h_j = \hat{h}_j} (h_{s_2} - \hat{h}_{s_2}) \\
&= -(\boldsymbol{h}_{t+1:t+k} - \hat{\boldsymbol{h}}_{t+1:t+k})' \hat{\boldsymbol{Q}} (\boldsymbol{h}_{t+1:t+k} - \hat{\boldsymbol{h}}_{t+1:t+k}).
\end{aligned}$$

A moda  $\hat{\boldsymbol{\eta}}_{t:t+k-1}$  pode ser encontrada realizando o seguinte algoritmo:

**Algoritmo:**

1. Inicialize  $\hat{\boldsymbol{\eta}}_{t:t+k-1}$  e calcule  $\hat{\boldsymbol{h}}_{t+1:t+k}$  utilizando (5.6),
2. Calcule  $\hat{d}_s, \hat{M}_s$  e  $\hat{N}_s$  usando as equações (5.13), (5.15) e (5.16), respectivamente,
3. Calcule  $G_s, J_s$  e  $b_s$ , para  $s = t + 2, \dots, t + k$ , recursivamente:

$$\begin{aligned} G_s &= \hat{M}_s - \hat{N}_s^2 G_{s-1}^{-1}, \quad G_{t+1} = \hat{M}_{t+1}, \\ J_s &= K_{s-1}^{-1} \hat{N}_s, \quad J_{t+k+1} = 0, \\ b_s &= \hat{d}_s - J_s K_{s-1}^{-1} b_{s-1}, \quad b_{t+1} = \hat{d}_{t+1}. \end{aligned}$$

onde  $K_s = \sqrt{G_s}$ .

4. Defina as variáveis auxiliares  $\hat{y}_s = \hat{\gamma}_s + G_s^{-1} b_s$ , onde

$$\hat{\gamma}_s = \hat{h}_s + K_s^{-1} J_{s+1} \hat{h}_{s+1}, \quad s = t + 1, \dots, t + k$$

5. Considere o modelo de espaço-estado Gaussiano:

$$\hat{y}_s = c_s + Z_s h_s + H_s \xi_s, \quad s = t + 1, \dots, t + k, \quad (5.17)$$

$$h_{s+1} = \alpha + \phi h_s + L_s \xi_s, \quad s = t, t + 1, \dots, t + k. \quad (5.18)$$

onde  $\xi_s \sim N(\mathbf{0}, \mathbf{I}_2)$ ,  $c_s = K_s^{-1} J_{s+1} \alpha$ ,  $Z_s = 1 + K_s^{-1} J_{s+1} \phi$ ,  $H_s = K_s^{-1} [1 J_{s+1} \sigma_\eta]$  e  $L_s = [0, \sigma_\eta]$ . Execute o filtro de Kalman e um disturbance smoother (Koopman, 1993) para o modelo de espaço-estado Gaussiano nas equações (5.17), (5.18) e obtenha a media a posteriori de  $\boldsymbol{\eta}_{t:t+k-1}(\boldsymbol{h}_{t:t+k})$  e faça  $\hat{\boldsymbol{\eta}}_{t:t+k-1}(\hat{\boldsymbol{h}}_{t:t+k})$  igual a este valor,

6. Retorne ao passo 2 e repita o processo até atingir convergência.

Aplicando o o algoritmos de Jong e Shephard para o modelo definido em (5.17) e (5.18) com as variáveis auxiliares definidas no passo 4 do Algoritmo acima permite-nos amostrar  $\boldsymbol{\eta}_{t:t+k-1}$  da densidade  $f^*$ . Já que  $f$  não é limitada por  $f^*$ , nós utilizamos o algoritmo ARMH (2.3.3) para amostrar de  $f$  como recomendado por Chib e Greenberg (1995).

Como  $\hat{\mathbf{Q}}$  é uma matriz positiva definida, há uma matriz triangular inferior  $\mathbf{R}$  tal que  $\hat{\mathbf{Q}} = \mathbf{R}\mathbf{R}'$ , utilizando a decomposição de Cholesky, onde

$$\mathbf{R} = \begin{bmatrix} K_{t+1} & 0 & 0 & \dots & 0 \\ J_{t+2} & K_{t+2} & 0 & \dots & 0 \\ 0 & J_{t+3} & K_{t+3} & \ddots & \vdots \\ \vdots & \ddots & \vdots & \vdots & 0 \\ 0 & \dots & 0 & J_{t+k} & K_{t+k} \end{bmatrix}$$

tal que

$$\hat{M}_s = J_s^2 + K_s^2, \quad s = t+1, \dots, t+k,$$

$$\hat{N}_s = J_s K_{s-1}, \quad s = t+2, \dots, t+k,$$

$$N_{t+1} = J_{t+1} = J_{t+k+1} = 0.$$

Considere  $F_s = J_s^2$  e  $G_s = K_s^2$ , então:

$$F_s = \hat{N}_s^2 G_{s-1}^{-1}, \quad s = t+2, \dots, t+k,$$

$$G_s = \hat{M}_s - F_s = \hat{M}_s - F_s = \hat{M}_s - \hat{N}_s^2 G_{s-1}^{-1}, \quad s = t+2, \dots, t+k$$

$$G_{t+1} = \hat{M}_{t+1}$$

onde  $K_s = \sqrt{G_s}$  e  $J_s = K_{s-1}^{-1} \hat{N}_s$ .

Considere  $\mathbf{K} = \text{diag}(K_{t+1}, \dots, K_{t+k})$ ,  $\mathbf{G} = \text{diag}(G_{t+1}, \dots, G_{t+k})$ ,  $\mathbf{b}_{t+1:t+k} = \mathbf{K}\mathbf{R}^{-1}\hat{\mathbf{d}}_{t+1:t+k} =$

$KR^{-1}\hat{\mathbf{d}}_{t+1:t+k}$ ,  $\gamma_{t+1:t+k} = K^{-1}R'h_{t+1:t+k}$  e  $\hat{\gamma}_{t+1:t+k} = K^{-1}R'\hat{h}_{t+1:t+k}$ . Então temos:

$$\begin{aligned} & \mathbf{d}_{t+1:t+k}^{\wedge}(\mathbf{h}_{t+1:t+k} - \hat{\mathbf{h}}_{t+1:t+k}) - \frac{1}{2}(\mathbf{h}_{t+1:t+k} - \hat{\mathbf{h}}_{t+1:t+k})'\hat{\mathbf{Q}}(\mathbf{h}_{t+1:t+k} - \hat{\mathbf{h}}_{t+1:t+k}) = \\ & = \mathbf{b}_{t+1:t+k}'(\gamma_{t+1:t+k} - \hat{\gamma}_{t+1:t+k}) - \frac{1}{2}(\mathbf{h}_{t+1:t+k} - \hat{\mathbf{h}}_{t+1:t+k})'UK^{-1}GK^{-1}R'(\mathbf{h}_{t+1:t+k} - \hat{\mathbf{h}}_{t+1:t+k}) \\ & = \mathbf{b}_{t+1:t+k}'(\gamma_{t+1:t+k} - \hat{\gamma}_{t+1:t+k}) - \frac{1}{2}(\gamma_{t+1:t+k} - \hat{\gamma}_{t+1:t+k})G(\gamma_{t+1:t+k} - \hat{\gamma}_{t+1:t+k}) \\ & = -\frac{1}{2}(\hat{\mathbf{y}}_{t+1:t+k} - \gamma_{t+1:t+k})G(\hat{\mathbf{y}}_{t+1:t+k} - \gamma_{t+1:t+k}) \end{aligned}$$

onde  $\hat{\mathbf{y}}_{t+1:t+k} = \hat{\gamma}_{t+1:t+k} + G^{-1}\mathbf{b}_{t+1:t+k}$  tal que  $\hat{\mathbf{y}}_s = \hat{\gamma}_s + G^{-1}\mathbf{b}_s$ . Além disso:

$$\gamma_s = h_s + K_s^{-1}J_{s+1}h_{s+1}, \quad s = t+1, \dots, t+k \text{ com } J_{t+k+1} = 0$$

$$b_s = \hat{d}_s - J_s K_{t-1}^{-1}b_{s-1}, \quad s = t+2, \dots, t+k \text{ com } b_{t+1} = d_{t+1}$$

Dado  $h_s$ , temos:

$$\begin{aligned} \hat{y}_s &= h_s + K_s^{-1}J_{s+1}h_{s+1} + K_s^{-1}\nu_s \\ &= h_s + K_s^{-1}J_{s+1}(\alpha + \phi h_s \sigma_\eta \eta_s) + K_s^{-1}\nu_s \\ &= c_s + Z_s h_s + H_s \xi_s, \quad s = t+1, \dots, t+k \\ h_{s+1} &= \alpha + \phi h_s + L_s \xi_s, \quad s = t, t+1, \dots, t+k \end{aligned}$$

onde  $\xi \sim N(\mathbf{0}, \mathbf{I}_2)$ ,  $c_s = K_s^{-1}J_{s+1}\alpha$ ,  $Z_s = 1 + K_s^{-1}J_{s+1}\phi$ ,  $H_s = K_s^{-1}[1, J_{s+1}\sigma_\eta]$  e  $L_s = [0, \sigma_\eta]$ .

Por fim, vamos descrever o processo de atualização dos nós  $h_{k_l}$  para  $l = 2, \dots, K$ . Como a densidade condicional  $p(h_{k_l}|h_{k_l-1}, h_{k_l+1})$  não possui forma fechada, utilizaremos o algoritmo MH (2.3.2) com a proposta  $N(\alpha(1-\phi) + \phi(h_{k_l-1} + h_{k_l+1})/(1+\phi^2), \sigma_\eta^2/(1+\phi^2))$ . Considere  $h_{k_l}^p$  e  $h_{k_l}^{(i-1)}$  denotam o valor proposto e o valor da iteração anterior. Assim, a probabilidade de aceitação é dada por  $\alpha_{MH} = \min\{1, Q(h_{k_l}^p)/Q(h_{k_l}^{(i-1)})\}$ , onde  $Q(h_{k_l})$  é o produto das densidades condicionais  $y_{k_l-1}|z_{k_l-1}, y_{k_l-2}, h_{k_l-1}, h_{k_l} \sim N(\mu_{k_l-1}, V_{k_l-1})$  e  $y_{k_l}|z_{k_l}, y_{k_l-1}, h_{k_l+1}, h_{k_l} \sim N(\mu_{k_l}, V_{k_l})$ , com  $\mu_s$  e  $V_s$  definidos pelas equações (5.8) e (5.9) respectivamente, para  $s = k_l - 1$  e  $k_l$ .



# Capítulo 6

## Aplicação

Nesta capítulo são apresentados os resultados encontrados nas aplicações com dados simulados e com dados reais do modelo de VEM-GHt introduzido no capítulo 5.

### 6.1 Aplicação com dados simulados

Com a finalidade de ilustrar o método de estimação proposto, gerou-se 2000 observações do modelo VEM-GHt através das equações (5.1) e de (5.5) até (5.6) assumindo os seguintes valores para os parâmetros  $\beta_0 = 0.25$ ,  $\beta_1 = 0.03$ ,  $\beta_3 = -0.2$ ,  $\alpha = -0.008$ ,  $\phi = 0.95$ ,  $\sigma^2 = 0.0225$ ,  $\rho = -0.35$  e  $\nu = 10$ , estes valores são frequentemente encontrados na literatura.

As seguintes distribuições a priori são assumidas para os parâmetros:

$$\begin{array}{lll} \beta_0 \sim \mathcal{N}(0, 100) & \alpha|\tau^2 \sim \mathcal{N}(0, \tau^2/0.002) & \phi|\tau^2 \sim \mathcal{N}(0.96, 100)\mathbb{I}_{\{\phi < 1\}} \\ \beta_1 \sim \mathcal{N}(0.1, 100)\mathbb{I}_{\{|\beta_1| < 1\}} & \delta \sim \mathcal{N}(0, 1) & \tau^2 \sim \mathcal{IG}(2.5, 0.025) \\ \beta_2 \sim \mathcal{N}(-0.1, 100) & \varphi|\tau^2 \sim \mathcal{N}(-0.3, \tau^2/0.005) & \mathcal{G}(12, 0.8)\mathbb{I}_{\{\nu > 4\}} \end{array}$$

As distribuições a priori acima foram escolhidas com base em resultados empíricos da literatura.

Para a amostragem das log-volatilidades por blocos assume-se que o número de blocos igual a 30 de tal modo que cada bloco contenha 66  $h_t$ 's em média. Foram realizadas 50000 iterações descartando as 10000 amostras iniciais para cada parâmetro. A fim de reduzir a

autocorrelação entre os valores sucessivos simulados, somente a cada 10 valores da cadeia foram salvos, resultando em uma amostra de tamanho 4000.

Para checar a eficiência do algoritmo o fator de ineficiência foi calculado. O fator de ineficiência é definido por  $1 + 2 \sum_{s=1}^{\infty} \rho_s$  onde  $\rho_s$  é a correlação amostral no lag  $s$ . O fator de ineficiência avalia o esquema de amostragem pelo aumento da variância da amostra a posteriori quando comparada com a de amostras independentes, quanto mais longe de 1, menos eficiente é a amostragem. Com as 4000 amostras consideradas foram calculados a média a posteriori, o intervalo de credibilidade de 95% e a estatística Diagnóstico de Convergência (DC) (Geweke, 1992).

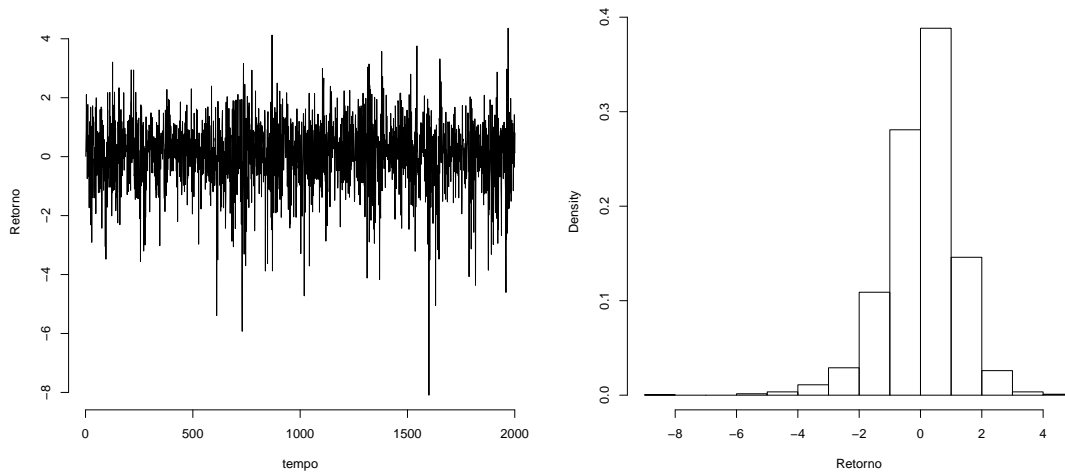


Figura 6.1: Retornos simulados do modelo VEM-HGt. O gráfico a esquerda mostra o plot da série dos dados e a direita tem-se o histograma dos retornos.

Tabela 6.1: Resultados da estimação do modelo VEM-HGt via MCMC.

Parâmetro	Verdadeiro	Media	Desvio	intervalo(95%)	Ineficiência	DC
$\beta_0$	0.25	0.281	0.085	[0.122, 0.465]	6.263	-0.116
$\beta_1$	0.03	0.026	0.022	[-0.017, 0.068]	1.288	0.948
$\beta_2$	-0.2	-0.250	0.094	[-0.445, -0.070]	5.378	0.008
$\alpha$	-0.008	-0.016	0.008	[-0.034, -0.003]	10.360	-0.938

$\phi$	0.95	0.921	0.024	[0.868, 0.961]	10.668	-1.014
$\sigma^2$	0.0225	0.033	0.010	[0.016, 0.055]	21.832	0.961
$\delta$	-0.5	-0.768	0.317	[-1.610, -0.340]	21.312	-0.322
$\rho$	-0.35	-0.235	0.099	[-0.427, -0.042]	7.022	0.522
$\nu$	10	12.443	2.812	[8.233, 19.552]	20.019	0.147

A tabela 6.1 mostra as médias a posteriori, os desvios padrões, o intervalo de credibilidade de 95% e os fatores de ineficiência.

Note que todos os intervalos de credibilidade de 95% contém os valores verdadeiros dos parâmetros (ver figura 6.3). No intervalo de credibilidade de  $\beta_1$  contém o valor 0 porém deve-se levar em consideração que o valor real escolhido para a simulação é proximo a 0. De acordo com a estatística DC a hipótese nula da sequencia de 4000 amostra ser estacionária é aceita em um nível de 5% ( $DC \in (-1.96, 1.96)$ ) para todos os parâmetros do modelo o que pode ser confirmado na figura 6.5 onde pode-se observar o rápido decaimento das autocorrelações.

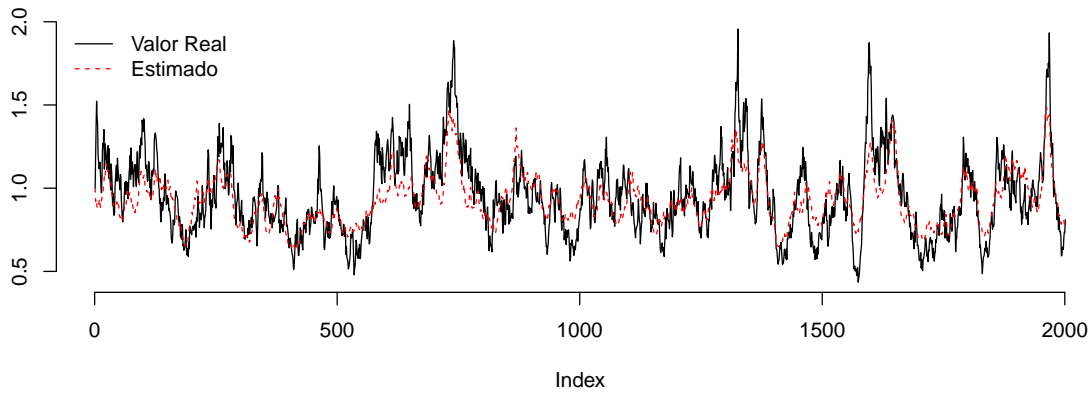


Figura 6.2: Volatilidade dos dados simulados  $e^{h_t/2}$

Como pode ser vista acima o valor médio estimado (linha pontilhada) da volatilidade segue bem o comportamento dos valores reais.

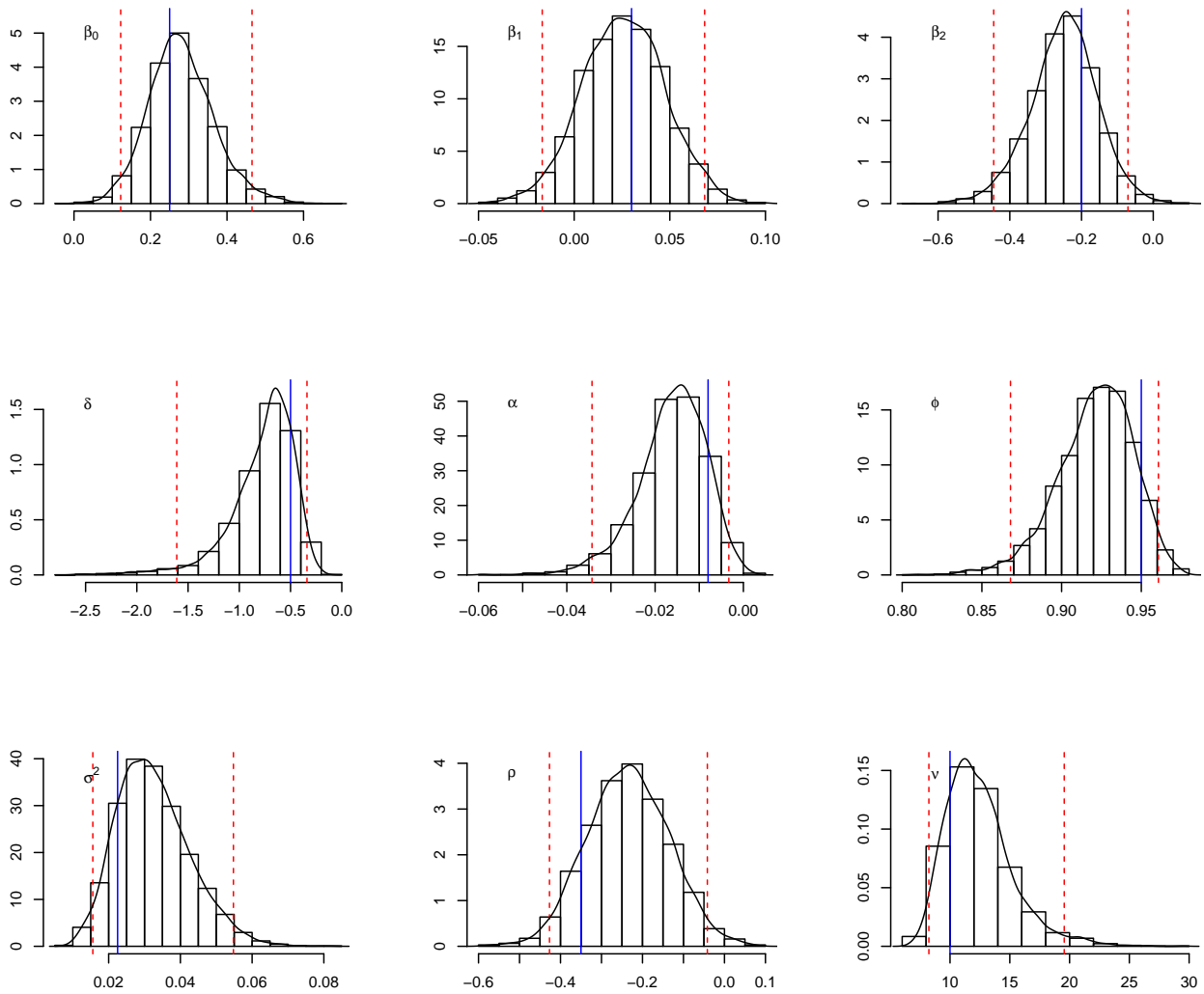


Figura 6.3: Histogramas e densidades estimadas a partir das saídas do MCMC. Linha pontilhada: intervalos de credibilidade de 95%. Linha solida: valor real

A figura 6.4 mostra as saídas do MCMC para cada um dos parâmetros e a figura 6.5 as respectivas funções de autorrelação, mostrando um rápido decrescimento das mesmas, o que é um indicador da eficiência do método de amostragem e da convergência.

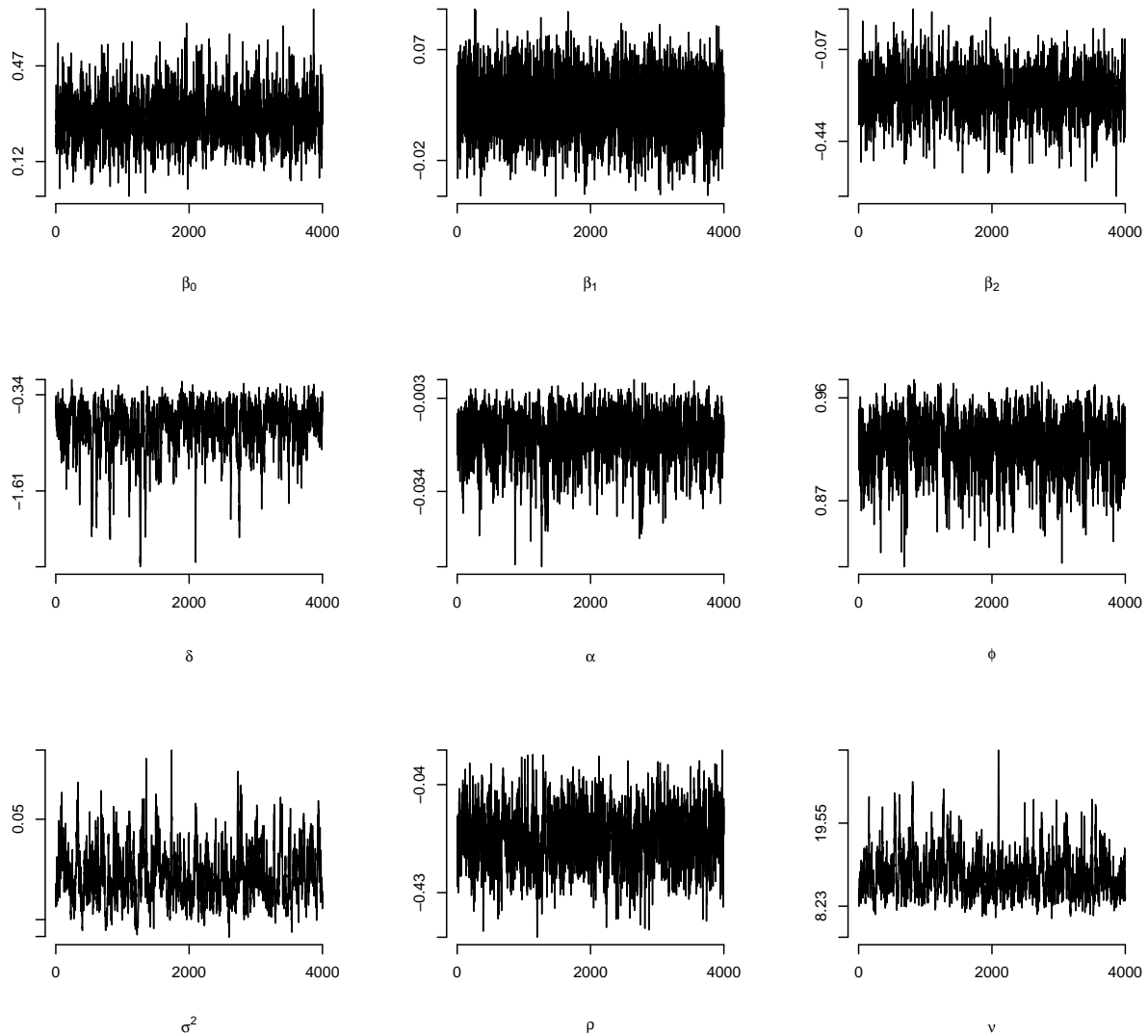


Figura 6.4: Saídas do MCMC

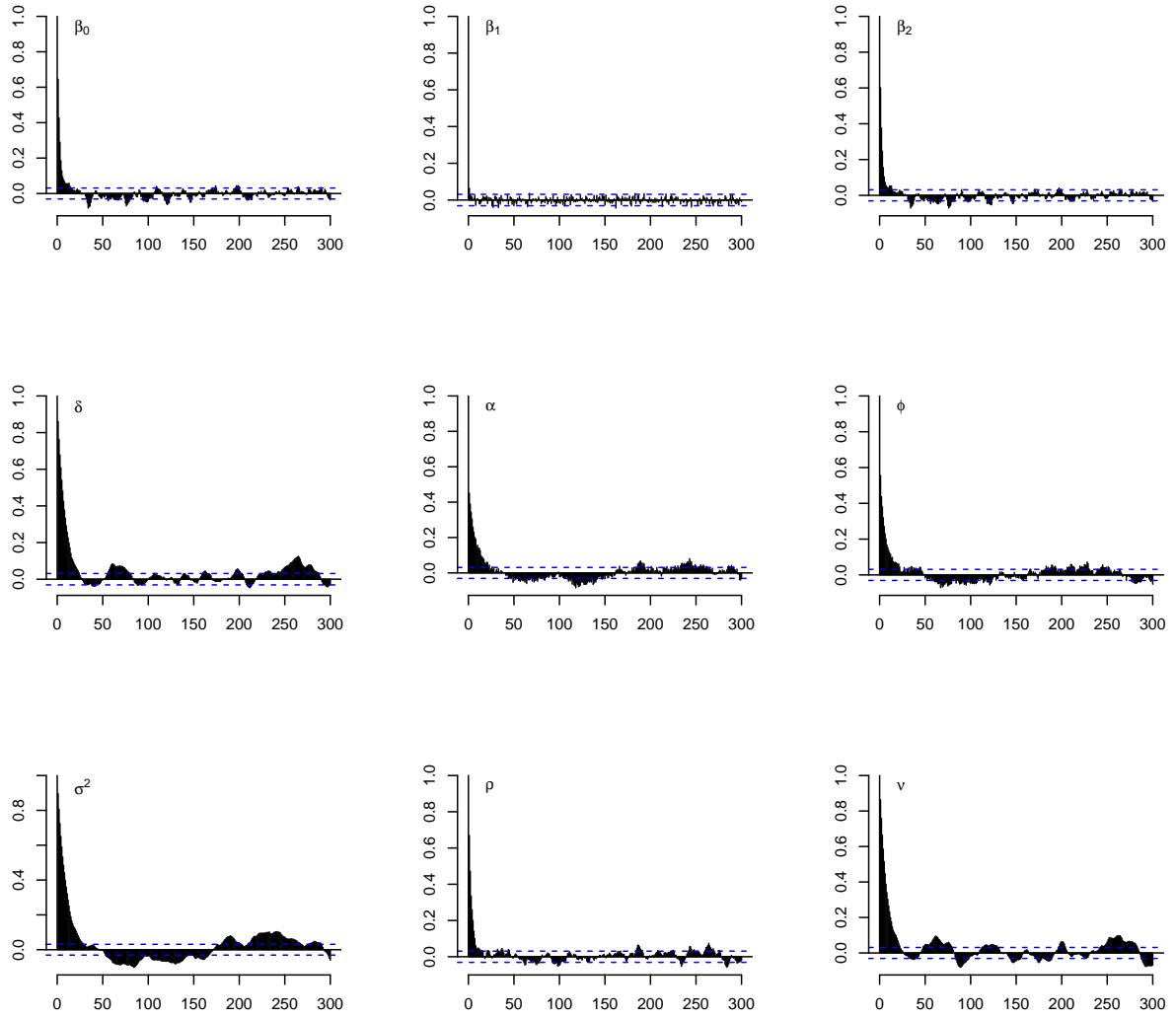


Figura 6.5: Funções de Autocorrelação

## 6.2 Aplicação com dados empíricos

Esta seção analisa os os preços de fechamento da Índice Bovespa (IBOVESPA). O IBOVESPA é o mais importante indicador do desempenho médio das cotações do mercado de ações brasileiro. O índice é composto por uma carteira teórica com as ações que representaram 80% do volume negociado nos últimos 12 meses e que foram negociadas pelo menos 80% dos dias de negociação. É revista trimestral, a fim de manter sua representatividade do volume negociado e em média os componentes do Ibovespa representam 70% do todas as ações valor negociado.

O período utilizado na análise contempla de 5 de Janeiro de 1998 a 3 de Outubro de 2005, sendo composto de 1917 observações. São utilizados os log-retornos percentuais,  $y_t = 100(\log P_t - \log P_{t-1})$ , onde  $P_t$  é o preço de fechamento no dia  $t$ .

Os retornos do IBOBESPA são apresentados na figura 6.6 como um gráfico série temporal e um histograma dos retornos. A tabela 6.2 fornece estatísticas descritivas para os retornos.

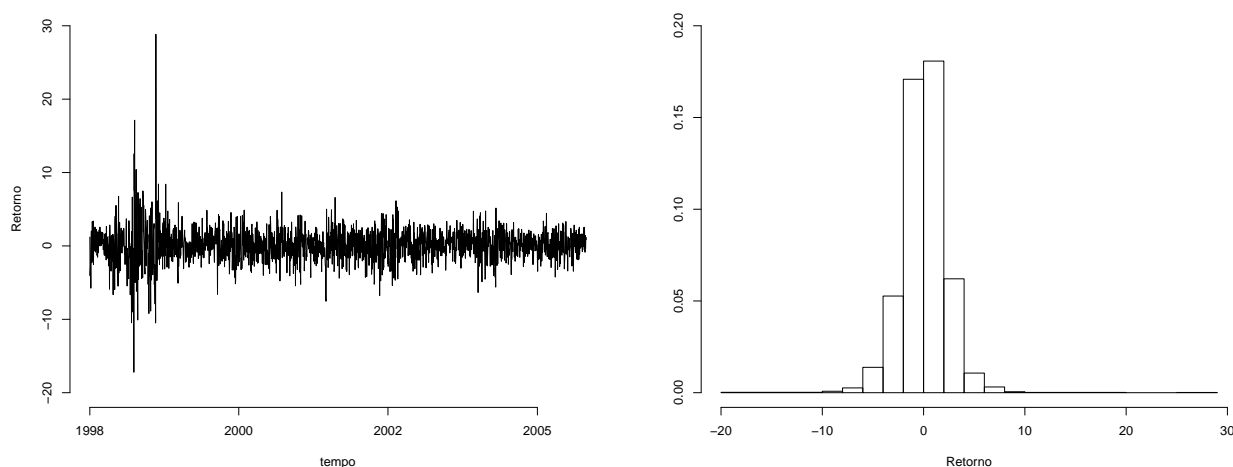


Figura 6.6: Retornos do IBOVESPA de 5 Janeiro de 1998 a 3 de setembro de 2005. O gráfico a esquerda mostra o plot da série dos retornos e a direita tem-se o histograma dos retornos.

Tabela 6.2: Estatísticas descritivas dos retornos do IBOVESPA.

Média	D.p.	Minimo	Máximo	Assimetria	Curtose
0.058	2.335	-17.208	28.830	0.834	19.165

Pode-se ver na figura 6.6 que o histograma apresenta uma leve assimetria (0.834) nas observações, e note também que os retornos possuem uma grande amplitude  $([-17.208, 28.830])$ . Algumas observações extremas podem ser observadas no plot da série de retornos, explicadas pela turbulência, devido a crises nas taxas de cambio, nos mercados financeiros russo e brasileiro que ocorreu em agosto de 1998 e janeiro de 1999 respectivamente, as quais contribuíram para a alta curtose (19.165) dos retornos do IBOVESPA. Assim provavelmente a suposição de normalidade não é adequada para os retornos. Portanto serão utilizados o modelo proposto e os processos de estimação apresentados no capítulo 5 para a modelagem desses dados e, além disso, será feita uma comparação com os modelos vistos em Abanto-Valle et al. (2011).

As prioris escolhidas para a estimação dos dados empíricos são as mesmas utilizadas para o caso com dados simulados.

Para a amostragem das log-volatilidades via amostragem por blocos, foi considerado o número de blocos igual a 30 de tal modo que cada bloco contenha 66  $h_t$ 's em média. Foram realizadas 50000 iterações descartando-se as 10000 amostras iniciais como burning para cada parâmetro e somente a cada 10 valores da cadeia foram salvos, resultando em uma amostra de tamanho 4000.

Da mesma forma que no caso com dados simulados o fator de ineficiência foi calculado para verificar a qualidade do ajuste.

Na tabela 6.3 são encontrados os resultados da estimação do modelo VEM-HGt e também dos modelos SVML-N e SVML-t encontrados em Abanto-Valle et al. (2011), os quais serão comparados entre si como pode ser visto mais adiante.

De acordo com a estatística DC a hipótese nula da sequência de 4000 amostra ser estacionária é aceita em um nível de 5% para todos os parâmetros do modelo o que pode ser confirmado na figura 6.9 onde pode-se observar o rápido decaimento das autocorrelações.



Tabela 6.3: Resultados da estimação dos retornos do IBOVESPA. Primeira Linha: Média a Posteriori. Segunda Linha: intervalo de credibilidade de 95% em parênteses. Terceira linha: estatística DC. Quarta linha: fator de ineficiência.

Parâmetro	SVML-N	SVML-t	VEM-HGt
$\beta_0$	0.1409	0.1801	0.151
	(0.0031,0.2820)	(0.0269,0.3388)	(0.018,0.292)
	0.61	-1.67	-1.38
	1.33	2.34	2.43
$\beta_1$	0.0299	0.0242	0.023
	(-0.0562,0.0239)	(-0.0219,0.0682)	(-0.022,0.069)
	-0.18	-0.18	-1.13
	1.49	1.60	1.91
$\beta_2$	-0.0179	-0.0343	-0.026
	(-0.0559,0.0193)	(-0.0894,0.0160)	(-0.070,0.014)
	0.69	0.66	1.21
	1.27	3.02	2.26
$\alpha$	0.0713	0.0407	0.049
	(0.0271, 0.1196)	(0.0147,0.0758)	(0.021,0.085)
	1.47	-1.49	1.65
	23.55	28.31	20.05
$\phi$	0.9368	0.9579	0.951
	(0.8940,0.9765)	(0.9184,0.9855)	(0.915,0.980)
	-1.33	1.47	-1.68
	25.29	36.05	25.00
$\sigma_\eta^2$	0.0708	0.0426	0.055
	(0.0250,0.1214)	(0.0146,0.0818)	(0.024,0.094)
	1.31	-1.46	1.74
	28.74	44.37	28.43
$\rho$	-0.3112	-0.3445	-0.350
	(-0.4677,-0.1774)	(-0.5319,-0.1779)	(-0.449,-0.206)
	1.36	-1.75	-0.830
	11.13	21.96	9.621
$\nu$	—	10.9988	18.715
	—	(6.9690,16.9087)	(10.30,29.56)
	—	1.32	0.156
	—	26.25	23.75
$\delta$	—	—	-0.381
	—	—	(-1.22,0.16)
	—	—	-1.79
	—	—	10.22

A estimativa  $\beta_2$ , a qual mede a relação entre retornos e volatilidade como também o efeito “feedback”, possui média negativa a posteriori. Embora o intervalo de credibilidade de  $\beta_2$  contenha 0, a distribuição a posteriori é primariamente localizada em domínio negativo como é mostrado na Tabela 6.4. Este resultado confirma os resultados anteriores encontrados na literatura que quando os investidores esperam altos níveis persistentes de volatilidade no futuro, eles exigem uma compensação desta na forma de maiores retornos esperados. O intervalo de credibilidade para  $\beta_1$  também contém 0. Uma vez que o intervalo de credibilidade contenha o 0, este coeficiente deve ser não significativo. Foi encontrado alto valor de  $\phi$  indicando alta persistência da volatilidade, além de ser encontrado valor negativo para  $\rho$  o que indica a presença de efeito de alavancagem para os retornos do IBOVESPA.

Tabela 6.4: Series de retornos do IBOVESPA:  $P(\beta_2 < 0)$  estimada.

	SVML-N	SVML-t.	VEM-HGt
$P(\beta_2 < 0)$	0.8295	0.9055	0.8948

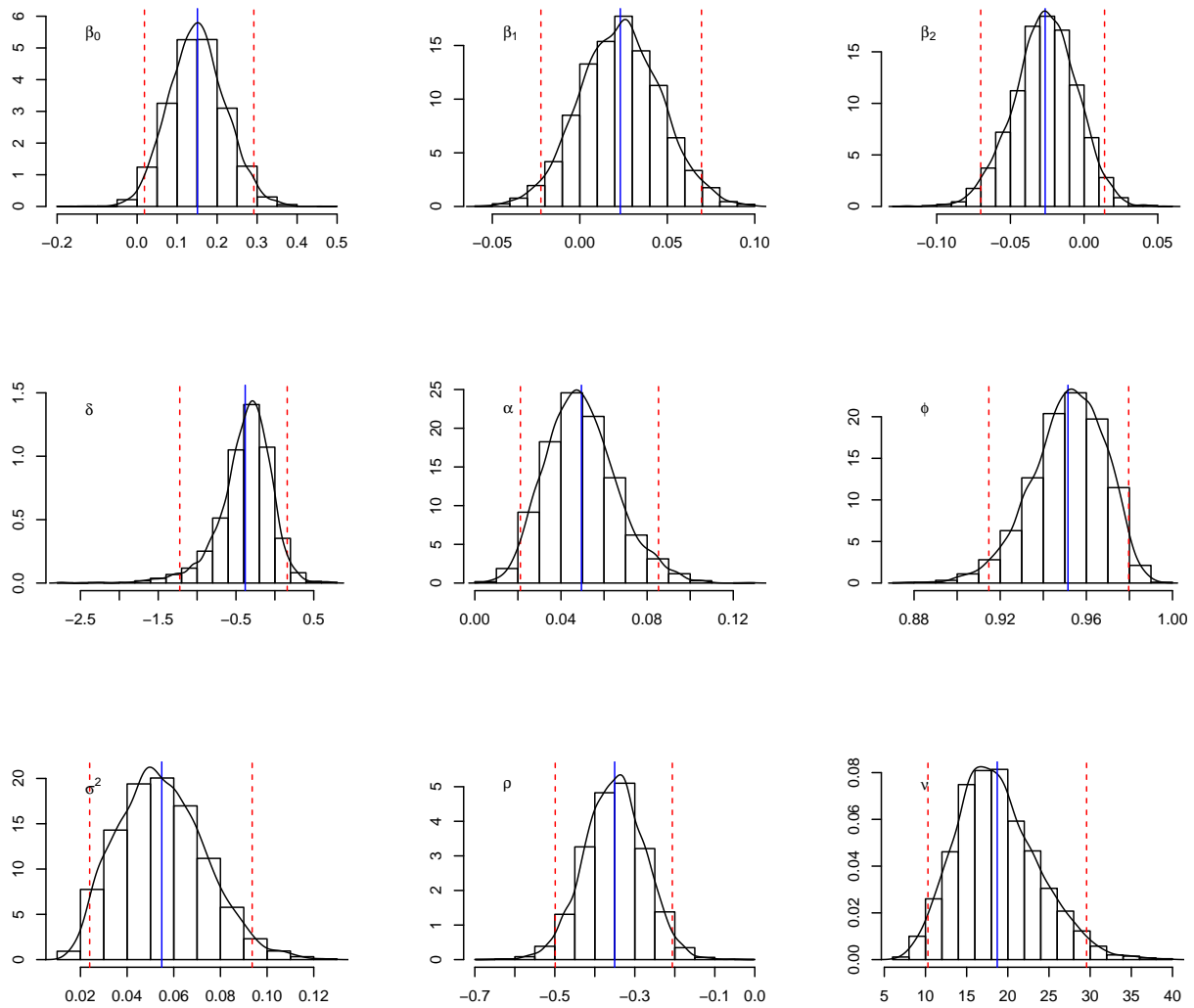


Figura 6.7: Histogramas e densidades estimadas a partir das saídas do MCMC. Linha pontilhada: intervalos de credibilidade de 95%. Linha solida: valor da média estimada

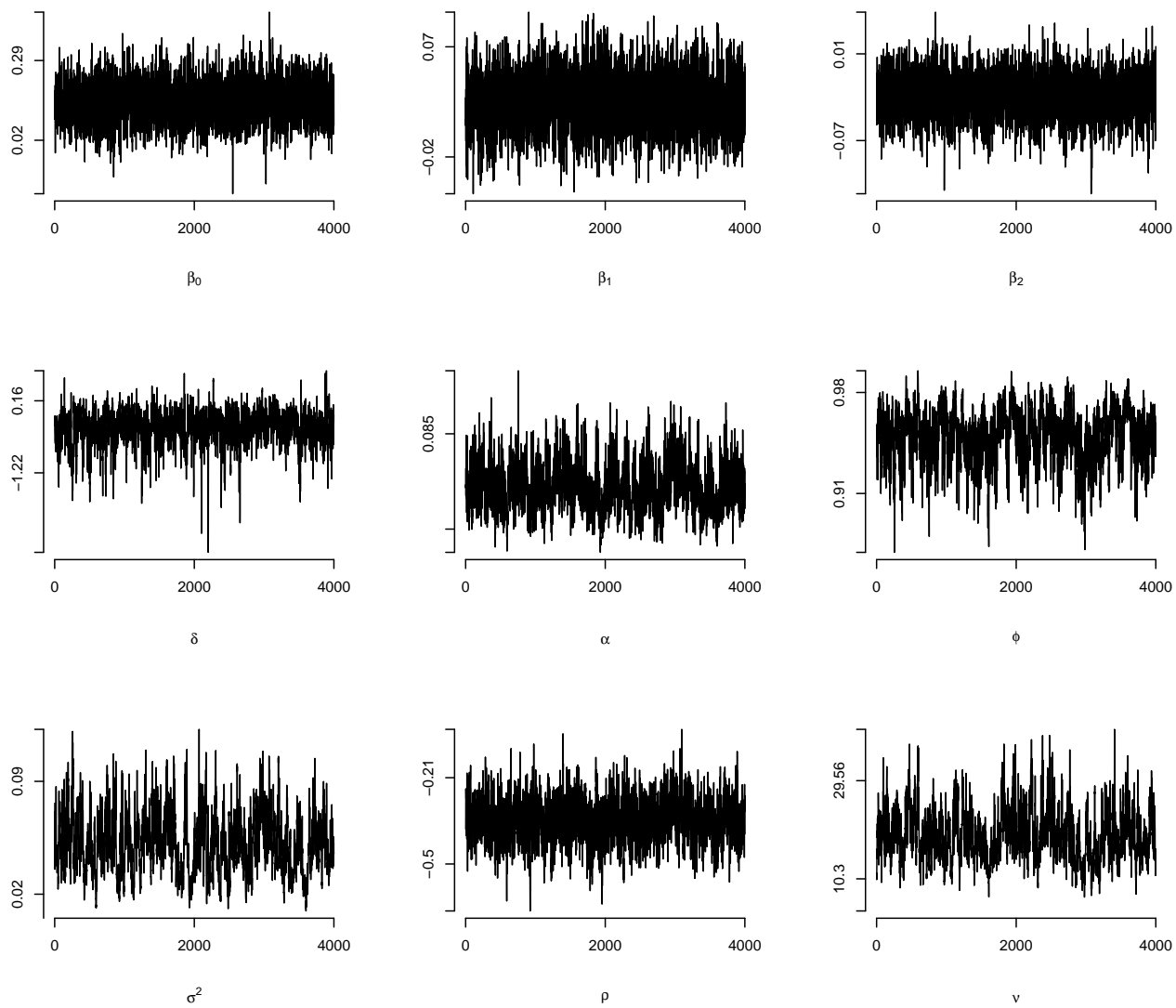


Figura 6.8: Saídas do MCMC

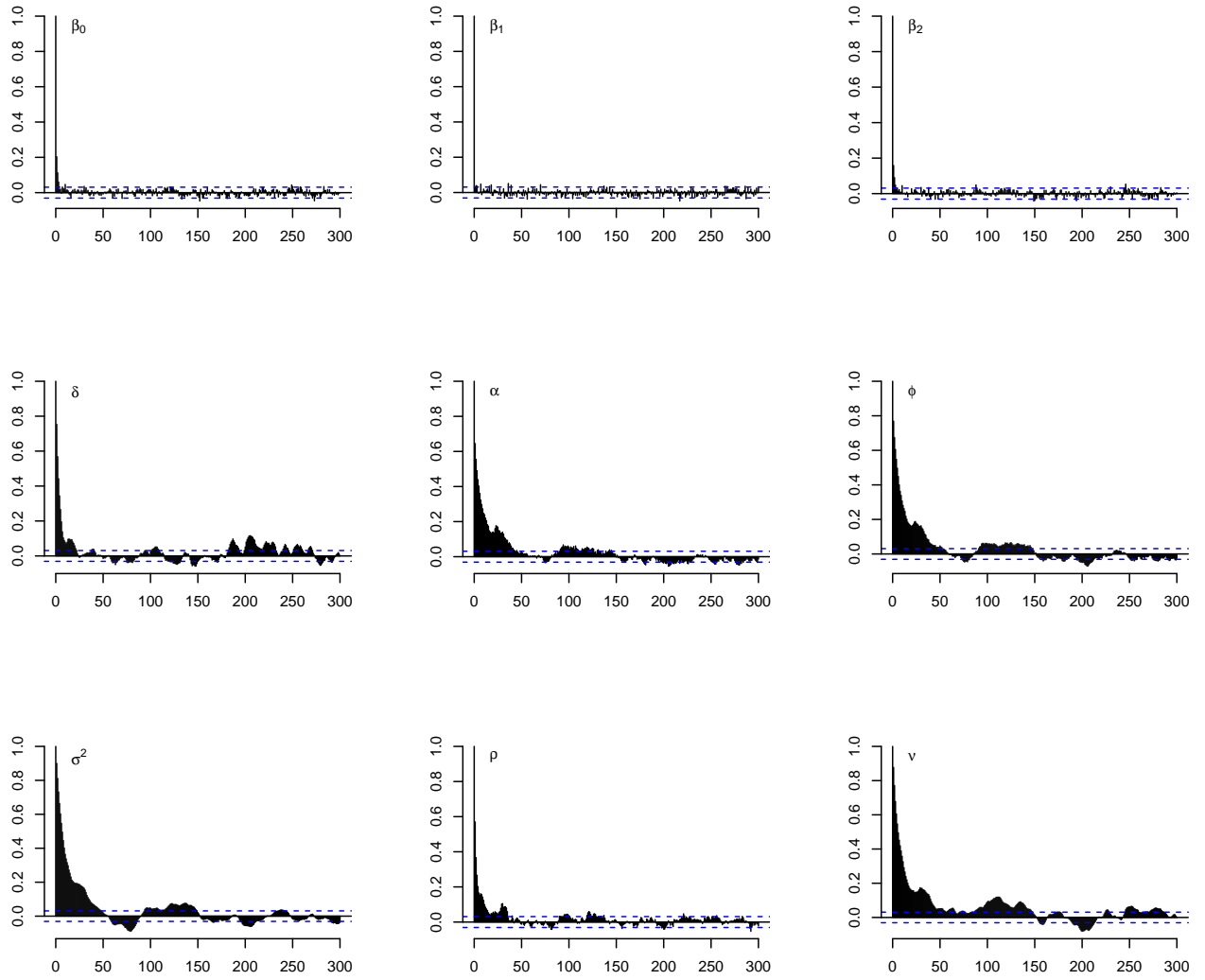


Figura 6.9: Funções de Autocorrelação

Aqui é realizada uma comparação do modelo proposto na dissertação (VEM-HGt) e os modelos SVML-N e SVML-t encontrados em Abanto-Valle et al. (2011).

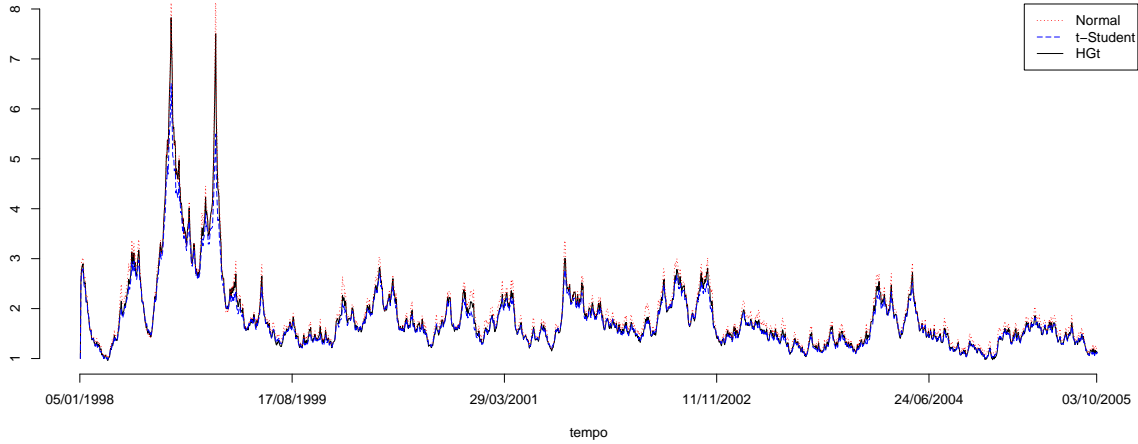


Figura 6.10: Média a posteriori suavizada de  $e^{h_t/2}$

Na figura 6.10 pode ser visto de forma que os retornos extremos nos períodos das crises da taxa de câmbio russa em agosto de 1998 e a brasileira em janeiro de 1999. Os modelos de caudas pesadas acomodam possíveis “outliers” de uma maneira um pouco diferente inflando a variância  $e^{h_t/2}$  de  $z_t e^{h_t/2}$ . Isto pode ter um substancial impacto em precificação de derivativos e em estratégia para alocação de ativos.

Tabela 6.5: Critério de Informação Preditiva Bayesiana (BPIC)

Modelo	BPIC	Ranking
SVML-N	8084.6	3
SVML-t	8081.8	2
VEM-HGt	8075.4	1

Para a comparação entre os modelos estimados, bondade de ajuste, é calculado o critério de informação preditiva Bayesiana BPIC. O BPIC mostra que o modelo proposto VEM-HGt é relativamente melhor que os demais modelos considerados, sugerindo que os retornos do IBOVESVA desvia da suposição de normalidade e que o modelo envolvendo simultaneamente assimetria e caudas pesadas ajusta melhor os dados.

# Capítulo 7

## Conclusões e Futuros Desenvolvimentos

Esta dissertação apresentou uma implementação Bayesiana robusta para estimação da volatilidade estocástica na média com erros correlacionados, confirmou a presença dos fatos estilizados conglomerados de volatilidade, caudas pesadas, assimetria negativa, os efeitos de alavancagem e “feedback”. Utilizando na distribuição condicional dos dados a distribuição hiperbólica generalizada t-Student assimétrica. Sobre o ponto de vista Bayesiano foi construído um algoritmo baseado em Simulação de Monte Carlo via Cadeias de Markov (MCMC) para estimar todos os parâmetros e as quantidades latentes no modelo proposto (VEM-HGt). O método de estimação proposto foi ilustrado através do uso de dados gerados artificialmente e uma aplicação empírica utilizando a série de retornos do IBOVESPA. A aplicação com dados simulados mostrou que método de estimação é eficiente para encontrar as quantidades desejadas e na aplicação empírica foram encontrados os fatos estilizados que retornos financeiros apresentam sendo que a estimativa  $\beta_2$ , a qual mede a relação entre retornos e volatilidade como também o efeito “feedback”, possui média negativa a posteriori confirmando que quando os investidores esperam altos níveis persistentes de volatilidade no futuro, os mesmos exigem uma compensação desta na forma de maiores retornos esperados. Foi encontrado alto valor de  $\phi$  indicando alta persistência da volatilidade além de ser encontrado valor negativo para  $\rho$  que indica a presença de efeito de alavancagem para os retornos do IBOVESPA.

Na comparação entre o modelo VEM-HGt mostrou-se, de acordo com o critério de comparações de modelos BPIC, superior ao modelo normal (SVML-N), mostrando que a suposição de normalidade não é adequada, e ao modelo t-Student (SVML-t), indicando que a modelagem de assimetria e caudas pesadas fornecem melhores resultados.

A discussão dos modelos de volatilidade estocástica na média com erros correlacionados é uma importante área para futuras pesquisas, pois os resultados claramente indicam que a presença da volatilidade na equação dos retornos é importante para entender o efeito “feedback”.

Existem algumas extensões imediatas ao presente trabalho. A primeira é incluir uma variável “threshold” da qual dependam os coeficientes dos parâmetros na equação da média e da volatilidade para estudar outras fontes de assimetria. A segunda é desenvolver métodos sequenciais para estimação “on-line” de parâmetros e quantidades latentes. Finalmente, usando justificativas econômicas e econométricas estender a classe de modelos introduzida a uma versão multivariada, pois o conhecimento da estrutura de correlação é vital em muitas aplicações financeiras, tais como gerenciamento e alocação de uma carteira de valores.



# Referências Bibliográficas

- Aas, k. e Haff, I. (2006) The generalized hyperbolic skew student's t-distribution. *Journal of Financial Econometrics*, **4**, 275–309.
- Abanto, A., Bandyopadhyay, D., Lachos, V. e Enriquez, I. (2010) Robust bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics & Data Analysis*, **54**, 2883–2898.
- Abanto-Valle, C., Migon, H. e Lachos, V. (2011) Stochastic volatility in mean models with scale mixtures of normal distributions and correlated erros: A bayesian approach. *Journal of Statistical Planning and Inference*, **141**, 1875–1887.
- Abramowitz, M. e Stegun, I. A. (1972) *Handbook of Mathematical Function*. New York:Dover.
- Ando, T. (2007) Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models. *Biometrika*, **94**, 443–458.
- (2009b) Bayesian portfolio selection using multifactor model and bayesian predictive information criterion. *International Journal of Forecasting*, **25**, 1925–1939.
- (2009c) Bayesian inference for hazard term structure with functional predictors using bayesian predictive information criterion. *Computational Statistical and Data Analysis*, **53**, 1925–1939.
- (2010) *Bayesian Model Selection and Statistical Modeling*. Chapman & Hall/CRC, 1 edn.

- Azzaline, A. e Capitanio, A. (2003) Distributions generated by perturbation of symmetric with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society, Series B* **65**, 579–602.
- Barndorff-Nielsen, O. (1977) Exponentially decreasing distributions for logarithm of particle size. *Proceedings of the Royal Society of London*, **A(353)**, 401–419.
- Bekaert, G. e Wu, G. (2000) Asymmetric volatility and risk in equity markets. *Review of Financial Economics*, **13**, 1–42.
- Black, F. (1976) Studies of stock price volatility changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistical Section*, **pp**, 177–181.
- Campbell, J. e Hentschel, L. (1992) No news is good news, an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, **31**, 281–318.
- Carnero, M., Peña, D. e Ruiz, E. (2004) Persistence and kurtosis in garch and stochastic volatility models. *Journal of Financial Econometrics*, **2**, 319–342.
- Chib, S. e Greenberg, E. (1995) Understanding the metropolis-hastings algorithm. *The American Statistician*, **49**, 327–335.
- Chib, S. e Jeliazkov, I. (2004) Accept-reject metropolis-hastings sampling and marginal likelihood estimation. *Relatório técnico*, Washington University.
- Cowles, M. K. e Carlin, B. P. (1996) Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, **91**, 883–904.
- Fernandez, C. e Steel, M. (1998) On bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359–371.
- French, K., Schert, W. e Stambough, R. (1987) Expected stock return and volatility. *Journal of Financial Economics*, **19**, 3–29.
- Gamerman, D. e Lopes, H. (2006) *Monte Carlo Markov Chain: Stochastic Simulation for Bayesian Inference*. New York:Chapman & Hall, 2 edn.

- Gelman, A., Carlin, J. B., Stern, H. S. e Rubin, D. B. (2003) *Bayesian Data Analysis*. Boca Raton: Chapman and Hall/CRC, 2 edn.
- Gelman, A. e Rubin, D. B. (1992) Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, **7**, 457–511.
- Geweke, J. (1992) *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, vol. 4. In J. M. Bernardo, J. O. Berger, A. P. David e A. F. M. Smith. Bayesian Statistics.
- Geyer, C. (1992) Pratical markov chain monte carlo. *Statistical Science*, **7**, 473–483.
- Gilks, W., Richardson, S. e Spiegelhalter, D. (1996) *Markov chain Monte Carlo in pratice*. London:Chapman & Hall.
- Hansen, B. (1994) Autoregressive conditiononal density estimation. *International Economic Review*, **35**, 705–730.
- Harrison, P. e Stevens (1976) Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society Series B*, **38**, 205–247.
- Hastings, W. (1970) Monte-carlo sampling methods using markov chains and their applications. *Biometrika*, **57**, 97–109.
- Heidelbelger, P. e Welch, P. (1983) Simulation run lenght control in the presence of a initial transient. *Operations Research*, **31**, 1109–1144.
- Koopman, S. (1993) Disturbance smoothers for state space models. *Biometrika*, **80**, 117–126.
- Koopman, S. e Uspensky, E. (2002) The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of Applied Econometrics*, **17**, 667–689.
- Madras, N. e Sokal, A. D. (1998) The pivot algorithm: a highly efficient monte carlo method for the self-avoiding walk. *Journal of Statistical Physics*, **50**, 109–186.

- Melino, A. e Turnbull, S. (1990) The variation of certain speculative prices. *Journal of Business*, **36**, 314–419.
- 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**, 1087–1092.
- Omori, Y. e Watanabe, T. (2008) Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics & Data Analysis*, **52**, 2892–2910.
- Prause, K. (1999) *The generalized hyperbolic models: estimation, financial derivatives and risk measurement*. Tese de Doutorado, University of Freiburg.
- Raftery, A. E. e Lewis, S. (1992) *How many iterations in the Gibbs sampler?* In J. M. Bernardo, J. O. Berger, A. P. David e A. F. M. Smith, vol. 4. Bayesian Statistics.
- Shephard, N. e Pitt, M. (1997) Likelihood analysis of non-gaussian measurements time series. *Biometrika*, **84**, 653–667.
- Tauchen, G. e Pitts, M. (1983) The price variability-volume relationships in speculative markets. *Econometrica*, **51**, 485–506.
- Taylor, S. (1982) Financial returns modelled by the product of two stochastic process—a study of the daily sugar prices 1961–75. in anderson, o. (ed.). *Time Series and Analyses: Theory and Practice, North -Holland, Amsterdam*, **1**, 203–226.
- (1986) Modelling financial times series. *Wiley, Chichester*.
- Tierney, L. (1994) Markov chains for exploring posterior distributions. *Annals of Statistics*, **22**, 1701–1728.
- Vitor, M. M. T. (2006) *Distribuições Hiperbólicas Generalizadas: Aplicação ao Mercado Português*. Dissertação de Mestrado, Universidade da Madeira - Departamento de Matemática e Engenharias.

- Wang, C. (2005) *On the Numerics of Estimation Generalized Hyperbolic Distributions*.  
Dissertação de Mestrado, Humboldt-Universität zu Berlin.
- Watanabe, T. e Omori, Y. (2004) A multi-move sampler for estimate non-gaussian time series model: Comments on shepard and pitt (1997). *Biometrika*, **91**, 246–248.
- West, M. e Harrison, P. (1997) *Bayesian Forecasting and Dynamic Models*. New York:Springer-Verlag, 2 edn.