

# UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

## Métodos de Monte Carlo Hamiltoniano aplicados em modelos GARCH

Cleber Martins Xavier

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em  
Estatística (PIPGEs)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Cleber Martins Xavier**

# Métodos de Monte Carlo Hamiltoniano aplicados em modelos GARCH

Tese apresentada ao Instituto de Ciências  
Matemáticas e de Computação – ICMC-USP e  
ao Departamento de Estatística – DEs-UFSCar,  
como parte dos requisitos para obtenção do título  
de Doutor em Estatística – Programa Interinstitucional  
de Pós-Graduação em Estatística. *VERSÃO  
REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Ricardo Sandes Ehlers

Coorientador: Prof. Dr. Marinho Gomes de  
Andrade Filho

**USP – São Carlos**  
**Maio de 2019**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

M386m      Martins Xavier, Cleber  
                Métodos de Monte Carlo Hamiltoniano aplicados em  
                modelos GARCH / Cleber Martins Xavier; orientador  
                Ricardo Sandes Ehlers ; coorientador Marinho Gomes  
                de Andrade Filho . -- São Carlos, 2019.  
                87 p.

Tese (Doutorado - Programa Interinstitucional de  
Pós-graduação em Estatística) -- Instituto de Ciências  
Matemáticas e de Computação, Universidade de São  
Paulo, 2019.

1. Séries Temporais. 2. Modelos GARCH. 3.  
Inferência Bayesiana. 4. Metrópolis-Hastings. 5.  
Hamiltoniano. I. , Ricardo Sandes Ehlers, orient.  
II. , Marinho Gomes de Andrade Filho, coorient.  
III. Título.

**Cleber Martins Xavier**

# Hamiltonian Monte Carlo methods in GARCH models

Thesis submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar – in accordance with the requirements of the Statistics Interagency Graduate Program, for the degree of Doctor in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Ricardo Sandes Ehlers

Co-advisor: Prof. Dr. Marinho Gomes de Andrade Filho

**USP – São Carlos**  
**May 2019**



*Aos meus pais, minha esposa e minha filha.*



# **AGRADECIMENTOS**

---

---

Gostaria de agradecer aos meus pais e minha família por todo apoio emocional durante minha vida acadêmica.

Agradecer a minha esposa por estar sempre me ajudando, principalmente nos momentos mais difíceis durante o meu doutorado.

Ao Professor Ricardo Ehlers pela paciência e orientação durante a construção desse trabalho.

Ao corpo Docente do departamento de estatística da UFSCar e USP por todo conhecimento ministrado durante o curso.

Aos meus colegas de doutorado, principalmente Vanessa, George, José Cleto, Daiane e Danila por toda companhia e diversão durante esta caminhada.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001



*“A ausência da evidência  
não significa evidência da ausência.*

*(The absence of evidence  
is not the evidence of absence.)”*

*Carl Sagan*



# RESUMO

XAVIER, C. M. **Métodos de Monte Carlo Hamiltoniano aplicados em modelos GARCH.** 2019. 87 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Uma das informações mais importantes no mercado financeiro é a variabilidade de um ativo. Diversos modelos foram propostos na literatura com o intuito de avaliar este fenômeno. Dentre eles podemos destacar os modelos GARCH. Este trabalho propõe o uso do método Monte Carlo Hamiltoniano (HMC) para a estimação dos parâmetros do modelo GARCH univariado e multivariado. Estudos de simulação são realizados e as estimativas comparadas com o método de estimação Metropolis-Hastings presente no pacote *BayesDccGarch*. Além disso, compara-se os resultados do método HMC com a metodologia adotada no pacote *rstan*. Por fim, é realizado uma aplicação a dados reais utilizando o DCC-GARCH bivariado e os métodos de estimação HMC e Metropolis-Hastings.

**Palavras-chave:** volatilidade, modelos GARCH, MCMC, Monte Carlo Hamiltoniano.



# ABSTRACT

XAVIER, C. M. **Hamiltonian Monte Carlo methods in GARCH models.** 2019. 87 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

One of the most important informations in financial market is variability of an asset. Several models have been proposed in literature with a view of to evaluate this phenomenon. Among them we have the GARCH models. This paper use Hamiltonian Monte Carlo (HMC) methods for estimation of parameters univariate and multivariate GARCH models. Simulation studies are performed and the estimatives compared with Metropolis-Hastings methods of the BayesDccGarch package. Also, we compared the results of HMC method with the methodology present in rstan package. Finally, a application with real data is performed using bivariate DCC-GARCH and the methods of estimation HMC and Metropolis-Hastings.

**Keywords:** volatility, GARCH models, MCMC, Hamiltonian Monte Carlo..



---

# LISTA DE ILUSTRAÇÕES

---

Figura 1 – Gráfico comparando as estimativas para os algoritmos HMC e MH em modelo GARCH(1,1).	45
Figura 2 – Gráfico comparando os valores do erro naive para os algoritmos HMC e MH em modelo GARCH(1,1)	45
Figura 3 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 1000 simulações.	46
Figura 4 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 1000 simulações.	46
Figura 5 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 3000 simulações.	47
Figura 6 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 3000 simulações.	47
Figura 7 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 5000 simulações.	48
Figura 8 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 5000 simulações.	48
Figura 9 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 10000 simulações.	49
Figura 10 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 10000 simulações.	49
Figura 11 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 15000 simulações.	50
Figura 12 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 15000 simulações.	50
Figura 13 – Boxplot utilizando o algoritmo HMC para $\omega = 0,9$ , $\alpha = 0,5$ e $\beta = 0,2$ em modelos GARCH(1,1) univariados.	51
Figura 14 – Boxplot utilizando o pacote BayesDccGarch para $\omega = 0,9$ , $\alpha = 0,5$ e $\beta = 0,2$ em modelos GARCH(1,1) univariados.	53
Figura 15 – Boxplot utilizando o algoritmo HMC para $\omega = 0,9$ , $\alpha = 0,5$ e $\beta = 0,4$ em modelos GARCH(1,1) univariados.	53
Figura 16 – Boxplot utilizando o pacote BayesDccGarch para $\omega = 0,9$ , $\alpha = 0,5$ e $\beta = 0,4$ em modelos GARCH(1,1) univariados.	54

Figura 17 – Boxplot utilizando o algoritmo HMC for $\omega = 0.9$ , $\alpha = 0.5$ e $\beta = 0.49$ em modelos GARCH(1,1) univariados. . . . .	54
Figura 18 – Boxplot utilizando pacote BayesDccGarch for $\omega = 0.9$ , $\alpha = 0.5$ e $\beta = 0.49$ em modelos GARCH(1,1) univariados. . . . .	56
Figura 19 – Boxplot utilizando o algoritmo HMC for $\omega = 0.1$ , $\alpha = 0.09$ e $\beta = 0.9$ em modelos GARCH(1,1) univariados. . . . .	56
Figura 20 – Boxplot utilizando o pacote BayesDccGarch for $\omega = 0.1$ , $\alpha = 0.09$ e $\beta = 0.9$ em modelos GARCH(1,1) univariados. . . . .	58
Figura 21 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 1000 simulações. . . . .	60
Figura 22 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo RSTAN com 1000 simulações. . . . .	61
Figura 23 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 5000 simulações. . . . .	61
Figura 24 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo RSTAN com 5000 simulações. . . . .	62
Figura 25 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 10000 simulações. . . . .	62
Figura 26 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo RSTAN com 10000 simulações. . . . .	64
Figura 27 – Boxplot utilizando HMC para $\omega = 0.1$ , $\alpha = 0.09$ e $\beta = 0.9$ em modelos GARCH(1,1) univariados. . . . .	66
Figura 28 – Boxplot utilizando pacote RSTAN para $\omega = 0.1$ , $\alpha = 0.09$ e $\beta = 0.9$ em modelos GARCH(1,1) univariados. . . . .	66
Figura 29 – Gráfico do traço das Cadeias de Markov e das autocorrelações para o algo- ritmo HMC com 100 mil simulações. . . . .	71
Figura 30 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o método presente no BayesDccGarch com 100 mil simulações. . . . .	72
Figura 31 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o método HMC com 100 mil simulações. . . . .	73
Figura 32 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o método HMC com 100 mil simulações. . . . .	74
Figura 33 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o método presente no BayesDccGarch com 100 mil simulações. . . . .	75

---

## LISTA DE ALGORITMOS

---

Algoritmo 1 – Algoritmo Metropolis-Hastings . . . . .	33
Algoritmo 2 – Algoritmo Monte Carlo Hamiltoniano - HMC . . . . .	37



# LISTA DE TABELAS

---

---

Tabela 1 – Valores médios das estimativas, viés e EQM para $\omega = 0,9$ , $\alpha = 0,5$ e $\beta = 0,2$ considerando os algoritmos HMC e MH. . . . .	52
Tabela 2 – Valores médios das estimativas, viés e EQM para $\omega = 0,9$ , $\alpha = 0,5$ e $\beta = 0,4$ considerando os algoritmo HMC e MH. . . . .	55
Tabela 3 – Valores médios das estimativas, viés e EQM para $\omega = 0,9$ , $\alpha = 0,5$ e $\beta = 0,49$ considerando os algoritmos HMC e MH. . . . .	57
Tabela 4 – Valores médios das estimativas, viés e EQM para $\omega = 0,1$ , $\alpha = 0,09$ e $\beta = 0,9$ considerando os algoritmos HMC e MH. . . . .	59
Tabela 5 – Valores médios das estimativas para $\omega = 0,1$ , $\alpha = 0,09$ e $\beta = 0,9$ considerando os algoritmos HMC e o RSTAN. . . . .	63
Tabela 6 – Valores médios das estimativas, viés e EQM para $\omega = 0,1$ , $\alpha = 0,09$ e $\beta = 0,9$ considerando os algoritmos HMC e RSTAN. . . . .	65
Tabela 7 – Estimativas para o DCC-GARCH(1,1) considerando os métodos HMC e o presente no pacote <i>BayesDccGarch</i> com 100 mil iterações; . . . . .	70
Tabela 8 – Estimativas para o DCC-GARCH(1,1) considerando os métodos HMC e o presente no pacote BayesDccGarch com 100 mil iterações; . . . . .	76



# SUMÁRIO

---

---

1	INTRODUÇÃO . . . . .	23
1.1	Modelos de volatilidade . . . . .	24
1.1.1	Modelos <i>GARCH univariados</i> . . . . .	24
1.1.2	Modelos <i>GARCH multivariados</i> . . . . .	26
2	TÉCNICAS BAYESIANAS DE ESTIMAÇÃO . . . . .	31
2.1	O princípio MCMC . . . . .	31
2.2	Monte Carlo via cadeias de Markov . . . . .	32
2.3	Metropolis-Hastings . . . . .	33
2.4	Monte Carlo Hamiltoniano . . . . .	34
2.4.1	<i>Dinâmica Hamiltoniana</i> . . . . .	35
2.4.2	<i>Algoritmo HMC</i> . . . . .	35
2.4.3	<i>Monte Carlo Hamiltoniano em variedade Riemanniana (RMHMC)</i> . . . . .	38
2.4.4	<i>STAN</i> . . . . .	39
3	MONTE CARLO HAMILTONIANO EM MODELOS GARCH UNIVARIADOS . . . . .	41
3.1	Introdução . . . . .	41
3.2	<i>GARCH(1,1) e o HMC</i> . . . . .	43
3.2.1	<i>Caso com uma série temporal de tamanho 1000</i> . . . . .	44
3.2.2	<i>Caso utilizando 500 séries temporais</i> . . . . .	50
3.2.3	<i>Primeiro caso : <math>\alpha + \beta = 0,7</math></i> . . . . .	51
3.2.4	<i>Segundo caso : <math>\alpha + \beta = 0,9</math></i> . . . . .	53
3.2.5	<i>Terceiro caso : <math>\alpha + \beta = 0,99</math></i> . . . . .	54
3.2.6	<i>Quarto caso: series de tamanho 1500 e <math>\alpha + \beta = 0,99</math></i> . . . . .	56
3.2.7	<i>HMC e o RSTAN</i> . . . . .	58
3.2.8	<i>HMC e o RSTAN usando 500 séries</i> . . . . .	64
4	MONTE CARLO HAMILTONIANO EM MODELOS GARCH MULTIVARIADOS . . . . .	67
4.1	<i>DCC-GARCH e o HMC</i> . . . . .	67
4.1.1	<i>Caso Bivariado</i> . . . . .	69
4.1.2	<i>Aplicação à Dados Reais</i> . . . . .	73

5	<b>CONCLUSÃO</b>	77
5.1	Trabalhos futuros	78
 <b>REFERÊNCIAS</b>		79
 <b>APÊNDICE A</b>		
DERIVADAS DA DISTRIBUIÇÃO A POSTERIORI PARA O MODELO DCC-GARCH		83
A.1	Derivadas DCC-GARCH	83
A.2	Derivadas caso com duas séries temporais	84



# INTRODUÇÃO

---



---

A volatilidade é muitas vezes modelada como o desvio padrão condicional aos retornos financeiros através das informações históricas e das volatilidades agrupadas o que sugere que os desvios padrão condicionais estão mudando continuamente de forma parcialmente previsível. Se os retornos tem sido grande nos últimos dias devido à excitação do mercado, então há razão para acreditar que a distribuição a partir da qual o retorno de amanhã é estimado deve ter uma grande variação. Neste sentido, o estudo de modelos temporais utilizados para avaliar a volatilidade é de grande utilidade no mercado financeiro.

Em séries financeiras é comum observar que a volatilidade em ativos financeiros mudam em determinados períodos onde são apresentados geralmente em grupos sugerindo alguma dependência no tempo. A previsão da volatilidade tem uma gama de aplicações em finanças, por exemplo, na otimização de portfólios, gestão de riscos, alocação de ativos, preços de ativos, etc. ([VIRBICKAITE; AUSÍN; GALEANO, 2015](#)).

Para estudar este tipo de comportamento utiliza-se os modelos heterocedásticos condicionais ARCH proposto por [Engle \(1982\)](#), os modelos GARCH proposto por [Bollerslev \(1986\)](#) e os modelos de volatilidade estocástica proposto por [Taylor \(1982\)](#). Diversos artigos sobre o tema vem sendo estudados na literatura propondo inovações como extensões e generalizações destes modelos. Além dessas linhas de pesquisas, alguns pesquisadores tentam abordar dois temas de grande relevância para o estudo da volatilidade: procurar uma melhor forma de especificação dos erros e selecionar uma abordagem que seja mais eficiente para realizar inferência e predição.

Na literatura, existem duas formas distintas de abordar este tipo de problema: a inferência frequentista e a inferência Bayesiana. Na inferência frequentista estima-se os parâmetros através da função de verossimilhança, enquanto que na inferência Bayesiana esta estimativa é realizada utilizando a distribuição a posteriori dos parâmetros.

Neste trabalho o foco principal é estudar novas formas de estimativa para realizar inferência em modelos GARCH. Diversos algoritmos foram propostos quando utiliza-se a abordagem

Bayesiana. Um dos mais conhecido é o Metropolis-Hastings. Todavia, com o advento da evolução computacional novos algoritmos de estimação estão sendo propostos como é o caso, por exemplo, do método Monte Carlo Hamiltoniano (HMC). Ao propor a utilização do método Monte Carlo Hamiltoniano como ferramenta de estimação, um estudo comparativo entre estes algoritmos de estimação é realizado.

Na [Seção 1.1](#) apresenta-se os modelos ARCH e os modelos GARCH. Neste trabalho devido a complexibilidade dos modelos GARCH utiliza-se de abordagens conhecidas como MCMC (Monte Carlo via Cadeia Markov) para realizar estimação, inferências e predição. No [Capítulo 2](#) alguns conceitos fundamentais sobre Cadeias de Markov e algoritmos de estimação são apresentados. No [Capítulo 3](#) é realizado um estudo de simulação em modelos GARCH univariados utilizando o método Monte Carlo Hamiltoniano e o Metropolis-Hastings. No [Capítulo 4](#) apresenta-se um estudo de simulação com o método Monte Carlo Hamiltoniano empregado em modelos GARCH multivariados. Além disso, uma aplicação a dados financeiros é realizada. Por fim, no [Capítulo 5](#) tem-se a conclusão e apresentação de propostas para continuidade do trabalho.

## 1.1 Modelos de volatilidade

Neste capítulo apresenta-se os modelos GARCH (Generalized Autoregressive Conditional Heteroscedasticity) um dos principais modelos utilizados para a modelagem estatística da volatilidade. Na [Subseção 1.1.1](#) tem-se o modelo de forma univariada. Já na [Subseção 1.1.2](#) define-se uma extensão desse modelo, os modelos GARCH multivariados.

### 1.1.1 Modelos GARCH univariados

Seja  $\mathbf{y} = \{y_1, \dots, y_n\}$  uma série temporal de tamanho  $n$ . [Engle \(1982\)](#) propôs o modelo  $ARCH(p)$  que assume que a variância conditional no tempo  $t$  depende de uma função linear do quadrado das  $p$  observações passadas da seguinte forma

$$\begin{aligned} y_t &= \varepsilon_t \sqrt{h_t}, \quad \varepsilon \sim \mathcal{D}(0, 1) \\ h_t &= w + \sum_{i=1}^p \alpha_i y_{t-i}^2 \end{aligned}$$

em que  $\omega > 0$ ,  $\alpha_i \geq 0$  para  $i = 1, \dots, p$  e  $\varepsilon_t$  é uma variável aleatória independente e identicamente distribuída e  $\mathcal{D}(0, 1)$  uma distribuição de probabilidades com média zero e variância um. Contudo, [Bollerslev \(1986\)](#) propôs uma generalização deste modelo, considerando que a volatilidade  $h_t$  depende tanto dos valores passados dos retornos quanto dos valores passados da volatilidade. O modelo  $GARCH(p, q)$  é definido por

$$y_t = \varepsilon_t \sqrt{h_t}, \quad \varepsilon_t \sim \mathcal{D}(0, 1) \tag{1.1}$$

$$h_t = w + \sum_{i=1}^p \alpha_i y_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} \tag{1.2}$$

em que as restrições de positividade e estacionariedade de  $h_t$  são  $w > 0$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, p$ ,  $\beta_j \geq 0$ ,  $j = 1, \dots, q$  e  $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ . Quando  $p = q = 1$  tem-se o modelo *GARCH*(1,1) e a variância condicional do modelo é equivalente a um modelo *ARCH*( $\infty$ ), mas com diferentes efeitos ARCH, ou seja,

$$\begin{aligned} h_t &= w + \sum_{k=1}^{\infty} \alpha_1 \beta_1^{k-1} y_{t-k}^2 \\ &= w + \alpha_1 y_{t-1}^2 + \alpha_1 \beta_1 y_{t-2}^2 + \alpha_1 \beta_1 y_{t-3}^2 + \dots . \end{aligned}$$

Disto, fica claro que para garantir que cada  $h_t$  seja estritamente positiva é suficiente que  $w > 0$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, q$  e  $\beta_j \geq 0$ ,  $j = 1, \dots, q$ . Para a esperança de  $y_t$  dado  $I_t = \{y_{t-1}, y_{t-2}, \dots\}$ , segue que

$$E[y_t | I_t] = \sqrt{h_t} E[\varepsilon_t | I_t] = \sqrt{h_t} E[\varepsilon_t] = 0 \quad (1.3)$$

$$\text{Var}[y_t | I_t] = E[y_t^2 | I_t] = h_t \text{Var}[\varepsilon_t] = h_t. \quad (1.4)$$

Assim, utilizando as equações (1.3) e (1.4) pode-se calcular a esperança e a variância incondicionais  $y_t$ , ou seja

$$\begin{aligned} E[y_t] &= E[E[y_t | I_t]] = E[0] = 0 \\ \text{Var}[y_t] &= E[y_t^2] = E[E[y_t^2 | I_t]] = E[h_t] \end{aligned}$$

e assumindo que  $y$  seja estacionária segue que

$$E[h_t] = \frac{w}{1 - \left( \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j \right)}. \quad (1.5)$$

Para que (1.5) seja estritamente positiva e finita  $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ . Em muitas aplicações com modelos GARCH,  $p = q = 1$  é suficiente para capturar o comportamento volátil presente em diversas séries, comparados a valores elevados de  $p$  para os modelos ARCH (para mais detalhes veja [Bollerslev, Chou e Kroner \(1992\)](#)).

Da expressão (1.1) pode-se escrever que  $\varepsilon_t = y_t / \sqrt{h_t}$ . Assim, a função densidade de probabilidades (fdp) condicional do retorno  $y_t$  será dada por

$$p(y_t | I_t) = (h_t)^{-1/2} p_\varepsilon(y_t / \sqrt{h_t}). \quad (1.6)$$

Considerando (1.6) pode-se escrever a função de verossimilhança como o produto das distribuições dos retornos, considerando os retornos passados. Logo,

$$\begin{aligned} p(y_1, \dots, y_n) &= p(y_1)p(y_2 | I_2)p(y_3 | I_3)\dots p(y_n | I_n) \\ L(\boldsymbol{\theta}) &= p(y_1) \prod_{t=2}^n (h_t)^{-1/2} p_\varepsilon(y_t / \sqrt{h_t}), \end{aligned}$$

sendo  $\boldsymbol{\theta} = (w, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)'$  os parâmetros associados ao modelo. Observe que como não se conhece a distribuição marginal de  $Y_1$ , a função de verossimilhança não pode ser obtida. Assim, deve-se utilizar a função de verossimilhança condicional dada por

$$p(y_2, \dots, y_n | y_1) = \prod_{t=2}^n (h_t)^{-1/2} p_{\varepsilon}(y_t / \sqrt{h_t}),$$

Uma característica deste modelo é que a distribuição incondicional de  $y_t$  possue caudas mais pesadas do que a Normal. Assim, vários autores tem proposto distribuições com caudas mais pesadas para  $\varepsilon_t$ , como por exemplo, a distribuição *t*-Student padronizada com  $v$  graus de liberdade (BAILLIE; BOLLERSLEV, 1989) e (ARDIA, 2008) e a distribuição exponencial potência (CHUANG; LU; LEE, 2007).

Na literatura surgem evidências empíricas que os retornos financeiros apresentam assimetria, que não são tratadas pelos modelos GARCH ((GOURIEROUX, 1997) e (ALEXANDER, 2001)). Devido a este fato, diversos autores propuseram o uso de distribuições possivelmente assimétricas ((BAUWENS; LUBRANO, 2002), (BAUWENS; LAURENT, 2005), (PIPIÉN, 2006) e (EHLERS, 2012)). Pode-se também evidenciar recentes propostas para modelar outras características como modelos com mudança de regime estocástico ((BAUWENS; PREMINGER; ROMBOUTS, 2010) e (ARDIA, 2009)) e mistura de distribuições para os erros (AUSIN; GALEANO, 2007)).

No mercado financeiro é de grande utilidade analisar várias séries temporais ao mesmo tempo. Desta forma, uma generalização destes modelos é considerar os modelos GARCH de forma multivariada. Por exemplo, pode-se estar interessado em estudar a relação entre volatilidades em diferentes cenários. A seguir defini-se algumas extensões multivariadas dos modelos *GARCH*.

### 1.1.2 Modelos *GARCH* multivariados

Seja  $\mathbf{y}_t = \{y_{t1}, \dots, y_{tm}\}$  um vetor de retornos no tempo  $t$  para  $m$  séries temporais. Logo, os modelos *GARCH* multivariados podem ser definidos por

$$\mathbf{y}_t = \mathbf{H}_t^{1/2} \boldsymbol{\varepsilon}_t \quad (1.7)$$

sendo  $\mathbf{H}_t^{1/2}$  uma matriz  $m \times m$  positiva definida de tal modo que  $\mathbf{H}_t$  é a matriz de variâncias condicionais de  $\mathbf{y}_t$  e  $\mathbf{H}_t^{1/2}$  pode ser obtido pela decomposição de Cholesky da matriz  $\mathbf{H}_t$ . O vetor de erros  $\boldsymbol{\varepsilon}_t$  tem ordem  $m \times 1$  com média e variâncias dadas por

$$\begin{aligned} E(\boldsymbol{\varepsilon}_t) &= \mathbf{0} \\ \text{Var}(\boldsymbol{\varepsilon}_t) &= \mathbf{I}_m \end{aligned}$$

com  $\mathbf{I}_m$  uma matriz identidade de ordem  $m$ . A média e a variância do vetor  $\mathbf{y}_t$  condicionado até o momento  $t$ , ( $I_t = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots\}$ ) podem ser definidas da seguinte forma

$$\begin{aligned} E(\mathbf{y}_t | I_t) &= E(\mathbf{H}_t^{1/2} \boldsymbol{\epsilon}_t | I_t) = \mathbf{H}_t^{1/2} E(\boldsymbol{\epsilon}_t) = 0 \\ \text{Var}(\mathbf{y}_t | I_t) &= \text{Var}(\mathbf{H}_t^{1/2} \boldsymbol{\epsilon}_t | I_t) = \mathbf{H}_t^{1/2} \text{Var}(\boldsymbol{\epsilon}_t) (\mathbf{H}_t^{1/2})' = \mathbf{H}_t. \end{aligned}$$

Os modelos GARCH multivariados mudam de acordo com a especificação da matriz  $\mathbf{H}_t$ . Estas abordagens podem ser relacionadas em três tipos principais:

- *Generalizações diretas dos modelos univariados* - Nesta categoria, tem-se os modelos VEC ([BOLLERSLEV; ENGLE; WOOLDRIDGE, 1988](#)), BEKK ([ENGLE; KRONER, 1995](#)) e de fatores ([\(ENGLE; NG; ROTHSCHILD, 1990\)](#) e ([BOLLERSLEV; ENGLE, 1993](#)));
- *Combinações lineares dos modelos GARCH univariados* - Nesta segunda categoria, tem-se os modelos ortogonais ([\(ALEXANDER, 2001\)](#)) e os modelos de fatores latentes;
- *Combinações não lineares dos modelos GARCH univariados* - Nesta última categoria pode-se citar os modelos de correlação condicionais (constantes, ([BOLLERSLEV, 1990](#)) e dinâmicos, ([ENGLE, 2002](#))), o modelo geral de covariância dinâmica ([KRONER; NG, 1998](#)) e o modelo cópula-GARCH ([\(JONDEAU; ROCKINGER, 2001\)](#)).

[Bauwens, Laurent e Rombouts \(2006\)](#) apresentam uma ótima revisão de todos os modelos citados acima. Neste trabalho o foco principal são os modelos de combinações não lineares do modelos GARCH univariados, mais especificamente os modelos GARCH multivariados de correlação dinâmicos - DCC GARCH.

### CCC e DCC GARCH

O modelo CCC-GARCH (*Constant Conditional Correlation GARCH* ) especifica a matriz  $H_t$  da seguinte forma

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t$$

sendo

$$\mathbf{D}_t = \text{diag}(h_{11,t}^{1/2}, \dots, h_{mm,t}^{1/2}) \quad (1.8)$$

e  $h_{ii,t}$  definido como um modelo GARCH univariado. Se for definido um GARCH(1,1) para cada  $h_{ii,t}$  em  $\mathbf{D}_t$ , segue que  $h_{ii,t} = w_i + \alpha_i y_{i,t-1}^2 + \beta_i h_{ii,t-1}$  com  $w_i > 0, \alpha_i > 0, \beta_i > 0$  e  $\alpha_i + \beta_i < 1$  para  $i = 1, \dots, m$ . A matriz de correlações  $\mathbf{R} = \{\rho_{ij}\}$ ,  $j = 1, \dots, m$  e  $i = 1, \dots, m$  é simétrica e positiva definida com  $\rho_{ii} = 1$ . Como  $\rho_{ij} = \rho_{ji}$ , tem-se que os parâmetros adicionais do modelo serão todos  $\rho_{ij}$  com  $i > j$ , para  $i, j = 1, \dots, m$ . Portanto, esse modelo terá  $m(m+5)/2$  parâmetros. Note que para 3 séries ( $m = 3$ ), tem-se 12 parâmetros no modelo, enquanto que para 4 séries ( $m = 4$ ), 18 parâmetros, um crescimento rápido à medida que aumenta-se o número de séries.

Devido a grande quantidade de parâmetros, surgiram na literatura modelos mais parcimoniosos. É o caso dos modelos DCC-GARCH (*Dynamic Conditional Correlation GARCH*) propostos simultaneamente por [Engle \(2002\)](#) e [Tse e Tsui \(2002\)](#). Segundo [Engle \(2002\)](#),

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \quad (1.9)$$

sendo  $\mathbf{D}_t$  definido em [\(1.8\)](#) e  $\mathbf{R}_t$  como

$$\begin{aligned} \mathbf{R}_t &= \text{diag}(\mathbf{Q}_t)^{-1/2} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2} \\ \mathbf{Q}_t &= (1 - a - b)\mathbf{R} + a\boldsymbol{\mu}_{t-1}\boldsymbol{\mu}'_{t-1} + b\mathbf{Q}_{t-1} \end{aligned}$$

com  $\boldsymbol{\mu}_t = \mathbf{D}_t^{-1} \mathbf{y}_t$  e  $\mathbf{R}$  a matriz de covariâncias incondicionais de  $\boldsymbol{\mu}_t$ . Para  $a$  e  $b$  as restrições de estacionariedade são  $a > 0, b > 0$  e  $a + b < 1$ .

Uma grande vantagem do modelo DCC-GARCH é ser mais parcimonioso do que o modelo CCC-GARCH. Este fato ocorre devido a inclusão de apenas dois parâmetros. Assim, se for definido o modelo GARCH(1,1) para as variâncias condicionais em  $\mathbf{D}_t$ , o número de parâmetros do modelo será  $3m + 2$ . Desta forma, para 3 séries ( $m = 3$ ) o número de parâmetros será 11, para 4 séries ( $m = 4$ ), 15 parâmetros.

Seja  $\boldsymbol{\theta}$  os parâmetros associados ao modelo, pode-se expressar a distribuição conjunta dos retornos por

$$p(\mathbf{y}_1, \dots, \mathbf{y}_T | \boldsymbol{\theta}) = p(\mathbf{y}_1)p(\mathbf{y}_2 | \mathbf{y}_1)p(\mathbf{y}_3 | \mathbf{y}_2, \mathbf{y}_1) \dots p(\mathbf{y}_T | I_T). \quad (1.10)$$

A fdp dos retornos  $\mathbf{y}_t$  em função da fdp dos erros  $\boldsymbol{\epsilon}_t$  será definida por

$$p(\mathbf{y}_t | I_t) = |\mathbf{H}_t|^{-1/2} p_{\boldsymbol{\epsilon}_t}(\mathbf{H}_t^{-1/2} \mathbf{y}_t), \quad t = 1, \dots, T. \quad (1.11)$$

A função de verossimilhança será definida pela substituição [\(1.10\)](#) em [\(1.11\)](#), ou seja

$$L(\boldsymbol{\theta}) = p(\mathbf{y}_1, \dots, \mathbf{y}_T) = \prod_{t=1}^T |\mathbf{H}_t|^{-1/2} |p_{\boldsymbol{\epsilon}_t}(\mathbf{H}_t^{-1/2} \mathbf{y}_t)|.$$

Considerando o modelo DCC-GARCH segue que

$$L(\boldsymbol{\theta}) = \prod_{t=2}^T \left[ \prod_{i=1}^m h_{ii,t}^{-1/2} \right] |\mathbf{R}_t|^{-1/2} p_{\boldsymbol{\epsilon}_t} \left( (\mathbf{D}_t \mathbf{R}_t \mathbf{D}_t)^{-1/2} \mathbf{y}_t \right). \quad (1.12)$$

Para realizar inferência e previsão nos modelos GARCH, existem na literatura duas formas de abordagem: clássica (frequentista) e Bayesiana. Entretanto, neste trabalho a abordagem utilizada é a Bayesiana. A princípio, as restrições de positividade sobre os parâmetros para garantir uma variância positiva, podem gerar alguns problemas em alguns procedimentos de otimização. Essas restrições, do ponto de vista Bayesiano, podem ser incorporados através de distribuições a *priori*.

Outra característica da abordagem Bayesiana é poder realizar a comparação dos modelos através de meios probabilísticos. Neste tipo de abordagem, pode-se construir probabilidades marginais e fatores de Bayes permitindo uma melhor comparação de modelos. A definição das distribuições a priori dos parâmetros do modelo será dada pelas distribuições propostas por [Fioruci, Ehlers e Andrade \(2014\)](#) e assumindo independência entre os parâmetros, logo

$$\begin{aligned} w_i &\sim \mathcal{N}(\mu_{w_i}, \sigma_{w_i}^2) I_{(w_i > 0)} \\ \alpha_i &\sim \mathcal{N}(\mu_{\alpha_i}, \sigma_{\alpha_i}^2) I_{(0 < \alpha_i < 1)} \\ \beta_i &\sim \mathcal{N}(\mu_{\beta_i}, \sigma_{\beta_i}^2) I_{(0 < \beta_i < 1)} \\ a &\sim \mathcal{N}(\mu_a, \sigma_a^2) I_{(0 < a < 1)} \\ b &\sim \mathcal{N}(\mu_b, \sigma_b^2) I_{(0 < b < 1)} \end{aligned}$$

para  $i = 1, \dots, m$ . A notação  $\mathcal{N}(\mu, \sigma^2) I_{(a < x < b)}$ , com  $b > a$ , representa a fdp da distribuição normal truncada dada por

$$\frac{1}{\phi(b) - \phi(a)} \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

em que  $\phi(\cdot)$  denota a fdp da distribuição normal com média  $\mu$  e variância  $\sigma^2$ . Ao utilizar a abordagem Bayesiana, algoritmos de simulação Bayesiana devem ser utilizados e são apresentados no [Capítulo 2](#).





# TÉCNICAS BAYESIANAS DE ESTIMAÇÃO

---



---

Neste capítulo são apresentadas algumas técnicas Bayesianas de estimação presentes na literatura. Além disso, algumas definições sobre as cadeias de Markov também são definidas. Na literatura existem excelentes livros que detalham e exemplificam estes métodos conhecidos como MCMC (Monte Carlo via Cadeias de Markov). Para mais detalhes veja, por exemplo, [Robert e Casella \(2004\)](#) e [Gamerman e Lopes \(2006\)](#).

## 2.1 O princípio MCMC

As técnicas de estimação Bayesiana partem do pressuposto que pode-se incorporar informações a priori e usá-las em conjunto com os dados que serão coletados para fazer inferência estatística. Ou seja, usando a distribuição a posteriori que é uma função dos dados coletados (na forma da função de verossimilhança) e uma distribuição a priori que reflete uma informação anterior segue a [Definição 1](#).

**Definição 1.** Seja  $\mathbf{y} = \{y_1, \dots, y_n\}$  uma amostra coletada de um conjunto de dados de tamanho  $n$  e seja  $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_m\}$  denotando  $m$  parâmetros associados ao modelo assumido para  $\mathbf{y}$ . A distribuição conjunta a posteriori de  $\boldsymbol{\theta}$  dado  $\mathbf{y}$  pode ser dada, usando a regra de Bayes, por

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) &= \frac{p(\boldsymbol{\theta}, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})} \\ &\propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \end{aligned} \quad (2.1)$$

em que  $p(\boldsymbol{\theta}, \mathbf{y})$  denota a distribuição conjunta de  $\boldsymbol{\theta}$  e  $\mathbf{y}$ ,  $p(\mathbf{y}|\boldsymbol{\theta})$  representa a função de verossimilhança,  $p(\boldsymbol{\theta})$  é a distribuição a priori e  $p(\mathbf{y})$  funciona como uma constante normalizadora de  $p(\boldsymbol{\theta}|\mathbf{y})$  para que a posteriori possa integrar em 1 (assumindo que  $\boldsymbol{\theta}$  é contínua) ou somar em 1 (caso  $\boldsymbol{\theta}$  seja discreta).

Note que na equação (2.1), o termo  $p(\mathbf{y})$  foi omitido e a igualdade anterior foi substituída por uma proporcionalidade. Esta é uma forma útil do teorema de Bayes para problemas

envolvendo a estimativa de parâmetros, uma vez que o denominador é apenas uma constante normalizadora.

Um dos maiores interesses na inferência Bayesiana é obter informações acerca da distribuição a posteriori  $p(\boldsymbol{\theta}|\mathbf{y})$  mas nem sempre é fácil obtê-la diretamente. Entretanto, sob certas condições de regularidade, utiliza-se a inferência Bayesiana para a construção de uma cadeia de Markov a qual será utilizada para resumir características da distribuição a posteriori  $p(\boldsymbol{\theta}|\mathbf{y})$ .

## 2.2 Monte Carlo via cadeias de Markov

Segundo [Gamerman e Lopes \(2006\)](#), em termos simples, uma cadeia de Markov é um processo estocástico em que dado o estado presente, o estado passado e o futuro são independentes. Esta propriedade pode ser definida formalmente por

$$P(\boldsymbol{\theta}^{(n+1)} \in A | \boldsymbol{\theta}^{(n)} = x, \boldsymbol{\theta}^{(n-1)} \in A_{n-1}, \dots, \boldsymbol{\theta}^{(0)} \in A_0) = P(\boldsymbol{\theta}^{(n+1)} \in A | \boldsymbol{\theta}^{(n)} = x)$$

para todo conjunto  $A_0, \dots, A_{n-1}, A \subset S$  e  $x \in S$  com  $S$  um espaço de estados. Para o uso dos métodos MCMC, as cadeias de Markov devem satisfazer as seguintes condições:

- homogeniedade, isto é, a probabilidade de transição de um estado para o outro seja invariante;
- irreducibilidade, isto é, cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações;
- aperiodicidade, isto é, não haja estados absorventes.

Para maiores detalhes veja [Robert e Casella \(2004\)](#). Estas condições são necessárias para que  $X_t$  seja amostrado de uma distribuição invariante quando  $t \rightarrow \infty$ .

**Definição 2.** (Distribuição invariante) Uma cadeia de Markov possui uma distribuição invariante  $\pi(x)$  se, para todo conjunto  $A \subset S$  e probabilidade de transição  $T(x, A)$

$$\int_A \pi(x) dx = \int T(x, A) \pi(x) dx.$$

Desta forma suponha que uma distribuição  $\pi(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^d$  seja complexa de tal forma que não seja possível obter uma amostra diretamente. Considerando as realizações  $\{\mathbf{X}^{(t)}, t = 0, 1, \dots\}$  de uma cadeia de Markov com  $\pi$  como distribuição de equilíbrio então temos que

$$\mathbf{X}^{(t)} \xrightarrow{t \rightarrow \infty} \pi(\mathbf{x}) \quad \text{e} \quad \frac{1}{n} \sum_{t=1}^n g(X_i^{(t)}) \xrightarrow{n \rightarrow \infty} \mathbb{E}(g(X_i)) \quad q.c.$$

Na literatura existem diversos métodos computacionais que utilizam princípios markovi-anos para estimativa. A seguir apresenta-se um dos mais conhecidos métodos de computação Bayesiana.

## 2.3 Metropolis-Hastings

O algoritmo Metropolis-Hastings ([HASTINGS, 1970](#)) foi desenvolvido para facilitar a simulação de valores aleatórios de uma distribuição de interesse que possua forma complexa, o que torna esta tarefa muito complicada.

Suponha que o interesse seja em gerar valores de uma distribuição conjunta  $\pi(\boldsymbol{\theta})$ . Considere uma distribuição auxiliar  $q(\boldsymbol{\theta})$  e que um valor  $\boldsymbol{\theta}'$  seja gerado desta distribuição. O algoritmo Metropolis-Hastings pode ser representado pelo [Algoritmo 1](#).

---

### Algoritmo 1 – Algoritmo Metropolis-Hastings

---

- 1: **procedimento** METROPOLIS-HASTINGS  $\triangleright$  Simular valores aleatórios de uma distribuição de interesse
  - 2:     Initialize  $j = 0$  e considere um valor inicial  $\boldsymbol{\theta}^{(0)}$
  - 3:     Gere um valor  $\boldsymbol{\theta}'$  da distribuição  $q(\boldsymbol{\theta})$ ;
  - 4:     Gere  $u \sim U(0, 1)$  e calcule
- $$p(\boldsymbol{\theta}', \boldsymbol{\theta}^{(j)}) = \min \left\{ \frac{\pi(\boldsymbol{\theta}') q(\boldsymbol{\theta}^{(j)} | \boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}^{(j)}) q(\boldsymbol{\theta}' | \boldsymbol{\theta}^{(j)})} \right\}$$
- 5:     **se**  $u < p(\boldsymbol{\theta}', \boldsymbol{\theta}^{(j)})$  **então**
  - 6:          $\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}'$
  - 7:     **senão**
  - 8:          $\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)}$
  - 9:     **fim se**
  - 10:    Faça  $j = j + 1$  e volte ao passo 3
  - 11: **fim procedimento**
- 

Repita 3 – 10 até obter a distribuição estacionária  $\pi(\boldsymbol{\theta})$ .

No segundo passo, deve-se estabelecer os valores iniciais para os parâmetros. Estes valores podem ser obtidos por meio de estimativas de máxima verossimilhança ou através de algum outro método (até mesmo de forma arbitrária). A teoria MCMC diz que a distribuição estacionária do algoritmo é a distribuição à qual a cadeia de Markov produzida converge, ou seja, será a distribuição a posteriori de interesse, independentemente dos valores iniciais escolhidos. No entanto, valores iniciais particularmente pobres podem fazer com que o algoritmo rejeite muitos candidatos e, portanto, não se move rapidamente no suporte da distribuição posteriori, levando a tempos de execução extremamente longos. Para maiores detalhes veja [Lynch \(2007\)](#).

No passo 3, um valor candidato para o parâmetro  $\boldsymbol{\theta}$  é obtido através da simulação de um valor a partir de uma densidade proposta  $q(\boldsymbol{\theta})$ . O valor simulado é considerado um "candidato", porque não é automaticamente aceito para a distribuição de interesse, pois este valor deve ser avaliado para a aceitação através da probabilidade  $p(\boldsymbol{\theta}', \boldsymbol{\theta}^{(j)})$ . A densidade proposta pode assumir uma forma que seja fácil de simular. Dado que a densidade proposta não é a densidade de interesse, tem-se de verificar e determinar se o parâmetro candidato pode ser considerado a

partir da distribuição alvo.

Já no quarto passo, calcula-se a probabilidade de aceitação do novo candidato. Quando a distribuição auxiliar é simétrica, isto é,  $q(\boldsymbol{\theta}^{(j)}|\boldsymbol{\theta}') = q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(j)})$ , a probabilidade de aceitação se reduz a  $\pi(\boldsymbol{\theta}')/\pi(\boldsymbol{\theta}^{(j)})$ . Neste caso, tem-se simplesmente a razão entre a densidade a posteriori avaliada no valor candidato sobre a densidade a posteriori avaliada no valor anterior do parâmetro. Mas se a distribuição auxiliar é assimétrica o termo  $q(\boldsymbol{\theta}^{(j)}|\boldsymbol{\theta}')/q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(j)})$  deve ser levado em consideração. Esta razão significa a probabilidade de que um valor  $\boldsymbol{\theta}^{(j)}$  seja proposto, uma vez que a cadeia está localizada em  $\boldsymbol{\theta}'$ .

Por fim, no passo 5, compara-se o valor gerado de uma distribuição Uniforme com a probabilidade de aceitação  $p(\boldsymbol{\theta}', \boldsymbol{\theta}^{(j)})$ . Se  $p(\boldsymbol{\theta}', \boldsymbol{\theta}^{(j)}) > u$ , aceita-se o valor candidato como um valor da distribuição a posteriori de interesse. Caso contrário, mantem-se o valor anterior do parâmetro.

Uma dificuldade desse algoritmo surge quando a dimensão do espaço de parâmetros torna-se grande e/ou a distribuição a posteriori possue uma forma analítica bastante complexa. Nesses casos, geralmente é necessário um longo tempo de simulação para explorar todas as regiões do espaço paramétrico o que pode levar a valores gerados bastante pobres. Além disso, a eficiência deste algoritmo ficará condicionada a um bom esquema de atualização e uma boa escolha da distribuição auxiliar, o que nem sempre é uma tarefa fácil. Devido à essas restrições, outros algoritmos de estimação Bayesiana vem surgindo na literatura. Um desses algoritmos é o Monte Carlo Hamiltoniano.

## 2.4 Monte Carlo Hamiltoniano

A ideia original do algoritmo Monte Carlo Hamiltoniano (HMC) é baseada na dinâmica Hamiltoniana, uma técnica geral da física teórica em que a energia total de um sistema de partículas é descrito por meio de uma função chamada função Hamiltoniana.

O método HMC foi inicialmente proposto por [Duane et al. \(1987\)](#) e chamado de Monte Carlo Híbrido, sendo utilizado para a simulação dinâmica de sistemas moleculares. Na literatura, o primeiro uso do método HMC foi em aplicações estatísticas de redes neurais ([NEAL, 1995](#)). [Neal \(2011\)](#) e [Gelman et al. \(2014\)](#) apresentam propriedades importantes desse algoritmo.

Para [Calderhead \(2012\)](#) o Monte Carlo Hamiltoniano pode ser entendido a partir de duas perspectivas: do ponto de vista de dinâmicas moleculares o HMC surge como um método que corrige erros introduzidos por uma discretização das equações de Hamilton. Já do ponto de vista estatístico, tem-se uma proposta para o uso da dinâmica Hamiltoniana como um mecanismo eficaz para realização de métodos MCMC, de tal forma que o novo estado proposto estará longe do estado atual da cadeia e será aceito com uma alta probabilidade.

### 2.4.1 Dinâmica Hamiltoniana

Para exemplificar a dinâmica Hamiltoniana considere um disco sem atrito deslizando sobre uma superfície de altura variável. O estado deste sistema consiste no posicionamento do disco, dada por um vector  $\boldsymbol{\theta}$ , e o movimento do disco (sua massa vezes sua velocidade), dada por um vetor  $\mathbf{p}$ . A energia potencial,  $U(\boldsymbol{\theta})$ , do disco é proporcional à altura da superfície na posição atual. Já sua energia cinética,  $K(\mathbf{p})$ , é igual a  $\mathbf{p}'\mathbf{p}/2m$ , em que  $m$  é a massa do disco.

Com o movimento do disco, em uma alguma parte do nível da superfície, o disco move-se a uma velocidade constante. Entretanto, se este encontrar uma inclinação ascendente, o movimento do disco permite que este continue com a sua energia cinética diminuindo e sua energia potencial crescente até que sua energia cinética (e, portanto,  $\mathbf{p}$ ) chegue a zero. Em seguida, o disco iniciará um movimento descendente com energia cinética aumentando e energia potencial diminuindo.

**Definição 3.** (Função Hamiltoniana) A energia total de um sistema é representada pela função Hamiltoniana

$$\begin{aligned} H(\boldsymbol{\theta}, \mathbf{p}) &= U(\boldsymbol{\theta}) + K(\mathbf{p}) \\ &= U(\boldsymbol{\theta}) + \mathbf{p}'M^{-1}\mathbf{p}, \end{aligned} \quad (2.2)$$

em que  $U(\boldsymbol{\theta})$  é a energia potencial e é definida como menos o log da densidade de probabilidades da distribuição de  $\boldsymbol{\theta}$  a qual se quer amostrar e  $K(\mathbf{p})$  é a energia cinética com  $M$  uma matriz de massa simétrica positiva definida, que é tipicamente diagonal, e é muitas vezes um múltiplo escalar da matriz identidade (NEAL, 2011).

Definida a função Hamiltoniana deve-se determinar como  $\boldsymbol{\theta}$  e  $\mathbf{p}$  se movem durante um tempo  $t$ . Disto segue a Definição 4.

**Definição 4.** (Equações de movimento) Para determinar o movimento de  $\boldsymbol{\theta}$  e  $\mathbf{p}$ , considere as seguintes derivadas parciais da função Hamiltoniana

$$\frac{\partial \boldsymbol{\theta}}{\partial t} = \frac{\partial H}{\partial \mathbf{p}} = \nabla_{\mathbf{p}}K(\mathbf{p}) \quad (2.3)$$

$$\frac{\partial \mathbf{p}}{\partial t} = -\frac{\partial H}{\partial \boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}}U(\boldsymbol{\theta}), \quad (2.4)$$

em que o símbolo  $\nabla_x$  indica o vetor gradiente em relação a uma variável  $x$ . As equações (2.3) e (2.4) determinam a posição e a velocidade de uma partícula para qualquer intervalo de tempo de duração  $s$ , ou seja, estas equações definem um mapeamento  $T_s$ , do estado, em qualquer instante  $t$  para o estado no tempo  $t + s$  (NEAL, 2011).

### 2.4.2 Algoritmo HMC

Nesta seção apresenta-se o algoritmo do método HMC. Esta síntese pode ser encontrada com maiores detalhes em Burda e Maheu (2013). Os autores apresentam o método HMC com

uma perspectiva mais probabilística, com foco Bayesiano, diferentemente de outros autores cujo foco, geralmente, se concentra em termos da leis físicas do movimento.

Considere um vetor de parâmetros de interesse  $\boldsymbol{\theta} \in \mathbb{R}^d$  distribuído de acordo com uma densidade a posteriori  $\pi(\boldsymbol{\theta})$ . Seja  $\mathbf{p} \in \mathbb{R}^d$  um vetor auxiliar de parâmetros com  $\mathbf{p} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{M})$ , distribuído como uma distribuição Normal com média  $\mathbf{0}$  e com uma matriz de covariâncias  $\mathbf{M}$ , independente de  $\boldsymbol{\theta}$ . Logo, o negativo do logaritmo da distribuição conjunta de  $(\boldsymbol{\theta}, \mathbf{p})$  é dado pela equação Hamiltoniana

$$H(\boldsymbol{\theta}, \mathbf{p}) = -\ln\pi(\boldsymbol{\theta}) + \frac{1}{2}\ln((2\pi)^d|\mathbf{M}|) + \frac{1}{2}\mathbf{p}'\mathbf{M}^{-1}\mathbf{p}. \quad (2.5)$$

Em aplicações práticas as equações (2.3) e (2.4), em geral, não podem ser resolvidas de forma analítica, sendo necessário a utilização de métodos numéricos. Um dos métodos mais populares na literatura é o método de Störmer-Verlet

$$\mathbf{p}(t + \varepsilon/2) = \mathbf{p}(t) + (\varepsilon/2)\nabla_{\boldsymbol{\theta}}\ln\pi(\boldsymbol{\theta}(t)) \quad (2.6)$$

$$\boldsymbol{\theta}(t + \varepsilon) = \boldsymbol{\theta}(t) + \varepsilon\mathbf{M}^{-1}\mathbf{p}(t + \varepsilon/2) \quad (2.7)$$

$$\mathbf{p}(t + \varepsilon) = \mathbf{p}(t + \varepsilon/2) + (\varepsilon/2)\nabla_{\boldsymbol{\theta}}\ln\pi(\boldsymbol{\theta}(t + \varepsilon)) \quad (2.8)$$

para algum valor de  $\varepsilon > 0$  pequeno. Utilizando as equações (2.6), (2.7) e (2.8) pode-se definir o algoritmo HMC de acordo com o [Algoritmo 2](#).

Ao gerar  $\mathbf{p}^*$  de uma distribuição normal proporciona-se uma inicialização estocástica. Este passo é necessário a fim de tornar a cadeia de Markov resultante irredutível e aperiódica ([BURDA; MAHEU, 2013](#)). Em termos do algoritmo HMC, o valor inicial de  $\mathbf{p}^*$  constitui um passo de Gibbs no espaço paramétrico de  $(\boldsymbol{\theta}, \mathbf{p})$  aceito com probabilidade 1. Esta natureza estocástica impede a cadeia de ficar presa no ponto original ou muito próximo a ele.

Em seguida, gera-se uma sequência de valores  $(\boldsymbol{\theta}^{(k)}, \mathbf{p}^{(k)})$  de acordo com a dinâmica Hamiltoniana a partir do estado atual da cadeia  $(\boldsymbol{\theta}^{(0)}, \mathbf{p}^*)$  e define-se o último membro da sequência como o novo estado proposto pelo HMC ( $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^S$ ). O papel da dinâmica Hamiltoniana é o de assegurar que a probabilidade de aceitação seja mantido próximo a 1.

Já a propriedade de transição do valor atual para o novo valor proposto pode ser conseguida tornando  $\boldsymbol{\theta}$  e  $\mathbf{p}$  funções de um tempo contínuo  $T$  e especificando suas respectivas evoluções usando as equações dinâmicas (2.3) e (2.4). Disto, a variável  $\mathbf{p}$  desempenha um papel de variável auxiliar que parametriza  $\pi(\boldsymbol{\theta}, \cdot)$  fornecendo-lhe um grau adicional de flexibilidade para manter a probabilidade de aceitação perto de um durante todo o processo. Mesmo que  $\ln\pi(\boldsymbol{\theta}^{(k)})$  possa divergir substancialmente do  $\ln\pi(\boldsymbol{\theta}^{(0)})$ , resultando em uma mistura (mixing) favorável para  $\boldsymbol{\theta}$ , os termos adicionais em (2.5) compensam este desvio para manter o nível geral de  $H(\boldsymbol{\theta}^{(k)}, \mathbf{p}^k)$  próximo de uma constante durante o processo, quando utilizado de acordo com (2.6), (2.7) e (2.8). Em contrapartida, sem a parametrização adicional com  $\mathbf{p}$  a probabilidade de aceitação MH, muitas vezes, cai para zero de forma relativamente rápida.

**Algoritmo 2 – Algoritmo Monte Carlo Hamiltoniano - HMC**


---

```

1: procedimento MONTE CARLO HAMILTONIANO      ▷ Simular valores aleatórios de uma
   distribuição de interesse
2:   Forneça uma posição inicial  $\boldsymbol{\theta}^{(0)}$  e o tamanho do passo  $\varepsilon$ ;
3:   Inicialize um contador  $i = 1$ ;
4:   Gere  $\mathbf{p}^* \sim \mathcal{N}_d(\mathbf{0}, \mathbf{M})$  e  $u \sim U(0, 1)$ ;
5:   Faça  $(\boldsymbol{\theta}^I, \mathbf{p}^I) = (\boldsymbol{\theta}^{(i-1)}, \mathbf{p}^*)$ ,  $H_0 = H(\boldsymbol{\theta}^I, \mathbf{p}^I)$ ;
6:   para  $i = 1, \dots, LF$  faz           ▷ LF é o número de vezes para resolver as equações de
   Störmer-Verlet
7:

$$\begin{aligned}\mathbf{p}^* &= \mathbf{p}^* + (\varepsilon/2)\nabla_{\boldsymbol{\theta}}\ln\pi(\boldsymbol{\theta}^{(j-1)}) \\ \boldsymbol{\theta}^{(j-1)} &= \boldsymbol{\theta}^{(j-1)} + \varepsilon\mathbf{M}^{-1}\mathbf{p}^* \\ \mathbf{p}^* &= \mathbf{p}^* + (\varepsilon/2)\nabla_{\boldsymbol{\theta}}\ln\pi(\boldsymbol{\theta}^{j-1})\end{aligned}$$

8:   fim para
9:   Faça  $(\boldsymbol{\theta}^S, \mathbf{p}^S) = (\boldsymbol{\theta}^{(j-1)}, \mathbf{p}^*)$  e  $H_1 = H(\boldsymbol{\theta}^S, \mathbf{p}^S)$ ;
10:  Calcule  $\alpha[(\boldsymbol{\theta}^S, \mathbf{p}^S), (\boldsymbol{\theta}^I, \mathbf{p}^I)] = \min[\exp(H_0 - H_1), 1]$ ;
11:  se  $u < \alpha[(\boldsymbol{\theta}^S, \mathbf{p}^S), (\boldsymbol{\theta}^I, \mathbf{p}^I)]$  então
12:     $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^S$ ;
13:  senão
14:     $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^I$ ;
15:  fim se
16:  Faça  $i = i + 1$  e volte ao passo 4.
17: fim procedimento

```

---

Na expressão  $\alpha[(\boldsymbol{\theta}^S, \mathbf{p}^S), (\boldsymbol{\theta}^I, \mathbf{p}^I)]$  a probabilidade de aceitação do HMC é dada pela diferença das equações Hamiltonianas  $H_0$  e  $H_1$ . Para manter esta probabilidade próxima de um, a diferença entre as equações Hamiltonianas deve tender a zero. Outra característica é que esta probabilidade só será mantida próxima a um se a evolução da trajetória for contínua. No entanto, devido a sua discretização em etapas individuais, a probabilidade de aceitação sempre se desvia de um devido a erros de discretização. Desta forma os parâmetros  $\varepsilon > 0$  e  $LF$  devem ser ajustados para atingir uma taxa de aceitação desejada.

Segundo [Girolami e Calderhead \(2011\)](#), embora as escolhas de  $\varepsilon$  e  $LF$  devam ser ajustadas, como já mencionado anteriormente, não está claro como selecionar os valores da matriz  $\mathbf{M}$  de uma forma automática ou com princípios que não necessitem de algum conhecimento prévio da distribuição alvo. Para melhorar a performance do HMC, poderíamos empregar as transições levando em conta a estrutura da densidade alvo, propondo mover-se para diferentes regiões de probabilidade empregando uma posição específica na covariância da densidade proposta  $\mathcal{N}(\mathbf{p}, \mathbf{M})$ , em vez de uma matriz de covariâncias global.

Conceitos geométricos de distância, curvatura, métrica e geodésica (curvas de comprimento mínimo entre dois pontos) são de interesse natural na metodologia estatística e são explorados no desenvolvimento e aprimoramento do método HMC.

### 2.4.3 Monte Carlo Hamiltoniano em variedade Riemanniana (RMHMC)

Uma característica do RMHMC é utilizar os conceitos do método HMC explorando conceitos geométricos da distribuição a posteriori. Girolami e Calderhead (2011) apresentam importantes propriedades do método RMHMC bem como realizam uma grande quantidade de aplicações mostrando a eficácia do método. Aqui o método será apresentado de uma forma mais simples, para mais detalhes veja também Calderhead (2012). Para maiores detalhes matemáticos sobre variedades diferenciáveis e Riemannianas veja Chavel (2006).

O RMHMC é uma extensão do HMC com o propósito de fornecer uma melhor convergência e mistura da cadeia. Este método explora a geometria riemanniana do espaço paramétrico adaptando a matriz de covariâncias  $\mathbf{M}$  utilizada no HMC, substituindo-a em (2.5) pela matriz de informação de Fisher  $M(\boldsymbol{\theta})$ . Desta forma, tem-se a seguinte equação Hamiltoniana

$$H(\boldsymbol{\theta}, \mathbf{p}) = -\ln\pi(\boldsymbol{\theta}) + \frac{1}{2}\ln|M(\boldsymbol{\theta})| + \frac{1}{2}\mathbf{p}'M(\boldsymbol{\theta})^{-1}\mathbf{p}. \quad (2.9)$$

Portanto, as equações de movimento (2.3) e (2.4) são definidas por

$$\frac{\partial \boldsymbol{\theta}}{\partial t} = \frac{\partial H}{\partial \mathbf{p}} = M(\boldsymbol{\theta})^{-1}\mathbf{p} \quad (2.10)$$

$$\frac{\mathbf{p}}{\partial t} = -\frac{\partial H}{\partial \boldsymbol{\theta}} = \nabla_{\boldsymbol{\theta}}\ln\pi(\boldsymbol{\theta}) - \frac{1}{2}tr\left[M(\boldsymbol{\theta})^{-1}\frac{\partial M(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] + \frac{1}{2}\mathbf{p}'M(\boldsymbol{\theta})^{-1}\frac{\partial M(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}M(\boldsymbol{\theta})^{-1}\mathbf{p}. \quad (2.11)$$

A equação (2.11) requer o cálculo de derivadas de segunda e terceira ordem, aumentando a complexidade computacional do algoritmo. Assim, também será necessário a utilização de métodos de solução numérica. Neste caso, utiliza-se uma generalização do método Störmer-Verlet definindo o algoritmo RMHMC da seguinte maneira,

1 - Forneça uma posição inicial  $\boldsymbol{\theta}^{(0)}$  e o tamanho do passo  $\varepsilon$ ;

2 - Inicialize um contador  $i = 1$ ;

3 - Posteriormente,

- Gere  $\mathbf{p}^* \sim \mathcal{N}_d(\mathbf{0}, M(\boldsymbol{\theta}^{(i-1)}))$  e  $u \sim U(0, 1)$ ;
- Faça  $(\boldsymbol{\theta}^I, \mathbf{p}^I) = (\boldsymbol{\theta}^{(i-1)}, \mathbf{p}^*)$ ,  $H_0 = H(\boldsymbol{\theta}^I, \mathbf{p}^I)$ ;
- Para  $j = 1, \dots, l$  ( $l$  representa o número de vezes do loop)
 
$$\hat{\mathbf{p}}^{(0)} = \mathbf{p}^{(j)}$$

Para  $k = 1, \dots, \text{nfps}$  (número de passos do ponto fixo )

% Atualização do momento com iterações de ponto fixo

$$\cdot \hat{\mathbf{p}}^{(k)} = \mathbf{p}^{(j)} - \frac{\varepsilon}{2}\nabla_{\boldsymbol{\theta}}H(\boldsymbol{\theta}^{(j)}, \hat{\mathbf{p}}^{(j-1)}) \text{ em que}$$

$$\nabla_{\theta_i}H = -\frac{\partial \ln\pi(\boldsymbol{\theta})}{\partial \theta_i} + \frac{1}{2}tr\left[M(\boldsymbol{\theta})^{-1}\frac{\partial M(\boldsymbol{\theta})}{\partial \theta_i}\right] + \frac{1}{2}\mathbf{p}'M(\boldsymbol{\theta})^{-1}\frac{\partial M(\boldsymbol{\theta})}{\partial \theta_i}M(\boldsymbol{\theta})^{-1}\mathbf{p}$$

$$\mathbf{p}^{(j+1/2)} = \hat{\mathbf{p}}^{(k)}$$

$$\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}^{(j)}$$

Para  $k = 1, \dots, \text{nfps}$

% Atualização dos parâmetros com iterações de ponto fixo

$$\begin{aligned} \cdot \quad \hat{\boldsymbol{\theta}}^{(k)} &= \boldsymbol{\theta}^{(j)} + \frac{\varepsilon}{2} \nabla_{\boldsymbol{\theta}} H(\boldsymbol{\theta}^{(j)}, \mathbf{p}^{(j+1/2)}) + \frac{\varepsilon}{2} \nabla_{\boldsymbol{\theta}} H(\hat{\boldsymbol{\theta}}^{(k-1)}, \mathbf{p}^{(j+1/2)}) \text{ em que } \nabla_{\boldsymbol{\theta}} H = \\ &M(\boldsymbol{\theta})^{-1} \mathbf{p} \\ \boldsymbol{\theta}^{(j+1)} &= \hat{\boldsymbol{\theta}}^{(k)} \end{aligned}$$

Atualize o momento

$$\cdot \quad \mathbf{p}^{(j+1)} = \mathbf{p}^{(j+1/2)} - \frac{\varepsilon}{2} \nabla_{\mathbf{p}} H(\boldsymbol{\theta}^{(j+1)}, \mathbf{p}^{(j+1/2)})$$

- Depois, faça  $\boldsymbol{\theta}^S = \boldsymbol{\theta}^{(j+1)}$ ,  $\mathbf{p}^S = \mathbf{p}^{(j+1)}$ ,  $H_1 = H(\boldsymbol{\theta}^S, \mathbf{p}^S)$ ;
- Calcule  $\alpha[(\boldsymbol{\theta}^S, \mathbf{p}^S), (\boldsymbol{\theta}^I, \mathbf{p}^I)] = \min[\exp(H_0 - H_1, 1)]$ ;
- Se  $u < \alpha[(\boldsymbol{\theta}^S, \mathbf{p}^S), (\boldsymbol{\theta}^I, \mathbf{p}^I)]$  então faça  $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^S$ . Caso contrário, faça  $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^I$ .

4 - Faça  $i = i + 1$  e volte ao passo 3.

O método RMHMC consiste em obter valores do momento a partir de uma distribuição Normal, cuja covariância é o tensor métrico dados os valores atuais dos parâmetros, determinando em um certo número de passos movimentos propostos para  $\boldsymbol{\theta}$  e  $\mathbf{p}^*$  que, em seguida, serão aceitos com uma probabilidade  $\alpha[(\boldsymbol{\theta}^S, \mathbf{p}^S), (\boldsymbol{\theta}^I, \mathbf{p}^I)] = \min[\exp(H_0 - H_1, 1)]$ . Desta forma, a cada iteração propõe-se um valor em que sua estrutura está ajustada a geometria local da distribuição alvo dado o atual estado da cadeia.

Assim como no HMC padrão, este esquema de amostragem produz uma cadeia de Markov reversível, ergódica que satisfaz a equação de balanço ([CALDERHEAD, 2012](#)). Neste caso, no entanto, não há nenhuma necessidade de selecionar e ajustar manualmente a matriz  $\mathbf{M}$ , uma vez que ela é definida em cada passo pela geometria riemanniana.

Devido a dificuldades na implementação da matriz  $\mathbf{M}$ , uma alternativa seria considerar a matriz  $\mathbf{M}$  como uma métrica fixa, ou seja, avaliando  $\mathbf{M}$  em algum ponto do espaço paramétrico. Assim, considerando  $\hat{\boldsymbol{\theta}}$  igual a moda a posteriori e calculamos  $\mathbf{M} = \mathbf{M}(\hat{\boldsymbol{\theta}})$  podemos obter uma ótima aproximação fornecendo uma boa performance ao algoritmo RMHMC.

A maior dificuldade na utilização do HMC é encontrar o valor de  $LF$  e  $\varepsilon$  que forneçam uma taxa de aceitação satisfatória. Essa escolha, muitas vezes, é realizada por tentativas de valores de  $LF$  e  $\varepsilon$  o que torna o método não muito atrativo. Entretanto, com o avanço tecnológico vem surgindo alternativas para a seleção desses valores como é o caso do STAN ([Stan Development Team, 2018b](#)).

#### 2.4.4 STAN

O STAN é uma plataforma para modelagem e computação estatística de alto desempenho. É um software muito utilizado para a construção de amostras MCMC, uma vez que possue em

sua implementação o HMC e NUTS, por exemplo.

O NUTS (*No-U-Turn-Sampler*) proposto por [Homan e Gelman \(2014\)](#) é uma variação do HMC que adapta automaticamente o número de passos  $LF$  a cada iteração. Essa adaptação torna o algoritmo mais atraente pois elimina a dificuldade de escolha do  $LF$  e  $\varepsilon$  por tentativas do usuário. Outra característica atraente do STAN é sua iteração com o software **R** através do pacote *rstan* ([Stan Development Team, 2018a](#)). O *rstan* permite uma ótima interligação entre a linguagem do STAN com o R.

Neste trabalho o *rstan* é utilizado como forma de comparação com o HMC proposto aplicado a modelos GARCH univariados. No caso do *rstan* as derivadas da função alvo são encontradas numericamente. Para os modelos propostos nesse trabalho, as derivadas são encontradas analiticamente. Vale ressaltar que a escolha do  $LF$  e  $\varepsilon$  para as amostras obtidas utilizando o HMC são encontradas depois de diversas tentativas. Como o *rstan* possui um método automático estes valores podem não coincidir levando a diferentes comportamentos nas Cadeias de Markov.



# MONTE CARLO HAMILTONIANO EM MODELOS GARCH UNIVARIADOS

---



---

## 3.1 Introdução

Dois dos principais problemas que surgem na inferência estatística envolvem problemas de integração e de otimização. Por exemplo, nem sempre é possível calcular analiticamente os estimadores associados aos parâmetros ou avaliar as suas propriedades distribucionais como a consistência e eficiência que envolvem cálculos de integrais e, assim, muitas vezes leva-se a considerar soluções numéricas. Uma abordagem geral é o uso de simulação para calcular estas quantidades de interesse. Através destas simulações pode-se utilizar resultados assintóticos e probabilísticos, a fim de avaliar, por exemplo, a convergência dos estimadores.

O problema genérico e abordado no método Monte Carlo é como avaliar a esperança matemática (integral)

$$E_f[h(X)] = \int_{\Delta} h(x)f(x)dx, \quad (3.1)$$

em que  $h$  é uma função mensurável e  $\Delta$  determina a região em que a variável aleatória  $X$  assume seus valores o que geralmente coincide com o suporte da densidade  $f$  (ROBERT; CASELLA, 2010). O princípio de Monte Carlo para a aproximação de (3.1) consiste em gerar uma amostra  $(X_1, \dots, X_n)$  da densidade  $f$  e propor como aproximação da integral à média empírica

$$\bar{h}_n = \frac{1}{n} \sum_{j=1}^n h(x_j),$$

em que  $x_j$  é o valor observado de  $X_j$  e desde que  $\bar{h}_n$  converja quase certamente (ou seja, para quase todas as sequências geradas) para  $E_f[h(X)]$ .

No Capítulo 2 define-se que

$$\mathbf{X}^{(t)} \xrightarrow{t \rightarrow \infty} \pi(\mathbf{x}) \quad \text{e} \quad \frac{1}{n} \sum_{t=1}^n g(X_i^{(t)}) \xrightarrow{n \rightarrow \infty} \mathbb{E}(g(X_i)) \quad q.c.$$

Como  $\theta \approx \hat{\theta}$ , pode-se calcular o viés do estimador de  $\theta$  (erro de aproximação) por

$$\text{viés}(\hat{\theta}) \approx \widehat{\text{viés}}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^N (\hat{\theta}^{(i)} - \theta)$$

em que  $\hat{\theta}^{(i)}$  é a estimativa do parâmetro  $\theta$  na  $i$ -ésima iteração com  $i = 1, \dots, m$ . Uma medida da qualidade de  $\hat{\theta}$  é o Erro Quadrático Médio (EQM), o qual é definido por

$$\begin{aligned} \text{EQM}(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2] \\ &= E\{[\hat{\theta} - E(\hat{\theta})]^2\} + 2E\{[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]\} + E\{[E(\hat{\theta}) - \theta]^2\} \\ &= E\{[\hat{\theta} - E(\hat{\theta})]^2\} + E\{[E(\hat{\theta}) - \theta]^2\} = \text{Var}(\hat{\theta}) + [\text{viés}(\hat{\theta})]^2, \end{aligned}$$

uma vez que  $E(\hat{\theta}) - \theta$  é uma constante e  $E[\hat{\theta} - E(\hat{\theta})] = 0$ . Na estimação do EQM pelo método de Monte Carlo segue que

$$\text{EQM}(\hat{\theta}) \approx \widehat{\text{EQM}}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^N (\hat{\theta}^{(i)} - \theta)^2.$$

Tomando a raiz quadrada de EQM encontra-se o Desvio Quadrático Médio que é uma medida de precisão também conhecida como Média Quadrática. Suponha que  $\hat{\theta}_1$  e  $\hat{\theta}_2$  são dois estimadores de  $\theta$ , então  $\hat{\theta}_1$  é melhor que  $\hat{\theta}_2$  se

$$\text{EQM}(\hat{\theta}_1) \leq \text{EQM}(\hat{\theta}_2).$$

No Erro Quadrático Médio avalia-se a diferença entre um estimador e o verdadeiro valor da quantidade estimada. No caso de  $\hat{\theta}_1$  e  $\hat{\theta}_2$  serem estimadores não viesados de  $\theta$ , isto é,  $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$ , então  $\hat{\theta}_1$  é melhor que  $\hat{\theta}_2$  se

$$\text{Var}(\hat{\theta}_1) \leq \text{Var}(\hat{\theta}_2).$$

Utilizando o viés e erro quadrático médio compara-se as estimativas de  $\omega$ ,  $\alpha$  e  $\beta$  em modelos GARCH(1,1) univariados em diferentes cenários, observando seus comportamentos através dos traços das Cadeias de Markov e funções de autocorrelação. Nesta comparação utiliza-se os métodos de estimação Monte Carlo Hamiltoniano (HMC) e Metropolis-Hastings presente no pacote *BayesDccGarch* ([FIORUCI; EHLERS; ANDRADE, 2014](#)).

Na Seção 3.2 apresenta-se a construção do método HMC considerando o modelo GARCH(1,1) em que os erros seguem distribuição Normal. Na [Subseção 3.2.1](#) e [Subseção 3.2.2](#) é realizado estudos de simulação para estudar o comportamento das estimativas dos parâmetros do modelo em diversos cenários. Para a [Subseção 3.2.7](#) e [Subseção 3.2.8](#) é realizado um estudo utilizando o HMC proposto neste trabalho e o pacote *rstan*. Em todos os casos foram utilizados

o mesmo computador com sistema operacional Ubuntu 16.04.1 LTS (Xenial Xerus), 8gb de memória ram e processador Intel® Core™ i5, com ambos os métodos escritos em linguagem C e utilizados no software R através do pacote *Rcpp* para uma comparação mais precisa das estimativas e tempos de simulação.

## 3.2 GARCH(1,1) e o HMC

Seja o modelo  $GARCH(p, q)$  definido como na [Subseção 1.1.1](#). Para exemplificar a comparação entre os métodos de estimação HMC (Monte Carlo Hamiltoniano) e MH (Metropolis Hastings) considere o modelo GARCH(1,1)

$$y_t = \varepsilon_t \sqrt{h_t}, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \quad (3.2)$$

$$h_t = w + \alpha y_{t-1}^2 + \beta h_{t-j} \quad (3.3)$$

em que as restrições de positividade e estacionariedade de  $h_t$  são  $w > 0$ ,  $\alpha \geq 0$ ,  $\beta \geq 0$ ,  $\alpha + \beta < 1$  e  $\mathcal{N}(0, 1)$  representa uma distribuição normal com média zero e variância um. As prioris dos parâmetros são dadas por  $w \sim \mathcal{N}(0, 10^2)I_{(w>0)}$ ,  $\alpha \sim \mathcal{N}(0, 10^2)I_{(0<\alpha<1)}$  e  $\beta \sim \mathcal{N}(0, 10^2)I_{(0<\beta<1)}$  ocasionando em pouca informação a priori sobre os parâmetros.

Uma característica do método HMC é que o espaço paramétrico para cada parâmetro deve pertencer aos reais. Desta forma, uma transformação é necessária para que o método possa ser utilizado. Disto, considere as seguintes transformações para  $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)' = (w, \alpha, \beta)'$

$$\phi_1 = \log(w), \phi_2 = \log\left(\frac{\alpha}{1-\alpha}\right) \text{ e } \phi_3 = \log\left(\frac{\beta}{1-\beta}\right).$$

Os valores na escala original são obtidos através da transformação inversa.

Considerando a distribuição normal como a distribuição associada ao erro do modelo (veja [Equação 3.2](#)) define-se a distribuição à posteriori de  $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)'$  por

$$\pi(\boldsymbol{\phi} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) |J(\boldsymbol{\phi})| \quad (3.4)$$

em que  $p(\mathbf{y} | \boldsymbol{\theta}) = \prod_{t=2}^n (\omega + \alpha y_{t-1}^2 + \beta h_{t-1})^{-1/2} p_\varepsilon(y_t / \sqrt{h_t})$ ,  $p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \prod_{i=1}^3 p_{\theta_i}(\theta_i)$  representa as distribuições a priori dos parâmetros do modelo como normais truncadas com  $\mu_w = \mu_\alpha = \mu_\beta = 0$  e  $\sigma_w^2 = \sigma_\alpha^2 = \sigma_\beta^2 = 100$ , ou seja,

$$\begin{aligned} p(\omega) &= \frac{1}{1 - \Phi(0)} \left( \frac{1}{\sqrt{2\pi(100)}} \exp\left\{-\frac{\omega^2}{2(100)}\right\} \right), \\ p(\alpha) &= \frac{1}{\Phi(1) - \Phi(0)} \left( \frac{1}{\sqrt{2\pi(100)}} \exp\left\{-\frac{\alpha^2}{2(100)}\right\} \right), \\ p(\beta) &= \frac{1}{\Phi(1) - \Phi(0)} \left( \frac{1}{\sqrt{2\pi(100)}} \exp\left\{-\frac{\beta^2}{2(100)}\right\} \right), \end{aligned}$$

e

$$|J(\boldsymbol{\phi})| = \begin{vmatrix} e^{\phi_1} & 0 & 0 \\ 0 & \frac{e^{\phi_2}}{(1+e^{\phi_2})^2} & 0 \\ 0 & 0 & \frac{e^{\phi_3}}{(1+e^{\phi_3})^2} \end{vmatrix} = \frac{e^{\phi_1+\phi_2+\phi_3}}{(1+e^{\phi_2})^2(1+e^{\phi_3})^2}.$$

Além disso, é necessário encontrar o logaritmo da distribuição a posteriori. Dada a equação (3.4) segue que

$$\log(\pi(\boldsymbol{\phi}|\mathbf{y})) = \log(p(\mathbf{y}|\boldsymbol{\theta})) + \log(p(\boldsymbol{\theta})) + \log(|J(\boldsymbol{\phi})|) + c \quad (3.5)$$

com  $c$  uma constante que não depende dos parâmetros e

$$\begin{aligned} \log(p(\mathbf{y}|\boldsymbol{\theta})) &= -\frac{1}{2} \sum_{t=2}^n \log(\omega + \alpha y_{t-1}^2 + \beta h_{t-1}) + \frac{y_t^2}{h_t}, \\ \log(p(\boldsymbol{\theta})) &= \sum_{i=1}^3 \log(p_{\theta_i}(\theta_i)), \\ \log(J|\boldsymbol{\phi}|) &= \phi_1 + \phi_2 + \phi_3 - 2\log(1+e^{\phi_2}) - 2\log(1+e^{\phi_3}). \end{aligned}$$

Além do logaritmo da distribuição à posteriori (Equação 3.5) deve-se encontrar as derivadas dessa função de acordo com cada parâmetro associado ao modelo. Logo,

$$\frac{\partial \log(\pi(\boldsymbol{\phi}|\mathbf{y}))}{\partial \phi_i} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} + \frac{\partial \log(p(\boldsymbol{\theta}))}{\partial \theta_i} + \frac{\partial \log|J(\boldsymbol{\phi})|}{\partial \phi_i} \quad (3.6)$$

com

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} &= \sum_{t=2}^n \frac{1}{2h_t} \frac{\partial h_t}{\partial \theta_i} \left( \frac{y_t^2}{h_t} - 1 \right) \\ \frac{\partial h_t}{\partial \theta_i} &= z_t + \beta \frac{\partial h_{t-1}}{\partial \theta_i}, \quad z_t = (1, y_{t-1}^2, h_{t-1})'. \end{aligned}$$

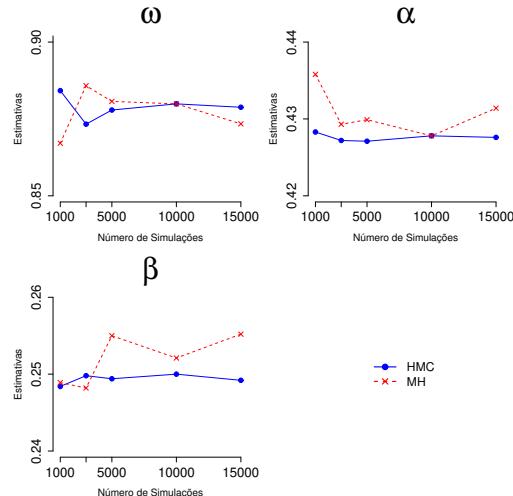
As equações (3.5) e (3.6) são utilizadas durante o processo de estimação utilizando o método HMC. Na Subseção 3.2.1 e Subseção 3.2.2 é apresentado estudos de simulação para analisar o comportamento dos estimadores usando os métodos de simulação HMC e MH.

### 3.2.1 Caso com uma série temporal de tamanho 1000

Considere um modelo GARCH(1,1) com  $\omega = 0.9$ ,  $\alpha = 0.5$ ,  $\beta = 0.2$  e  $\varepsilon_t \sim \mathcal{N}(0, 1)$ . Simulado uma série artificial de tamanho 1000, foi realizado um estudo de simulação comparando as estimativas dos algoritmos Monte Carlo Hamiltoniano (HMC) e Metropolis-Hastings (MH). Para o número de simulações considere os valores 1000, 3000, 5000, 10000 e 15000. Em cada um dos casos, foi decartado metade do número de simulações como amostra de aquecimento (*burn-in*).

Na [Figura 1](#) observa-se as estimativas dos parâmetros considerando os algoritmos HMC (linha azul) e MH (linha vermelha). Na maioria dos casos tem-se que as estimativas do algoritmo HMC estão mais próximas do valor real dos parâmetros quando comparados com o algoritmo MH.

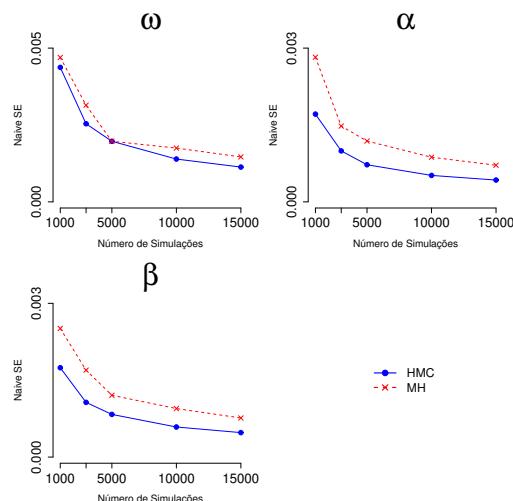
Figura 1 – Gráfico comparando as estimativas para os algoritmos HMC e MH em modelo GARCH(1,1).



Fonte: Elaborada pelo autor.

Já na [Figura 2](#) observa-se os valores dos erros naives para cada tamanho de simulação e considerando cada um dos algoritmos de estimação. Ambos apresentam um decaimento dos valores do erros naives a medida que aumentamos o número de simulações. Entretanto, as estimativas utilizando o algoritmo HMC apresentam, na maioria dos casos, valores menores quando comparado com as estimativas utilizando o MH.

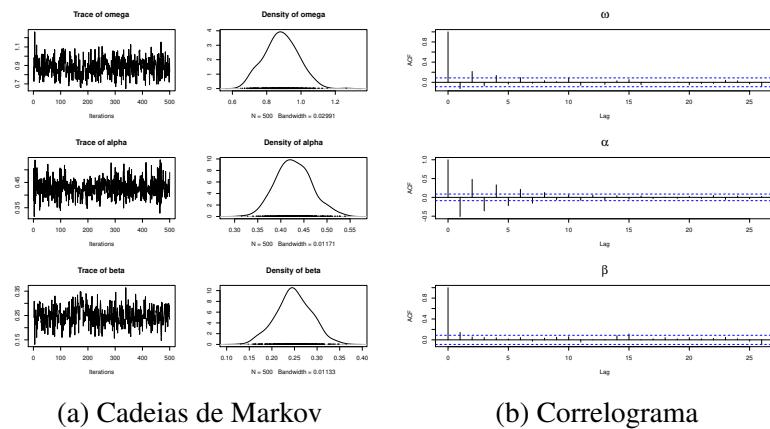
Figura 2 – Gráfico comparando os valores do erro naive para os algoritmos HMC e MH em modelo GARCH(1,1)



Fonte: Elaborada pelo autor.

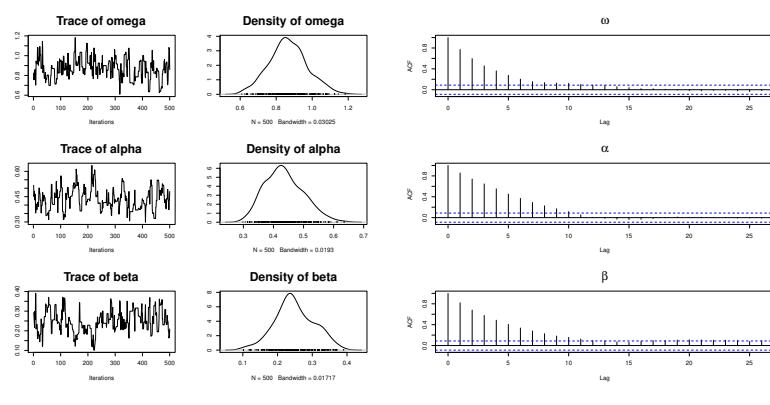
Além de estudar os valores das estimativas dos parâmetros e seus erros naives para ambos os algoritmos também realizou-se estudos através dos gráficos do comportamento das Cadeias de Markov e funções de autocorrelação para todos os parâmetros do modelo GARCH(1,1). As Figura 3 e 4 contem o comportamento das Cadeias de Markov e funções de autocorrelação para  $\omega$ ,  $\alpha$  e  $\beta$  considerando 1000 simulações e os algoritmos HMC e MH. Desta forma, as Cadeias de Markov considerando o algoritmo HMC apresentam um comportamento mais persistente se comparado as estimativas utilizando o algoritmo MH. Ademais, as estimativas com o algoritmo HMC apresentam uma menor autocorrelação serial, principalmente para  $\omega$  e  $\beta$ , enquanto que para o algortimo MH essa autocorrelação se mostra mais presente nos primeiros lags para todos os parâmetros indicando uma dependência entre os valores simulados.

Figura 3 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 1000 simulações.



Fonte: Elaborada pelo autor.

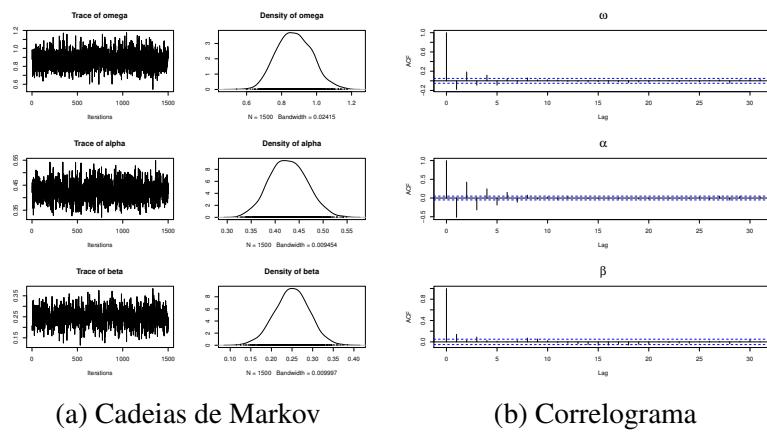
Figura 4 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 1000 simulações.



Fonte: Elaborada pelo autor.

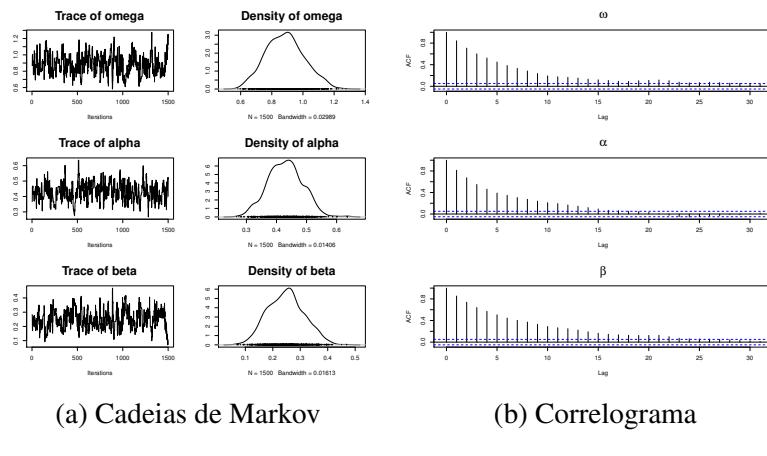
As Figuras 5 e 6 apresentam os comportamentos para as Cadeias de Markov considerando 3000 simulações. Neste caso, as Cadeias de Markov no algoritmo HMC apresentam um comportamento mais persistente se comparado ao MH e ao caso com 1000 simulações. As funções de autocorrelação continuam apresentando fraca autocorrelação serial se comparado com o MH utilizando 3000 simulações.

Figura 5 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 3000 simulações.



Fonte: Elaborada pelo autor.

Figura 6 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 3000 simulações.

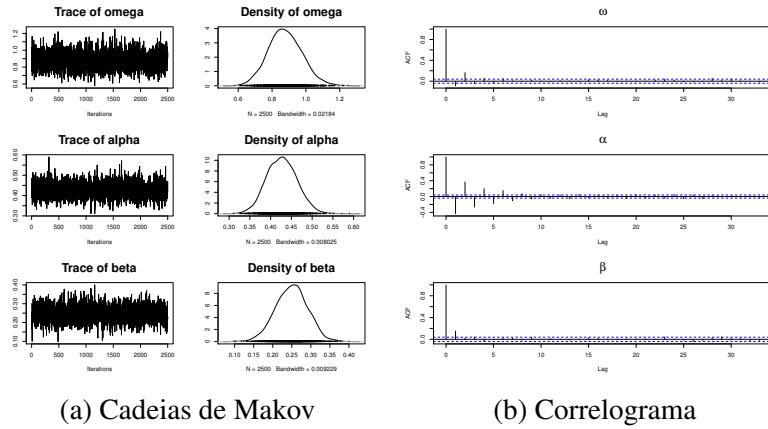


Fonte: Elaborada pelo autor.

As Figuras 7 e 8 mostram o comportamento das Cadeias de Markov e função de autocorrelação para os algoritmos HMC e MH com 5000 simulações. Desta forma, observa-se que o comportamento das estimativas dos parâmetros apresentam o mesmo desempenho dos casos com 1000 e 3000 simulações. Ou seja, as Cadeias de Markov para o algoritmo HMC apresentam

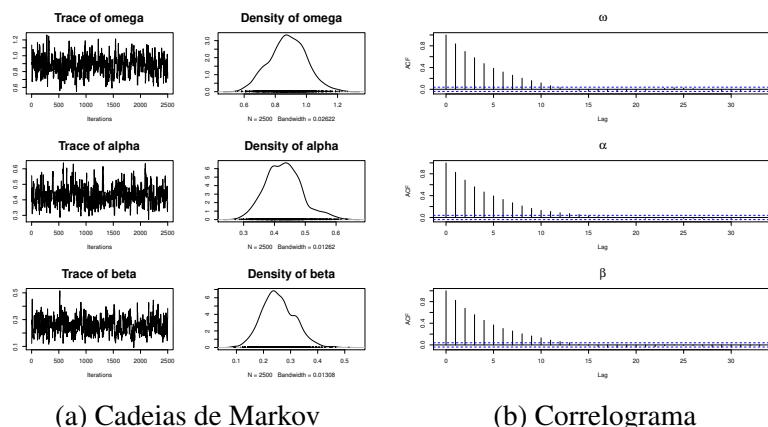
um comportamento mais persistente e com menos autocorrelação serial para as estimativas dos parâmetros quando comparado com o MH.

Figura 7 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 5000 simulações.



Fonte: Elaborada pelo autor.

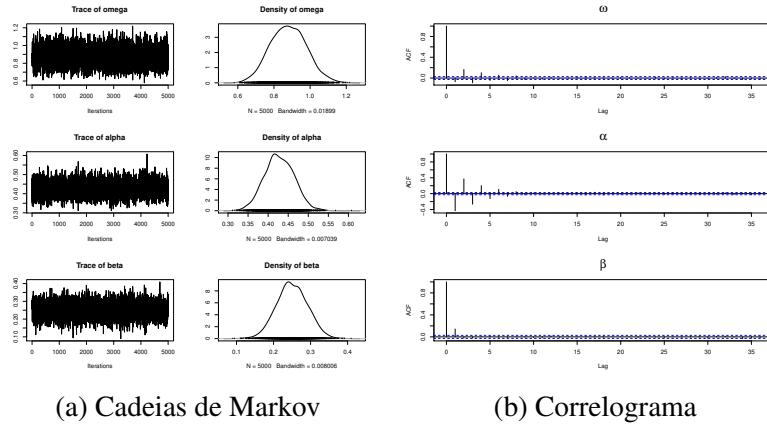
Figura 8 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 5000 simulações.



Fonte: Elaborada pelo autor.

Considerando 10000 simulações observa-se nas Figuras 9 e 10 que o comportamento das Cadeias de Markov de ambos os algoritmos apresentam um comportamento persistente mas, considerando as funções de autocorrelação, o algoritmo MH continua à apresentar uma forte autocorrelação serial nos primeiros lags, comportamento este que permanece em todos os parâmetros do modelo GARCH(1,1).

Figura 9 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 10000 simulações.

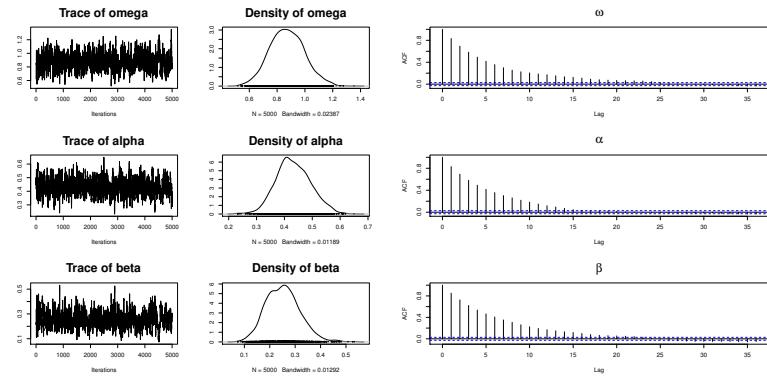


(a) Cadeias de Markov

(b) Correlograma

Fonte: Elaborada pelo autor.

Figura 10 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 10000 simulações.



(a) Cadeias de Markov

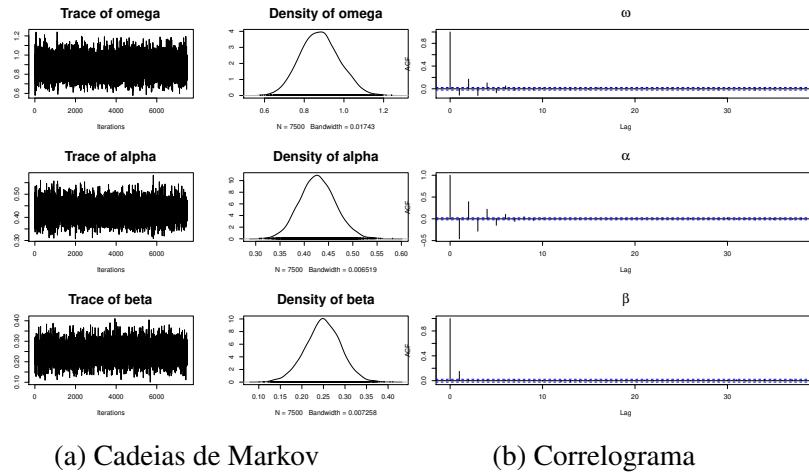
(b) Correlograma

Fonte: Elaborada pelo autor.

Finalmente considerando 15000 simulações observa-se nas Figuras 11 e 12 que ambos os algoritmos apresentam Cadeias de Markov com comportamento persistente mas com o algoritmo MH permanecendo com forte autocorrelação serial nos primeiros lags em todos os parâmetros.

Neste estudo foi utilizado uma série temporal de tamanho 1000 e diversos números de simulação, mostrando que o algoritmo HMC apresenta Cadeias de Markov com comportamentos mais persistentes e com pouca autocorrelação no primeiros lags da função de autocorrelação. Entretanto, uma melhor análise deste comportamento pode ser feita considerando mais séries temporais. Na Subseção 3.2.2 um estudo em que é simulado 500 séries temporais é realizado, construindo valores de viés e EQM para comparar as estimativas dos parâmetros utilizando os dois métodos de estimação.

Figura 11 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 15000 simulações.

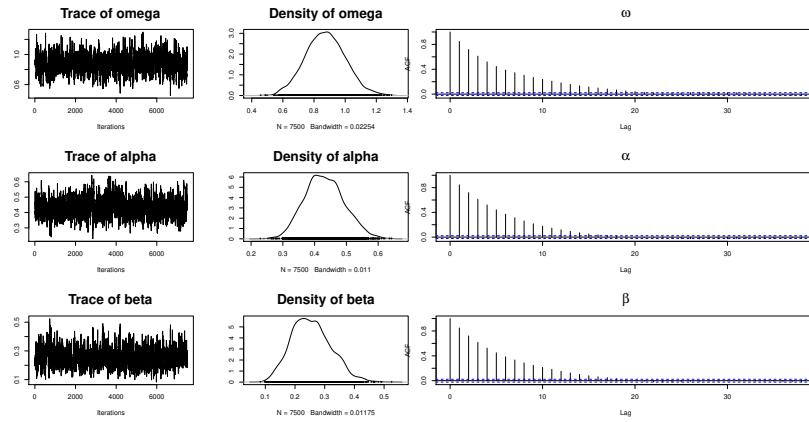


(a) Cadeias de Markov

(b) Correlograma

Fonte: Elaborada pelo autor.

Figura 12 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo MH com 15000 simulações.



(a) Cadeias de Markov

(b) Correlograma

Fonte: Elaborada pelo autor.

### 3.2.2 Caso utilizando 500 séries temporais

Neste estudo de simulação considere o modelo GARCH(1,1) com  $\omega$ ,  $\alpha$ ,  $\beta$  e  $\varepsilon_t \sim \mathcal{N}(0, 1)$ . Nas Seções 3.2.3, 3.2.4 e 3.2.5 foram simuladas 500 séries temporais de tamanho 1000 com  $\omega = 0,9$  e três diferentes configurações para  $\alpha$  e  $\beta$ . No primeiro caso considere  $\alpha + \beta = 0,7$  com  $\alpha = 0,5$  e  $\beta = 0,2$ . Para o segundo caso,  $\alpha + \beta = 0,9$  com  $\alpha = 0,5$  e  $\beta = 0,4$ . Já no terceiro caso, admita  $\alpha + \beta = 0,99$  com  $\alpha = 0,5$  e  $\beta = 0,49$ . Nos três casos, os valores de  $\alpha$  e  $\beta$  foram escolhidos aleatoriamente desde que a soma deles fosse igual aos valores previamente selecionados. Na seção 3.2.6 foi realizado um estudo com 500 séries temporais de tamanho

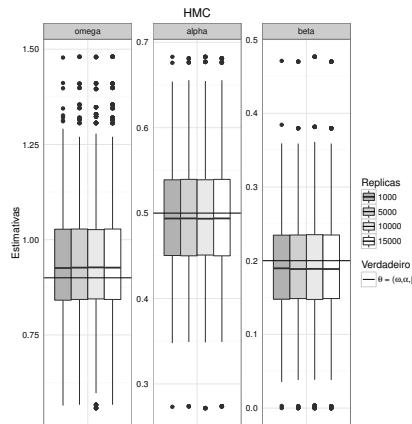
1500 com  $\omega = 0,1$  e  $\alpha + \beta = 0,99$  com  $\alpha = 0,09$  e  $\beta = 0,9$ , fato este que ocorre com mais frequência na prática.

Em todos os casos as simulações foram realizadas com tamanhos 1000, 5000, 10000 e 15000 considerando metade destes valores como amostras de aquecimento (*burn-in*). Para o HMC foi considerado LF=12 e valores de  $\epsilon$  que proporcione uma taxa de aceitação entre 85% e 95%. Em seguida, calculou-se a média das estimativas bem como o viés e o EQM para os dois algoritmos de simulação. Além disso, gráficos boxplot foram construídos para estudar o comportamento distribucional dessas estimativas. Considere também o tempo total de simulação em minutos.

### 3.2.3 Primeiro caso : $\alpha + \beta = 0,7$

Neste estudo de simulação considere  $\omega = 0,9$  e  $\alpha + \beta = 0,7$  com  $\alpha = 0,5$  e  $\beta = 0,2$ . Nas Figuras 13 e 14 apresenta-se os boxplot das estimativas considerando o algoritmo HMC e o MH. Na Figura 13 observa-se que o comportamento das estimativas dos parâmetros considerando os diferentes tamanhos de simulação são bastante parecidos. Quando compara-se com o MH (Figura 14) este comportamento também é bastante similar. Entretanto, em termos de mediana o MH tende a diminuir e aumentar o seu valor a medida que o tamanho de amostra aumenta.

Figura 13 – Boxplot utilizando o algoritmo HMC para  $\omega = 0,9$ ,  $\alpha = 0,5$  e  $\beta = 0,2$  em modelos GARCH(1,1) univariados.



Fonte: Elaborada pelo autor.

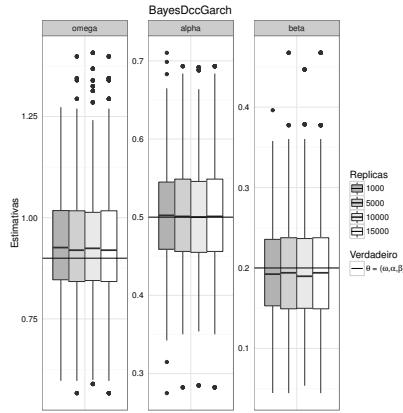
Na Tabela 1 apresenta-se os valores da média das estimativas para os dois algoritmos de estimação bem como o viés e o EQM. Disto, ambos apresentam valores bem próximos em termos das médias das estimativas. Entretanto o método MH apresenta menor valor de viés e EQM quando comparado com HMC. Vale ressaltar que o tempo de simulação total considerando o HMC é bem menor para todos os tamanhos de simulação.

Tabela 1 – Valores médios das estimativas, viés e EQM para  $\omega = 0,9$ ,  $\alpha = 0,5$  e  $\beta = 0,2$  considerando os algoritmos HMC e MH.

Estimativas	HMC			BayesDccGarch				
	Número de iterações MCMC 1000	Número de iterações MCMC 5000	Número de iterações MCMC 10000	Número de iterações MCMC 15000	Número de iterações MCMC 1000	Número de iterações MCMC 5000	Número de iterações MCMC 10000	Número de iterações MCMC 15000
$\hat{\omega}$	0,936768	0,936967	0,936965	0,936977	0,932567	0,930240	0,932397	0,932177
$\hat{\alpha}$	0,497241	0,497174	0,497183	0,497218	0,504541	0,502506	0,502876	0,502656
$\hat{\beta}$	0,189997	0,189866	0,189861	0,189848	0,192964	0,194096	0,193065	0,193287
Viés( $\hat{\omega}$ )	0,036768	0,036967	0,036965	0,036977	0,032567	0,030240	0,032397	0,032177
Viés( $\hat{\alpha}$ )	-0,002759	-0,002826	-0,002817	-0,002782	0,004541	0,002506	0,002876	0,002656
Viés( $\hat{\beta}$ )	-0,010003	-0,010134	-0,010139	-0,010152	-0,007036	-0,005904	-0,006935	-0,006713
EQM( $\hat{\omega}$ )	0,021221	0,021164	0,021188	0,021162	0,016910	0,017680	0,017872	0,017865
EQM( $\hat{\alpha}$ )	0,004083	0,004086	0,004090	0,004081	0,004036	0,004160	0,004188	0,004107
EQM( $\hat{\beta}$ )	0,004093	0,004067	0,004086	0,004072	0,003413	0,003786	0,003712	0,003696
Tempo (min)	6,973933	34,8982	94,61813	122,0494	124,3073	137,1924	191,8649	208,0577

Fonte: Elaborada pelo autor.

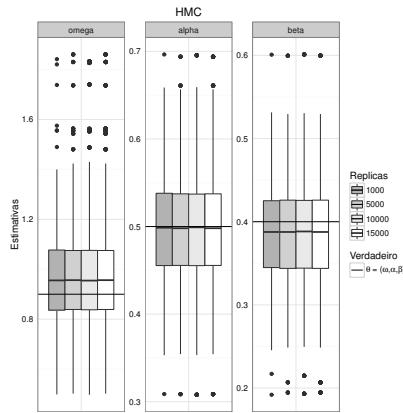
Figura 14 – Boxplot utilizando o pacote BayesDccGarch para  $\omega = 0,9$ ,  $\alpha = 0,5$  e  $\beta = 0,2$  em modelos GARCH(1,1) univariados.



### 3.2.4 Segundo caso : $\alpha + \beta = 0,9$

Para este caso, considere  $\omega = 0,9$  e  $\alpha + \beta = 0,9$  com  $\alpha = 0,5$  e  $\beta = 0,4$ . A Figura 15 apresenta o comportamento da estimativas através do gráfico boxplot. Disto, pode-se observar que para  $\alpha$  os valores da estimativas são muito próximos ao valor real em termos das medianas, com um comportamento aparentemente mais simétrico. Já para  $\omega$  e  $\beta$  os valores medianos das estimativas se mostram um pouco distantes do valor real, com comportamentos distribucionais assimétricos.

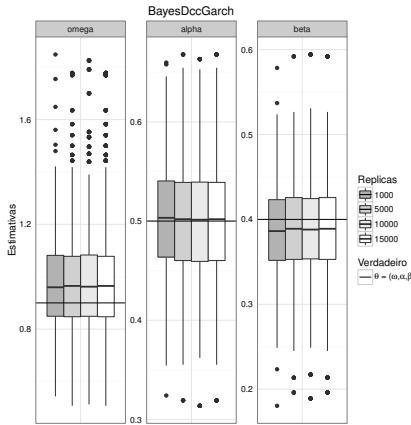
Figura 15 – Boxplot utilizando o algoritmo HMC para  $\omega = 0,9$ ,  $\alpha = 0,5$  e  $\beta = 0,4$  em modelos GARCH(1,1) univariados.



Fonte: Elaborada pelo autor.

A Figura 16 apresenta o gráfico boxplot considerando o algoritmo MH. Observando os valores para o parâmetro  $\beta$  observa-se um comportamento parecido com o HMC, entretanto em termos de mediana estes valores são mais distantes do valor real do parâmetro. Analisando o comportamento para  $\omega$  e  $\alpha$  observa-se, também, comportamentos similares aos que encontram-se no boxplot utilizando o algoritmo HMC.

Figura 16 – Boxplot utilizando o pacote BayesDccGarch para  $\omega = 0,9$ ,  $\alpha = 0,5$  e  $\beta = 0,4$  em modelos GARCH(1,1) univariados.



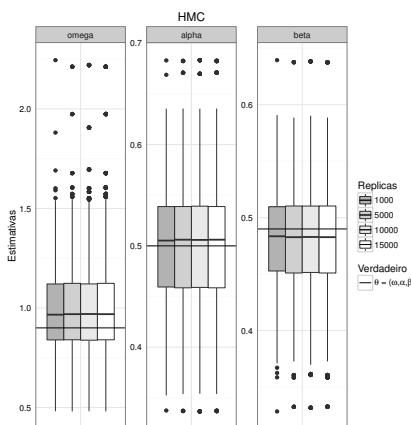
Fonte: Elaborada pelo autor.

A Tabela 2 apresenta os valores das médias das estimativas para os parâmetros do modelo considerando os algoritmos HMC e MH. Para o parâmetro  $\omega$ , observa-se que, em média, ambos os algoritmos tendem a superestimar o valor real do parâmetro. Entretanto, o valor médio estimado pelo HMC está mais próximo do valor real de  $\omega$ . Observando o parâmetro  $\alpha$ , tem-se que o valor médio considerando o HMC tende a subestimar o valor real do parâmetro, enquanto que o MH tende a superestimá-lo. Já para  $\beta$ , ambos subestimam o valor real do parâmetro mas o MH possui valores mais próximos do valor real do parâmetro.

### 3.2.5 Terceiro caso : $\alpha + \beta = 0,99$

Para o terceiro caso considere os valores  $\alpha + \beta = 0,99$  com  $\alpha = 0,5$  e  $\beta = 0,49$ , tornando-o bem mais próximo do valor 1. A Figura 17 apresenta o caso do modelo GARCH(1,1) utilizando o algoritmo HMC. Observa-se que os valores medianos das estimativas estão mais próximos do verdadeiro valor se comparados com as estimativas encontradas com o algoritmo MH (Figura 18).

Figura 17 – Boxplot utilizando o algoritmo HMC for  $\omega = 0.9$ ,  $\alpha = 0.5$  e  $\beta = 0.49$  em modelos GARCH(1,1) univariados.



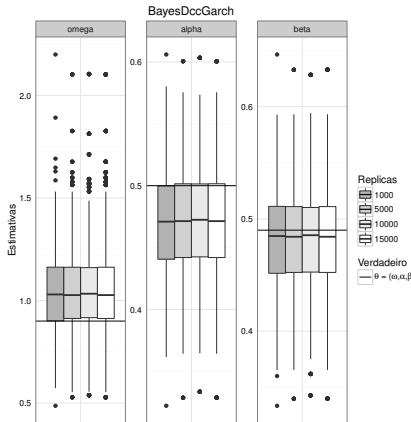
Fonte: Elaborada pelo autor.

Tabela 2 – Valores médios das estimativas, viés e EQM para  $\omega = 0.9$ ,  $\alpha = 0.5$  e  $\beta = 0.4$  considerando os algoritmo HMC e MH.

Estimativas	HMC			BayesDccGarch		
	1000	5000	10000	15000	1000	5000
$\hat{\omega}$	0,966073	0,966540	0,966490	0,966357	0,973928	0,971213
$\hat{\alpha}$	0,499040	0,499057	0,499053	0,499046	0,502021	0,501146
$\hat{\beta}$	0,384674	0,384644	0,384586	0,384552	0,386395	0,387387
Viés( $\hat{\omega}$ )	0,066073	0,066540	0,066490	0,066357	0,073928	0,071213
Viés( $\hat{\alpha}$ )	-0,000960	-0,000943	-0,000947	-0,000954	0,002021	0,001146
Viés( $\hat{\beta}$ )	-0,015326	-0,015356	-0,015414	-0,015448	-0,013605	-0,012613
EQM( $\hat{\omega}$ )	0,039005	0,038998	0,038949	0,038875	0,037168	0,036222
EQM( $\hat{\alpha}$ )	0,003451	0,003442	0,003440	0,003435	0,003039	0,003044
EQM( $\hat{\beta}$ )	0,003537	0,003529	0,037374	0,037357	0,003251	0,038089
Tempo (min)	7,205867	40,10807	69,9092	110,5291	58,6364	87,08153
					110,3011	142,9889

Fonte: Elaborada pelo autor.

Figura 18 – Boxplot utilizando pacote BayesDccGarch for  $\omega = 0.9$ ,  $\alpha = 0.5$  e  $\beta = 0.49$  em modelos GARCH(1,1) univariados.



Fonte: Elaborada pelo autor.

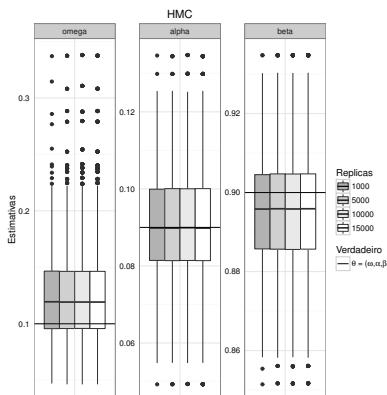
A Tabela 3 apresenta os resultados das estimativas, viés e EQM para os algoritmos HMC e MH. Neste caso, os valores do viés são menores quando considera-se o algoritmo HMC para os três parâmetros do modelo.

### 3.2.6 Quarto caso: series de tamanho 1500 e $\alpha + \beta = 0,99$

Para este caso considere 500 séries temporais de tamanho 1500 com  $\omega = 0,1$  e  $\alpha + \beta = 0,99$  com  $\alpha = 0,09$  e  $\beta = 0,9$ .

Na Figura 19 observa-se que os valores das estimativas utilizando o algoritmo HMC estão próximos do verdadeiro valor do parâmetro se comparados com o algoritmo MH (ver Figura 20).

Figura 19 – Boxplot utilizando o algoritmo HMC for  $\omega = 0.1$ ,  $\alpha = 0.09$  e  $\beta = 0.9$  em modelos GARCH(1,1) univariados.



Fonte: Elaborada pelo autor.

Tabela 3 – Valores médios das estimativas, viés e EQM para  $\omega = 0.9$ ,  $\alpha = 0.5$  e  $\beta = 0.49$  considerando os algoritmos HMC e MH.

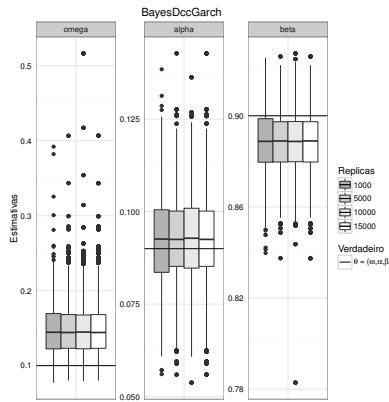
Estimativas	HMC			BayesDccGarch			
	Número de iterações MCMC 1000	Número de iterações MCMC 5000	Número de iterações MCMC 10000	Número de iterações MCMC 15000	Número de iterações MCMC 1000	Número de iterações MCMC 5000	Número de iterações MCMC 10000
$\hat{\omega}$	0,989311	0,989913	0,989753	0,989540	1,040409	1,042659	1,043461
$\hat{\alpha}$	0,501620	0,501603	0,501604	0,501611	0,471589	0,472393	0,472270
$\hat{\beta}$	0,480415	0,480309	0,480307	0,480326	0,482344	0,481627	0,481670
Viés( $\hat{\omega}$ )	0,089311	0,089913	0,089753	0,089540	0,140409	0,142659	0,143461
Viés( $\hat{\alpha}$ )	0,001620	0,001603	0,001604	0,001611	-0,028411	-0,027607	-0,027730
Viés( $\hat{\beta}$ )	-0,009585	-0,009691	-0,009693	-0,009674	-0,007656	-0,008373	-0,008330
EQM( $\hat{\omega}$ )	0,054749	0,054855	0,054622	0,054447	0,059283	0,058293	0,058521
EQM( $\hat{\alpha}$ )	0,003090	0,003088	0,003087	0,003082	0,002620	0,002519	0,002510
EQM( $\hat{\beta}$ )	0,002000	0,002001	0,002002	0,001996	0,002016	0,001952	0,001936
Tempo (min.)	7,6404	42,41133	72,55007	108,0329	52,70847	72,35673	100,4702
							133,405

Fonte: Elaborada pelo autor.

É importante notar que o algoritmo MH apresenta estimativas para  $\omega$  e  $\beta$  em que o valor real do parâmetro está abaixo do 1º quartil e acima do 3º quartil, respectivamente. Fato esse que não ocorre da mesma maneira com o algoritmo HMC.

Na [Tabela 4](#) observa-se que o algoritmo HMC apresenta melhores valores de viés e EQM para os parâmetros quando comparados com o MH. Outro fato importante é o tempo de simulação menor na maioria dos casos quando utiliza-se o algoritmo HMC.

Figura 20 – Boxplot utilizando o pacote BayesDccGarch for  $\omega = 0.1$ ,  $\alpha = 0.09$  e  $\beta = 0.9$  em modelos GARCH(1,1) univariados.



Fonte: Elaborada pelo autor.

Outra alternativa de estimação Bayesiana é considerar o uso da plataforma Stan apresentada na [Subseção 2.4.4](#). Assim um estudo de simulação é realizado para comparar as estimativas do HMC proposto neste trabalho com a estimação numérica realizada pelo pacote *RSTAN*.

### 3.2.7 HMC e o RSTAN

Nesta seção considere uma série de retornos de tamanho 1500 com  $\omega = 0,1$ ,  $\alpha = 0,09$  e  $\beta = 0,9$ . As simulações foram realizadas com tamanhos 1000, 5000 e 10000 descartando-se metade das observações como amostra de aquecimento. Para o HMC admita  $LF = 10$  e  $\varepsilon$  para uma taxa de aceitação entre 85% e 95%. O tempo total de simulação e o ESS (*Effective Size*) foram calculados para uma melhor comparação entre o HMC proposto e o RSTAN.

O ESS é dado por ([KASS et al., 1998](#))

$$ESS = \frac{M}{1 + 2 \sum_{i=1}^j \rho(k)}$$

em que  $M$  é o número de amostras MCMC pós *burn-in* e  $\sum_{i=1}^j \rho(k)$  é a soma das autocorrelações de  $k$  amostras monótonas estimadas pelo estimador de sequência monótona inicial ([GEYER, 1992](#)). O ESS é uma estimativa do número de amostras independentes necessárias para obter uma estimativa do parâmetro com a mesma precisão que uma estimativa MCMC considerando

Tabela 4 – Valores médios das estimativas, viés e EQM para  $\omega = 0.1$ ,  $\alpha = 0.09$  e  $\beta = 0.9$  considerando os algoritmos HMC e MH.

Estimativas	HMC			BayesDccGarch		
	Número de iterações MCMC			Número de iterações MCMC		
	1000	5000	10000	15000	1000	5000
$\hat{\omega}$	0,124734	0,124753	0,124752	0,124736	0,149480	0,149955
$\hat{\alpha}$	0,090644	0,090653	0,090654	0,090649	0,092849	0,092957
$\hat{\beta}$	0,895572	0,895570	0,895572	0,895575	0,888479	0,888290
Viés( $\hat{\omega}$ )	0,024734	0,024753	0,024752	0,024736	0,049480	0,049955
Viés( $\hat{\alpha}$ )	0,000644	0,000653	0,000654	0,000649	0,002849	0,002957
Viés( $\hat{\beta}$ )	-0,004428	-0,004430	-0,004428	-0,004425	-0,011521	-0,011710
EQM( $\hat{\omega}$ )	0,002252	0,002254	0,002257	0,002251	0,004051	0,004068
EQM( $\hat{\alpha}$ )	0,000208	0,000207	0,000207	0,000207	0,000202	0,000183
EQM( $\hat{\beta}$ )	0,000253	0,000253	0,000253	0,000252	0,000372	0,000372
Tempo (min.)	26,3833	100,5901	205,3958	295,9489	102,8350	152,2609
					180,4155	233,3863

Fonte: Elaborada pelo autor.

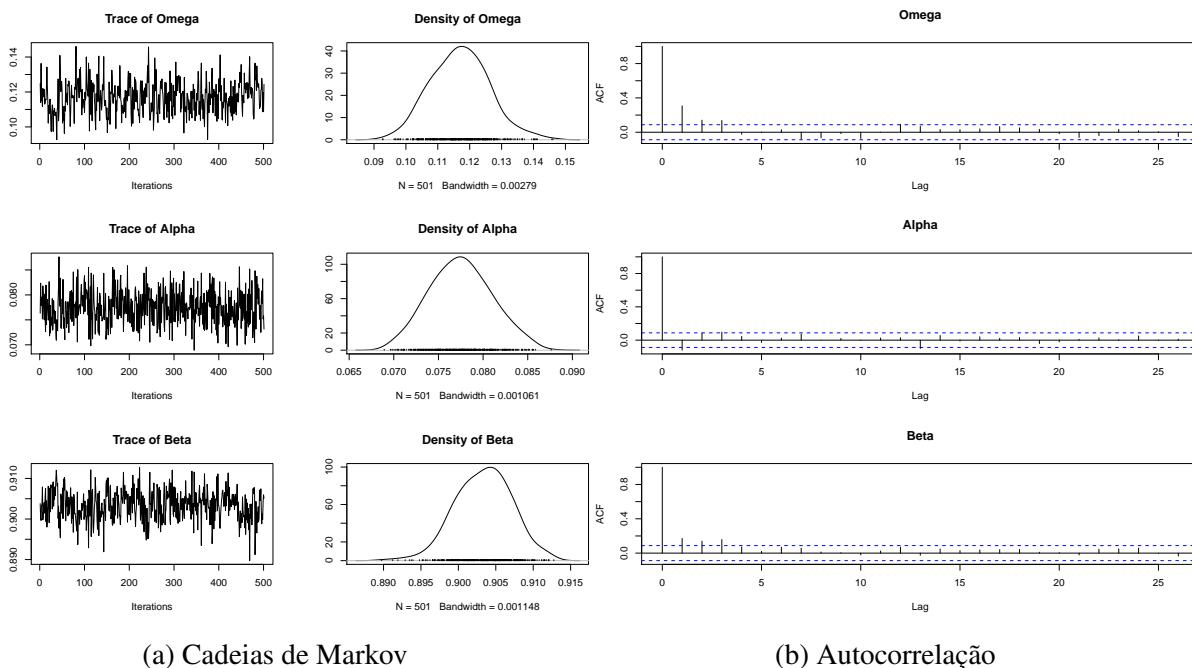
$M$  amostras dependentes. Quanto mais próximo o ESS for do tamanho de amostra pós *burn-in* mais amostras independentes estão sendo fornecidas pelo algoritmo.

Na Figura 21 tem-se o traço das Cadeias de Markov e o gráficos da função de autocorrelação para os parâmetros considerando o algoritmo HMC com 1000 simulações. Pode-se observar que este comportamento é muito similar ao apresentado pelo pacote *RSTAN* (Figura 22), tanto o comportamento persistente da Cadeia de Markov quanto as autocorrelações seriais presentes nas estimativas.

Na Figura 23 observa-se que o comportamento das estimativas dos parâmetros do modelo para o algoritmo HMC continuam similares aos comportamentos quando utiliza-se o pacote *RSTAN* (Figura 24). Com 10000 simulações essa similariedade continua presente ao se analisar as Figuras 25 e 26.

De acordo com a Tabela 5 pode-se observar que as estimativas são bastante próximas dos valores reais dos parâmetros para ambos os algoritmos. Entretanto, ao analisar o ESS tem-se que o HMC com as derivadas analíticas são maiores que o ESS calculado utilizando o pacote *rstan*. Dessa forma, pode-se afirmar que o HMC possui estimativas menos correlacionadas. Outra característica importante é o tempo de simulação que foi bem menor para o HMC proposto.

Figura 21 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 1000 simulações.

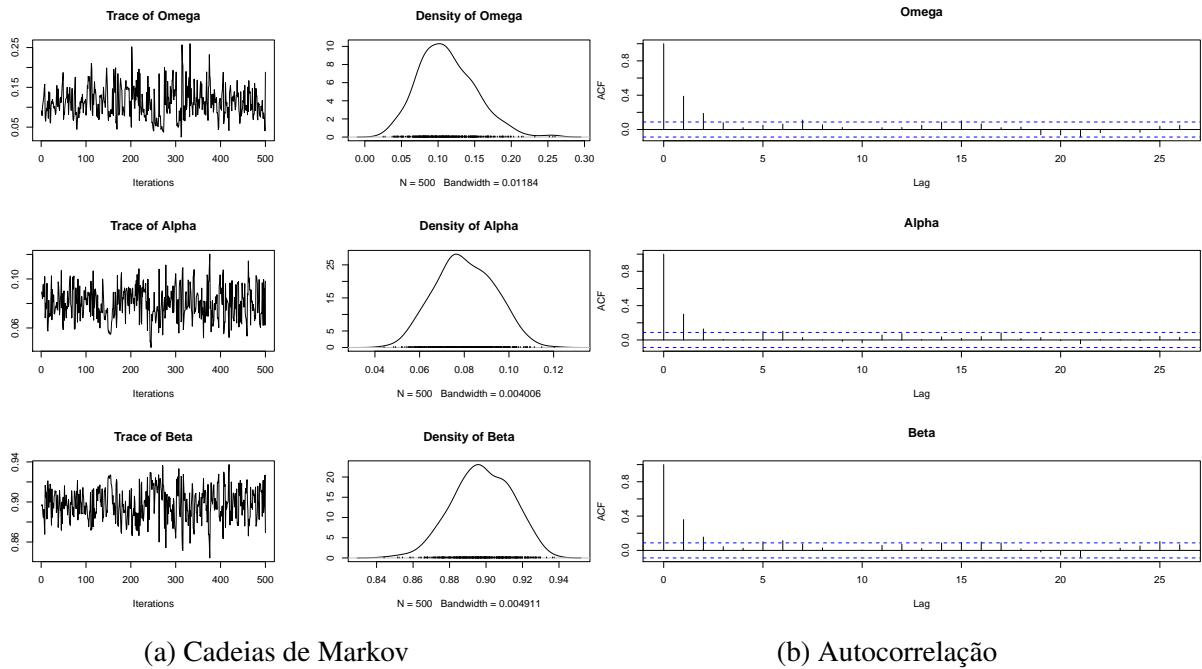


(a) Cadeias de Markov

(b) Autocorrelação

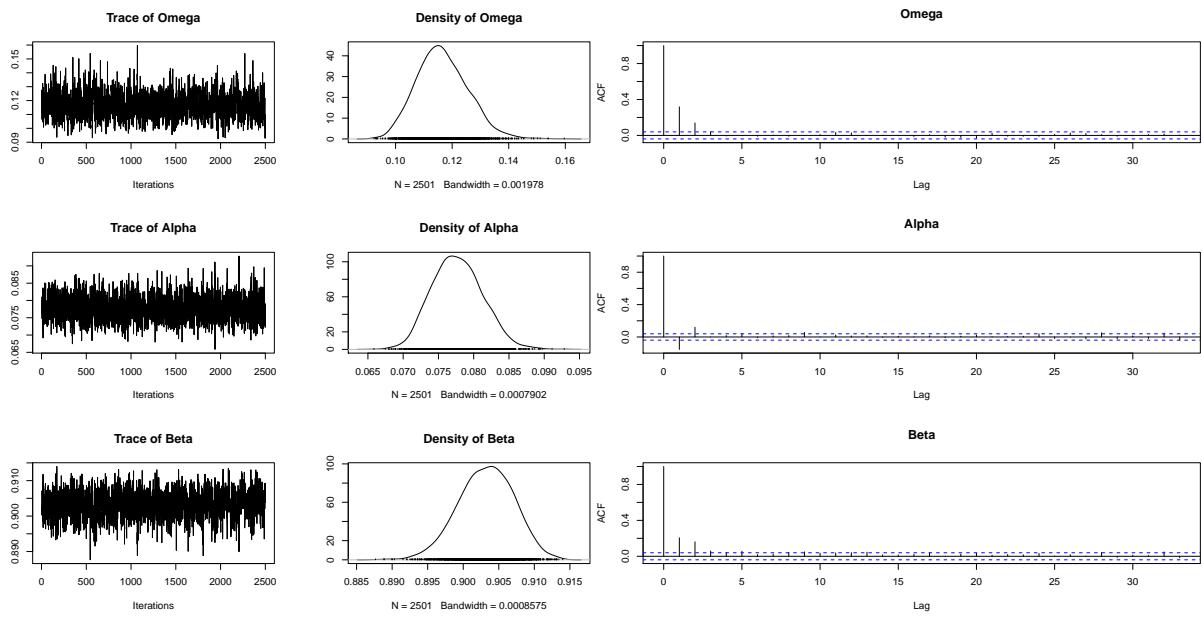
Fonte: Elaborada pelo autor.

Figura 22 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo RSTAN com 1000 simulações.



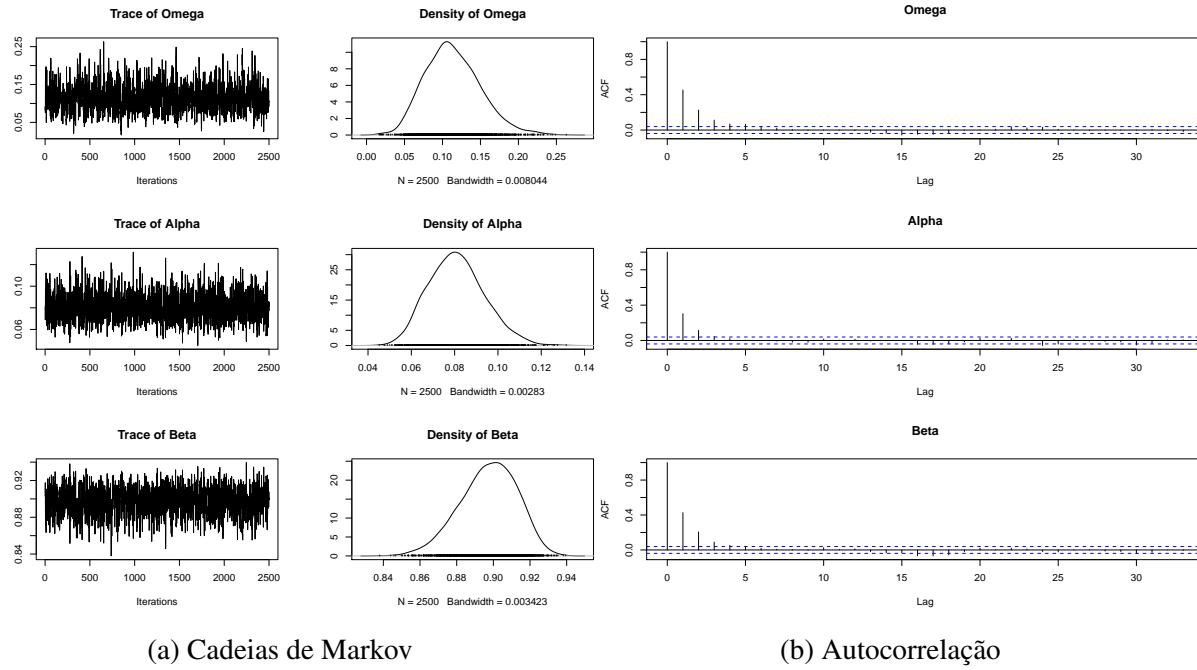
Fonte: Elaborada pelo autor.

Figura 23 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 5000 simulações.



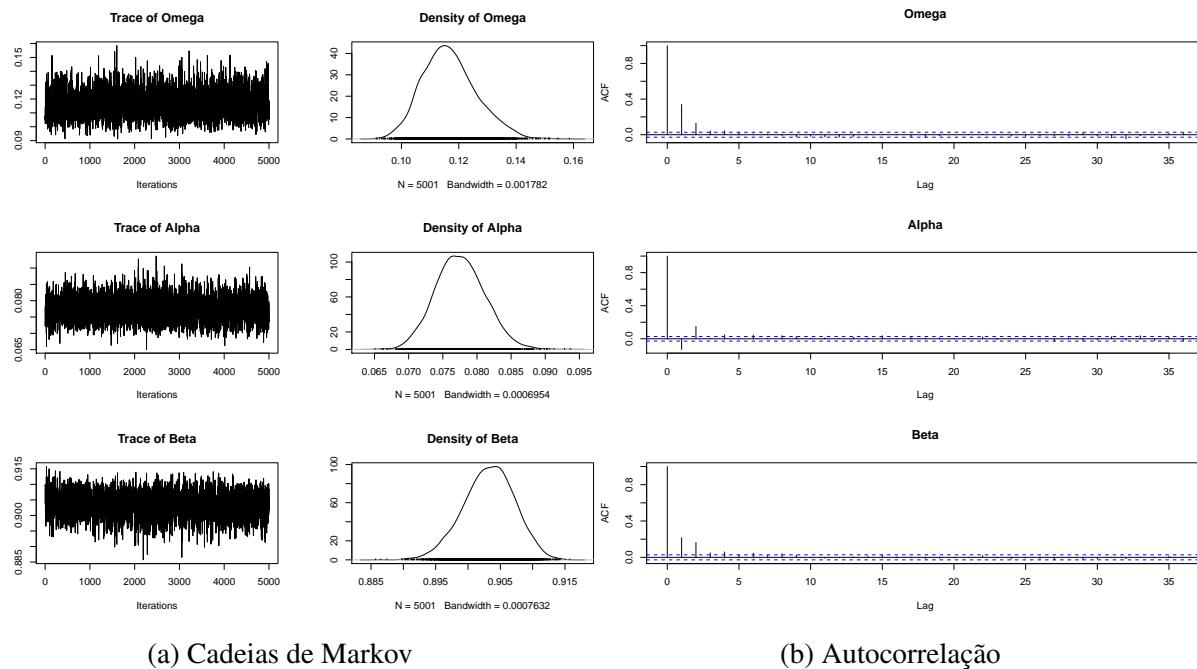
Fonte: Elaborada pelo autor.

Figura 24 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo RSTAN com 5000 simulações.



Fonte: Elaborada pelo autor.

Figura 25 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo HMC com 10000 simulações.



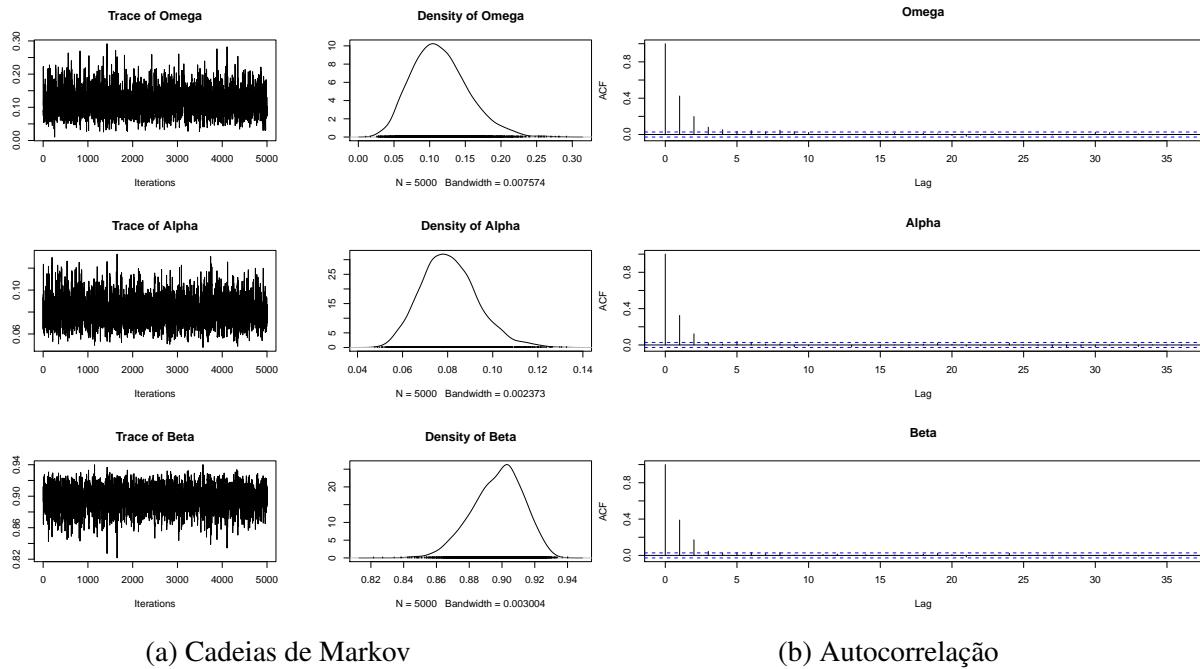
Fonte: Elaborada pelo autor.

Tabela 5 – Valores médios das estimativas para  $\omega = 0.1$ ,  $\alpha = 0.09$  e  $\beta = 0.9$  considerando os algoritmos HMC e o RSTAN

Estimativas	HMC						RSTAN		
	Número de iterações MCMC			Número de iterações MCMC			ESS/seg	10000	ESS/seg
	1000	ESS/seg	5000	ESS/seg	10000	ESS/seg			
$\hat{\omega}$	0,1169	170.36	0,1166	164.65	0,1168	158.02	0,1122	3.63	0,1138
$\hat{\alpha}$	0,0775	263.37	0,0776	273.43	0,0775	239.98	0,0803	3.49	0,0810
$\hat{\beta}$	0,9034	152.44	0,9032	302.09	0,9032	155.49	0,8980	3.88	0,8973
Tempo (seg.)	1,44	-	7,22	-	14,39	-	60,63	-	67,34
								-	77,64

Fonte: Elaborada pelo autor.

Figura 26 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o algoritmo RSTAN com 10000 simulações.



(a) Cadeias de Markov

(b) Autocorrelação

Fonte: Elaborada pelo autor.

### 3.2.8 HMC e o RSTAN usando 500 séries

Para uma comparação mais precisa entre os métodos utilizados foram geradas 500 séries de retornos de tamanho 1500 e calculados o viés e o EQM para as estimativas dos parâmetros do modelo GARCH(1,1). Além disso, gráficos Boxplots foram gerados comparando as estimativas para ambos os algoritmos. Neste estudo considere  $\omega = 0,1$ ,  $\alpha = 0,09$  e  $\beta = 0,9$  e simulações de tamanho 1000, 5000 e 10000 descartando-se metade das observações como *burn-in*. Utilizando o HMC admita  $LF = 10$  e  $\varepsilon$  de tal forma que a taxa de aceitação fique entre 85% e 95%.

A Figura 27 e a Figura 28 apresenta o Boxplot das estimativas para o algoritmo HMC e *RSTAN*, respectivamente. Observa-se que em termos da mediana as estimativas do parâmetro  $\alpha$  estão bem próximas ao valor real utilizando o algoritmo HMC. Esse comportamento não aparece da mesma maneira utilizando o algoritmo do pacote *RSTAN*. Para  $\omega$  ambos os algoritmos tendem a superestimar o verdadeiro valor do parâmetro. Entretanto ao observar o algoritmo do pacote *RSTAN* o valor real do parâmetro encontra-se antes do 1º quartil das estimativas. Esse comportamento não é observado com o método HMC. Para  $\beta$  ambos os métodos, em termos da mediana, tendem a subestimar o verdadeiro valor do parâmetro.

A Tabela 6 apresenta os valores da média das estimativas, bem como o viés e o EQM, para todos os parâmetros do modelo e tamanhos de simulação em ambos os algoritmos. Pode-se

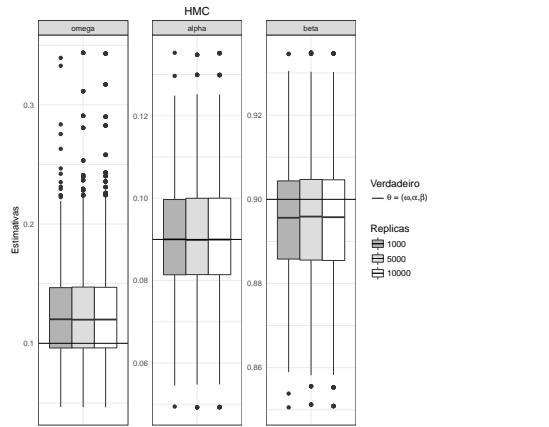
Tabela 6 – Valores médios das estimativas, viés e EQM para  $\omega = 0,1$ ,  $\alpha = 0,09$  e  $\beta = 0,9$  considerando os algoritmos HMC e RSTAN.

Estimativas	HMC			RSTAN		
	1000	5000	10000	1000	5000	10000
$\hat{\omega}$	0,12535	0,12537	0,12536	0,15008	0,15017	0,15007
$\hat{\alpha}$	0,09072	0,09070	0,09071	0,09492	0,09498	0,09496
$\hat{\beta}$	0,89543	0,89544	0,89544	0,88284	0,88279	0,88281
Viés( $\hat{\omega}$ )	0,02536	0,02537	0,02536	0,05008	0,05017	0,05007
Viés( $\hat{\alpha}$ )	0,00072	0,00070	0,00071	0,00492	0,00498	0,00496
Viés( $\hat{\beta}$ )	-0,00457	-0,00456	-0,00456	-0,01716	-0,01721	-0,01719
EQM( $\hat{\omega}$ )	0,00233	0,00232	0,00232	0,00755	0,00738	0,00735
EQM( $\hat{\alpha}$ )	0,00021	0,00021	0,00021	0,00021	0,00020	0,00020
EQM( $\hat{\beta}$ )	0,00026	0,00025	0,00025	0,00072	0,00071	0,00070
Tempo (min.)	11,95	59,54	119,32	19,21	81,59	160,72

Fonte: Elaborada pelo autor.

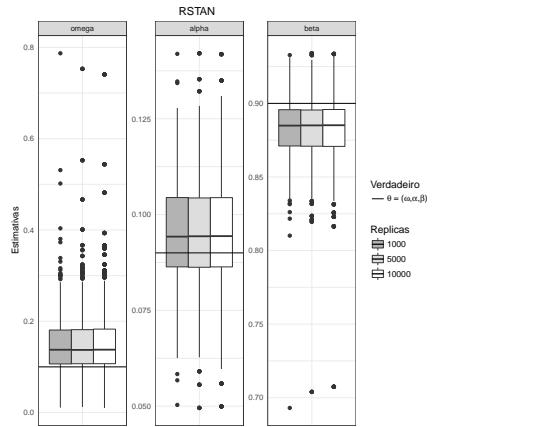
observar que o viés considerando o algoritmo HMC é menor que os calculados com o *RSTAN* para todos os parâmetros do modelo e todos os tamanhos de simulação. Nota-se também que o tempo de simulação total para o algoritmo HMC é menor que o tempo de simulação para o *RSTAN*.

Figura 27 – Boxplot utilizando HMC para  $\omega = 0.1$ ,  $\alpha = 0.09$  e  $\beta = 0.9$  em modelos GARCH(1,1) univariados.



Fonte: Elaborada pelo autor.

Figura 28 – Boxplot utilizando pacote RSTAN para  $\omega = 0.1$ ,  $\alpha = 0.09$  e  $\beta = 0.9$  em modelos GARCH(1,1) univariados.



Fonte: Elaborada pelo autor.



# MONTE CARLO HAMILTONIANO EM MODELOS GARCH MULTIVARIADOS

---



---

Neste capítulo estuda-se a aplicação do método HMC a modelos GARCH multivariados, mais precisamente ao modelo DCC-GARCH definido [Subseção 1.1.2](#). Um estudo de simulação é realizado considerando um modelo DCC-GARCH(1,1) bivariado ([Subseção 4.1.1](#)). Neste estudo uma série de retornos bivariados é simulado e os resultados do método HMC comparados com o método presente no pacote *BayesDccGarch*. Além disso, uma aplicação a dados reais é realizada na [Subseção 4.1.2](#).

## 4.1 DCC-GARCH e o HMC

Suponha o modelo DCC-GARCH apresentado na [Subseção 1.1.2](#). Considere agora que  $\boldsymbol{\varepsilon}_t$  segue uma distribuição Normal multivariada, denotada por  $\mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ , cuja função densidade de probabilidades é dada por

$$f(\mathbf{x}) = |\mathbf{I}_p|^{-1/2} (2\pi)^{p/2} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mathbf{0})' \mathbf{I}_p (\mathbf{x} - \mathbf{0}) \right]$$

com  $\mathbf{0}$  um vetor  $p \times 1$  de zeros e  $\mathbf{I}_p$  uma matriz identidade de ordem  $p$ . Assim, pode-se construir a função de verossimilhança do modelo da seguinte forma

$$\begin{aligned} L(\boldsymbol{\theta}) &= p(\mathbf{y}_1, \dots, \mathbf{y}_T | \boldsymbol{\theta}) = \prod_{t=2}^T \frac{1}{(2\pi)^{p/2} |\mathbf{H}_t|^{1/2}} \exp \left\{ -\frac{1}{2} \mathbf{y}_t' \mathbf{H}_t^{-1} \mathbf{y}_t \right\} \\ &= \prod_{t=1}^T |\mathbf{H}_t|^{-1/2} (2\pi)^{-p/2} \exp \left\{ -\frac{1}{2} \mathbf{y}_t' \mathbf{H}_t^{-1} \mathbf{y}_t \right\} \end{aligned}$$

em que  $\boldsymbol{\theta} = (w_1, \alpha_1, \beta_1, \dots, w_m, \alpha_m, \beta_m, a, b)$ . Definida a função de verossimilhança e utilizando

as prioris para os parâmetros dadas na [Subseção 1.1.2](#), a distribuição a posteriori será dada por

$$\begin{aligned}\pi(\boldsymbol{\theta}|\mathbf{y}) &\propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &\propto L(\boldsymbol{\theta})\mathcal{N}(\mu_{w_i}, \sigma_{w_i}^2)I_{(w_i>0)}\mathcal{N}(\mu_{\alpha_i}, \sigma_{\alpha_i}^2)I_{(0<\alpha_i<1)}\mathcal{N}(\mu_{\beta_i}, \sigma_{\beta_i}^2)I_{(0<\beta_i<1)} \\ &\quad \times \mathcal{N}(\mu_a, \sigma_a^2)I_{(0<a<1)}\mathcal{N}(\mu_b, \sigma_b^2)I_{(0<b<1)}\end{aligned}$$

para  $i = 1, \dots, m$  e  $(\mu_{w_i}, \sigma_{w_i}^2, \mu_{\alpha_i}, \sigma_{\alpha_i}^2, \mu_{\beta_i}, \sigma_{\beta_i}^2, \mu_a, \sigma_a^2, \mu_b, \sigma_b^2)$  são hiperparâmetros. Para aplicar o método Monte Carlo Hamiltoniano é necessário que todos os parâmetros sejam simulados em linha reta. Disto, considere as seguintes transformações nos parâmetros do modelo  $\phi_1 = \log(w_1), \phi_2 = \log\left(\frac{\alpha_1}{1-\alpha_1}\right), \phi_3 = \log\left(\frac{\beta_1}{1-\beta_1}\right), \dots, \phi_{3m-2} = \log(w_m), \phi_{3m-1} = \log\left(\frac{\alpha_2}{1-\alpha_2}\right), \phi_{3m} = \log\left(\frac{\beta_m}{1-\beta_m}\right), \phi_{3m+1} = \log\left(\frac{a}{1-a}\right), \phi_{3m+2} = \log\left(\frac{b}{1-b}\right)$ . Os resultados na escala original são encontrados realizando a transformação inversa nos valores simulados.

Ao realizar a transformação, uma nova parametrização será utilizada para encontrar a distribuição a posteriori dos parâmetros  $\boldsymbol{\phi}$ . Considere  $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_{3m+1}, \phi_{3m+2})'$  e utilizando o Método Jacobiano,

$$\pi(\boldsymbol{\phi}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})|J(\boldsymbol{\phi})| \quad (4.1)$$

em que

$$p(\mathbf{y}|\boldsymbol{\theta}) = \prod_{t=1}^T \left[ \prod_{i=1}^m \left( \omega_i + \alpha_i y_{i,t-1}^2 + \beta_i h_{ii,t-1} \right)^{-1/2} \right] |\mathbf{R}_t|^{-1/2} p_{\boldsymbol{\epsilon}_t} \left( (\mathbf{D}_t \mathbf{R}_t \mathbf{D}_t)^{-1/2} \mathbf{y}_t \right),$$

$p(\boldsymbol{\phi})$  representam as prioris dos parâmetros definidos na [Subseção 1.1.2](#) e

$$|J(\boldsymbol{\phi})| = \begin{vmatrix} e^{\phi_1} & 0 & \cdots & 0 \\ 0 & \frac{e^{\phi_2}}{(1+e^{\phi_2})^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{e^{\phi_{3m+2}}}{(1+e^{\phi_{3m+2}})^2} \end{vmatrix} = \frac{e^{\phi_1+\phi_2+\cdots+\phi_{3m+2}}}{(1+e^{\phi_2})^2 \cdots (1+e^{\phi_{3m+2}})^2}$$

Aplicando o logaritmo na [Equação 4.1](#)

$$\log \pi(\boldsymbol{\phi}|\mathbf{y}) = \log(p(\mathbf{y}|\boldsymbol{\theta})) + \log(p(\boldsymbol{\theta})) + \log(|J(\boldsymbol{\phi})|) + c$$

em que  $c$  representa uma constante que não depende dos parâmetros,

$$\log(p(\mathbf{y}|\boldsymbol{\theta})) = \ell(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^T [p \ln(2\pi) + \ln(|\mathbf{H}_t|) + \mathbf{y}'_t \mathbf{H}_t^{-1} \mathbf{y}_t],$$

$\log(p(\boldsymbol{\phi}))$  representa o logaritmo da distribuição a priori dos parâmetros transformados e

$$\log |J(\boldsymbol{\phi})| = \phi_1 + \phi_2 + \cdots + \phi_{3m+2} - \log(1+e^{\phi_2})^2 - \cdots - \log(1+e^{\phi_{3m+2}})^2$$

Desta forma o escore da função log-verossimilhança será dado por

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} = \sum_{t=1}^n -\frac{1}{2} Tr \left( \mathbf{H}_t^{-1} \frac{\partial \mathbf{H}_t}{\partial \theta_i} \right) - \frac{1}{2} \mathbf{y}'_t \left( -\mathbf{H}_t^{-1} \frac{\partial \mathbf{H}_t}{\partial \theta_i} \mathbf{H}_t^{-1} \right) \mathbf{y}_t.$$

Todos os detalhes e cálculos para definir as derivadas com relação a  $\boldsymbol{\theta}$  encontram-se no Apêndice A. As derivadas do logaritmo da distribuição à posteriori serão dadas por

$$\frac{\partial \log(\pi(\boldsymbol{\phi}|y))}{\partial \phi_i} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} + \frac{\partial \log(p(\boldsymbol{\theta}))}{\partial \theta_i} + \frac{\partial \log|J(\boldsymbol{\phi})|}{\partial \phi_i}. \quad (4.2)$$

Definidos a função (4.1) e as derivadas (4.2) da distribuição à posteriori, o método HMC dado na Subseção 2.4.2 pode ser utilizado. Desta forma, na Subseção 4.1.1 um estudo de simulação é realizado utilizando o algoritmo HMC para a estimativa dos parâmetros do modelo DCC-GARCH(1,1). Além disso, os resultados são comparados com o algoritmo Metropolis-Hastings presente no pacote *BayesDccGarch*.

### 4.1.1 Caso Bivariado

Considere um vetor de dados simulados  $y_t$  gerado por um processo DCC-GARCH(1,1). Os valores são simulados considerando a seguinte configuração

$$\begin{aligned} \mathbf{R} &= \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \\ h_{11,t} &= 0.03 + 0.07y_{1,t-1}^2 + 0.88h_{11,t-1} \\ h_{22,t} &= 0.04 + 0.04y_{2,t-1}^2 + 0.91h_{22,t-1} \\ \mathbf{Q}_t &= (1 - 0.03 - 0.9)\mathbf{R} + 0.03\mathbf{u}_{t-1}\mathbf{u}'_{t-1} + 0.9\mathbf{Q}_{t-1}. \end{aligned}$$

Neste exemplo considera-se dois modelos GARCH univariados com alta persistência. Assim,  $\alpha + \beta$  do primeiro e do segundo modelo GARCH são próximos de 1.

Durante o processo de simulação foi gerado um conjunto de dados de tamanho 2000. Além disso, foram geradas Cadeias de Markov de tamanho 100 mil descartando-se as primeiras 50 mil observações como amostra de aquecimento. Para os hiperparâmetros das distribuições à priori foram escolhidos os valores  $\mu_{\omega_i} = \mu_{\alpha_i} = \mu_{\beta_i} = \mu_a = \mu_b = 0$  e  $\sigma_{\omega_i}^2 = \sigma_{\alpha_i}^2 = \sigma_{\beta_i}^2 = \sigma_a^2 = \sigma_b^2 = 100$  com  $i = 1, 2$ . No processo de geração dos valores simulados foi utilizado o pacote *ccgarch*. Este pacote além de simular valores DCC-GARCH possui rotinas de estimativa dos parâmetros utilizando a inferência frequentista. Todo o processo de simulação e estimativa foi realizado na linguagem R com os algoritmos de estimativa implementados na linguagem C.

Para o método HMC considere  $LF = 30$  e  $\varepsilon = 0.005$  para uma taxa de aceitação em torno de 95%. O tempo total de simulação para o método HMC foi de 6h e 45 min. Já para o método presente no pacote *BayesDccGarch* foi de 3 min e 51seg. Essa diferença em relação

ao tempo de simulação ocorre devido o valor alto do LF e pelas derivadas analíticas para o DCC-GARCH(1,1) que envolvem vários produtos matriciais.

Na [Figura 29](#) apresenta-se o traço das Cadeias de Markov e função de autocorrelação para os parâmetros do modelo DCC-GARCH(1,1) considerando o algoritmo HMC. Ao utilizar o modelo DCC-GARCH(1,1) com duas séries temporais um total de 8 parâmetros devem ser estimados. Para os parâmetros  $\omega_1$ ,  $\alpha_1$  e  $\beta_1$  observa-se um comportamento persistente das cadeias e funções de autocorrelação com baixa autocorrelação serial. Observando  $\omega_2$ ,  $\alpha_2$  e  $\beta_2$  tem-se um comportamento similar. Já para o parâmetro  $b$  o traço da Cadeia de Markov apresenta um comportamento pouco persistente e com forte autocorrelação serial.

Na [Figura 30](#) utiliza-se o algoritmo presente no pacote *BayesDccGarch* para a estimativa dos parâmetros. Ao observar as estimativas para os parâmetros  $\omega_1$ ,  $\alpha_1$  e  $\beta_1$ , as Cadeias de Markov apresentam um comportamento persistente mas com autocorrelação maior que as estimativas utilizando o algoritmo HMC. Mesmo comportamento presente para os parâmetros  $\omega_2$ ,  $\alpha_2$  e  $\beta_2$ . Entretanto, para o parâmetro  $b$  o traço da Cadeia de Markov apresenta um comportamento mais persistente e com menor autocorrelação ao compará-lo com o algoritmo HMC.

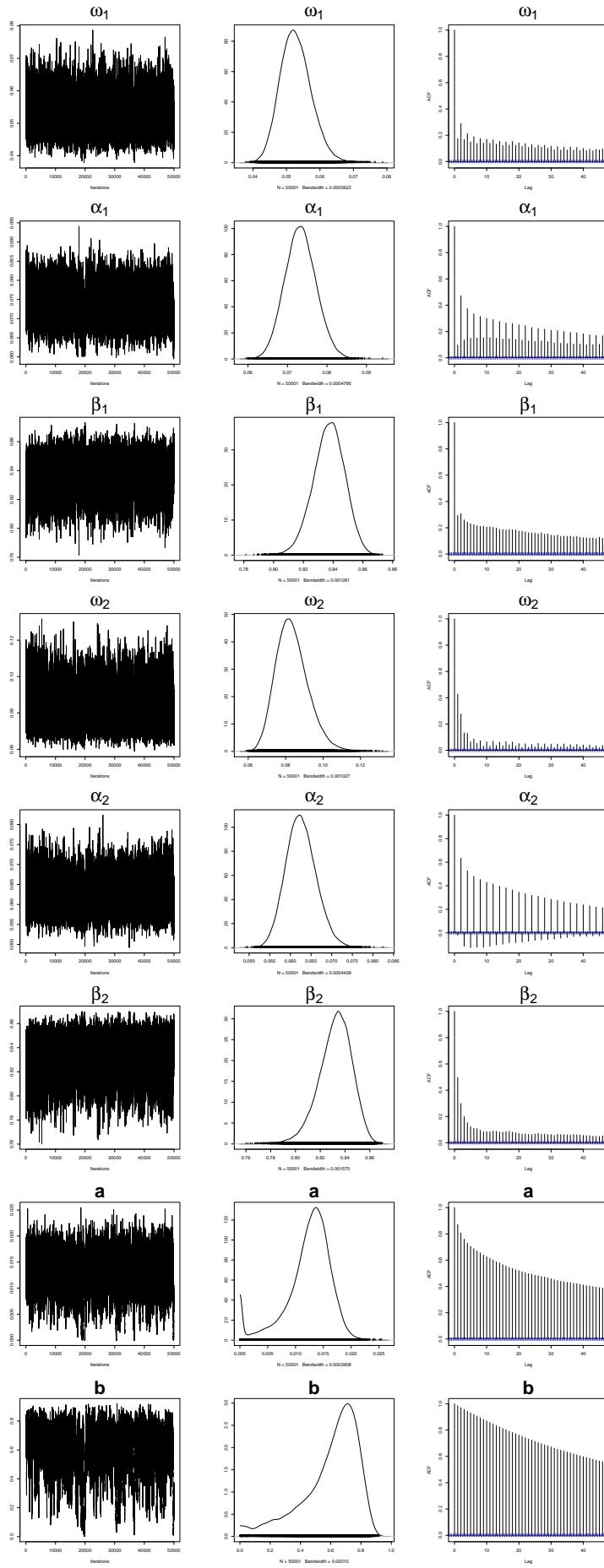
A [Tabela 7](#) apresenta a média das estimativas e os quantis amostrais de 5%, 50% (Mediana) e 95% para os parâmetros do modelo considerando o método HMC e *BayesDccGarch*, respectivamente.

Tabela 7 – Estimativas para o DCC-GARCH(1,1) considerando os métodos HMC e o presente no pacote *BayesDccGarch* com 100 mil iterações;

HMC						
Estimativas	Média	Desvio Padrão	2.5%	Mediana	97.5%	
$\hat{\omega}_1$	0.05282	0.00463	0.044536	0.05254	0.06267	
$\hat{\alpha}_1$	0.07333	0.00393	0.065834	0.07326	0.08130	
$\hat{\beta}_1$	0.83713	0.01060	0.815432	0.83749	0.85690	
$\hat{\omega}_2$	0.08354	0.00865	0.068807	0.08276	0.10260	
$\hat{\alpha}_2$	0.06258	0.00367	0.055797	0.06242	0.07019	
$\hat{\beta}_2$	0.83213	0.01328	0.803196	0.83318	0.85526	
$\hat{a}$	0.01243	0.00394	0.000775	0.01306	0.01849	
$\hat{b}$	0.60024	0.18436	0.123135	0.64770	0.84115	
BayesDccGarch						
$\hat{\omega}_1$	0.05859	0.01504	0.035913	0.05636	0.09458	
$\hat{\alpha}_1$	0.07427	0.01279	0.052090	0.07343	0.10198	
$\hat{\beta}_1$	0.82650	0.03318	0.748786	0.83053	0.87935	
$\hat{\omega}_2$	0.10457	0.03449	0.055958	0.09789	0.18828	
$\hat{\alpha}_2$	0.06683	0.01308	0.043651	0.06584	0.09509	
$\hat{\beta}_2$	0.80206	0.05040	0.681615	0.81090	0.87561	
$\hat{a}$	0.01842	0.01261	0.001152	0.01635	0.04794	
$\hat{b}$	0.50022	0.23456	0.044605	0.52727	0.88160	

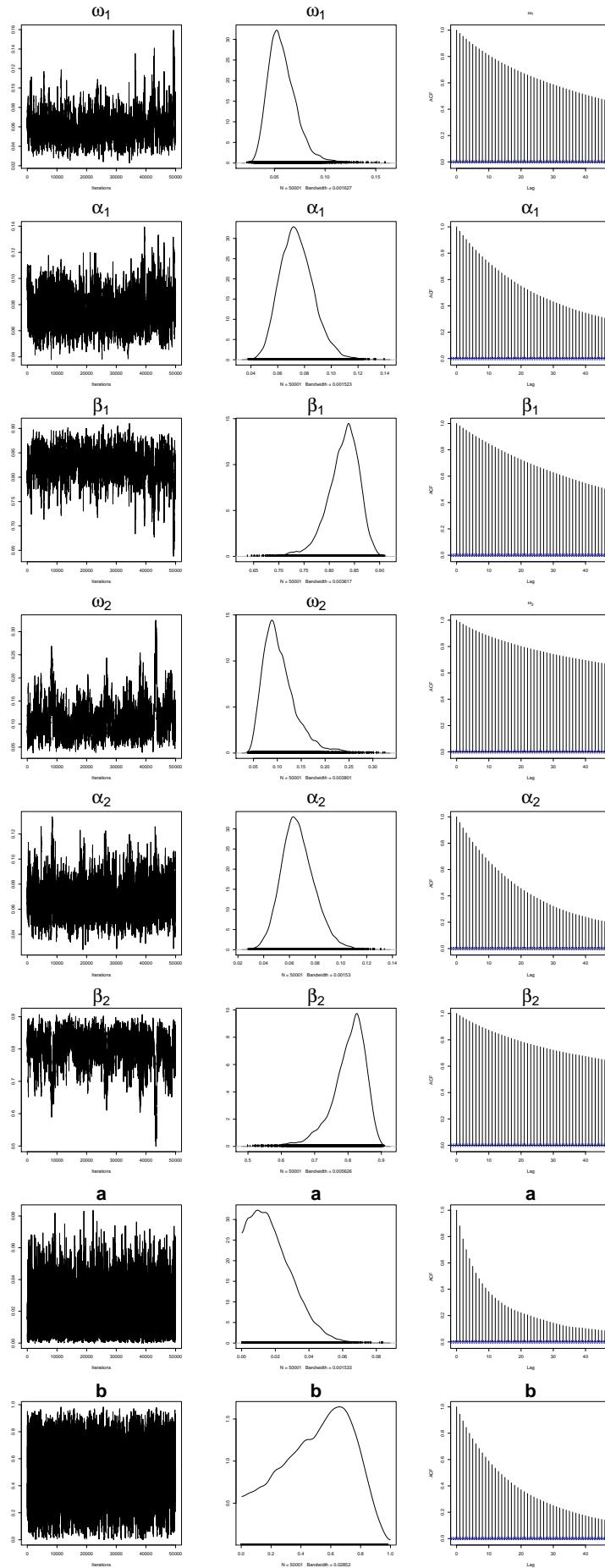
Fonte: Elaborada pelo autor.

Figura 29 – Gráfico do traço das Cadeias de Markov e das autocorrelações para o algoritmo HMC com 100 mil simulações.



Fonte: Elaborada pelo autor.

Figura 30 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o método presente no BayesDccGarch com 100 mil simulações.



Fonte: Elaborada pelo autor.

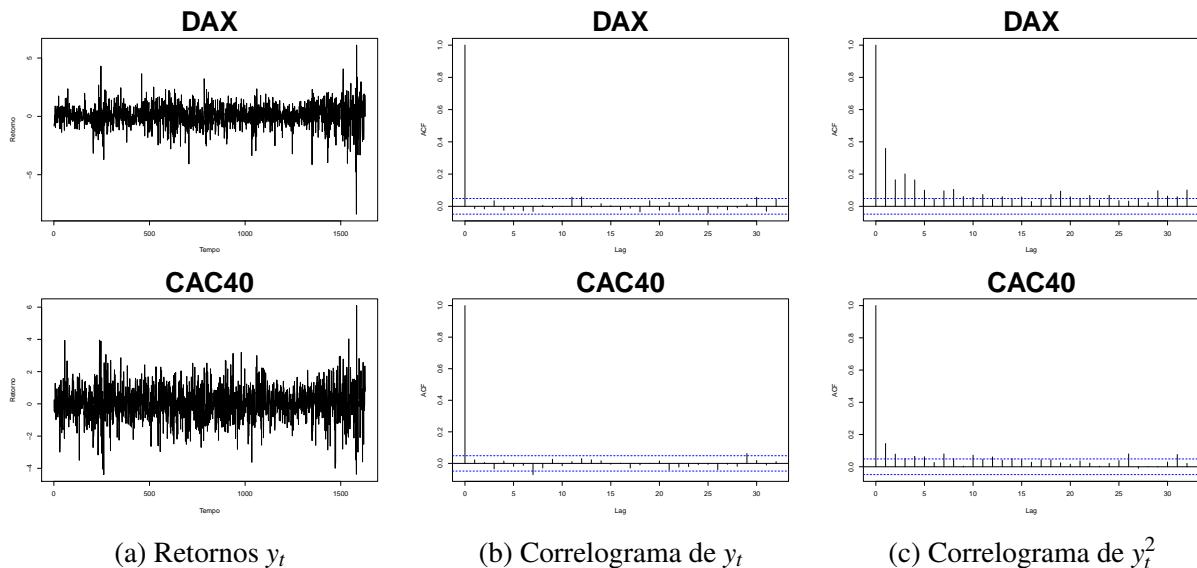
### 4.1.2 Aplicação à Dados Reais

Nesta seção um estudo é realizado com dados reais. Esse conjunto de dados está presente no pacote *BayesDccGarch* e é formado pelos retornos diários das bolsas de valores de Frankfurt (DAX) e Paris (CAC40). Um total de 1627 dias foram observados. Para mais detalhes sobre esse conjunto de dados veja [Fioruci, Ehlers e Andrade \(2014\)](#).

Nesta aplicação considere  $LF = 30$  e  $\varepsilon = 0.002$  para uma taxa de aceitação próxima a 95%. As distribuições à priori e os valores para os hiperâmetros são os mesmos definidos na [Subseção 4.1.1](#). O tempo total de simulação para o HMC foi de 5h e 24 min. Para o BayesDccGarch o tempo total foi de 4min e 13seg.

A [Figura 31](#) apresenta os gráficos dos retornos financeiros, funções de autocorrelação dos retornos e dos retornos ao quadrado para o DAX e CAC40. Observe que tanto para o DAX quanto para o CAC40 apresentam correlações significativas quando leva-se em consideração os retornos ao quadrado.

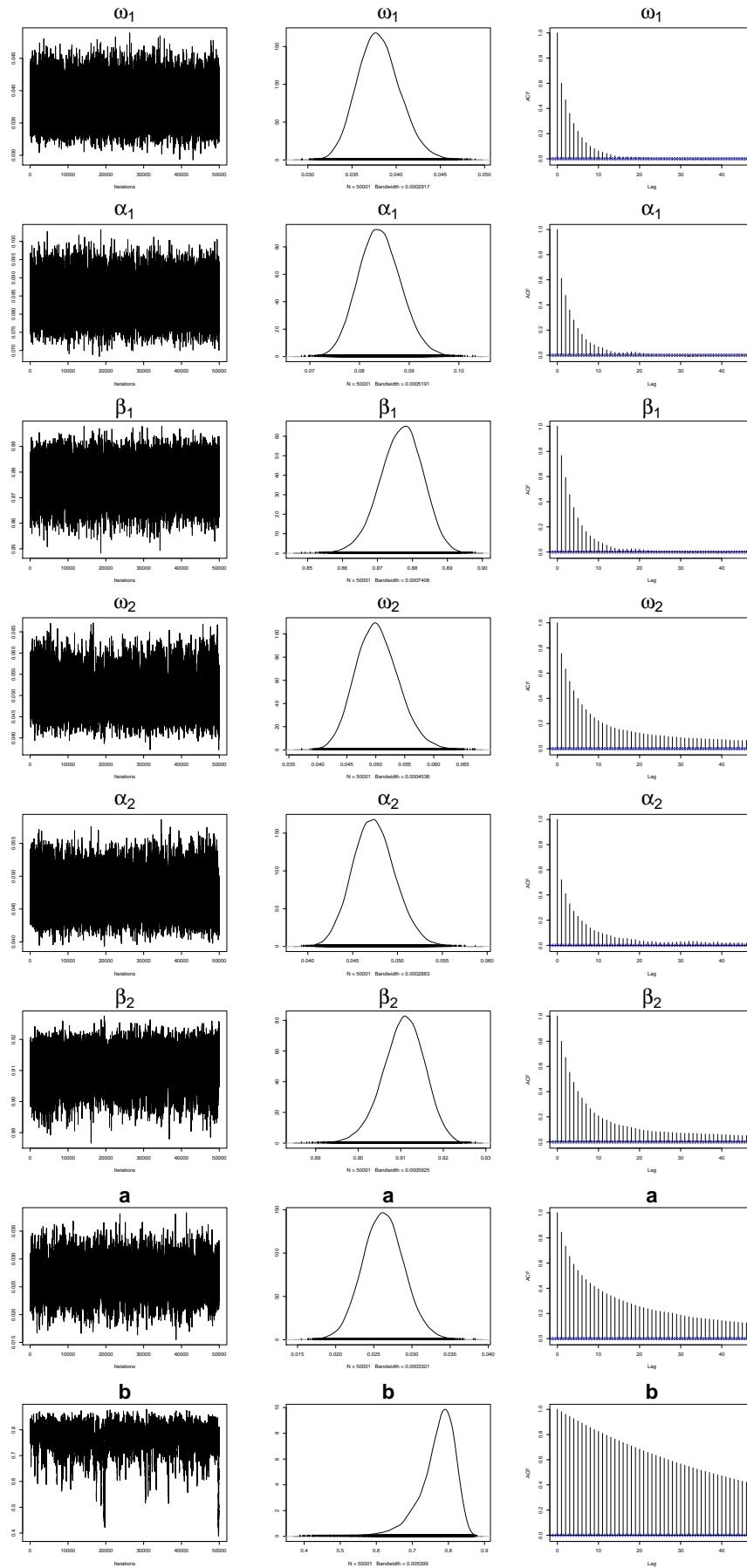
Figura 31 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o método HMC com 100 mil simulações.



Fonte: [Fioruci, Ehlers e Andrade \(2014\)](#).

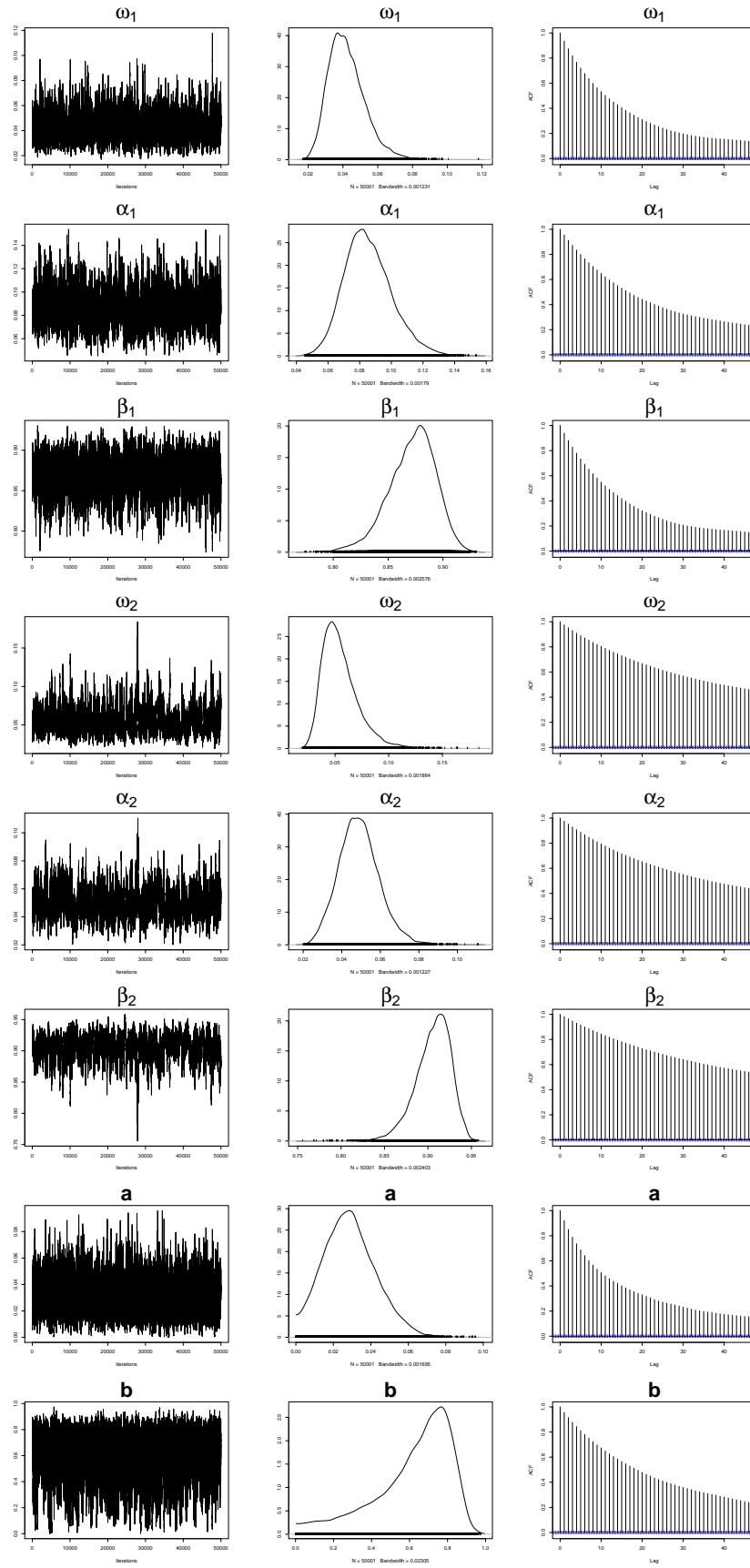
A [Figura 32](#) apresenta os gráficos do traço das Cadeias de Markov e funções de autocorrelação para as estimativas dos parâmetros do modelo DCC-GARCH(1,1) utilizando o método HMC. As estimativas para  $\omega_1$ ,  $\alpha_1$  e  $\beta_1$  apresentam um comportamento persistente com baixa autocorrelação serial. Já para  $\omega_2$ ,  $\alpha_2$ ,  $\beta_2$  e  $a$  observa-se um comportamento mais persistente com maior presença de autocorrelação. Ao analisar o parâmetro  $b$  tem-se um comportamento persistente mas com forte autocorrelação serial.

Figura 32 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o método HMC com 100 mil simulações.



Fonte: Elaborada pelo autor.

Figura 33 – Gráfico das Cadeias de Markov (a) e das autocorrelações (b) para o método presente no BayesDccGarch com 100 mil simulações.



Fonte: Elaborada pelo autor.

A [Figura 33](#) apresenta o traço das Cadeias de Markov e funções de autocorrelação para os parâmetros do modelo utilizando o método presente no BayesDccGarch. Os parâmetros  $\omega_1$ ,  $\alpha_1$  e  $\beta_1$  possuem um comportamento estacionário mas com menos intensidade se comparado ao método HMC. Além disso, há uma maior autocorrelação serial quando comparados a [Figura 32](#). Para os parâmetros  $\omega_2$ ,  $\alpha_2$  e  $\beta_2$  este comportamento permanece. Mas ao analisar o parâmetro  $b$  nota-se um comportamento mais estacionário e menos autocorrelacionado utilizando o método BayesDccGarch.

A [Tabela 8](#) exibe a média das estimativas dos parâmetros e os quantis 5%, 50% (Mediana) e 95% para o método HMC e BayesDccGarch.

Tabela 8 – Estimativas para o DCC-GARCH(1,1) considerando os métodos HMC e o presente no pacote BayesDccGarch com 100 mil iterações;

Estimativas	HMC				
	Média	Desvio Padrão	2.5%	Mediana	97.5%
$\hat{\omega}_1$	0.03806	0.002402	0.03367	0.03794	0.04310
$\hat{\alpha}_1$	0.08408	0.004264	0.07611	0.08395	0.09283
$\hat{\beta}_1$	0.87683	0.006082	0.86427	0.87705	0.88812
$\hat{\omega}_2$	0.05027	0.003727	0.04349	0.05009	0.05813
$\hat{\alpha}_2$	0.04740	0.002375	0.04297	0.04732	0.05230
$\hat{\beta}_2$	0.91065	0.004906	0.90035	0.91084	0.91970
$\hat{a}$	0.02626	0.002717	0.02105	0.02622	0.03174
$\hat{b}$	0.76886	0.053395	0.63713	0.77901	0.84236
BayesDccGarch					
$\hat{\omega}_1$	0.04218	0.01051	0.025102	0.04093	0.06648
$\hat{\alpha}_1$	0.08486	0.01502	0.058550	0.08360	0.11770
$\hat{\beta}_1$	0.87226	0.02138	0.825437	0.87411	0.90935
$\hat{\omega}_2$	0.05490	0.01665	0.030871	0.05201	0.09502
$\hat{\alpha}_2$	0.04892	0.01051	0.030132	0.04825	0.07184
$\hat{\beta}_2$	0.90571	0.02040	0.859626	0.90824	0.93871
$\hat{a}$	0.02868	0.01394	0.004886	0.02765	0.05866
$\hat{b}$	0.62413	0.19921	0.118255	0.67669	0.87954

Fonte: Elaborada pelo autor.



## CONCLUSÃO

---



---

Neste trabalho foi apresentado uma nova forma de estimação Bayesiana para modelos GARCH utilizando o método Monte Carlo Hamiltoniano (HMC). Foi realizado uma comparação entre este método e o método Metropolis-Hastings presente no pacote *BayesDccGarch*. Nesse estudo observou-se que o método HMC, na maioria das vezes, possui Cadeias de Markov com comportamento mais persistente, menos autocorrelacionadas e com convergência mais rápida ao utilizar um número menor de observações.

Ao utilizar o modelo GARCH univariado, notou-se que na maioria dos casos o método HMC apresentou Cadeias de Markov mais persistentes e com menor autocorrelação serial quando comparado ao método presente no pacote *BayesDccGarch*. Além disso, apresentou um menor tempo de simulação. Outra característica marcante foi um maior valor de ESS quando comparado com a metodologia presente no pacote *rstan*.

No [Capítulo 3](#) um estudo com 500 séries temporais simuladas é realizado para comparação do HMC com o Metropolis-Hastings. Neste estudo observou-se que o método HMC possui menor viés e EQM para a maioria dos casos estudados. Ademais, foram construídos gráficos boxplot sugerindo que em termos de mediana as estimativas do método HMC estão mais próximas do valor real dos parâmetros para a maioria dos casos.

Quando utiliza-se o HMC em modelos GARCH multivariados exige-se um número alto de LF o que torna o tempo de simulação elevado. Alinhado a este fato há também a dificuldade no cálculo das derivadas que envolvem cálculos matriciais que contribuem para o aumento do tempo de simulação total. Ao realizar um estudo de simulação o método HMC apresentou cadeias com comportamentos mais persistentes e com menor autocorrelação serial quando comparado ao *BayesDccGarch* para a maioria dos parâmetros. Entretanto, este fato não ocorre ao analisar o parâmetro  $b$  do modelo DCC-GARCH(1,1). Por fim, um estudo com dados reais é produzido em que o método HMC apresenta melhores resultado do que o método presente no *BayesDccGarch* para a maioria dos parâmetros.

## 5.1 Trabalhos futuros

Algumas propostas para trabalhos futuros:

- Propor distribuições para os erros além da Normal. Por exemplo, distribuição *t*-Student e GED;
- Propor distribuições assimétricas para o erro. Por exemplo, *t*-Student assimétrica;
- Atualizar o método HMC utilizando como proposta o NUTS (*No-U-Turn Sampler*);
- Utilizar outras variações do HMC, como por exemplo o Monte Carlo Hamiltoniano em variedade Riemanniana (RMHMC);
- Calcular estimativas VaR (Valor em Risco);
- Propor o uso do HMC para outros modelos: modelos de volatilidade estocástica e modelos da família GARCH assimétrica, por exemplo os modelos EGARCH.

## REFERÊNCIAS

---



---

- ALEXANDER, C. **Market Models: A Guide to Financial Data Analysis**. J. Wiley & Sons, 2001. Disponível em: <<http://www.amazon.com/Market-Models-Guide-Financial-Analysis/dp/0471899755>>. Citado nas páginas 26 e 27.
- ARDIA, D. **Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications**, Springer, 2008. Disponível em: <<http://www.springer.com/economics/econometrics/book/978-3-540-78656-6>>. Citado na página 26.
- \_\_\_\_\_. Bayesian estimation of a Markov-switching threshold asymmetric GARCH model with Student-t innovations. **Econometrics Journal**, v. 12, n. 1, p. 105–126, 2009. Citado na página 26.
- AUSIN, M. C.; GALEANO, P. Bayesian estimation of the Gaussian mixture GARCH model. **Computational Statistics & Data Analysis**, v. 51, p. 2636–2652, 2007. Citado na página 26.
- BAILLIE, R.; BOLLERSLEV, T. The message in daily exchange rates: A conditional variance tale. **Journal of Business and Economic Statistics**, v. 7, p. 297–305, 1989. Disponível em: <[http://public.econ.duke.edu/~boller/Published\\_Papers/jbes\\_89.pdf](http://public.econ.duke.edu/~boller/Published_Papers/jbes_89.pdf)>. Citado na página 26.
- BAUWENS, L.; LAURENT, S. A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. **Journal of Business & Economic Statistics**, v. 23, p. 346–342, 2005. Disponível em: <<http://amstat.tandfonline.com/doi/abs/10.1198/073500104000000523#.U25wfXWx3t4>>. Citado na página 26.
- BAUWENS, L.; LAURENT, S.; ROMBOUTS, J. V. K. Multivariate GARCH models: a survey. **Journal of Applied Econometrics**, v. 21, p. 79–109, 2006. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.132.5170&rep=rep1&type=pdf>>. Citado na página 27.
- BAUWENS, L.; LUBRANO, M. Bayesian option pricing using asymmetric GARCH models. **Journal of Empirical Finance**, v. 9, p. 321–342, 2002. Citado na página 26.
- BAUWENS, L.; PREMINGER, A.; ROMBOUTS, J. V. Theory and inference for a Markov switching GARCH model. **The Econometrics Journal**, v. 13, p. 218–244, 2010. Citado na página 26.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, v. 31, p. 307–327, 1986. Citado nas páginas 23 e 24.
- \_\_\_\_\_. Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. **Review of Economics and Statistics**, v. 72, p. 498–505, 1990. Disponível em: <[http://public.econ.duke.edu/~boller/Published\\_Papers/restat\\_90.pdf](http://public.econ.duke.edu/~boller/Published_Papers/restat_90.pdf)>. Citado na página 27.
- BOLLERSLEV, T.; CHOU, R.; KRONER, K. ARCH modelling in finance: A review of the theory and empirical evidence. **Journal of Econometrics**, v. 52(1), p. 5–59, 1992. Citado na página 25.

BOLLERSLEV, T.; ENGLE, R. F. Common persistence in conditional variances. **Econometrica**, v. 61, p. 167–186, 1993. Citado na página 27.

BOLLERSLEV, T.; ENGLE, R. F.; WOOLDRIDGE, J. A capital asset pricing model with time varying covariances. **Journal of Political Economy**, v. 96, p. 116–131, 1988. Citado na página 27.

BURDA, M.; MAHEU, J. M. Bayesian adaptively updated Hamiltonian Monte Carlo with an application to high-dimensional BEKK GARCH models. **Studies in Nonlinear Dynamics and Econometrics**, v. 17, p. 345–372, 2013. Citado nas páginas 35 e 36.

CALDERHEAD, B. **Differential geometric MCMC methods and applications**. Tese (Doutorado) — University of Glasgow, 2012. Citado nas páginas 34, 38 e 39.

CHAVEL, I. **Riemannian Geometry: A Modern Introduction**. Second edition. [S.l.]: Cambridge University Press, 2006. Citado na página 38.

CHUANG, I.-Y.; LU, J.-R.; LEE, P.-H. Forecasting volatility in the financial markets: a comparison of alternative distributional assumptions. **Applied Financial Economics**, n. 17:13, p. 1051–1060, 2007. Citado na página 26.

DUANE, S.; KENNEDY, A. D.; PENDLETON, B. J.; ROWETH, D. Hybrid Monte Carlo. **Physics Letter B**, v. 195 (2), p. 216 – 222, 1987. Citado na página 34.

EHLERS, R. S. Computational tools for comparing asymmetric GARCH models via Bayes factors. **Mathematics and Computers in Simulation**, v. 82, p. 858–867, 2012. Citado na página 26.

ENGLE, R. F. Autoregressive conditional heterokedasticity with estimates of the variance of U.K. inflation. **Econometrica**, v. 50, p. 987–1007, 1982. Disponível em: <<http://www.jstor.org/discover/10.2307/1912773?uid=2134&uid=2&uid=70&uid=4&sid=21104133807973>>. Citado nas páginas 23 e 24.

\_\_\_\_\_. Dynamic conditional correlation - a simple class of multivariate GARCH models. **Journal of Business and Economic Statistics**, v. 20, p. 339–350, 2002. Disponível em: <<http://archive.nyu.edu/bitstream/2451/26482/2/02-38.pdf>>. Citado nas páginas 27 e 28.

ENGLE, R. F.; KRONER, K. Multivariate simultaneous generalized ARCH. **Econometric Theory**, v. 11, p. 122–150, 1995. Citado na página 27.

ENGLE, R. F.; NG, V.; ROTHSCHILD, M. Asset pricing with a factor-ARCH covariance structure: empirical estimates for treasury bills. **Journal of Econometrics**, v. 45, p. 213–238, 1990. Citado na página 27.

FIORUCI, J. A.; EHLERS, R. S.; ANDRADE, M. G. Bayesian multivariate GARCH models with dynamic correlations and asymmetric error distributions. **Journal of Applied Statistics**, v. 41, n. 2, p. 320–331, 2014. Citado nas páginas 29, 42 e 73.

GAMERMAN, D.; LOPES, H. F. **Markov Chain Monte Carlo**. 2nd. ed. [S.l.]: Chapman and Hall/CRC, 2006. Citado nas páginas 31 e 32.

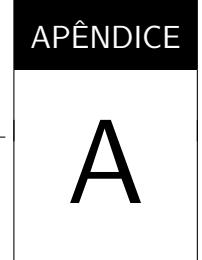
GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. [S.l.]: Taylor & Francis, 2014. Citado na página 34.

- GEYER, C. Practical markov chain monte carlo. **Statistical Science**, n. 7, p. 473 – 511, 1992. Citado na página 58.
- GIROLAMI, M.; CALDERHEAD, B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 73, p. 123 – 214, 2011. Citado nas páginas 37 e 38.
- GOURIEROUX, C. **ARCH Models and Financial Applications**. Springer Series in Statistics, 1997. Disponível em: <<http://link.springer.com/book/10.1007/978-1-4612-1860-9>>. Citado na página 26.
- HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, v. 57, p. 97–109, 1970. Citado na página 33.
- HOMAN, M. D.; GELMAN, A. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. **The Journal of Machine Learning Research**, v. 15, n. 1, p. 1593–1623, 2014. Disponível em: <<http://dl.acm.org/citation.cfm?id=2627435.2638586>>. Citado na página 40.
- JONDEAU, E.; ROCKINGER, M. The copula-GARCH model of conditional dependencies: an international stock-market application. **Journal of International Money and Finance**, 2001. Citado na página 27.
- KASS, R.; CARLIN, B.; GELMAN, A.; NEAL, R. Markov chain monte carlo in practice: A roundtable discussion. **The American Statistician**, n. 52, p. 93 – 100, 1998. Citado na página 58.
- KRONER, F.; NG, V. Modelling asymmetric comovements of asset returns. **The Review of Financial Studies**, v. 11, p. 817–844, 1998. Citado na página 27.
- LYNCH, S. M. **Introduction to Applied Bayesian Statistics and Estimation for Social Scientists**. [S.l.]: Springer-Verlag New York, 2007. Citado na página 33.
- NEAL, R. **Bayesian Learning of Neural Networks**. Tese (Doutorado) — University of Toronto, 1995. Citado na página 34.
- NEAL, R. MCMC using hamiltonian dynamics. **Handbook of Markov Chain Monte Carlo**, 2011. Citado nas páginas 34 e 35.
- PIPIEŃ, M. Bayesian comparison of GARCH processes with skewness mechanism in conditional distributions. **Acta Physica Polonica**, v. 37, p. 3105–3121, 2006. Disponível em: <<http://arxiv.org/pdf/physics/0606253v4.pdf>>. Citado na página 26.
- ROBERT, C.; CASELLA, G. **Monte Carlo Statistical Methods**. [S.l.]: Springer Text in Statistics, 2004. Citado nas páginas 31 e 32.
- ROBERT, C. P.; CASELLA, G. **Introducing Monte Carlo Methods with R**. [S.l.: s.n.], 2010. Citado na página 41.
- Stan Development Team. **RStan: the R interface to Stan**. [S.l.], 2018. R package version 2.18.2. Disponível em: <<http://mc-stan.org/>>. Citado na página 40.
- \_\_\_\_\_. **Stan Modeling Language Users Guide and Reference Manual**. [S.l.], 2018. Versão 2.18.0. Citado na página 39.

TAYLOR, S. Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961-75. **Time Series Analysis: Theory and Practice**, v. 1, p. 203–226, 1982. Citado na página 23.

TSE, Y. K.; TSUI, A. K. C. A multivariate GARCH model with time-varying correlations. **Journal of Business and Economic Statistics**, v. 20, p. 351–362, 2002. Disponível em: <[http://staff.mysmu.edu/yktse/VC\\_R2.pdf](http://staff.mysmu.edu/yktse/VC_R2.pdf)>. Citado na página 28.

VIRBICKAITE, A.; AUSÍN, M. C.; GALEANO, P. Bayesian inference methods for univariate and multivariate garch models: a survey. **Journal of Economic Surveys**, v. 29, p. 76–96, 2015. Citado na página 23.



# DERIVADAS DA DISTRIBUIÇÃO A POSTERIORI PARA O MODELO DCC-GARCH

---

## A.1 Derivadas DCC-GARCH

O DCC-GARCH é dado por

$$\mathbf{y}_t = \mathbf{H}_t^{1/2} \boldsymbol{\epsilon}_t$$

em que  $\mathbf{H}_t$  é a matriz de variâncias condicionais e  $\mathbf{H}_t^{1/2}$  é uma matriz  $m \times m$  positiva definida.

No DCC-GARCH tem-se que

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t$$

com

$$\mathbf{D}_t = \text{diag}(h_{11,t}^{1/2}, \dots, h_{mm,t}^{1/2}),$$

em que  $h_{ii,t} = w_{1,i} + \alpha_i y_{i,t-1}^2 + \beta_i h_{ii,t-1}$ ,  $w_i > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_i \geq 0$ ,  $\alpha_i + \beta_i < 1$  para  $i = 1, \dots, m$ ;

$$\begin{aligned}\mathbf{R}_t &= \text{diag}(\mathbf{Q}_t)^{-1/2} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2} \\ \mathbf{Q}_t &= (1 - a - b) \mathbf{R} + a \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1} + b \mathbf{Q}_{t-1}\end{aligned}$$

em que  $\boldsymbol{\mu}_t = \mathbf{D}_t^{-1} \mathbf{y}_t$ ,  $\mathbf{R}$  é matriz de covariâncias incondicionais de  $\boldsymbol{\mu}_t$  e  $a > 0$ ,  $b > 0$  and  $a + b < 1$ .

Considerando as transformações dadas na [Seção 4.1](#), a distribuição a posteriori de  $\boldsymbol{\theta}$  será

$$\pi(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) |J(\boldsymbol{\phi})|$$

cujas derivadas da log-posteriori serão

$$\frac{\partial \log(\pi(\boldsymbol{\theta} | \mathbf{y}))}{\partial \theta_i} \propto \frac{\partial \ell(\boldsymbol{\phi})}{\partial \theta_i} + \frac{\partial \log(p(\boldsymbol{\theta}))}{\partial \theta_i} + \frac{\partial \log|J(\boldsymbol{\phi})|}{\partial \phi_i}$$

com

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} = \sum_{t=2}^n -\frac{1}{2} Tr \left( \mathbf{H}_t^{-1} \frac{\partial \mathbf{H}_t}{\partial \theta_i} \right) - \frac{1}{2} \mathbf{y}'_t \left( -\mathbf{H}_t^{-1} \frac{\partial \mathbf{H}_t}{\partial \theta_i} \mathbf{H}_t^{-1} \right) \mathbf{y}_t.$$

A seguir apresenta-se alguns resultados para o cálculo das derivadas de  $\mathbf{H}_t$  para os parâmetros do modelo DCC-GARCH.

- $\frac{\partial \mathbf{H}_t}{\partial \theta_i} = \frac{\partial \mathbf{D}_t}{\partial \theta_i} \mathbf{R}_t \mathbf{D}_t + \mathbf{D}_t \frac{\partial \mathbf{R}_t}{\partial \theta_i} \mathbf{D}_t + \mathbf{D}_t \mathbf{R}_t \frac{\partial \mathbf{D}_t}{\partial \theta_i}$
- $\frac{\partial \mathbf{D}_t}{\partial \theta_i} = \begin{cases} \frac{1}{2\sqrt{h_{jj,t}}} \frac{\partial h_{jl,t}}{\partial \theta_i} & \text{se } j = l \text{ com } l = 1, 2, \dots, m \text{ e } i = 1, 2, \dots, 3m+2 \\ 0, c.c. & \end{cases}$
- $\frac{\partial \mathbf{R}_t}{\partial \theta_i} = \frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{-1/2}}{\partial \theta_i} \boldsymbol{\Omega}_t \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} + \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} \frac{\partial \boldsymbol{\Omega}_t}{\partial \theta_i} \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} + \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} \boldsymbol{\Omega}_t \frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{-1/2}}{\partial \theta_i}$
- $\frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{-1/2}}{\partial \theta_i} = \frac{\partial \text{diag}((\boldsymbol{\Omega}_t)^{1/2})^{-1}}{\partial \theta_i} = -\text{diag}(\boldsymbol{\Omega}_t)^{-1/2} \frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{1/2}}{\partial \theta_i} \text{diag}(\boldsymbol{\Omega}_t)^{-1/2}$
- $\frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{1/2}}{\partial \theta_i} = \frac{1}{2} \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} \frac{\partial \text{diag}(\boldsymbol{\Omega}_t)}{\partial \theta_i}$
- $\frac{\partial \boldsymbol{\Omega}_t}{\partial \theta_i} = \begin{cases} a \frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \theta_i} + b \frac{\partial \boldsymbol{\Omega}_{t-1}}{\partial \theta_i}, & \text{se } \theta_i \in (\omega_1, \alpha_1, \beta_1, \dots, \beta_m) \\ -\mathbf{R} + \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1} + b \frac{\partial \boldsymbol{\Omega}_{t-1}}{\partial \theta_i}, & \text{se } \theta_i = a \\ -\mathbf{R} + \boldsymbol{\Omega}_{t-1} + b \frac{\partial \boldsymbol{\Omega}_{t-1}}{\partial \theta_i}, & \text{se } \theta_i = b \end{cases}$
- $\frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \theta_i} = \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \theta_i} \mathbf{y}_{t-1} \boldsymbol{\mu}'_{t-1} + \boldsymbol{\mu}_{t-1} \mathbf{y}'_{t-1} \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \theta_i}$
- $\frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \theta_i} = -\mathbf{D}_{t-1}^{-1} \frac{\partial \mathbf{D}_{t-1}}{\partial \theta_i} \mathbf{D}_{t-1}^{-1}$

## A.2 Derivadas caso com duas séries temporais

Considerando o caso da Subseção 4.1.1 segue as derivadas para os parâmetros do modelo. Primeiro considere  $\omega_1$ ,

- $\frac{\partial \mathbf{H}_t}{\partial \omega_1} = \frac{\partial \mathbf{D}_t}{\partial \omega_1} \mathbf{R}_t \mathbf{D}_t + \mathbf{D}_t \frac{\partial \mathbf{R}_t}{\partial \omega_1} \mathbf{D}_t + \mathbf{D}_t \mathbf{R}_t \frac{\partial \mathbf{D}_t}{\partial \omega_1}$
- $\frac{\partial \mathbf{D}_t}{\partial \omega_1} = \frac{1}{2\sqrt{h_{11,t}}} \left( 1 + \beta \frac{\partial h_{11,t-1}}{\partial \omega_1} \right)$
- $\frac{\partial \mathbf{R}_t}{\partial \omega_1} = \frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{-1/2}}{\partial \omega_1} \boldsymbol{\Omega}_t \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} + \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} \frac{\partial \boldsymbol{\Omega}_t}{\partial \omega_1} \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} + \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} \boldsymbol{\Omega}_t \frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{-1/2}}{\partial \omega_1}$
- $\frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{-1/2}}{\partial \omega_1} = \frac{\partial \text{diag}((\boldsymbol{\Omega}_t)^{1/2})^{-1}}{\partial \omega_1} = -\text{diag}(\boldsymbol{\Omega}_t)^{-1/2} \frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{1/2}}{\partial \omega_1} \text{diag}(\boldsymbol{\Omega}_t)^{-1/2}$
- $\frac{\partial \text{diag}(\boldsymbol{\Omega}_t)^{1/2}}{\partial \omega_1} = \frac{1}{2} \text{diag}(\boldsymbol{\Omega}_t)^{-1/2} \frac{\partial \text{diag}(\boldsymbol{\Omega}_t)}{\partial \omega_1}$

- $\frac{\partial \mathbf{Q}_t}{\partial \omega_1} = a \frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \omega_1} + b \frac{\partial \mathbf{Q}_{t-1}}{\partial \omega_1}$
- $\frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \omega_1} = \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \omega_1} \mathbf{y}_{t-1} \boldsymbol{\mu}'_{t-1} + \boldsymbol{\mu}_{t-1} \mathbf{y}'_{t-1} \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \omega_1}$
- $\frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \omega_1} = -\mathbf{D}_{t-1}^{-1} \frac{\partial \mathbf{D}_{t-1}}{\partial \omega_1} \mathbf{D}_{t-1}^{-1}$

Considerando o parâmetro  $\alpha_1$  temos que

- $\frac{\partial \mathbf{H}_t}{\partial \alpha_1} = \frac{\partial \mathbf{D}_t}{\partial \alpha_1} \mathbf{R}_t \mathbf{D}_t + \mathbf{D}_t \frac{\partial \mathbf{R}_t}{\partial \alpha_1} \mathbf{D}_t + \mathbf{D}_t \mathbf{R}_t \frac{\partial \mathbf{D}_t}{\partial \alpha_1}$
- $\frac{\partial \mathbf{D}_t}{\partial \alpha_1} = \frac{1}{2\sqrt{h_{11,t}}} \left( y_{1,t-1}^2 + \beta_1 \frac{\partial h_{11,t-1}}{\partial \alpha_1} \right)$
- $\frac{\partial \mathbf{R}_t}{\partial \alpha_1} = \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \alpha_1} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \mathbf{Q}_t}{\partial \alpha_1} \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \mathbf{Q}_t \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \alpha_1}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \alpha_1} = \frac{\partial \text{diag}((\mathbf{Q}_t)^{1/2})^{-1}}{\partial \alpha_1} = -\text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \alpha_1} \text{diag}(\mathbf{Q}_t)^{-1/2}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \alpha_1} = \frac{1}{2} \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)}{\partial \alpha_1}$
- $\frac{\partial \mathbf{Q}_t}{\partial \alpha_1} = a \frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \alpha_1} + b \frac{\partial \mathbf{Q}_{t-1}}{\partial \alpha_1}$
- $\frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \alpha_1} = \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \alpha_1} \mathbf{y}_{t-1} \boldsymbol{\mu}'_{t-1} + \boldsymbol{\mu}_{t-1} \mathbf{y}'_{t-1} \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \alpha_1}$
- $\frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \alpha_1} = -\mathbf{D}_{t-1}^{-1} \frac{\partial \mathbf{D}_{t-1}}{\partial \alpha_1} \mathbf{D}_{t-1}^{-1}$

Para  $\beta_1$  segue que

- $\frac{\partial \mathbf{H}_t}{\partial \beta_1} = \frac{\partial \mathbf{D}_t}{\partial \beta_1} \mathbf{R}_t \mathbf{D}_t + \mathbf{D}_t \frac{\partial \mathbf{R}_t}{\partial \beta_1} \mathbf{D}_t + \mathbf{D}_t \mathbf{R}_t \frac{\partial \mathbf{D}_t}{\partial \beta_1}$
- $\frac{\partial \mathbf{D}_t}{\partial \beta_1} = \frac{1}{2\sqrt{h_{11,t}}} \left( h_{11,t-1} + \beta_1 \frac{\partial h_{11,t-1}}{\partial \beta_1} \right)$
- $\frac{\partial \mathbf{R}_t}{\partial \beta_1} = \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \beta_1} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \mathbf{Q}_t}{\partial \beta_1} \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \mathbf{Q}_t \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \beta_1}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \beta_1} = \frac{\partial \text{diag}((\mathbf{Q}_t)^{1/2})^{-1}}{\partial \beta_1} = -\text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \beta_1} \text{diag}(\mathbf{Q}_t)^{-1/2}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \beta_1} = \frac{1}{2} \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)}{\partial \beta_1}$
- $\frac{\partial \mathbf{Q}_t}{\partial \beta_1} = a \frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \beta_1} + b \frac{\partial \mathbf{Q}_{t-1}}{\partial \beta_1}$
- $\frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \beta_1} = \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \beta_1} \mathbf{y}_{t-1} \boldsymbol{\mu}'_{t-1} + \boldsymbol{\mu}_{t-1} \mathbf{y}'_{t-1} \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \beta_1}$
- $\frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \beta_1} = -\mathbf{D}_{t-1}^{-1} \frac{\partial \mathbf{D}_{t-1}}{\partial \beta_1} \mathbf{D}_{t-1}^{-1}$

Considere agora  $\omega_2$ ,

- $\frac{\partial \mathbf{H}_t}{\partial \omega_2} = \frac{\partial \mathbf{D}_t}{\partial \omega_2} \mathbf{R}_t \mathbf{D}_t + \mathbf{D}_t \frac{\partial \mathbf{R}_t}{\partial \omega_2} \mathbf{D}_t + \mathbf{D}_t \mathbf{R}_t \frac{\partial \mathbf{D}_t}{\partial \omega_2}$
- $\frac{\partial \mathbf{D}_t}{\partial \omega_2} = \frac{1}{2\sqrt{h_{22,t}}} \left( 1 + \beta_2 \frac{\partial h_{22,t-1}}{\partial \omega_2} \right)$
- $\frac{\partial \mathbf{R}_t}{\partial \omega_2} = \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \omega_2} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \mathbf{Q}_t}{\partial \omega_2} \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \mathbf{Q}_t \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \omega_2}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \omega_2} = \frac{\partial \text{diag}((\mathbf{Q}_t)^{1/2})^{-1}}{\partial \omega_2} = -\text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \omega_2} \text{diag}(\mathbf{Q}_t)^{-1/2}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \omega_2} = \frac{1}{2} \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)}{\partial \omega_2}$
- $\frac{\partial \mathbf{Q}_t}{\partial \omega_2} = a \frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \omega_2} + b \frac{\partial \mathbf{Q}_{t-1}}{\partial \omega_2}$
- $\frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \omega_2} = \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \omega_2} \mathbf{y}_{t-1} \boldsymbol{\mu}'_{t-1} + \boldsymbol{\mu}_{t-1} \mathbf{y}'_{t-1} \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \omega_2}$
- $\frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \omega_2} = -\mathbf{D}_{t-1}^{-1} \frac{\partial \mathbf{D}_{t-1}}{\partial \omega_2} \mathbf{D}_{t-1}^{-1}$

Considerando o parâmetro  $\alpha_2$ ,

- $\frac{\partial \mathbf{H}_t}{\partial \alpha_2} = \frac{\partial \mathbf{D}_t}{\partial \alpha_2} \mathbf{R}_t \mathbf{D}_t + \mathbf{D}_t \frac{\partial \mathbf{R}_t}{\partial \alpha_2} \mathbf{D}_t + \mathbf{D}_t \mathbf{R}_t \frac{\partial \mathbf{D}_t}{\partial \alpha_2}$
- $\frac{\partial \mathbf{D}_t}{\partial \alpha_2} = \frac{1}{2\sqrt{h_{22,t}}} \left( y_{2,t-1}^2 + \beta \frac{\partial h_{22,t-1}}{\partial \alpha_2} \right)$
- $\frac{\partial \mathbf{R}_t}{\partial \alpha_2} = \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \alpha_2} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \mathbf{Q}_t}{\partial \alpha_2} \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \mathbf{Q}_t \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \alpha_2}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \alpha_2} = \frac{\partial \text{diag}((\mathbf{Q}_t)^{1/2})^{-1}}{\partial \alpha_2} = -\text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \alpha_2} \text{diag}(\mathbf{Q}_t)^{-1/2}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \alpha_2} = \frac{1}{2} \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)}{\partial \alpha_2}$
- $\frac{\partial \mathbf{Q}_t}{\partial \alpha_2} = a \frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \alpha_2} + b \frac{\partial \mathbf{Q}_{t-1}}{\partial \alpha_2}$
- $\frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \alpha_2} = \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \alpha_2} \mathbf{y}_{t-1} \boldsymbol{\mu}'_{t-1} + \boldsymbol{\mu}_{t-1} \mathbf{y}'_{t-1} \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \alpha_2}$
- $\frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \alpha_2} = -\mathbf{D}_{t-1}^{-1} \frac{\partial \mathbf{D}_{t-1}}{\partial \alpha_2} \mathbf{D}_{t-1}^{-1}$

Para  $\beta_2$  segue que

- $\frac{\partial \mathbf{H}_t}{\partial \beta_2} = \frac{\partial \mathbf{D}_t}{\partial \beta_2} \mathbf{R}_t \mathbf{D}_t + \mathbf{D}_t \frac{\partial \mathbf{R}_t}{\partial \beta_2} \mathbf{D}_t + \mathbf{D}_t \mathbf{R}_t \frac{\partial \mathbf{D}_t}{\partial \beta_2}$
- $\frac{\partial \mathbf{D}_t}{\partial \beta_2} = \frac{1}{2\sqrt{h_{22,t}}} \left( h_{22,t-1} + \beta_2 \frac{\partial h_{22,t-1}}{\partial \beta_2} \right)$
- $\frac{\partial \mathbf{R}_t}{\partial \beta_2} = \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \beta_2} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \mathbf{Q}_t}{\partial \beta_2} \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \mathbf{Q}_t \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \beta_2}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \beta_2} = \frac{\partial \text{diag}((\mathbf{Q}_t)^{1/2})^{-1}}{\partial \beta_2} = -\text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \beta_2} \text{diag}(\mathbf{Q}_t)^{-1/2}$

- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \beta_2} = \frac{1}{2} \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)}{\partial \beta_2}$
- $\frac{\partial \mathbf{Q}_t}{\partial \beta_2} = a \frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \beta_2} + b \frac{\partial \mathbf{Q}_{t-1}}{\partial \beta_2}$
- $\frac{\partial \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1}}{\partial \beta_2} = \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \beta_2} \mathbf{y}_{t-1} \boldsymbol{\mu}'_{t-1} + \boldsymbol{\mu}_{t-1} \mathbf{y}'_{t-1} \frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \beta_2}$
- $\frac{\partial \mathbf{D}_{t-1}^{-1}}{\partial \beta_2} = -\mathbf{D}_{t-1}^{-1} \frac{\partial \mathbf{D}_{t-1}}{\partial \beta_2} \mathbf{D}_{t-1}^{-1}$

Para  $a$

- $\frac{\partial \mathbf{H}_t}{\partial a} = \mathbf{D}_t \frac{\partial \mathbf{R}_t}{\partial \phi_7} \mathbf{D}_t$
- $\frac{\partial \mathbf{R}_t}{\partial a} = \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial a} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \mathbf{Q}_t}{\partial a} \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \mathbf{Q}_t \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial a}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial \phi_7} = \frac{\partial \text{diag}((\mathbf{Q}_t)^{1/2})^{-1}}{\partial a} = -\text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial a} \text{diag}(\mathbf{Q}_t)^{-1/2}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial \phi_7} = \frac{1}{2} \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)}{\partial \phi_7}$
- $\frac{\partial \mathbf{Q}_t}{\partial \phi_7} = -\mathbf{R} + \boldsymbol{\mu}_{t-1} \boldsymbol{\mu}'_{t-1} + b \frac{\partial \mathbf{Q}_{t-1}}{\partial a}$

Por fim, para  $b$

- $\frac{\partial \mathbf{H}_t}{\partial b} = \mathbf{D}_t \frac{\partial \mathbf{R}_t}{\partial b} \mathbf{D}_t$
- $\frac{\partial \mathbf{R}_t}{\partial b} = \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial b} \mathbf{Q}_t \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \mathbf{Q}_t}{\partial b} \text{diag}(\mathbf{Q}_t)^{-1/2} + \text{diag}(\mathbf{Q}_t)^{-1/2} \mathbf{Q}_t \frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial b}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{-1/2}}{\partial b} = \frac{\partial \text{diag}((\mathbf{Q}_t)^{1/2})^{-1}}{\partial b} = -\text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial b} \text{diag}(\mathbf{Q}_t)^{-1/2}$
- $\frac{\partial \text{diag}(\mathbf{Q}_t)^{1/2}}{\partial b} = \frac{1}{2} \text{diag}(\mathbf{Q}_t)^{-1/2} \frac{\partial \text{diag}(\mathbf{Q}_t)}{\partial b}$
- $\frac{\partial \mathbf{Q}_t}{\partial b} = -\mathbf{R} + \mathbf{Q}_{t-1} + b \frac{\partial \mathbf{Q}_{t-1}}{\partial b}$

