

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

**Inferência Bayesiana em Modelos de Volatilidade Estocástica  
usando Métodos de Monte Carlo Hamiltoniano**

**David de Souza Dias**

Dissertação de Mestrado do Programa Interinstitucional de  
Pós-Graduação em Estatística (PIPGEs)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

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

**David de Souza Dias**

# Inferência Bayesiana em Modelos de Volatilidade Estocástica usando Métodos de Monte Carlo Hamiltoniano

Dissertação 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 Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Ricardo Sandes Ehlers

**USP – São Carlos**  
**Setembro de 2018**



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia  
Programa Interinstitucional de Pós-Graduação em Estatística

---

Folha de Aprovação

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato David de Souza Dias, realizada em 10/08/2018:

---

Prof. Dr. Ricardo Sandes Ehlers  
USP

---

Prof. Dr. Mauricio Enrique Zevallos Herencia  
UNICAMP

---

Profa. Dra. Maria Sílvia de Assis Moura  
UFSCar

**David de Souza Dias**

**Bayesian Inference in Stochastic Volatility Models using  
Hamiltonian Monte Carlo Methods**

Master dissertation submitted to the Institute of  
Mathematics and Computer Sciences – ICMC-USP  
and to the Department of Statistics – DEs-UFSCar, in  
partial fulfillment of the requirements for the degree of  
the Master Interagency Program Graduate in Statistics.  
*FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Ricardo Sandes Ehlers

**USP – São Carlos  
September 2018**



# RESUMO

DIAS, D. S. **Inferência Bayesiana em Modelos de Volatilidade Estocástica usando Métodos de Monte Carlo Hamiltoniano**. 2018. 92 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

Este trabalho apresenta um estudo através da abordagem Bayesiana em modelos de volatilidade estocástica, para modelagem de séries temporais financeiras, com o uso do método de Monte Carlo Hamiltoniano (HMC). Propomos o uso de outras distribuições para os erros da equação de observações dos modelos de volatilidade estocástica, além da distribuição Gaussiana, para tratar problemas como caudas pesadas e assimetria nos retornos. Além disso, utilizamos critérios de informações, recentemente desenvolvidos, WAIC e LOO que aproximam a metodologia de validação cruzada, para realizar a seleção de modelos. No decorrer do trabalho, estudamos a qualidade do método HMC através de exemplos, estudo de simulação e aplicação a conjunto de dados. Adicionalmente, avaliamos a performance dos modelos e métodos propostos através do cálculo de estimativas para o Valor em Risco (VaR) para múltiplos horizontes de tempo.

**Palavras-chave:** Inferência Bayesiana, Modelos de Volatilidade Estocástica, Método HMC, WAIC & LOO, Valor em Risco (VaR).



# ABSTRACT

DIAS, D. S. **Bayesian Inference in Stochastic Volatility Models using Hamiltonian Monte Carlo Methods**. 2018. 92 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2018.

This paper presents a study using Bayesian approach in stochastic volatility models for modeling financial time series, using Hamiltonian Monte Carlo methods (HMC). We propose the use of other distributions for the errors of the equation at stochastic volatility model, besides the Gaussian distribution, to treat the problem as heavy tails and asymmetry in the returns. Moreover, we use recently developed information criteria WAIC and LOO that approximate the cross-validation methodology, to perform the selection of models. Throughout this work, we study the quality of the HMC methods through examples, simulation study and application to dataset. In addition, we evaluated the performance of the proposed models and methods by calculating estimates for Value at Risk (VaR) for multiple time horizons.

**Keywords:** Bayesian Inference, Stochastic Volatility Models, HMC methods, WAIC and LOO, Value at Risk (VaR).



# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Densidade simulada $\pi(\phi)$ com os parâmetros $\alpha = 20$ e $\beta = 1,5$ . . . . .	31
Figura 2 – Curva de nível representando amostras estimadas de $p(\mu, \sigma X)$ , em que uma amostra de tamanho $n = 20$ foi gerada de $\mathcal{N}(X \mu = 0, \sigma = 10)$ . Ambos os métodos MH e HMC foram simulados a partir de pontos iniciais $\mu = 3$ e $\sigma = 40$ para 200 passos (neste caso, HMC com $\varepsilon = 0,3$ e $L = 15$ ). . . . .	40
Figura 3 – Mesmo gráfico da Figura 2, mas a partir de pontos iniciais $\mu = 15$ e $\sigma = 2$ para 200 passos (neste caso, HMC com $\varepsilon = 0,8$ e $L = 10$ ). . . . .	41
Figura 4 – Gráficos da série financeira de preços diários de fechamento do IBOVESPA	42
Figura 5 – Gráficos da série financeira de preços diários de fechamento do IBOVESPA (Retornos). . . . .	43
Figura 6 – Gráficos de convergência dos parâmetros na série de retornos do IBOVESPA.	44
Figura 7 – Gráfico de log-retornos em percentual e volatilidades estimadas para série de retornos do IBOVESPA, usando método HMC. . . . .	45
Figura 8 – Gráficos das taxas de retornos diários em percentagens da série £/USD. . . . .	52
Figura 9 – Gráficos das taxas de retornos diários em percentagens da série EUR/USD. . . . .	52
Figura 10 – Gráficos das taxas de retornos diários em percentagens da série S&P500. . . . .	52
Figura 11 – Correlogramas das séries de retornos £/USD, EUR/USD e S&P500. . . . .	53
Figura 12 – Gráficos de convergência dos parâmetros na série de retornos £/USD com erros Gaussianos. . . . .	56
Figura 13 – Gráficos de convergência dos parâmetros na série de retornos £/USD com erros Skew-Normal. . . . .	57
Figura 14 – Gráficos de convergência dos parâmetros na série de retornos EUR/USD com erros t-Student. . . . .	58
Figura 15 – Gráficos de convergência dos parâmetros na série de retornos EUR/USD com erros GED. . . . .	59
Figura 16 – Gráficos de convergência dos parâmetros na série de retornos IBOVESPA com erros Gaussianos. . . . .	60
Figura 17 – Gráficos de convergência dos parâmetros na série de retornos IBOVESPA com erros Skew-Normal. . . . .	61
Figura 18 – Gráficos de convergência dos parâmetros na série de retornos S&P500 com erros Gaussianos. . . . .	62
Figura 19 – Gráficos de convergência dos parâmetros na série de retornos S&P500 com erros Skew-Normal. . . . .	63

Figura 20 – Retornos absolutos em percentual para a série £/USD e volatilidades estimadas com bandas de credibilidade de 95% com erros Gaussianos. . . . .	64
Figura 21 – Retornos absolutos em percentual para a série £/USD e volatilidades estimadas com bandas de credibilidade de 95% com erros Skew-Normal. . . . .	64
Figura 22 – Retornos absolutos em percentual para a série EUR/USD e volatilidades estimadas com bandas de credibilidade de 95% com erros t-Student. . . . .	65
Figura 23 – Retornos absolutos em percentual para a série EUR/USD e volatilidades estimadas com bandas de credibilidade de 95% com erros GED. . . . .	65
Figura 24 – Retornos absolutos em percentual para a série do IBOVESPA e volatilidades estimadas com bandas de credibilidade de 95% com erros Gaussianos. . . . .	66
Figura 25 – Retornos absolutos em percentual para a série do IBOVESPA e volatilidades estimadas com bandas de credibilidade de 95% com erros Skew-Normal. . . . .	66
Figura 26 – Retornos absolutos em percentual para a série do S&P500 e volatilidades estimadas com bandas de credibilidade de 95% com erros Gaussianos. . . . .	67
Figura 27 – Retornos absolutos em percentual para a série do S&P500 e volatilidades estimadas com bandas de credibilidade de 95% com erros Skew-Normal. . . . .	67
Figura 28 – Valor em Risco (VaR) com erros Gaussianos e Skew-Normais, para a série de retornos £/USD. . . . .	70
Figura 29 – Valor em Risco (VaR) com erros t-Student e GED, para a série de retornos EUR/USD. . . . .	71
Figura 30 – Valor em Risco (VaR) com erros Gaussianos e Skew-Normais, para a série de retornos IBOVESPA. . . . .	72
Figura 31 – Valor em Risco (VaR) com erros Gaussianos e Skew-Normais, para a série de retornos S&P500. . . . .	73
Figura 32 – Medida $\hat{R}$ para as log-volatilidades da série do IBOVESPA. . . . .	83
Figura 33 – Medida $\hat{R}$ para as log-volatilidades da série do S&P500. . . . .	84
Figura 34 – Gráfico de convergência para uma amostra de log-volatilidades <i>a posteriori</i> estimadas para a série do IBOVESPA, com erros Gaussianos. . . . .	85
Figura 35 – Gráfico de convergência para uma amostra de log-volatilidades <i>a posteriori</i> estimadas para a série do IBOVESPA, com erros Skew-Normais. . . . .	86
Figura 36 – Gráfico de convergência para uma amostra de log-volatilidades <i>a posteriori</i> estimadas para a série do S&P500, com erros Gaussianos. . . . .	87
Figura 37 – Gráfico de convergência para uma amostra de log-volatilidades <i>a posteriori</i> estimadas para a série do S&P500, com erros Skew-Normais. . . . .	88
Figura 38 – Gráficos de convergência dos parâmetros na série de retornos S&P500 com erros Gaussianos, considerando 50mil simulações. . . . .	90
Figura 39 – Gráficos de convergência dos parâmetros na série de retornos S&P500 com erros Skew-Normais, considerando 50mil simulações. . . . .	90

Figura 40 – Medida $\hat{R}$ para as log-volatilidades da série do S&P500, considerando 50mil simulações. . . . .	91
Figura 41 – Retornos absolutos em percentual para a série £/USD e volatilidades estimadas com bandas de credibilidade de 95%. . . . .	91



# LISTA DE TABELAS

---

---

Tabela 1	– Estatísticas descritivas da série de retornos do IBOVESPA . . . . .	42
Tabela 2	– Viés e raiz quadrada do erro quadrático médio ( <i>smse</i> ) das médias a posteriori. Parâmetros: $\mu = -9$ , $\phi = 0,99$ e $\sigma_\eta = 0,15$ . . . . .	48
Tabela 3	– Viés e raiz quadrada do erro quadrático médio ( <i>smse</i> ) das médias a posteriori. Parâmetros: $\mu = -9$ , $\phi = 0,99$ e $\sigma_\eta = 0,05$ . . . . .	48
Tabela 4	– Viés e raiz quadrada do erro quadrático médio ( <i>smse</i> ) das médias a posteriori. Parâmetros: $\mu = -9$ , $\phi = 0,95$ e $\sigma_\eta = 0,15$ . . . . .	48
Tabela 5	– Viés e raiz quadrada do erro quadrático médio ( <i>smse</i> ) das médias a posteriori. Parâmetros: $\mu = -9$ , $\phi = 0,95$ e $\sigma_\eta = 0,05$ . . . . .	49
Tabela 6	– Viés e <i>smse</i> das médias a posteriori, método HMC. Parâmetros: $\mu = -10$ , $\nu = 5$ (t-Student), $\nu = 0,8$ (Skew-Normal), $\nu = 1,5$ (GED), $\phi = 0,99$ e $\sigma_\eta = 0,05$ . . . . .	50
Tabela 7	– Viés e <i>smse</i> das médias a posteriori, método HMC. Parâmetros: $\mu = -10$ , $\nu = 5$ (t-Student), $\nu = 0,8$ (Skew-Normal), $\nu = 1,5$ (GED), $\phi = 0,99$ e $\sigma_\eta = 0,15$ . . . . .	50
Tabela 8	– Viés e <i>smse</i> das médias a posteriori, método HMC. Parâmetros: $\mu = -10$ , $\nu = 5$ (t-Student), $\nu = 0,8$ (Skew-Normal), $\nu = 1,5$ (GED), $\phi = 0,95$ e $\sigma_\eta = 0,05$ . . . . .	50
Tabela 9	– Viés e <i>smse</i> das médias a posteriori, método HMC. Parâmetros: $\mu = -10$ , $\nu = 5$ (t-Student), $\nu = 0,8$ (Skew-Normal), $\nu = 1,5$ (GED), $\phi = 0,95$ e $\sigma_\eta = 0,15$ . . . . .	50
Tabela 10	– Tempos médios de simulação para cada uma das distribuições. . . . .	51
Tabela 11	– Sumário das séries em estudo . . . . .	51
Tabela 12	– Comparação entre os modelos propostos via critérios de seleção. . . . .	54
Tabela 13	– Estimativas das médias (desvios padrões) dos parâmetros a <i>posteriori</i> . . . . .	55
Tabela 14	– Comparação entre os modelos propostos via critérios de seleção para estudo de sensibilidade, série £/USD. . . . .	68
Tabela 15	– Comparação entre os modelos propostos via critérios de seleção para estudo de sensibilidade, série S&P500. . . . .	69
Tabela 16	– Comparação entre os modelos propostos via critérios de seleção para série do S&P500. . . . .	89
Tabela 17	– Estimativas das médias (desvios padrões) dos parâmetros a <i>posteriori</i> . . . . .	89



# SUMÁRIO

---

---

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
1.1	Organização da Monografia	18
1.2	Séries Financeiras	19
1.2.1	<i>Retornos</i>	19
1.3	Abordagem Bayesiana	20
1.3.1	<i>Método de Monte Carlo via Cadeias de Markov</i>	21
1.3.1.1	<i>Cadeias de Markov</i>	21
1.3.1.2	<i>Algoritmo de Metropolis-Hastings</i>	22
1.3.1.3	<i>Algoritmo Gibbs Sampling</i>	23
<b>2</b>	<b>MODELOS PARA VOLATILIDADE</b>	<b>25</b>
2.1	Modelos de Volatilidade	25
2.1.1	<i>Modelos ARCH</i>	25
2.1.1.1	<i>Identificação</i>	26
2.1.2	<i>Modelos GARCH</i>	27
2.1.2.1	<i>Identificação</i>	27
2.2	Modelos de Volatilidade Estocástica	27
2.2.1	<i>Funções de Verossimilhança</i>	29
2.2.2	<i>Distribuições a Priori</i>	30
2.3	Seleção de Modelos	32
2.3.1	<i>Deviance Information Criterion</i>	32
2.3.2	<i>Watanabe Information Criterion</i>	33
2.3.3	<i>Approximate Leave-One-Out Cross-Validation</i>	33
2.4	Valor em Risco (VaR)	35
<b>3</b>	<b>MONTE CARLO HAMILTONIANO</b>	<b>37</b>
3.1	Dinâmica Hamiltoniana	37
3.1.1	<i>Algoritmo</i>	39
3.1.2	<i>NUTS e Stan</i>	41
<b>4</b>	<b>ESTUDO DE SIMULAÇÃO E APLICAÇÕES</b>	<b>47</b>
4.1	Estudo de Simulação 1	47
4.2	Estudo de Simulação 2	49
4.3	Aplicação a Conjuntos de Dados	51

4.3.1	<i>Descrição dos Dados e Análise Descritiva</i> . . . . .	51
4.3.2	<i>Resultados</i> . . . . .	54
4.4	Sensibilidade à escolha da distribuição a priori . . . . .	68
4.5	Valor em Risco (VaR) . . . . .	70
5	<b>CONSIDERAÇÕES FINAIS</b> . . . . .	75
5.1	Propostas Futuras . . . . .	77
5.1.1	<i>Kullback–Leibler Divergence</i> . . . . .	77
5.1.2	<i>Reversible Jump</i> . . . . .	77
5.1.3	<i>Outras distribuições e abordagens</i> . . . . .	77
5.1.4	<i>Método RMHMC</i> . . . . .	78
	<b>REFERÊNCIAS</b> . . . . .	79
<b>APÊNDICE A</b>	<b>ESTUDOS ADICIONAIS E CÓDIGOS COMPUTACIONAIS</b> . . . . .	<b>83</b>
A.1	Avaliação de Convergência das volatilidades . . . . .	83
A.2	Análise da série S&P500 . . . . .	89
A.3	Códigos Computacionais . . . . .	92

---

# INTRODUÇÃO

---

A análise de séries temporais é área da estatística dedicada ao estudo de fenômenos caracterizados por uma evolução temporal. Atualmente existe uma grande demanda no estudo deste tipo de análise, pois sua aplicabilidade pode estar presente nas mais diversas áreas. Em epidemiologia pode-se estar interessado em modelar o número de indivíduos que sofrem de uma determinada doença. Em controle da qualidade, o interesse pode estar na modelagem do número de peças defeituosas em uma linha de produção. No mercado financeiro é necessário quantificar riscos de perdas ou ganhos, pois o aporte de grandes somas em investimentos causa impactos diretos na economia de um país.

A utilização de um modelo capaz de capturar todas as situações oriundas de um estudo é de fundamental importância para a obtenção de resultados bons e confiáveis. Diante disso, diversos modelos tem sido propostos na literatura para tentar descrever a dinâmica desses tipos de dados ((MORETTIN; TOLOI, 2006), (BROCKWELL; DAVIS, 2013), (BOX *et al.*, 2015)). As características dessas abordagens consistem no fato de que os modelos são capazes de incorporar as incertezas associadas a estrutura temporal e adaptar-se aos diferentes comportamentos da série. Em muitas aplicações assume-se que a variância dos erros é constante ao longo do tempo, no entanto, em diversos casos as séries temporais em estudo exibem períodos de grande volatilidade seguidos de períodos de estabilidade.

Neste caso, sua variância constante pode estar mudando ao longo do tempo e a suposição de variância constante (homoscedasticidade) pode não ser apropriada. [...] Além disso, pode-se estar interessado em prever a variância condicional da série, por exemplo, no mercado de ações o interesse é não apenas prever a taxa de retorno mas também a sua variância ao longo de um certo período (EHLERS, 2007).

Neste sentido, foram desenvolvidos modelos estatísticos apropriados para séries financeiras. Dentre os mais conhecidos estão os modelos ARCH (*autoregressive conditional*

*heteroscedasticity*) e GARCH (*generalized ARCH*) introduzidos, respectivamente, por Engle (1982) e Bollerslev (1986). Existem outras extensões, como por exemplo, os modelos EGARCH (*exponential GARCH*) (NELSON, 1991) e TARARCH (*threshold ARCH*) (ZAKOIAN, 1994). Estes modelos propõem modelar a variância condicional considerando que esta depende de retornos passados por meio de uma função quadrática, todavia, pode ser razoável assumir que esta variância condicional varia estocasticamente ao longo do tempo. Assim, uma alternativa a classe GARCH são os modelos de Volatilidade Estocástica (VE), apresentados por Taylor (1986), que consistem em assumir um processo estocástico para modelagem da volatilidade. Ambos os modelos tem sido amplamente utilizados no meio econômico, pois auxiliam direta ou indiretamente no mercado financeiro como um dos principais critérios utilizados para análise de riscos e para potencializar o portfólio de investimentos ((HESTON, 1993), (ENGLE, 2001), (BARNDORFF-NIELSEN; SHEPHARD, 2002)).

Os modelos VE são mais flexíveis em relação aos modelos ARCH e GARCH, pois possuem dois processos para os ruídos, sendo um para as observações e outro para as volatilidades. Na literatura é possível encontrar diferentes estudos comparativos entre as referidas classes de modelos (e.g., Taylor (1994), Kim, Shephard e Chib (1998)). Devido a complexidade da função de verossimilhança, na estimação desses modelos, diversos métodos tem sido propostos na literatura, por exemplo, método de quasi-verossimilhança (RUIZ, 1994), método generalizado dos momentos, Monte Carlo via Cadeia de Markov (MCMC) ((JACQUIER; POLSON; ROSSI, 2002), (CHIB; NARDARI; SHEPHARD, 2002)), aproximação de Laplace Aninhadas integradas, entre outros. Especificamente, o método MCMC é considerado um dos mais eficientes (ZEVALLOS; GASCO; EHLERS, 2016).

Propomos neste trabalho a utilização dos modelos de volatilidade estocástica na análise de séries temporais financeiras utilizando a abordagem Bayesiana. Portanto, sendo necessário especificar a distribuição *a priori* para os parâmetros do modelo. Neste caso, toda inferência será baseada na distribuição *a posteriori*. Conseqüentemente para obter as distribuições *a posteriori* será proposto o uso do método iterativo de Monte Carlo Hamiltoniano (HMC) introduzido por Neal *et al.* (2011).

## 1.1 Organização da Monografia

A organização deste texto está da seguinte forma:

**Capítulo 1** Nas próximas seções são apresentadas algumas definições preliminares para maior compreensão do tema abordado;

**Capítulo 2** Exibe os principais conceitos e ferramentas a serem utilizadas;

**Capítulo 3** Apresenta o método de Monte Carlo Hamiltoniano, abordando alguns exemplos;

**Capítulo 4** É realizado o estudo de Simulação e uma aplicação a séries de retornos financeiros.

## 1.2 Séries Financeiras

Uma característica presente em séries de ativos financeiros é que estas apresentam variância condicional (volatilidade) com evolução temporal. Para levar em conta a presença de grupos de volatilidade em séries financeiras é necessário recorrer a modelos heteroscedásticos condicionais. Esses modelos são não lineares no que se refere à variância (MORETTIN, 2008). Neste caso, estamos interessados em modelar a volatilidade de um retorno, que está definido a seguir.

### 1.2.1 Retornos

Em finanças, geralmente, deseja-se avaliar os riscos de uma carteira de ativos financeiros. Estes riscos podem ser medidos considerando as variações de preços dos ativos.

Seja  $P_t$  o preço de um ativo no instante  $t$ . Então, define-se:

- i. A variação de preços,  $\Delta P_t = P_t - P_{t-1}$ ;
- ii. A variação relativa de preços ou retorno líquido simples,  $R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{\Delta P_t}{P_{t-1}}$ ;
- iii. Retorno bruto simples ou taxa de retorno,  $R_t + 1 = \frac{P_t}{P_{t-1}}$ ;
- iv. Retorno composto continuamente ou log-retorno,  $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(R_t + 1)$ .

A definição iv. é usualmente utilizada e, quase sempre,  $r_t$  é chamado simplesmente de retorno.

Os retornos financeiros possuem características peculiares, que outras séries não apresentam. Raramente retornos financeiros exibem tendência ou sazonalidade, com exceção, eventualmente, em retornos intradiários (veja Morettin (2008) para maiores detalhes).

De acordo com Morettin (2008), as principais características empíricas relacionadas à séries de retornos financeiros são:

- (a) São comumente não autocorrelacionados;
- (b) Possuem média amostral próxima de zero;
- (c) Os quadrados dos retornos são autocorrelacionados, apresentando uma correlação *lag* inicialmente pequena com um decaimento lento das demais;
- (d) Exibem agrupamento (*clusters*) de volatilidades ao longo do tempo;
- (e) A distribuição incondicional dos retornos frequentemente aparenta ser simétrica, porém é leptocúrtica;

- (f) Em alguns casos são não-lineares (reagem de maneira diferente a choques grandes ou pequenos, ou ainda, positivos ou negativos).

Considerando uma série de retornos  $\{r_t\}$ , define-se:

$$\mu_t = E(r_t | \mathcal{F}_{t-1}) \quad \text{e} \quad \sigma_t^2 = \text{Var}(r_t | \mathcal{F}_{t-1}), \quad (1.1)$$

a média e a variância condicional de  $r_t$ , em que  $\mathcal{F}_{t-1} = \{r_{t-1}, \dots, r_1\}$  denota o conjunto de informações até o instante  $t - 1$ . Em algumas situações é razoável supor que  $\mu_t = 0$ , de modo que  $\sigma_t^2 = E(r_t^2 | \mathcal{F}_{t-1})$ .

### 1.3 Abordagem Bayesiana

Ao trabalhar com análises estatísticas há interesse em quantificar a cerca de diferentes graus de incerteza sobre proposições através de dados observados. Uma maneira formal para analisar dados é através da inferência científica ou estatística. Neste sentido, os diferentes graus de incerteza podem ser representados através de modelos probabilísticos. Em outras palavras, se  $A$  indica a proposição “choverá amanhã”, então  $P(A)$  indica o quão se acredita nesta afirmação. Além disso, é natural que diferentes pesquisadores possam ter distintos graus de incerteza.

Formalmente é comum assumir que o modelo probabilístico gerador dos dados pode ser explicado por várias hipóteses. Neste caso, a informação disponível sobre determinada hipótese ou característica  $\theta$  está resumida probabilisticamente a *priori* através de  $\pi(\theta)$ . O interesse está, então, em utilizar observações de uma quantidade aleatória  $X$  para aprender sobre  $\theta$ . O teorema de Bayes permite tratar todas as características de interesse desconhecidas como variáveis aleatórias, possibilitando que informações externas sejam incorporadas a análise dos dados. Assim, a atualização, a partir dos dados, é determinada por

$$\pi(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{p(x)} = \frac{p(x|\theta)\pi(\theta)}{\int_{\Theta} p(x|\theta)\pi(\theta)d\theta} \quad (1.2)$$

Note que  $1/p(x)$  não depende de  $\theta$  e funciona como uma constante normalizadora de  $\pi(\theta|x)$ .  $p(x)$  pode ser colocada como probabilidade marginal dos dados ou distribuição preditiva, ou seja, esta é a distribuição esperada para a observação  $x$  dado  $\theta$ .

Para um valor fixo  $x$  a função  $p(x|\theta)$  fornece a plausibilidade ou verossimilhança de cada um dos possíveis valores de  $\theta$ . Combinando as fontes de informação a *priori* e a verossimilhança, é possível obter uma forma mais usual do teorema de Bayes que leva à distribuição a *posteriori*. Desta forma,

$$\pi(\theta|x) \propto p(x|\theta)\pi(\theta), \quad (1.3)$$

esta forma simplificada é útil em problemas que envolvem estimação de parâmetros, já que o denominador é apenas uma constante normalizadora. No entanto, em algumas situações, como seleção e comparação de modelos, este termo possui um papel importante.

Inferências sobre funções dos parâmetros são obtidas através dos cálculos de esperanças *a posteriori*, deste modo, é necessário determinar integrais do tipo,

$$E_{\theta|x}[g(\theta)] = \int_{\Theta} g(\theta)\pi(\theta|x)d\theta. \quad (1.4)$$

Porém, em diversas aplicações práticas é extremamente complicado calcular essas integrais.

Existem diversos métodos aproximados baseados em simulações para obtenção dessas integrais, dentre os mais conhecidos está o método de Monte Carlo, descrito em [Robert e Casella \(2009\)](#), baseado na Lei Forte dos Grandes números que garante que, se é possível obter uma amostra de  $\pi(\theta|x)$  pode-se aproximar o valor de (1.4) por,

$$E_{\theta|x}[\widehat{g(\theta)}] = \frac{1}{n} \sum_{i=1}^n g(\theta_i) \xrightarrow{q.c.} E_{\theta|x}[g(\theta)], \quad (1.5)$$

em que  $g : \Theta \rightarrow \mathbb{R}$ . Em particular, note que decorre da equação (1.5) que  $E_{\theta|x}[\widehat{g(\theta)}^2] = \frac{1}{n} \sum_{i=1}^n g(\theta_i)^2$ . Com isso, pode-se escrever

$$Var_{\theta|x}[\widehat{g(\theta)}] = \frac{1}{n} \sum_{i=1}^n g(\theta_i)^2 - \left( \frac{1}{n} \sum_{i=1}^n g(\theta_i) \right)^2 \xrightarrow{q.c.} Var_{\theta|x}[g(\theta)]. \quad (1.6)$$

No entanto, obter valores de  $\pi(\theta|x)$  em dimensões elevadas pode não ser trivial. Com isso, são propostos métodos de Monte Carlo via Cadeia de Markov (MCMC) como alternativa em casos onde é difícil obter valores da distribuição *a posteriori*.

Na próxima seção serão apresentadas algumas definições sobre Cadeias de Markov e o método MCMC através do algoritmos de Metropolis-Hastings (MH) e Gibbs *Sampling*.

### 1.3.1 Método de Monte Carlo via Cadeias de Markov

A ideia é basicamente simular valores da distribuição de interesse através de técnicas baseadas em cadeias de Markov e calcular estimativas amostrais de características desta. O primeiro membro da família MCMC é o Metropolis proposto por [Metropolis et al. \(1953\)](#), como parte da investigação das propriedades de equilíbrio a sistemas de partículas, e generalizado por [Hastings \(1970\)](#), denominado Metropolis-Hastings, tornando o método uma dominante em aplicações bayesianas, juntamente com o Gibbs *Sampling* apresentado por [Gelfand e Smith \(1990\)](#), para simulações de valores da distribuição *a posteriori*.

#### 1.3.1.1 Cadeias de Markov

**Definição 1.** Um processo estocástico  $\{X_t : t \in T\}$  é uma coleção de variáveis aleatórias indexadas no tempo. O conjunto dos possíveis valores de  $X_t$ ,  $S = \{s_1, s_2, \dots\}$ , é denominado espaço dos estados e  $T$  denota o espaço de tempo.

Considere que o processo estocástico está no estado  $s_i$  no tempo  $t - 1$ . A probabilidade do processo estar no estado  $s_j$  no tempo  $t$ , dada por

$$\mathbb{P}(X_t = s_j | X_{t-1} = s_i, \dots, X_0 = s_0), \quad (1.7)$$

será denominada probabilidade de transição (em um passo) do estado  $s_i$  para o estado  $s_j$ .

**Definição 2.** Uma Cadeia de Markov é um processo estocástico tal que,

$$\mathbb{P}(X_t = s_j | X_{t-1} = s_i, \dots, X_0 = s_0) = \mathbb{P}(X_t = s_j | X_{t-1} = s_i) = p_{ij}, \quad (1.8)$$

ou seja, o estado atual da cadeia depende apenas do estado no qual o processo esteve no tempo anterior.

### *Algumas propriedades importantes*

**Irredutibilidade** Uma cadeia de Markov é dita ser irredutível se para quaisquer dois estados,  $s_i, s_j \in \mathcal{S}$ , existe um inteiro  $t$  tal que  $\mathbb{P}(X_t = s_j | X_{t-1} = s_i) > 0$ . Isso significa que cada estado pode ser alcançado a partir de qualquer outro, usando apenas probabilidades positivas;

**Periodicidade** Seja  $\tau(s_i) := \{t \geq 1, \mathbb{P}(X_t = s_i | X_1 = s_i) = p_{ii} > 0\}$  o conjunto dos tempos em que é possível que a cadeia retorne a posição inicial  $s_i$ . O período do estado  $s_i$  é definido ser o máximo divisor comum de  $\tau(s_i)$ ;

**Reversibilidade** Um cadeia de Markov é dita reversível se as probabilidades de transição entre os estados são invariantes,  $p_{ij} = p_{ji}$ , ou seja, a probabilidade de estar em um estado  $s_i$  e ir para  $s_j$  é igual a probabilidade de estar no estado  $s_j$  e ir para  $s_i$ .

**Lema 1.** Se uma cadeia de Markov é irredutível, então  $\text{mdc}\tau(s_i) = \text{mdc}\tau(s_j)$  para todo  $s_i, s_j \in \mathcal{S}$ .

Para uma cadeia de Markov irredutível, o período é definido ser comum para todos os estados. A cadeia é dita ser aperiódica se todos os estados tem período 1.

Para utilizar o método de Monte Carlo de maneira adequada é necessário que a Cadeia de Markov gerada seja, irredutível, aperiódica e reversível. Essas condições garantem que, independente do valor inicial da cadeia,  $X_t$  converge para a distribuição invariante. Para mais detalhes veja [Levin e Peres \(2017\)](#).

#### *1.3.1.2 Algoritmo de Metropolis-Hastings*

O algoritmo Metropolis-Hastings (MH) utiliza uma distribuição auxiliar para simular valores aleatórios e aceitá-los com uma certa probabilidade. Este mecanismo garante a convergência da cadeia para a distribuição de equilíbrio, que neste caso é a distribuição de interesse.

Suponha que deseja-se simular valores de uma distribuição qualquer  $\pi(\theta)$  a partir de uma distribuição auxiliar  $q$  no mesmo domínio de  $\pi$ . O algoritmo MH é dado por

- i. Inicialize o contador de iterações  $t = 0$  e especifique um valor inicial  $\theta^{(0)}$ .
- ii. Gere o novo valor  $\theta'$  da distribuição proposta  $q(\cdot|\theta)$ .
- iii. Calcule a probabilidade de aceitação

$$\alpha(\theta, \theta') = \min \left( 1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right) \quad (1.9)$$

- iv. Gere  $u \sim U(0, 1)$ . Se  $u \leq \alpha(\theta, \theta')$  aceite o novo valor e faça  $\theta^{(n+1)} = \theta'$ , caso contrário rejeite e faça  $\theta^{(n+1)} = \theta$ .
- v. Incremente o contador de  $t$  para  $t + 1$  e volte ao passo ii., repetindo os passos o número desejado de vezes.

A ideia é que conforme o número de iterações aumenta, a cadeia gradualmente “esquece” os valores iniciais e eventualmente converge para uma distribuição de equilíbrio.

Note que o algoritmo MH faz a transição de  $\theta \rightarrow \theta'$  de acordo com  $q(\cdot)$  e aceita com probabilidade  $\alpha(\theta, \theta')$ , ou seja, a probabilidade de transição fica dada por,

$$p_{\theta, \theta'} = q(\theta'|\theta)\alpha(\theta, \theta'). \quad (1.10)$$

A equação (1.10) satisfaz a condição de reversibilidade (homogeneidade), assim,

$$\begin{aligned} p_{\theta, \theta'} \pi(\theta) &= q(\theta'|\theta)\alpha(\theta, \theta')\pi(\theta) \\ &= q(\theta'|\theta) \min \left( 1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right) \pi(\theta) \\ &= \min (\pi(\theta)q(\theta'|\theta), \pi(\theta')q(\theta|\theta')) \\ &= p_{\theta', \theta} \pi(\theta'). \end{aligned} \quad (1.11)$$

Em aplicações bayesianas o interesse está em simular valores da distribuição a *posteriori*, i.e.,  $\pi(\theta|x)$ , assim, a probabilidade de aceitação (1.9) assume a forma particular,

$$\alpha(\theta, \theta') = \min \left( 1, \frac{p(x|\theta')\pi(\theta')q(\theta|\theta')}{p(x|\theta)\pi(\theta)q(\theta'|\theta)} \right). \quad (1.12)$$

### 1.3.1.3 Algoritmo Gibbs Sampling

O Gibbs *Sampling* não apresenta o mecanismo de aceitação do Metropolis, fazendo com que a cadeia se mova sempre para um novo valor. Considere um vetor aleatório  $\theta = (\theta_1, \dots, \theta_d)$  com distribuição conjunta  $\pi(\theta_1, \dots, \theta_d)$  sendo que as distribuições condicionais são conhecidas, ou seja,

$$\pi(\theta_i|\theta_{-i}), \quad \text{em que} \quad \theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)^\top. \quad (1.13)$$

Assim, as transições de um estado para outro são feitas de acordo com as distribuições condicionais. O algoritmo Gibbs *Sampling* é dado por

- i. Inicialize o contador de iterações  $t = 0$  e especifique valores iniciais  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_d^{(0)})^\top$ .
- ii. Obtenha o valor de  $\theta^{(t+1)}$  a partir de  $\theta^{(t)}$  através da geração sucessiva

$$\begin{aligned}\theta_1^{(t+1)} &\sim \pi(\theta_1 | \theta_2^{(t)}, \theta_3^{(t)}, \dots, \theta_d^{(t)}) \\ \theta_2^{(t+1)} &\sim \pi(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_d^{(t)}) \\ &\vdots \\ \theta_d^{(t+1)} &\sim \pi(\theta_d | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{d-1}^{(t+1)})\end{aligned}$$

- iii. Incremente o contador  $t$  e retorne ao passo ii. até obter a convergência.

Desta forma, cada iteração se completa após  $d$  movimentos ao longo dos eixos coordenados das componentes de  $\theta$ . Após a convergência, os valores resultantes formam uma amostra de  $\pi(\theta)$ . Vale notar que mesmo em problemas de alta dimensionalidade as simulações são univariadas, o que, de maneira geral, pode ser visto como uma vantagem computacional, pois o método utiliza as distribuições condicionais completas. Além disso, o Gibbs *Sampling* pode ser considerado como um caso particular do Metropolis-Hastings, no qual os elementos de  $\theta$  são atualizados um de cada vez (ou em blocos), tomando a distribuição condicional como proposta e probabilidade de aceitação igual a 1. Mais detalhes sobre os métodos até agora descritos podem ser obtidos em [Gamerman e Lopes \(2006\)](#) e [Robert e Casella \(2009\)](#).

Recentemente, [Neal et al. \(2011\)](#) introduziu o método de Monte Carlo Hamiltoniano que trata-se de um algoritmo de simulação Monte Carlo via Cadeia de Markov (MCMC) trazido para sistemas dinâmicos hamiltonianos. Este método possui vantagens em relação aos algoritmos já citados, evitando, por exemplo, o comportamento de passeio aleatório e consequentemente reduzindo a dependência excessiva dos valores gerados, o que faz com que a convergência para distribuição estacionária seja demorada. Este método será abordado com maiores detalhes no Capítulo 3.

O Capítulo 2 apresentará os principais conceitos relacionados a modelos de Volatilidade para séries que apresentam heteroscedasticidade.

---

## MODELOS PARA VOLATILIDADE

---

Existem variadas formas de especificar como a variância condicional (volatilidade) varia com o tempo. Neste capítulo serão descritos alguns dos modelos mais conhecidos para séries financeiras que apresentam evolução temporal na volatilidade.

### 2.1 Modelos de Volatilidade

#### 2.1.1 Modelos ARCH

Os modelos ARCH ou autoregressivos com heteroscedasticidade condicional, foram introduzidos por [Engle \(1982\)](#). A estratégia utilizada para estimar a variância condicional de uma série de retornos ( $r_t$ ), consiste em assumir que ela depende de retornos passados por meio de uma função quadrática, embora o retorno  $r_t$  seja não correlacionado serialmente.

**Definição 3.** O modelo ARCH( $p$ ) é definido por

$$r_t = \sqrt{\sigma_t^2} \varepsilon_t, \quad (2.1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_p r_{t-p}^2, \quad (2.2)$$

em que  $\{\varepsilon_t\}$  é um processo puramente aleatório com média zero e variância igual a 1,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0, i = 1, \dots, p$ .

Usualmente supõe-se que  $\varepsilon_t \sim \mathcal{N}(0, 1)$ ,  $\varepsilon_t \sim t_\nu$  ou uma distribuição que descreva as caudas pesadas da série ([MORETTIN, 2008](#)).

Um caso particular, em que é possível investigar algumas propriedades desses modelos, é o ARCH(1) ou autorregressivo condicionalmente heteroscedástico de ordem 1, dado por

$$r_t = \sqrt{\sigma_t^2} \varepsilon_t, \quad (2.3)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2, \quad (2.4)$$

com  $\alpha_0 > 0$  e  $\alpha_1 \geq 0$ . Note que quando  $\alpha_1 = 0$  a variância condicional é constante e  $r_t$  é um processo condicionalmente homoscedástico. Neste caso, deseja-se garantir a estacionariedade da autorregressão de modo que deve-se impor a restrição  $0 < \alpha_1 < 1$ . Pode-se notar também que (2.4) não inclui um termo de erro.

Além disso, se o processo  $\{r_t\}$  for estacionário, tem-se que a média, variância e covariâncias incondicionais da série são obtidas por:

- (i)  $E(r_t) = E(E(r_t|\mathcal{F}_{t-1})) = 0$ ;
- (ii)  $Var(r_t) = E(r_t^2) = E(E(r_t^2|\mathcal{F}_{t-1})) = E(\alpha_0 + \alpha_1 r_{t-1}^2) = \frac{\alpha_0}{1-\alpha_1}$ ;
- (iii)  $Cov(r_t, r_{t+k}) = E(r_t r_{t+k}) = E(E(r_t r_{t+k}|\mathcal{F}_{t+k-1})) = E(r_t E(\sqrt{h_{t+k}} \varepsilon_{t+k}|\mathcal{F}_{t+k-1})) = 0$ , para  $k > 0$ , pois  $r_t$  está em  $\mathcal{F}_{t+k-1}$  e  $E(\varepsilon_{t+k}|\mathcal{F}_{t+k-1}) = 0$ .

Lembrando que  $\mathcal{F}_{t-1} = \{r_{t-1}, \dots, r_1\}$  denota o conjunto de informações até o instante  $t - 1$ .

Vale ressaltar que geralmente os retornos apresentam caudas longas, de maneira que a curtose,  $\kappa$ , é maior que 3. Pode-se mostrar que o coeficiente de curtose, supondo que  $\varepsilon_t$  sejam normais, é, então,

$$\kappa = \frac{E(r_t^4)}{[Var(r_t)]^2} = \frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2}. \quad (2.5)$$

Uma desvantagem deste modelo é tratar retornos positivos e negativos de forma similar, uma vez que os quadrados dos retornos entram na fórmula da volatilidade, o que não é muito razoável sabendo que a volatilidade reage de modo diferente a retornos positivos e negativos.

### 2.1.1.1 Identificação

A característica fundamental nos modelos ARCH é que a volatilidade dos retornos  $r_t$  se comporta como um processo autorregressivo. Sendo assim, para a construção do modelo, inicialmente seria interessante ajustar modelos ARMA, para remover a correlação serial na série, se ela existir. Dessa forma, se este for o caso, tem-se que

$$\phi(B)r_t = \theta_0 + \theta(B)a_t,$$

sendo que  $a_t \sim \text{ARCH}(p)$ . Portanto, no que se refere a  $r_t$ , estaremos supondo ou que a série é não correlacionada, ou então ela é o resíduo da aplicação de um modelo ARMA à série original (MORETTIN, 2008). Particularmente, se o processo ajustado for adequado então a Função de Autocorrelações (FAC) e a Função de Autocorrelação Parcial (PACF) dos resíduos devem indicar um processo puramente aleatório, no entanto, se a FAC dos quadrados dos resíduos tiver um decaimento característico de uma autorregressão, isso é uma indicação de que um modelo ARCH pode ser apropriado (EHLERS, 2007).

### 2.1.2 Modelos GARCH

Os modelos GARCH (*generalized ARCH*) são uma generalização dos modelos ARCH, apresentada por [Bollerslev \(1986\)](#), que consiste em assumir que a FAC de  $r_t^2$  se comporta como um processo ARMA, isto é, depende também de seus valores passados, o que pode tornar o modelo mais parcimonioso, no sentido de apresentar menos parâmetros em relação ao ARCH.

**Definição 4.** O modelo GARCH(p,q) é definido por

$$r_t = \sqrt{\sigma_t^2} \varepsilon_t, \quad (2.6)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i r_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad (2.7)$$

em que  $\{\varepsilon_t\}$  são v.a. i.i.d., com média zero e variância igual a 1,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0, i = 1, \dots, p$ ,  $\beta_i \geq 0, j = 1, \dots, q$ ,  $\sum_{i=1}^m (\alpha_i + \beta_i) < 1$ ,  $m = \max(p, q)$ .

#### 2.1.2.1 Identificação

As ferramentas básicas para identificação são mais uma vez a FAC e a FACP dos quadrados dos resíduos, onde estas devem indicar um processo ARMA(p,q). Neste caso a identificação pode não ser trivial, entretanto, na prática em diversos casos o modelo GARCH(1,1) é suficiente, recomenda-se também o uso de modelos de ordem baixa, como (1,2), (2,1) ou (2,2), e depois escolher o modelo com base em variados critérios (veja [Morettin e Tolo \(2006\)](#) para mais detalhes).

## 2.2 Modelos de Volatilidade Estocástica

Os modelos da classe GARCH supõem que a variância condicional depende de retornos passados e a equação de volatilidades apresentadas por esses modelos são determinísticas, no sentido de que não assumem um componente aleatório. No entanto, é razoável assumir que a equação da variância condicional varia estocasticamente no tempo, especialmente quando existem mudanças inesperadas.

Os modelos de volatilidade estocástica (VE) foram introduzidos por [Taylor \(1986\)](#) e consistem em assumir um processo estocástico para uma função da equação da variância condicional. Estes modelos tem como premissa o fato de que a volatilidade presente depende de seus valores anteriores, mas é independente dos retornos passados.

**Definição 5.** O modelo VE na forma canônica de [Kim, Shephard e Chib \(1998\)](#) é definido por

$$r_t = \beta \exp(h_t/2) \varepsilon_t, \quad (2.8)$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \quad (2.9)$$

$$h_1 \sim \mathcal{N} \left( \mu, \frac{\sigma_\eta^2}{1 - \phi^2} \right),$$

em que  $h_t$  é um processo estacionário do logaritmo da volatilidade no tempo  $t$ ,  $\{\varepsilon_t\}$  é uma sequência de variáveis aleatórias i.i.d. com média zero e variância unitária,  $\{\eta_t\}$  é uma sequência i.i.d., tal que  $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$ , em que  $\varepsilon_t$  e  $\eta_t$  são independentes para todo  $t$ . Tem-se que  $\mu$  é o nível das log-volatilidades e  $|\phi| < 1$  é o parâmetro de persistência. O parâmetro  $\beta$  desempenha o papel de fator escala e por questões de identificabilidade  $\beta$  deve ser configurado para um ou  $\mu$  para zero.

A equação de log-volatilidades do modelo (2.9) pode ser estendida para um caso mais geral, ou seja,

$$h_t = \mu + \sum_{i=1}^p \phi_i (h_{t-i} - \mu) + \eta_t, \quad (2.10)$$

Além disso, algumas propriedades relacionadas aos momentos incondicionais são enumeradas abaixo:

- (i)  $E(r_t) = E(\beta e^{h_t/2} \varepsilon_t) = \beta E(e^{h_t/2}) E(\varepsilon_t) = 0$ , já que  $h_t$  e  $\varepsilon_t$  são independentes;
- (ii)  $Var(r_t) = E(r_t^2) = E(\beta^2 e^{h_t} \varepsilon_t^2) = \beta^2 E(e^{h_t}) E(\varepsilon_t^2) = \beta^2 E(e^{h_t})$ . Mas, como assume-se que  $h_t$  é um processo estacionário segue que,

$$E(h_t) = E(\mu + \phi(h_{t-1} - \mu) + \eta_t) \Rightarrow E(h_t) = \mu + \phi E(h_t) - \phi \mu = \mu \quad e$$

$$Var(h_t) = Var(\mu + \phi(h_{t-1} - \mu) + \eta_t) \Rightarrow Var(h_t) = \frac{\sigma_\eta^2}{1 - \phi^2} = \sigma^2.$$

A distribuição incondicional das log-volatilidades é  $h_t \sim \mathcal{N}(\mu, \sigma^2)$ , assim,  $e^{h_t}$  segue uma distribuição log-normal com,

$$E(e^{h_t}) = e^{\mu + \sigma^2/2} \quad e \quad Var(e^{h_t}) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}.$$

Portanto,  $Var(r_t) = \beta^2 e^{\mu + \sigma^2/2}$ .

- (iii)  $E(r_t^4) = E(\beta^4 e^{2h_t} \varepsilon_t^4) = \beta^4 E(e^{2h_t}) E(\varepsilon_t^4)$ . Se  $\varepsilon_t \sim \mathcal{N}(0, 1)$ , então,  $E(\varepsilon_t^4) = 3$  e  $E(r_t^4) = 3\beta^4 E(e^{2h_t})$ . Mas,

$$E(e^{2h_t}) = Var(e^{h_t}) + E^2(e^{h_t}) = e^{2\mu + 2\sigma^2}.$$

Logo,  $E(r_t^4) = 3\beta^4 e^{2\mu + 2\sigma^2}$ . Assim, a curtose é dada por

$$\kappa = \frac{E(r_t^4)}{[Var(r_t)]^2} = \frac{3\beta^4 e^{2\mu + 2\sigma^2}}{\beta^4 e^{2\mu + \sigma^2}} = 3e^{\sigma^2},$$

que é sempre maior que 3, pois  $e^{\sigma^2} > 1$ . Um resultado mais geral é que  $\kappa = E(\varepsilon_t^4) e^{\sigma^2}$ , ou seja, a curtose induzida por este modelo é sempre maior do que a curtose de  $\varepsilon_t$ .

Nos modelos VE, condicional ao conjunto de informações  $\mathcal{F}_t = \{r_t, r_{t-1}, \dots\}$ , o desvio padrão de  $r_t$  é dado por,

$$\sigma_t = \beta \exp(h_t/2).$$

Em finanças, se  $r_t$  representa o  $t$ -ésimo retorno, então  $\sigma_t$  é a volatilidade no tempo  $t$ .

Na formulação original dos modelos, introduzida por Taylor (1986), supõe-se  $\varepsilon_t$  seguindo uma distribuição normal padrão. No entanto, muitos estudos empíricos indicam que este modelo não considera a curtose observada nos retornos em diversas séries temporais financeiras. Com isso, considerou-se também outras distribuições para a equação de observações do modelo, são elas:

A distribuição de potência exponencial (ou *generalized error distribution*, GED) com média zero e variância unitária com função densidade dada por,

$$p(\varepsilon_t) = \frac{\nu}{\lambda 2^{(1+\frac{1}{\nu})} \Gamma(\frac{1}{\nu})} \exp \left\{ -\frac{1}{2} \left| \frac{\varepsilon_t}{\lambda} \right|^\nu \right\}, \quad (2.11)$$

em que  $\lambda^2 = 2^{-\frac{2}{\nu}} \Gamma(\frac{1}{\nu}) / \Gamma(\frac{3}{\nu})$  e o parâmetro de forma  $\nu > 0$ . Importantes casos são, a distribuição de Laplace para  $\nu = 1$  e a normal padrão quando  $\nu = 2$ . A curtose é dada por  $\Gamma(\frac{1}{\nu}) \Gamma(\frac{5}{\nu}) \Gamma(\frac{3}{\nu})^2 - 3$  de modo que, quando  $\nu < 2$ , esta distribuição apresenta caudas pesadas.

Distribuição t-Student com  $\nu$  graus de liberdade e função densidade dada por,

$$p(\varepsilon_t) = \frac{1}{\sqrt{\pi(\nu-2)}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left\{ 1 + \frac{\varepsilon_t^2}{\nu-2} \right\}^{-\frac{\nu+1}{2}}. \quad (2.12)$$

Quando  $\nu \rightarrow \infty$  esta distribuição se aproxima de uma Gaussiana padrão.

E adicionalmente, a distribuição Skew-Normal padrão com parâmetro de forma  $\nu$  e densidade dada por,

$$p(\varepsilon_t) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{\varepsilon_t^2}{2} \right\} \left\{ 1 + \operatorname{erf} \left[ \nu \left( \frac{\varepsilon_t}{\sqrt{2}} \right) \right] \right\}, \quad (2.13)$$

em que  $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ . Essa distribuição é assimétrica à esquerda para  $\nu < 0$  e à direita quando  $\nu > 0$ . A Gaussiana padrão é um caso particular se  $\nu = 0$ .

### 2.2.1 Funções de Verossimilhança

A função de verossimilhança conjunta  $L_{r|\theta} = p(r, h|\theta)$ , em que  $\theta = (\beta, \mu, \phi, \sigma_\eta, \nu)$ , é dada por

$$p(r, h|\beta, \mu, \phi, \sigma_\eta, \nu) = p(h_1|\mu, \phi, \sigma_\eta) \prod_{t=2}^T p(h_t|h_{t-1}, \mu, \phi, \sigma_\eta) \prod_{t=1}^T p(r_t|h_t, \beta, \nu), \quad (2.14)$$

em que  $h_1 | (\mu, \phi, \sigma_\eta) \sim \mathcal{N}(\mu, \sigma_\eta^2 / (1 - \phi^2))$  e  $h_t | (h_{t-1}, \mu, \phi, \sigma_\eta) \sim \mathcal{N}(\mu + \phi(h_{t-1} - \mu), \sigma_\eta^2)$ . Para erros Normais, temos,

$$p(r_t | h_t, \beta) = \frac{1}{\sqrt{2\pi}\beta \exp(h_t/2)} \exp\left\{-\frac{r_t^2}{2\beta^2 \exp(h_t)}\right\}, \quad (2.15)$$

para erros GED,

$$p(r_t | h_t, \beta, \nu) = \frac{\nu}{\beta \lambda 2^{(1+\frac{1}{\nu})} \Gamma(\frac{1}{\nu})} \exp\left\{-\frac{h_t}{2} - \frac{1}{2\lambda^\nu} \left|\frac{r_t}{\beta \exp(h_t/2)}\right|^\nu\right\}, \quad (2.16)$$

para erros t-Student,

$$p(r_t | h_t, \beta, \nu) = \frac{1}{\beta \sqrt{\pi(\nu-2)}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left\{1 + \frac{r_t^2}{\beta^2(\nu-2) \exp(h_t)}\right\}^{-\frac{\nu+1}{2}} \exp(-h_t/2), \quad (2.17)$$

e para erros Skew-Normal,

$$p(r_t | h_t, \beta, \nu) = \frac{1}{\sqrt{2\pi}\beta \exp(h_t/2)} \exp\left\{-\frac{r_t^2}{2\beta^2 \exp(h_t)}\right\} \left\{1 + \operatorname{erf}\left[\nu \left(\frac{r_t}{\sqrt{2}\beta \exp(h_t/2)}\right)\right]\right\}. \quad (2.18)$$

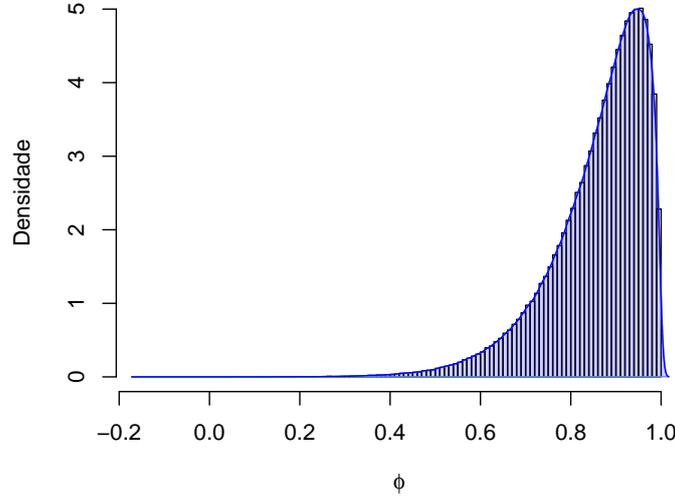
## 2.2.2 Distribuições a Priori

Para completar a especificação do modelo, sob o paradigma bayesiano, definimos a *priori* para  $\phi$  de acordo com Kim, Shephard e Chib (1998). Sendo  $\phi = 2\phi^* - 1$  em que  $\phi^* \sim \mathcal{B}(\alpha; \beta)$  temos,

$$\pi(\phi) = \frac{1}{2B(\alpha, \beta)} \left(\frac{1+\phi}{2}\right)^{\alpha-1} \left(\frac{1-\phi}{2}\right)^{\beta-1}, \quad \alpha, \beta > \frac{1}{2}, \quad (2.19)$$

assim o suporte da distribuição pertence ao intervalo  $[-1, 1]$ . Note que  $E(\phi) = 2\frac{\alpha}{\alpha+\beta} - 1$ , neste caso, foram selecionados os parâmetros  $\alpha = 20$  e  $\beta = 1,5$ , implicando em uma média a *priori* de 0,86. O gráfico estimado desta distribuição é dado pela Figura 1. A maior parte dos autores escolhe esta *priori* ((GIROLAMI; CALDERHEAD, 2011), (KASTNER; FRÜHWIRTH-SCHNATTER, 2014), (ZEVALLOS; GASCO; EHLERS, 2016)), pois na prática geralmente os valores de  $\phi$  são próximos de um. Alternativamente, outras distribuições a *priori* podem ser utilizadas, como por exemplo,  $\pi(\phi) \propto 1$  que conduz uma densidade condicional analiticamente tratável. Porém, se  $\phi = 1$  os termos de  $\mu$  da equação de log-volatilidades do modelo (2.8) se cancelam e, portanto,  $\mu$  se torna identificável a partir dos dados. A *priori* selecionada evita este tipo de problema (KIM; SHEPHARD; CHIB, 1998).

Por questões de identificabilidade o modelo é configurado com o parâmetro  $\beta = 1$  ou  $\mu = 0$ . No caso em que  $\mu = 0$ , para o parâmetro  $\beta$  foi proposta uma distribuição exponencial com média um. Quando se configura  $\beta = 1$ , propomos  $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$ .

Figura 1 – Densidade simulada  $\pi(\phi)$  com os parâmetros  $\alpha = 20$  e  $\beta = 1,5$ .

Para o parâmetro  $\sigma_\eta \in \mathbb{R}^+$  usualmente emprega-se a *priori* Gama Inversa, no entanto, foram propostas outras duas distribuições a *priori*, sendo

$$\sigma_\eta^2 \sim \mathcal{G}^{-1}\left(\frac{a_\sigma}{2}, \frac{b_\sigma}{2}\right), \quad (2.20)$$

$$\sigma_\eta^2 \sim B_\sigma \times \chi_1^2 = \mathcal{G}\left(\frac{1}{2}, \frac{1}{2B_\sigma}\right), \quad (2.21)$$

$$\sigma_\eta^2 \sim \text{Inv} - \chi^2(c_\sigma, s_\sigma). \quad (2.22)$$

Consideramos a distribuição a *priori* (2.22) de acordo com [Girolami e Calderhead \(2011\)](#) e (2.21) conforme utilizado em [Kastner \(2016\)](#), por efeito de comparação e pela escolha desta ser menos influente quando a verdadeira volatilidade da volatilidade é pequena.

Para o parâmetro  $v$  a distribuição a *priori* depende da distribuição adotada para os erros da equação de observações do modelo. No caso da distribuição GED,

$$v \sim \text{Inv} - \chi^2(10, 0.05), \quad (2.23)$$

quando os erros seguem distribuição t-Student, foi considerada a distribuição exponencial truncada, com densidade

$$\pi(v) = \lambda \exp\{-\lambda(v-4)\}, \quad v > 4, \quad (2.24)$$

neste caso, foi especificado  $\lambda = 1/3$ . As distribuições (2.23) e (2.24) foram tomadas de acordo com [Zevallos, Gasco e Ehlers \(2016\)](#). E para a distribuição Skew-Normal,

$$v \sim \mathcal{N}(0, 5). \quad (2.25)$$

Portanto, a densidade a *priori* é denotada por  $\pi(\theta) = \pi(\beta)\pi(\mu)\pi(\phi)\pi(\sigma_\eta)\pi(v)$ .

## 2.3 Seleção de Modelos

Explorar e encontrar um modelo que representa a dinâmica dos dados de forma mais adequada (entre vários candidatos) é um problema de grande importância em uma pesquisa estatística. A escolha natural para medir o desempenho do modelo é utilizar uma técnica chamada *cross-validation*. O conceito central do *cross-validation* é o particionamento do conjunto de dados em subconjuntos mutuamente exclusivos e, posteriormente, utilizar alguns destes subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento) e o restante dos subconjuntos são empregados no processo de validação do modelo.

O método *cross-validation k-fold* consiste em dividir o conjunto total de dados em  $n/k$  subconjuntos e, a partir disto, um subconjunto é utilizado para estimação dos parâmetros e calcula-se a acurácia do modelo. Quando  $k = n$ , tem-se um *leave-one-out cross-validation*, onde o conjunto de validação consiste em apenas um único ponto de dados de cada vez.

Embora atraente, o *cross-validation* tem a desvantagem de ser computacionalmente caro, sendo indicado para situações onde poucos dados estão disponíveis. Neste caso, uma alternativa é a utilização de critérios de informação. Dentre os mais conhecidos, provavelmente, o critério de informação mais utilizado é o *Akaike Information Criterion* AIC proposto por [Akaike \(1974\)](#).

Abaixo são descritos os critérios a serem usados neste trabalho, de acordo com a metodologia de inferência Bayesiana.

### 2.3.1 Deviance Information Criterion

Uma popular escolha na literatura inferencial Bayesiana é o *Deviance Information Criterion* (DIC) proposto por [Spiegelhalter et al. \(2002\)](#), que é definido como,

$$\text{DIC} = -2 \log p(y|\hat{\theta}_{\text{Bayes}}) + 2p_D,$$

em que  $\hat{\theta}_{\text{Bayes}} = E[\theta|y]$  e  $p_D$  é uma penalização em termos do desvio médio a posteriori que mede a qualidade do ajuste do modelo aos dados. O  $p_D$  é dado por

$$p_D = 2 \left( \log p(y|\hat{\theta}_{\text{Bayes}}) - E_{\theta|y}[\log p(y|\theta)] \right).$$

Então, dado  $S$  simulações da distribuição a *posteriori* de  $\theta$  é possível aproximar diretamente o  $p_D$  como,

$$\hat{p}_D = 2 \left( \log p(y|\hat{\theta}_{\text{Bayes}}) - \frac{1}{S} \sum_{s=1}^S \log p(y|\theta^s) \right). \quad (2.26)$$

Uma observação relevante é que em alguns casos o DIC tende a selecionar modelos super ajustados e está aberto a diferentes variações. [Celeux et al. \(2006\)](#) apresentam e comparam várias extensões do critério no cenário de misturas de distribuições e modelos de efeitos aleatórios.

### 2.3.2 Watanabe Information Criterion

Um critério de informação mais recentemente desenvolvido é o chamado *Watanabe information criterion* (WAIC), que foi descrito por [Watanabe \(2010\)](#). O WAIC tem a vantagem sobre o DIC, em termos do cálculo da média sobre a distribuição a *posteriori* e também devido ao fato de não depender da teoria assintótica de Fisher, como apontado por [Gelman et al. \(2014\)](#). Por conseguinte, não assume que a *posteriori* converge para um único ponto tornando-se uma opção atraente para os modelos hierárquicos e de mistura. O WAIC pode ser interpretado com uma aproximação computacionalmente conveniente para validação cruzada e é definido com base em,

$$\text{elpd}_{\text{waic}} = \text{lpd} - p_{\text{waic}},$$

em que  $p_{\text{waic}} = \sum_{i=1}^n \text{Var}_{\theta|y}(\log p(y_i|\theta))$  é uma penalização para o número efetivo de parâmetros e  $\text{lpd} = \sum_{i=1}^n \log p(y_i|y)$  é a log densidade preditiva pontual, tal como definido por [Gelman et al. \(2014\)](#). Pode-se reescrever,

$$\log p(y_i|y) = \log \int_{\Theta} p(y_i|\theta) p(\theta|y) d\theta.$$

Então, dado  $S$  simulações pode aproximar a variância posteriori da log densidade para cada ponto de  $y_i|\theta$  utilizando a variância amostral, isto é,  $V_{s=1}^S a_s = \frac{1}{S-1} \sum_{s=1}^S (a_s - \bar{a})^2$ . Assim,

$$\hat{p}_{\text{waic}} = \sum_{i=1}^n V_{s=1}^S (\log p(y_i|\theta)). \quad (2.27)$$

E também a log densidade preditiva pontual por,

$$\widehat{\text{lpd}} = \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i|\theta^s) \right). \quad (2.28)$$

Com isso, o WAIC é definido por,

$$\text{WAIC} = -2\text{elpd}_{\text{waic}}. \quad (2.29)$$

[Watanabe \(2010\)](#) mostrou, que sob algumas condições de regularidade, o WAIC é assintoticamente equivalente ao *leave-one-out cross-validation* ([GELMAN; HWANG; VEHTARI, 2014](#)).

### 2.3.3 Approximate Leave-One-Out Cross-Validation

A estimativa Bayesiana *Approximate Leave-One-Out Cross-Validation* (LOO) é,

$$\text{lpd}_{\text{loo}} = \sum_{i=1}^n \log p(y_i|y_{-i}) = \sum_{i=1}^n \log \int_{\Theta} p(y_i|\theta) p(\theta|y_{-i}) d\theta,$$

que é a log densidade preditiva pontual sem o  $i$ -ésimo ponto dos dados no cálculo da distribuição a *posteriori*.

De acordo com [Gelfand \(1996\)](#), se os  $n$  pontos forem condicionalmente independentes no modelo de dados, o LOO pode ser calculado facilmente usando amostragem por importância com pesos,

$$w_s = \frac{1}{p(y_i|\theta^s)} \approx \frac{p(\theta^s|y_{-i})}{p(\theta^s|y)}. \quad (2.30)$$

Porém, o uso direto desses pesos pode induzir à estimativas ruidosas com variância alta ou infinita nas caudas da distribuição. Em vez disso, estabiliza-se os pesos, substituindo  $w_s$  por,

$$\tilde{w}_s = \min(w_s, \sqrt{S\bar{w}}).$$

conforme recomendado por [Ionides \(2008\)](#). No entanto, a desvantagem deste truncamento é que ele induz um viés, que pode ser grande em diversos casos ([VEHTARI; GELMAN; GABRY, 2017](#)). Com isso, [Vehtari, Gelman e Gabry \(2015\)](#) propõem uma abordagem inovadora que fornece estimativas mais precisas e confiáveis, usando pesos suavizados ao ajustar uma distribuição de Pareto generalizada para a parte superior da distribuição dos pesos de importância. Este procedimento pode ser resumido nos seguintes passos:

1. Ajustar a distribuição de Pareto Generalizada para os 20% maiores pesos de importância  $w_s$  conforme calculado em (2.30);
2. Estabilize os pesos de importância, substituindo as  $M$  maiores taxas pelos valores esperados das estatísticas de ordem da distribuição de Pareto Generalizada ajustada,

$$F^{-1}\left(\frac{z-1/2}{M}\right), \quad z = 1, \dots, M,$$

em que  $M$  é o número de simulações usadas para ajustar a Pareto (Neste caso,  $M = 0,2 \times S$ ) e  $F^{-1}$  é a inversa da distribuição acumulada;

3. Para garantir variância finita da estimativa, trunca-se os pesos suavizados em  $S^{3/4}\bar{w}$ , em que  $\bar{w}$  é a média dos pesos suavizados.

As etapas acima devem ser realizadas para cada ponto de dados  $i$ . O resultado é um vetor de pesos  $w_i^s$ ,  $s = 1, \dots, S$ , para cada  $i$ .

A estimativa LOO estabilizada utilizando amostragem por importância suavizada de Pareto (PSIS) da log densidade preditiva esperada é dada por,

$$\widehat{\text{elpd}}_{\text{psis-loo}} = \sum_{i=1}^n \log \left( \frac{\sum_{s=1}^S p(y_i|\theta^s) w_i^s}{\sum_{s=1}^S w_i^s} \right). \quad (2.31)$$

Seguindo a literatura sobre os critérios de informação, a penalização para o número efetivo de parâmetros é estimado pela diferença,

$$\widehat{p}_{\text{psis-loo}} = \widehat{\text{lpd}} - \widehat{\text{elpd}}_{\text{psis-loo}}, \quad (2.32)$$

que é (2.28) menos (2.31). Com isso, o LOO é dado por,

$$\text{LOO} = -2\widehat{\text{elpd}}_{\text{psis-loo}}. \quad (2.33)$$

## Erro Padrão e Biblioteca loo

As estimativas calculadas  $\widehat{\text{elpd}}_{waic}$  e  $\widehat{\text{elpd}}_{psis-loo}$  são definidas como a soma de  $n$  componentes independentes, por isso é trivial calcular seu erro padrão, em cada caso, calculando o desvio padrão dos  $n$  componentes e multiplicando por  $\sqrt{n}$ . Por exemplo, defina

$$\widehat{\text{elpd}}_{waic_i} = \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i | \theta^s) \right) - \left( V_{s=1}^S (\log p(y_i | \theta)) \right),$$

de modo que  $\widehat{\text{elpd}}_{waic}$  é a soma desses  $n$  termos independentes. Então, o erro padrão é definido por

$$\text{EP}(\widehat{\text{elpd}}_{waic}) = \sqrt{n V_{s=1}^S (\widehat{\text{elpd}}_{waic})},$$

e de forma similar para  $\widehat{\text{elpd}}_{psis-loo}$ . Os penalizações para o número efetivo de parâmetros  $\hat{p}_{waic}$  e  $\hat{p}_{psis-loo}$  também são somas de termos independentes e, portanto, pode-se calcular os erros padrões associados da mesma maneira.

[Vehtari, Gelman e Gabry \(2015\)](#) desenvolveram uma biblioteca em R chamada `loo` que fornece estimativas dos critérios de informação WAIC e LOO através de objetos `stanfit` que, por sua vez, são obtidos com o auxílio do pacote `rstan` que será mencionado no próximo Capítulo. A biblioteca retorna, além das estimativas com seus respectivos erros padrões, uma avaliação da confiabilidade das estimativas LOO. Para poder obter os resultados dos critérios de informação é necessário fornecer à biblioteca a medida  $\widehat{\text{lpd}}$  que pode ser facilmente implementada junto ao modelo de interesse (veja [Vehtari, Gelman e Gabry \(2017\)](#) para mais detalhes).

## 2.4 Valor em Risco (VaR)

Uma forma de avaliar o desempenho dos modelos propostos é estimando o Valor em Risco (VaR) para múltiplos horizonte de tempo. O VaR que é um conceito empregado para medir o risco de mercado, esta medida avalia a exposição de uma posição financeira em relação às variações dos retornos, ou seja, mede a pior perda esperada que um ativo ou carteira de portfólio pode incorrer devido a eventos atípicos do mercado. A necessidade de quantificar o risco geralmente ocorre quando uma instituição financeira precisa determinar o montante de capital para manter como proteção contra perdas inesperadas ([FRANCQ; ZAKOIAN, 2011](#)).

O VaR é definido de forma a mensurar o grau de incerteza na obtenção do retorno esperado em uma determinada aplicação, visando quantificar o montante que um certo investimento pode perder. Seja  $r_t$  uma série de retornos financeiros, então, para um dado horizonte de tempo  $l$ , o VaR é dado por,

$$P(r_l \leq VaR) = \alpha^{(L)},$$

em que  $r_l = \ln(\Delta P_l)$  é a série de retornos para o período de tempo  $l$ . Assim, para a expressão do quantil da distribuição da variação do preço do ativo, por exemplo, na prática, estima-se

este quantil pela distribuição empírica dos retornos (MORETTIN, 2008). De uma perspectiva Bayesiana, dados os valores de retornos, as estimativas pontuais do VaR um passo a frente podem ser obtidas usando uma amostra de valores extraídos de sua distribuição preditiva, isto é,

$$E(\text{VaR}_{T+1|y}) \approx \frac{1}{S} \sum_{s=1}^S \text{VaR}_{T+1}^{(s)},$$

em que  $\text{VaR}_{T+1}^{(s)}$  é o valor em risco predito um passo a frente na iteração  $s$  conforme descrito em Zevallos, Gasco e Ehlers (2016). Como não há uma determinação analítica desta quantidade, adotamos o seguinte procedimento:

1. Dados os valores dos parâmetros e log-volatilidades na  $s$ -ésima iteração se obtém os valores de  $h_{T+1}^{(s)}$  a partir de  $\eta_{T+1}^{(s)} \sim \mathcal{N}(0, \sigma_\eta^{(s)})$  e definindo  $h_{T+1}^{(s)} = \mu^{(s)} + \phi^{(s)}(h_T^{(s)} - \mu^{(s)}) + \eta_{T+1}^{(s)}$ ;
2. Em seguida, simulamos  $K$  replicações  $\{\varepsilon_{T+1}^{(s,1)}, \dots, \varepsilon_{T+1}^{(s,K)}\}$  da distribuição dos erros (com parâmetro  $v^{(s)}$  para as distribuições GED, t-Student e Skew-Normal) ;
3. Finalmente, forma-se uma amostra de retornos, estabelecendo  $r_{T+1}^{(s,k)} = \beta^{(s)} \exp(h_{T+1}^{(s)}/2) \varepsilon_{T+1}^{(s,k)}$ .

Este procedimento permite aproximar o  $\text{VaR}_{T+1}^{(s)}$  de confiança  $\alpha$  pelo valor da amostra  $(1-\alpha)$ -quantil.

---

## MONTE CARLO HAMILTONIANO

---

O método de Monte Carlo Hamiltoniano (HMC) é uma técnica baseada na dinâmica de Hamilton que utiliza de forma simultânea a ideia de aceitação do Metropolis para realizar transições na Cadeia de Markov. Este método não faz uso de distribuições auxiliares (funções proposta), mas é capaz de percorrer rapidamente o suporte e produzir de modo eficiente simulações de valores aleatórios da distribuição alvo. Além disso, a forma de configuração do algoritmo é mais próxima do ideal para simular valores de distribuições com dimensões um tanto elevadas e com alta estrutura de correlação.

O método HMC foi originalmente chamado de híbrido Monte Carlo e trata-se de um algoritmo de amostragem cadeia de Markov Monte Carlo (MCMC), que foi trazido para sistemas dinâmicos hamiltonianos e integrado a estatística por [Duane \*et al.\* \(1987\)](#). O termo Halmiltoniano, introduzido por [Neal \*et al.\* \(2011\)](#) é mais utilizado na literatura recente.

A ideia consiste em resolver problemas de simulação dinâmica no qual as variáveis do estudo original são transformadas em variáveis de posição e variáveis artificiais são introduzidas no sistema. Essas variáveis artificiais geralmente possuem distribuição Gaussiana e referem-se às variáveis de momento do sistema Hamiltoniano ([NEAL \*et al.\*, 2011](#)). Basicamente, simula-se o movimento do deslocamento de uma partícula sob uma energia potencial igual ao logaritmo negativo da densidade de probabilidade de interesse. A cada iteração a velocidade da partícula é aleatorizada representando seu movimento durante determinado tempo. Após isso é obtida uma nova posição, que é justamente o valor proposto da distribuição alvo. Este valor é aceito ou não de acordo com a regra de aceitação do Metropolis.

### 3.1 Dinâmica Hamiltoniana

As equações de Hamilton formam um conjunto de equações diferenciais ordinárias vistas como uma reformulação das leis de Newton da mecânica clássica e descrevem de modo

determinístico como partículas (ou corpos) evoluem no tempo em um sistema conservativo e fechado. A dinâmica hamiltoniana possui uma interpretação física que pode fornecer intuições úteis. Em duas dimensões, podemos visualizar a dinâmica como um disco sem fricção que desliza sobre uma superfície de altura variável. O estado deste sistema consiste na posição do disco, dada por  $\theta$ , e o impulso do disco (sua massa vezes sua velocidade), dada por  $p$ .

A energia potencial  $U(\theta)$  do disco e a energia cinética  $K(p) = |p|^2/2m$  ( $m$  é a massa do disco) variam de acordo com o movimento do disco sobre a superfície. Em uma parte nivelada, o disco se move a uma velocidade constante igual a  $p/m$ . Se encontrar uma inclinação ascendente, o impulso do disco permite com que sua energia potencial aumente e a energia cinética diminua até o ponto em que seja nula. Neste momento, o disco inicia um movimento descendente passando por uma energia cinética máxima e potencial mínima.

A energia total deste sistema é representada matematicamente pela função hamiltoniana que é definida a seguir

**Definição 6** (Função Hamiltoniana). A energia total de um sistema fechado e conservativo é dada pela função Hamiltoniana,

$$H(\theta, p) = U(\theta) + K(p) \quad (3.1)$$

em que  $U(\theta)$  é a energia potencial e  $K(p)$  é a energia cinética do sistema.

A evolução determinística de partículas no decorrer do tempo é dada pela solução da Dinâmica Hamiltoniana

**Definição 7** (Dinâmica Hamiltoniana). As equações de movimento de Hamilton são definidas pelo sistema de equações diferenciais,

$$\begin{aligned} \frac{d\theta}{dt} &= \frac{\partial H}{\partial p} = \nabla_p K(p) \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial \theta} = -\nabla_\theta U(\theta) \end{aligned} \quad (3.2)$$

A descrição do movimento por estas equações é denominada Dinâmica Hamiltoniana.

As derivadas parciais da função Hamiltoniana determinam como  $\theta$  e  $p$  mudam ao longo do tempo, por exemplo, suponha que para um intervalo qualquer de duração  $s$ , deseja-se saber a posição de uma partícula após a evolução  $t + s$ . Então, fixados os valores  $[\theta(t), p(t)]$  as equações definem um mapeamento para determinar a posição da partícula no tempo  $[\theta(t + s), p(t + s)]$ .

No entanto, as equações Hamiltonianas requerem a solução de um sistema de equações diferenciais que, exceto em casos muito simples, só pode ser aproximada numericamente empregando uma classe especial para solução aproximada de sistemas de equações diferenciais. O método de Störmer-Verlet (*leapfrog*) é o método ideal por possuir propriedades fundamentais na garantia da convergência das cadeias de Markov (LEIMKUHNER B. & REICH, 2005).

### 3.1.1 Algoritmo

Seja  $\theta \in \mathbb{R}^d$  e  $\pi(\theta|y)$  a distribuição de interesse. Considere que o logaritmo negativo da densidade a *posteriori* é dado por  $U(\theta) = -\ell(\theta) = -\log(p(y|\theta)) - \log(\pi(\theta)) + c$ . Assuma que o momento na função Hamiltoniana  $p$  tem distribuição Normal multivariada  $\mathcal{N}_d(0, M)$ , com  $M$  conhecido. Então, a densidade conjunta de  $(\theta, p)$  é definida pela exponencial negativa da função Hamiltoniana,

$$\begin{aligned} f(\theta, p) &\propto \exp\{-H(\theta, p)\} \\ &\propto p(y|\theta)\pi(\theta)\exp\{-p^\top M^{-1}p\}. \end{aligned} \quad (3.3)$$

Note que (3.3) é uma função de densidade fatorável e pela definição (7) pode-se obter valores da distribuição a *posteriori*. Para valores  $(\theta^{(\tau)}, p^{(\tau)})$  em um dado tempo  $\tau$ , inicialmente  $\tau = 0$ , é aplicada a solução numérica *leapfrog*,

$$\begin{aligned} p^{(\tau+\frac{\varepsilon}{2})} &= p^{(\tau)} + \frac{\varepsilon}{2}\nabla_{\theta}\ell(\theta^{(\tau)}) \\ \theta^{(\tau+\varepsilon)} &= \theta^{(\tau)} + \varepsilon\nabla_p K(p^{(\tau+\frac{\varepsilon}{2})}) \\ p^{(\tau+\varepsilon)} &= p^{(\tau+\frac{\varepsilon}{2})} + \frac{\varepsilon}{2}\nabla_{\theta}\ell(\theta^{(\tau+\varepsilon)}). \end{aligned} \quad (3.4)$$

Assim, o método faz uso do gradiente da log-*posteriori* e, portanto, aponta para regiões de maior probabilidade, fazendo o algoritmo alcançar de forma mais rápida a distribuição de equilíbrio da cadeia de Markov.

Ao aplicar a etapa *leapfrog* um número  $L$  de vezes, a diferença entre os Hamiltonianos é próxima de zero devido a discretização do sistema de equações diferenciais e o estado resultante ao final da trajetória é definido por  $(\theta^*, p^*)$ . Após este procedimento aplica-se a regra de aceitação de Metropolis para eventuais correções do erro introduzido no sistema de equações. Desta forma, a probabilidade de manter a proposta é dada por

$$\alpha\left[(\theta^{(0)}, p^{(0)}), (\theta^*, p^*)\right] = \min\left\{\exp\left[H(\theta^{(0)}, p^{(0)}) - H(\theta^*, p^*)\right], 1\right\}. \quad (3.5)$$

A variável  $\varepsilon$  representa a discretização do sistema e assim como  $M$  é de livre escolha. Ambos determinam a rapidez com que a cadeia alcança a distribuição estacionária. Em geral,  $M$  é tomado como matriz identidade ou matrizes determinísticas positivas definidas, pois em casos mais gerais é de difícil especificação.

O algoritmo HMC, tomando  $M = I$ , por exemplo, pode ser descrito através dos seguintes passos,

1. Forneça a posição inicial  $\theta^{(0)}$ ;
2. Defina  $i = 1, \dots, S$

- Gere  $p \sim \mathcal{N}_d(0, I)$  e  $u \sim \mathcal{U}(0, 1)$ ;
  - Faça  $(\theta^I, p^I) = (\theta^{(i-1)}, p)$  e  $H^I = H(\theta^I, p^I)$ ;
  - Repita o número adequado da solução numérica *leapfrog* descrita em (3.4).
3. Ao final da trajetória, faça  $(\theta^F, p^F) = (\theta^*, p^*)$  e  $H^F = H(\theta^F, p^F)$ ;
  4. Calcule  $\alpha \left[ (\theta^F, p^F), (\theta^I, p^I) \right] = \min \left\{ \exp [H^I - H^F], 1 \right\}$
  5. Faça  $\theta^{(i)} = \theta^F$  com probabilidade  $u < \alpha(\cdot)$  e  $\theta^{(i)} = \theta^I$  caso contrário.

O número de *leapfrog* representa o quão longe se avança no tempo. Por exemplo, para valores pequenos, o algoritmo irá percorrer pequenas regiões do espaço paramétrico, já para valores grandes o algoritmo explora maiores regiões.

O HMC possui uma vantagem sobre os algoritmos tradicionais Metropolis-Hasting, pois evita o comportamento de passeio aleatório e obtém, assim, menor estrutura de correlação.

**Exemplo 1.** Considere que há interesse em se amostrar valores *a posteriori* dos parâmetros  $\mu$  e  $\sigma$  considerando um modelo normal. As Figuras 2 e 3 ilustram o comportamento de ambos os métodos.

Figura 2 – Curva de nível representando amostras estimadas de  $p(\mu, \sigma|X)$ , em que uma amostra de tamanho  $n = 20$  foi gerada de  $\mathcal{N}(X|\mu = 0, \sigma = 10)$ . Ambos os métodos MH e HMC foram simulados a partir de pontos iniciais  $\mu = 3$  e  $\sigma = 40$  para 200 passos (neste caso, HMC com  $\varepsilon = 0,3$  e  $L = 15$ ).

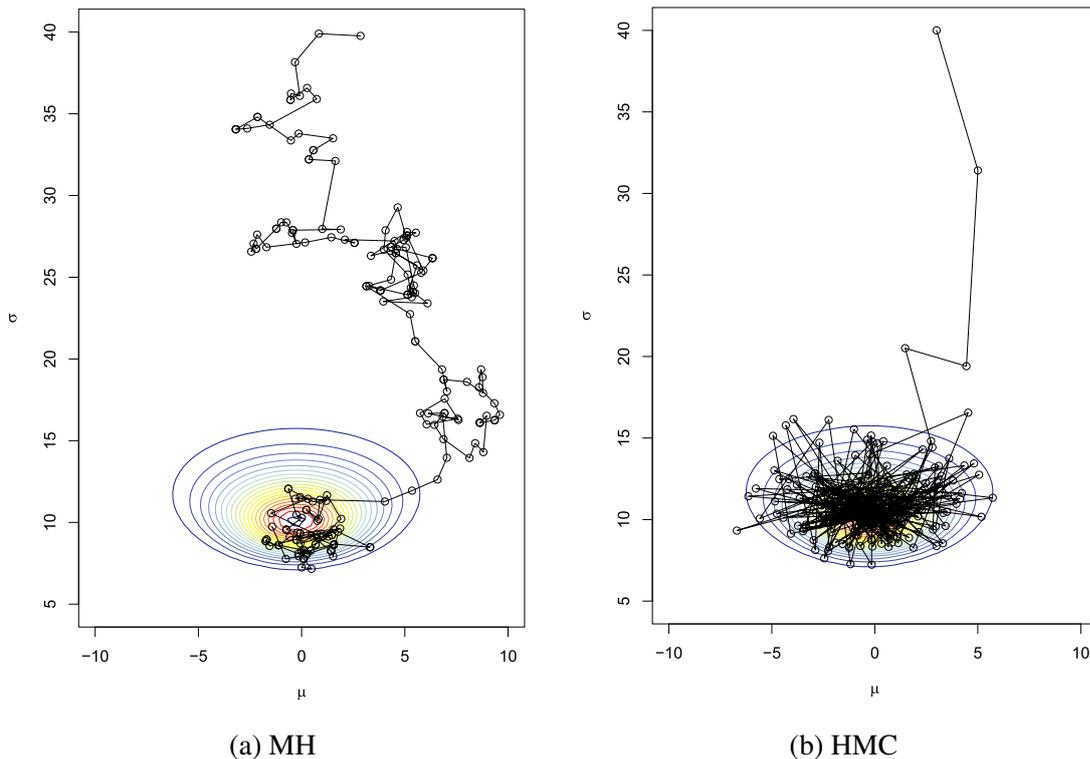
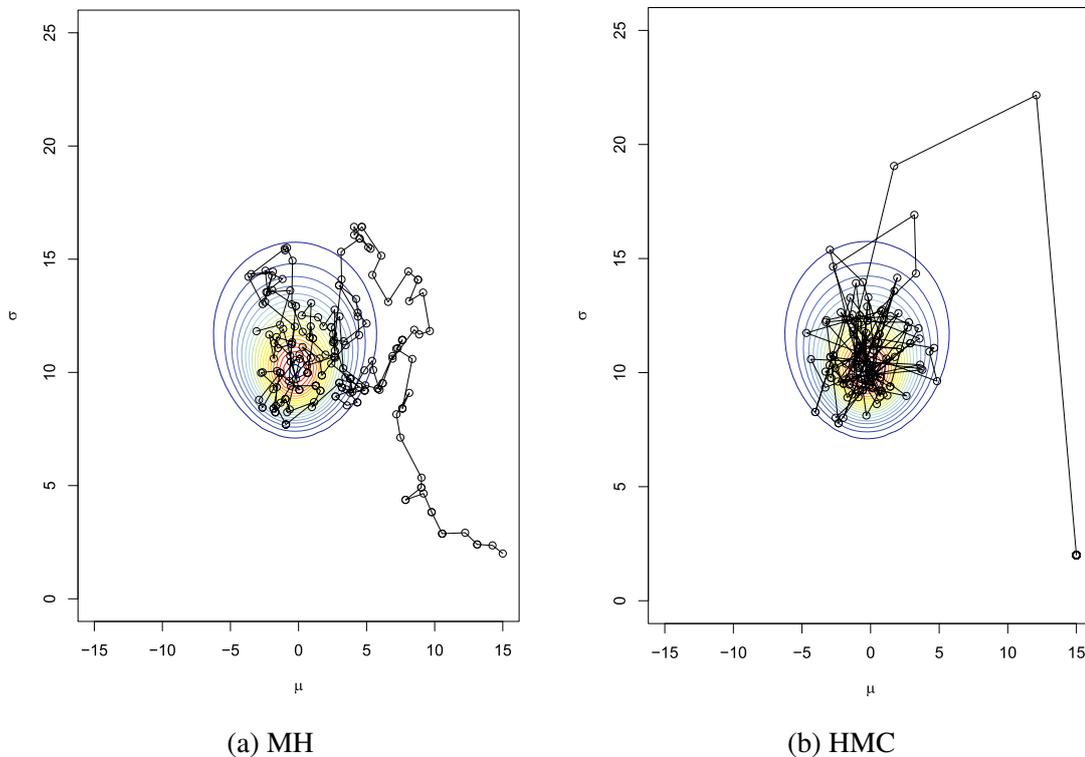


Figura 3 – Mesmo gráfico da Figura 2, mas a partir de pontos iniciais  $\mu = 15$  e  $\sigma = 2$  para 200 passos (neste caso, HMC com  $\varepsilon = 0,8$  e  $L = 10$ ).



Observa-se que o método de Monte Carlo Hamiltoniano converge de forma mais rápida em relação ao Metropolis-Hastings, amostrando valores dentro da região de interesse. Nota-se ainda que o MH apresenta um comportamento de passeio aleatório o que faz com que os valores simulados apresentem maior estrutura de correlação.

Vale ressaltar que a taxa de aceitação do Metropolis-Hastings varia em torno de 70%, enquanto esta mesma taxa para o método de Monte Carlo Hamiltoniano, dependendo da escolha de  $\varepsilon$  e  $L$ , varia entre 80 e 99%. No entanto, deve-se levar em consideração que o HMC utiliza mais passos por iteração durante o processo de simulação, o que pode tornar o procedimento de amostragem um pouco mais demorado em relação ao MH, dependendo da complexidade do modelo.

### 3.1.2 NUTS e Stan

Para usar HMC, é preciso escolher basicamente duas coisas: o número de passos *leapfrog* por iteração e o  $\varepsilon$  de cada etapa. Existem algumas estratégias para selecionar esses parâmetros e também, em vez de escolher um número global de passos *leapfrog*, há uma variação do HMC chamada NUTS (*No-U-Turn Sampler*) descrito em [Homan e Gelman \(2014\)](#), que se adapta automaticamente a este número de passos a cada iteração.

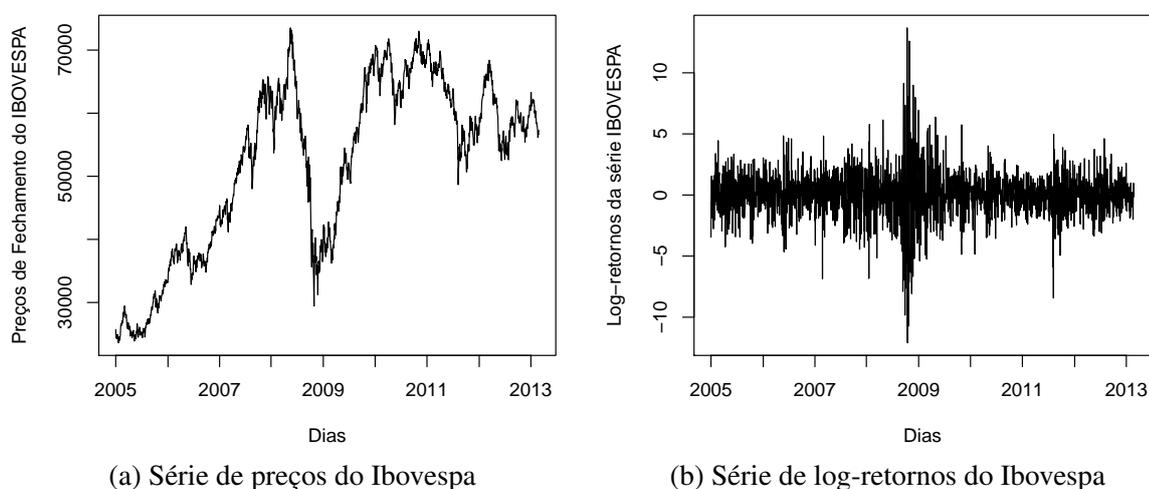
O Stan ([Stan Development Team and others, 2014](#)) é um *software* estatístico para

obtenção de amostras MCMC (e também para a otimização) usando o HMC e o algoritmo NUTS. Uma característica atraente do Stan é que ele pode automaticamente usar estratégias de adaptação para escolher parâmetros ideais. A implementação dos modelos foi realizada com o auxílio do pacote `rstan` (interface R (R Core Team, 2017) para Stan) que permite a comunicação entre as linguagens de forma eficiente.

Comparado a Gibbs ou Metropolis-Hastings, o Monte Carlo Hamiltoniano pode levar mais tempo por iteração (como normalmente possui muitos “passos de salto” dentro de cada iteração), mas as iterações tipicamente têm menor autocorrelação. Assim, Stan pode funcionar bem com 10.000 iterações em um exemplo onde BUGS exigiria 100.000 para uma boa mistura.

**Exemplo 2.** Para visualizar a performance do método HMC ao utilizar a ferramenta citada, será proposta a modelagem de uma série financeira. O objetivo desta análise é ajustar o modelo generalizado autorregressivo com heteroscedasticidade condicional (GARCH) a um conjunto de dados correspondentes a preços diários de fechamento do IBOVESPA de 3 de janeiro de 2005 a 28 de fevereiro de 2013.

Figura 4 – Gráficos da série financeira de preços diários de fechamento do IBOVESPA



(a) Série de preços do Ibovespa

(b) Série de log-retornos do Ibovespa

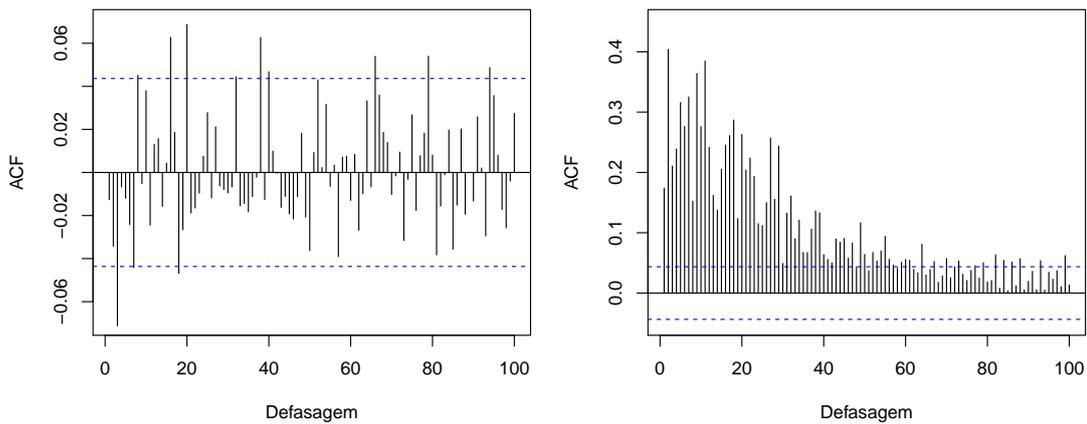
A Figura 4a apresenta o gráfico da série diária de preços de fechamento do IBOVESPA no período citado no enunciado do exemplo. Na Figura 4b é apresentada a série de log-retornos, onde é possível notar algumas propriedades como, por exemplo, estacionariedade. Observa-se, de acordo com a Tabela 1, que quase não há assimetria, mas existe uma curtose um tanto elevada comparada a da distribuição normal.

Tabela 1 – Estatísticas descritivas da série de retornos do IBOVESPA

T	Média	Desv. Padrão	Assimetria	Exc. Curtose
2016	0,00041	0,01876	-0,04264	5,88799

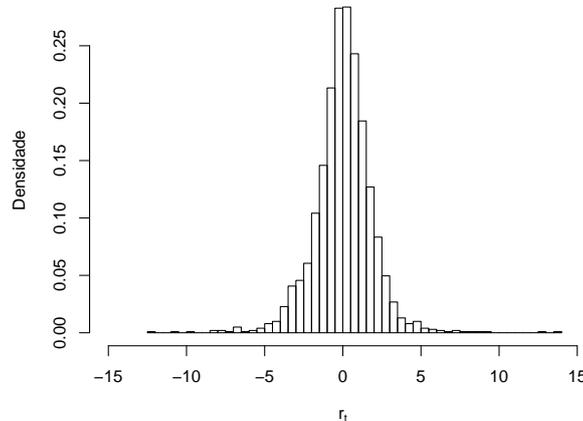
As Figuras 5a e 5b mostram os gráficos de autocorrelação dos retornos e retornos ao quadrado, respectivamente. É possível observar que quase não existe estrutura de autocorrelação no gráfico (a), enquanto o (b) mostra uma autocorrelação de fraca a moderada em muitas das defasagens. No histograma dos log-retornos, apresentado na Figura 5c mostra indícios de uma distribuição com caudas mais pesadas que a normal (curva em azul), o que já havia sido observado na Tabela 1 ao verificar a curtose.

Figura 5 – Gráficos da série financeira de preços diários de fechamento do IBOVESPA (Retornos).



(a) Correlograma da série  $r_t$ .

(b) Correlograma da série  $r_t^2$ .



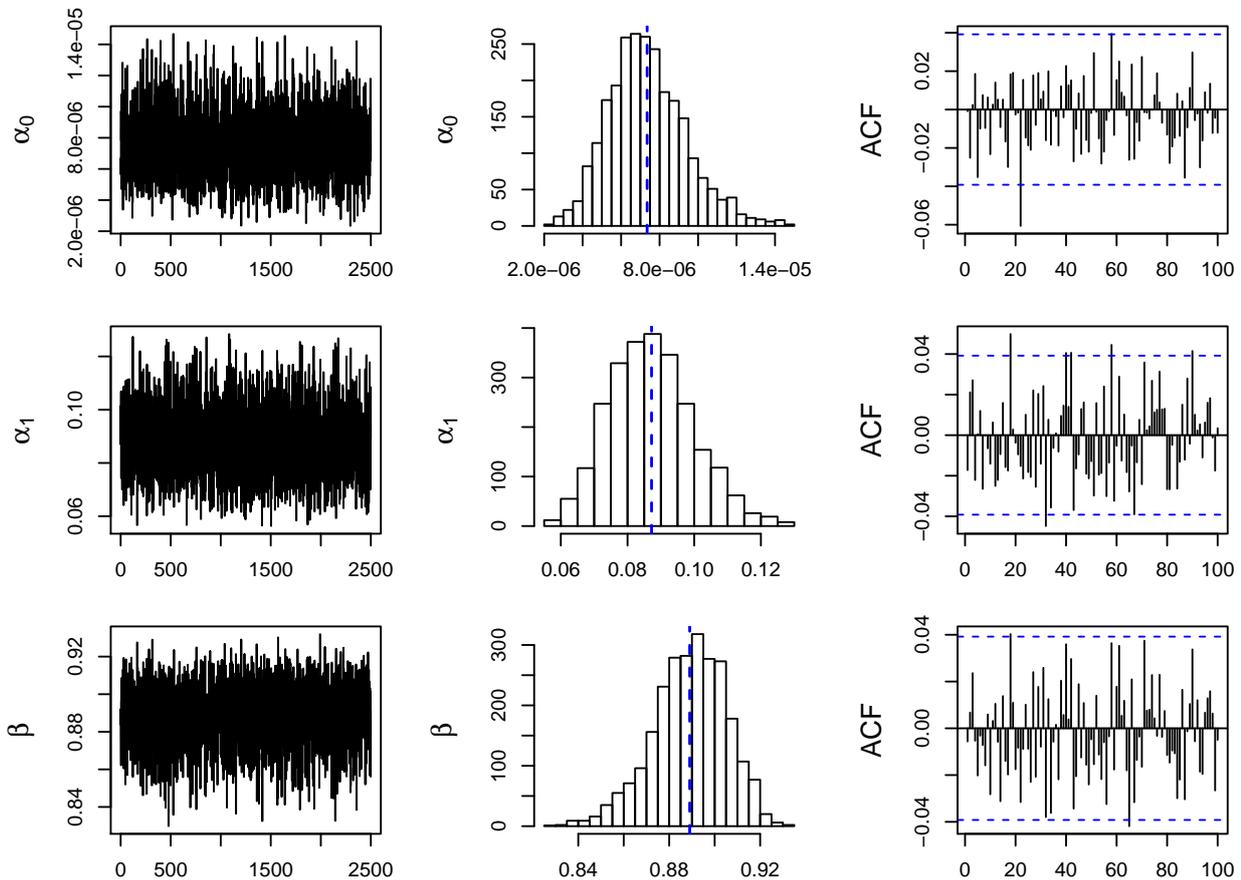
(c) Histograma de % dos log-retornos.

O modelo utilizado para capturar a dependência na série de retornos foi o GARCH(p,q) definido na equação (2.6). Neste caso, foi considerado  $p = q = 1$ , pois ao realizar estudos nesta série com este tipo de modelo, essa configuração, apresentou melhor performance em termos de ajuste. Associando ao que foi mostrado na seção 3.1.1, temos que  $\theta = (\alpha_0, \alpha_1, \beta)$ . Na especificação, usando a biblioteca `rstan`, foram consideradas distribuições a priori não informativas por questão de simplicidade e por ser padrão do pacote na construção do modelo.

Foram simuladas 5 mil valores das quais foram descartados 50% como *burn-in*. Os gráficos de convergência da cadeia referentes aos parâmetros de interesse são apresentados na Figura 6, em que mostra o gráfico de traços, autocorrelações e densidades a *posteriori* estimadas

marginais para as amostras dos parâmetros na série de retornos do IBOVESPA. É possível observar que mesmo sem realizar saltos as amostras simuladas apresentam baixa estrutura de correlação serial.

Figura 6 – Gráficos de convergência dos parâmetros na série de retornos do IBOVESPA.



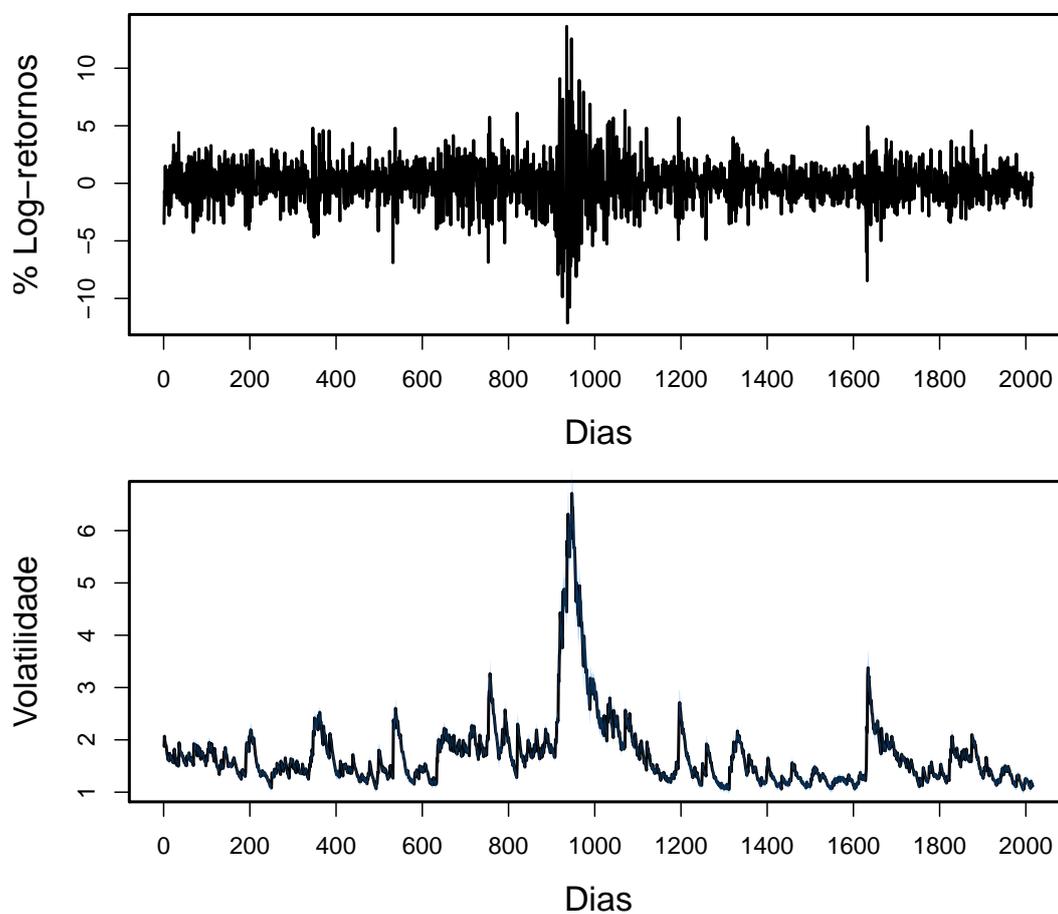
A equação de volatilidades do modelo GARCH(1,1) com as médias *a posteriori* dos parâmetros é dada por

$$\sigma_t^2 = 0,000007 + 0,087203r_{t-1}^2 + 0,889022\sigma_{t-1}^2,$$

Podemos notar que as volatilidades estimadas descrevem bem o comportamento da série de retornos do IBOVESPA. Vale ressaltar que seria razoável testar, também, o uso de outras distribuições com caudas mais pesadas ou até mesmo com assimetria para os erros.

A Figura 7 apresenta as volatilidades estimadas através do modelo em exemplo.

Figura 7 – Gráfico de log-retornos em percentual e volatilidades estimadas para série de retornos do IBOVESPA, usando método HMC.



No Capítulo 4 será apresentado um estudo de simulação dos modelos de volatilidade estocástica e aplicação a conjuntos de dados.



## ESTUDO DE SIMULAÇÃO E APLICAÇÕES

### 4.1 Estudo de Simulação 1

Este estudo de simulação foi realizado considerando os modelos de volatilidade estocástica. Para isso, foram geradas  $m = 100$  repetições de  $T \in \{500, 1000, 1500\}$  observações do modelo proposto, descrito na equação (2.8), considerando  $\beta = 1$  por questões de identificabilidade. Com isso, foram fixados os parâmetros da seguinte forma:  $\mu = -9$ ,  $\phi \in \{0,95; 0,99\}$  e  $\sigma_\eta \in \{0,05; 0,15\}$ . Para efeito de comparação utilizou-se uma biblioteca recentemente desenvolvida, específica para modelos de volatilidade estocástica, chamada `stochvol` que foi proposta por [Kastner \(2016\)](#). Uma das características fundamentais do algoritmo usado nesta biblioteca é a amostragem em todas as volatilidades instantâneas *all without a loop* (AWOL), uma técnica discutida com maiores detalhes em [McCausland, Miller e Pelletier \(2011\)](#) e [Kastner e Frühwirth-Schnatter \(2014\)](#). Além disso, para evitar o custo da interpretação do código dentro de cada iteração do MCMC, os cálculos do núcleo são implementados em linguagem C, o que a torna bastante eficiente.

Foram assumidas as distribuições *a priori* (2.19) para o parâmetro  $\phi$ ,  $\mu \sim \mathcal{N}(-10, 1)$  e (2.21) para  $\sigma_\eta$  com  $B_\sigma = 0, 1$ , conforme utilizado em [Kastner \(2016\)](#), em ambos os métodos. Sendo assim, os parâmetros de entrada de ambas as bibliotecas foram os mesmos. Para cada série foram simuladas 10.000 amostras MCMC, em que as 5.000 primeiras foram descartadas como *burn-in*. A performance dos métodos de estimação foi avaliada considerando dois critérios conforme utilizado por [Zevallos, Gasco e Ehlers \(2016\)](#) que são: O viés e a raiz quadrada do erro quadrático médio (*smse*), definidos por

$$\text{viés} = \frac{1}{m} \sum_{i=1}^m \hat{\theta}^{(i)} - \theta, \quad \text{smse}^2 = \frac{1}{m} \sum_{i=1}^m (\hat{\theta}^{(i)} - \theta)^2 \quad (4.1)$$

em que  $\hat{\theta}^{(i)}$  é a média *a posteriori* estimada na  $i$ -ésima replicação,  $i = 1, \dots, m$ .

Na Tabela 2, para  $\sigma_\eta = 0.15$ , obtivemos bons resultados em termos de viés e *smse* em

Tabela 2 – Viés e raiz quadrada do erro quadrático médio (*smse*) das médias a posteriori. Parâmetros:  $\mu = -9$ ,  $\phi = 0,99$  e  $\sigma_\eta = 0,15$ 

T	Método	$\mu$		$\phi$		$\sigma_\eta$	
		viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>
500	Stochvol	0,213	0,474	0,020	0,030	-0,137	0,121
	HMC	0,491	0,491	0,068	0,066	-0,117	0,118
1000	Stochvol	0,223	0,415	0,007	0,011	-0,019	0,031
	HMC	0,055	0,056	0,002	0,002	-0,006	0,007
1500	Stochvol	0,129	0,366	0,004	0,007	-0,012	0,024
	HMC	0,120	0,267	0,003	0,007	-0,007	0,021

todos os parâmetros com relativa superioridade do método HMC quando o tamanho da série aumenta. Em contra partida, nesta configuração, quando o tamanho da série é menor,  $n = 500$ , os ajustes realizados foram próximos, no entanto, com uma certa vantagem do *stochvol*.

Tabela 3 – Viés e raiz quadrada do erro quadrático médio (*smse*) das médias a posteriori. Parâmetros:  $\mu = -9$ ,  $\phi = 0,99$  e  $\sigma_\eta = 0,05$ 

T	Método	$\mu$		$\phi$		$\sigma_\eta$	
		viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>
500	Stochvol	0,063	0,218	0,106	0,121	-0,071	0,082
	HMC	0,090	0,093	0,132	0,133	0,023	0,029
1000	Stochvol	0,041	0,146	0,042	0,064	-0,039	0,049
	HMC	0,018	0,134	-0,001	0,002	-0,010	0,012
1500	Stochvol	0,017	0,122	0,027	0,047	-0,027	0,037
	HMC	0,017	0,123	0,018	0,034	-0,015	0,026

De acordo com a Tabela 3, observamos que, para a configuração da série com o parâmetro  $\sigma_\eta = 0,05$ , os valores de viés e *smse* foram bem menores em ambos os algoritmos. Mesmo assim, obteve-se melhores resultados considerando o método HMC na maioria dos parâmetros.

Tabela 4 – Viés e raiz quadrada do erro quadrático médio (*smse*) das médias a posteriori. Parâmetros:  $\mu = -9$ ,  $\phi = 0,95$  e  $\sigma_\eta = 0,15$ 

T	Método	$\mu$		$\phi$		$\sigma_\eta$	
		viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>
500	Stochvol	0,049	0,151	0,050	0,066	-0,042	0,059
	HMC	0,042	0,150	0,030	0,049	0,000	0,051
1000	Stochvol	0,022	0,107	0,025	0,045	-0,022	0,044
	HMC	0,019	0,106	0,014	0,035	-0,003	0,038
1500	Stochvol	0,027	0,083	0,018	0,030	-0,023	0,042
	HMC	0,024	0,082	0,011	0,024	-0,011	0,035

Tabela 5 – Viés e raiz quadrada do erro quadrático médio (*smse*) das médias a posteriori. Parâmetros:  $\mu = -9$ ,  $\phi = 0,95$  e  $\sigma_\eta = 0,05$

T	Método	$\mu$		$\phi$		$\sigma_\eta$	
		viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>
500	Stochvol	0,025	0,079	0,119	0,124	-0,050	0,060
	HMC	0,011	0,076	0,086	0,091	0,009	0,037
1000	Stochvol	0,008	0,059	0,118	0,123	-0,046	0,060
	HMC	-0,001	0,057	0,087	0,092	0,001	0,043
1500	Stochvol	0,005	0,042	0,116	0,122	-0,037	0,049
	HMC	-0,001	0,044	0,087	0,097	-0,006	0,037

Já nas Tabelas 4 e 5 notamos que os resultados das simulações apresentam-se com melhor qualidade para o método HMC, evidenciado pelo viés e raiz quadrada do erro quadrático médio, em quase todos os casos. Vale ressaltar que, baseado na diminuição dos valores das estatísticas de ajuste, houve uma melhora nas estimativas em ambos os métodos, considerando estas configurações. Porém, as cadeias obtidas através do método HMC (não mostradas aqui) apresentam baixa estrutura de autocorrelação serial e tendem a alcançar a distribuição estacionária de interesse de forma rápida, fornecendo melhores estimativas para os parâmetros.

## 4.2 Estudo de Simulação 2

Este estudo de simulação foi feito considerando todas as distribuições propostas para os erros da equação de observação do modelo de volatilidade estocástica. Neste caso, foram geradas  $m = 100$  repetições de  $T = 1000$  observações do modelo proposto, descrito na equação (2.8), considerando  $\beta = 1$  por questões de identificabilidade. Por conseguinte, os parâmetros do modelo foram fixados da seguinte forma:  $\mu = -10$ ,  $\phi \in \{0,95; 0,99\}$ ,  $\sigma_\eta \in \{0,05; 0,15\}$ ,  $\nu = 5$  (t-Student),  $\nu = 0,8$  (Skew-Normal) e  $\nu = 1,5$  (GED).

Foram assumidas as distribuições a priori do Estudo de Simulação 1. Para cada série foram simuladas 10.000 amostras do método HMC e descartou-se 50% dessas como *burn-in*, resultando em 5.000 amostras efetivas. As medidas usadas para avaliar a performance foram o viés e a raiz quadrada do erro quadrático médio (*smse*), definidos de acordo com a equação (4.1).

As Tabelas 6 e 7 mostram os resultados obtidos com as simulações realizadas, considerando  $\phi = 0,99$  e  $\sigma_\eta \in \{0,05; 0,15\}$ . Observa-se que o viés e a raiz quadrada do erro quadrático médio apresentam valores baixos, indicando que as estimativas obtidas para a média a posteriori dos parâmetros está próxima dos valores pré-fixados na simulação das séries de retorno. Nota-se que o parâmetro  $\nu$  para os erros t-Student apresentam maiores valores para o viés e *smse* ao considerar essas configurações.

Nas Tabelas 8 e 9 são apresentados os estudos de simulação realizados, considerando

$\phi = 0,95$  e  $\sigma_\eta \in \{0,05; 0,15\}$ .

Tabela 6 – Viés e *smse* das médias a posteriori, método HMC. Parâmetros:  $\mu = -10$ ,  $\nu = 5$  (t-Student),  $\nu = 0,8$  (Skew-Normal),  $\nu = 1,5$  (GED),  $\phi = 0,99$  e  $\sigma_\eta = 0,05$ .

Distribuição	$\mu$		$\phi$		$\sigma_\eta$		$\nu$	
	viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>
(1) Gaussiana	-0,007	0,150	0,059	0,090	-0,052	0,073		
(2) t-Student	-0,025	0,167	0,051	0,071	-0,033	0,056	-0,409	1,058
(3) Skew-Normal	-0,028	0,146	0,037	0,058	-0,015	0,033	-0,007	0,055
(4) GED	-0,016	0,084	0,024	0,037	-0,014	0,018	-0,018	0,068

Tabela 7 – Viés e *smse* das médias a posteriori, método HMC. Parâmetros:  $\mu = -10$ ,  $\nu = 5$  (t-Student),  $\nu = 0,8$  (Skew-Normal),  $\nu = 1,5$  (GED),  $\phi = 0,99$  e  $\sigma_\eta = 0,15$ .

Distribuição	$\mu$		$\phi$		$\sigma_\eta$		$\nu$	
	viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>
(1) Gaussiana	0,042	0,379	0,006	0,010	-0,010	0,033		
(2) t-Student	-0,058	0,412	0,006	0,011	-0,010	0,031	-0,412	0,936
(3) Skew-Normal	0,019	0,381	0,005	0,010	-0,005	0,022	-0,002	0,065
(4) GED	-0,149	0,331	0,010	0,011	-0,021	0,037	0,015	0,057

Tabela 8 – Viés e *smse* das médias a posteriori, método HMC. Parâmetros:  $\mu = -10$ ,  $\nu = 5$  (t-Student),  $\nu = 0,8$  (Skew-Normal),  $\nu = 1,5$  (GED),  $\phi = 0,95$  e  $\sigma_\eta = 0,05$ .

Distribuição	$\mu$		$\phi$		$\sigma_\eta$		$\nu$	
	viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>
(1) Gaussiana	-0,005	0,057	0,085	0,092	0,000	0,035		
(2) t-Student	-0,013	0,066	0,095	0,100	0,002	0,037	-0,385	0,871
(3) Skew-Normal	0,003	0,060	0,087	0,096	0,001	0,038	0,001	0,057
(4) GED	-0,010	0,071	0,087	0,092	0,008	0,031	0,052	0,106

Tabela 9 – Viés e *smse* das médias a posteriori, método HMC. Parâmetros:  $\mu = -10$ ,  $\nu = 5$  (t-Student),  $\nu = 0,8$  (Skew-Normal),  $\nu = 1,5$  (GED),  $\phi = 0,95$  e  $\sigma_\eta = 0,15$ .

Distribuição	$\mu$		$\phi$		$\sigma_\eta$		$\nu$	
	viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>	viés	<i>smse</i>
(1) Gaussiana	-0,021	0,118	0,011	0,028	-0,007	0,041		
(2) t-Student	0,013	0,120	0,025	0,047	-0,014	0,066	-0,569	1,098
(3) Skew-Normal	0,001	0,101	0,014	0,031	-0,007	0,037	0,008	0,057
(4) GED	0,001	0,112	0,017	0,036	0,008	0,051	0,010	0,136

Nota-se que o viés e a raiz quadrada do erro quadrático médio (*smse*) apresentam valores ainda menores em relação a configuração anterior. No entanto, o parâmetro  $\nu$ , para os erros t-Student, os valores ainda continuam relativamente mais elevados quando comparados aos demais parâmetros. Vale ressaltar que nesta configuração para os valores obtidos de *smse* para os parâmetros entre as distribuições  $\phi$  e  $\sigma_\eta$ , nas Tabelas 8 e 9 são similares. A Tabela 10 mostra os

Tabela 10 – Tempos médios de simulação para cada uma das distribuições.

Distribuição	Segundos	Minutos
(1) Gaussiana	99,89	1,66
(2) t-Student	953,50	15,89
(3) Skew-Normal	446,16	7,44
(4) GED	2889,87	48,16

tempos médios de estimação dos modelos considerando cada uma das distribuições, levando em conta as configurações adotadas. Nota-se que ao utilizar a distribuição GED o tempo médio para se obter os valores simulados é bem maior em relação as demais distribuições. Cabe destacar que a distribuição GED não está implementada no pacote *rstan* e talvez isso seja um dos fatores que elevam o período para gerar as distribuições *a posteriori* simuladas.

## 4.3 Aplicação a Conjuntos de Dados

### 4.3.1 Descrição dos Dados e Análise Descritiva

A ilustração é realizada utilizando quatro conjuntos de dados de séries financeiras. Os retornos diários em percentuais da série £/USD cobrem o período de 01/10/81 a 28/06/85 e já foi utilizada por outros autores (por exemplo, (HARVEY; RUIZ; SHEPHARD, 1994) e (ZEVALLOS; GASCO; EHLERS, 2016)), da série EUR/USD<sup>1</sup> é referente ao período de 03/01/2004 a 04/04/2012, já da séries do IBOVESPA e S&P500<sup>2</sup> cobrem, respectivamente, os períodos de 03/01/2005 a 28/02/2013 e de 03/01/2012 a 31/07/2017.

Tabela 11 – Sumário das séries em estudo

Série	T	Média	Desvio Padrão	Assimetria	Exc. Curtose
£/USD	945	-0,03531	0,71108	0,60321	4,84528
EUR/USD	2120	0,00002	0,00660	-0,15238	3,17629
IBOVESPA	2016	0,00041	0,01876	-0,04264	5,88799
S&P500	1402	0,00047	0,00778	-0,31108	2,27725

Observamos, de acordo com a Tabela 11, que a curtose amostral é um pouco elevada em todas as séries, o que ressalta o uso outras distribuições além da normal para a equação de observações (2.8) do modelo. É possível notar uma leve assimetria positiva na série £/USD e negativa nas séries EUR/USD e S&P500.

<sup>1</sup> A série EUR/USD está disponível na biblioteca `stochvol`

<sup>2</sup> As séries IBOVESPA e S&P500 estão disponíveis através do site: <https://br.investing.com>

Figura 8 – Gráficos das taxas de retornos diários em percentagens da série £/USD.

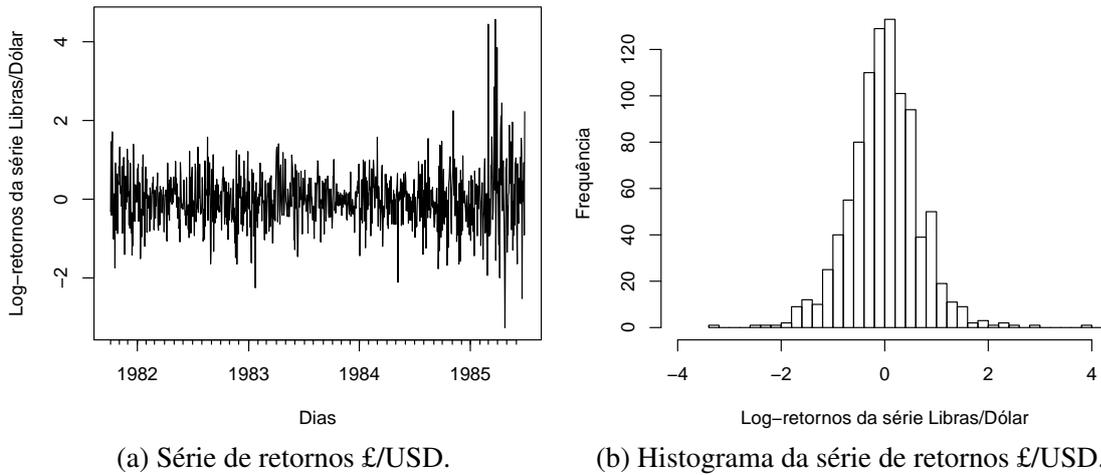


Figura 9 – Gráficos das taxas de retornos diários em percentagens da série EUR/USD.

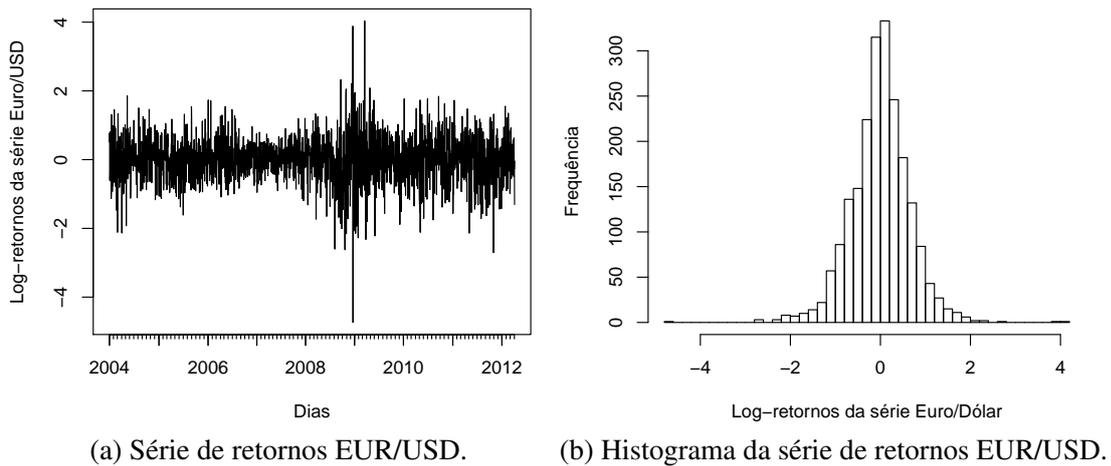
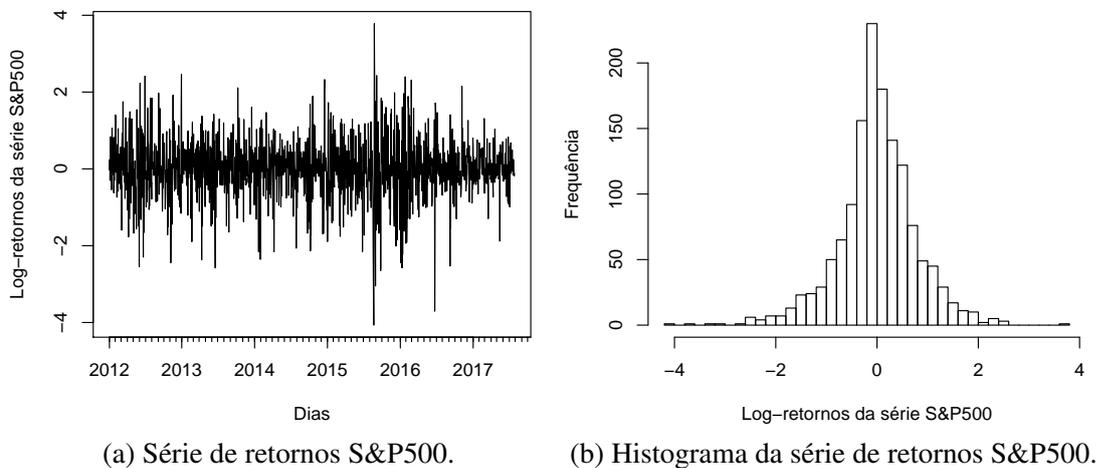
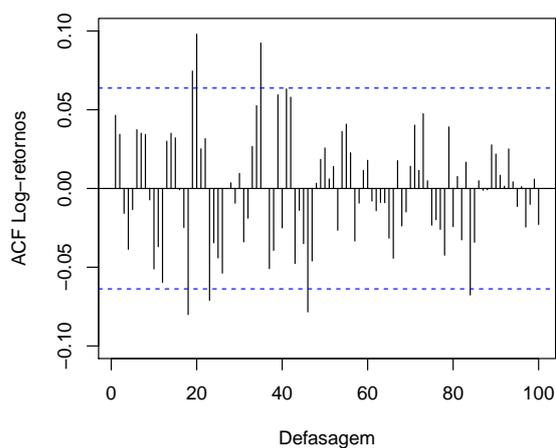


Figura 10 – Gráficos das taxas de retornos diários em percentagens da série S&amp;P500.

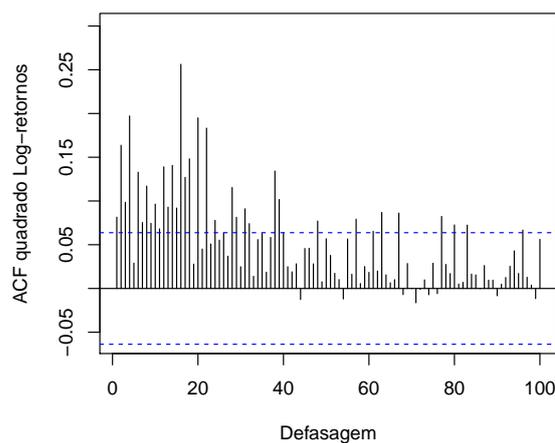


Nas Figuras 8, 9 e 10 é possível observar, analisando os histogramas, que os dados se distribuem em uma amplitude pouco maior que -4 e 4 dando indícios de uma distribuição com caudas mais pesadas que a Gaussiana. Nota-se ainda períodos de alta volatilidade no fim de 2008 a 2009 na série EUR/USD e na série S&P500 no período de Agosto de 2015 a Março de 2016.

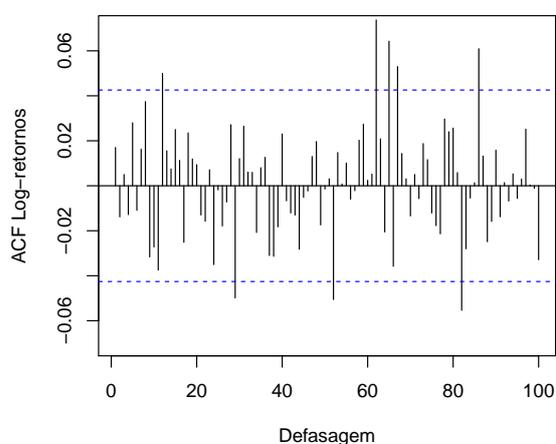
Figura 11 – Correlogramas das séries de retornos £/USD, EUR/USD e S&P500.



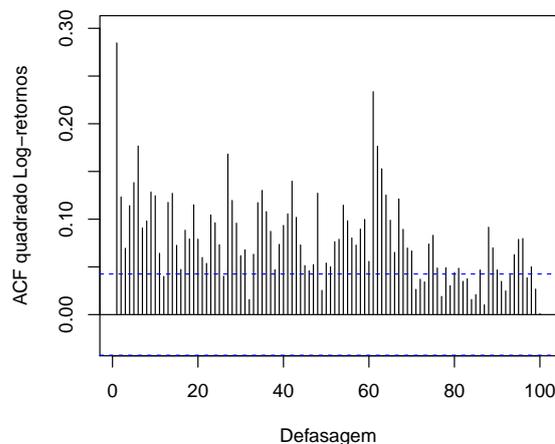
(a) ACF dos retornos £/USD.



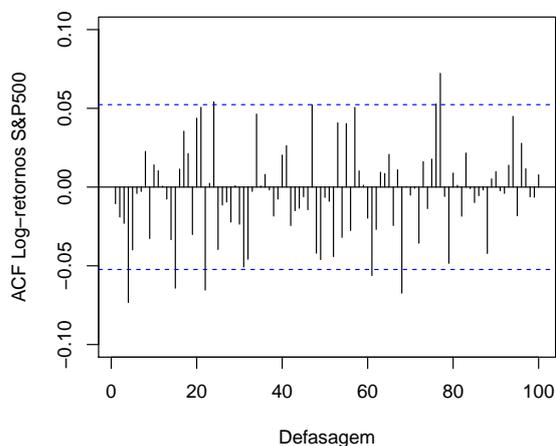
(b) ACF do quadrado dos retornos £/USD.



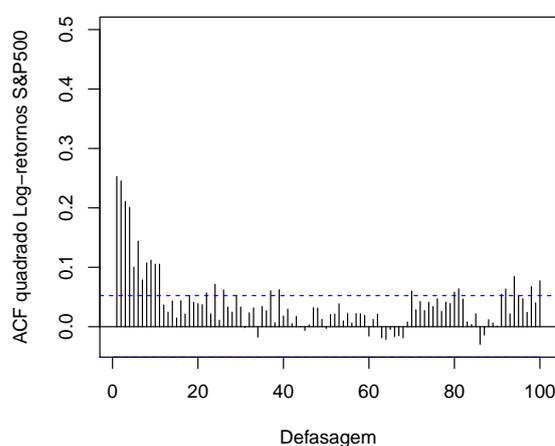
(c) ACF dos retornos EUR/USD.



(d) ACF do quadrado dos retornos EUR/USD.



(e) ACF dos retornos S&P500.



(f) ACF do quadrado dos retornos S&P500.

Os correlogramas das séries de retornos (gráficos 11a, 11c e 11e) apresentam baixa estrutura de correlação serial. Por outro lado, os quadrados dos retornos (gráficos 11b, 11d e 11f) são claramente autocorrelacionados, o que é característico em retorno de séries temporais financeiras, conforme visto no Capítulo 1.

### 4.3.2 Resultados

No modelo de volatilidade estocástica estimado, utilizamos a série de retornos depreciada em termos da média e consideramos as quatro distribuições para os erros, descritas no Capítulo 2, da equação de observações, pois percebemos que as séries em estudo apresentam uma curtose elevada indicando uma distribuição leptocúrtica nos retornos. Consideramos as distribuições a priori (2.19) para o parâmetro  $\phi$ ,  $\mu \sim \mathcal{N}(-10, 1)$  e (2.21) para  $\sigma_\eta$  com hiperparâmetro  $B_\sigma = 0, 1$ . Já para o parâmetro  $\nu$  nos casos GED, t-Student e Skew-Normal as distribuições a priori utilizadas são descritas, respectivamente, nas equações (2.23), (2.24) e (2.25).

Para cada série foram simuladas 10000 amostras MCMC para os parâmetros e volatilidades, em que as primeiras 5000 foram descartadas como *burn-in*. Para selecionar um modelo mais adequado entre os propostos, foram utilizados os critérios descritos na seção (2.3).

Tabela 12 – Comparação entre os modelos propostos via critérios de seleção.

Série	Dist.	DIC	WAIC	EP <sub>waic</sub>	LOO	EP <sub>loo</sub>
£/USD	(1) Gaussiana	1802,7	1807,4	53,5	1811,8	54,3
	(2) t-Student	1805,3	1812,8	52,8	1813,5	52,9
	(3) Skew-Normal	1803,1	1807,4	53,3	1812,2	54,2
	(4) GED	1810,6	1811,3	53,2	1814,0	53,7
EUR/USD	(1) Gaussiana	-15647,9	-15640,2	77,0	-15638,2	77,3
	(2) t-Student	-15664,9	-15647,8	76,4	-15647,4	76,4
	(3) Skew-Normal	-15646,7	-15638,9	77,2	-15636,6	77,5
	(4) GED	-15651,2	-15650,9	76,6	-15650,2	76,7
IBOVESPA	(1) Gaussiana	-10986,7	-10983,7	74,9	-10975,9	75,6
	(2) t-Student	-10974,6	-10970,3	75,7	-10968,0	75,9
	(3) Skew-Normal	-10984,3	-10981,0	75,1	-10973,7	75,8
	(4) GED	-10977,3	-10979,1	75,4	-10974,4	75,8
S&P500	(1) Gaussiana	-10016,2	-10014,7	65,3	- 9981,1	67,8
	(2) t-Student	- 9986,1	- 9979,0	65,3	- 9970,5	68,7
	(3) Skew-Normal	-10016,1	-10014,8	65,4	- 9982,2	67,8
	(4) GED	- 9985,8	- 9990,6	67,5	- 9976,6	68,6

De acordo com a Tabela 12, vimos que na a série £/USD os critérios de seleção apresentaram valores próximos para as distribuição Gaussiana e Skew-Normal. Verificamos ainda que para a distribuição Normal as estimativas dos critérios foram um pouco menores. Para a série EUR/USD observamos que, de acordo com os critérios de seleção, o modelo de Volatilidade Estocástica com as distribuições t-Student e GED se destacaram. Notamos que o valor do critério

DIC na distribuição t-Student foi menor, no entanto, os critérios WAIC e LOO retornaram valores menores para a distribuição GED. Já para as séries do IBOVESPA e S&P500 os critérios indicam que o modelo com erros Gaussianos e Skew-Normal apresentam melhores resultados.

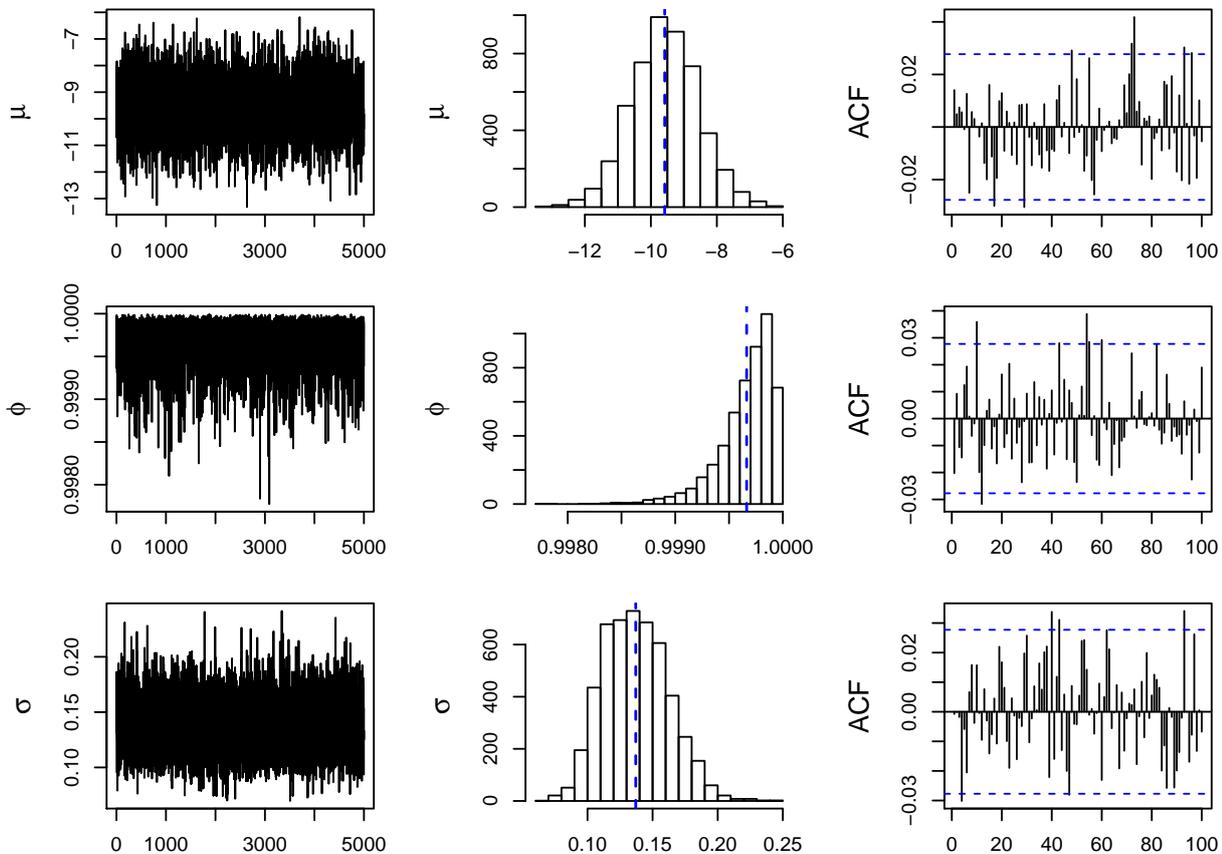
Tabela 13 – Estimativas das médias (desvios padrões) dos parâmetros *a posteriori*.

Série	Dist.	$\mu$	$\phi$	$\sigma_{\eta}$	$\nu$
£/USD	(1)	-9,5805 (1,0149)	0,9997 (0,0003)	0,1370 (0,0249)	
	(2)	-9,5912 (1,0460)	0,9997 (0,0002)	0,1132 (0,0248)	12,1437 (3,5816)
	(3)	-9,6067 (1,0266)	0,9996 (0,0003)	0,1416 (0,0262)	-0,0386 (0,0426)
	(4)	-9,5928 (1,0432)	0,9997 (0,0002)	0,1238 (0,0254)	1,7630 (0,1415)
EUR/USD	(1)	-10,1972 (0,2882)	0,9940 (0,0029)	0,0654 (0,0105)	
	(2)	-10,3211 (0,2892)	0,9945 (0,0027)	0,0598 (0,0103)	14,0905 (3,3934)
	(3)	-10,1990 (0,2907)	0,9940 (0,0029)	0,0652 (0,0107)	0,0166 (0,0275)
	(4)	-10,1758 (0,3128)	0,9953 (0,0024)	0,0555 (0,0101)	1,6814 (0,0820)
IBOVESPA	(1)	- 8,4164 (0,1999)	0,9805 (0,0065)	0,1487 (0,0193)	
	(2)	- 8,5462 (0,2364)	0,9848 (0,0056)	0,1281 (0,0173)	16,2022 (4,2292)
	(3)	- 8,4169 (0,2113)	0,9811 (0,0065)	0,1457 (0,0202)	0,0209 (0,0284)
	(4)	- 8,4176 (0,2167)	0,9837 (0,0058)	0,1325 (0,0186)	1,8019 (0,1075)
S&P500	(1)	-10,1100 (0,1206)	0,9037 (0,0223)	0,3844 (0,0455)	
	(2)	-10,2592 (0,1417)	0,9326 (0,0202)	0,3030 (0,0491)	11,0003 (3,3114)
	(3)	-10,1117 (0,1198)	0,9056 (0,0221)	0,3806 (0,0478)	0,0365 (0,0351)
	(4)	-10,0525 (0,1314)	0,9269 (0,0216)	0,3126 (0,0521)	1,5898 (0,1271)

As estimativas obtidas para os parâmetros para as quatro séries estão apresentadas na Tabela 13. Podemos ver que o desvio padrão para o parâmetro  $\mu$  nas séries EUR/USD, IBOVESPA e S&P500 são bem menores em relação aos da série £/USD. O parâmetro de persistência  $\phi$  é bastante elevado para as três primeiras séries, indicando que as volatilidades observadas no dia anterior tendem a impactar a volatilidade dos preços, das taxas de câmbio £/USD, EUR/USD e na bolsa de valores do IBOVESPA, no dia posterior. É possível perceber que o parâmetro de forma  $\nu$  da distribuição Skew-Normal (3), nas séries de retornos, foi próximo de zero e seus intervalos de credibilidade foram respectivamente, (-0,1232; 0,0454), (-0,0365; 0,0717), (-0,0334; 0,0765) e (-0,0314; 0,1060), indicando que há cerca de 95% de probabilidade de conter o zero. No entanto, para as séries £/USD, IBOVESPA e S&P500 o modelo apresentou bons resultados, de acordo com os critérios de seleção, pois este segue bem os movimentos das log-volatilidades. É possível observar que a média *a posteriori* para os parâmetros  $\phi$  e/ou  $\sigma_{\eta}$ , nos modelos que apresentam menores valores nos critérios de seleção, tendem a serem próximos e um pouco diferentes em relação a média observada para outros modelos.

A partir disso, analisamos as convergências dos parâmetros, graficamente, dos modelos de volatilidade estocástica com os erros que obtiveram os melhores ajustes, de acordo com os critérios de seleção. Portanto, para a série £/USD, IBOVESPA e S&P500 consideramos as distribuições Gaussianas e Skew-Normal. Para a série EUR/USD o modelo com erros t-Student e GED.

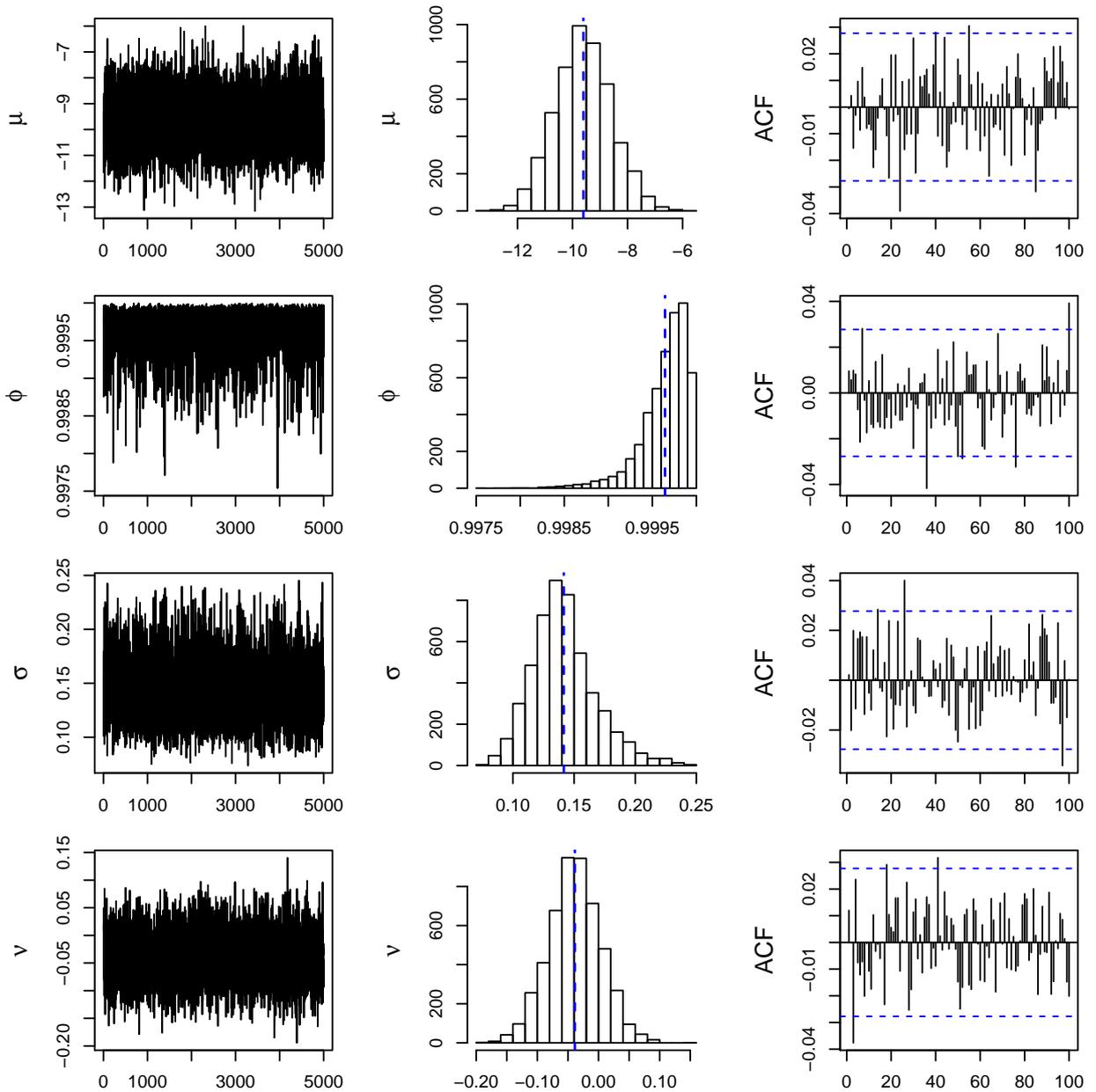
Figura 12 – Gráficos de convergência dos parâmetros na série de retornos £/USD com erros Gaussianos.



A Figura 12 mostra o gráfico de traços, autocorrelações e histogramas das distribuições *a posteriori* marginais estimadas para os parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta$  para a série £/USD usando o esquema de amostragem HMC com erros normais para a equação de observações do modelo. É possível observar que mesmo sem realizar saltos as amostras não apresentam estrutura de correlação serial. Na Figura 13 são mostrados os gráficos de convergência dos parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta$  para a série £/USD usando o esquema de amostragem HMC sob erros com distribuição Skew-Normal. Notamos também, que houve uma boa convergência da série mesmo sem realizar saltos entre as simulações.

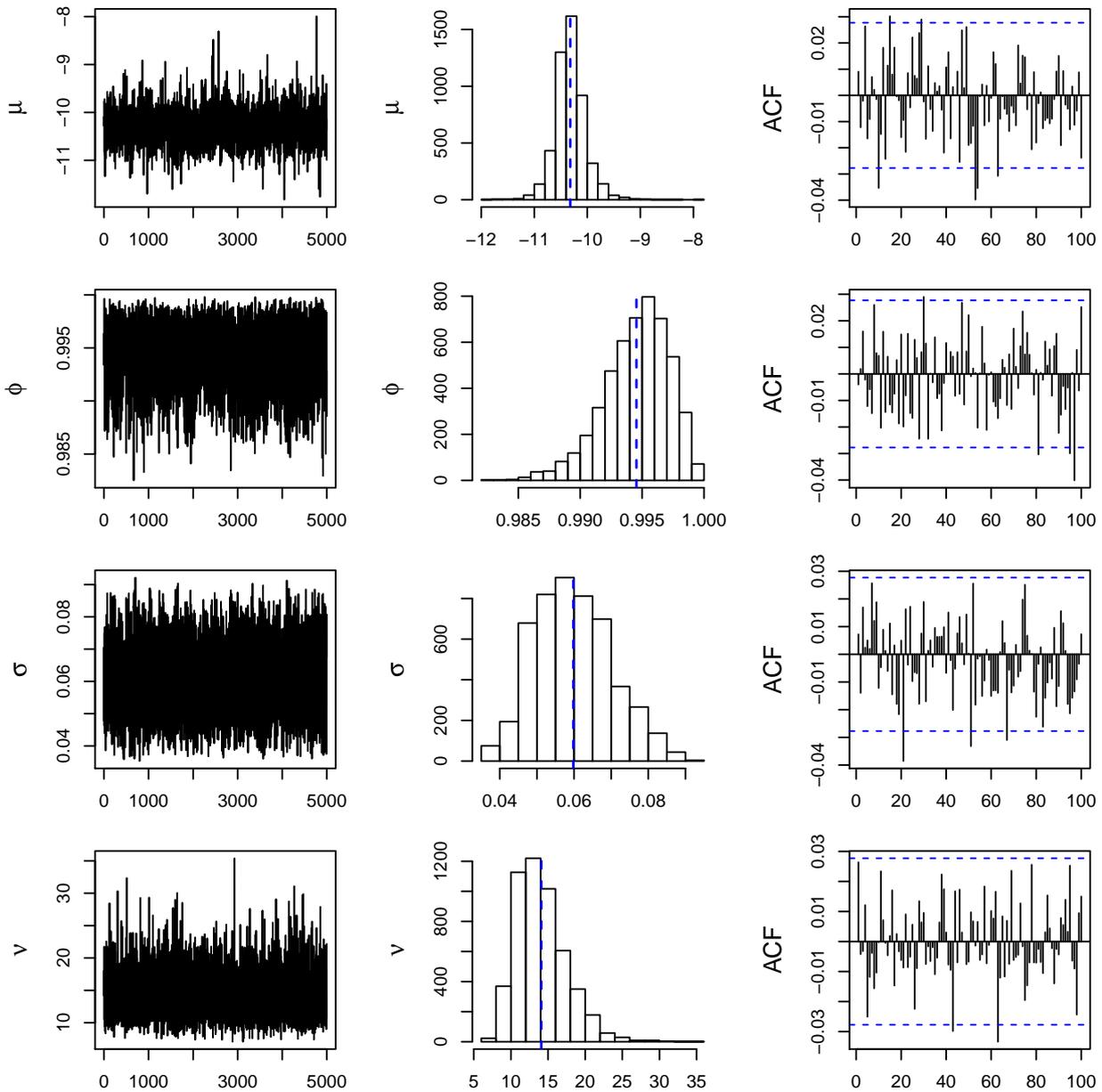
Nas Figuras 14 e 15 são apresentados os gráficos de traços, histogramas e correlogramas (sem o primeira autocorrelação, que vale um) para os parâmetros do modelo de volatilidade, na série EUR/USD, com erros t-Student e GED, respectivamente. Observamos que o método HMC apresentou uma boa convergência para o modelo considerando ambas as distribuições. O gráfico de traços do parâmetro  $\mu$  apresenta uma amplitude pequena e o histograma mostra que os valores simulados estão mais agrupados em torno da média *a posteriori* do que na série £/USD. O histograma para o parâmetro  $\sigma_\eta$  na distribuição GED (Figura 15) apresenta um pico próximo ao valor da média (linha azul). No entanto, o critério  $\hat{R}$  de Gelman e Rubin (1992) foi de aproximadamente 1 em todos os parâmetros, indicando uma boa convergência da cadeia.

Figura 13 – Gráficos de convergência dos parâmetros na série de retornos £/USD com erros Skew-Normal.



As Figuras 16 e 17 apresentam os gráficos para verificação de convergência das cadeias, para os parâmetros do modelo de Volatilidade Estocástica com erros Gaussianos e Skew-Normais, na série do IBOVESPA. Novamente nota-se uma baixa estrutura de correlação e boa convergência em ambos os modelos. Observamos ainda, que os valores das médias *a posteriori* marginais dos parâmetros são próximos e o valor médio do parâmetro de assimetria da distribuição Skew-Normal é perto de zero, dando indícios de que os modelos são de certa forma parecidos. Os valores do critério  $\hat{R}$  de Gelman e Rubin (1992) (não mostrados aqui) foram aproximadamente 1 em todos os parâmetros em ambos os casos, o que indica uma boa convergência das cadeias simuladas. Vale lembrar que no caso do algoritmo HMC não há necessidade da realização de saltos nos valores das cadeias, resultando assim, após o *burn-in*, em 5 mil simulações efetivas.

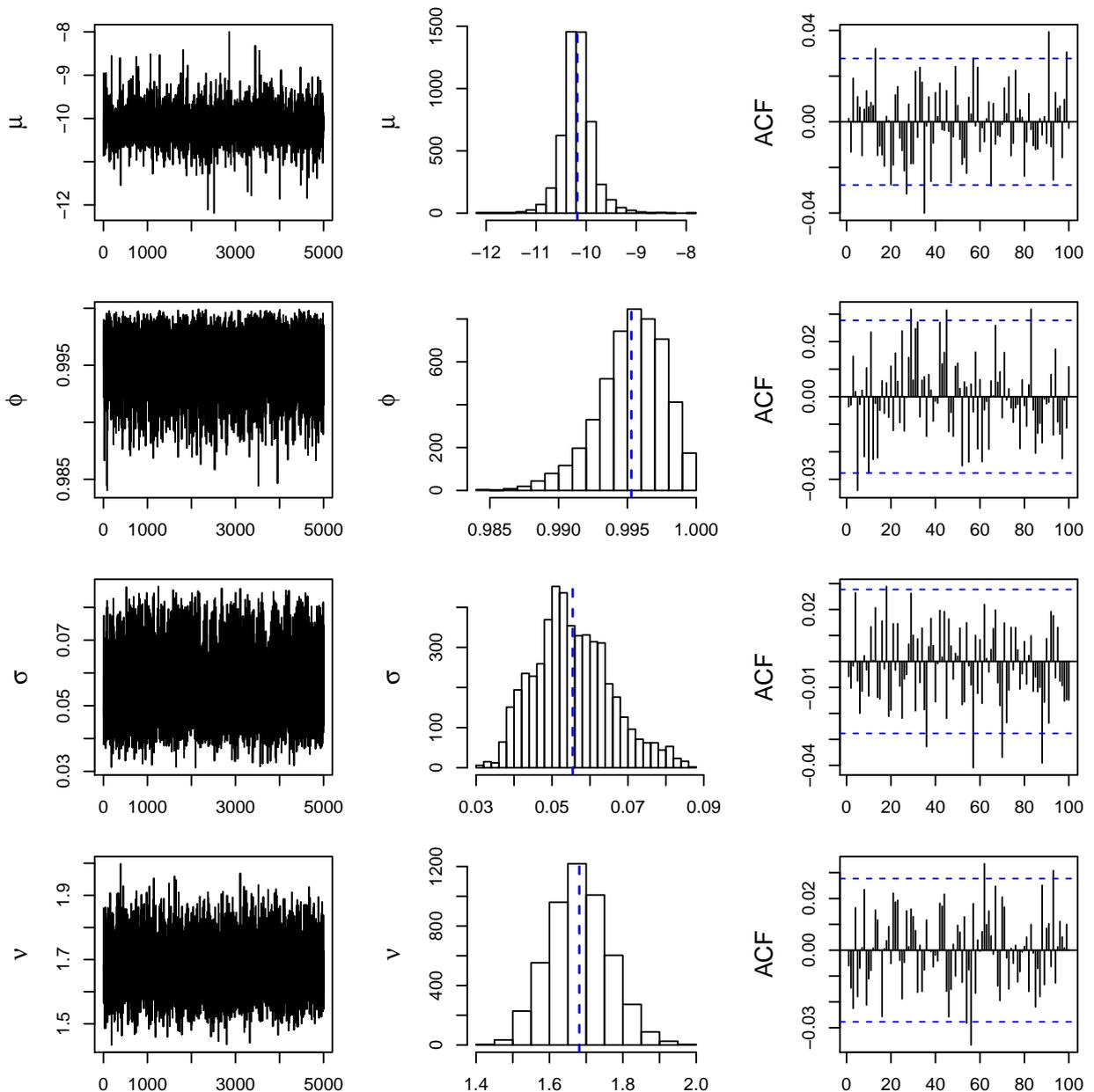
Figura 14 – Gráficos de convergência dos parâmetros na série de retornos EUR/USD com erros t-Student.



Nas Figuras 18 e 19 são mostradas os gráficos de convergência das cadeias simuladas para os parâmetros da série de retornos do S&P500. Mais uma vez é observada uma boa convergência, de acordo com os gráficos, em ambos os modelos, pois as cadeias não parecem apresentar estrutura de correlação serial. O critério  $\hat{R}$  de Gelman e Rubin (1992) foi de aproximadamente 1 em todos os parâmetros. Assim como na série do IBOVESPA o valor do parâmetro de forma, no modelo com erros Skew-Normais contém o zero e os valores das médias *a posteriori* dos demais parâmetros é próximo, dando indícios de que os ajustes são parecidos.

Nas Figuras 20 e 21 são apresentadas as volatilidades estimadas em percentuais  $\hat{\sigma} = 100 \exp(\hat{h}_t/2)$  tomando as médias e os quantis amostrais de 5% e 95% da distribuição *a posteriori* de  $h_t$  e sobrepondo sobre os retornos absolutos em percentuais. É possível notar que as volatilida-

Figura 15 – Gráficos de convergência dos parâmetros na série de retornos EUR/USD com erros GED.

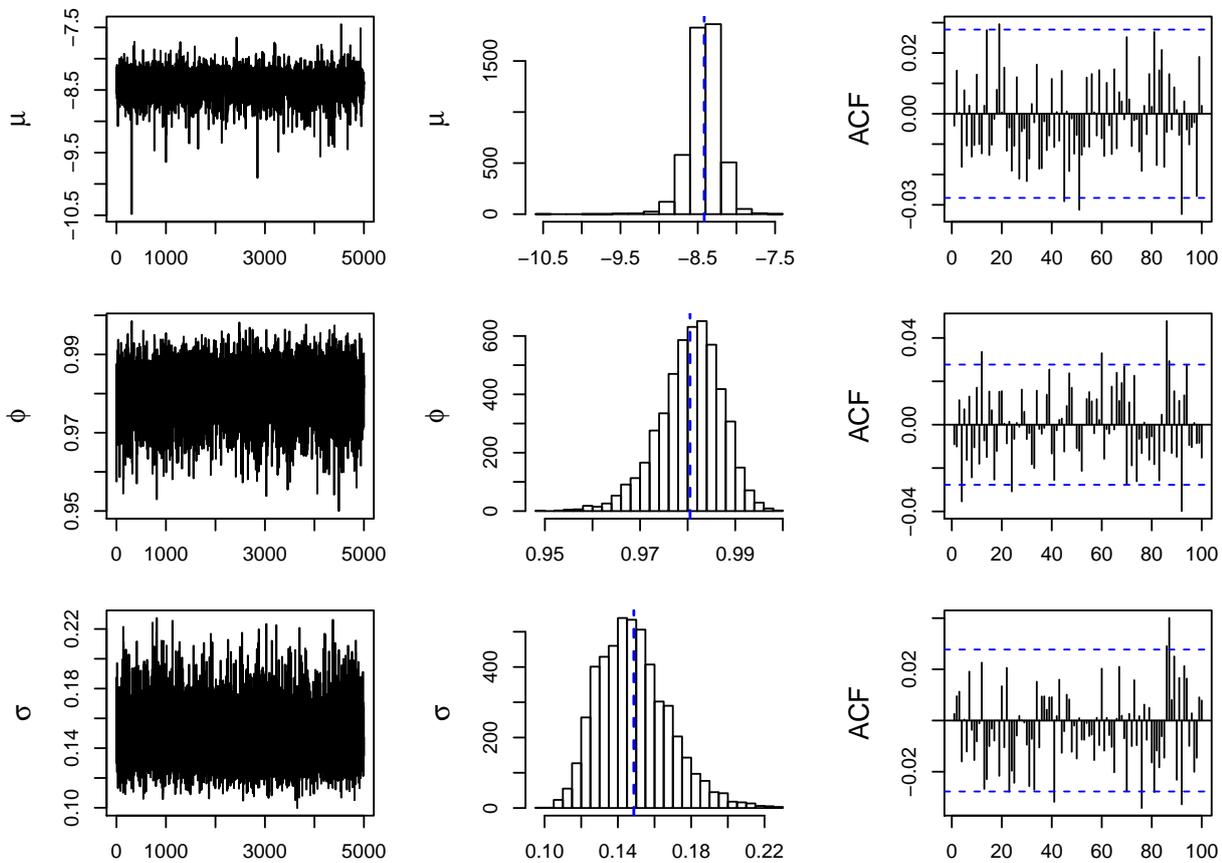


des estimadas descrevem bem o comportamento dos retornos £/USD e as bandas de credibilidade tem maior amplitude quando as volatilidades aumentam em ambas as distribuições.

De acordo com as Figuras 22 e 23, as volatilidades estimadas também seguem bem o agrupamento de volatilidades observadas da série de retornos EUR/USD absolutos em percentuais considerando as duas distribuições selecionadas para a equação (2.8) do modelo proposto. Verificamos que as bandas de credibilidade tem menor amplitude que na série £/USD.

As Figuras 24 e 25 apresentam as volatilidades estimadas em percentuais  $\hat{\sigma} = 100 \exp(\hat{h}_t/2)$  tomando as médias e os quantis amostrais de 5% e 95% da distribuição *a posteriori* de  $h_t$ , juntamente com a série de retornos do IBOVESPA em valores absolutos. Podemos notar que as volatilidades parecem seguir bem o comportamento dos retornos do IBOVESPA. Notamos ainda

Figura 16 – Gráficos de convergência dos parâmetros na série de retornos IBOVESPA com erros Gaussianos.



que para esta série as estimativas parecem menos suavizadas em relação as séries de retornos de taxas de câmbio.

Uma observação interessante é que na série  $\text{£/USD}$  e na série do IBOVESPA, as volatilidades estimadas foram bem parecidas, independente da distribuição utilizada para o erro. Isso de certa forma é esperado, considerando o fato de que os critérios de seleção obtiveram valores muito próximos para as distribuições selecionadas para o erro. Já para a série  $\text{EUR/USD}$  há uma leve elevação nos valores estimados utilizando a distribuição GED.

Apesar de apresentar uma boa convergência, de acordo com os gráficos, e a medida de  $\hat{R}$  de Gelman e Rubin (1992), com aproximadamente 1 em todos os parâmetros do modelo, os gráficos de volatilidade estimados apresentam uma forma menos suavizada para esta série, em relação as demais. Estudamos a convergência das cadeias simuladas nesta aplicação e observou-se, ao realizar uma amostragem nas volatilidades estimadas, que houve convergência. Para confirmar o estudo, realizou-se a aplicação considerando um número maior de simulações (50 mil, com *burn-in* de 50%) e obteve-se resultados muito similares, conforme apresentado no Apêndice A.

Figura 17 – Gráficos de convergência dos parâmetros na série de retornos IBOVESPA com erros Skew-Normal.

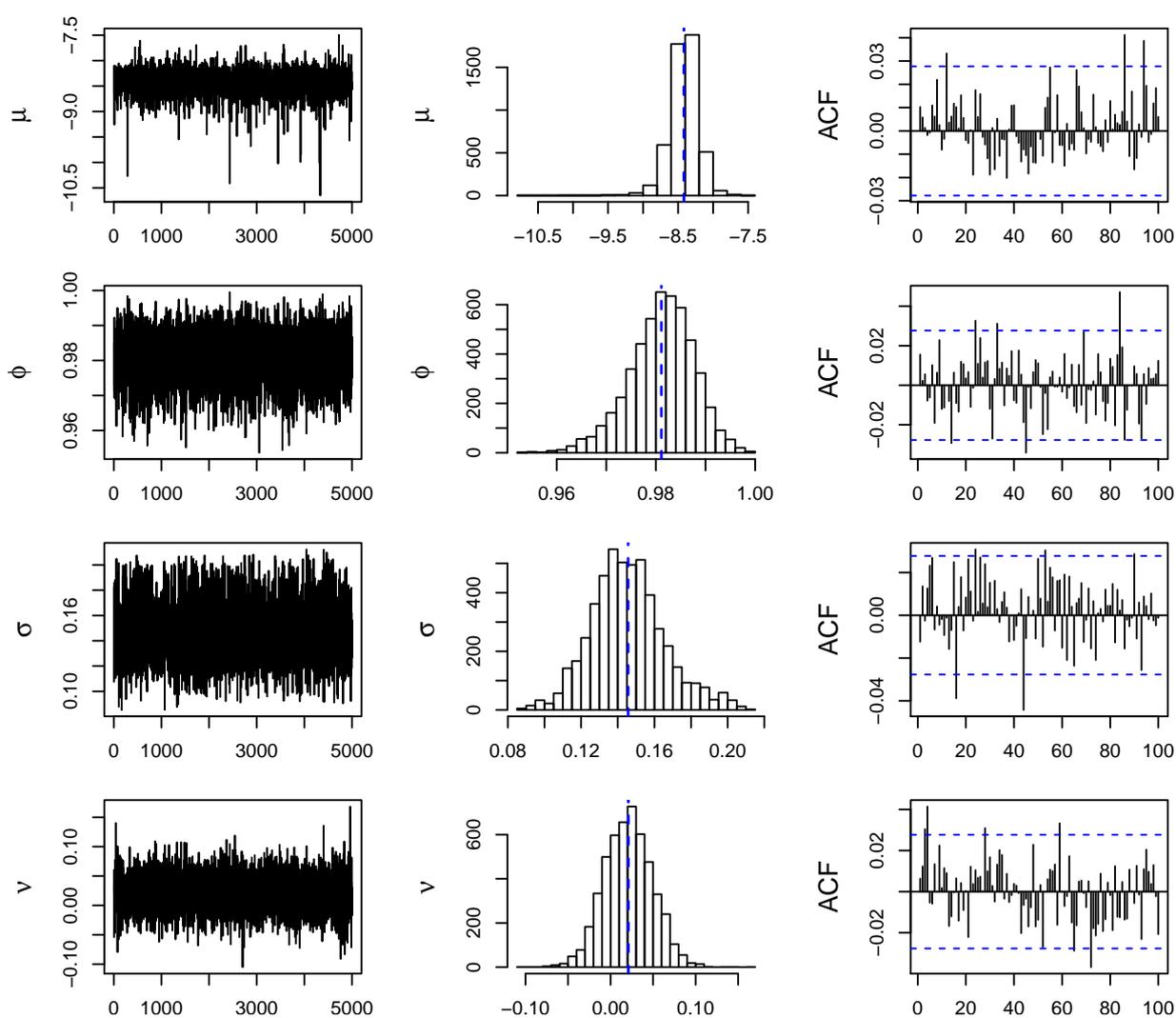


Figura 18 – Gráficos de convergência dos parâmetros na série de retornos S&amp;P500 com erros Gaussianos.

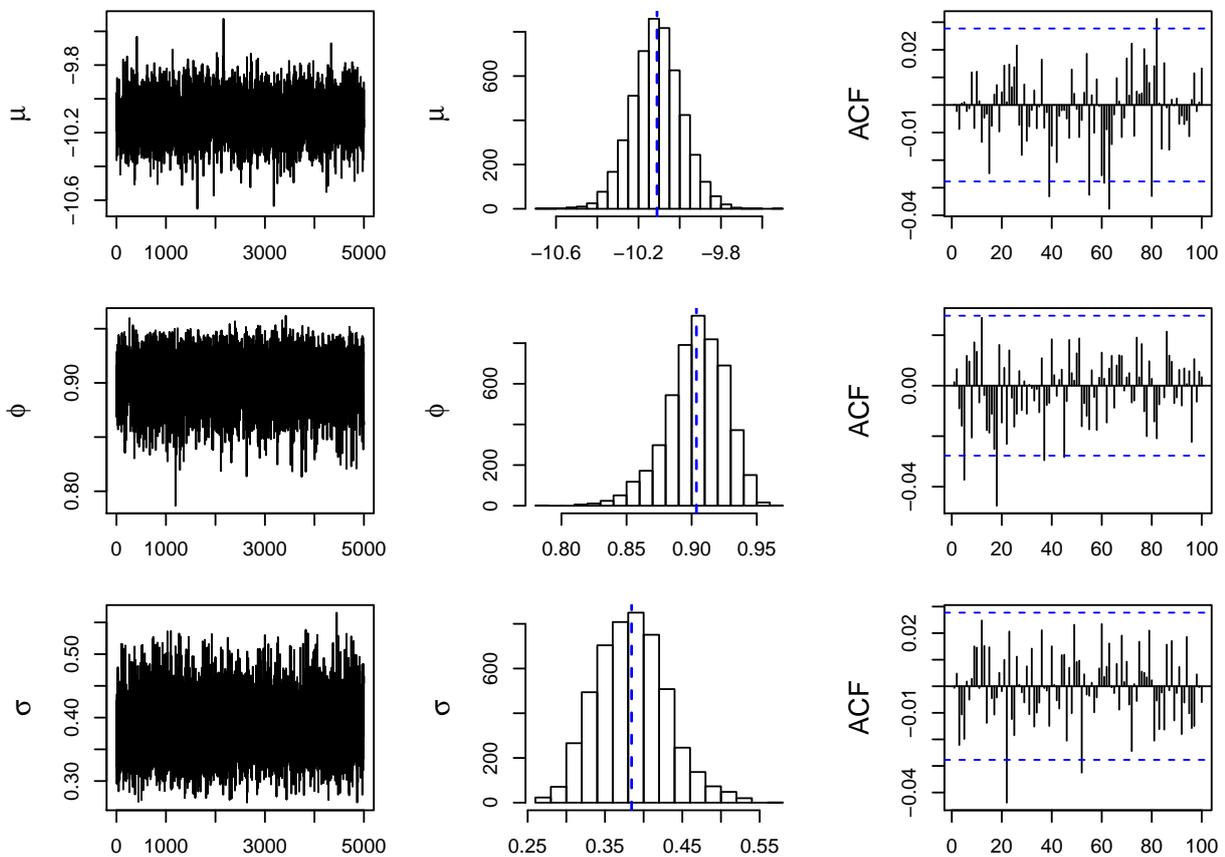


Figura 19 – Gráficos de convergência dos parâmetros na série de retornos S&amp;P500 com erros Skew-Normal.

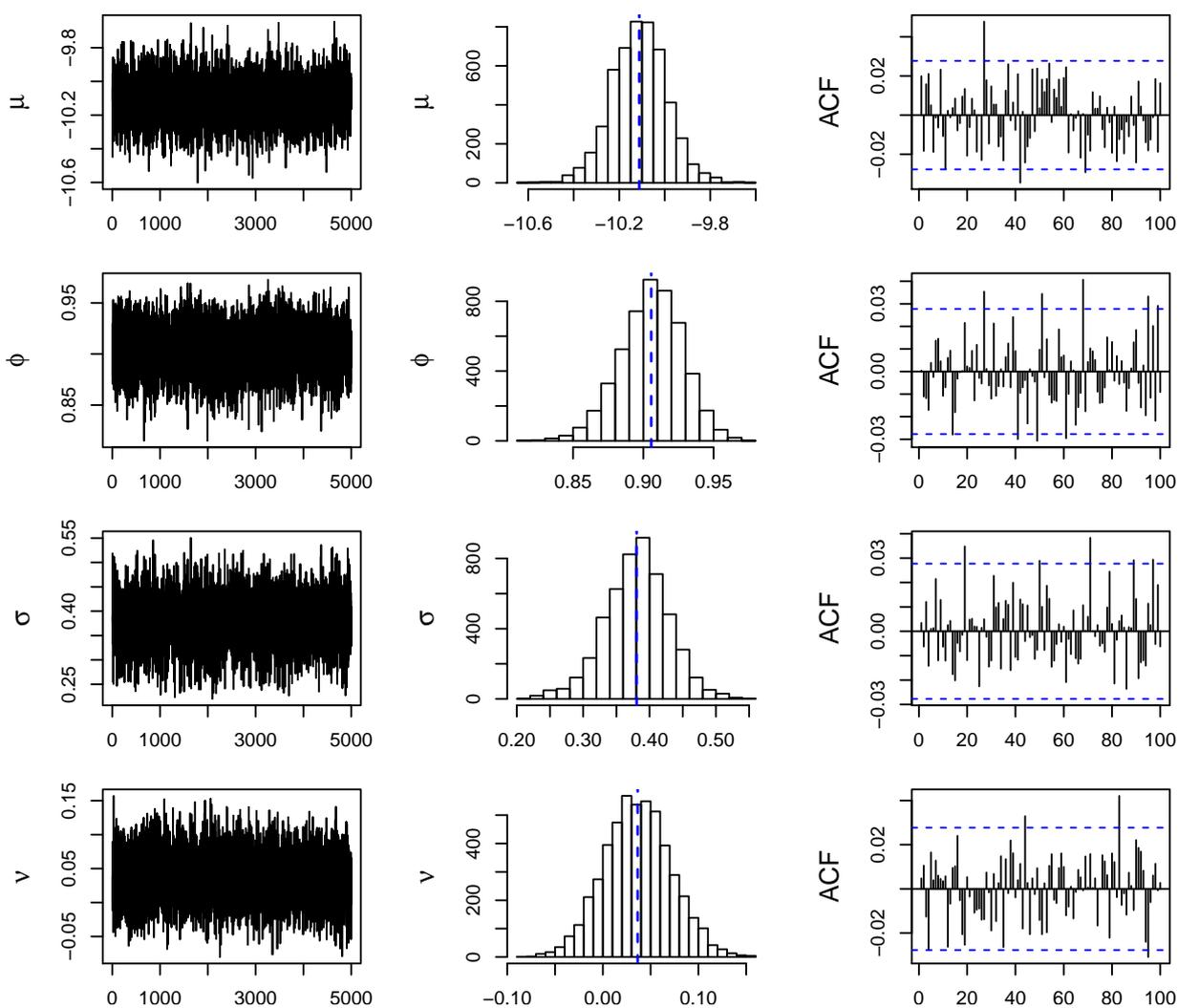


Figura 20 – Retornos absolutos em percentual para a série £/USD e volatilidades estimadas com bandas de credibilidade de 95% com erros Gaussianos.

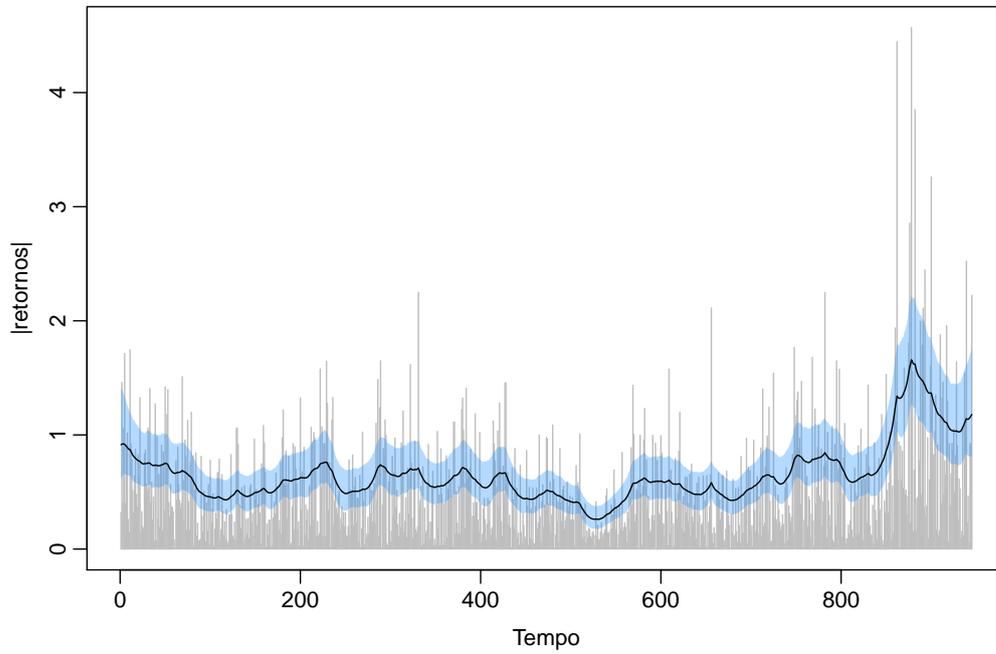


Figura 21 – Retornos absolutos em percentual para a série £/USD e volatilidades estimadas com bandas de credibilidade de 95% com erros Skew-Normal.

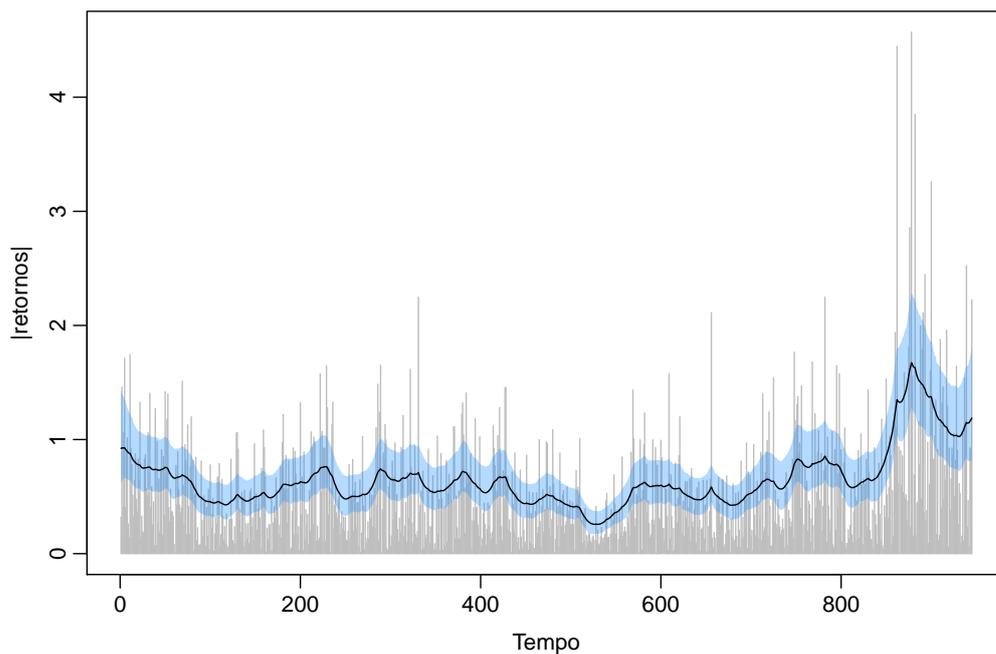


Figura 22 – Retornos absolutos em percentual para a série EUR/USD e volatilidades estimadas com bandas de credibilidade de 95% com erros t-Student.

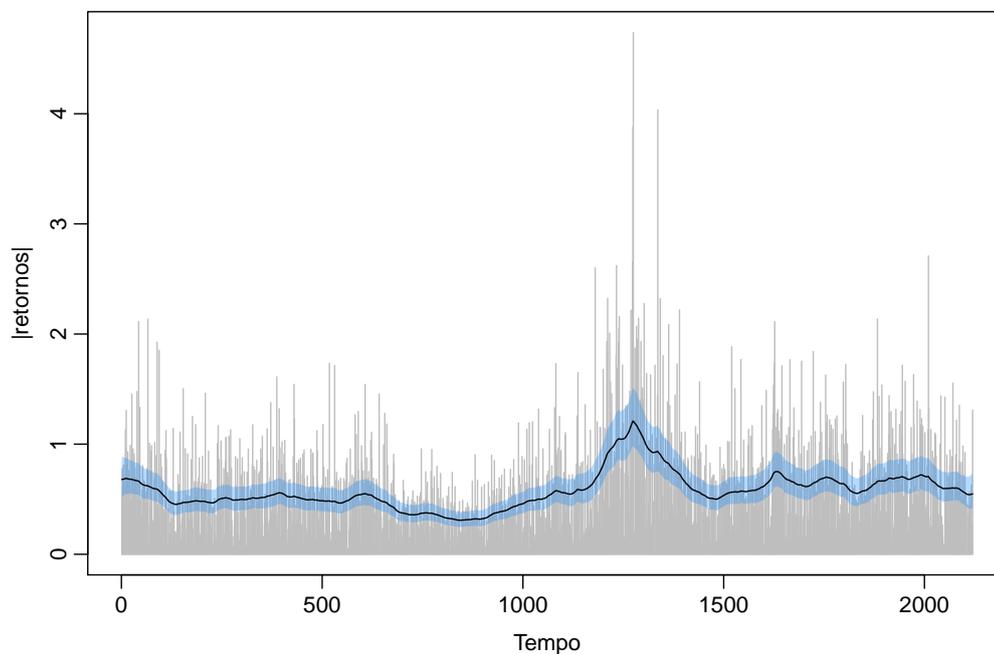


Figura 23 – Retornos absolutos em percentual para a série EUR/USD e volatilidades estimadas com bandas de credibilidade de 95% com erros GED.

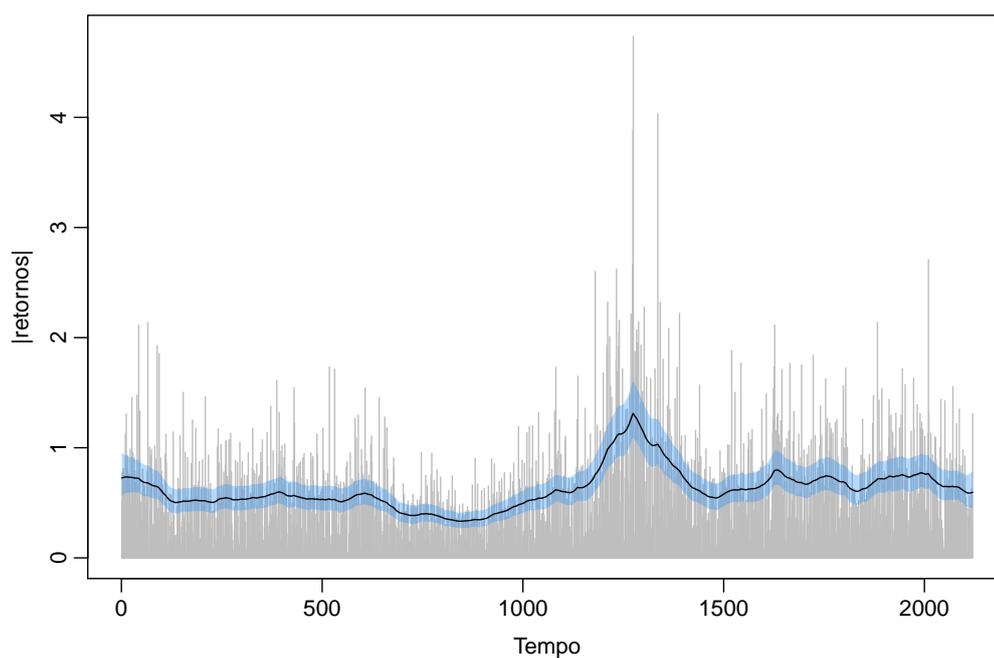


Figura 24 – Retornos absolutos em percentual para a série do IBOVESPA e volatilidades estimadas com bandas de credibilidade de 95% com erros Gaussianos.

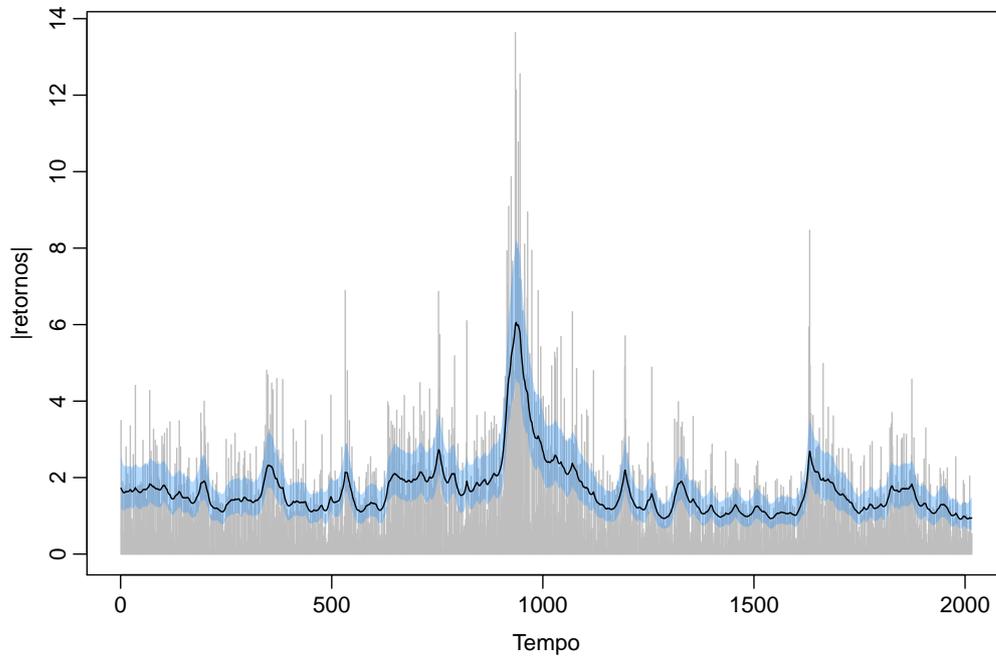


Figura 25 – Retornos absolutos em percentual para a série do IBOVESPA e volatilidades estimadas com bandas de credibilidade de 95% com erros Skew-Normal.

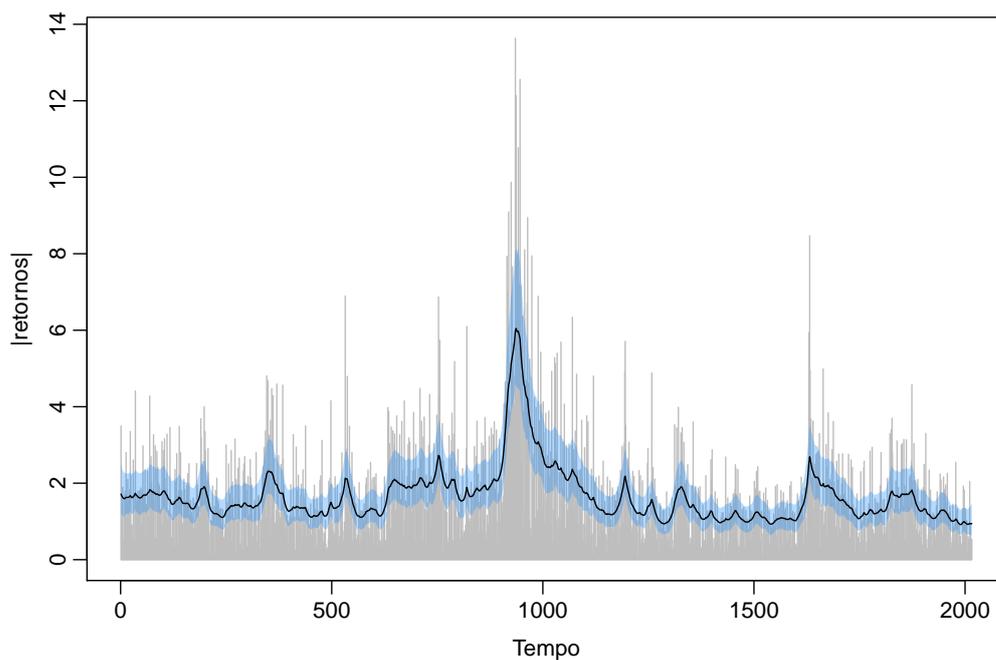


Figura 26 – Retornos absolutos em percentual para a série do S&P500 e volatilidades estimadas com bandas de credibilidade de 95% com erros Gaussianos.

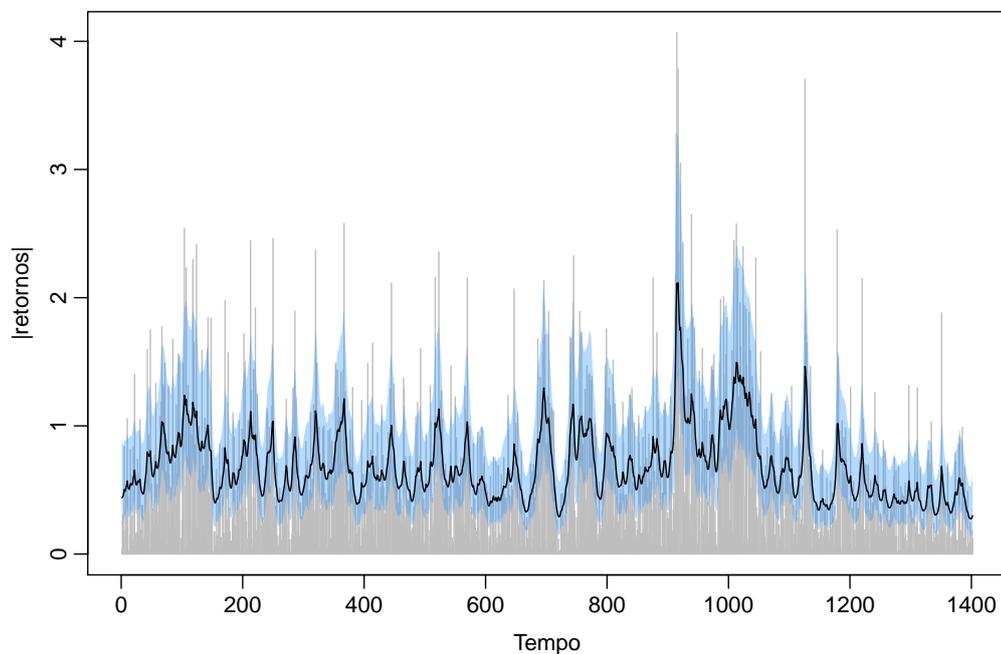
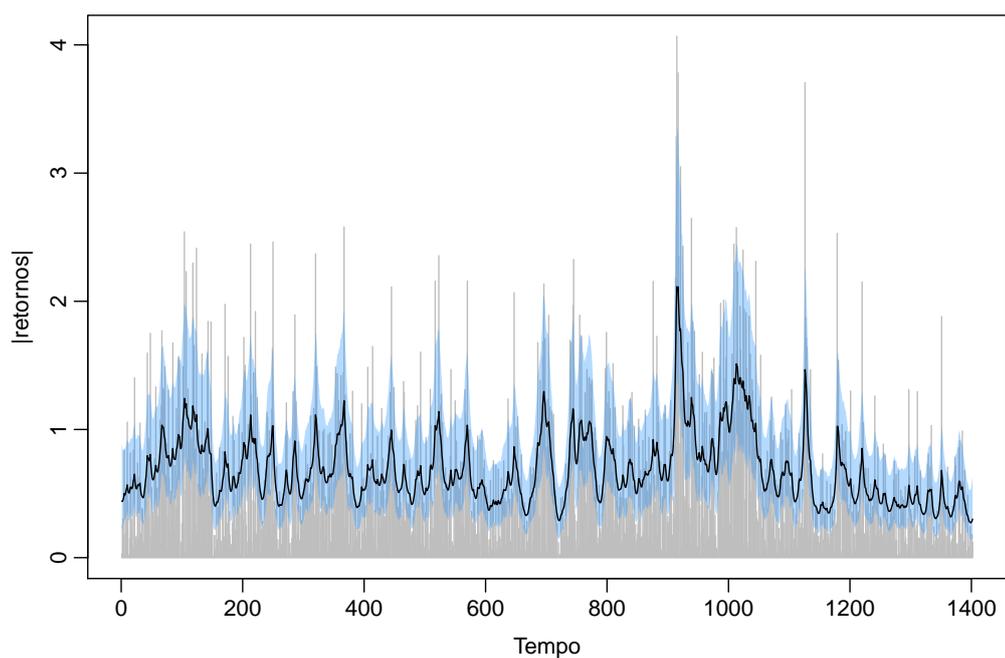


Figura 27 – Retornos absolutos em percentual para a série do S&P500 e volatilidades estimadas com bandas de credibilidade de 95% com erros Skew-Normal.



## 4.4 Sensibilidade à escolha da distribuição a priori

Nesta seção verificou-se se a escolha da priori para o parâmetro  $\sigma_\eta$  pode influenciar nos resultados. Com isso, estimou-se o modelo de volatilidade estocástica utilizando as séries £/USD e S&P500 considerando as distribuições propostas para este parâmetro que estão descritas nas equações (2.20), (2.21) e (2.22).

A sensibilidade foi avaliada considerando os critérios de seleção dos modelos, apresentados na seção (2.3), para aferir, se devido a escolha da priori, o resultado é discrepante em relação aos que foram obtidos. Na estimação, foram simuladas 5000 amostras e descartadas 50% após o aquecimento, resultando em cadeias de 2500 simulações. Na estimação considerou-se as quatro distribuições propostas para os erros da equação de observações e realizou-se este procedimento duas vezes.

Tabela 14 – Comparação entre os modelos propostos via critérios de seleção para estudo de sensibilidade, série £/USD.

Priori para $\sigma_\eta$	Dist.	DIC	WAIC	EP <sub>waic</sub>	LOO	EP <sub>loo</sub>
$\mathcal{G}^{-1}(2.5, 0.025)$	(1) Gaussiana	1805,0	1810,3	53,8	1811,0	54,5
		1801,4	1805,8	53,4	1809,6	54,0
	(2) t-Student	1804,5	1811,5	52,8	1812,3	52,9
		1805,7	1813,4	52,8	1814,2	52,9
	(3) Skew-Normal	1803,7	1808,3	53,4	1812,0	54,0
		1803,4	1807,7	53,3	1811,4	53,9
	(4) GED	1809,7	1810,4	53,2	1813,2	53,0
		1812,6	1813,3	53,2	1815,5	53,6
$\mathcal{G}\left(\frac{1}{2}, 5\right)$	(1) Gaussiana	1800,9	1805,0	53,1	1809,3	53,8
		1801,7	1806,8	53,4	1811,1	54,1
	(2) t-Student	1805,3	1812,8	52,8	1813,4	52,9
		1804,0	1810,7	52,9	1811,5	52,9
	(3) Skew-Normal	1804,0	1808,8	53,5	1812,7	54,2
		1803,9	1808,2	53,3	1812,1	54,0
	(4) GED	1809,0	1809,5	53,0	1812,0	53,4
		1811,4	1812,2	53,2	1815,0	53,7
$Inv - \chi^2(10, 0.05)$	(1) Gaussiana	1803,4	1808,0	53,4	1811,7	54,0
		1804,2	1809,0	53,5	1813,0	54,3
	(2) t-Student	1805,8	1813,5	52,8	1814,1	52,9
		1807,3	1815,1	52,9	1815,7	53,0
	(3) Skew-Normal	1804,2	1808,5	53,4	1811,6	53,9
		1802,9	1807,6	53,3	1811,2	54,0
	(4) GED	1811,3	1812,0	53,2	1813,9	53,5
		1814,8	1815,5	53,3	1817,0	53,6

De acordo com os resultados mostrados na Tabela 14, notamos que a escolha da priori não influencia na seleção do modelo. Observamos que os critérios apresentam valores próximos, contudo, mantendo o resultado de considerar o modelo com erros Gaussianos e Skew-Normal

como os mais bem ajustados. Algo interessante a destacar é que, mesmo com 2500 simulações efetivas, o método HMC obteve excelentes resultados (não mostrados aqui) de convergência e ajuste. Os Valores de DIC, WAIC e LOO apresentados são similares aos obtidos para 5000 simulações efetivas (após o aquecimento da cadeia).

Tabela 15 – Comparação entre os modelos propostos via critérios de seleção para estudo de sensibilidade, série S&P500.

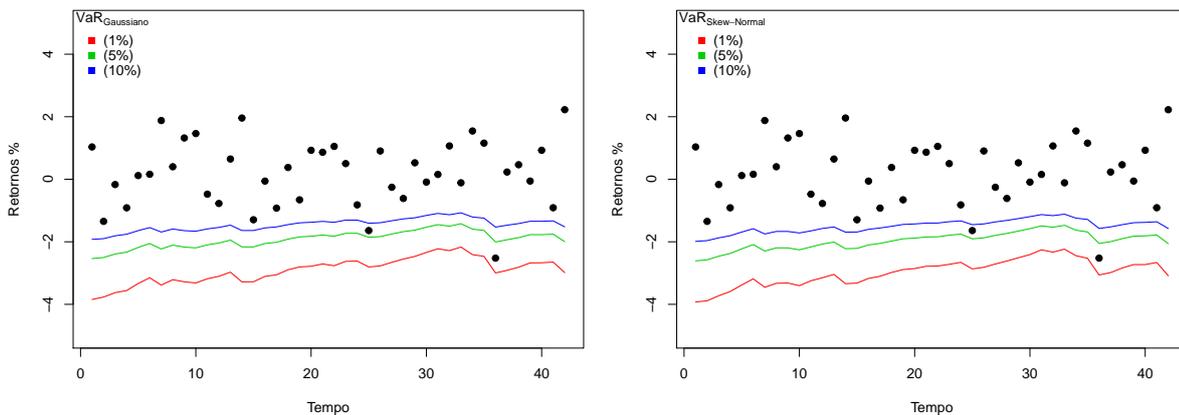
Priori para $\sigma_\eta$	Dist.	DIC	WAIC	EP <sub>waic</sub>	LOO	EP <sub>loo</sub>
$\mathcal{G}^{-1}(2.5, 0.025)$	(1) Gaussiana	-10016,1	-10014,6	65,3	-9982,4	67,7
		-10016,4	-10016,1	65,1	-9983,4	67,5
	(2) t-Student	-9985,4	-9979,9	68,1	-9970,5	68,7
		-9984,7	-9978,5	68,1	-9969,3	68,8
	(3) Skew-Normal	-10016,5	-10015,5	65,3	-9982,1	67,8
		-10015,8	-10014,2	65,5	-9982,9	67,9
	(4) GED	-9986,0	-9990,7	67,5	-9977,8	68,5
		-9983,7	-9987,9	67,6	-9974,0	68,7
$\mathcal{G}(\frac{1}{2}, 5)$	(1) Gaussiana	-10017,3	-10017,3	65,1	-9985,2	67,4
		-10017,6	-10016,4	65,3	-9987,1	67,4
	(2) t-Student	-99862,9	-9982,0	67,9	-9972,3	68,6
		-99848,3	-9981,3	68,1	-9971,3	68,7
	(3) Skew-Normal	-10014,2	-10012,8	65,4	-9983,0	67,5
		-10013,5	-10011,9	65,4	-9981,2	67,6
	(4) GED	-99815,0	-9985,4	67,7	-9973,2	68,7
		-99836,8	-9988,3	67,5	-9978,0	68,3
$Inv - \chi^2(10, 0.05)$	(1) Gaussiana	-10016,3	-10013,7	65,4	-9984,3	67,6
		-10015,4	-10013,7	65,4	-9981,6	67,8
	(2) t-Student	-99857,7	-9979,3	68,1	-9970,9	68,7
		-99854,3	-9979,4	68,1	-9970,6	68,7
	(3) Skew-Normal	-10015,3	-10014,5	65,4	-9984,9	67,5
		-10014,9	-10012,4	65,5	-9981,5	67,8
	(4) GED	-99823,8	-9986,0	67,7	-9975,0	68,6
		-99836,3	-9987,6	67,6	-9975,9	68,5

Conforme os resultados apresentados na Tabela 15, é possível observar que os valores obtidos para os critérios de seleção favorecem a escolha das distribuições Gaussiana e Skew-Normal. Notamos ainda que para a distribuição  $\mathcal{G}(\frac{1}{2}, 5)$  nos dois casos foi escolhida sempre a priori Gaussiana, que obteve valores um pouco inferiores aos da Skew-Normal. Já para as outras prioris os valores apresentados são bem mais próximos e em um dos casos ( $Inv - \chi^2(10, 0.05)$ ) a escolhida, através dos critérios WAIC e LOO, é a distribuição Skew-Normal. Contudo, os resultados de escolha das distribuições para os erros do modelo é mantido.

## 4.5 Valor em Risco (VaR)

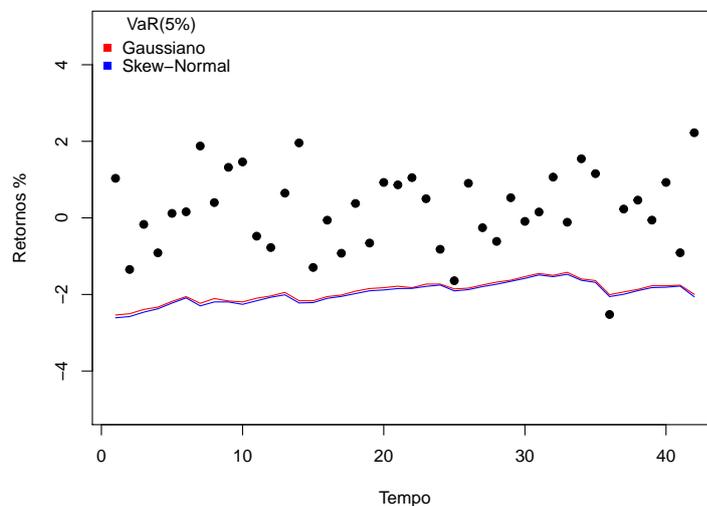
Para ilustrar o método, trabalhamos com as séries utilizando o procedimento de estimar o VaR de 1%, 5% e 10% um dia a frente para aproximadamente dois meses de bolsa, sendo que para cada uma das aplicações foram utilizadas, respectivamente, 42, 45, 41 e 41 observações. Para reproduzir um cenário real, os parâmetros do modelo são estimados novamente a partir das observação após se obter o cálculo do  $\text{VaR}_{T+1}$ , ou seja, consideramos janela móvel.

Figura 28 – Valor em Risco (VaR) com erros Gaussianos e Skew-Normais, para a série de retornos £/USD.



(a) VaR Gaussiano de 1%, 5% e 10%.

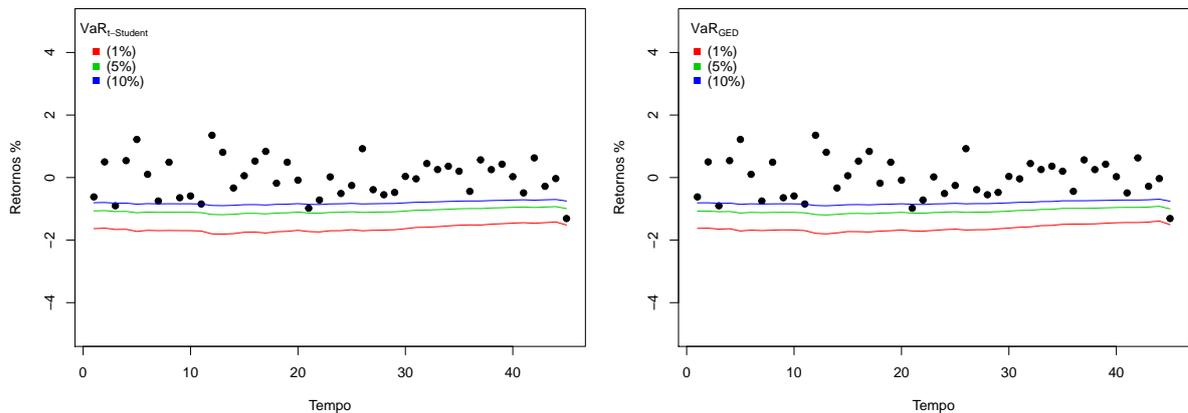
(b) VaR Skew-Normal de 1%, 5% e 10%.



(c) VaR 5% para erros Gaussianos e Skew-Normais.

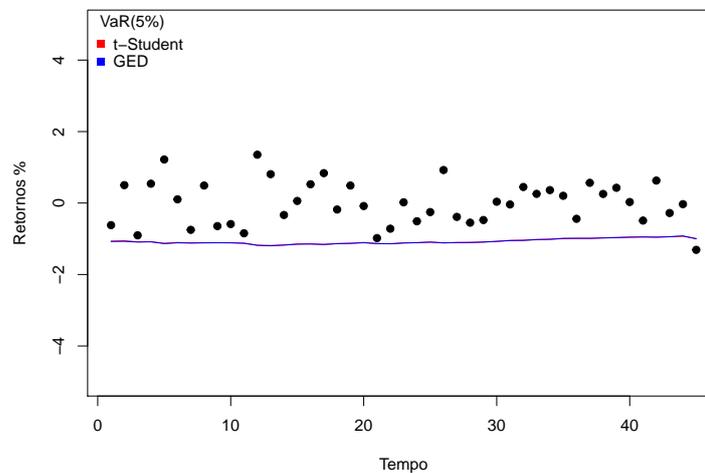
Observamos, conforme a Figura 28, que as estimativas VaR seguem muito bem a volatilidade no mercado e reagem bem aos movimentos extremos para baixo, ou seja, grandes valores negativos de retorno. Além disso, obtivemos apenas uma observação para ambas as distribuições fora dos limites de 5% do VaR para as distribuições Gaussiana e Skew-Normal, respectivamente, menos do que o esperado, 2,1 (cinco por cento de 42 observações). Nota-se que as estimativas VaR obtidas através da distribuição Gaussiana apresentam valores um pouco maiores, porém, próximos dos obtidos através da distribuição Skew-Normal.

Figura 29 – Valor em Risco (VaR) com erros t-Student e GED, para a série de retornos EUR/USD.



(a) VaR t-Student de 1%, 5% e 10%.

(b) VaR GED de 1%, 5% e 10%.

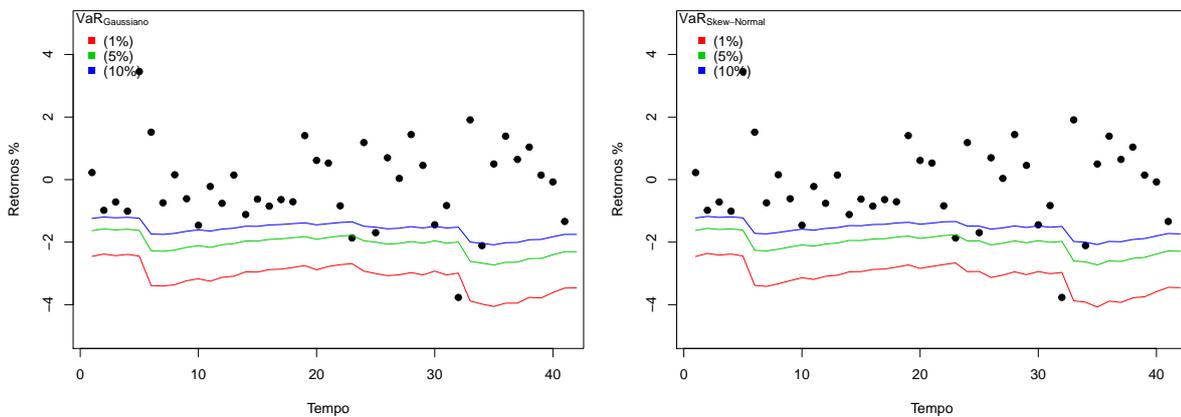


(c) VaR 5% para erros t-Student e GED.

Na série de retornos EUR/USD, o VaR, para as distribuições t-Student e GED, apresentado na Figura 29, notamos um comportamento suavizado nas projeções do valor em risco. Observa-se que apenas um ponto aparece abaixo dos limites de 5% em ambas as distribuições, já ao considerar os 10% notamos quatro pontos abaixo dos limites de pré-estabelecidos pelo valor em risco. Estes valores são abaixo, no entanto, próximos do esperado, 2,25 e 4,5, respectivamente (cinco e dez por cento de 45 observações). O gráfico apresentado na Figura 29c mostra que os limites de 5% do VaR para as distribuições t-Student e GED são muito similares.

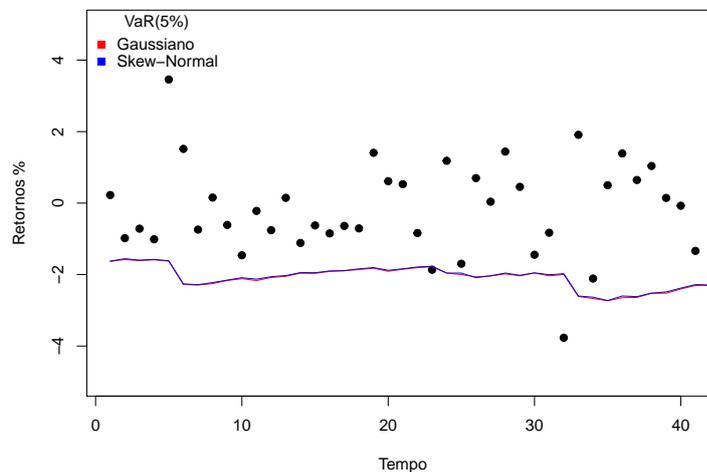
A Figura 30 apresenta as estimativas para o Valor em Risco (VaR) para a série do IBOVESPA, com erros Gaussianos e Skew-Normais. Os resultados mostram que as estimativas para o VaR tanto para a distribuição Gaussiana quanto para a Skew-Normal são muito próximas. De certa forma já era esperado de acordo com as estimativas obtidas na Tabela 13, dado que o valor estimado para o parâmetro de forma da distribuição Skew-Normal apresenta valor próximo de zero e seu intervalo empírico indica que há cerca de 95% de credibilidade de conter o valor zero. Cabe destacar que um ponto aparece abaixo do limite de 1%, mais que o esperado 0,41 (um por cento de 41 observações). Já para os demais limites, 5% e 10%, os valores correspondentes

Figura 30 – Valor em Risco (VaR) com erros Gaussianos e Skew-Normais, para a série de retornos IBOVESPA.



(a) VaR Gaussiano de 1%, 5% e 10%.

(b) VaR Skew-Normal de 1%, 5% e 10%.



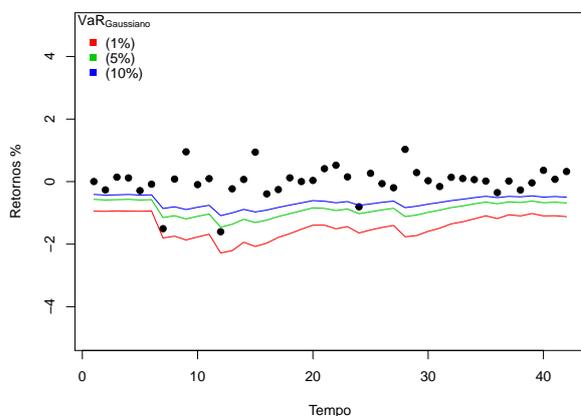
(c) VaR 5% para erros Gaussianos e Skew-Normais.

aos pontos que violaram os limites está de acordo com o esperado, ou seja, respectivamente 2 e 4 observações aparecem fora dos extremos pré-estabelecidos.

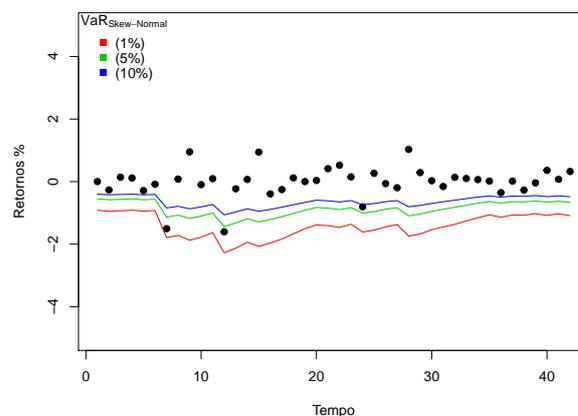
Assim como para as demais aplicação no cálculo do VaR, de acordo com a Figura 31, quantitativamente a performance observada através do modelo proposto é boa, visto que apenas três pontos violam os limites de 5% e 10%. Novamente nota-se que os valores apresentados para o VaR em ambos os erros, Gaussianos e Skew-Normais, são similares e, de acordo com a Tabela 13, as estimativas obtidas para os parâmetros do modelo para essas distribuições são próximas. Vale ressaltar que o valor estimado para o parâmetro de forma da distribuição Skew-Normal apresenta valor muito baixo e seu intervalo empírico indica que há cerca de 95% de credibilidade de conter o valor zero, ou seja, neste caso, seria esperado que os Valores em Risco estimados, em ambos os caso, fossem próximos.

Portanto, em nosso estudo empírico, em termos de aplicações, o VaR estimado para ambas as distribuições a partir do modelo de volatilidade estocástica proposto segue bem as volatilidades do mercado e reagem com boa performance aos movimentos extremos baixos, isto

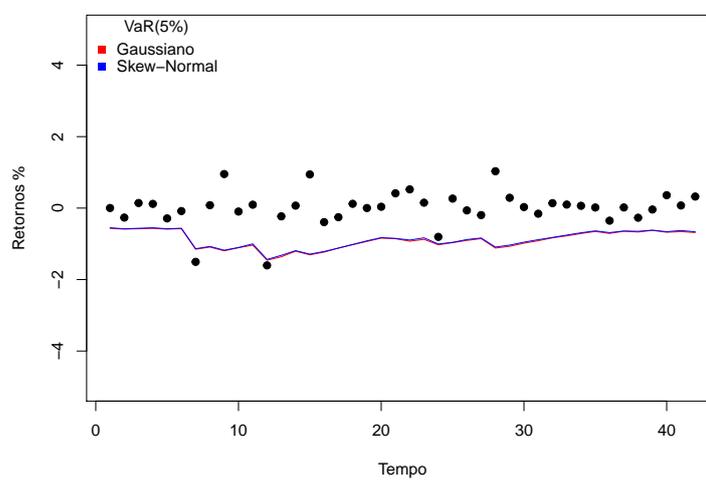
Figura 31 – Valor em Risco (VaR) com erros Gaussianos e Skew-Normais, para a série de retornos S&amp;P500.



(a) VaR Gaussiano de 1%, 5% e 10%.



(b) VaR Skew-Normal de 1%, 5% e 10%.



(c) VaR 5% para erros Gaussianos e Skew-Normais.

é, retornos negativos grandes.



---

## CONSIDERAÇÕES FINAIS

---

Neste trabalho foi discutida a estimação Bayesiana em modelos de volatilidade estocástica com o uso do método de Monte Carlo Hamiltoniano (HMC). Foi possível observar que o método HMC possui relativa superioridade em relação aos métodos tradicionalmente utilizados como *Gibbs Sampling* e Metropolis-Hastings, além de proporcionar uma amostragem com baixa correlação serial, convergir com um número menor de observações e não necessitar de amostragem após a geração das cadeias, o que o torna relevante do ponto de vista teórico. Por exigir um número maior de passos antes de realizar uma proposta, o algoritmo HMC pode levar mais tempo por iteração, o que o torna, na perspectiva prática, uma complicação dependendo do tipo de aplicação.

No estudo de simulação 1, notou-se que o método HMC, considerando os critérios de viés e *smse*, obteve melhores resultados para a maioria dos parâmetros em ambas as configurações, em relação à metodologia proposta na biblioteca `stochvol`. Já no estudo de simulação 2, os resultados apresentados mostram uma boa convergência das cadeias para obtenção das distribuições a posteriori dos parâmetros. Entretanto, o tempo estimado para as gerações das cadeias simuladas, no caso do modelo de volatilidade estocástica, está relacionado com a distribuição considerada para os erros da equação de observações.

Em termos de generalização, foram utilizadas diferentes distribuições para os erros da equação de observação do modelo de volatilidade estocástica. Com isso, é possível obter uma vantagem em relação aos pacotes que consideram apenas a distribuição Gaussiana, visto que algumas séries apresentam curtose elevada ou assimetria e, nestes casos, o modelo com distribuição Gaussiana pode não conseguir captar da melhor forma estas particularidades. Além disso, apresentou-se dois critérios de seleção de modelos que são bastante robustos, pois além de possuírem um erro padrão associado, buscam aproximar o método de validação cruzada. Do ponto de vista prático, estes métodos, podem ser utilizados em diversas aplicações, não somente em modelos de volatilidade estocástica.

Apesar de considerar outras distribuições para a equação de observações do modelo de volatilidade, na aplicação aos conjuntos de dados observamos que apenas para a série de câmbio EUR/USD os critérios de seleção favorecem a escolha de distribuições com caudas mais pesadas que a distribuição normal. Verificamos também que os critérios de seleção propostos apresentam valores de EP consideravelmente elevados, o que nos leva a crer que os critérios são bem próximos em termos de erro de Monte Carlo. Por outro lado, os critérios utilizados apontam para os modelos com distribuições que apresentam os melhores ajustes. Contudo, vale ressaltar que os indicadores são medidos empiricamente e que a escolha do modelo que melhor representa a dinâmica dos dados cabe ao pesquisador. Mostramos ainda um estudo de sensibilidade a escolha da distribuição *a priori* para o parâmetro  $\sigma_\eta$ , visto que diversos autores propõem diferentes *prioris* para este parâmetro. Observamos que o método HMC apresentou resultados bastante eficientes na aplicação a este tipo de análise, no entanto, dependendo do tamanho da série, em alguns casos, como o da distribuição GED, por exemplo, que não está implementada no pacote `rstan`, o processo de simulação pode se tornar bastante demorado, o que não é factível na prática. Uma opção simples e viável é estudar se para tamanhos maiores da série o método converge com um menor número de simulações, tendo em vista que a série forneceria maior informação através das observações.

Adicionalmente, calculamos os valores em risco (VaR) para as séries, com base nas estimativas obtidas para as médias a posteriori dos parâmetros dos modelos que apresentaram melhores performances, de acordo com os critérios de seleção. Ilustrou-se o procedimento de estimar o VaR um dia a frente para aproximadamente dois meses de bolsa, ou seja, cada vez que foi estimado o valor em risco no dia  $T$ , o valor do retorno deste dia foi considerado na amostra e simulou-se novamente as cadeias de Markov para obter novas médias estimadas a posteriori e, a partir destas simular novamente o  $\text{VaR}_{T+1}$  (janela móvel). Com este estudo empírico foi possível analisar que os modelos possuem um bom desempenho em detectar movimentos extremos baixos, ou seja, retornos negativos elevados, dado que os valores em risco estimados seguem bem as volatilidades do mercado.

Vale ressaltar que as diferenças apresentadas a partir das análises gráficas do VaR, considerando as aplicações para as séries £/USD, IBOVESPA e S&P500, nas distribuições que possuem os melhores valores nas métricas utilizadas para escolha do modelo, não foram expressivas. O mesmo ocorre para a série do EUR/USD, em que foi possível notar que os valores de VaR estão bem suavizados, contudo, os modelos selecionados apresentam ótimos resultados em termos de VaR.

## 5.1 Propostas Futuras

### 5.1.1 *Kullback–Leibler Divergence*

Uma propriedade comum em séries temporais financeiras é que os pontos de dados irregulares tendem a se agrupar em determinados intervalos. Esses *clusters* podem ser visualizados como correspondendo a períodos de alta ou baixa volatilidade. No entanto, de acordo com [Hao et al. \(2016\)](#), estudos empíricos existentes revelam que muitos métodos de estimativas, em séries temporais, não são robustos a observações influentes e algumas suposições sobre o modelo podem não ser razoáveis em alguns casos práticos, violações que podem levar a conclusões incorretas. Com isso, a ênfase deve ser colocada não apenas na inferência ou ajuste das séries, mas também no diagnóstico de possíveis observações influentes. Partindo do princípio de que as observações são dependentes, ou seja, indexadas no tempo, não é conveniente avaliar pontos de influência baseado em técnicas como distância de Cook ou DFFITS, onde a ideia é verificar se um ponto é influente a partir de sua exclusão. A literatura para análise de influência em modelos de séries temporais com abordagem bayesiana é escassa. Dessa maneira, uma das propostas futuras para este trabalho é desenvolver uma análise de influência em casos bayesianos para modelos de volatilidade estocástica, baseado no artigo apresentado por [Hao et al. \(2016\)](#) que desenvolveram uma análise de influência direcionada especificamente para modelos GARCH na abordagem bayesiana.

### 5.1.2 *Reversible Jump*

Os métodos de monte Carlos para Cadeias de Markov utilizando o algoritmo *Reversible Jump*, proposto por [Green \(1995\)](#), está diretamente associado a seleção de modelos considerando uma probabilidade marginal a *posteriori* do modelo. Assim, é possível utilizar métodos computacionais MCMC para estimar conjuntamente a incerteza sobre os parâmetros e a probabilidade dos modelos concorrentes estendendo o espaço de estados da cadeia de Markov para permitir um indicador de modelo aplicando-se o algoritmo *Reversible Jump* ([HASTIE; GREEN, 2012](#)).

### 5.1.3 *Outras distribuições e abordagens*

Há uma possibilidade de se construir modelos de volatilidade estocástica considerando outras distribuições, tais como, t-Student Generalizada, t-Student Assimétrica, mistura de distribuições, entre outros, com o auxílio da biblioteca `rstan`. Entretanto, algumas distribuições ainda não estão disponíveis neste pacote e ao construir uma função que ainda não está implementada pode fazer com que, dependendo da complexidade desta, a execução seja um pouco mais demorada. Todavia, pode-se estudar uma abordagem não paramétrica ou modelos de volatilidade determinística/estocástica com mudança de regime, utilizando o método de Monte Carlo Hamiltoniano, com o auxílio da biblioteca `rstan`.

### 5.1.4 Método RMHMC

Adicionalmente, pretende-se utilizar variações que melhorem o desempenho do método HMC, como o Monte Carlo Hamiltoniano em Variedade Riemanniana (*Riemann Manifold Hamiltonian Monte Carlo* - RMHMC) descrito em [Girolami e Calderhead \(2011\)](#), e estender o uso destes métodos a outros modelos.

Cabe destacar que as propostas *Reversible Jump* e Método RMHMC não estão implementadas na biblioteca `rstan` e, com isso, teriam que ser programadas. A medida Kullback–Leibler *Divergence* pode ser construída com o auxílio da biblioteca, assim como o uso de outras abordagens (veja [Stan Development Team and others \(2016\)](#) para maiores detalhes).

## REFERÊNCIAS

---

---

- AKAIKE, H. A new look at the statistical model identification. **Automatic Control, IEEE Transactions on**, Ieee, v. 19, n. 6, p. 716–723, 1974. Citado na página 32.
- BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 64, n. 2, p. 253–280, 2002. Citado na página 18.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of econometrics**, Elsevier, v. 31, n. 3, p. 307–327, 1986. Citado nas páginas 18 e 27.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015. Citado na página 17.
- BROCKWELL, P. J.; DAVIS, R. A. **Time series: theory and methods**. [S.l.]: Springer Science & Business Media, 2013. Citado na página 17.
- CELEUX, G.; FORBES, F.; ROBERT, C. P.; TITTERINGTON, D. M. *et al.* Deviance information criteria for missing data models. **Bayesian analysis**, International Society for Bayesian Analysis, v. 1, n. 4, p. 651–673, 2006. Citado na página 32.
- CHIB, S.; NARDARI, F.; SHEPHARD, N. Markov chain monte carlo methods for stochastic volatility models. **Journal of Econometrics**, Elsevier, v. 108, n. 2, p. 281–316, 2002. Citado na página 18.
- DUANE, S.; KENNEDY, A. D.; PENDLETON, B. J.; ROWETH, D. Hybrid monte carlo. **Physics letters B**, Elsevier, v. 195, n. 2, p. 216–222, 1987. Citado na página 37.
- EHLERS, R. S. Análise de séries temporais. **Universidade Federal do Paraná**, 2007. Citado nas páginas 17 e 26.
- ENGLE, R. Garch 101: The use of arch/garch models in applied econometrics. **The Journal of Economic Perspectives**, JSTOR, v. 15, n. 4, p. 157–168, 2001. Citado na página 18.
- ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. **Econometrica: Journal of the Econometric Society**, JSTOR, p. 987–1007, 1982. Citado nas páginas 18 e 25.
- FRANCQ, C.; ZAKOIAN, J.-M. **GARCH models: structure, statistical inference and financial applications**. [S.l.]: John Wiley & Sons, 2011. Citado na página 35.
- GAMERMAN, D.; LOPES, H. F. **Markov chain Monte Carlo: stochastic simulation for Bayesian inference**. [S.l.]: Chapman and Hall/CRC, 2006. Citado na página 24.
- GELFAND, A. E. Model determination using sampling-based methods. **Markov chain Monte Carlo in practice**, Chapman and Hall, London, p. 145–161, 1996. Citado na página 34.

- GELFAND, A. E.; SMITH, A. F. Sampling-based approaches to calculating marginal densities. **Journal of the American statistical association**, Taylor & Francis Group, v. 85, n. 410, p. 398–409, 1990. Citado na página 21.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. [S.l.]: Taylor & Francis, 2014. v. 2. Citado na página 33.
- GELMAN, A.; HWANG, J.; VEHTARI, A. Understanding predictive information criteria for bayesian models. **Statistics and Computing**, Springer, v. 24, n. 6, p. 997–1016, 2014. Citado na página 33.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, v. 7, p. 457–511, 1992. Citado nas páginas 56, 57, 58, 60, 83 e 84.
- GIROLAMI, M.; CALDERHEAD, B. Riemann manifold langevin and hamiltonian monte carlo methods. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 73, n. 2, p. 123–214, 2011. Citado nas páginas 30, 31 e 78.
- GREEN, P. J. Reversible jump markov chain monte carlo computation and bayesian model determination. **Biometrika**, Oxford University Press, v. 82, n. 4, p. 711–732, 1995. Citado na página 77.
- HAO, H.-X.; LIN, J.-G.; WANG, H.-X.; HUANG, X.-F. Bayesian case influence analysis for garch models based on kullback–leibler divergence. **Journal of the Korean Statistical Society**, Elsevier, v. 45, n. 4, p. 595–609, 2016. Citado na página 77.
- HARVEY, A.; RUIZ, E.; SHEPHARD, N. Multivariate stochastic variance models. **The Review of Economic Studies**, Wiley-Blackwell, v. 61, n. 2, p. 247–264, 1994. Citado na página 51.
- HASTIE, D. I.; GREEN, P. J. Model choice using reversible jump markov chain monte carlo. **Statistica Neerlandica**, Wiley Online Library, v. 66, n. 3, p. 309–338, 2012. Citado na página 77.
- HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. **Biometrika**, Biometrika Trust, v. 57, n. 1, p. 97–109, 1970. Citado na página 21.
- HESTON, S. L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. **Review of financial studies**, Soc Financial Studies, v. 6, n. 2, p. 327–343, 1993. Citado na página 18.
- HOMAN, M. D.; GELMAN, A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. **The Journal of Machine Learning Research**, JMLR. org, v. 15, n. 1, p. 1593–1623, 2014. Citado na página 41.
- IONIDES, E. L. Truncated importance sampling. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 17, n. 2, p. 295–311, 2008. Citado na página 34.
- JACQUIER, E.; POLSON, N. G.; ROSSI, P. E. Bayesian analysis of stochastic volatility models. **Journal of Business & Economic Statistics**, Taylor & Francis, v. 20, n. 1, p. 69–87, 2002. Citado na página 18.
- KASTNER, G. Dealing with stochastic volatility in time series using the r package stochvol. **Journal of Statistical Software**, Foundation for Open Access Statistics, v. 69, n. 5, p. 1–30, 2016. Citado nas páginas 31 e 47.

- KASTNER, G.; FRÜHWIRTH-SCHNATTER, S. Ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc estimation of stochastic volatility models. **Computational Statistics & Data Analysis**, Elsevier, v. 76, p. 408–423, 2014. Citado nas páginas 30 e 47.
- KIM, S.; SHEPHARD, N.; CHIB, S. Stochastic volatility: likelihood inference and comparison with arch models. **The review of economic studies**, Oxford University Press, v. 65, n. 3, p. 361–393, 1998. Citado nas páginas 18, 27 e 30.
- LEIMKUHNER B. & REICH, S. **Simulating Hamiltonian Dynamics**. [S.l.]: Cambridge University Press., 2005. Citado na página 38.
- LEVIN, D. A.; PERES, Y. **Markov chains and mixing times**. [S.l.]: American Mathematical Soc., 2017. v. 107. Citado na página 22.
- MCCAUSLAND, W. J.; MILLER, S.; PELLETIER, D. Simulation smoothing for state–space models: A computational efficiency analysis. **Computational Statistics & Data Analysis**, Elsevier, v. 55, n. 1, p. 199–212, 2011. Citado na página 47.
- METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. **The journal of chemical physics**, AIP, v. 21, n. 6, p. 1087–1092, 1953. Citado na página 21.
- MORETTIN, P. A. **Econometria financeira: um curso em séries temporais financeiras**. [S.l.]: Edgard Blücher, 2008. Citado nas páginas 19, 25, 26 e 36.
- MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. [S.l.]: Blucher, 2006. Citado nas páginas 17 e 27.
- NEAL, R. M. *et al.* Mcmc using hamiltonian dynamics. **Handbook of Markov Chain Monte Carlo**, v. 2, p. 113–162, 2011. Citado nas páginas 18, 24 e 37.
- NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. **Econometrica: Journal of the Econometric Society**, JSTOR, p. 347–370, 1991. Citado na página 18.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>. Citado na página 42.
- ROBERT, C.; CASELLA, G. **Introducing Monte Carlo Methods with R**. [S.l.]: Springer Science & Business Media, 2009. Citado nas páginas 21 e 24.
- RUIZ, E. Quasi-maximum likelihood estimation of stochastic volatility models. **Journal of econometrics**, Elsevier, v. 63, n. 1, p. 289–306, 1994. Citado na página 18.
- SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002. Citado na página 32.
- Stan Development Team and others. Stan: a c++ library for probability and sampling. **Version 2.5.0**, 2014. Citado na página 41.
- \_\_\_\_\_. Rstan: the r interface to stan. **R package version 2.14.1**, 2016. Citado na página 78.
- TAYLOR, S. J. **Modelling financial time series**. [S.l.]: Wiley, 1986. Citado nas páginas 18, 27 e 29.

\_\_\_\_\_. Modeling stochastic volatility: A review and comparative study. **Mathematical finance**, Wiley Online Library, v. 4, n. 2, p. 183–204, 1994. Citado na página 18.

VEHTARI, A.; GELMAN, A.; GABRY, J. Efficient implementation of leave-one-out cross-validation and waic for evaluating fitted bayesian models. **arXiv preprint arXiv:1507.04544**, Technical report, 2015. Citado nas páginas 34 e 35.

\_\_\_\_\_. Practical bayesian model evaluation using leave-one-out cross-validation and waic. **Statistics and Computing**, Springer, v. 27, n. 5, p. 1413–1432, 2017. Citado nas páginas 34 e 35.

WATANABE, S. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. **The Journal of Machine Learning Research**, JMLR.org, v. 11, p. 3571–3594, 2010. Citado na página 33.

ZAKOIAN, J.-M. Threshold heteroskedastic models. **Journal of Economic Dynamics and control**, Elsevier, v. 18, n. 5, p. 931–955, 1994. Citado na página 18.

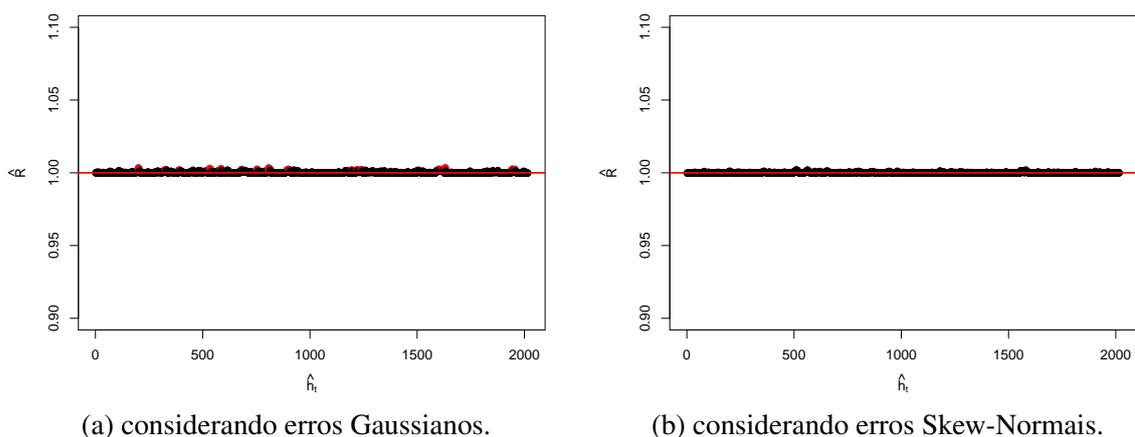
ZEEVALLOS, M.; GASCO, L.; EHLERS, R. Riemann manifold langevin methods on stochastic volatility estimation. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, n. just-accepted, 2016. Citado nas páginas 18, 30, 31, 36, 47 e 51.

## ESTUDOS ADICIONAIS E CÓDIGOS COMPUTACIONAIS

### A.1 Avaliação de Convergência das volatilidades

Estudamos os casos das séries do IBOVESPA e S&P500 para verificar se houve convergência em relação as log-volatilidades, ao considerar o número de simulações igual a 10mil. Para realizar a análise, tomamos uma amostra dentro das log-volatilidades estimadas a *posteriori*, analisamos os gráficos e a medida de convergência  $\hat{R}$  proposta por [Gelman e Rubin \(1992\)](#).

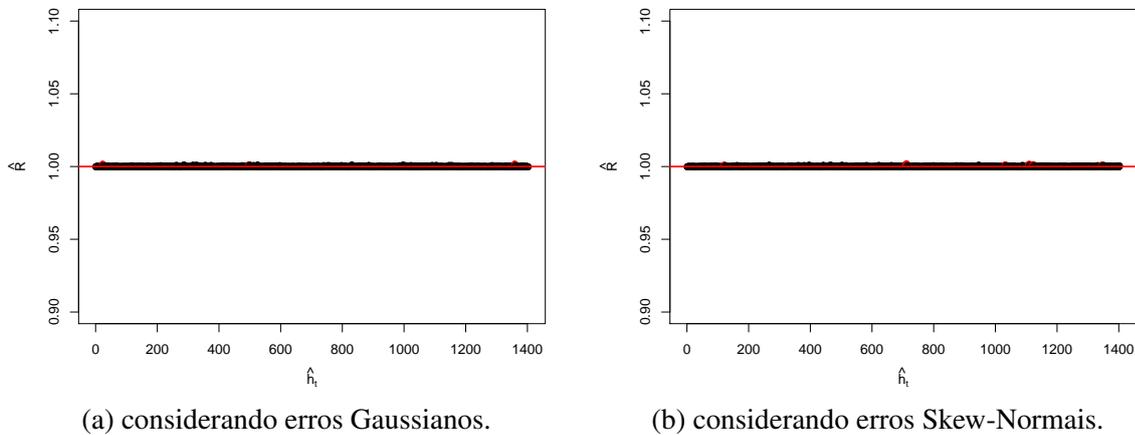
Figura 32 – Medida  $\hat{R}$  para as log-volatilidades da série do IBOVESPA.



De acordo com as Figuras 32a e 32b, é possível observar que a maioria das log-volatilidades é de aproximadamente 1, dando, novamente, indícios de uma boa convergência. Os pontos em destaque, foram os que ficaram com valores pouco maiores que 1,003. No caso do modelo considerando erros Gaussianos, 48 pontos ficaram com valores maiores que 1,002, sendo que destes, 10 ( $h_{200}$ ,  $h_{201}$ ,  $h_{202}$ ,  $h_{532}$ ,  $h_{807}$ ,  $h_{808}$ ,  $h_{809}$ ,  $h_{811}$ ,  $h_{1631}$  e  $h_{1632}$ ) obtiveram va-

lores de  $\hat{R}$  pouco maiores que 1,003. Já para o modelo considerando erros Skew-Normais, das log-volatilidades simuladas apenas uma apresenta valor maior que 1,002 ( $h_{510}$ ).

Figura 33 – Medida  $\hat{R}$  para as log-volatilidades da série do S&P500.



As Figuras 33a e 33b mostram que a maioria das log-volatilidades é de aproximadamente 1, dando indícios de uma boa convergência. Os pontos em destaque, foram os que ficaram com valores pouco maiores que 1,001, no entanto, menores que 1,002. No caso do modelo considerando erros Gaussianos os pontos são  $h_{23}$ ,  $h_{497}$  e  $h_{1359}$ , já para o modelo considerando erros Skew-Normais as log-volatilidades simuladas que ficaram em destaque são  $h_{120}$ ,  $h_{710}$ ,  $h_{711}$ ,  $h_{1031}$ ,  $h_{1108}$ ,  $h_{1120}$ ,  $h_{1122}$  e  $h_{1345}$ .

Foi considerada uma amostra das log-volatilidades simuladas para a série do IBOVESPA, englobando os casos em que o critério de convergência  $\hat{R}$ , proposto por Gelman e Rubin (1992), obtiveram valores pouco maiores que 1,003, no caso da distribuição Gaussiana, e pouco maiores que 1,001, no caso em que os erros seguem uma distribuição Skew-Normal, para analisar graficamente as convergências. Os gráficos de convergência para as log-volatilidades considerando erros Gaussianos é apresentado na Figura 34 e para erros Skew-Normais é apresentado na Figura 35, em que é possível observar que os valores simulados apresentam baixa estrutura de correlação e reforçam os indícios de uma boa convergência, não somente dos parâmetros do modelo como também das log-volatilidades.

Assim como para a série do IBOVESPA, foi considerada uma amostra das log-volatilidades simuladas na série do S&P500, englobando os casos em que o critério de convergência  $\hat{R}$  proposto por Gelman e Rubin (1992) obtiveram valores pouco maiores que 1, para analisar graficamente as convergências. Os gráficos de convergência para as log-volatilidades considerando erros Gaussianos é apresentado na Figura 36 e para erros Skew-Normais é apresentado na Figura 37. É possível observar que os valores simulados apresentam baixa estrutura de correlação e reforçam os indícios de uma boa convergência, assim como no caso da série do IBOVESPA, não somente dos parâmetros do modelo como também das log-volatilidades.

Figura 34 – Gráfico de convergência para uma amostra de log-volatilidades a *posteriori* estimadas para a série do IBOVESPA, com erros Gaussianos.

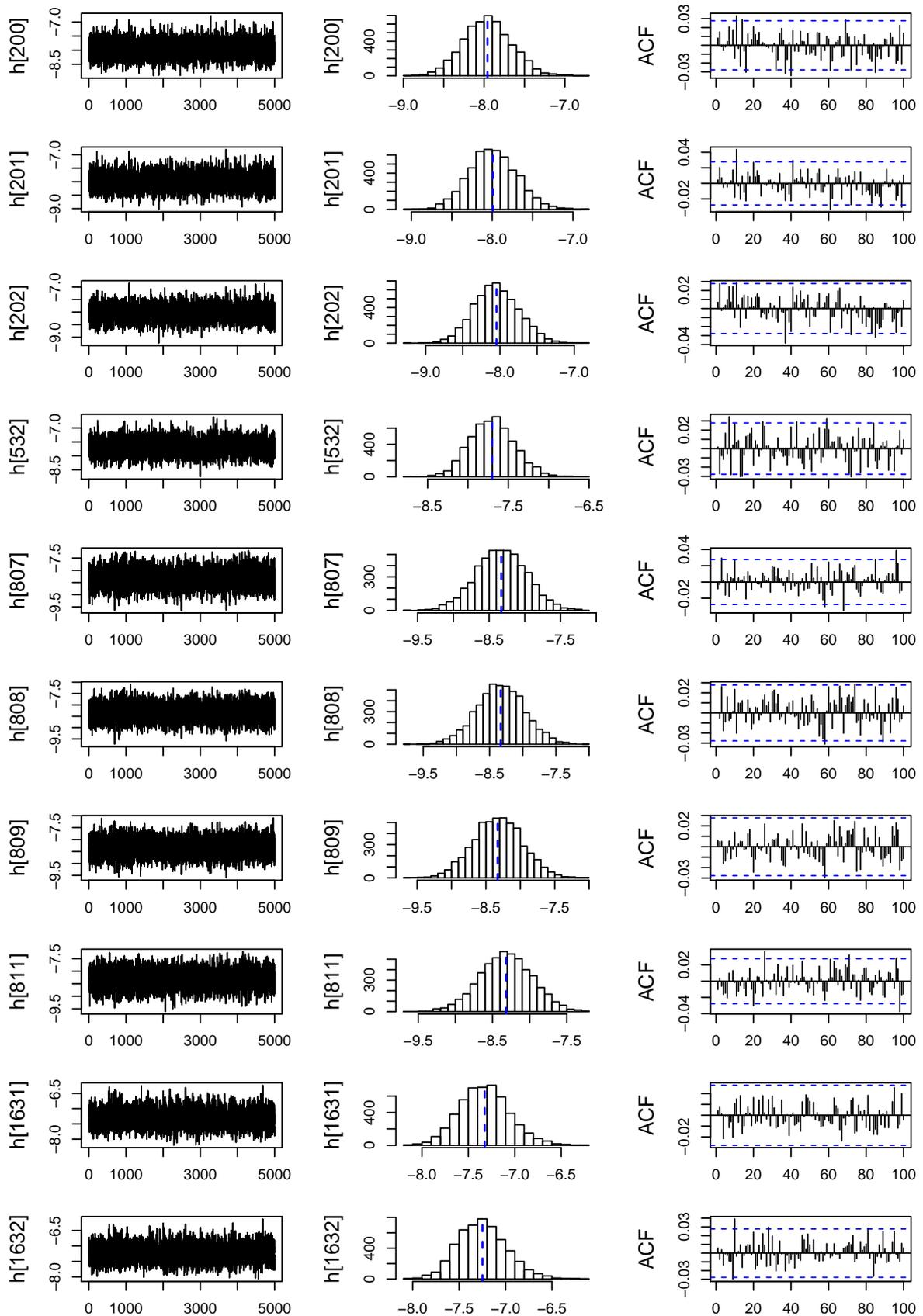


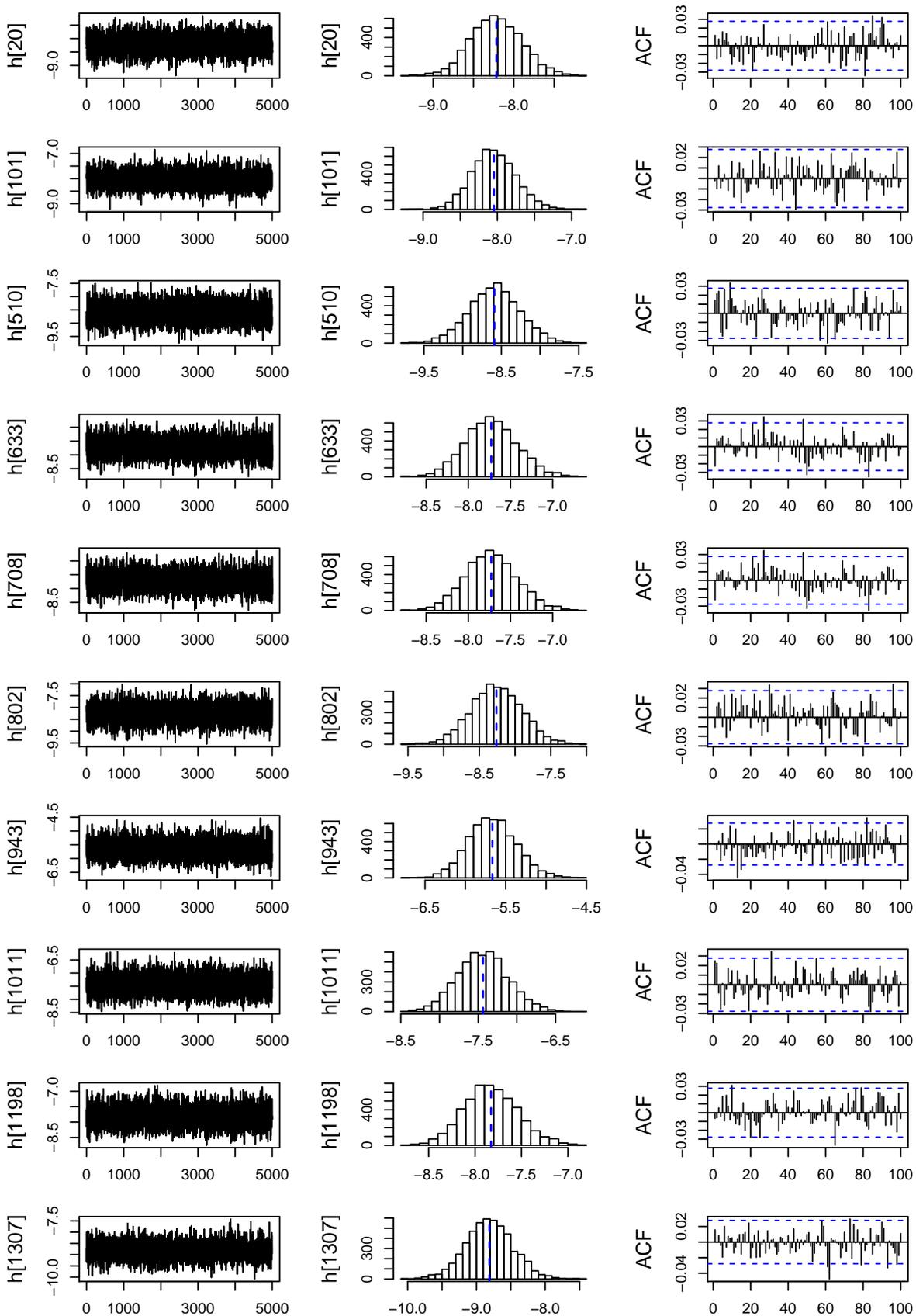
Figura 35 – Gráfico de convergência para uma amostra de log-volatilidades a *posteriori* estimadas para a série do IBOVESPA, com erros Skew-Normais.

Figura 36 – Gráfico de convergência para uma amostra de log-volatilidades a *posteriori* estimadas para a série do S&P500, com erros Gaussianos.

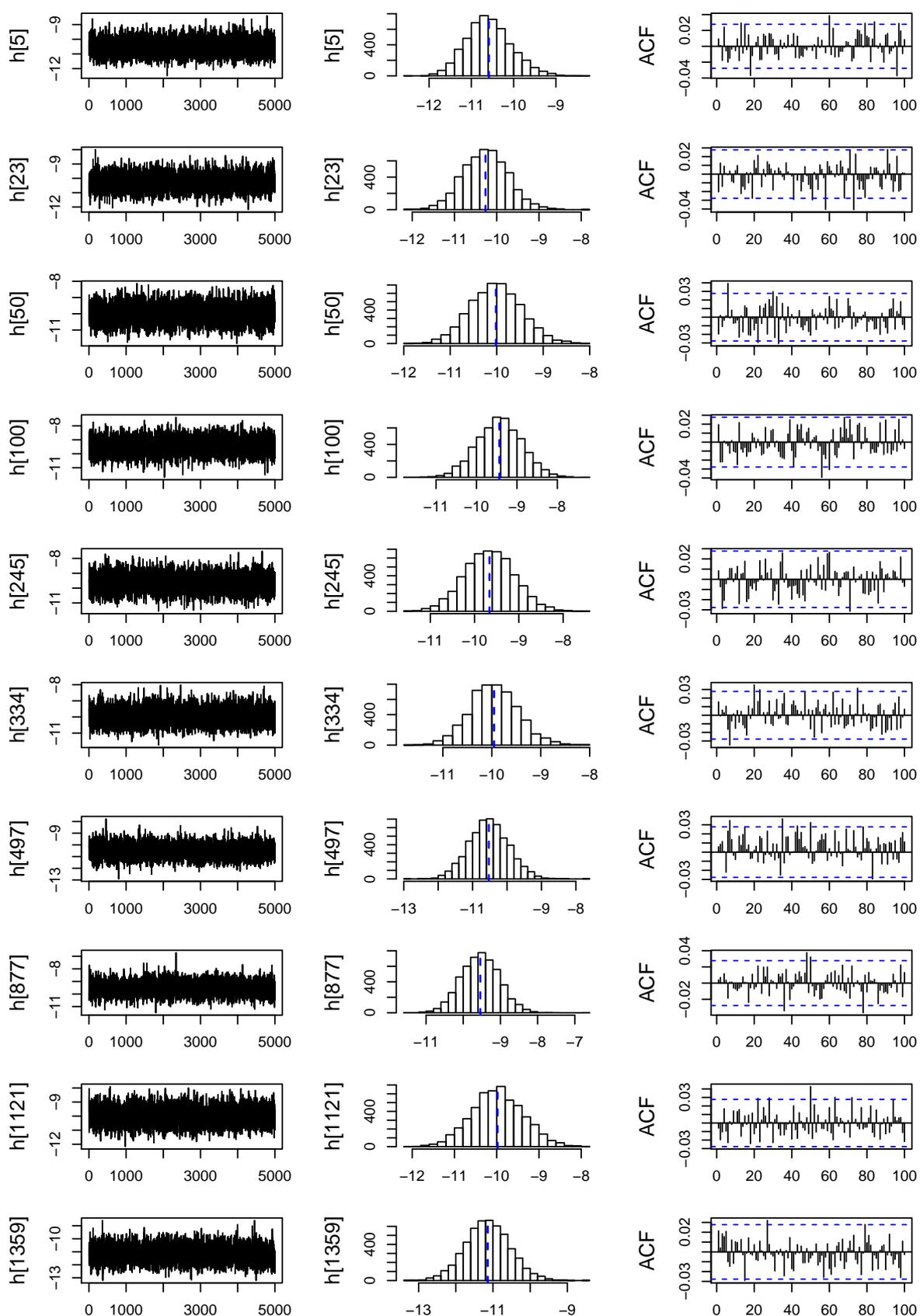
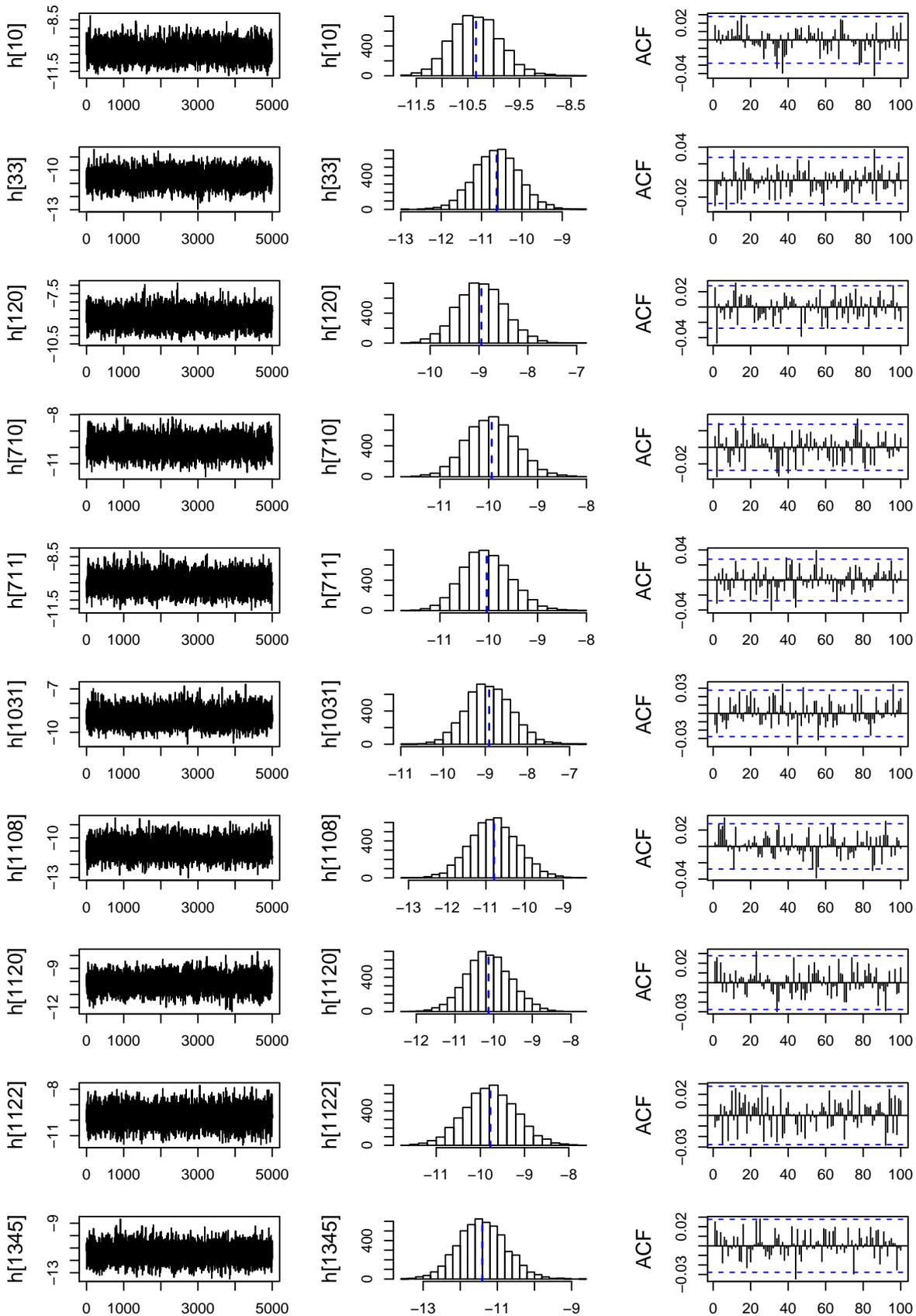


Figura 37 – Gráfico de convergência para uma amostra de log-volatilidades a *posteriori* estimadas para a série do S&P500, com erros Skew-Normais.

## A.2 Análise da série S&P500

Apesar de ter analisado não somente a convergência dos parâmetros do modelo, mas também das log-volatilidade da série do S&P500, fizemos mais uma análise, no entanto, aumentando o número de simulações, pois as volatilidades estimadas não apresentaram uma forma suavizada como para as demais aplicações. Além disso, verificamos se ao aumentar o número de simulações a convergência dos parâmetros e log-volatilidades, assim como a escolha da distribuição dos erros seria mantida. Os resultados estão descritos abaixo:

Tabela 16 – Comparação entre os modelos propostos via critérios de seleção para série do S&P500.

Dist.	DIC	WAIC	EP <sub>waic</sub>	LOO	EP <sub>loo</sub>
(1) Gaussiana	-10016.07	-10014.3	65.3	-9978.0	68.1
(2) t-Student	-99880.24	-9979.4	68.2	-9970.2	68.8
(3) Skew-Normal	-10015.53	-10014.1	65.4	-9977.5	68.2
(4) GED	-99841.07	-9988.7	67.6	-9974.9	68.7

De acordo com a Tabela 16, mesmo ao aumentar o número de simulações, os critérios de seleção continuam favorecendo a escolha das distribuições Gaussiana e Skew-Normal. As estimativas obtidas, considerando 50mil simulações e descartando 50% como *burn-in*, são mostradas na Tabela 17. Nota-se que as médias a *posteriori* apresentadas continuam próximas aos valores que foram obtidos considerando 10mil simulações.

Tabela 17 – Estimativas das médias (desvios padrões) dos parâmetros a *posteriori*.

Dist.	$\mu$	$\phi$	$\sigma_{\eta}$	$\nu$
(1)	-10.1092 (0.1185)	0.9066 (0.0224)	0.3772 (0.0478)	
(2)	-10.2654 (0.1402)	0.9307 (0.0197)	0.3092 (0.0467)	10.2869 (2.1799)
(3)	-10.1103 (0.1196)	0.9054 (0.0229)	0.3813 (0.0494)	0.0362 (0.0341)
(4)	-10.0503 (0.1321)	0.9305 (0.0208)	0.3022 (0.0509)	1.5769 (0.1253)

As Figuras 38 e 39 apresentam os gráficos de convergência das distribuições a *posteriori* simuladas para os parâmetros dos modelos, com erros Gaussianos e Skew-Normais, respectivamente. Nota-se que os gráficos apresentam uma baixa estrutura de correlação, além de apresentar uma boa concentração dos valores no gráfico de traços, dando indícios de uma boa convergência.

As Figuras 40a e 40b são mostradas as log-volatilidades simuladas a *posteriori*, nos casos dos modelos com erros seguindo as distribuições Gaussiana e Skew-Normal, respectivamente. No caso do modelo considerando erros Gaussianos alguns pontos possui valores pouco maiores que 1,001 e menores que 1,002, sendo estes  $h_{151}$ ,  $h_{152}$ ,  $h_{153}$ ,  $h_{154}$ ,  $h_{155}$ ,  $h_{231}$ ,  $h_{232}$ ,  $h_{233}$ ,  $h_{234}$ ,  $h_{250}$ ,  $h_{286}$ ,  $h_{910}$ ,  $h_{979}$ ,  $h_{980}$ ,  $h_{981}$ ,  $h_{982}$ ,  $h_{1126}$  e  $h_{1174}$ . Já para o modelo considerando erros Skew-Normais as log-volatilidades não apresentam nenhum valor maior que 1,001. Entretanto, observa-se que a maioria das log-volatilidades apresenta critério  $\hat{R}$  de aproximadamente 1.

Figura 38 – Gráficos de convergência dos parâmetros na série de retornos S&P500 com erros Gaussianos, considerando 50mil simulações.

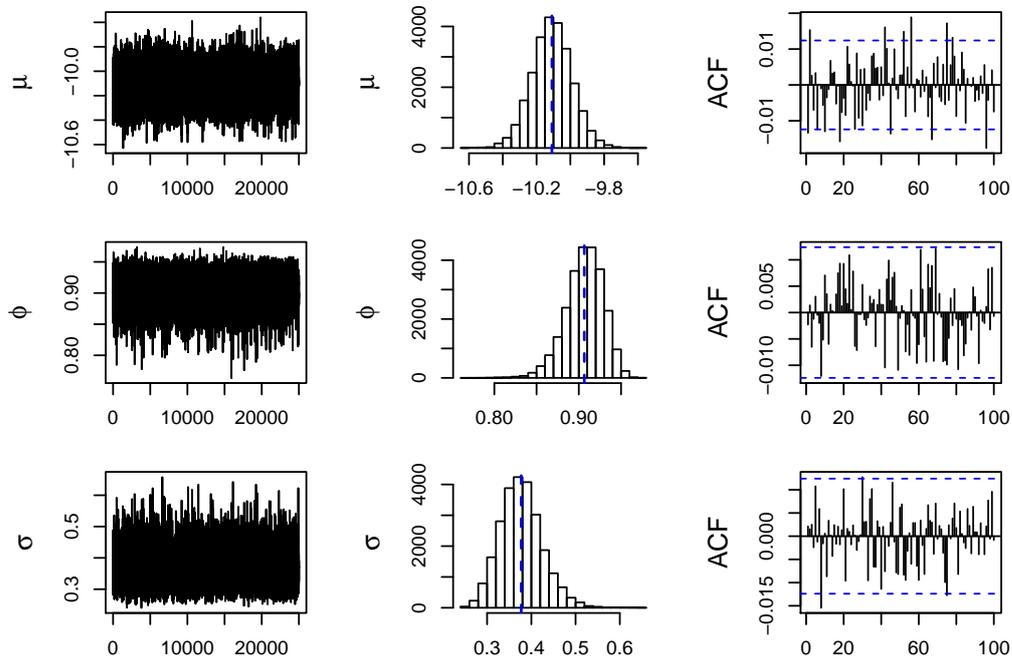


Figura 39 – Gráficos de convergência dos parâmetros na série de retornos S&P500 com erros Skew-Normais, considerando 50mil simulações.

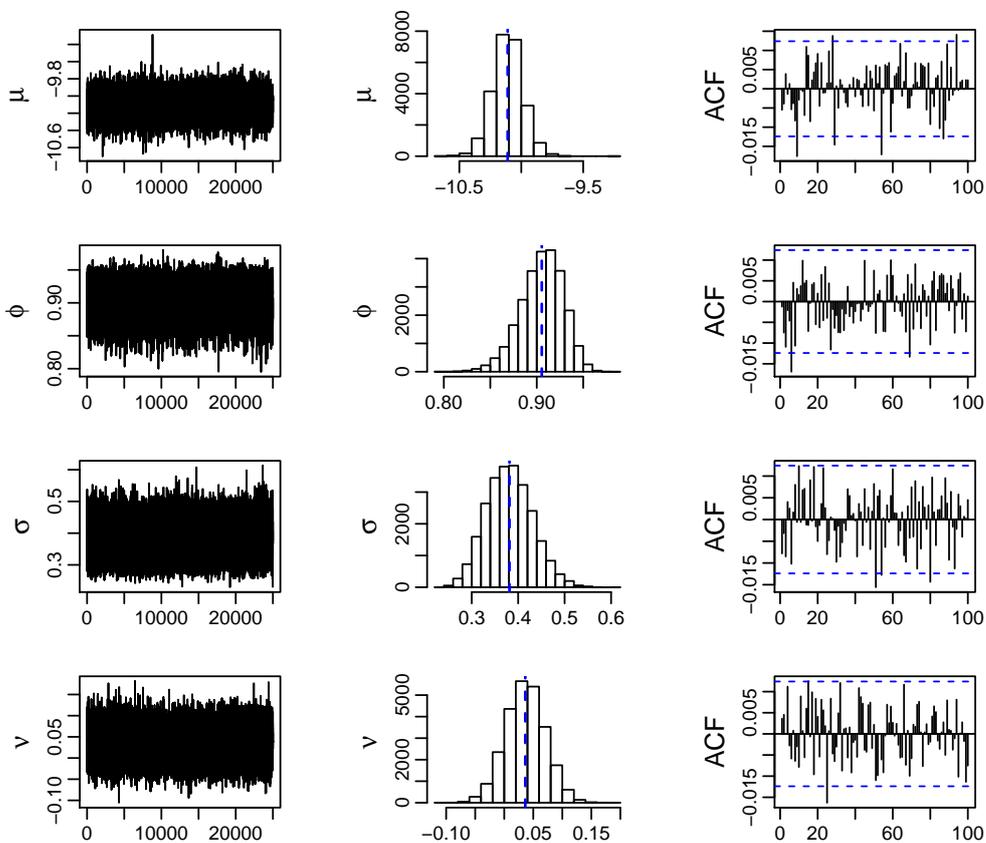


Figura 40 – Medida  $\hat{R}$  para as log-volatilidades da série do S&P500, considerando 50mil simulações.

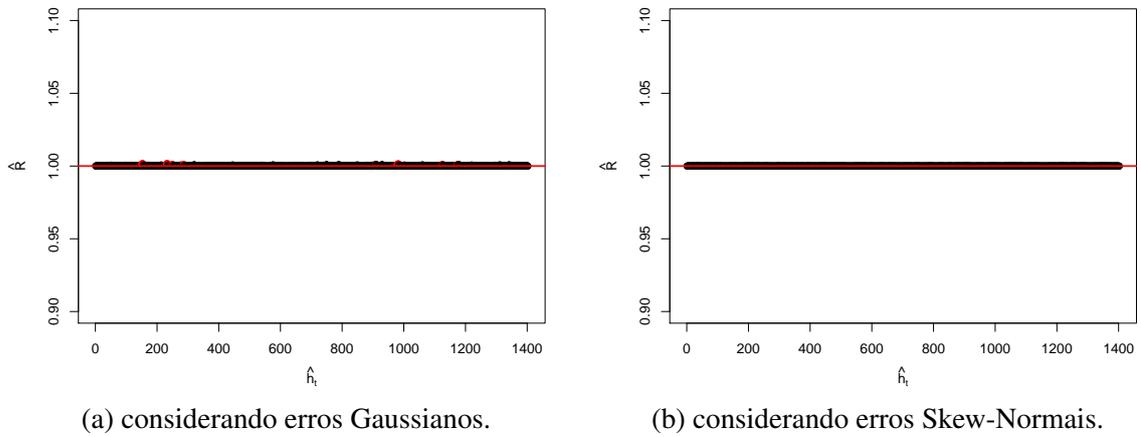
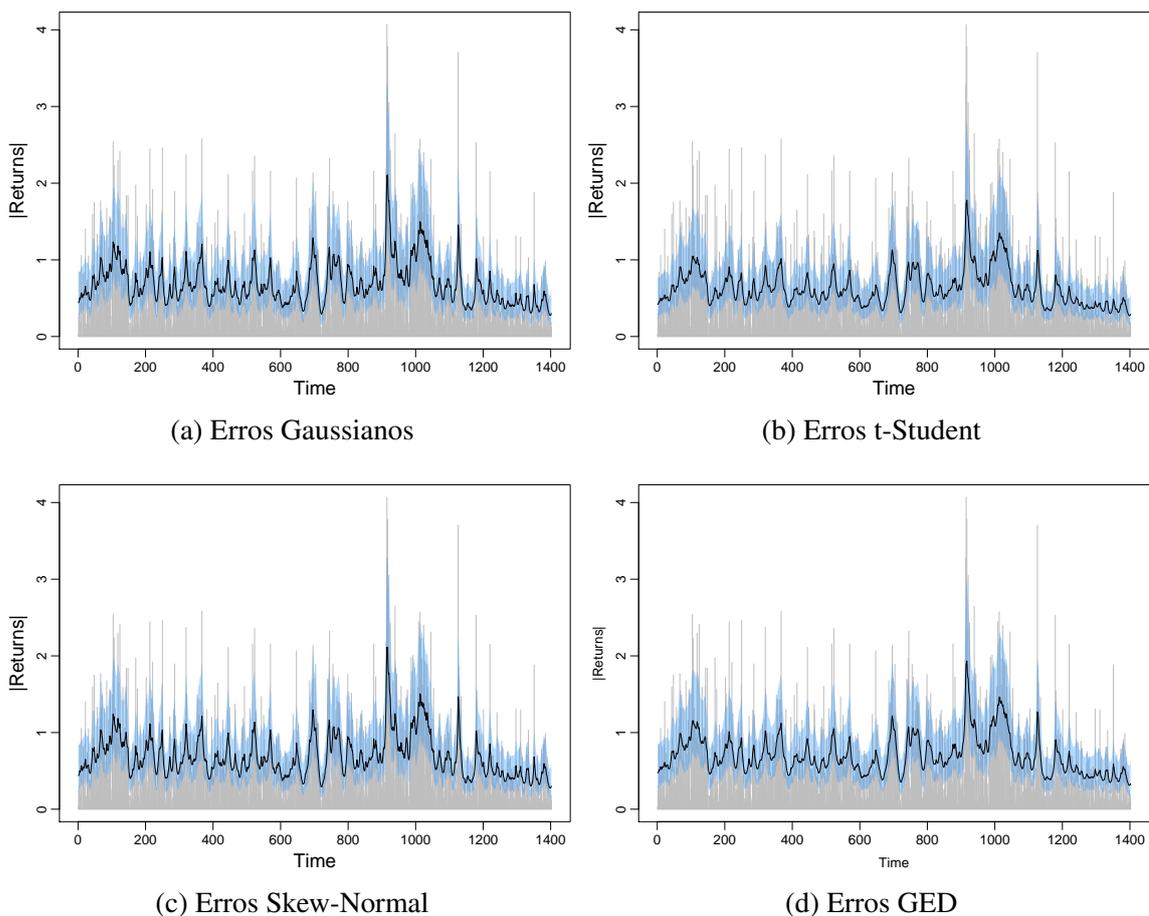


Figura 41 – Retornos absolutos em percentual para a série £/USD e volatilidades estimadas com bandas de credibilidade de 95%.



A Figura 41 são apresentadas as médias *a posteriori* das volatilidades estimadas com seus intervalos de credibilidade, considerando todas as distribuições para os erros. Nota-se que os valores apresentados considerando as distribuições Gaussiana e Skew-Normal são muito próximos e mostram comportamento similar aos gráficos estimados considerando 10mil simulações.

## **A.3 Códigos Computacionais**

Disponibilizamos os códigos computacionais utilizados na dissertação no GitHub, através do seguinte endereço: <<https://github.com/ddias7/svmodelsrstan>>

