



Universidade de Brasília  
Instituto de Ciências Exatas  
Departamento de Estatística

Dissertação de Mestrado

Um Estudo sobre Modelos para Volatilidade  
Estocástica

por

André Silva de Queiroz

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Cibele Queiroz da Silva  
Dezembro de 2015

André Silva de Queiroz

# Um Estudo sobre Modelos para Volatilidade Estocástica

Dissertação apresentada ao Departamento de  
Estatística do Instituto de Ciências Exatas  
da Universidade de Brasília como requisito  
parcial à obtenção do título de Mestre em  
Estatística.

Universidade de Brasília  
Brasília, Dezembro de 2015

TERMO DE APROVAÇÃO

André Silva de Queiroz

## UM ESTUDO SOBRE MODELOS PARA VOLATILIDADE ESTOCÁSTICA

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Data da defesa: 10 de Dezembro de 2015

Orientadora:

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Cibele Queiroz da Silva  
Departamento de Estatística, UnB

Comissão Examinadora:

---

Prof. Dr. Raul Yukihiro Matsushita  
Departamento de Estatística, UnB

---

Prof. Dr. Jorge Alberto Achcar  
Departamento de Medicina Social, FMRP-USP

Brasília, Dezembro de 2015

## Ficha Catalográfica

**DE QUEIROZ, ANDRÉ SILVA**

Um Estudo sobre Modelos para Volatilidade Estocástica, (UnB - IE, Mestre em Estatística, 2015).

Dissertação de Mestrado - Universidade de Brasília. Departamento de Estatística - Instituto de Ciências Exatas.

1. Estatística Bayesiana 2. Modelo de Volatilidade Estocástica 3. Modelos Lineares Dinâmicos 4. Ancillarity-Sufficiency Interweaving Strategy 5. JAGS

É concedida à Universidade de Brasília a permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

André Silva de Queiroz

*À efêmera jornada humana no Universo.*

# Agradecimentos

Agradeço a Deus, o Criador de todas as coisas e detentor de toda ciência, pelo dom da vida.

Agradeço aos meus pais, Manoel e Sandra, meus principais exemplos, por cuidar de mim, me amar e me ensinar o caminho a ser seguido.

Agradeço à Sarah, meu amor, por ser minha companheira e melhor amiga, a razão de eu levantar todas as manhãs.

Agradeço aos meus amigos, Bernardo, Leonardo e Tomaz, por, além de todas as alegrias compartilhadas, preencherem, também, a lacuna destinada aos meus irmãos.

Agradeço ao Billy e a Geneviève, por passar em minha vida, e me mostrar que os anseios humanos não são tão importante como pensamos.

Por fim, agradeço à Cibele, minha orientadora, por me mostrar pacientemente como é produzido o conhecimento, e me motivar a sempre querer saber hoje, algo a mais do que ontem.

Todos vocês, estarão sempre em minhas memórias, enquanto durar a minha jornada.

# Sumário

<b>Lista de Figuras</b>	<b>4</b>
<b>Lista de Tabelas</b>	<b>5</b>
<b>Lista de Códigos</b>	<b>6</b>
<b>Resumo</b>	<b>7</b>
<b>Abstract</b>	<b>8</b>
<b>Introdução</b>	<b>9</b>
<b>1 Modelo Linear Dinâmico</b>	<b>11</b>
1.1 Definição do Modelo . . . . .	11
1.2 Breve Resumo do Teorema de Bayes . . . . .	14
1.3 Equações de Atualização . . . . .	15
1.4 Previsões . . . . .	16
<b>2 Modelo de Volatilidade Estocástica</b>	<b>18</b>
2.1 Definição do Modelo Canônico . . . . .	18
2.2 Um Modelo de Espaço-Estado . . . . .	19
2.2.1 Equação das Observações . . . . .	19
2.2.2 Equação do Sistema . . . . .	21
2.2.3 Modelo Completo . . . . .	21
2.3 Modelo Parcialmente Não Centralizado . . . . .	23

2.4	Modelo Não Centralizado . . . . .	24
<b>3</b>	<b>Proposta de Estimação dos Parâmetros do MVE</b>	<b>27</b>
3.1	Definição do Modelo Bayesiano . . . . .	28
3.2	Estimando $\mu$ , $\phi$ e $\sigma_\eta^2$ . . . . .	29
3.2.1	Amostrando $\sigma_\eta^2$ . . . . .	31
3.2.2	Amostrando $\phi$ . . . . .	32
3.2.3	Amostrando $\mu$ . . . . .	33
3.2.4	Reamostrando $\sigma_\eta^2$ . . . . .	34
3.2.5	Reamostrando $\mu$ . . . . .	35
3.3	Estimando $h_{1:N}$ . . . . .	36
3.3.1	O Método de McCormick et al. (2012) . . . . .	37
3.3.2	Adaptação do Método de McCormick et al. (2012) ao MVE . . . . .	39
<b>4</b>	<b>Aplicação da Metodologia Proposta</b>	<b>42</b>
4.1	Definição dos Dados Simulados . . . . .	42
4.2	Estimação dos Parâmetros para os Dados Simulados . . . . .	43
4.3	Comparação dos Resultados Obtidos com o JAGS . . . . .	62
4.4	Um Exemplo com Dados Reais . . . . .	70
	<b>Conclusão</b>	<b>73</b>
	<b>A Códigos-Fonte</b>	<b>79</b>



# Lista de Figuras

1.1	Exemplo simulado de 100 observações geradas a partir de um modelo polinomial de primeira ordem. . . . .	13
2.1	Comparação das distribuições de $\ln \delta_t^2$ , $\varepsilon_t$ e $\xi_t$ . . . . .	21
4.1	Conjunto de dados simulados a partir do modelo canônico em (2.3), $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,5^2$ . . . . .	43
4.2	Valores médios estimados de $h_t$ do modelo (2.3) quando os parâmetros $\mu$ , $\phi$ e $\sigma_\eta^2$ são conhecidos ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,5^2$ ). . . . .	45
4.3	Valores médios estimados de $h_t$ do modelo (2.3) quando apenas o parâmetro $\phi$ é conhecido ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,5^2$ ). . . . .	46
4.4	Valores médios estimados de $h_t$ do modelo (2.3) quando o valor inicial de $\phi$ é igual ao valor verdadeiro do parâmetro ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,5^2$ ). . . . .	46
4.5	Valores médios estimados de $h_t$ do modelo (2.3) quando $\lambda_t = 0,50$ ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,5^2$ ). . . . .	47
4.6	Valores estimados ( $\mu = -5,4$ , $\phi = 0,99$ e $\sigma_\eta^2 = 0,5^2$ ). . . . .	50
4.7	Valores estimados ( $\mu = -5,4$ , $\phi = 0,99$ e $\sigma_\eta^2 = 0,3^2$ ). . . . .	51
4.8	Valores estimados ( $\mu = -5,4$ , $\phi = 0,99$ e $\sigma_\eta^2 = 0,1^2$ ). . . . .	52
4.9	Valores estimados ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,5^2$ ). . . . .	53
4.10	Valores estimados ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,3^2$ ). . . . .	54
4.11	Valores estimados ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,1^2$ ). . . . .	55
4.12	Valores estimados ( $\mu = -5,4$ , $\phi = 0,80$ e $\sigma_\eta^2 = 0,5^2$ ). . . . .	56
4.13	Valores estimados ( $\mu = -5,4$ , $\phi = 0,80$ e $\sigma_\eta^2 = 0,3^2$ ). . . . .	57
4.14	Valores estimados ( $\mu = -5,4$ , $\phi = 0,80$ e $\sigma_\eta^2 = 0,1^2$ ). . . . .	58

4.15	Valores estimados ( $\mu = -5,4$ , $\phi = 0,50$ e $\sigma_\eta^2 = 0,5^2$ ).	59
4.16	Valores estimados ( $\mu = -5,4$ , $\phi = 0,50$ e $\sigma_\eta^2 = 0,3^2$ ).	60
4.17	Valores estimados ( $\mu = -5,4$ , $\phi = 0,50$ e $\sigma_\eta^2 = 0,1^2$ ).	61
4.18	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,99$ e $\sigma_\eta^2 = 0,5^2$ ).	64
4.19	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,99$ e $\sigma_\eta^2 = 0,3^2$ ).	64
4.20	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,99$ e $\sigma_\eta^2 = 0,1^2$ ).	65
4.21	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,5^2$ ).	65
4.22	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,3^2$ ).	66
4.23	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,90$ e $\sigma_\eta^2 = 0,1^2$ ).	66
4.24	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,80$ e $\sigma_\eta^2 = 0,5^2$ ).	67
4.25	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,80$ e $\sigma_\eta^2 = 0,3^2$ ).	67
4.26	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,80$ e $\sigma_\eta^2 = 0,1^2$ ).	68
4.27	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,50$ e $\sigma_\eta^2 = 0,5^2$ ).	68
4.28	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,50$ e $\sigma_\eta^2 = 0,3^2$ ).	69
4.29	Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ , $\phi = 0,5$ e $\sigma_\eta^2 = 0,1^2$ ).	69
4.30	Valores estimados do MVE aplicado aos dados reais.	71
4.31	Valores estimados via JAGS vs método proposto aplicados aos dados reais.	72

# Lista de Tabelas

2.1	Parâmetros da distribuição de $\xi_t$ em (2.9), (Omori et al., 2007). . . . .	20
3.1	Métodos de estimação dos parâmetros do modelo de volatilidade estocástica e principais referências, (Bos, 2012). . . . .	27
4.1	Estatísticas dos valores estimados dos parâmetros. . . . .	49
4.2	Estatísticas dos valores estimados dos parâmetros do MVE aplicado aos dados reais. . . . .	72

# Lista de Códigos

A.1	Função Responsável por Gerar os Dados Simulados. . . . .	79
A.2	Implementação da Proposta de Estimação dos Parâmetros do Modelo de Volatilidade Estocástica. . . . .	79
A.3	Modelo Canônico de Volatilidade Estocástica em JAGS definido em (2.3). .	85

# Resumo

O modelo de volatilidade estocástica (Kim et al., 1998) constitui uma classe de modelos bastante importante, pois visa ajustar séries de dados temporais cuja variabilidade é aleatória. Tal modelo tem sido abordado tanto do ponto de vista clássico, quanto Bayesiano. Porém, várias de suas especificações ainda carecem de uma solução mais adequada.

O recente trabalho de Kastner e Frühwirth-Schnatter (2014) apresentou desenvolvimentos importantes quanto ao método Bayesiano de estimação dos parâmetros do modelo. Os autores apresentam uma ideia bastante inovadora denominada, em inglês, de *Ancillarity-Sufficiency Interweaving Strategy*, que consiste em alternar as parametrizações do modelo durante o processo de estimação. Entretanto, a proposta de estimação da variável latente, que rege o modelo, é um pouco complicada. Com o presente trabalho de dissertação propõe-se uma forma alternativa, e mais simples, de se estimar a variável latente do processo baseada no trabalho de McCormick et al. (2012).

Nesta dissertação, foram adaptadas as metodologias propostas pelos dois últimos autores de modo a sugerir uma forma alternativa de estimar os parâmetros do modelo de volatilidade estocástica. Utilizando dados simulados, foi avaliada a efetividade da nova proposta. Ao final desse trabalho foram destacados os problemas emergentes do procedimento proposto e sugeridas algumas novas alternativas para resolvê-los.

**Palavras Chave:** *Estatística Bayesiana, Modelo de Volatilidade Estocástica, Modelos Lineares Dinâmicos, Ancillarity-Sufficiency Interweaving Strategy, JAGS*

# Abstract

The stochastic volatility model (Kim et al., 1998) is a very important class of models, because it fits time series data with random variability. This model has been studied through the classical point of view as through the Bayesian one. Nevertheless, its many specifications still needs an adequated solution.

The recent work of Kastner e Frühwirth-Schnatter (2014) presented important developments of the Bayesian approach for the estimation of the model parameters. The authors introduce a very innovative idea called Ancillarity-Sufficiency Interweaving Strategy, which consists in switching the model's parametrization during the estimation process. However, the proposed estimation of the latent variable, which governs the model, is a bit tricky. So, the present work proposes an alternative way, also simpler, to estimate the latent variable of the process based on the work of McCormick et al. (2012).

The methodologies proposed by the last two authors were adapted in this dissertation to suggest an alternative way to estimate the parameters of the stochastic volatility model. The effectiveness of the new proposal was evaluated by using simulated data. The emerging issues of the proposed procedure were highlighted and some new alternatives to solve them were suggested at the end of this work.

**Key Words:** *Bayesian Statistics, Stochastic Volatility Models, Dynamic Linear Models, Ancillarity-Sufficiency Interweaving Strategy, JAGS*

# Introdução

Modelos de dados temporais são de extrema importância para a sociedade, seja para entender um processo natural ou um fenômeno econômico. Sua aplicação adequada é fundamental para gerar boas estimativas do comportamento do sistema. Afinal, predição de valores futuros é um dos principais objetivos de um modelo desse tipo (Brockwell e Davis, 2002). Eventualmente, o conjunto de dados de interesse pode apresentar uma variância distinta ao longo do tempo. Os modelos ARCH e GARCH são as soluções tradicionais para esse tipo de situação. Porém, tratam a variância de forma determinística. Em um panorama mais geral, a variância ou volatilidade dos dados pode ser considerada de natureza estocástica. Assim, surgem os modelos de volatilidade estocástica, que é o objeto principal de estudo nesse trabalho.

A inferência Bayesiana tem se mostrado uma excelente ferramenta para análise de dados. Como cita Pearl (2009), ela se conecta perfeitamente com a maneira humana de raciocínio diante da incerteza. Pensar em problemas sob essa perspectiva num contexto de séries temporais é possível através dos modelos dinâmicos.

No Capítulo 1, é introduzida uma pequena apresentação dos modelos lineares dinâmicos com seus conceitos básicos, como em West e Harrison (1997). No Capítulo 2, o modelo de volatilidade estocástica (MVE) é definido em sua parametrização canônica, e também são mostradas algumas parametrizações alternativas. No Capítulo 3 é apresentada uma proposta de estimação dos parâmetros do MVE que busca simplificar a proposta de Kastner e Frühwirth-Schnatter (2014). Para tanto, os estados latentes do MVE são estimados através de uma adaptação às ideias propostas por McCormick et al. (2012), desenvolvida também no Capítulo 3. No Capítulo 4, a proposta é avaliada através de um

conjunto de dados simulados e, posteriormente, comparada com um processo de estimação baseado no amostrador de Gibbs. Por fim, apresenta-se uma aplicação, com dados reais, inspirada em Achcar et al. (2011). São identificados os desafios e problemas que eventualmente surgiram no árduo processo de estimação dos parâmetros do modelo de volatilidade estocástica.



# Capítulo 1

## Modelo Linear Dinâmico

O uso de modelos lineares dinâmicos (MLD) cresceu bastante nas últimas décadas. Sua aplicabilidade tem se estendido nos campos da biologia, genética, geofísica, economia entre outros, (Petrís et al., 2009). O desenvolvimento computacional das técnicas Bayesianas de estimação dos parâmetros é um dos fatores que impulsionou esse crescimento recente.

Os MLD constituem uma família muito importante, pois, além de trazerem os modelos tradicionais de séries de tempo para a abordagem Bayesiana, expandem a gama de possibilidades de aplicações através da sua estrutura flexível e do próprio paradigma Bayesiano em si. São, ainda, um caso particular linear e Gaussiano dos modelos mais gerais de espaço-estado.

### 1.1 Definição do Modelo

O modelo linear dinâmico mais geral é definido em relação a um vetor de observações  $\mathbf{Y}_t$ . Porém, para efeitos de simplificação, será considerado o caso univariado  $Y_t$  como em West e Harrison (1997). Portanto, sejam, para os tempos  $t = 1, \dots, N$ :

- $Y_t$  uma variável observada de interesse;
- $\boldsymbol{\theta}_t$  um vetor latente de variáveis que representam o sistema gerador de  $Y_t$ ; e
- $D_t$  toda informação disponível a respeito do sistema.

Desse modo, para todo tempo  $t$ , o modelo linear dinâmico é descrito por:

$$\text{Equação das Observações:} \quad Y_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, V_t), \quad (1.1)$$

$$\text{Equação do Sistema:} \quad \boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t). \quad (1.2)$$

A quádrupla  $\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$  é composta pelas matrizes que caracterizam o modelo.  $\mathbf{F}_t$  é a matriz de *design*, ou o vetor de regressão, no caso univariado.  $\mathbf{G}_t$  é a matriz de evolução do sistema e através dela é possível a especificação de alguns efeitos nos estados latentes, tais como nível, tendência e sazonalidade.  $V_t$  é a variância observacional e  $\mathbf{W}_t$  é a matriz de variância da evolução do sistema. Quando esses elementos são invariantes no tempo, o modelo  $\{\mathbf{F}, \mathbf{G}, V, \mathbf{W}\}$  é chamado de constante, e engloba, essencialmente, todos os modelos lineares tradicionais de séries temporais.

Os termos  $\nu_t$  e  $\boldsymbol{\omega}_t$  são os erros associados às equações (1.1) e (1.2). São denominados erro observacional e erro do sistema, respectivamente. Por definição, eles são independentes, no tempo e entre si, e assumem distribuição normal com média zero. Modelos ainda mais gerais podem ser definidos com  $\nu_t$  e  $\boldsymbol{\omega}_t$  autocorrelacionados e correlacionados entre si. Entretanto, West e Harrison (1997) destaca que tais modelos sempre podem ser reescritos em termos mais simples, satisfazendo as condições de independência.

A informação inicial sobre  $\boldsymbol{\theta}_t$  é representada por:

$$\text{Informação Inicial:} \quad (\boldsymbol{\theta}_0 | D_0) \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0), \quad (1.3)$$

em que  $\mathbf{m}_0$  e  $\mathbf{C}_0$  são o vetor de médias e a matriz de variâncias e covariâncias da distribuição proposta, respectivamente.

A Figura 1.1 ilustra um exemplo de 100 observações simuladas a partir de um modelo linear dinâmico constante definido pela quádrupla:

$$\{\mathbf{F}, \mathbf{G}, V, \mathbf{W}\} = \left\{ \begin{bmatrix} 1 \\ \pi \end{bmatrix}, \begin{bmatrix} 0,15 & -0,08 \\ -0,08 & 0,15 \end{bmatrix}, 1, \begin{bmatrix} 1 & 0,3 \\ 0,3 & 1 \end{bmatrix} \right\}. \quad (1.4)$$

A série partiu da distribuição inicial:

$$(\boldsymbol{\theta}_0|D_0) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}\right). \quad (1.5)$$

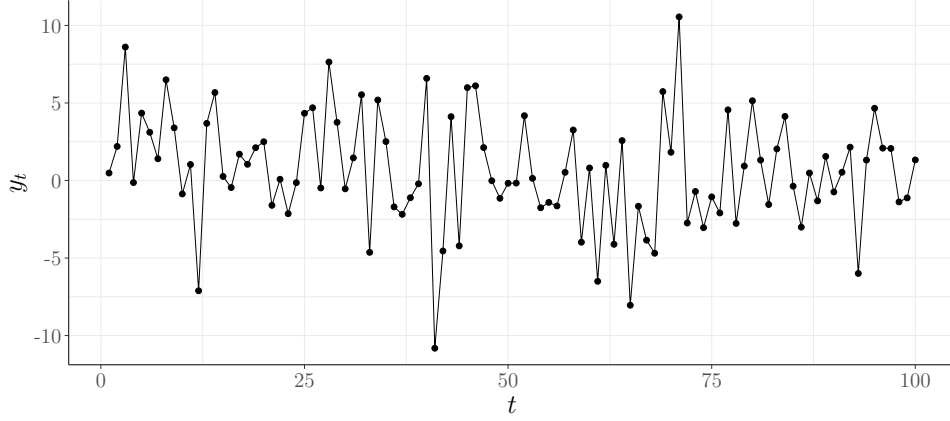


Figura 1.1: Exemplo simulado de 100 observações geradas a partir de um modelo polinomial de primeira ordem.

Os valores dos parâmetros nesse exemplo foram selecionados arbitrariamente, apenas tomando-se o cuidado para que não fossem muito grandes de modo a gerar observações numa escala elevada.

Dentre as finalidades de um modelo de dados temporais está a previsão de observações futuras. Sob esse aspecto, o objetivo a ser alcançado agora é definir as distribuições de probabilidade da previsão de  $Y$  no tempo  $t$  dado todo o conhecimento anterior a esse tempo ( $Y_t|D_{t-1}$ ) e da distribuição do vetor  $\boldsymbol{\theta}$  no tempo  $t$  dado todo conhecimento disponível até então ( $\boldsymbol{\theta}_t|D_t$ ). No entanto, tais desenvolvimentos demandam o uso da Inferência Bayesiana e, dessa forma, serão apresentadas algumas noções fundamentais sobre esse tipo de inferência a seguir.

## 1.2 Breve Resumo do Teorema de Bayes

Sejam dois eventos distintos  $A$  e  $B$ . A probabilidade de ocorrência conjunta deles é dada pela regra do produto:

$$P(A \cap B) = P(A|B)P(B). \quad (1.6)$$

Naturalmente, a ordem dos eventos pode ser mudada, isto é,  $P(B \cap A) = P(B|A)P(A)$ . A partir dessa equação e (1.6) surge a relação:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (1.7)$$

O resultado em (1.7) tem sua origem nas ideias do Reverendo Thomas Bayes no século XVIII, porém Pierre Simon de Laplace é o autor dessa equação conhecida nos dias atuais como Teorema de Bayes, (Sivia e Skilling, 2006). No contexto dos modelos lineares dinâmicos,

$$f(\boldsymbol{\theta}_t|y_t, D_{t-1}) = \frac{g(y_t|\boldsymbol{\theta}_t, D_{t-1})\pi(\boldsymbol{\theta}_t|D_{t-1})}{h(y_t|D_{t-1})}. \quad (1.8)$$

O processo latente definido por  $\boldsymbol{\theta}_t$  é estimado através da realização  $y_t$  de  $Y$  e do conjunto de informações relevantes,  $D_{t-1}$ . Porém, a incerteza (ou certeza) inicial existente sobre  $\boldsymbol{\theta}_t$  deve ser expressa através de uma distribuição de probabilidade adequada (Jaynes, 2003), no caso,  $\pi(\boldsymbol{\theta}_t|D_{t-1})$ , que é denotada por distribuição *a priori* de  $(\boldsymbol{\theta}_t|D_{t-1})$ .

Conforme surjam novos dados, o conhecimento sobre o processo latente deve ser atualizado. Isso é feito pela função  $g(y_t|\boldsymbol{\theta}_t, D_{t-1})$ , denominada de *verossimilhança*.

A função  $h(y_t|D_{t-1})$  é chamada de distribuição preditiva *a priori* de  $y_t$ . Ela é muito útil quando é desejável fazer inferências a respeito de uma observação ainda desconhecida, (Gelman et al., 2014). No caso, quando o interesse é estimar o valor  $y_t$  de  $Y$ , até então não observado.

Por fim,  $f(\boldsymbol{\theta}_t|y_t, D_{t-1})$  é a distribuição resultante da composição das três anteriores. Ela permite fazer inferências sobre o processo latente e recebe o nome de distribuição *a posteriori* de  $(\boldsymbol{\theta}_t|D_t)$ , uma vez que o conhecimento de  $y_t$  e  $D_{t-1}$  dá origem a  $D_t$ .

Recursivamente, utilizando a distribuição *a posteriori* de  $(\boldsymbol{\theta}_t|D_t)$  obtem-se a distribuição *a priori* de  $(\boldsymbol{\theta}_{t+1}|D_t)$ . Assim, o Teorema de Bayes flui naturalmente no contexto dos modelos lineares dinâmicos, como é visto através das equações de atualização a seguir.

### 1.3 Equações de Atualização

O modelo linear dinâmico oferece um conjunto de equações que são atualizadas ao longo do tempo, e que permitem estimar observações futuras. O raciocínio a seguir se baseia em West e Harrison (1997). Apenas para simplificar a álgebra a ser apresentada, será considerado o modelo linear dinâmico constante.

Seja, para algum  $\mathbf{m}_t$  e  $\mathbf{C}_t$ , a *posteriori* de  $\boldsymbol{\theta}_t$ :

$$(\boldsymbol{\theta}_t|D_t) \sim \mathcal{N}(\mathbf{m}_t, \mathbf{C}_t). \quad (1.9)$$

As equações (1.2) e (1.9) permitem calcular a distribuição *a priori* de  $\boldsymbol{\theta}_{t+1}$  dada a informação em  $t$ , que é:

$$\begin{aligned} \boldsymbol{\theta}_{t+1} &= \mathbf{G}\boldsymbol{\theta}_t + \boldsymbol{\omega}_{t+1}, \\ (\boldsymbol{\theta}_{t+1}|D_t) &\sim \mathbf{G}\mathcal{N}_{(\boldsymbol{\theta}_t|D_t)}(\mathbf{m}_t, \mathbf{C}_t) + \mathcal{N}_{\boldsymbol{\omega}_{t+1}}(\mathbf{0}, \mathbf{W}), \\ (\boldsymbol{\theta}_{t+1}|D_t) &\sim \mathcal{N}(\mathbf{a}_{t+1}, \mathbf{R}_{t+1}), \end{aligned} \quad (1.10)$$

em que  $\mathbf{a}_{t+1} = \mathbf{G}\mathbf{m}_t$  e  $\mathbf{R}_{t+1} = \mathbf{G}\mathbf{C}_t\mathbf{G}' + \mathbf{W}$ .

A distribuição da previsão da observação de  $Y_{t+1}$  dada a informação em  $t$ , ou  $(Y_{t+1}|D_t)$ , pode ser calculada por meio das equações (1.1) e (1.10):

$$\begin{aligned} Y_{t+1} &= \mathbf{F}'\boldsymbol{\theta}_{t+1} + \nu_{t+1}, \\ (Y_{t+1}|D_t) &\sim \mathbf{F}'\mathcal{N}_{(\boldsymbol{\theta}_{t+1}|D_t)}(\mathbf{a}_{t+1}, \mathbf{R}_{t+1}) + \mathcal{N}_{\nu_{t+1}}(0, V), \\ (Y_{t+1}|D_t) &\sim \mathcal{N}(f_{t+1}, Q_{t+1}), \end{aligned} \quad (1.11)$$

em que  $f_{t+1} = \mathbf{F}'\mathbf{a}_{t+1}$  e  $Q_{t+1} = \mathbf{F}'\mathbf{R}_{t+1}\mathbf{F} + V$ .

A distribuição *a posteriori* de  $\boldsymbol{\theta}_{t+1}$  dada a informação em  $t + 1$ , que será *a priori* na próxima iteração, é obtida via o Teorema de Bayes, onde:

$$p(\boldsymbol{\theta}_{t+1}|D_{t+1}) \propto p(Y_{t+1}|\boldsymbol{\theta}_{t+1}, D_t)p(\boldsymbol{\theta}_{t+1}|D_t) \quad (1.12)$$

e obtem-se:

$$(\boldsymbol{\theta}_{t+1}|D_{t+1}) \sim \mathcal{N}(\mathbf{m}_{t+1}, \mathbf{C}_{t+1}), \quad (1.13)$$

onde,  $\mathbf{m}_{t+1} = \mathbf{a}_{t+1} + \mathbf{A}_{t+1}e_{t+1}$  e  $\mathbf{C}_{t+1} = \mathbf{R}_{t+1} - \mathbf{A}_{t+1}Q_{t+1}\mathbf{A}'_{t+1}$ , com  $\mathbf{A}_{t+1} = \mathbf{R}_{t+1}\mathbf{F}Q_{t+1}^{-1}$  e  $e_{t+1} = Y_{t+1} - f_{t+1}$ .

## 1.4 Previsões

A previsão para um passo a frente é dada pela distribuição (1.11), descrita anteriormente. Para definir a predição  $k$  passos a frente é necessário, antes, encontrar a distribuição de  $(\boldsymbol{\theta}_{t+k}|D_t)$ . Isso é feito por intermédio da distribuição *a priori* (1.10), aplicada na equação do sistema (1.2), sucessivamente. Assim, dado  $D_t$  que é ainda a última informação disponível sobre o sistema, e para  $k = 2$ :

$$\begin{aligned} \boldsymbol{\theta}_{t+2} &= \mathbf{G}\boldsymbol{\theta}_{t+1} + \boldsymbol{\omega}_{t+2}, \\ (\boldsymbol{\theta}_{t+2}|D_t) &\sim \mathbf{G}\mathcal{N}_{(\boldsymbol{\theta}_{t+1}|D_t)}(\mathbf{a}_{t+1}, \mathbf{R}_{t+1}) + \mathcal{N}_{\boldsymbol{\omega}_{t+2}}(0, \mathbf{W}), \\ (\boldsymbol{\theta}_{t+2}|D_t) &\sim \mathcal{N}(\mathbf{a}_{t+2}, \mathbf{R}_{t+2}), \end{aligned} \quad (1.14)$$

em que  $\mathbf{a}_{t+2} = \mathbf{G}\mathbf{a}_{t+1}$  e  $\mathbf{R}_{t+2} = \mathbf{G}\mathbf{R}_{t+1}\mathbf{G}' + \mathbf{W}$ .

Para um valor de  $k \geq 2$ , Pole et al. (1994) mostra que:

$$(\boldsymbol{\theta}_{t+k}|D_t) \sim \mathcal{N}(\mathbf{a}_{t+k}, \mathbf{R}_{t+k}), \quad (1.15)$$

onde  $\mathbf{a}_{t+k} = \mathbf{G}^{k-1}\mathbf{a}_{t+1}$  e  $\mathbf{R}_{t+k} = \mathbf{G}^{k-1}\mathbf{R}_{t+1}(\mathbf{G}^{k-1})' + \sum_{j=2}^k \mathbf{G}^{k-j}\mathbf{W}(\mathbf{G}^{k-j})'$ . E, com isso,

a distribuição da  $k$ -ésima predição dada a informação  $D_t$  será:

$$(Y_{t+k}|D_t) \sim \mathcal{N}(f_{t+k}, Q_{t+k}), \quad (1.16)$$

em que  $f_{t+k} = \mathbf{F}'\mathbf{a}_{t+k}$  e  $Q_{t+k} = \mathbf{F}'\mathbf{R}_{t+k}\mathbf{F} + V$ .

## Capítulo 2

# Modelo de Volatilidade Estocástica

O modelo de volatilidade estocástica (MVE) surgiu como uma alternativa aos tradicionais modelos ARCH (Engle, 1982) e GARCH (Bollerslev, 1986). Nesse modelo, a variabilidade dos dados deixa de ter um padrão determinístico e passa, também, a ser interpretada como uma variação aleatória.

### 2.1 Definição do Modelo Canônico

A definição do modelo canônico de volatilidade estocástica foi inicialmente proposta por Taylor (1982). Seja uma variável observada,  $Y_t$ , que segue a seguinte formulação:

$$Y_t = \rho_t \delta_t, \quad (2.1)$$

em que  $\rho_t$  é a *volatilidade* no tempo  $t$  e  $\delta_t$  é um ruído com distribuição normal padrão. Seja o logaritmo do quadrado da volatilidade no tempo  $t$  modelado por um processo latente auto-regressivo de primeira ordem. Dessa maneira,

$$\ln \rho_t^2 = h_t \quad (2.2)$$

é descrito por um modelo AR(1), ver equação (2.3).

Kim et al. (1998) apresenta o modelo canônico para dados igualmente espaçados. O modelo é basicamente composto pela equação que descreve a variável observada  $Y_t$  em



função de uma variável latente que capta a variabilidade dos dados no decorrer do tempo, e pelo processo auto-regressivo de primeira ordem que modela a estrutura da volatilidade. Assim, para  $t = 1, \dots, N$ , o modelo canônico é expresso por:

$$\text{Modelo Canônico : } \begin{cases} Y_t &= e^{\frac{h_t}{2}} \delta_t, \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \eta_t, \\ h_0 &\sim \mathcal{N}\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right). \end{cases} \quad (2.3)$$

Os termos  $\delta_t$  e  $\eta_t$  são ruídos Gaussianos, independentes no tempo e entre si, com distribuições:

$$\delta_t \sim \mathcal{N}(0, 1) \quad \text{e} \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2). \quad (2.4)$$

Observe que,  $Y_t \sim \mathcal{N}(0, e^{h_t})$ , ou como requerido pelo modelo, a variância das observações é totalmente descrita no tempo  $t$  em função da variável latente  $h_t$ . O conjunto de parâmetros do modelo a serem estimados é definido pelo vetor:

$$\boldsymbol{\psi} = (\mu, \phi, \sigma_\eta^2). \quad (2.5)$$

## 2.2 Um Modelo de Espaço-Estado

O modelo de volatilidade estocástica pode ser reescrito utilizando a notação de um modelo de espaço-estado proposta por West e Harrison (1997), descrita no Capítulo 1. Uma das vantagens disso é poder utilizar o aparato ferramental e metodológico já implementado computacionalmente, (Petrís et al., 2009 e Petris, 2010). Tal formulação será útil nas subseções 3.2.4, 3.2.5 e 3.3.2.

### 2.2.1 Equação das Observações

A redefinição do modelo canônico, como um modelo de espaço-estado, pode ser feita pela extensão da proposta em Zivot e Yollin (2012) descrita a seguir.

Pela definição em (2.3):

$$Y_t = e^{\frac{h_t}{2}} \delta_t. \quad (2.6)$$

Então,

$$\ln Y_t^2 = h_t + \ln \delta_t^2. \quad (2.7)$$

Ainda, pode-se fazer  $\varepsilon_t = \ln \delta_t^2 - E(\ln \delta_t^2)$  e escrever:

$$\ln Y_t^2 = h_t + E(\ln \delta_t^2) + \varepsilon_t. \quad (2.8)$$

de modo que  $\varepsilon_t \sim (0, \text{Var}(\ln \delta_t^2))$  seja um ruído em torno de zero.

Apesar dessa transformação linearizar o problema, o pressuposto Gaussiano de  $\delta_t$  não pode mais ser mantido. Durbin e Koopman (2012) afirma que isso não impede a utilização das técnicas associadas aos modelos Gaussianos, uma vez que se assuma que a distribuição de  $\varepsilon_t$  seja aproximadamente normal com média zero. Entretanto, Omori et al. (2007) sugere o uso de uma mistura de  $r = 10$  normais, que se aproxima da distribuição do erro  $\ln \delta_t^2$  na equação (2.7), isto é:

$$\ln \delta_t^2 \approx \xi_t \sim \sum_{j=1}^{10} w_j \mathcal{N}(m_j, s_j^2). \quad (2.9)$$

Os pesos, as médias e as variâncias,  $w_j$ ,  $m_j$  e  $s_j^2$ , respectivamente para  $j = 1, \dots, 10$ , das 10 distribuições sugeridas por Omori et al. (2007) são detalhados na Tabela 2.1.

Tabela 2.1: Parâmetros da distribuição de  $\xi_t$  em (2.9), (Omori et al., 2007).

$j$	1	2	3	4	5	6	7	8	9	10
$w_j$	0,006	0,048	0,131	0,207	0,227	0,188	0,120	0,056	0,016	0,001
$m_j$	1,927	1,347	0,735	0,023	-0,852	-1,973	-3,468	-5,552	-8,684	-14,650
$s_j^2$	0,113	0,178	0,268	0,406	0,627	0,986	1,575	2,545	4,166	7,333

A Figura 2.1 mostra a comparação entre as distribuições de  $\ln \delta_t^2$ , que é o verdadeiro erro do modelo, de  $\varepsilon_t$ , que corresponde à aproximação Gaussiana ingênua, e de  $\xi_t$ , que é a mistura de 10 normais sugerida por Omori et al. (2007). Pode-se observar que a aproximação Gaussiana ingênua é bastante insatisfatória quando comparada à aproximação proposta por Omori et al. (2007).

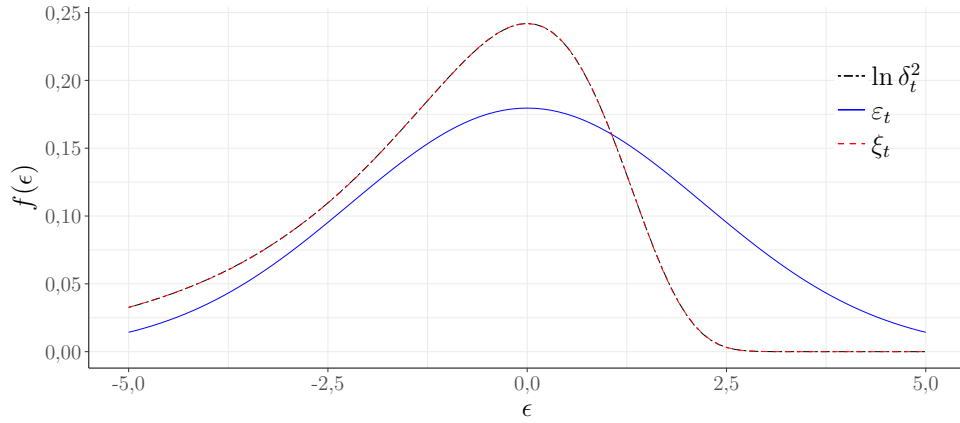


Figura 2.1: Comparação das distribuições de  $\ln \delta_t^2$ ,  $\varepsilon_t$  e  $\xi_t$ .

## 2.2.2 Equação do Sistema

A equação que define a evolução da variável latente,  $h_t$ , em (2.3) pode ser reescrita:

$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \quad (2.10)$$

então,

$$h_t = (1 - \phi)\mu + \phi h_{t-1} + \eta_t. \quad (2.11)$$

Como essa transformação é um mero rearranjo dos termos, a distribuição Gaussiana dos erros em  $\eta_t$  não foi perdida, como ocorre no caso da equação das observações em (2.7).

## 2.2.3 Modelo Completo

Utilizando a notação de espaço-estado para os modelos lineares dinâmicos em West e Harrison (1997), descrita no capítulo anterior, a equação das observações (2.7) pode ser escrita na forma da equação (1.1):

$$Z_t = \ln Y_t^2 = h_t + E(\ln \delta_t^2) + \varepsilon_t, \quad (2.12)$$

logo,

$$Z_t = \ln Y_t^2 = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_t \\ \mu \\ E(\ln \delta_t^2) \end{bmatrix} + \varepsilon_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \nu_t. \quad (2.13)$$

Já a equação do sistema (2.11) pode ser escrita na forma da equação (1.2):

$$h_t = \phi h_{t-1} + (1 - \phi)\mu + \eta_t, \quad (2.14)$$

então,

$$\boldsymbol{\theta}_t = \begin{bmatrix} h_t \\ \mu \\ E(\ln \delta_t^2) \end{bmatrix} = \begin{bmatrix} \phi & 1 - \phi & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} h_{t-1} \\ \mu \\ E(\ln \delta_t^2) \end{bmatrix} + \begin{bmatrix} \eta_t \\ 0 \\ 0 \end{bmatrix} = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t. \quad (2.15)$$

O modelo de volatilidade estocástica é definido, então, como um modelo de espaço-estado constante através das matrizes:

$$\boldsymbol{\theta}_t = \begin{bmatrix} h_t \\ \mu \\ E(\ln \delta_t^2) \end{bmatrix}, \quad \mathbf{F}_t = \mathbf{F} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{e} \quad \mathbf{G}_t = \mathbf{G} = \begin{bmatrix} \phi & 1 - \phi & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2.16)$$

Os termos que representam os erros nas equações das observações e do sistema são:

$$\nu_t = \varepsilon_t \quad \text{e} \quad \boldsymbol{\omega}_t = \begin{bmatrix} \eta_t \\ 0 \\ 0 \end{bmatrix}. \quad (2.17)$$

Consequentemente, as variâncias de  $\nu_t$  e  $\boldsymbol{\omega}_t$  são definidas, respectivamente, por:

$$V_t = V = \text{Var}(\varepsilon_t) \quad \text{e} \quad \mathbf{W}_t = \mathbf{W} = \begin{bmatrix} \sigma_\eta^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (2.18)$$

A variância  $V$  é obtida diretamente através da distribuição de  $\varepsilon_t$ . A matriz  $\mathbf{W}$ , por sua vez, é determinada a partir da matriz  $\boldsymbol{\omega}_t$  em (2.17) e da distribuição de  $\eta_t$  em (2.4).

Há de se destacar que, apesar de  $h_t$  ser a variável que expressa a evolução do sistema pela definição do próprio modelo em (2.3), o vetor  $\boldsymbol{\theta}_t$ , que capta essa característica, só pôde ser escrito em função de  $h_t$ , que é dinâmico, e dos dois termos estáticos, o parâmetro a ser estimado  $\mu$  e a constante  $E(\ln \delta_t^2)$ .

A distribuição *a priori* de  $\boldsymbol{\theta}_t$  é definida por  $\boldsymbol{\theta}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0)$ , onde:

$$\mathbf{m}_0 = \begin{bmatrix} \mu \\ \mu \\ E(\ln \delta_t^2) \end{bmatrix} \quad \text{e} \quad \mathbf{C}_0 = \begin{bmatrix} \frac{\sigma_\eta^2}{1-\phi^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (2.19)$$

Em suma, o modelo de volatilidade estocástica de acordo com a notação de West e Harrison (1997) é escrito como:

$$\begin{array}{l} \text{Modelo de} \\ \text{Espaço-Estado :} \end{array} \left\{ \begin{array}{l} \ln Y_t^2 = \mathbf{F}'\boldsymbol{\theta}_t + \varepsilon_t, \\ \boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \\ \boldsymbol{\theta}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0), \end{array} \right. \quad (2.20)$$

em que a quádrupla  $\{\mathbf{F}, \mathbf{G}, V, \mathbf{W}\}$  é dada por:

$$\{\mathbf{F}, \mathbf{G}, V, \mathbf{W}\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} \phi & 1-\phi & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{Var}(\varepsilon_t), \begin{bmatrix} \sigma_\eta^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\}. \quad (2.21)$$

## 2.3 Modelo Parcialmente Não Centralizado

O modelo canônico em (2.3) também é conhecido como modelo centralizado. Uma forma equivalente de redefinir esse modelo é levando o parâmetro  $\mu$ , da equação do sistema, para a equação das observações. Seja,

$$h_t^* = h_t - \mu, \quad (2.22)$$

uma nova variável latente, que é uma mera transformação linear da variável latente  $h_t$  em (2.3). Então pode-se reescrever a equação do sistema na forma alternativa:

$$\begin{aligned} h_t &= \mu + \phi(h_{t-1} - \mu) + \eta_t, \\ h_t - \mu &= \phi(h_{t-1} - \mu) + \eta_t, \\ h_t^* &= \phi h_{t-1}^* + \eta_t. \end{aligned} \tag{2.23}$$

De maneira semelhante, a equação das observações será reescrita como:

$$\begin{aligned} Y_t &= e^{\frac{h_t}{2}} \delta_t, \\ Y_t &= e^{\frac{h_t^* + \mu}{2}} \delta_t, \\ Y_t &= e^{\frac{\mu}{2}} e^{\frac{h_t^*}{2}} \delta_t. \end{aligned} \tag{2.24}$$

E a distribuição inicial da variável latente passa a ser:

$$h_0^* \sim \mathcal{N}\left(0, \frac{\sigma_\eta^2}{1 - \phi^2}\right). \tag{2.25}$$

O conjunto das equações (2.23), (2.24) e (2.25), resumidas em (2.26), define o modelo parcialmente não centralizado (Prado e West, 2010):

$$\begin{array}{l} \text{Modelo Parcialmente} \\ \text{Não Centralizado :} \end{array} \left\{ \begin{array}{l} Y_t = e^{\frac{\mu}{2}} e^{\frac{h_t^*}{2}} \delta_t, \\ h_t^* = \phi h_{t-1}^* + \eta_t, \\ h_0^* \sim \mathcal{N}\left(0, \frac{\sigma_\eta^2}{1 - \phi^2}\right). \end{array} \right. \tag{2.26}$$

## 2.4 Modelo Não Centralizado

A última reparametrização do modelo canônico em (2.3), a ser apresentada, é proposta em Kastner e Frühwirth-Schnatter (2014). Ele recebe o nome de modelo **completamente** não centralizado. Para efeitos de simplificação, este será referido apenas como modelo não centralizado. A transformação linear na variável latente, que conduz

ao modelo não centralizado é definida por:

$$\tilde{h}_t = \frac{h_t - \mu}{\sigma_\eta}. \quad (2.27)$$

A partir de (2.27) é possível reescrever a equação do sistema, que passa a ser:

$$\begin{aligned} h_t &= \mu + \phi(h_{t-1} - \mu) + \eta_t, \\ h_t - \mu &= \phi(h_{t-1} - \mu) + \eta_t, \\ \left( \frac{h_t - \mu}{\sigma_\eta} \right) &= \phi \left( \frac{h_{t-1} - \mu}{\sigma_\eta} \right) + \frac{\eta_t}{\sigma_\eta}, \\ \tilde{h}_t &= \phi \tilde{h}_{t-1} + \tilde{\eta}_t. \end{aligned} \quad (2.28)$$

Diferentemente da transformação (2.22) no caso parcialmente não centralizado, a transformação (2.27) para o caso não centralizado altera a distribuição do erro  $\eta_t$ . Antes, Gaussiano com média zero e variância  $\sigma_\eta^2$ , conforme definido em (2.4), o novo erro,  $\tilde{\eta}_t$ , passa a ter distribuição normal padrão:

$$\tilde{\eta}_t \sim \mathcal{N}(0, 1). \quad (2.29)$$

A equação das observações é reescrita como:

$$\begin{aligned} Y_t &= e^{\frac{h_t}{2}} \delta_t, \\ Y_t &= e^{\frac{\sigma_\eta \tilde{h}_t + \mu}{2}} \delta_t, \\ Y_t &= e^{\frac{\mu}{2}} e^{\sigma_\eta \frac{\tilde{h}_t}{2}} \delta_t. \end{aligned} \quad (2.30)$$

E, por fim, a distribuição inicial de  $\tilde{h}_t$  é dada por:

$$\tilde{h}_0 \sim \mathcal{N}\left(0, \frac{1}{1 - \phi^2}\right). \quad (2.31)$$

De forma sintetizada, o modelo de volatilidade estocástica canônico reparametri-

zado para a forma completamente não centralizada é expresso por:

$$\begin{array}{l} \textbf{Modelo} \\ \textbf{Não Centralizado :} \end{array} \left\{ \begin{array}{l} Y_t = e^{\frac{\mu}{2}} e^{\sigma_\eta \frac{\tilde{h}_t}{2}} \delta_t, \\ \tilde{h}_t = \phi \tilde{h}_{t-1} + \tilde{\eta}_t, \\ \tilde{h}_0 \sim \mathcal{N}\left(0, \frac{1}{1-\phi^2}\right). \end{array} \right. \quad (2.32)$$



## Capítulo 3

# Proposta de Estimação dos Parâmetros do MVE

Existe uma grande desvantagem do modelo de volatilidade estocástica em relação aos modelos ARCH e GARCH em termos de aplicações, como cita Bos (2012). Isso ocorre pois os modelos ARCH e GARCH possuem muitas variações, porém, basicamente apenas uma maneira de se estimar os parâmetros, que está presente em grande parte dos *softwares* estatísticos. Já o modelo de volatilidade estocástica apresenta poucas variações, porém diversos autores sugerem maneiras distintas de se estimar os parâmetros. No entanto, quase nenhuma destas está disponível facilmente através de algum pacote computacional. A Tabela 3.1 foi extraída de Bos (2012) e apresenta, de forma sintetizada, as principais referências sobre os diversos métodos de estimação publicados.

Tabela 3.1: Métodos de estimação dos parâmetros do modelo de volatilidade estocástica e principais referências, (Bos, 2012).

Método	Referência	Paradigma
Quasi-Maximum Likelihood (QML)	Harvey et al. (1994)	Clássico
Gaussian Mixture Sampling (GMS)	Kim et al. (1998)	Bayesiano
Simulated Method of Moments (SMM)	Gallant e Tauchen (1996)	Clássico
Importance Sampling (IS)	Durbin e Koopman (1997)	Clássico
Efficient Importance Sampling (EIS)	Richard e Zhang (2007)	Clássico
Improved Importance Sampling (IIS)	Nguyen (2007)	Clássico
Single Site Sampler (SSS)	Carter e Kohn (1994)	Bayesiano
MultiMove Sampler (MMS)	Shephard e Pitt (1997)	Bayesiano

Como pode ser percebido, a estimação dos parâmetros do modelo de volatilidade

estocástica é um problema atacado de diversas maneiras distintas. O foco desta dissertação é apresentar um método Bayesiano, alternativo àqueles da Tabela 3.1, que combina as técnicas propostas em Kastner e Frühwirth-Schnatter (2014) e McCormick et al. (2012). O primeiro artigo sugere uma metodologia inovadora que envolve, basicamente, a alternância da especificação do modelo no processo de estimação dos parâmetros. O segundo artigo, por sua vez, inspirou uma proposta para modelar a variável latente,  $h_t$ , do modelo de volatilidade estocástica. Ambos os processos serão detalhados nas próximas seções.

### 3.1 Definição do Modelo Bayesiano

A estimação Bayesiana dos parâmetros do modelo canônico (ou centralizado) de volatilidade estocástica em (2.3) consiste em determinar a distribuição *a posteriori* conjunta de:

$$\boldsymbol{\psi} = (\mu, \phi, \sigma_\eta^2). \quad (3.1)$$

Apesar de  $\boldsymbol{\psi}$  ser um vetor com três elementos, é tecnicamente conveniente supor que as componentes de  $\boldsymbol{\psi}$  sejam independentes *a priori*. Portanto, foram tomadas três distribuições, uma para cada elemento de  $\boldsymbol{\psi}$ , independentes entre si:

$$p_{\boldsymbol{\psi}}(\boldsymbol{\psi}) = p_\mu(\mu)p_\phi(\phi)p_{\sigma_\eta^2}(\sigma_\eta^2). \quad (3.2)$$

Como o paradigma Bayesiano sugere, deve-se definir as distribuições *a priori* dos parâmetros, que, atualizadas pela função de verossimilhança, irão gerar a distribuição *a posteriori* de  $(\boldsymbol{\psi}|D)$ , que é o ponto principal no processo de inferência estatística. Neste trabalho foram utilizadas as mesmas distribuições *a priori* descritas em Kastner e Frühwirth-Schnatter (2014) e Kim et al. (1998) para os parâmetros em  $\boldsymbol{\psi}$ .

O nível  $\mu$  da volatilidade tem seu suporte em  $\mathbb{R}$ , e será atribuída uma distribuição *a priori*,  $\pi(\mu)$ , com densidade Gaussiana para esse parâmetro:

$$\mu \sim \mathcal{N}(a_\mu, B_\mu). \quad (3.3)$$

O parâmetro  $\phi$  determina a persistência da volatilidade, e  $|\phi| < 1$ . A distribuição beta é bastante flexível, sendo, dessa forma, a escolha mais comum quando se deseja modelar alguma variável contínua cujo valor esteja descrito num intervalo  $(a, b)$ . Com a finalidade de aproveitar essa vantagem da distribuição beta, seja uma nova variável  $\phi_0$  com distribuição *a priori* é  $\mathcal{B}(a_\phi, b_\phi)$ . A persistência, então, é descrita por  $\phi = 2\phi_0 - 1$  e sua distribuição *a priori* será:

$$\pi(\phi) = \frac{\Gamma(a_\phi + b_\phi)}{2\Gamma(a_\phi)\Gamma(b_\phi)} \left(\frac{1 + \phi}{2}\right)^{a_\phi - 1} \left(\frac{1 - \phi}{2}\right)^{b_\phi - 1}. \quad (3.4)$$

O terceiro e último parâmetro do modelo é  $\sigma_\eta^2$ , a variância da volatilidade, e seus possíveis valores estão em  $\mathbb{R}^+$ . Sua distribuição *a priori* será:

$$\sigma_\eta^2 \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2B_\sigma}\right). \quad (3.5)$$

### 3.2 Estimando $\mu$ , $\phi$ e $\sigma_\eta^2$

Dada a complexidade do problema, as formas fechadas das distribuições condicionais completas *a posteriori* dos parâmetros em  $\psi$  serão estimadas por meio de algoritmos via métodos de Monte Carlo em cadeias de Markov (do inglês, *Markov chain Monte Carlo*, MCMC).

Kastner e Frühwirth-Schnatter (2014) propõe uma estratégia que é chamada de *Ancillarity-Sufficiency Interweaving Strategy* (ASIS). Em tradução livre para o português seria *Estratégia de Entrelaçamento Ancilar-Suficiente*. Apesar de não convencional, a ideia é bastante simples. Dados os valores da série da variável latente,  $h_t$ , os parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$  devem ser estimados a partir do modelo canônico (também conhecido como centralizado). Então, a série de  $h_t$  deve ser transformada para a parametrização não centralizada e os valores dos parâmetros devem ser novamente amostrados via MCMC. Os passos do algoritmo são resumidos da seguinte forma:

**Algoritmo Proposto por Kastner e Frühwirth-Schnatter (2014).**

1. Iniciar os valores de  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$ ;
2. Estimar a série de  $h_t$  (modelo canônico);
3. A partir dos valores de  $h_t$ , estimar os valores de  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$ ;
4. Transformar a série de  $h_t$  para o modelo não centralizado,  $h_t^*$ ;
5. A partir dos valores de  $h_t^*$ , estimar novamente os valores de  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$ .

O porquê dessa alternância de especificações do modelo vem do fato que a variável latente  $h_t$ , no modelo centralizado, forma uma estatística suficiente para  $\mu$  e  $\sigma_\eta^2$ , ao passo que  $h_t$ , transformada para o modelo não centralizado, forma uma estatística ancilar para esses parâmetros. Então, alternar entre essas especificações do modelo aumenta a eficiência do amostrador via MCMC. Kastner e Frühwirth-Schnatter (2014) sugere essa melhora no processo de estimação com base em Yu e Meng (2011), que apresenta a ASIS, relacionando-a com o teorema de Basu (Basu, 1955), como uma forma de aprimorar a eficiência de algoritmos baseados em MCMC.

No caso do modelo de volatilidade estocástica, Kastner e Frühwirth-Schnatter (2014) mostra que o modelo centralizado, sozinho, apresenta melhores resultados quando os valores da persistência,  $\phi$ , e da variância da volatilidade,  $\sigma_\eta^2$ , são altos. Enquanto que o modelo não centralizado, também sozinho, apresenta melhor performance quando o valor de  $\phi$  é menor. Então, a estratégia, na média, apresenta bons resultados num maior espectro de possíveis valores dos parâmetros.

A ideia mais intuitiva é “descentralizar” o modelo canônico (2.3), no passo 4 do algoritmo, através da transformação (2.22):

$$h_t^* = h_t - \mu, \quad (3.6)$$

o que leva ao modelo parcialmente não centralizado (2.26). Porém Kim et al. (1998) e

Prado e West (2010) advertem que esse modelo é ineficiente para se estimar  $\mu$ . Desse modo, o modelo completamente não centralizado (2.32) é proposto por Kastner e Frühwirth-Schnatter (2014), para completar o processo de estimação. Neste modelo,

$$\tilde{h}_t = \frac{h_t - \mu}{\sigma_\eta}. \quad (3.7)$$

Os passos detalhados do algoritmo de estimação, conforme Kastner e Frühwirth-Schnatter (2014), são descritos nas subseções que seguem.

### 3.2.1 Amostrando $\sigma_\eta^2$

A variância da volatilidade,  $\sigma_\eta^2$ , é estimada com base no algoritmo de Metropolis-Hastings (Metropolis et al. (1953); Hastings (1970)). Esse algoritmo parte de uma função geradora de candidatos a valores do parâmetro em questão. Então, gerado um provável valor, ele é aceito como observação *a posteriori* com probabilidade dada por  $\min(1, R)$ . Caso o valor proposto seja rejeitado, o valor da iteração anterior é repetido na iteração atual.

A partir do modelo canônico (2.3) e, conseqüentemente, dos valores de  $h_t$ , a distribuição geradora de candidatos de  $\sigma_\eta^2$  foi construída considerando uma distribuição *a priori* auxiliar conjugada  $p(\sigma_\eta^2) \propto \sigma_\eta^{-1}$ , de modo que a distribuição condicional completa de  $\sigma_\eta^2$  tenha uma distribuição geradora de candidatos tal que:

$$\sigma_\eta^2 | h_{1:N}, \mu, \phi \sim \mathcal{G}^{-1}(c_N, C_N). \quad (3.8)$$

Assim, a distribuição geradora de candidatos é Gama-Inversa com parâmetros  $c_N$  e  $C_N$ , em que  $N$  é a quantidade de observações,

$$c_N = \frac{N}{2}, \quad (3.9)$$

e

$$C_N = \frac{1}{2} \left( \sum_{t=1}^N ((h_t - \mu) - \phi(h_{t-1} - \mu))^2 + (h_0 - \mu)^2 (1 - \phi^2) \right). \quad (3.10)$$

O valor gerado é aceito com probabilidade  $\min(1, R)$ , onde:

$$R = \exp \left\{ \frac{\sigma_{\eta;*}^2 - \sigma_{\eta;0}^2}{2B_\sigma} \right\}, \quad (3.11)$$

$\sigma_{\eta;0}^2$  é o candidato gerado,  $\sigma_{\eta;*}^2$  é o valor gerado na iteração anterior e  $B_\sigma$  é o hiperparâmetro da distribuição *a priori* de  $\sigma_\eta^2$  em (3.5).

Se o valor  $\sigma_{\eta;0}^2$  não for aceito, o valor  $\sigma_{\eta;*}^2$  é repetido.

### 3.2.2 Amostrando $\phi$

De maneira semelhante, a persistência,  $\phi$ , do modelo de volatilidade estocástica é estimada, sob o modelo canônico, através do algoritmo de Metropolis-Hastings utilizado na estimação de  $\sigma_\eta^2$ .

Utilizando uma distribuição *a priori* hierárquica para  $\phi$  tal que, *a priori*  $(\phi|\sigma_\eta^2) \sim \mathcal{N}(0, \sigma_\eta^2 B_0^{22})$  com  $B_0^{22}$  sendo uma constante a ser definida, Kastner e Frühwirth-Schnatter (2014) controla a seguinte distribuição geradora de candidatos para  $\phi$ :

$$\phi|h_{1:N}, \gamma, \sigma_\eta^2 \sim \mathcal{N} \left( \frac{\sum_{t=1}^N h_{t-1}h_t - \gamma \sum_{t=0}^{N-1} h_t}{\sum_{t=0}^{N-1} h_t^2 + \frac{1}{B_0^{22}}}, \frac{\sigma_\eta^2}{\sum_{t=0}^{N-1} h_t^2 + \frac{1}{B_0^{22}}} \right) I_{(-1,1)}(\phi), \quad (3.12)$$

onde,  $N$  é a quantidade de observações,  $B_0^{22}$  é a constante definida anteriormente, e  $\gamma = (1 - \phi)\mu$  é uma transformação a ser detalhada na próxima subseção. A densidade de probabilidade proposta em (3.12) é multiplicada pela função indicadora,  $I_{(-1,1)}(\phi)$ , pois  $|\phi| < 1$ . Dessa maneira, a distribuição Gaussiana é *truncada* nos possíveis valores do parâmetro.

O candidato a valor do parâmetro gerado é aceito como observação *a posteriori* de  $\phi$  com probabilidade  $\min(1, R)$ , onde:

$$R = \frac{p(h_o|\mu, \phi_o, \sigma_\eta^2)\pi(\phi_o)}{p_{\text{aux}}(\phi_o|\sigma_\eta^2)} \times \frac{p_{\text{aux}}(\phi_*|\sigma_\eta^2)}{p(h_o|\mu, \phi_*, \sigma_\eta^2)\pi(\phi_*)}, \quad (3.13)$$

$\phi_o$  é o candidato gerado,  $\phi_*$  é a observação *a posteriori* do parâmetro  $\phi$  amostrada na

iteração anterior,  $p(h_0 | \dots)$  é a distribuição *a priori* de  $h_0$  sob o modelo canônico dada em (2.3),  $\pi(\phi)$  é a distribuição *a priori* de  $\phi$  definida em (3.4) e  $p_{\text{aux}}(\phi | \sigma_\eta^2)$  é a distribuição hierárquica *a priori* definida em Kastner e Frühwirth-Schnatter (2014):

$$p_{\text{aux}}(\phi | \sigma_\eta^2) \sim \mathcal{N}(0, \sigma_\eta^2 B_0^{22}). \quad (3.14)$$

Assim como (3.12), a distribuição anterior depende da constante  $B_0^{22}$  a ser definida.

Se o candidato  $\phi_0$  não for aceito, o valor de  $\phi_*$  é repetido como valor da distribuição *a posteriori* de  $\phi$ .

### 3.2.3 Amostrando $\mu$

A estimação de  $\mu$  descrita em Kastner e Frühwirth-Schnatter (2014) é feita por intermédio da transformação:

$$\gamma = (1 - \phi)\mu. \quad (3.15)$$

A nova variável  $\gamma$ , que aparece pela primeira vez na distribuição geradora de candidatos de  $\phi$  na subseção anterior, é proposta por Kastner e Frühwirth-Schnatter (2014) como uma via indireta de se estimar  $\mu$ . A distribuição *a priori* de  $(\gamma | \phi)$  é dada por:

$$(\gamma | \phi) \sim \mathcal{N}((1 - \phi)a_\mu, (1 - \phi)^2 B_\mu). \quad (3.16)$$

O algoritmo de Metropolis-Hastings também é utilizado como nas etapas anteriores. Utilizando uma distribuição *a priori* hierárquica para  $\gamma$  tal que,  $(\gamma | \sigma_\eta^2) \sim \mathcal{N}(0, \sigma_\eta^2 B_0^{11})$ , sendo  $B_0^{11}$  uma constante a ser definida, é possível construir a distribuição geradora de candidatos de  $\gamma$  dada por:

$$\gamma | h_{1:N}, \phi, \sigma_\eta^2 \sim \mathcal{N}\left(\frac{\sum_{t=1}^N h_t - \phi \sum_{t=0}^{N-1} h_t}{N + \frac{1}{B_0^{11}}}, \frac{\sigma_\eta^2}{N + \frac{1}{B_0^{11}}}\right), \quad (3.17)$$

onde  $N$  é o número de observações.

O valor gerado então é aceito com probabilidade  $\min(1, R)$ , onde:

$$R = \frac{p(h_o|\gamma_0, \phi, \sigma_\eta^2)\pi(\gamma_0|\phi)}{p_{\text{aux}}(\gamma_0|\sigma_\eta^2)} \times \frac{p_{\text{aux}}(\gamma_*|\sigma_\eta^2)}{p(h_o|\gamma_*, \phi, \sigma_\eta^2)\pi(\gamma_*|\phi)}, \quad (3.18)$$

$\gamma_0$  é o candidato gerado,  $\gamma_*$  é o valor gerado a partir da transformação (3.15) de  $\mu$  da iteração anterior e  $p(h_o|\dots)$  é a distribuição *a priori* de  $h_o$  sob o modelo canônico dada em (2.3) e modificada para  $\gamma$ :

$$h_o \sim \mathcal{N}\left(\frac{\gamma}{1-\phi}, \frac{\sigma_\eta^2}{1-\phi^2}\right). \quad (3.19)$$

$\pi(\gamma|\phi)$  é a distribuição *a priori* de  $\gamma$  definida em (3.16) e  $p_{\text{aux}}(\gamma|\sigma_\eta^2)$  é a distribuição *a priori* hierárquica definida em Kastner e Frühwirth-Schnatter (2014):

$$p_{\text{aux}}(\gamma|\sigma_\eta^2) \sim \mathcal{N}(0, \sigma_\eta^2 B_0^{11}). \quad (3.20)$$

Se o valor  $\gamma_0$  for rejeitado, o valor de  $\mu$  resultante da iteração anterior é repetido. A partir do valor amostral de  $\gamma$ , a observação *a posteriori* de  $\mu$  é definida resolvendo (3.15):

$$\mu = \frac{\gamma}{1-\phi}. \quad (3.21)$$

### 3.2.4 Reamostrando $\sigma_\eta^2$

Após a seleção das amostras como descrito nas subseções 3.2.1 a 3.2.3, os parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$  foram estimados sob o modelo centralizado, ou canônico. Na sequência, os valores atuais da variável  $h_t$  devem ser transformados para o modelo não centralizado (2.32), para que façam-se novas amostragens dos parâmetros. Isso é feito por intermédio da transformação:

$$\tilde{h}_t = \frac{h_t - \mu}{\sigma_\eta}. \quad (3.22)$$

O novo valor de  $\sigma_\eta^2$  é gerado diretamente da distribuição condicional completa do



parâmetro sob o modelo não centralizado:

$$\sigma_\eta | \tilde{y}_{1:N}, \tilde{h}_{1:N}, \mu \sim \mathcal{N}(a_{N,\sigma_\eta}, B_{N,\sigma_\eta}), \quad (3.23)$$

onde,

$$a_{N,\sigma_\eta} = B_{N,\sigma_\eta} \sum_{t=1}^N \frac{\tilde{h}_t (\tilde{y}_t - m_r - \mu)}{s_r^2}, \quad (3.24)$$

e

$$B_{N,\sigma_\eta} = \frac{1}{\sum_{t=1}^N \frac{\tilde{h}_t^2}{s_r^2} + \frac{1}{B_\sigma}}. \quad (3.25)$$

Os termos em  $\tilde{y}_t$  são os valores observados transformados de acordo com a proposta de linearização (2.8) no modelo de espaço-estado:

$$\tilde{y}_t = \ln y_t^2. \quad (3.26)$$

Os valores  $m_r$  e  $s_r^2$  correspondem aos parâmetros da mistura de 10 normais sugeridas (Tabela 2.1) para modelar o erro que surge na linearização do modelo de volatilidade estocástica.  $B_\sigma$  é o hiperparâmetro da distribuição *a priori* de  $\sigma_\eta^2$  em (3.5).

### 3.2.5 Reamostrando $\mu$

De maneira análoga, o valor de  $\mu$  é reamostrado diretamente através da distribuição condicional completa de  $\mu$ :

$$\mu | \tilde{y}_{1:N}, \tilde{h}_{1:N}, \sigma_\eta \sim \mathcal{N}(a_{N,\mu}, B_{N,\mu}), \quad (3.27)$$

onde,

$$a_{N,\mu} = B_{N,\mu} \sum_{t=1}^N \frac{\tilde{y}_t - m_r - \sigma_\eta \tilde{h}_t}{s_r^2} + \frac{a_\mu}{B_\mu}, \quad (3.28)$$

e

$$B_{N,\mu} = \frac{1}{\sum_{t=1}^N \frac{1}{s_r^2} + \frac{1}{B_\mu}}. \quad (3.29)$$

Novamente,  $\tilde{y}_t$  representa os valores observados transformados conforme a proposta de linearização do problema em (2.8). Os termos  $m_r$  e  $s_r^2$  são os parâmetros da mistura de 10 normais que modela o erro referente à mesma transformação, e os valores  $a_\mu$  e  $B_\mu$  são os hiperparâmetros da distribuição *a priori* de  $\mu$  em (3.3).

O valor de  $\phi$  não precisa ser reamostrado, pois esse parâmetro não é explicitamente envolvido na transformação (3.22) da variável latente. Como cita Kastner e Frühwirth-Schnatter (2014), apenas os valores de  $\mu$  e  $\sigma_\eta^2$  precisam ser reestimados segundo essa estratégia.

### 3.3 Estimando $h_{1:N}$

Na seção anterior apresentou-se a estratégia de estimação dos parâmetros do modelo de volatilidade estocástica, conforme a sugestão em Kastner e Frühwirth-Schnatter (2014). Desta feita, o procedimento consiste em alternar a parametrização do modelo ao longo das iterações, sendo conhecidos (ou estimados) os valores da série  $h_t$ . Os autores apresentam uma maneira de estimar a variável latente,  $h_t$ , através de uma distribuição normal  $N$ -variada (onde  $N$  é a quantidade de observações) bastante específica. Além disso a implementação computacional do processo de estimação não é trivial. O objetivo aqui é sugerir uma outra maneira de se estimar  $h_t$  em substituição àquela presente no artigo original.

A proposta a ser apresentada nessa seção é inspirada em McCormick et al. (2012). No trabalho citado, os autores utilizam aproximações Gaussianas para estimar os estados latentes de um modelo de regressão dinâmica de Bernoulli. A grande vantagem dessa abordagem é que se trata de um processo de estimação mais simples do que o proposto em Kastner e Frühwirth-Schnatter (2014). A adaptação da técnica de McCormick et al. (2012) ao modelo de volatilidade estocástica é desenvolvida a seguir. Antes, porém, o método de McCormick et al. (2012) será sumarizado de uma forma geral.

### 3.3.1 O Método de McCormick et al. (2012)

Nesta seção será descrito o método proposto por McCormick et al. (2012), originalmente para trata de regressão dinâmica para dados binários, de uma forma mais geral, a ser usada posteriormente no modelo de volatilidade estocástica.

Seja  $t = 1, \dots, N$ , os tempos discretos em que uma série temporal,  $y_t$  é monitorada e o seguinte modelo dinâmico:

$$\begin{aligned} \text{Equação das Observações:} \quad & y_t \sim f(\cdot | \mu_t), \\ \text{Função de Ligação:} \quad & \mu_t = \eta \boldsymbol{\theta}_t = \mathbf{x}_t^T \boldsymbol{\theta}_t, \\ \text{Equação do Sistema:} \quad & \boldsymbol{\theta}_t = \mathbf{G} \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \end{aligned} \tag{3.30}$$

onde  $\mathbf{x}_t = (\mathbf{x}_{1,t}, \dots, \mathbf{x}_{d,t})$  é um vetor de preditores,  $\boldsymbol{\theta}_t$  é um vetor  $d$ -dimensional de variáveis latentes,  $\eta(\boldsymbol{\theta}_t)$  é uma função de ligação,  $\mathbf{G}$  é uma matriz como definida na Seção 1.1 e os  $\boldsymbol{\omega}_t$ 's são vetores aleatórios independentes e identicamente distribuídos tais que  $\boldsymbol{\omega}_t \sim \mathcal{N}(0, \mathbf{W}_t)$ .

Para os dados observados  $D_{t-1} = (y_1, \dots, y_{t-1})$  e considerando pontos de partida razoáveis, a estimação recursiva começa supondo que (McCormick et al., 2012):

$$(\boldsymbol{\theta}_{t-1} | D_{t-1}) \sim \mathcal{N}(\hat{\boldsymbol{\theta}}_{t-1}, \hat{\boldsymbol{\Sigma}}_{t-1}). \tag{3.31}$$

O processo de estimação é, então, feito em dois passos: predição e atualização. Baseado em (3.30) e (3.31), a *equação de predição* é descrita por:

$$(\boldsymbol{\theta}_t | D_{t-1}) \sim \mathcal{N}(\mathbf{G} \hat{\boldsymbol{\theta}}_{t-1}, \mathbf{R}_t), \tag{3.32}$$

em que

$$\mathbf{R}_t = \frac{\mathbf{G} \hat{\boldsymbol{\Sigma}}_{t-1} \mathbf{G}^T}{\lambda_t}, \tag{3.33}$$

onde  $\lambda_t$  é um *fator de desconto*, com  $0 < \lambda_t < 1$  especificado através da equação (3.42) e  $\hat{\boldsymbol{\Sigma}}_{t-1}$  é obtido pela equação (3.37).

Após o passo da predição, a *atualização* é feita ao obter a distribuição *a posteriori* de  $(\boldsymbol{\theta}_t|D_t)$ , que é aproximada, usando-se (3.32) e o Teorema de Bayes, por:

$$p(\boldsymbol{\theta}_t|D_t) \propto p(y_t|\boldsymbol{\theta}_t)\mathcal{N}(\mathbf{G}\hat{\boldsymbol{\theta}}_{t-1}, \mathbf{R}_t). \quad (3.34)$$

O lado direito de (3.34) não possui forma fechada. Então,  $\boldsymbol{\theta}_t$  é estimado utilizando-se um procedimento de Newton-Raphson. Seja,

$$l(\boldsymbol{\theta}_t) = \ln p(\boldsymbol{\theta}_t|D_t) \approx \ln p(y_t|\boldsymbol{\theta}_t) - \frac{1}{2} \left[ \boldsymbol{\theta}_t^T \mathbf{R}_t^{-1} - 2\hat{\boldsymbol{\theta}}_{t-1}^T \mathbf{G}^T \mathbf{R}_t^{-1} \boldsymbol{\theta}_t \right], \quad (3.35)$$

e a estimativa de  $\boldsymbol{\theta}_t$  dada por:

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} - \left[ D^2 l(\hat{\boldsymbol{\theta}}_{t-1}) \right]^{-1} D l(\hat{\boldsymbol{\theta}}_{t-1}), \quad (3.36)$$

onde  $D l(\boldsymbol{\theta}_t)$  e  $D^2 l(\boldsymbol{\theta}_t)$  são a primeira e a segunda derivadas de  $l(\boldsymbol{\theta}_t)$ , respectivamente.

Para atualizar a matriz de covariância  $\boldsymbol{\Sigma}_t$ , utiliza-se:

$$\hat{\boldsymbol{\Sigma}}_t = - \left[ D^2 l(\hat{\boldsymbol{\theta}}_{t-1}) \right]^{-1}. \quad (3.37)$$

A distribuição preditiva  $p(Y_t|D_{t-1})$  é obtida através de:

$$p(Y_t|D_{t-1}) = \int p(Y_t|\boldsymbol{\theta}_t, D_{t-1}) p(\boldsymbol{\theta}_t|D_{t-1}) d\boldsymbol{\theta}_t. \quad (3.38)$$

No entanto, a expressão em (3.38) não possui forma fechada, mas é possível calcular o seu valor por aproximações de Laplace (Tierney e Kadane, 1986):

$$f(y_t|D_{t-1}) \approx (2\pi)^{\frac{d}{2}} \left| \left( D^2 l(\hat{\boldsymbol{\theta}}_t) \right)^{-1} \right| p(y_t|\hat{\boldsymbol{\theta}}_t) p(\hat{\boldsymbol{\theta}}_t|D_{t-1}), \quad (3.39)$$

em que,

$$p(\hat{\boldsymbol{\theta}}_t|D_{t-1}) \approx \mathcal{N}(\mathbf{G}\hat{\boldsymbol{\theta}}_t, \mathbf{R}_t(\hat{\boldsymbol{\theta}}_t)), \quad (3.40)$$

onde

$$\mathbf{R}_t(\hat{\boldsymbol{\theta}}_t) = \frac{\mathbf{G} \left[ D^2 l(\hat{\boldsymbol{\theta}}_t) \right]^{-1} \mathbf{G}^T}{\lambda_t}, \quad (3.41)$$

e  $p(y_t|\hat{\boldsymbol{\theta}}_t) = f(y_t|\eta(\hat{\boldsymbol{\theta}}))$ .

Na estimação do fator de desconto,  $\lambda_t$ , é escolhido o valor de  $\lambda_t$  que maximiza a expressão (3.39), isto é,

$$\lambda_t = \underset{\lambda_t}{\operatorname{argmax}} p(Y_t|D_{t-1}, \lambda_t). \quad (3.42)$$

Agora esta técnica é utilizada na estimação da variável latente,  $h_t$ , do modelo de volatilidade estocástica.

### 3.3.2 Adaptação do Método de McCormick et al. (2012) ao MVE

A partir da equação do sistema em (2.3), a distribuição  $(h_t|D_{t-1}, \mu, \phi)$  é encontrada:

$$\begin{aligned} h_t &= \mu + \phi(h_{t-1} - \mu) + \eta_t, \\ (h_t|D_{t-1}, \mu, \phi) &= (\mu|D_{t-1}) + \phi(h_{t-1}|D_{t-1}) + (\phi\mu|D_{t-1}) + (\eta_t|D_{t-1}), \\ (h_t|D_{t-1}, \mu, \phi) &= \mu + \phi(h_{t-1}|D_{t-1}) + \phi\mu + \eta_t, \\ (h_t|D_{t-1}, \mu, \phi) &= \mu(1 - \phi) + \phi(h_{t-1}|D_{t-1}) + \eta_t. \end{aligned} \quad (3.43)$$

Seja, por hipótese:

$$(h_{t-1}|D_{t-1}) \sim \mathcal{N}(\hat{m}_{t-1}, \hat{C}_{t-1}). \quad (3.44)$$

Como nas equações de atualização dos modelos lineares dinâmicos descritas na Seção 1.3 do Capítulo 1, a distribuição de  $(h_t|D_{t-1}, \mu, \phi)$ , através de (3.43) e (3.44), será:

$$(h_t|D_{t-1}, \mu, \phi) \sim \mathcal{N}(\mu(1 - \phi) + \phi\hat{m}_{t-1}, \phi^2\hat{C}_{t-1} + \sigma_\eta^2). \quad (3.45)$$

Porém, seja a variância da distribuição em (3.45) aproximada por:

$$R_t = \frac{\phi^2\hat{C}_{t-1}}{\lambda_t}. \quad (3.46)$$

Nesse passo, o acréscimo aditivo natural na variância que surge das equações de atualização foi substituído por um fator de desconto  $\lambda_t$  a ser determinado.

Pelo Teorema de Bayes, a distribuição *a posteriori*,  $(h_t|D_t)$ , é proporcional à distribuição *a priori*,  $(h_t|D_{t-1})$ , atualizada pela verossimilhança:

$$p(h_t|D_t) \propto p(y_t|h_t)p(h_t|D_{t-1}). \quad (3.47)$$

Dessa maneira, a distribuição *a posteriori* no lado esquerdo de (3.47) será aproximada por uma distribuição normal cuja média será a moda (ou o máximo) do produto no lado direito.

As distribuições no lado direito de (3.47) são dadas por (2.3) e (3.45). Então, seja  $L(h_t|\mu, \phi)$  dada por:

$$\begin{aligned} L(h_t|\mu, \phi) &\propto p(y_t|h_t)p(h_t|D_{t-1}), \\ &\propto e^{-\frac{h_t}{2}} \exp \left\{ -\frac{1}{2} e^{-h_t} y_t^2 \right\} \times \\ &\times R_t^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} R_t^{-1} [h_t - (\mu(1 - \phi) + \phi \hat{m}_{t-1})]^2 \right\}. \end{aligned} \quad (3.48)$$

Toma-se o logaritmo de (3.48), de modo que,  $\ln L(h_t|\mu, \phi) = l(h_t|\mu, \phi)$ . Assim,

$$l(h_t|\mu, \phi) \propto -\frac{h_t}{2} - \frac{y_t^2}{2e^{h_t}} - \frac{1}{2} \ln R_t - \frac{1}{2R_t} [h_t - (\mu(1 - \phi) + \phi \hat{m}_{t-1})]^2. \quad (3.49)$$

Com isso, iterativamente ao longo das  $N$  observações, o estimador,  $\hat{h}_t$ , da variável latente,  $h_t$ , será:

$$\hat{h}_t = \hat{m}_t = \hat{m}_{t-1} - \frac{l'(\hat{m}_{t-1})}{l''(\hat{m}_{t-1})}. \quad (3.50)$$

Para completar o processo de estimação de  $h_t$  é preciso calcular as derivadas de (3.49). A algebra é muito simples e a primeira derivada é:

$$\frac{dl(h_t|\mu, \phi)}{dh_t} = l'(h_t|\mu, \phi) = -\frac{1}{2} + \frac{y_t^2}{2e^{h_t}} - \frac{1}{R_t} [h_t - (\mu(1 - \phi) + \phi \hat{m}_{t-1})]. \quad (3.51)$$

A segunda derivada, por sua vez, é:

$$\frac{d^2 l(h_t|\mu, \phi)}{dh_t^2} = l''(h_t|\mu, \phi) = -\frac{y_t^2}{2e^{h_t}} - \frac{1}{R_t}. \quad (3.52)$$

A variância da distribuição de  $h_t$  ao longo do processo é atualizada por:

$$\hat{C}_t = -\frac{1}{l''(\hat{m}_{t-1})}. \quad (3.53)$$

O valor do fator de desconto,  $\lambda_t$ , é calculado de modo a maximizar a distribuição preditiva ( $y_t|D_{t-1}$ ):

$$p(y_t|D_{t-1}, \lambda_t) = \int_{h_t} p(y_t|h_t, D_{t-1})p(h_t|D_{t-1}, \lambda_t)dh_t. \quad (3.54)$$

Mas como não existe forma fechada para a integral em (3.54), é tomada uma aproximação de Laplace (Tierney e Kadane, 1986) para essa distribuição preditiva:

$$p(y_t|D_{t-1}, \lambda_t) \approx (2\pi)^{\frac{1}{2}} \left| \frac{1}{l''(\hat{m}_t)} \right|^{\frac{1}{2}} p(y_t|\hat{h}_t, D_{t-1})p(\hat{h}_t|D_{t-1}, \lambda_t). \quad (3.55)$$

Onde,  $p(y_t|\hat{h}_t, D_{t-1})$  é a distribuição das observações dada em (2.3) e  $p(\hat{h}_t|D_{t-1}, \lambda_t)$  é dada em (3.45).

Assim, a metodologia proposta para se estimar os parâmetros do modelo de volatilidade estocástica a partir da adaptação inspirada em McCormick et al. (2012) acrescida à estratégia ASIS (Kastner e Frühwirth-Schnatter, 2014) está concluída. As próximas etapas consistem em avaliar a consistência desse método alternativo utilizando dados simulados e testar também em dados reais.

# Capítulo 4

## Aplicação da Metodologia Proposta

A proposta de estimação dos parâmetros do modelo de volatilidade estocástica foi testada inicialmente num conjunto de dados artificiais e, posteriormente, num conjunto de dados reais. Os relatos que seguem nesse capítulo mostram como o processo foi implementado e avaliado.

### 4.1 Definição dos Dados Simulados

Foram geradas algumas séries de dados com parâmetros conhecidos segundo o modelo canônico em (2.3). As populações se caracterizam pelo produto cruzado dos possíveis valores predeterminados para  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$  a seguir:

- $\mu = -5,4$ ,
- $\phi \in \{0,50; 0,80; 0,90; 0,99\}$ ,
- $\sigma_\eta^2 \in \{0,1^2; 0,3^2; 0,5^2\}$ .

No caso de  $\mu$ , a opção de se trabalhar com um único valor fixado é devido ao conhecimento prévio de que tal parâmetro não apresenta maiores dificuldades técnicas em sua estimação, como cita (Kastner, 2015).

A geração dos dados foi implementada em R (Ihaka e Gentleman, 1996). O Código A.1 no apêndice A apresenta a função `sv_data` que é responsável por gerar os dados



simulados do modelo de volatilidade estocástica definido em (2.3). Essa função possui quatro parâmetros: o tamanho da amostra desejada `n`, o valor `mu` do nível  $\mu$ , o valor `phi` da persistência  $\phi$  e o valor `sigma` do desvio padrão  $\sigma_\eta$ . O resultado da função é um objeto do tipo `data.frame`, que na verdade é um conjunto de dados com os valores simulados `y` da variável observada  $Y_t$  e `h` da variável latente  $h_t$ .

Foram gerados alguns conjuntos de dados para os 12 possíveis valores distintos de  $\psi = (\mu, \phi, \sigma_\eta^2)$ . Todas as populações tinham o mesmo tamanho de  $n = 4 \times 365,25 = 1.461$  elementos, o que corresponde a quatro anos de observações diárias.

A Figura 4.1 mostra um exemplo de um conjunto de dados simulados com a persistência  $\phi = 0,90$ , e a volatilidade  $\sigma_\eta^2 = 0,5^2$ .

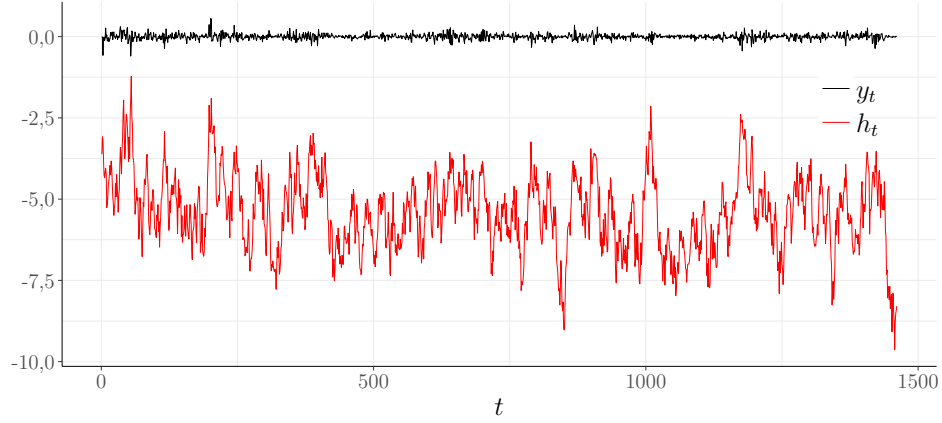


Figura 4.1: Conjunto de dados simulados a partir do modelo canônico em (2.3),  $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,5^2$ .

## 4.2 Estimação dos Parâmetros para os Dados Simulados

Inicialmente, a metodologia proposta também foi implementada em R. Porém, devido ao processo ser extremamente demorado, o projeto final foi totalmente programado em C (Kernighan e Ritchie, 1988). O Código A.2 no apêndice A mostra o código-fonte da implementação do algoritmo de estimação dos parâmetros.

Nessa e em todas as demais execuções do algoritmo, foram tomadas 15.000 iterações, sendo as 5.000 iniciais descartadas (*burn-in*), além de 9 a cada 10 iterações consecuti-

vas (*thin*). O resultado, portanto, foi sempre uma amostra de 1.000 observações “válidas”. A escolha desses valores foi baseada em Achcar et al. (2011). As distribuições *a priori* de  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$  são definidas como sugere Kim et al. (1998) e Kastner e Frühwirth-Schnatter (2014):

$$\begin{aligned}\mu &\sim \mathcal{N}(0, 100), \\ \phi_0 &\sim \mathcal{B}\left(20, \frac{3}{2}\right), \\ \sigma_\eta^2 &\sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2}\right).\end{aligned}\tag{4.1}$$

O que chama atenção é a distribuição *a priori* de  $\phi$ , que é bastante informativa. Kim et al. (1998) justifica isso no fato de que uma *priori* não informativa gera problemas quando a série de dados está próxima da não estacionariedade.

O primeiro passo foi avaliar a performance do processo de estimação da variável latente,  $h_t$ , conhecidos os valores reais dos parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$ . Para isso, os valores reais de  $h_t$  foram comparados através de gráficos com os valores médios dos 1.000 valores estimados de  $\hat{h}_t$ :

$$\bar{\hat{h}}_t = \frac{\sum_{j=1}^{1000} \hat{h}_{t,j}}{1000}.\tag{4.2}$$

O intervalo dos possíveis valores do fator de desconto,  $\lambda_t$ , foi limitado ao intervalo (0,75; 1). Tal intervalo foi definido pois, em estudos preliminares, foi constatado uma tendência de obtenção de valores de  $\lambda_t$  bem baixos, o que aumenta o ruído nas estimativas. O valor inicial,  $h_0$ , foi tomado aleatoriamente de sua respectiva distribuição inicial conforme o modelo (2.3).

A Figura 4.2 mostra os perfis da série real de  $h_t$  e de sua estimativa  $\bar{\hat{h}}_t$ . Como a própria visualização das séries sugere, os valores médios estimados,  $\bar{\hat{h}}_t$ , convergem de forma aceitável aos valores reais,  $h_t$ . E a aplicação da metodologia inspirada em McCormick et al. (2012) proporcionou resultados bastante satisfatórios para esse conjunto de dados. O intervalo de credibilidade de 95% foi calculado, e omitido do gráfico intencionalmente por uma questão estética, pois sua amplitude média foi inferior a 0,001, corroborando

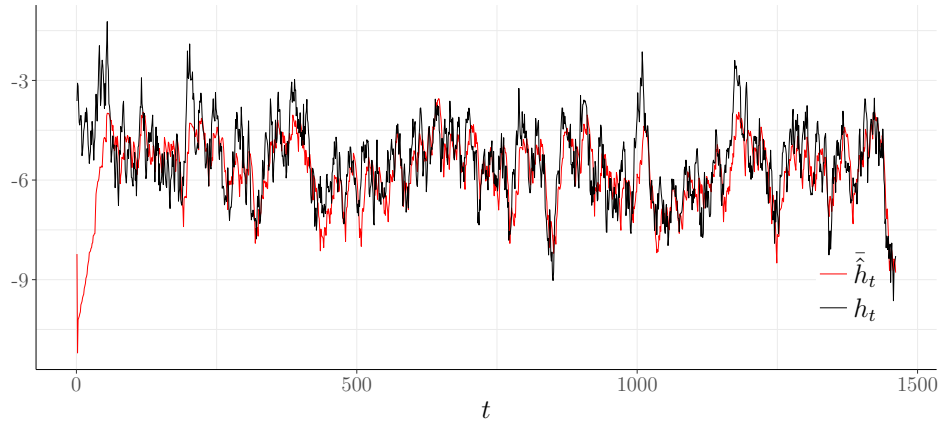


Figura 4.2: Valores médios estimados de  $h_t$  do modelo (2.3) quando os parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$  são conhecidos ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,5^2$ ).

mais uma vez com a avaliação positiva sobre a aplicabilidade da técnica.

Todavia, vale destacar que o processo se deu com os valores reais de  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$ . A próxima etapa então é agregar o algoritmo de estimação de  $h_t$  via McCormick et al. (2012) com a estratégia proposta em Kastner e Frühwirth-Schnatter (2014).

A primeira execução resultou num erro inesperado já de início. A variância de  $(h_t|D_t)$  definida em (3.53) vai diminuindo a cada iteração, até chegar num valor computacionalmente igual a zero. Assim, as estimativas de  $h_t$  são todas nulas.

Num processo de tentativa e erro, foi constatado que o parâmetro  $\phi$  é extremamente dominante no algoritmo proposto para se estimar a variável latente  $h_t$  inspirado em McCormick et al. (2012). A Figura 4.3 mostra o resultado da estimação da variável latente quando todos os parâmetros são desconhecidos, exceto  $\phi$ . Claramente é difícil perceber, a olho nú, a diferença na série de valores estimados de  $h_t$  do procedimento quando todos os parâmetros são conhecidos (Figura 4.2) e apenas a persistência  $\phi$  é conhecida (Figura 4.3).

A primeira alternativa de estimar todos os parâmetros estáticos, inclusive  $\phi$ , baseada no trabalho de Kastner e Frühwirth-Schnatter (2014) (no entanto, outros autores também a adotam), foi de tomar todos os parâmetros desconhecidos, porém partir do valor inicial de  $\phi$  igual ao verdadeiro valor deste parâmetro. Curiosamente, o processo volta a funcionar de forma satisfatória, conforme é mostrado na Figura 4.4, o que prova, mais uma vez, a extrema importância de  $\phi$  nessa metodologia.

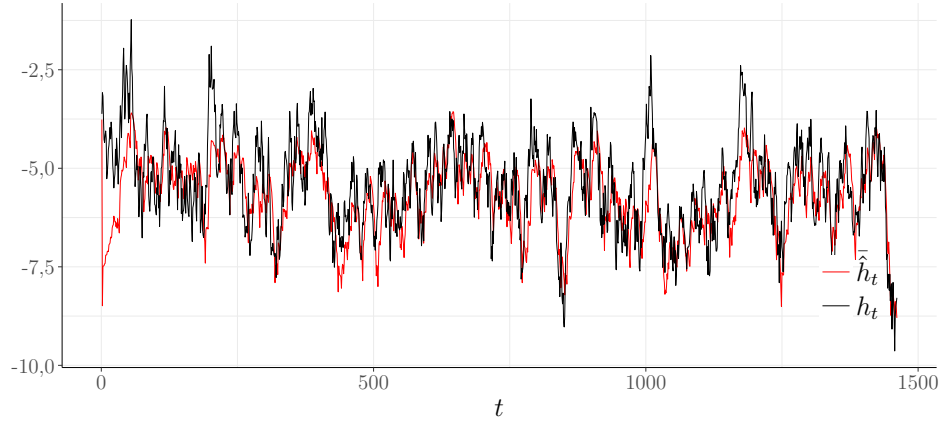


Figura 4.3: Valores médios estimados de  $h_t$  do modelo (2.3) quando apenas o parâmetro  $\phi$  é conhecido ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,5^2$ ).

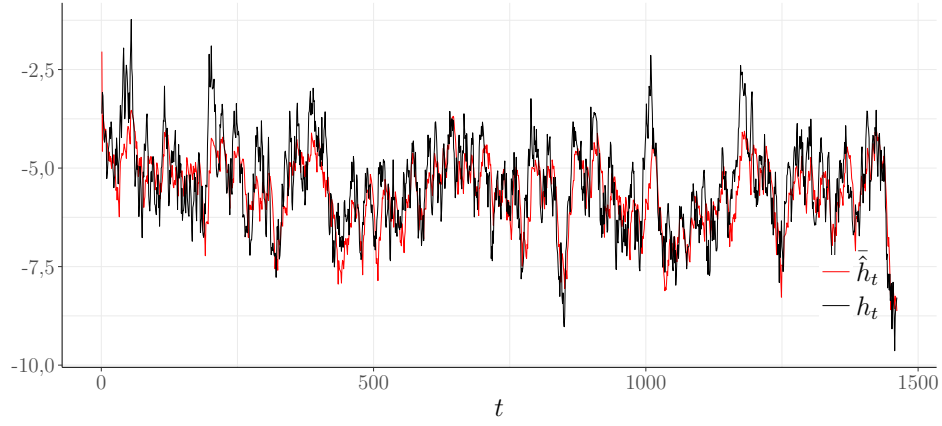


Figura 4.4: Valores médios estimados de  $h_t$  do modelo (2.3) quando o valor inicial de  $\phi$  é igual ao valor verdadeiro do parâmetro ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,5^2$ ).

Uma vez que iniciar o algoritmo a partir dos valores reais (mesmo que de uma variável somente) não é uma condição desejável, o plano então foi o de tentar resolver esse problema da estimação de  $\phi$  de outra maneira.

O termo em (3.46),

$$R_t = \frac{\phi^2 \hat{C}_{t-1}}{\lambda_t}, \quad (4.3)$$

presente nas derivadas (3.51) e (3.52) que surgem no algoritmo para estimar a variável latente  $h_t$ , é quem está fazendo com que a variância do candidato à observação *a posteriori* de  $h_t$  fique próxima de zero. Como  $\phi$  é um parâmetro do MVE e  $\hat{C}_{t-1}$  é um valor a ser estimado ao longo das iterações, o único termo que sobra para “contrabalancear” os baixos valores gerados de  $R_t$  é o fator  $\lambda_t$ .

Como citado anteriormente, os valores do fator de desconto,  $\lambda_t$ , foram limitados ao intervalo  $(0,75; 1)$ , devido a constatação inicial de uma tendência de polarização dos valores de  $\lambda_t$ , especialmente no limite inferior do intervalo. Desse modo, numa nova tentativa, esse intervalo limitante foi deixado de lado e o valor do fator de desconto foi fixado em  $\lambda_t = 0,50$ .

A Figura 4.5 mostra os valores médios de  $\hat{h}_t$  obtidos no processo de estimação considerando  $\lambda_t = 0,50$  e  $\phi$  estimado a partir de um valor inicial diferente do valor real. Observa-se que a série estimada de  $\hat{h}_t$  apresenta valores que subestimam o valores reais da variável latente,  $h_t$ , e apresenta incerteza maior do que os casos analisados em que o valor verdadeiro de  $\phi$  foi fixado.

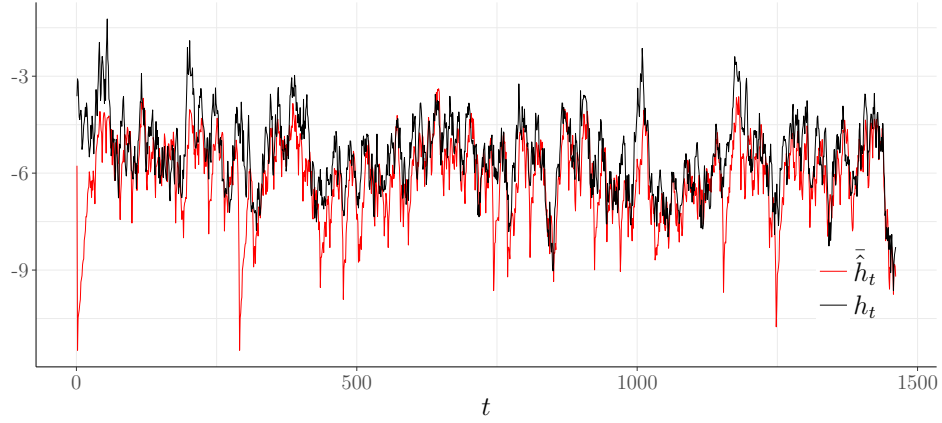


Figura 4.5: Valores médios estimados de  $h_t$  do modelo (2.3) quando  $\lambda_t = 0,50$  ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,5^2$ ).

Desse modo, os parâmetros do modelo de volatilidade estocástica foram estimados para o conjunto de populações simuladas fixando-se o valor de  $\lambda_t$  próximo do máximo valor possível, em cada caso. Os resultados podem ser vistos na Tabela 4.1 e nas Figuras 4.6 a 4.17, a seguir.

A Tabela 4.1 mostra o valor real do parâmetro, a média, o quantil de 2,5%, a mediana e o quantil de 95% dos valores estimados de cada um dos três parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$  do modelo de volatilidade estocástica obtidos através da estimativa proposta nesta dissertação. Já as Figuras 4.6 a 4.17 mostram respectivamente a série de valores “válidos” sequencialmente obtidos e o histograma representando a distribuição empírica, também dos parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$  do modelo.

O que é possível aprender pela Tabela 4.1 e pelas Figuras 4.6 a 4.17 é que os valores do parâmetro  $\mu$  são bem estimados em todas as circunstâncias. Os valores médios apresentam pouquíssimo viés, independente do valor verdadeiro dos parâmetros do conjunto de dados simulados, e o intervalo de dispersão é praticamente o mesmo, exceto quando a persistência  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,25$ . Os valores estimados de  $\sigma_\eta^2$  apresentam valores aceitáveis quando o verdadeiro valor deste parâmetro não é muito pequeno ( $< 0,09$ ). Já o parâmetro  $\phi$  parece preso num poço de potencial. Independente do seu valor verdadeiro, são amostrados valores altíssimos da persistência. Na tentativa de investigar o porquê desses valores discrepantes estimados para o parâmetro  $\phi$ , foram feitas novas escolhas de distribuições *a priori* para  $\phi$ . Porém, essa situação insistiu em persistir.

Um outro ponto a ser notado é que a escolha inicial de executar o algoritmo com 15.000 iterações (sendo as 5.000 iniciais descartadas) baseada em Achcar et al. (2011) se provou adequada. As séries geradas de valores sequenciais dos parâmetros nas Figuras 4.6 a 4.17 mostram que, em todos os casos, os parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$  estão estáveis em torno de um ponto fixo.

O modelo canônico em (2.3) será implementado e estimado também via JAGS na próxima seção, para efeitos de comparação com a técnica proposta nesta dissertação.

Tabela 4.1: Estatísticas dos valores estimados dos parâmetros.

Parâmetro	Real	Média	2,5%	Mediana	97,5%
$\mu$	-5,4	-5,672	-9,511	-5,462	-2,951
$\phi$	0,99	0,990	0,982	0,990	0,996
$\sigma_\eta^2$	0,25	0,268	0,187	0,264	0,374
$\mu$	-5,4	-5,557	-7,835	-5,477	-3,943
$\phi$	0,99	0,991	0,984	0,991	0,997
$\sigma_\eta^2$	0,09	0,061	0,042	0,058	0,094
$\mu$	-5,4	-5,400	-6,223	-5,364	-4,658
$\phi$	0,99	0,981	0,971	0,981	0,991
$\sigma_\eta^2$	0,01	0,071	0,044	0,069	0,112
$\mu$	-5,4	-5,585	-6,944	-5,471	-4,632
$\phi$	0,9	0,976	0,965	0,977	0,987
$\sigma_\eta^2$	0,25	0,228	0,095	0,236	0,341
$\mu$	-5,4	-5,388	-6,324	-5,337	-4,589
$\phi$	0,9	0,981	0,970	0,980	0,991
$\sigma_\eta^2$	0,09	0,095	0,055	0,093	0,148
$\mu$	-5,4	-5,548	-6,200	-5,523	-4,971
$\phi$	0,9	0,965	0,943	0,966	0,981
$\sigma_\eta^2$	0,01	0,154	0,085	0,151	0,244
$\mu$	-5,4	-5,575	-6,749	-5,473	-4,751
$\phi$	0,8	0,973	0,960	0,973	0,985
$\sigma_\eta^2$	0,25	0,275	0,122	0,283	0,409
$\mu$	-5,4	-5,390	-6,203	-5,341	-4,737
$\phi$	0,8	0,977	0,965	0,977	0,988
$\sigma_\eta^2$	0,09	0,110	0,050	0,107	0,191
$\mu$	-5,4	-5,481	-6,000	-5,470	-4,992
$\phi$	0,8	0,958	0,932	0,959	0,978
$\sigma_\eta^2$	0,01	0,136	0,073	0,131	0,232
$\mu$	-5,4	-5,461	-6,588	-5,359	-4,758
$\phi$	0,5	0,970	0,957	0,970	0,982
$\sigma_\eta^2$	0,25	0,282	0,107	0,292	0,446
$\mu$	-5,4	-5,445	-6,179	-5,401	-4,832
$\phi$	0,5	0,974	0,960	0,974	0,986
$\sigma_\eta^2$	0,09	0,127	0,056	0,124	0,224
$\mu$	-5,4	-5,499	-6,262	-5,451	-4,859
$\phi$	0,5	0,975	0,962	0,975	0,987
$\sigma_\eta^2$	0,01	0,127	0,053	0,122	0,228

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,5^2$

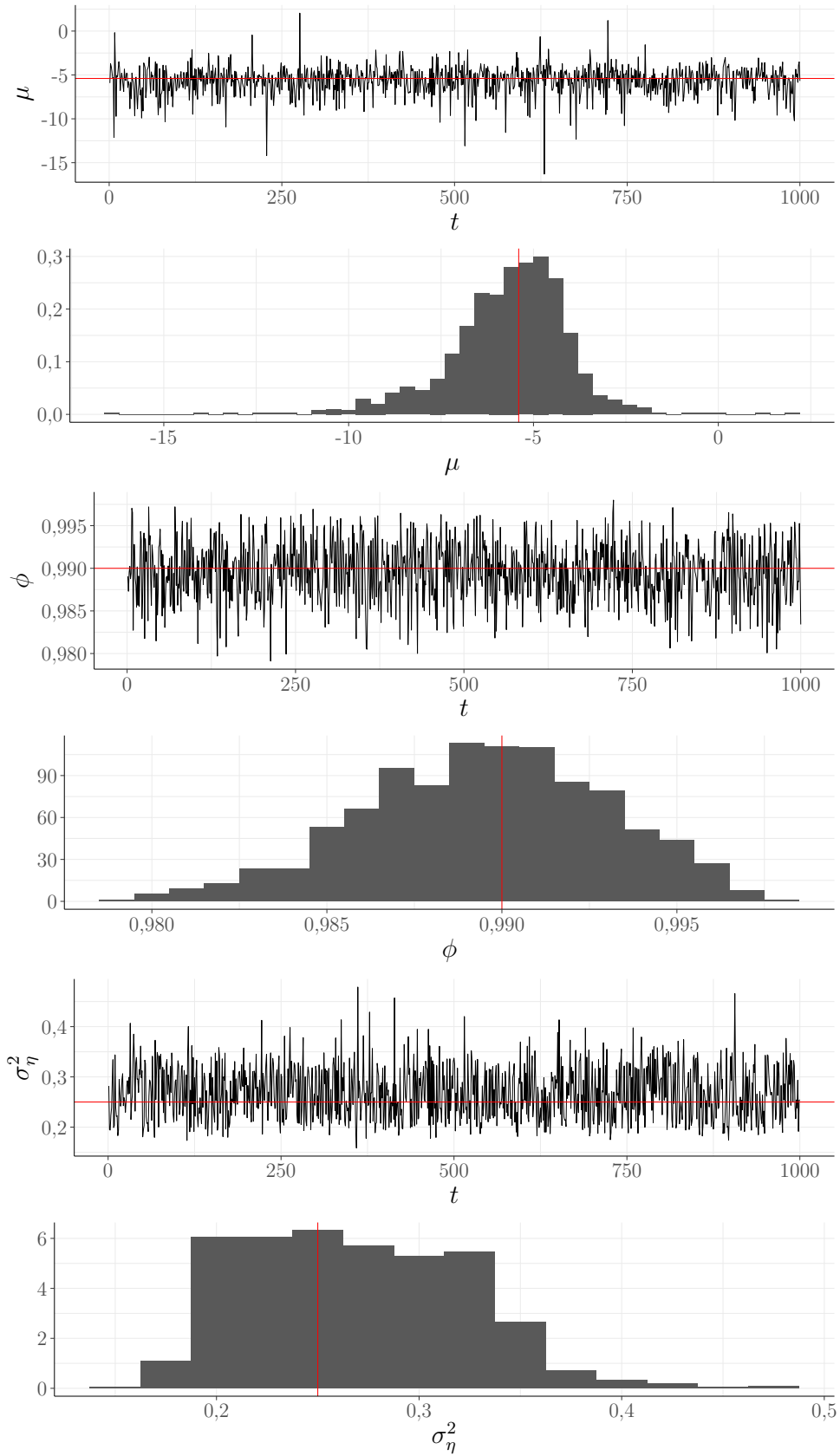


Figura 4.6: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,5^2$ ).



Valores reais:  $\mu = -5,4$ ,  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,3^2$

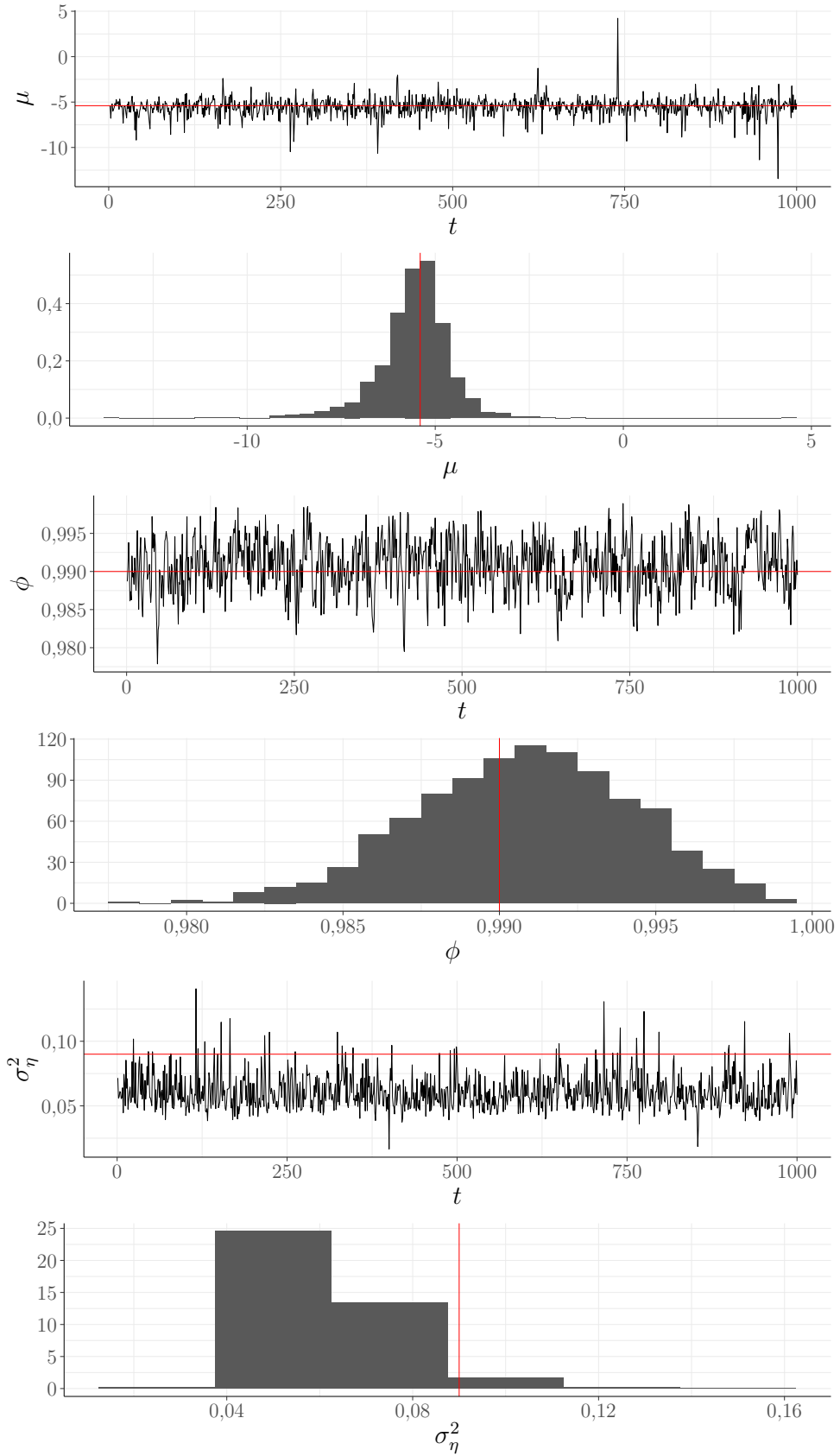


Figura 4.7: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,3^2$ ).

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,1^2$

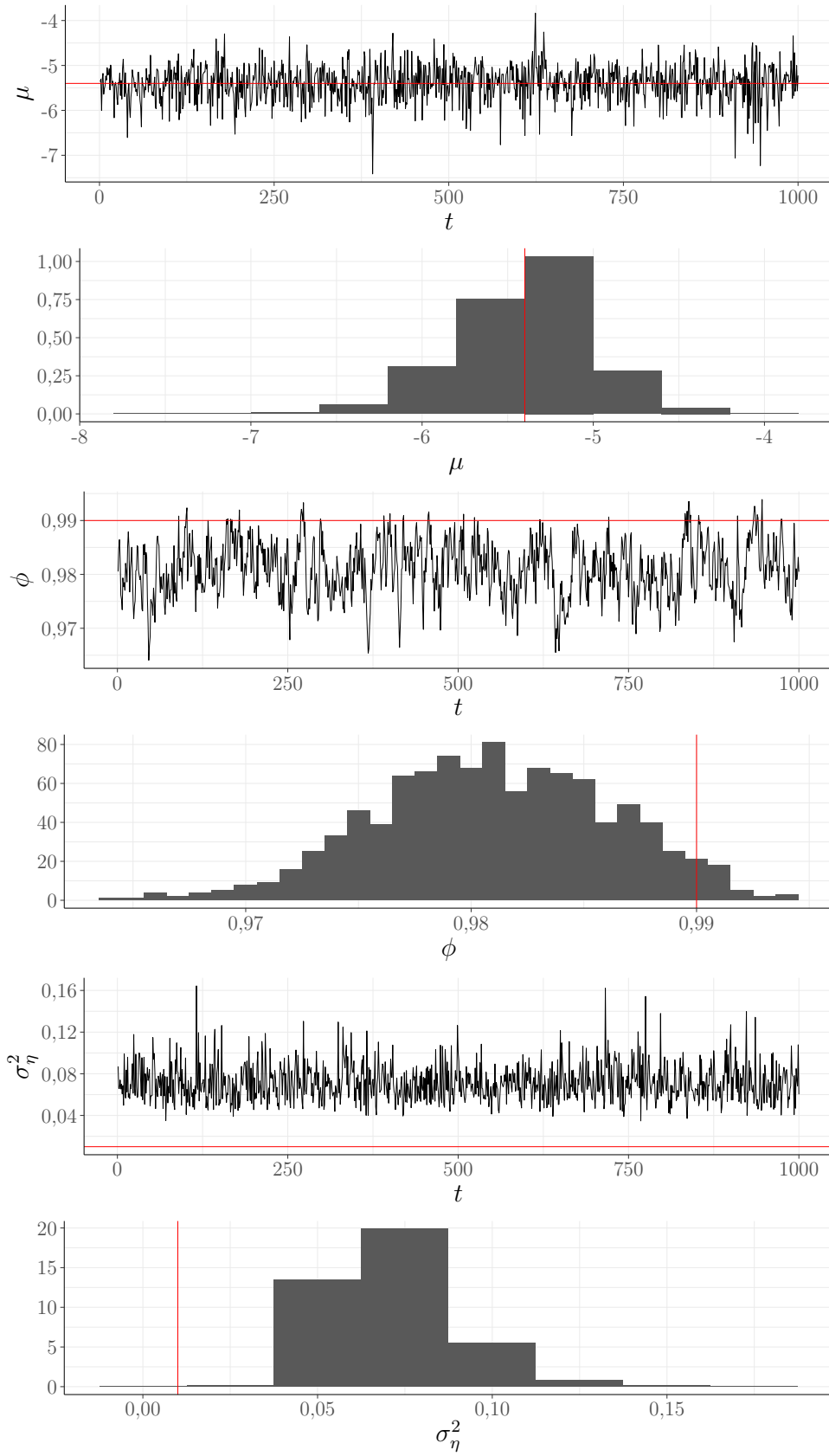


Figura 4.8: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,1^2$ ).

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,5^2$

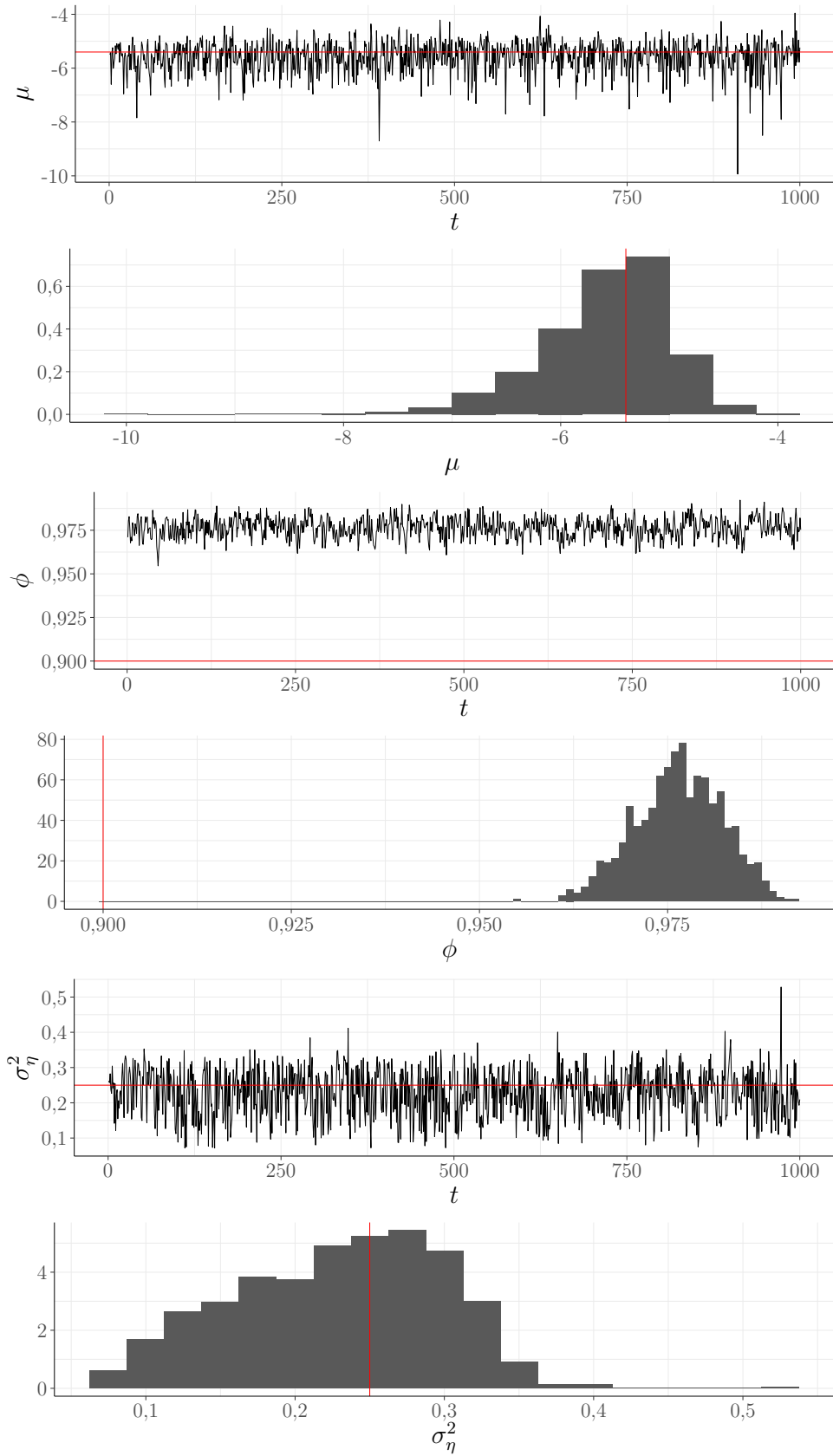


Figura 4.9: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,5^2$ ).

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,3^2$

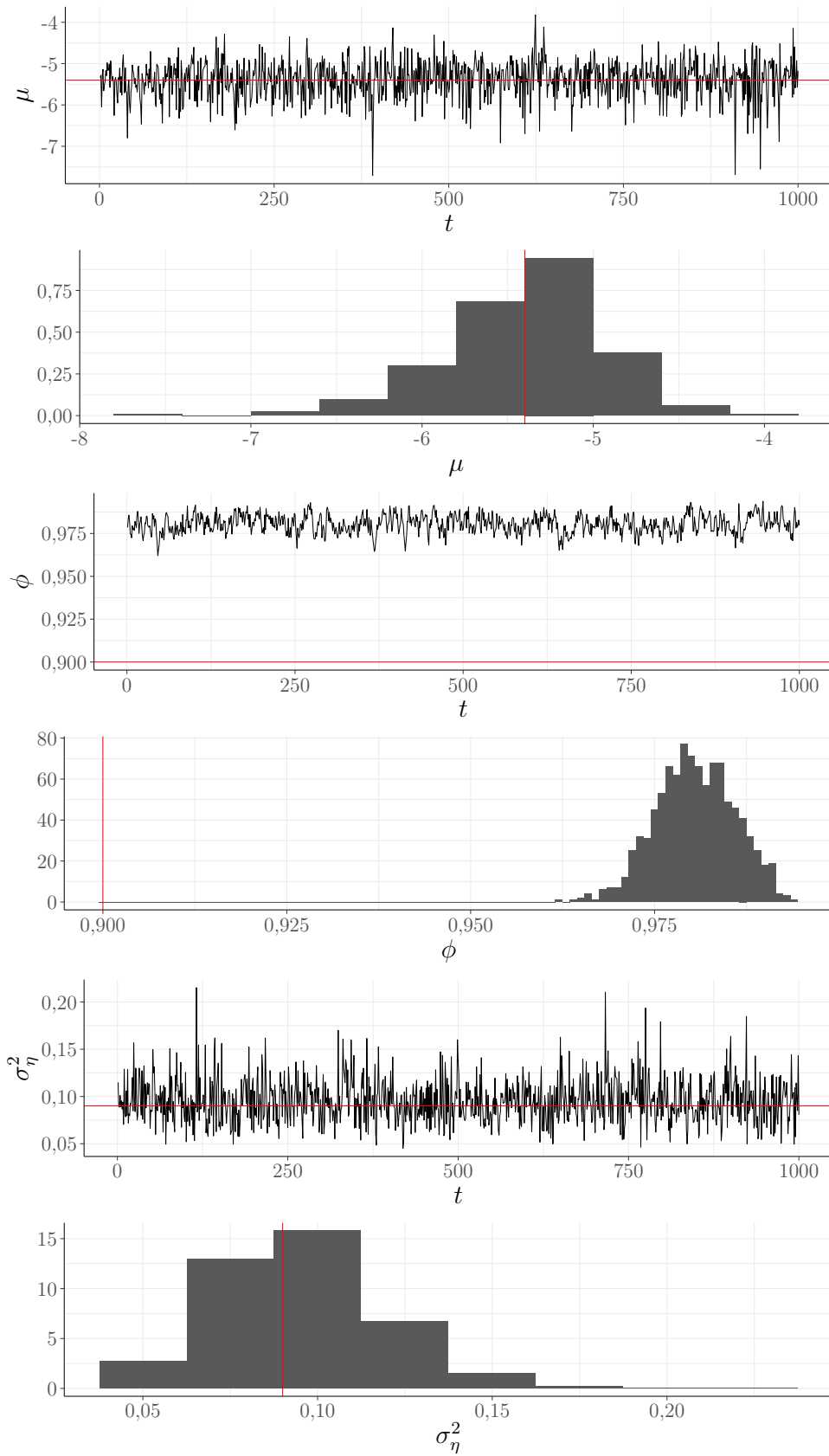


Figura 4.10: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,3^2$ ).

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,1^2$

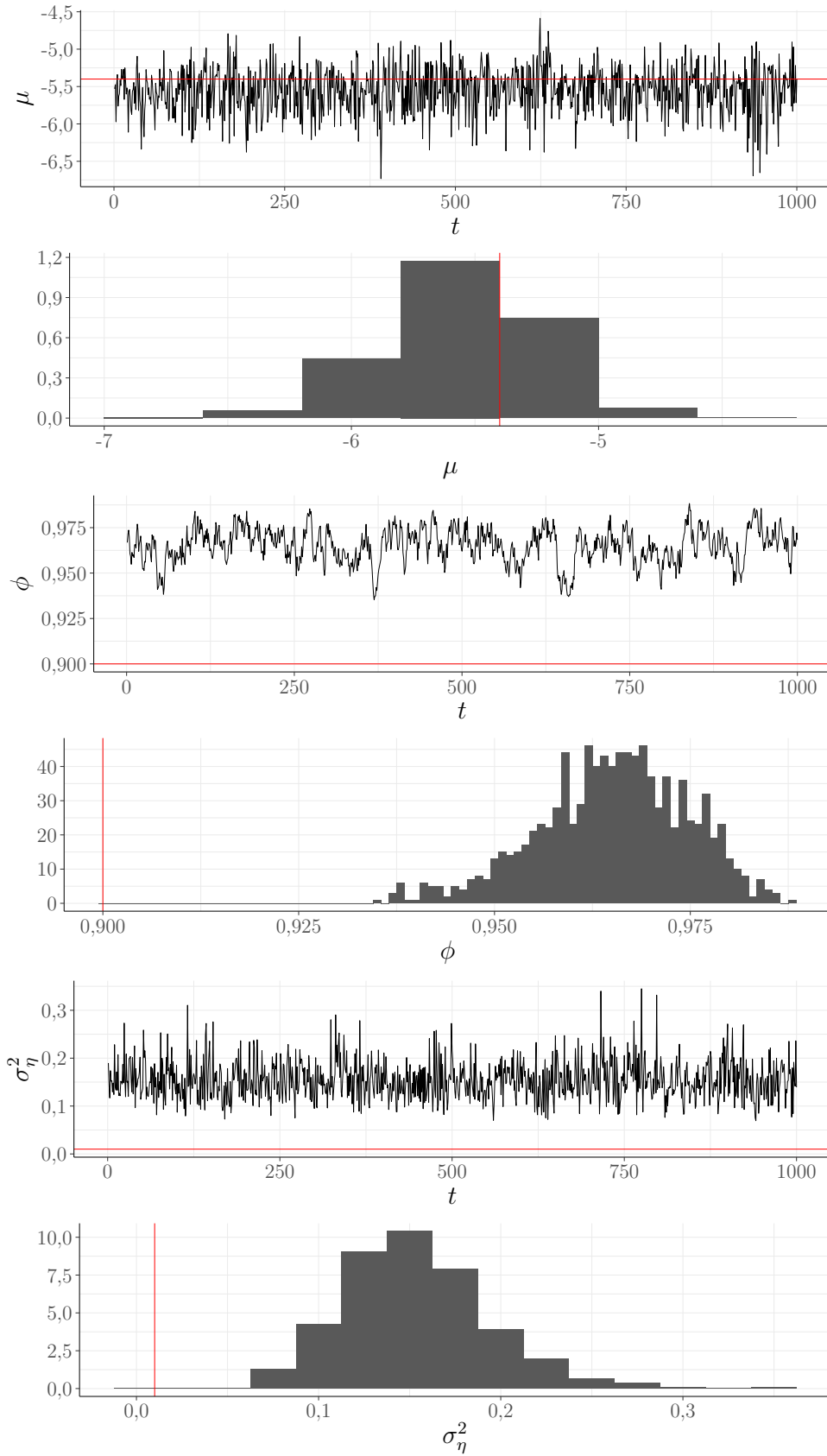


Figura 4.11: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,1^2$ ).

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,80$  e  $\sigma_\eta^2 = 0,5^2$

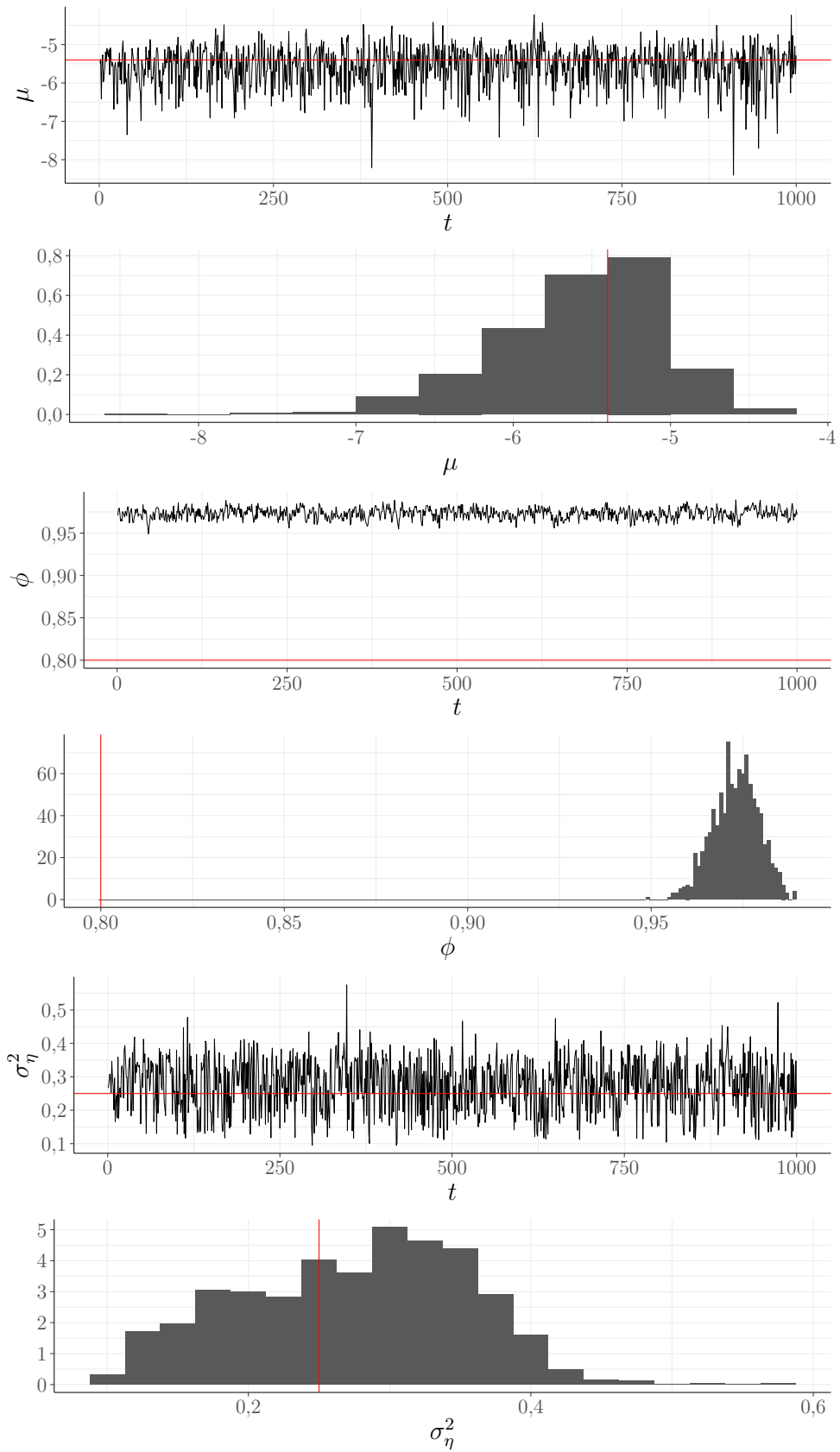


Figura 4.12: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,80$  e  $\sigma_\eta^2 = 0,5^2$ ).

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,80$  e  $\sigma_\eta^2 = 0,3^2$

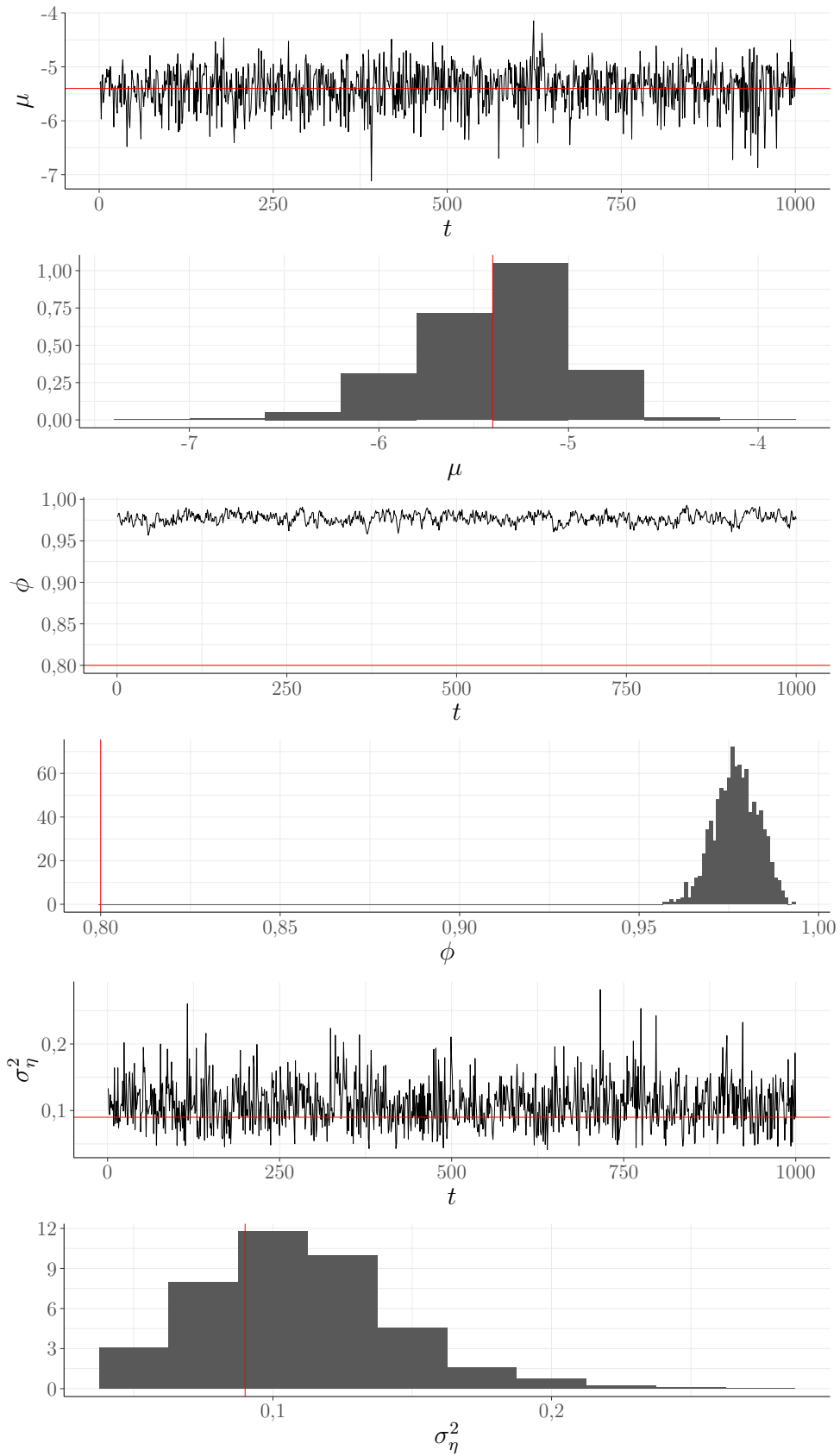


Figura 4.13: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,80$  e  $\sigma_\eta^2 = 0,3^2$ ).

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,80$  e  $\sigma_\eta^2 = 0,1^2$

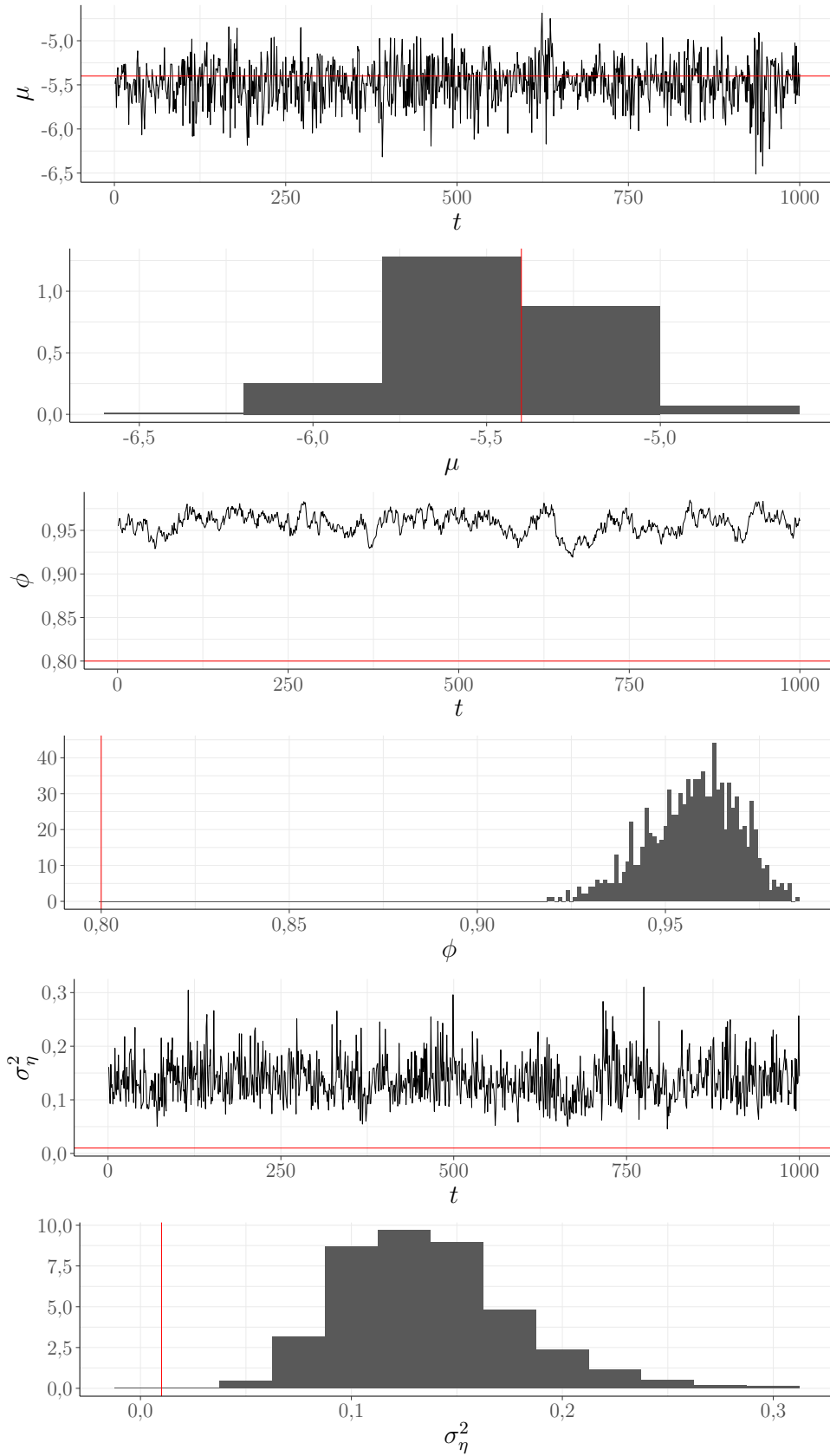


Figura 4.14: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,80$  e  $\sigma_\eta^2 = 0,1^2$ ).



Valores reais:  $\mu = -5,4$ ,  $\phi = 0,50$  e  $\sigma_\eta^2 = 0,5^2$

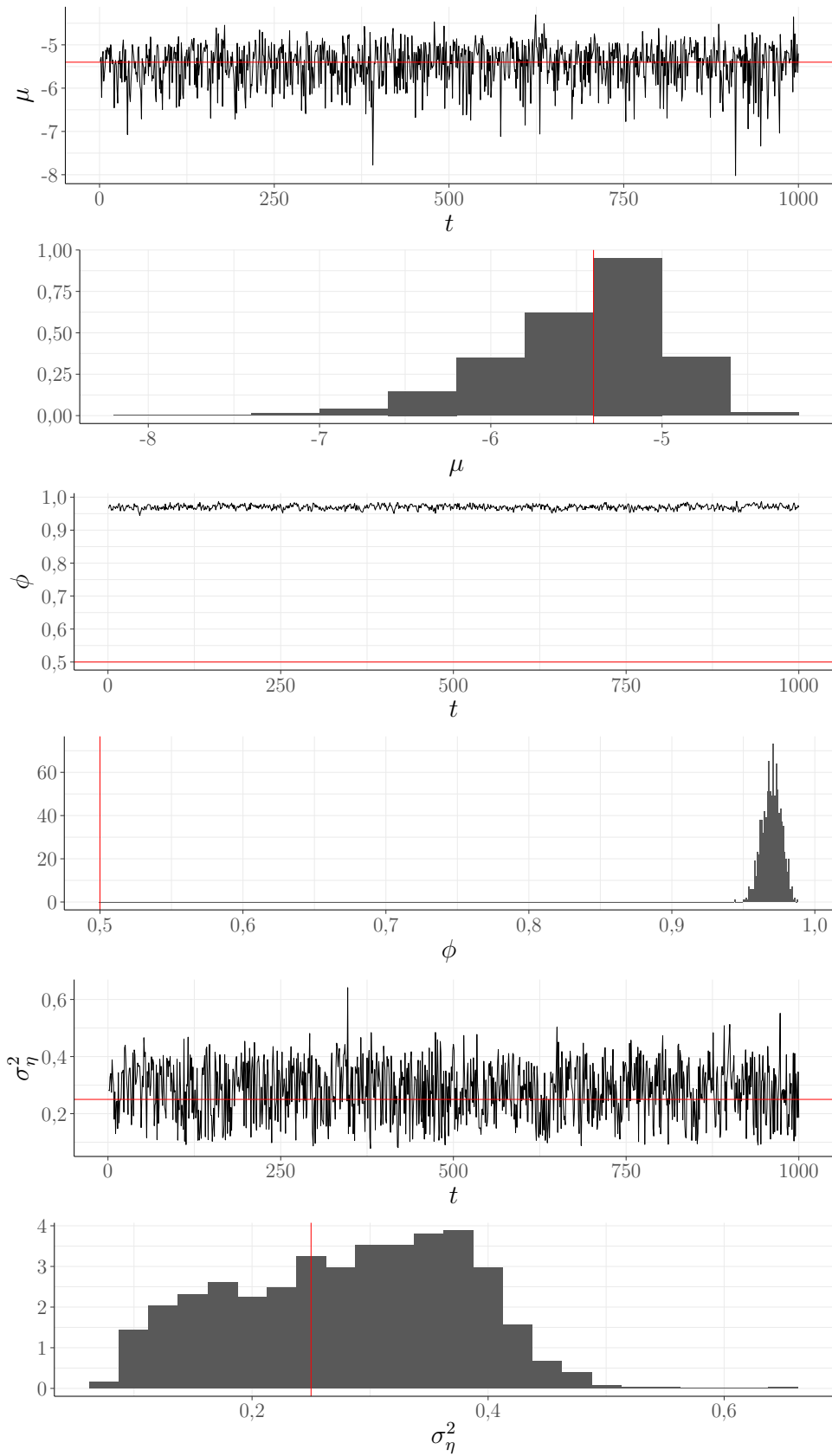


Figura 4.15: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,50$  e  $\sigma_\eta^2 = 0,5^2$ ).

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,50$  e  $\sigma_\eta^2 = 0,3^2$

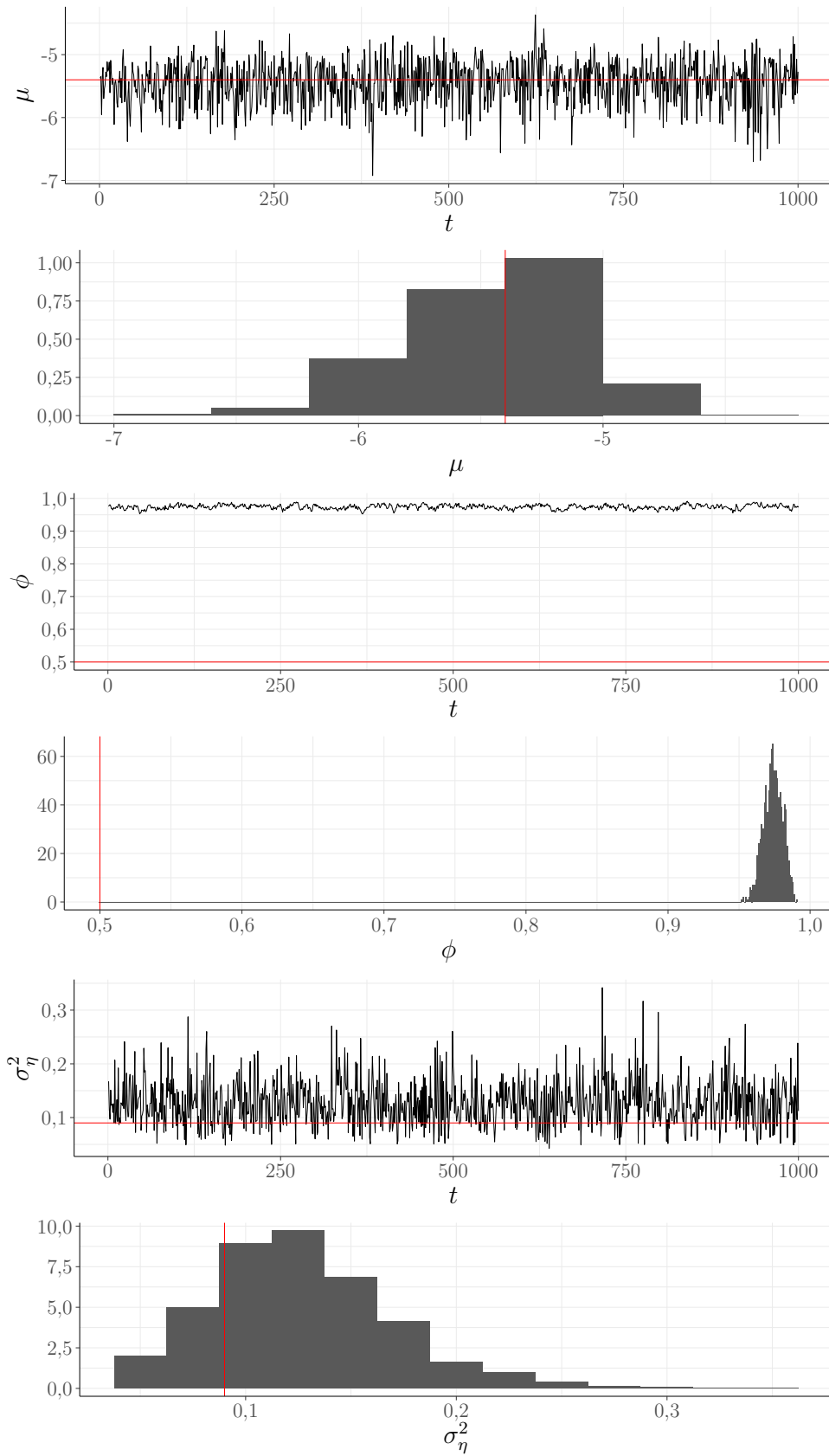


Figura 4.16: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,50$  e  $\sigma_\eta^2 = 0,3^2$ ).

Valores reais:  $\mu = -5,4$ ,  $\phi = 0,50$  e  $\sigma_\eta^2 = 0,1^2$

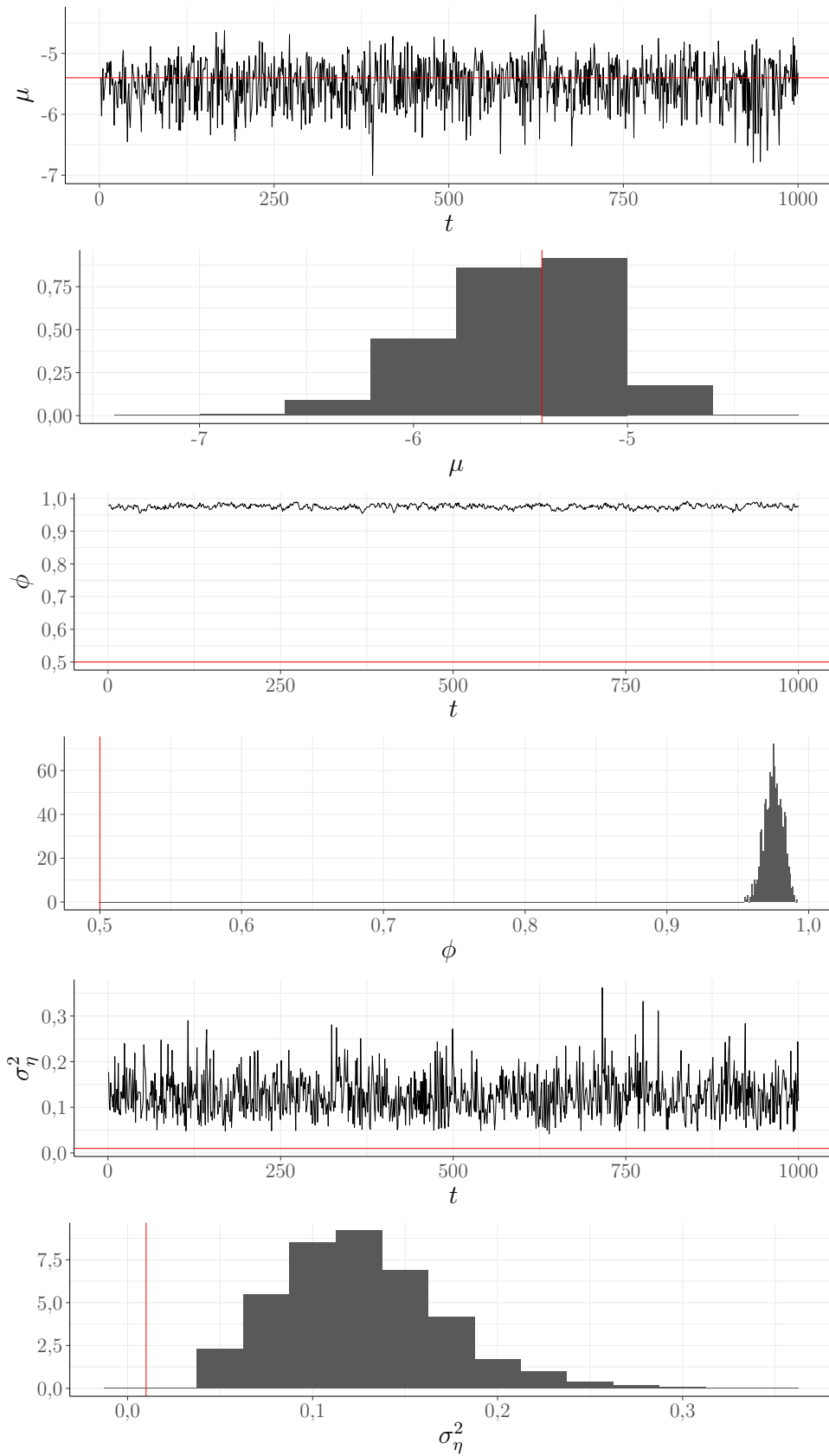


Figura 4.17: Valores estimados ( $\mu = -5,4$ ,  $\phi = 0,50$  e  $\sigma_\eta^2 = 0,1^2$ ).

### 4.3 Comparação dos Resultados Obtidos com o JAGS

A estimação dos parâmetros do modelo de volatilidade estocástica será feita nessa etapa a partir do amostrador de Gibbs implementado no *software* JAGS, Plummer (2003). JAGS é um acrônimo para a expressão em inglês *Just Another Gibbs Sampler*, e sua escolha foi motivada por dois fatores.

O primeiro motivo da escolha do JAGS se deve ao desenvolvimento e manutenção do programa. Enquanto a solução mais tradicional, WinBUGS (Lunn et al., 2000), foi descontinuada em 2007, o desenvolvimento do JAGS permanece ativo. Assim, a princípio, é um programa menos suscetível a *bugs* e em constante aprimoramento. O segundo motivo é devido a ótima integração com o R, através do pacote `R2jags` (Su e Yajima, 2015).

O JAGS, apesar de integrado ao R, possui uma linguagem própria, descrita em Plummer (2015). As variáveis do modelo estudado devem ser declaradas segundo a sua natureza (determinística ou estocástica). Sua sintaxe é muito parecida com R. Contudo, deve-se tomar cuidado quanto à implementação das distribuições. A normal, por exemplo, foi programada a partir de sua precisão ( $\tau = \sigma^{-2}$ ).

O código-fonte do modelo canônico de volatilidade estocástica (2.3) em JAGS está no Código A.3 do apêndice A. As distribuições de  $Y_t$  e  $h_t$ , que são a base dessa implementação são:

$$Y_t \sim \mathcal{N}(0, e^{h_t}) \text{ e} \quad (4.4)$$

$$(h_t | h_{t-1}) \sim \mathcal{N}(\mu + \phi(h_{t-1} - \mu), \sigma_\eta^2). \quad (4.5)$$

Apesar das facilidades computacionais, especialmente pela simplicidade de se implementar modelos bastante complexos, que o JAGS proporciona, os problemas de estimação dos parâmetros do modelo de volatilidade estocástica persistem e são bem similares com aqueles encontrados através da técnica proposta nessa dissertação.

O procedimento foi exatamente o mesmo feito anteriormente, isto é, o modelo foi calibrado através de um *burn-in* igual a 5.000. Então, mais 10.000 iterações foram executadas, e uma a cada 10 observações foi salva. Resultando, assim, numa amostra  $a$

*posteriori* de tamanho 1.000 para cada um dos três parâmetros em cada população. As distribuições *a priori* foram idênticas às aquelas já descritas em (4.1).

As Figuras 4.18 a 4.29, a seguir, mostram a comparação entre as estimativas do JAGS e as estimativas feitas pelo método proposto nessa dissertação. São apresentadas as densidades empíricas dos valores estimados dos parâmetros  $\mu$ ,  $\phi$  e  $\sigma_\eta^2$  através dos dois métodos.

Da mesma maneira que na metodologia proposta através da composição do método ASIS (Kastner e Frühwirth-Schnatter, 2014) com o método de estimação da variável latente,  $h_t$ , via (McCormick et al., 2012), a estimação através do JAGS gerou boas estimativas para  $\mu$  em todos os casos. Para o parâmetro  $\sigma_\eta^2$ , semelhante ao método anterior, nem sempre obteve-se boas estimativa. À medida que o valor real da persistência  $\phi$  cresce, o JAGS tende a subestimar os valores de  $\sigma_\eta^2$ . E por fim, para o parâmetro  $\phi$ , que apresentou a pior performance pelo método proposto nesta dissertação, obteve-se uma discreta melhora via JAGS. Muito embora esse aprimoramento não deva ser considerado excelente. Nota-se que, mesmo utilizando o JAGS, as estimativas do parâmetro  $\phi$  tiveram uma dispersão bem alta, e as vezes, até mesmo, apresentaram-se bastante viesadas. Sugerindo assim, mais uma vez, que o parâmetro da persistência  $\phi$  seja o “calcanhar de Aquiles” do modelo de volatilidade estocástica.

A próxima, e última, seção desse capítulo apresenta uma aplicação com dados reais da metodologia resultante da composição da estratégia ASIS (Kastner e Frühwirth-Schnatter, 2014) com a técnica adaptada de McCormick et al. (2012) para estimação da variável latente  $h_t$ . Os resultados obtidos também são comparados com as estimativas geradas através do JAGS.

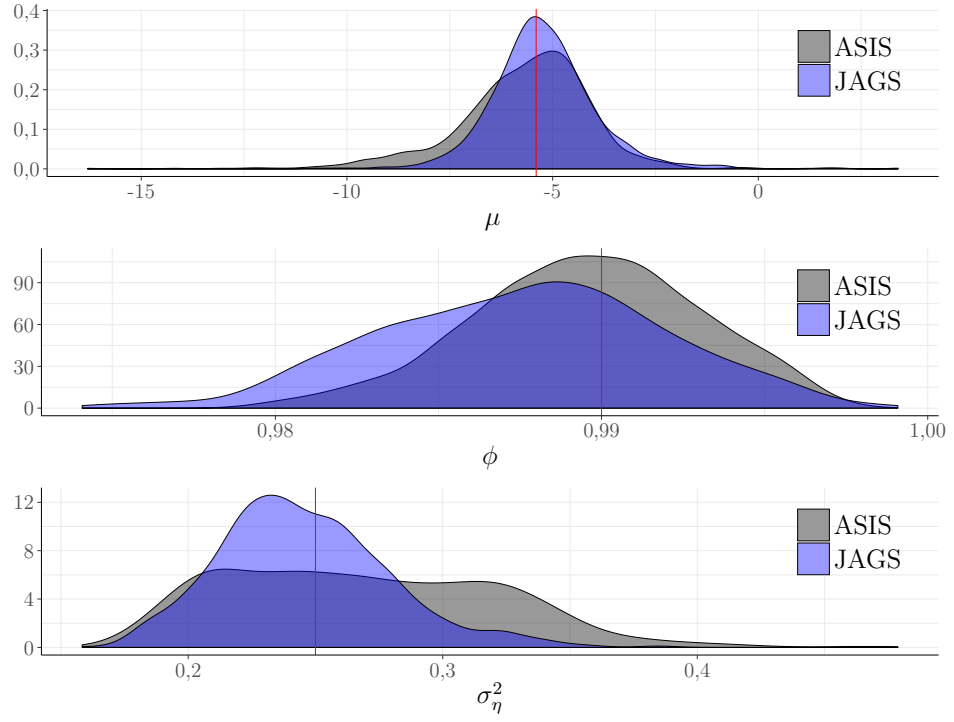


Figura 4.18: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,5^2$ ).

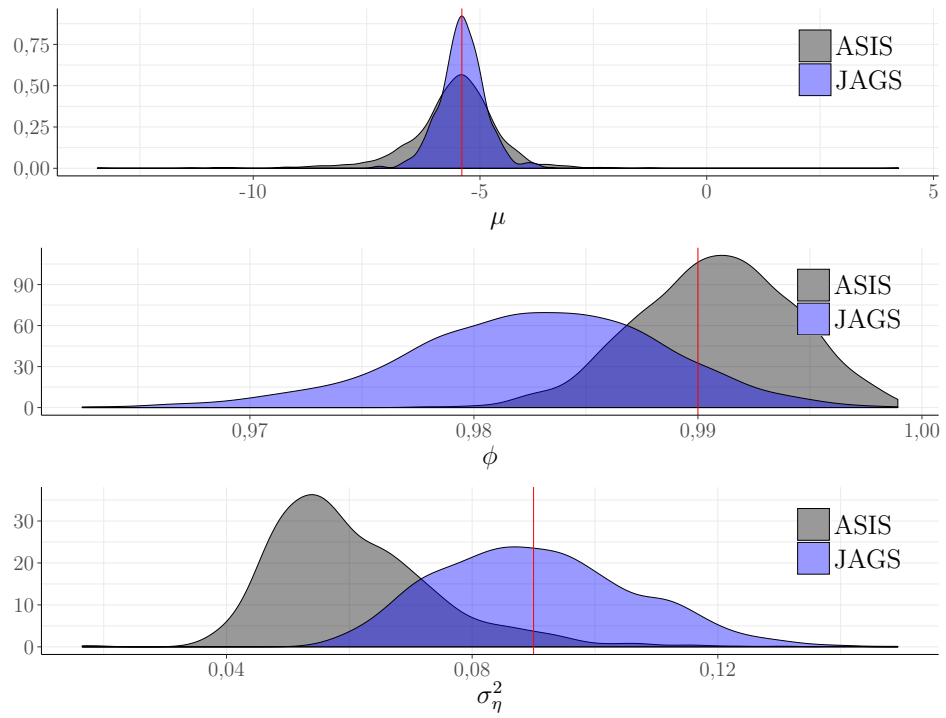


Figura 4.19: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,3^2$ ).

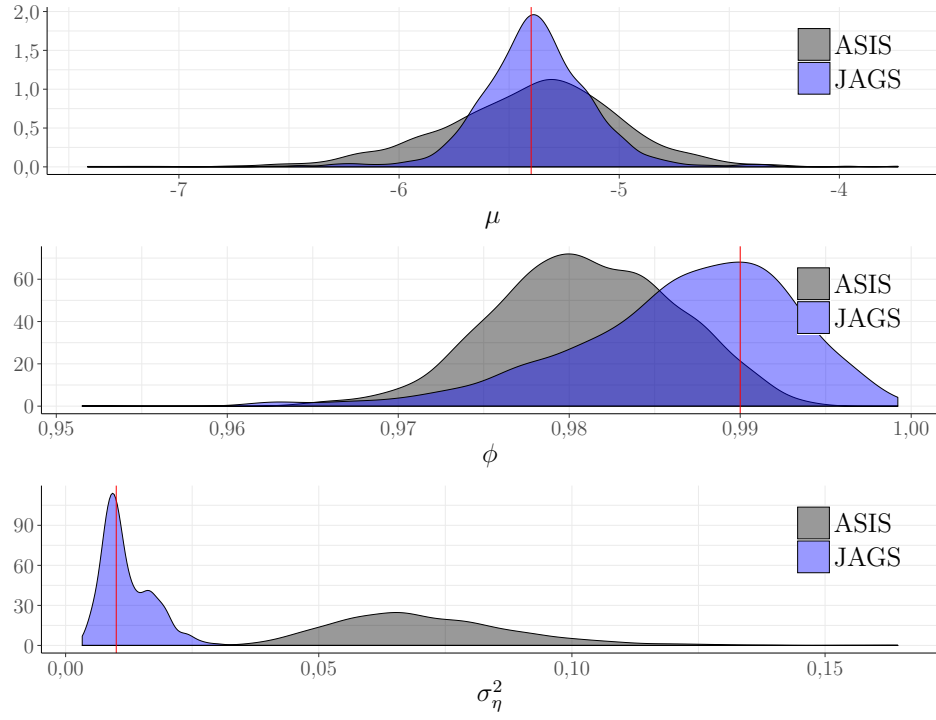


Figura 4.20: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,99$  e  $\sigma_\eta^2 = 0,1^2$ ).

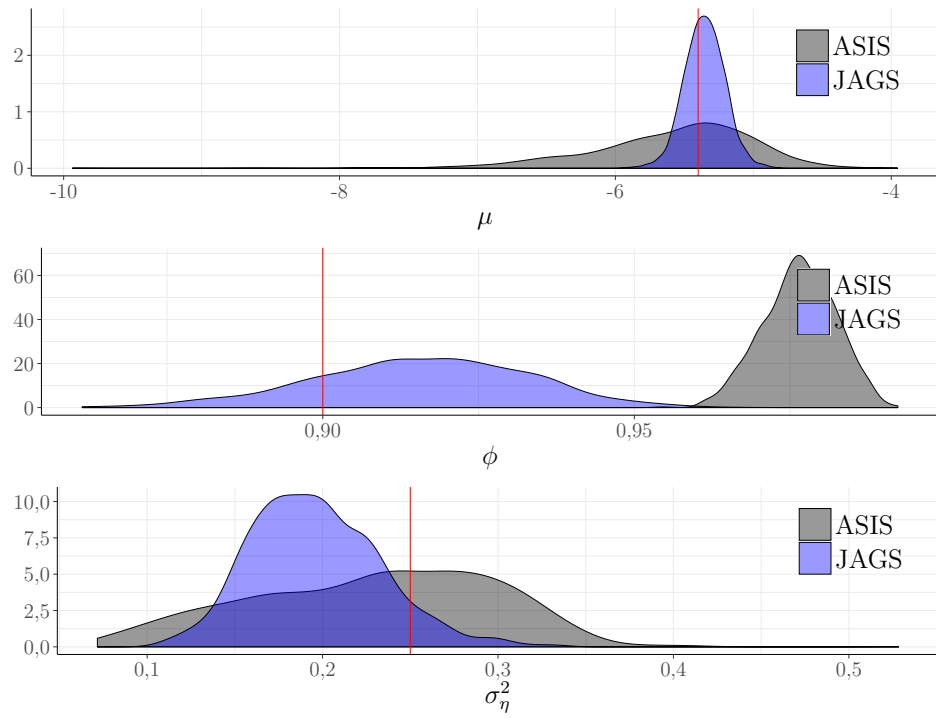


Figura 4.21: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,5^2$ ).

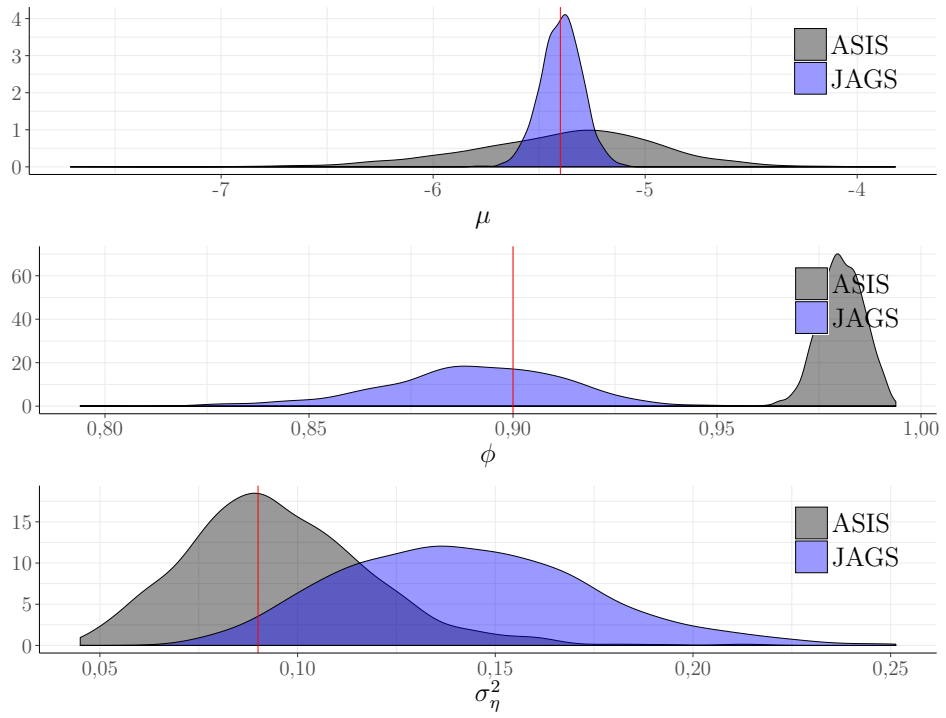


Figura 4.22: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,3^2$ ).

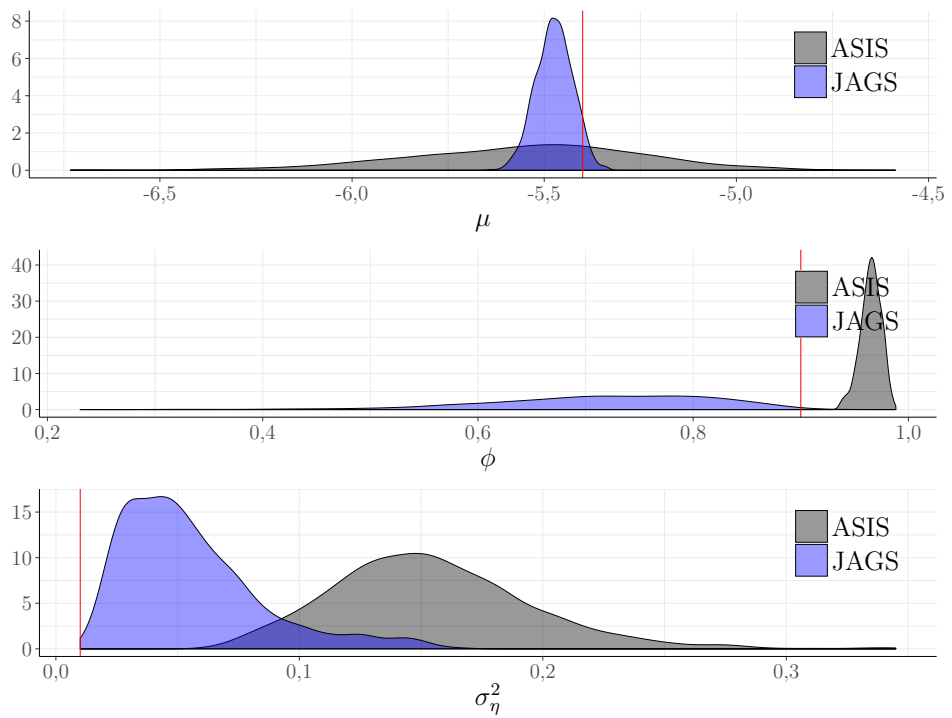


Figura 4.23: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,90$  e  $\sigma_\eta^2 = 0,1^2$ ).



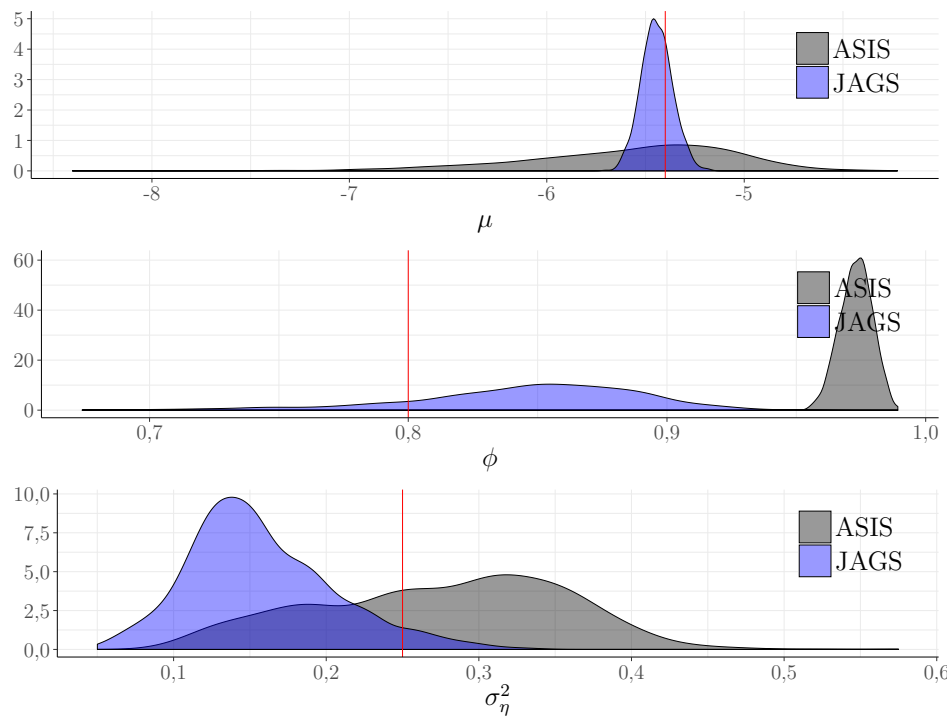


Figura 4.24: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,80$  e  $\sigma_\eta^2 = 0,5^2$ ).

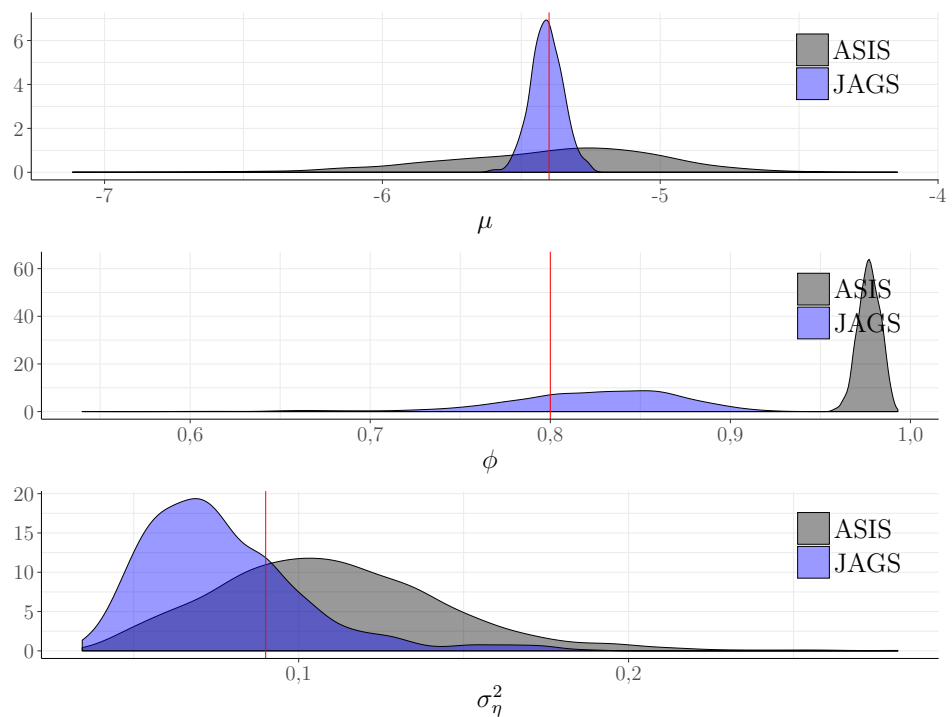


Figura 4.25: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,80$  e  $\sigma_\eta^2 = 0,3^2$ ).

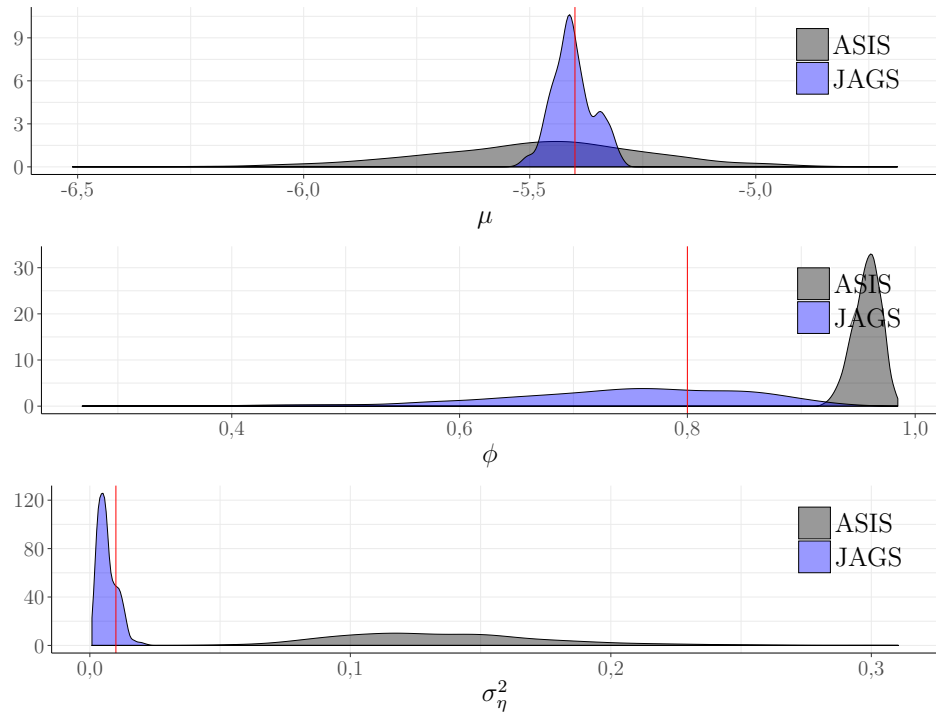


Figura 4.26: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,80$  e  $\sigma^2_\eta = 0,1^2$ ).

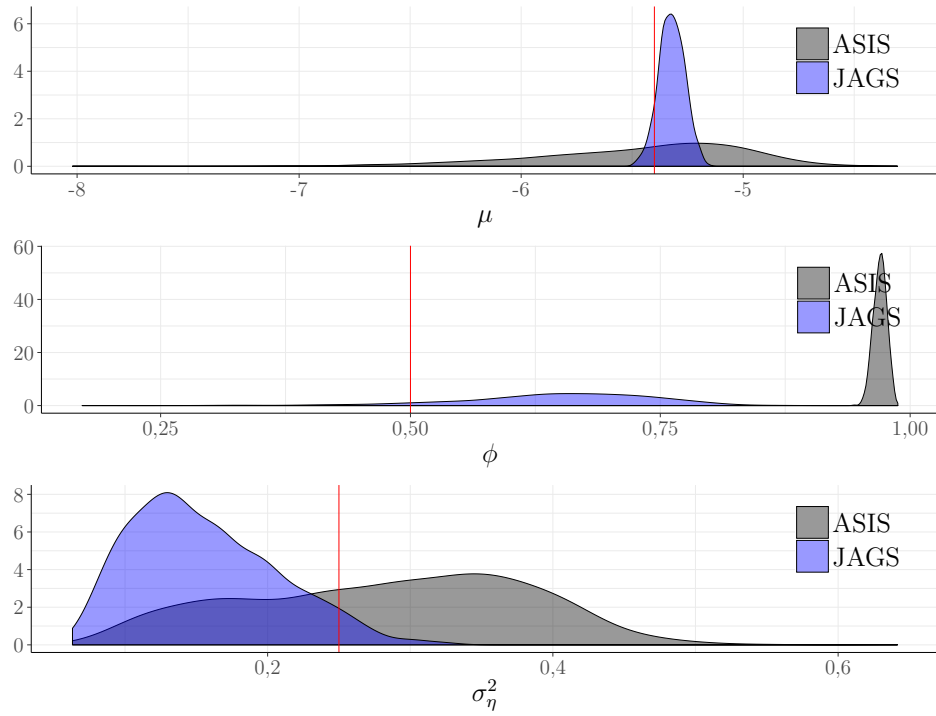


Figura 4.27: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,50$  e  $\sigma^2_\eta = 0,5^2$ ).

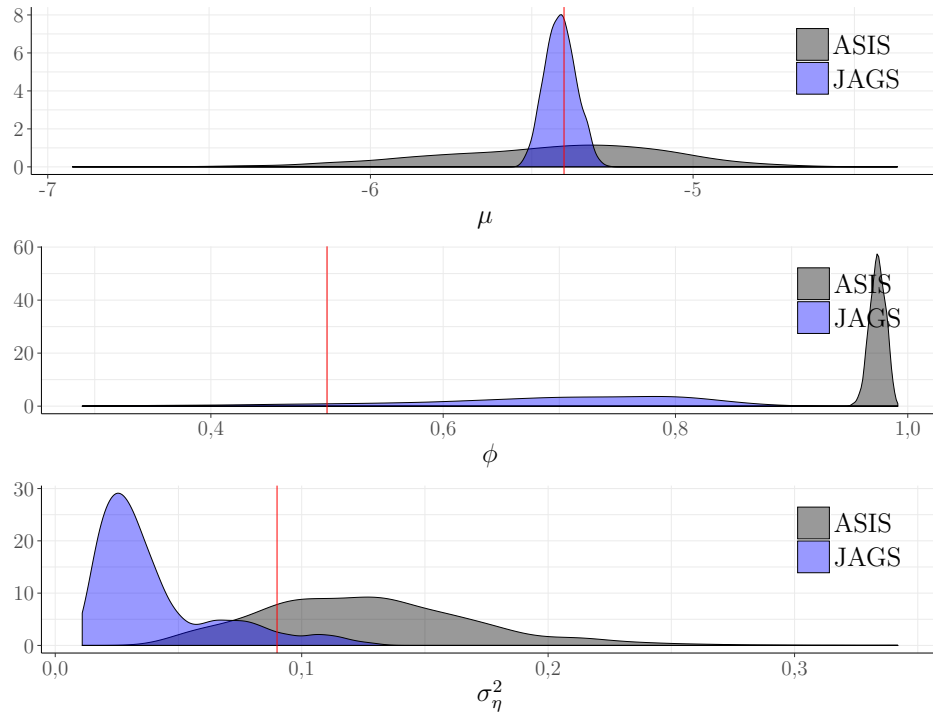


Figura 4.28: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,50$  e  $\sigma^2_\eta = 0,3^2$ ).

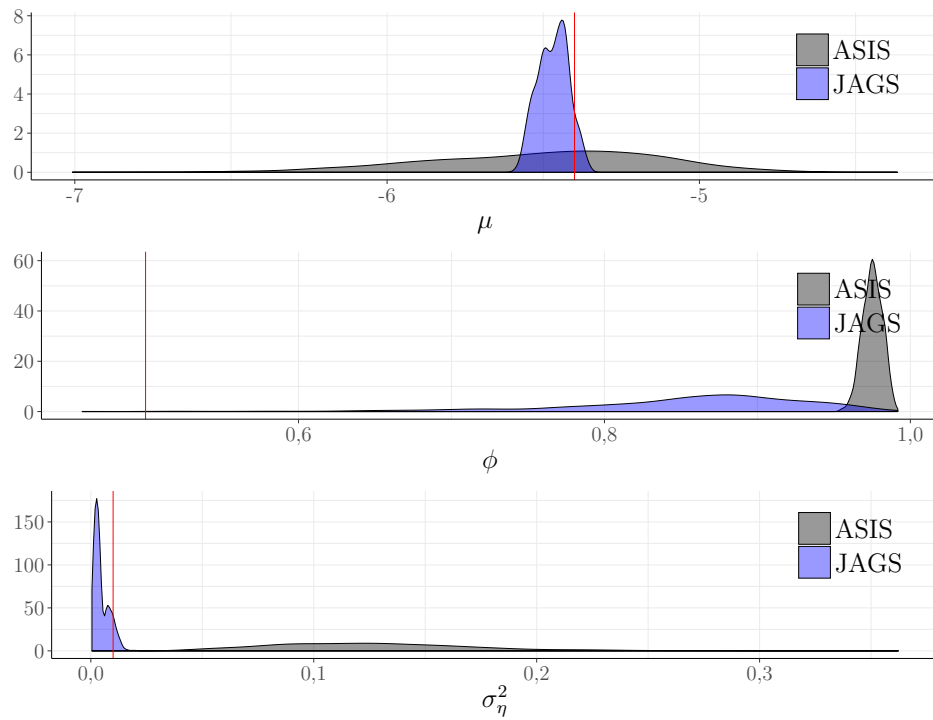


Figura 4.29: Valores estimados via JAGS vs método proposto ( $\mu = -5,4$ ,  $\phi = 0,5$  e  $\sigma^2_\eta = 0,1^2$ ).

## 4.4 Um Exemplo com Dados Reais

Achcar et al. (2011) propõe dois modelos de volatilidade estocástica levemente diferentes do modelo canônico em (2.3) para estudar a concentração de ozônio na Cidade do México. A base de dados utilizada no artigo corresponde a 16 anos (1º de janeiro 1990 a 31 de dezembro 2005) de medições da média semanal do nível diário máximo de ozônio em cinco regiões (NW, NE, CE, SE e SW) da Cidade do México. Os dados foram obtidos através do site <http://www.aire.df.gob.mx/>.

Os modelos propostos pelos autores são multivariados, pois levam em consideração a evolução das séries temporais da concentração de ozônio nas cinco regiões conjuntamente, sendo que, inclusive, um dos modelos apresentados propõe um parâmetro de volatilidade comum a todas as regiões, que adiciona-se à volatilidade individual de cada um dos setores. Por essa mesma razão não é prudente seguir com uma comparação direta dos resultados publicados em Achcar et al. (2011) com os resultados obtidos nesta dissertação em que nas análises considerou-se as séries temporais de forma univariada. Entretanto, o artigo citado foi muito importante para o entendimento inicial do modelo de volatilidade estocástica, e a base de dados referida foi utilizada nesta etapa como exemplo prático de aplicação da técnica.

O algoritmo foi executado nos dados da região CE (pois as séries das demais regiões continham dados perdidos) e os resultados pode ser vistos a seguir.

Como era esperado, o parâmetro de persistência  $\phi$ , no método proposto, ficou preso em torno dos valores acima de 0,9. E no caso do JAGS a estimativa ficou em torno de 0,8 (que é próximo da média *a priori* - 0,86), mas com uma dispersão bastante alta. Foram testadas outras distribuições *a priori* de  $\phi$  e foi constatado que os valores gerados deste parâmetro através do JAGS são totalmente determinados por sua distribuição *a priori*. As estimativas do parâmetro da variância da volatilidade  $\sigma_\eta^2$  foram bem próximas para os dois métodos, como também era esperado. Porém, não obteve-se concordância entre as estimativas de  $\mu$  via JAGS e via o método proposto. Muito provavelmente, diferente dos dados simulados, essa série corresponde a dados reais que foram gerados por um processo natural, absolutamente desconhecido.

## Região CE

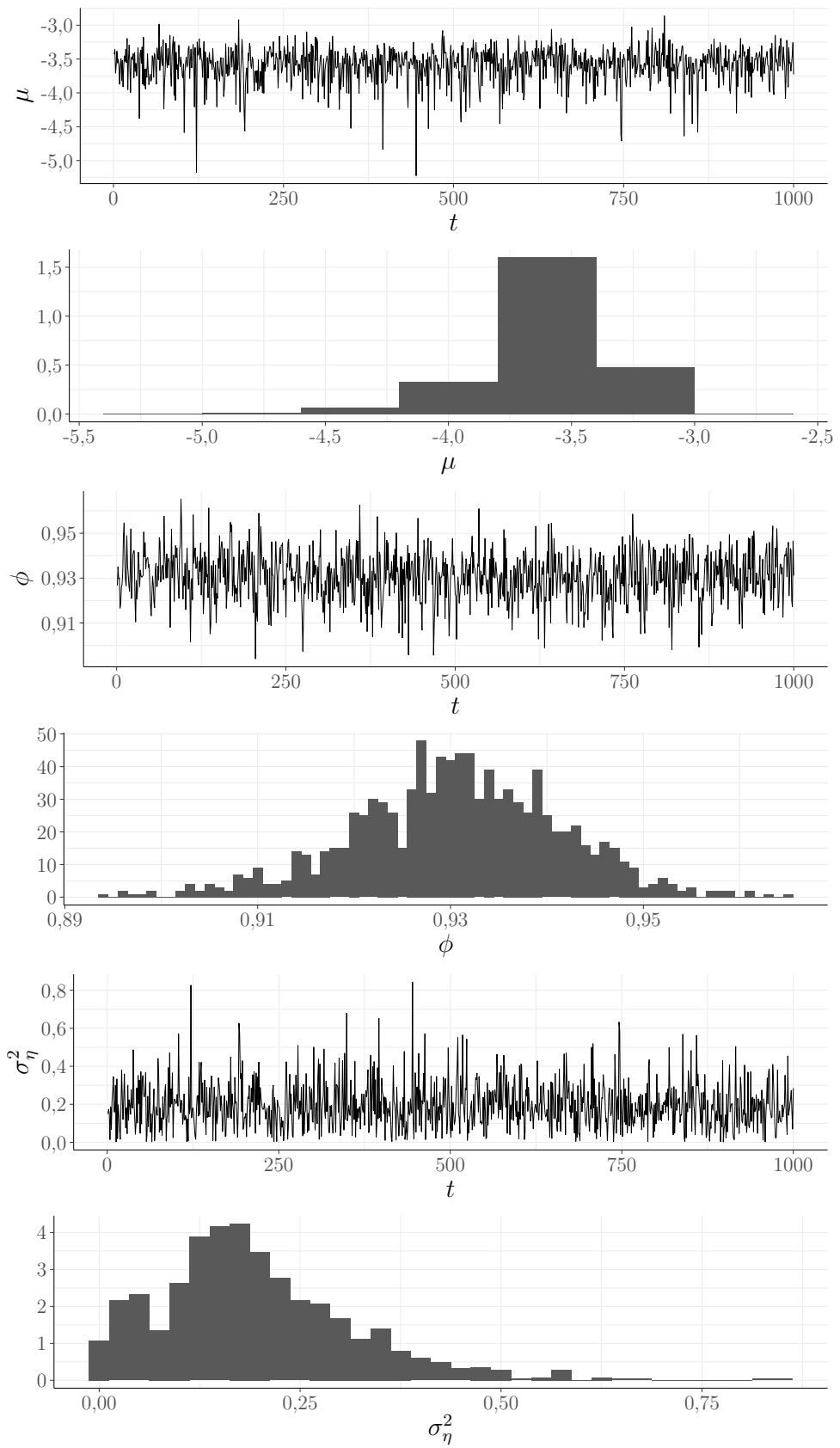


Figura 4.30: Valores estimados do MVE aplicado aos dados reais.

A Tabela 4.2 mostra algumas estatísticas dos valores estimados dos parâmetros e a Figuras 4.30 apresenta a comparação dos resultados obtidos também com o JAGS.

Tabela 4.2: Estatísticas dos valores estimados dos parâmetros do MVE aplicado aos dados reais.

Parâmetro	Média	2,5%	Mediana	97,5%
$\mu$	-3,594	-4,232	-3,557	-3,197
$\phi$	0,931	0,908	0,931	0,951
$\sigma_\eta^2$	0,194	0,011	0,178	0,478

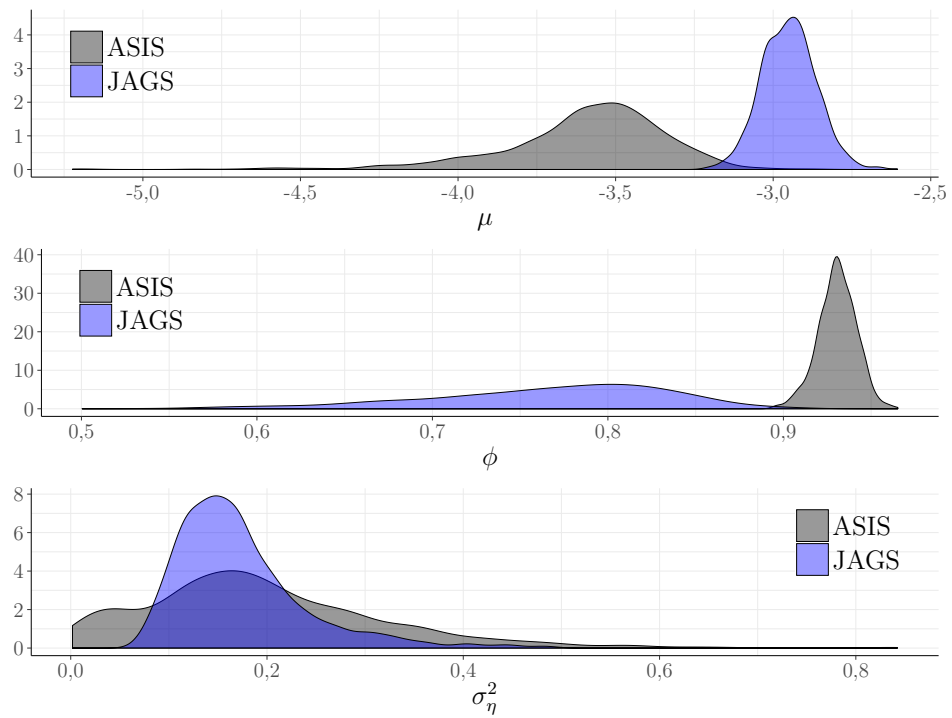


Figura 4.31: Valores estimados via JAGS vs método proposto aplicados aos dados reais.

# Conclusão

Este trabalho versa sobre o estudo de procedimentos Bayesianos de estimação para modelos de volatilidade estocástica (MVE). Nesta dissertação constatou-se as dificuldades e limitações de algumas metodologias propostas, recentemente, para os MVE (Kastner e Frühwirth-Schnatter, 2014) bem como as deficiências de alguns *softwares* existentes para lidar com o problema sob o paradigma Bayesiano, como o JAGS. Na tentativa de propor alternativas eficientes na estimação dos estados latentes do MVE, neste trabalho adaptou-se a metodologia descrita em McCormick et al. (2012). Posteriormente, combinou-se a metodologia desenvolvida neste trabalho com os procedimentos propostos por Kastner e Frühwirth-Schnatter (2014), para possibilitar a estimação de todos os parâmetros do MVE.

É bem verdade que a metodologia proposta, baseada na junção de Kastner e Frühwirth-Schnatter (2014) e McCormick et al. (2012), não resolveu o problema de forma definitiva. Porém, ela pode ser considerada um marco inicial para trabalhos futuros. É necessário avaliar, com detalhes, a forma como  $\phi$  é estimado, uma vez que, como apontado anteriormente a estimação desse parâmetro parece o problema mais desafiador de um MVE. É fundamental que se faça estudos mais aprofundados da associação entre o parâmetro de persistência  $\phi$  e o fator de desconto  $\lambda_t$ , além de investigar se a distribuição normal truncada (geradora de candidatos *a posteriori* de  $\phi$  no algoritmo de Metropolis-Hastings) é adequada nesse contexto.

Os demais parâmetros,  $\mu$  e  $\sigma_\eta^2$ , não apresentaram grandes problemas, especialmente  $\mu$ , pois como pôde ser observado, os seus valores estimados foram bastante satisfatórios nos estudos com os dados simulados.

Destaca-se, também, o desafio que é trabalhar com o modelo de volatilidade estocástica. Trata-se de um modelo computacionalmente caro, o que representa uma grande limitação no desenvolvimento de novas metodologias, e implementação computacional de novos algoritmos numa linguagem de alto nível, como o R, por exemplo.

Como citado anteriormente, há diversas propostas na literatura sobre a estimação dos parâmetros dos MVE, tanto do ponto de vista clássico quanto do Bayesiano. No entanto, até o momento, nenhuma destas propostas possibilitou a estimação satisfatória de todos os parâmetros do modelo (em especial do parâmetro de persistência  $\phi$ ), indicando que ainda há um campo vasto para pesquisas nesta área.

Conclui-se este trabalho com a expectativa de termos contribuído com algumas informações adicionais que podem ser úteis no desenvolvimento dos modelos de volatilidade estocástica.



# Referências Bibliográficas

- Achcar, J. A., Rodrigues, E. R., e Tzintzun, G. (2011). Using stochastic volatility models to analyse weekly ozone averages in Mexico City. *Environmental and Ecological Statistics*, 18(2):271–290.
- Basu, D. (1955). On statistics independent of a complete sufficient statistic. *The Indian Journal of Statistics*, 15(4):377–380.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.
- Bos, C. S. (2012). *Handbook of Volatility Models and their Applications*, chapter Relating Stochastic Volatility Estimation Methods, pages 147–174. John Wiley & Sons.
- Brockwell, P. J. e Davis, R. A. (2002). *Introduction to Time Series and Forecasting*. Springer Texts in Statistics. Springer, New York, 2nd edition.
- Carter, C. K. e Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81:541–553.
- Durbin, J. e Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation for non-gaussian state-space models. *Biometrika*, 89:603–615.
- Durbin, J. e Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, 2nd edition.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

- Gallant, A. R. e Tauchen, G. (1996). Which moments to match? *Econometric Theory*, 12:657–681.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., e Rubin, D. B. (2014). *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, 3rd edition.
- Harvey, A., Ruiz, E., e Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.
- Ihaka, R. e Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphics Statistics*, 5(3):299–314.
- Jaynes, E. T. (2003). *Probability Theory – The Logic of Science*. Cambridge University Press, Cambridge, 1st edition.
- Kastner, G. (2015). Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, (forthcoming):to appear.
- Kastner, G. e Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics and Data Analysis*, 76:408–423.
- Kernighan, B. W. e Ritchie, D. M. (1988). *The C Programming Language*. Prentice Hall, Upper Saddle River, 2nd edition.
- Kim, S., Shepard, N., e Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *The Review of Economic Studies*, 65(3):361–393.
- Lunn, D. J., Thomas, A., Best, N., e Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure and extensibility. *Statistics and Computing*, 10(4):325–337.

- McCormick, T. H., Raftery, A. E., Madigan, D., e Burd, R. S. (2012). Dynamic logistic regression and dynamic model averaging for binary classification. *Biometrics*, 68(1):23–30.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., e Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092.
- Nguyen, T. M. (2007). A new efficient algorithm for the analysis with non-linear and non-gaussian state space models. Master’s thesis, VU University Amsterdam.
- Omori, Y., Chib, S., Shephard, N., e Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140:425–449.
- Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, 2nd edition.
- Petris, G. (2010). An R package for dynamic linear models. *Journal of Statistical Software*, 36(12):1–16.
- Petris, G., Petrone, S., e Campagnoli, P. (2009). *Dynamic Linear Models with R*. Springer, New York, 1st edition.
- Plummer, M. (2003). JAGS: A program for analysis of bayesian graphical models using Gibbs sampling. In *Distributed Statistical Computing*, Vienna University of Technology.
- Plummer, M. (2015). *JAGS Version 4.0.0. user manual*.
- Pole, A., West, M., e Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. Chapman & Hall/CRC, Boca Raton, 1st edition.
- Prado, R. e West, M. (2010). *Time Series: Modeling, Computation, and Inference*. Chapman & Hall/CRC, Boca Raton, 1st edition.
- Richard, J. F. e Zhang, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics*, 141:1385–1411.

- Shephard, N. e Pitt, M. K. (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84:653–667.
- Sivia, D. e Skilling, J. (2006). *Data Analysis: A Bayesian Tutorial*. Oxford University Press, Oxford, 2nd edition.
- Su, Y. e Yajima, M. (2015). *R2jags: Using R to Run JAGS*. R package version 0.5-6.
- Taylor, S. (1982). Financial returns modelled by the product of two stochastic processes – a study of daily sugar prices 1961-75. *Time Series Analysis: Theory and Practice*, 1:203–226.
- Tierney, L. e Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.
- West, M. e Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, New York, 2nd edition.
- Yu, Y. e Meng, X. (2011). To center or not to center: That is not the question - an ancillarity-sufficiency interweaving strategy (asis) for boosting mcmc efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570.
- Zivot, E. e Yollin, G. (2012). Time series forecasting with state space models. In *R/Finance Workshop, University of Chicago*.

# Apêndice A

## Códigos-Fonte

Código A.1: Função Responsável por Gerar os Dados Simulados.

```
1 sv_data <- function(n, mu, phi, sigma) {  
2   h0 <- rnorm(1, mu, sigma / sqrt(1 - phi ** 2))  
3   h <- rnorm(n, mu + phi * (h0 - mu) , sigma)  
4   y <- rnorm(n, 0, exp(h[1] / 2))  
5   for (i in 2:n) {  
6     h[i] <- rnorm(1, mu + phi * (h[i -1] - mu), sigma)  
7     y[i] <- rnorm(1, 0, exp(h[i] / 2))  
8   }  
9   return(data.frame(y = y, h = h))  
10 }
```

Código A.2: Implementação da Proposta de Estimação dos Parâmetros do Modelo de Volatilidade Estocástica.

```
1 #include <stdio.h>  
2 #include <stdlib.h>  
3 #include <string.h>  
4 #include <math.h>  
5 #include <time.h>  
6 #include <gsl/gsl_randist.h>  
7  
8 #define p_A 0  
9 #define p_B 100  
10 #define p_C 20  
11 #define p_D 1.5  
12 #define p_E 1  
13  
14 double D1(double x, double *y, double *p, double m, double C,  
15           double lambda){  
16   double R, result;  
17   R = pow(*(p + 1), 2) * C / lambda;  
18   result = -0.5 + pow(*y, 2) / (2 * exp(x));  
19   result -= (x - (*p * (1 - *(p + 1)) + *(p + 1) * m)) / R;  
20   return result;  
21 }
```

```

20 };
21
22 double D2(double x, double *y, double *p, double C, double
    lambda){
23     double R, result;
24     R = pow(*(p + 1), 2) * C / lambda;
25     result = -pow(*y, 2) / (2 * exp(x)) - 1 / R;
26     return result;
27 };
28
29 double mu_media(double *h, double *p, double h0, int n){
30     double result;
31     result = *h - *(p + 1) * h0;
32     for(int i = 0; i < (n - 1); i++)
33         result += *(h + i + 1) - *(p + 1) * *(h + i);
34     return result;
35 }
36
37 double mu_cte(double B, int n){
38     double result;
39     result = n + 1 / B;
40     return result;
41 }
42
43 double mu_R(double x, double *p, double B, double h0){
44     double result;
45     result = exp(-pow(h0 - x / (1 - *(p + 1)), 2) * (1 - pow(*(p
        + 1), 2)) / (2 * *(p + 2)));
46     result *= exp(-pow(x - p_A * (1 - *(p + 1)), 2) / (2 * p_B *
        pow(1 - *(p + 1), 2)));
47     return result;
48 }
49
50 double phi_media(double *h, double *p, double h0, int n){
51     double result;
52     result = -(1 - *(p + 1)) * *p * h0;
53     for(int i = 0; i < (n - 1); i++)
54         result -= (1 - *(p + 1)) * *p * *(h + i);
55     result += h0 * *h;
56     for(int i = 0; i < (n - 1); i++)
57         result += *(h + i + 1) * *(h + i);
58     return result;
59 }
60
61 double phi_cte(double *h, double B, double h0, int n){
62     double result;
63     result = pow(h0, 2) + 1 / B;
64     for(int i = 0; i < (n - 1); i++)

```

```

65     result += pow(*(h + i), 2);
66     return result;
67 }
68
69 double phi_R(double x, double *p, double B, double h0){
70     double result;
71     result = sqrt(1 - pow(x, 2)) * pow((1 + x) / 2, p_C - 1) *
72         pow((1 - x) / 2, p_D - 1);
73     result *= exp(-(1 - pow(x, 2)) * pow(h0 - *p, 2) / (2 * *(p +
74         2)));
75     result /= exp(-pow(x, 2) / (2 * *(p + 2) * B));
76     return result;
77 }
78
79 double CT_sigma2(double *h, double *p, int n, double h0){
80     double result;
81     result = pow(h0 - *p, 2) * (1 - pow(*(p + 1), 2));
82     result += pow(*h - *p - *(p + 1) * (h0 - *p), 2);
83     for(int i = 1; i < n; i++)
84         result += pow(*(h + i) - *p - *(p + 1) * (*(h + i - 1) -
85             *p), 2);
86     return result;
87 }
88
89 double BT_sigma2(double *h, int n, double sr2){
90     double result;
91     result = 1 / p_E;
92     for(int i = 0; i < n; i++)
93         result += pow(*(h + i), 2) / sr2;
94     result = 1 / result;
95     return result;
96 }
97
98 double bT_sigma2(double *y, double *h, double *p, int n, double
99     mr, double sr2){
100     double result;
101     result = 0;
102     for(int i = 0; i < n; i++)
103         result += *(h + i) * (log(pow(*(y + i), 2)) - mr - *p) /
104             sr2;
105     return result;
106 }
107
108 double BT_mu(int n, double sr2){
109     double result;
110     result = 1 / p_B;
111     for(int i = 0; i < n; i++)
112         result += 1 / sr2;

```

```

108     result = 1 / result;
109     return result;
110 }
111
112 double bT_mu(double *y, double *h, double *p, int n, double mr,
113             double sr2){
114     double result;
115     result = p_A / p_B;
116     for(int i = 0; i < n; i++){
117         result += (log(pow(*(y + i), 2)) - mr - sqrt(*(p + 2)) *
118                 *(h + i)) / sr2;
119     }
120     return result;
121 }
122
123 int main(void){
124     unsigned long int t = time(NULL);
125     int n = 1461;
126     double *y;
127     y = malloc(n * sizeof (double));
128
129     // setup mcmc
130     int burnin = 5000, mc = 10000, thin = 10;
131     double h0, *h, p[3], m, C, lambda, x1, x2, candidate, s[3];
132     float a[3];
133     h = malloc(n * sizeof (double));
134     a[0] = a[1] = a[2] = 0;
135     s[0] = s[1] = s[2] = 0;
136
137     // valores iniciais dos parâmetros
138     p[0] = 0.0;
139     p[1] = 0.8;
140     p[2] = 0.09;
141
142     // erro do modelo
143     double mr = -1.2702800088, sr2 = 4.9337311973;
144
145     const gsl_rng_type *T;
146     gsl_rng *engine;
147     gsl_rng_env_setup();
148     T = gsl_rng_default;
149     engine = gsl_rng_alloc(T);
150
151     // mcmc (McCormick + ASIS)
152     scanf("%lf", &lambda);
153     for(int i = 0; i < burnin + mc; i++){
154         printf("%06d: %9.5lf %7.5lf %.5lf\n", i + 1, p[0], p[1],
155                p[2]);
156     }

```



```

153 // estimar h (McCormick)
154 m = 0.0;
155 C = p[2] / (1 - pow(p[1], 2));
156
157 for(int j = 0; j < n; j++){
158     x1 = D1(m, y + j, p, m, C, lambda);
159     x2 = D2(m, y + j, p, C, lambda);
160     m -= x1 / x2;
161     C = -1 / x2;
162     *(h + j) = m;
163 }
164 h0 = p[0] + gsl_ran_gaussian(engine, sqrt(*(p + 2) / (1 -
    pow(*(p + 1), 2))));
165
166 // (ASIS)
167 // C
168 // estimar sigma2 (MH)
169 x1 = (double) n;
170 x2 = CT_sigma2(h, p, n, h0);
171 candidate = 1 / gsl_ran_gamma(engine, x1 / 2, 2 / x2);
172 if(gsl_ran_flat(engine, 0, 1) < exp((p[2] - candidate) / (2
    * p_E))){
173     p[2] = candidate;
174     a[0]++;
175 }
176
177 // estimar phi (MH)
178 x1 = phi_media(h, p, h0, n);
179 x2 = phi_cte(h, 10e8, h0, n);
180 do{
181     candidate = x1 / x2 + gsl_ran_gaussian(engine, sqrt(p[2] /
        x2));
182 } while(candidate < -1 || candidate > 1);
183 x1 = phi_R(candidate, p, 10e8, h0);
184 x2 = phi_R(p[1], p, 10e8, h0);
185 if(gsl_ran_flat(engine, 0, 1) < x1 / x2){
186     p[1] = candidate;
187     a[1]++;
188 }
189
190 // estimar mu (MH)
191 x1 = mu_media(h, p, h0, n);
192 x2 = mu_cte(10e12, n);
193 candidate = x1 / x2 + gsl_ran_gaussian(engine, sqrt(p[2] /
    x2));
194 x1 = mu_R(candidate, p, 10e12, h0);
195 x2 = mu_R(p[0], p, 10e12, h0);
196 if(gsl_ran_flat(engine, 0, 1) < x1 / x2){

```

```

197     p[0] = candidate / (1 - p[1]);
198     a[2]++;
199 }
200
201 // NC
202 // transformar h
203 h0 = (h0 - *p) / sqrt(*(p + 2));
204 for(int i = 0; i < n; i++)
205     *(h + i) = (*(h + i) - *p) / sqrt(*(p + 2));
206
207 // re-estimar sigma2
208 x2 = BT_sigma2(h, n, sr2);
209 x1 = x2 * bT_sigma2(y, h, p, n, mr, sr2);
210 candidate = x1 + gsl_ran_gaussian(engine, sqrt(x2));
211 p[2] = pow(candidate, 2);
212
213 // re-estimar mu
214 x2 = BT_mu(n, sr2);
215 x1 = x2 * bT_mu(y, h, p, n, mr, sr2);
216 candidate = x1 + gsl_ran_gaussian(engine, sqrt(x2));
217 p[0] = candidate;
218 if(i > (burnin - 1) && (i + 1) % thin == 0){
219     s[0] += p[0];
220     s[1] += p[1];
221     s[2] += p[2];
222 }
223 }
224
225 gsl_rng_free(engine);
226 free(y);
227 free(h);
228 printf("\nmu: %11.5lf (%.2f%%)\nphi: %10.5lf
    (%.2f%%)\nsigma2: %.5lf (%.2f%%)\n", s[0] / (mc / thin),
    a[2] * 100 / (burnin + mc), s[1] / (mc / thin), a[1] * 100
    / (burnin + mc), s[2] / (mc / thin), a[0] * 100 / (burnin
    + mc));
229 t = time(NULL) - t;
230 printf("\nTempo de execucao: %ldmin %lds (%0.2f
    Rodadas/s)\n", t / 60, t % 60, (float) (burnin + mc) / t);
231
232 return 0;
233 }

```

Código A.3: Modelo Canônico de Volatilidade Estocástica em JAGS definido em (2.3).

```
1 model {
2   y[1] ~ dnorm(0, exp(-h[1]))
3   h[1] ~ dnorm(mu + phi * (h0 - mu), 1 / sigma2)
4   for (t in 2:N) {
5     y[t] ~ dnorm(0, exp(-h[t]))
6     h[t] ~ dnorm(mu + phi * (h[t - 1] - mu), 1 / sigma2)
7   }
8   h0 ~ dnorm(mu, (1 - phi^2) / sigma2)
9   mu ~ dnorm(0, 1 / 100)
10  phi0 ~ dbeta(20.0, 1.5)
11  phi <- 2 * phi0 - 1
12  sigma2 ~ dgamma(0.5, 0.5)
13 }
```