

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA
Departamento de Estatística

Modelos de Espaço de Estado Não-Gaussianos e o Modelo de Volatilidade Estocástica

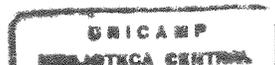
Anderson Carlos Oliveira Motta

Prof. Dr. Luiz Koodi Hotta
Orientador

Dissertação apresentada junto ao Departamento de Estatística do Instituto de Matemática Estatística e Computação Científica da Universidade Estadual de Campinas, para obtenção do título de Mestre em Estatística.

Campinas
2001

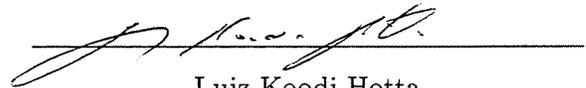
0115797



Modelos de Espaço de Estado Não-Gaussianos e o Modelo de Volatilidade Estocástica

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Anderson Carlos Oliveira Motta e aprovada pela comissão julgadora.

Campinas, 01 de Março de 2001



Luiz Koodi Hotta

Orientador

Banca Examinadora:

1. Prof. Dr. Luiz Koodi Hotta – Orientador — IMECC/UNICAMP
2. Prof. Dr. Pedro Luiz Valls Pereira – IBMEC/SP
3. Prof. Dr. Ronaldo Dias – IMECC/UNICAMP

Dissertação apresentada ao Instituto de Matemática Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do título de MESTRE em ESTATÍSTICA.

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP

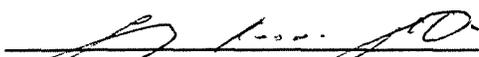
Motta, Anderson Carlos Oliveira
M858M Modelos de espaço de estado não-gaussianos e o modelo de
volatilidade estocástica /Anderson Carlos Oliveira Motta. – Campinas,
[S.P. :s.n.], 2001.

Orientador: Luiz Koodi Hotta.
Dissertação (mestrado) – Universidade Estadual de Campinas, Insti-
tuto de Matemática, Estatística e Computação Científica.

1. Derivativos(Finanças) 2. Análise de séries temporais I. Hotta,
Luiz Koodi. II. Universidade Estadual de Campinas, Instituto de Ma-
temática, Estatística e Computação Científica. III. Título.

Dissertação de Mestrado defendida em 01 de março de 2001 e aprovada

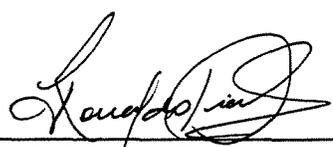
Pela Banca Examinadora composta pelos Profs. Drs.



Prof (a). Dr (a). LUIZ KOODI HOTTA



Prof (a). Dr (a). PEDRO LUIZ VALLS PEREIRA



Prof (a). Dr (a). RONALDO DIAS

Resumo

O objetivo deste trabalho é apresentar alguns métodos de estimação de modelos que podem ser vistos como um modelo dinâmico e na forma de espaço de estados. São consideradas as abordagens clássica e bayesiana, e aplicado ao modelo de volatilidade estocástica.

Os métodos foram aplicados à algumas séries financeiras. Foram consideradas séries simuladas para verificar o comportamento dos diversos métodos na presença de outliers . Na aplicação das metodologias ao mercado financeiro brasileiro foi utilizada a série de observações diárias do Índice da Bolsa de Valores de São Paulo (IBOVESPA), no período de 02/Janeiro/1995 a 27/Dezembro/2000 (1500 observações).

Abstract

The aim of this work is to present some methods of estimation of models that may be seen either as a dynamic model or a state space model and to apply them to some stochastic volatility models, considering classical and bayesian approaches.

Those methods were applied to the daily São Paulo Stock Exchange Index (IBOVES-PA) series and to simulated series for verify their behaviors in presence of outliers. The IBOVESPA series was collected between january 2nd, 1995 and december 27th, 2000, totaling 1,500 observations.

Agradecimentos

Agradeço a FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) pelo financiamento deste projeto durante o período de 24 meses em que foi realizado.

Agradeço ao meu orientador, Prof. Dr. Luiz Koodi Hotta, pela paciência, amizade e exemplo de dedicação.

Agradeço aos amigos da E2A, Anderson, Bovo, Luiz Carlos e Vitor pela força direta e indireta e pela amizade desses seis anos.

Agradeço aos meus amigos de mestrado Edimilson, Mariana, Fabiana, Marcelo Malucão "o(!)", Colugnati e a todos da pós 95-98 pela amizade e apoio durante o curso. E também àqueles que apesar de não estarem aqui, sei que torcem pelo meu trabalho, Valdecir e Jaqueline. Ao amigo Fukui por me ensinar (ainda na graduação) o que era espaço de estado.

Agradeço ao meu irmão, pai, primo e amigo Cláudio Fabiano Motta Toledo; sem você nada teria acontecido. A minha irmã Valdinéia e meu amigo Júlio César (*The law's man*), pelo carinho e também pela comida rápida (batata com sardinha) e por me acordar 5:00am. quando eu estudava para o vestibular. A minha irmã Adriana pelo carinho e exemplos de ousadia.

Agradeço a minha avó Esther pelo carinho, amor e dedicação, minha tia Shirley pela dedicação e carinho e a minha mãe.

Agradeço a Ana Paula Soares Lopez e a Marianinha pela compreensão e pelo incentivo incondicional, principalmente no final deste trabalho onde, mesmo por telefone, conseguiram me dar forças para continuar.

Campinas
01 de Março de 2001

Anderson Motta

SUMÁRIO

1	Introdução	1
2	Modelos Dinâmicos e Forma de Espaço de Estado	5
2.1	Introdução	6
2.2	Modelos Dinâmicos	6
2.2.1	Modelos Lineares Dinâmicos	6
2.2.2	Modelos Dinâmicos Generalizados	11
2.3	Forma de Espaço de Estado	12
2.4	Forma de Espaço de Estado Linear e Gaussiana	14
2.4.1	Filtro de Kalman	15
2.4.2	Filtro de Kalman Estendido	17
2.4.3	Estimativa Suavizada Através do Filtro de Kalman	19
2.4.4	Suavizador de de Jong e Shephard	20
2.5	Forma de Espaço de Estado Não-Gaussiano	21
2.5.1	Formas de Representação	21
2.5.2	Estimação, Filtragem e Suavizamento	23
2.5.3	Filtro de Partículas	25
2.6	Conclusões	27
3	Modelo de Volatilidade Estocástica	29
3.1	Retornos Diários das Séries Financeiras	30
3.2	Séries Financeiras	31
3.2.1	Modelo Básico para Retornos de Séries Financeiras	33
3.2.2	Volatilidade Estocástica	34
3.3	O Modelo de Volatilidade Estocástica	34

3.3.1	Propriedades do Modelo de Volatilidade Estocástica	37
3.3.2	Formulação na Forma de Espaço de Estado	41
3.4	Conclusões	43
4	Análise Clássica do Modelo de Volatilidade Estocástica	45
4.1	Introdução	46
4.2	Correção da quasi-Verossimilhança Através da Amostragem dos Estados	47
4.2.1	A Seleção do Modelo Aproximado	50
4.2.2	Estimação da Logverossimilhança e Correção de Vício	52
4.2.3	Algoritmo de Estimação da Verossimilhança	57
4.3	Correção da Quasi-Verossimilhança Através da Amostragem das Per- turbações	57
4.4	Modelo Aproximado para o Modelo de Volatilidade Estocástica	58
5	Análise Bayesiana do Modelo de Volatilidade Estocástica	61
5.1	Introdução	62
5.2	Amostragem da Variável Latente	63
5.2.1	Amostragem Simples da Variável Latente	65
5.2.2	Amostragem da Variável Latente por Blocos	73
5.2.3	Amostragem Conjunta da Variável Latente	78
5.3	Amostragem dos Coeficientes do Modelo	83
5.4	Conclusões	87
6	Aplicação dos Métodos	89
6.1	Introdução	89
6.2	Análise das Séries Simuladas	90
6.2.1	Cálculo da Verossimilhança Corrigida	94
6.2.2	Estudo da Convergência na Análise Bayesiana	95
6.2.3	Estimativas dos Parâmetros	99
6.2.4	Estimativas da Volatilidade	104
6.3	Análise da Série IBOVESPA	115
6.4	Conclusão	120
A	Métodos de Simulação	123
A.1	Método de Aceitação-Rejeição	124
A.2	MCMC	125
A.2.1	O Algoritmo de Metropolis-Hastings	125
A.2.2	O Amostrador de Gibbs	128
A.2.3	Amostragem por Rejeição Adaptativa com Metropolis Hastings Amostrador de Gibbs	128
A.3	Integração por Monte Carlo	130

A.3.1	Amostragem por Importância	131
A.3.2	Variáveis Antitéticas	131
A.3.3	Variáveis Controle	132
B	Amostragem das Densidades Condicionais	133
B.1	Introdução	133
B.1.1	Aproximação Monte Carlo: Tipo 1 (MA1)	134
B.1.2	Aproximação Monte Carlo: Tipo 2 (MA2)	139
B.1.3	Suavizamento Intervalo-Fixo	142
B.1.4	Amostragem por Importância para Filtro Não Linear	145
C	Resultados em Inferência Bayesiana	151
C.1	Distribuições Conjugadas	151
C.1.1	Famílias Conjugadas para Famílias Exponenciais	152
	Referências	153
	ÍNDICE	163

LISTA DE TABELAS

5.1	Seleção das misturas de normais para aproximar a distribuição $\log \chi_1^2$	80
6.1	Valores para o conjunto de parâmetros para o processo gerador das séries, com Δ = tamanho do AO e Λ = tamanho do IO	90
6.2	Estatísticas descritivas das Séries Simuladas para o Caso I	91
6.3	Estatísticas descritivas das Séries Simuladas para o Caso IV	91
6.4	Valores das Log-verossimilhanças corrigidas para os Casos I e II	94
6.5	Valores das Log-verossimilhanças corrigidas para os Casos III e IV	94
6.6	Estimativas dos parâmetros para as séries simuladas e abaixo seus respectivos erros padrão Caso I.	100
6.7	Estimativas dos parâmetros para as séries simuladas e seus respectivos erros padrão Caso II.	101
6.8	Estimativas dos parâmetros para as séries simuladas e seus respectivos erros padrão Caso III.	102
6.9	Estimativas dos parâmetros para as séries simuladas e seus respectivos erros padrão Caso IV.	104
6.10	Estatísticas descritivas dos retornos do IBOVESPA	115
6.11	Estimativas dos parâmetros para a série IBOVESPA e seus respectivos erros padrão	116

LISTA DE FIGURAS

5.1	Diagrama de Recursão para Estimar $p(\mathbf{h} \mathbf{y})$	64
5.2	Diagrama do Algoritmo de Amostragem dos Estados	70
6.1	Gráficos das séries simuladas com o conjunto de parâmetro para modelo para Caso I.	92
6.2	Gráficos das séries simuladas com o conjunto de parâmetro para modelo para Caso IV.	93
6.3	Gráficos dos valores dos parâmetros, histogramas das densidades posteriores e correlograma para amostragem simples com a presença de <i>outliers</i> do tipo AO e IO. Tempo para estimativa dos parâmetros em um Pentium III 550 foi de 399.42 segundos	96
6.4	Gráficos dos valores dos parâmetros, histogramas das densidades posteriores e correlograma para amostragem utilizando o algoritmo de mistura com a presença de <i>outliers</i> do tipo AO e IO. Tempo para estimativa dos parâmetros em um Pentium III 550 foi de 2176.4 segundos	97
6.5	Gráficos dos valores dos parâmetros, histogramas das densidades posteriores e correlograma para amostragem utilizando o algoritmo de integração com a presença de <i>outliers</i> do tipo AO e IO. Tempo para estimativa dos parâmetros em um Pentium III 550 foi de 38.840 segundos	97
6.6	Gráfico do valor absoluto dos retornos e da estimativa suavizada da sem outliers para o caso I utilizando a metodologia de Durbin e Koopman.	105
6.7	Gráfico do valor absoluto dos retornos e da volatilidade verdadeira e estimativa, filtrada e suavizada sem outliers para o caso I utilizando a metodologia de Kim <i>et al.</i>	106

6.8	Gráfico do valor absoluto dos retornos, da volatilidade verdadeira e suavizada considerando AO na posição $t = 250$ e IO em $t = 600$ para o caso I utilizando o método de Durbin e Koopman.	107
6.9	Gráfico dos retornos e volatilidades verdadeira e estimativa suavizada próximo ao AO em $t = 250$ para o caso I utilizando o método de Durbin e Koopman.	107
6.10	Gráfico dos retornos e volatilidades verdadeira e estimativa suavizada próximo ao IO em $t = 600$ para o caso I utilizando o método de Durbin e Koopman.	108
6.11	Volatilidade verdadeira e estimativas, filtrada e suavizada com <i>outliers</i> do tipo AO e IO utilizando o método de Kim <i>et al.</i>	109
6.12	Gráfico dos retornos e volatilidades verdadeira e estimativas, filtrada e suavizada próximo ao AO em $t = 250$ utilizando o método de Kim <i>et al.</i>	109
6.13	Gráfico dos retornos e volatilidades verdadeira e estimativas, filtrada e suavizada próximo ao IO em $t = 600$ para o caso I utilizando o método de Kim <i>et al.</i>	110
6.14	Volatilidade verdadeira e estimativa suavizada com <i>outliers</i> do tipo AO e IO para o caso IV utilizando método proposto por Durbin e Koopman.	111
6.15	Gráfico dos retornos e volatilidades verdadeira e estimativa suavizada próximo ao AO em $t = 250$ para o caso IV utilizando método proposto por Durbin e Koopman.	111
6.16	Gráfico dos retornos e volatilidades verdadeira e estimativa suavizada próximo ao IO em $t = 600$ para o caso IV utilizando método proposto por Durbin e Koopman.	112
6.17	Volatilidade verdadeira e estimativas, filtrada e suavizada com <i>outliers</i> do tipo AO e IO para o caso IV utilizando método proposto por Kim <i>et al.</i>	113
6.18	Gráfico dos retornos e volatilidades verdadeira e estimativas, filtrada e suavizada próximo ao AO em $t = 250$ para o caso IV utilizando método proposto por Kim <i>et al.</i>	113
6.19	Gráfico dos retornos e volatilidades verdadeira e estimativas, filtrada e suavizada próximo ao IO em $t = 600$ para o caso IV utilizando método proposto por Kim <i>et al.</i>	114
6.20	Gráfico dos retornos e histograma do IBOVESPA	116
6.21	Gráfico do valor absoluto dos retornos e da estimativa Suavizada pelo método de Durbin e Koopman.	117
6.22	Gráfico dos retornos e das estimativas filtradas e suavizadas da volatilidade sobreposto ao do valor absoluto dos retornos utilizando o método de Kim <i>et al.</i>	117

6.23	Gráficos resultante das simulações usando amostragem simples. Na primeira linha as iterações, na segunda os histogramas das distribuições a posterioris e por último o correlograma da amostra.	118
6.24	Gráficos resultante das simulações usando algoritmo de mistura. Na primeira linha as iterações, na segunda os histogramas das distribuições a posterioris e por último o correlograma da amostra.	119
6.25	Gráficos resultante das simulações usando algoritmo de integração. Na primeira linha as iterações, na segunda os histogramas das distribuições a posterioris e por último o correlograma da amostra.	119

CAPÍTULO 1

Introdução

Nesta tese, serão consideradas as abordagens clássicas e bayesianas a modelos de séries temporais, que podem ser escritos na forma de espaço de estado (**FEE**), mais especificamente o modelo de volatilidade estocástica (**VE**). Muitos modelos, incluindo os de séries temporais, podem ser representados na forma de espaço de estado. Este modelo, formulado por Kalman (1960) para problemas de engenharia, permite considerarmos os vetores de estados não observáveis evoluindo de acordo com um processo linear Markoviano.

Sob suposição de que as perturbações são Gaussianas temos o modelo na forma de espaço de estado Gaussiano (**FEEG**). Kalman (1960) propôs um algoritmo recursivo, conhecido como filtro de Kalman, que nos fornece estimativa da densidade de filtragem e suavização para os modelos na **FEEG**. Neste caso, é conhecido que este filtro, através do princípio da decomposição do erro de predição, nos produz a log-verossimilhança permitindo assim a análise clássica (ver Harvey (1989), pág.125). As integrais consideradas no passo de filtragem podem ser tratadas analiticamente e métodos eficientes para obter a densidade suavizada, a qual nos fornece a densidade

de todos estados condicionada a todas as observações, são encontrados na literatura, (ver, por exemplo, de Jong, 1989).

Mais recentemente o interesse está na ampliação das técnicas que tratam os modelos na **FEEG** que não possuem perturbações com distribuições Gaussianas. Dentre muitos autores, na área mais específica de Estatística, vale ressaltar West, Harrison e Migon (1985) que colaboraram através da teoria de modelos dinâmicos generalizados, considerando a equação de observação com erros pertencentes a família exponencial e Harvey e Fernandes (1989) que consideraram dados de contagem. Apesar da relevância dos trabalhos realizados, o enfoque será dado as técnicas relacionadas mais diretamente ao modelo de **VE**.

Devido ao avanço computacional, a representação na **FEE** tem recebido a atenção de muitos autores. O modelo de volatilidade estocástica é um dos modelos que após a linearização pode ser escrito nessa forma. Este modelo, proposto por Taylor (1986), é uma alternativa aos modelos da família ARCH proposto por Engle (1982) e recentemente está sendo amplamente explorado em finanças (Hull e White, 1987). A facilidade de se trabalhar analiticamente com a verossimilhança nos modelos ARCH é ainda um dos grandes motivos da preferência em trabalhar com esta família de modelos.

Os modelos de **VE**, quando representados na forma de espaço de estados, possuem os erros da equação de observação com distribuição log-qui-quadrado. Assim sua verossimilhança não pode ser obtida através do filtro de Kalman. Os artigos de Ruiz (1994) e Harvey *et al.*(1994) despertaram interesses na análise desse modelo, ao sugerirem o estimador de quasi-verossimilhança, sendo a quasi-verossimilhança calculada utilizando a representação na **FEE**. Posteriormente vários autores, entre

eles Durbin e Koopman (1997a) consideraram a estimação da verossimilhança para modelos de espaço de estado não Gaussianos.

Por outro lado, Jacquier *et al.*(1994) deram início a pesquisa deste modelo com enfoque bayesiano. Shephard e Pitt (1997) propuseram um novo e mais eficiente algoritmo de maneira a melhorar a abordagem bayesiana dos modelos na **FEE**. Mas recentemente, Kim *et al.*(1998) aprimoraram os métodos de amostragem a fim de obter melhores resultados. No artigo Kim *et al.*(1998) fizeram análise bayesiana completa dos modelos **VE** e compararam-no com os modelos ARCH.

Utilizando integração numérica, Kitagawa (1987) propôs um algoritmo para estimar as densidades de filtragem e suavização, e conseqüentemente a verossimilhança. Esse método tem uma grande desvantagem quando aplicado a problemas com grande dimensão nos vetores das observações, problema este que Kitagawa (1996) propôs uma forma de reduzir este problema. Dentro desta abordagem, um grupo de pesquisadores ligados a Tanizaki tem contribuído bastante na análise de modelos representados na **FEE**. Tanizaki e Mariano (1994), Tanizaki e Mariano (1998), dentre outros, vem desenvolvendo diversos algoritmos a fim de estimar as densidades de filtragem e suavização, e assim a verossimilhança para estimar os parâmetros do modelo. Dessa forma, pode-se fazer a análise clássica dos modelos na forma de espaço de estados.

A estrutura da tese é dada como segue. No segundo capítulo serão apresentados os modelos dinâmicos e a representação na forma de espaço de estado, onde será apontadas as divergências e similaridades entre o enfoque clássico e o bayesiano no tratamento dos modelos que podem ser representados na forma de espaço de estados. No capítulo 3 o modelo de **VE** e suas propriedades são apresentadas, assim como

sua representação na forma de espaço de estado. No capítulo 4 são apresentadas as metodologias utilizadas na abordagem frequentista do modelo na forma de espaço de estados não Gaussianos. O capítulo 5 trata os enfoques bayesianos apresentados na literatura. No último capítulo as técnicas são aplicadas a séries simuladas e a dados empíricos, onde foi considerada a série do Índice da Bolsa de Valores do Estado de São Paulo (IBOVESPA) no período de 02/Jan/1995 a 27/Dez/2000 com 1500 observações. Os apêndices apresentam algumas técnicas de simulação, resultados em inferência bayesiana e o método de amostragem das densidades condicionais. Este último método, embora não tenha sido implementado, é apresentado por estar intimamente relacionado com o tema da tese.

CAPÍTULO 2

Modelos Dinâmicos e Forma de Espaço de Estado

Neste capítulo serão apresentados os conceitos de modelos dinâmicos e forma de espaço de estados. Será feita uma revisão considerando o caso Gaussiano e apresentadas algumas ferramentas que são utilizadas nesse caso e que são úteis no caso não Gaussiano. A abordagem bayesiana será apresentada, revisando os modelos lineares dinâmicos Gaussianos e apresentando o caso de maior interesse neste trabalho, o caso não Gaussiano.

2.1 Introdução

Apesar da abordagem através de modelos dinâmicos e forma de espaço de estados pertencerem a escolas diferentes, elas tem em comum o fato da representação ser basicamente a mesma, e muitas vezes utilizarem as mesmas técnicas probabilística e estatística.

Neste capítulo serão apresentadas as duas formulações para o problema. Os modelos dinâmicos e a forma de espaço de estado serão apresentados em uma forma geral com o propósito de fornecer uma estrutura teórica para desenvolvimentos abordados nos capítulos posteriores.

2.2 Modelos Dinâmicos

Harrison e Stevens (1976) apresentaram uma classe de modelos com parâmetros variando no tempo, definida como modelos dinâmicos. Esses modelos são claramente adequados à modelagem de regressão e séries temporais. Nesta seção serão apresentados alguns conceitos básicos desses modelos de forma a deixar clara a analogia direta com os modelos na forma de espaço de estados.

2.2.1 Modelos Lineares Dinâmicos

Uma das principais vantagens dos modelos lineares dinâmicos (MLD) sobre os modelos de regressão usuais é que ele permite que os coeficientes de regressão tenha variação temporal. Essa característica é explorada por pesquisadores na área de séries temporais financeiras, pois a flexibilidade dos coeficientes de regressão torna o modelo mais eficiente para capturar as tendências do mercado. A definição formal dada por West e Harrison (1997, pág. 107) é dada como segue

Definição 2.2.1. *O Modelo Linear Dinâmico Geral Normal é caracterizado pela quadrada*

$$\{\mathbf{F}, \mathbf{G}, \mathbf{V}, \mathbf{W}\}_t = \{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$$

onde:

- (i) \mathbf{F}_t matriz conhecida ($n \times r$)
- (ii) \mathbf{G}_t matriz conhecida ($n \times n$)
- (iii) \mathbf{V}_t matriz conhecida ($r \times r$)
- (iv) \mathbf{W}_t matriz conhecida ($n \times n$)

Essas matrizes definem a relação de Y_t com o vetor de parâmetros θ_t no tempo t , e o vetor θ_t no tempo da seguinte forma:

$$\text{(Equação de Observação)} \quad Y_t = \mathbf{F}'_t \theta_t + \nu_t, \quad \nu_t \sim N(0, \mathbf{V}_t) \quad (2.2.1)$$

$$\text{(Equação de Sistema)} \quad \theta_t = \mathbf{G}_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, \mathbf{W}_t) \quad (2.2.2)$$

$$\text{(Priori Inicial)} \quad (\theta_0 | D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0) \quad (2.2.3)$$

com momento a priori \mathbf{m}_0 e \mathbf{C}_0 . Assume-se que as perturbações ν_t e ω_t são independentes serialmente, mutuamente independentes e independente de (θ_0) .

A matriz \mathbf{F}_t desempenha o papel semelhante ao da matriz de desenho no modelo de regressão da variável independente, θ_t é o vetor dinâmico dos parâmetros de regressão, conhecido como vetor de estado ou vetor de sistema do modelo e ν_t é a perturbação das observações no tempo t , o qual possui média zero e variância \mathbf{V}_t .

Na equação de sistema, \mathbf{G}_t é chamada de matriz de evolução de estado, podendo depender de parâmetros desconhecidos e que variam com o tempo. A perturbação da equação (2.2.2), ω_t , possui média zero e variância \mathbf{W}_t .

A definição 2.2.1 considera o caso onde as perturbações são Gaussianas, pode-se supor erros com distribuição de caudas pesadas ou misturas de distribuição (ver Carter e Kohn, 1994).

O modelo da definição 2.2.1 possui a estrutura padrão de sinal mais ruído. O sinal é modelado através do vetor de estado que varia no tempo de acordo com a equação de sistema. Na prática o sinal pode ser interpretado como componentes não observados da série, a saber, tendências, periodicidade, dentre outros.

Equação de Atualização:

Será considerado o caso onde temos o **MLD** univariado. Dessa forma, dada uma informação a priori D_0 no tempo $t = 0$, West e Harrison (1997, pág. 27) denotaram a informação obtida até o tempo t por:

$$D_t = \{Y_t, D_{t-1}\}$$

onde Y_t é o valor observado da série no tempo t . Apesar de já existirem métodos computacionais para tratar esses modelos na ausência de normalidade, a veracidade de tal suposição facilita o cálculo das distribuições condicionadas do vetor de estado. O teorema a seguir será apresentado sem demonstração¹.

Teorema 2.2.1. *Para o **MLD** univariado da definição 2.2.1, as distribuições da previsão um passo à frente e distribuições posterioris são dadas como segue:*

➤ Posteriori em $t - 1$: Para alguma média \mathbf{m}_{t-1} e matriz de variância \mathbf{C}_{t-1} ,

$$(\theta_{t-1}|D_{t-1}) \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}). \quad (2.2.4)$$

➤ Priori em t :

$$(\theta_t|D_{t-1}) \sim N(\mathbf{a}_t, \mathbf{R}_t). \quad (2.2.5)$$

¹Ver West e Harrison, 1997, pág. 112-114)

onde

$$\begin{aligned} \mathbf{a}_t &= \mathbf{G}_t \mathbf{m}_{t-1} \\ \mathbf{R}_t &= \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t. \end{aligned}$$

> Previsão um passo à frente:

$$(Y_t | D_{t-1}) \sim N(f_t, Q_t), \quad (2.2.6)$$

onde

$$\begin{aligned} f_t &= \mathbf{F}_t' \mathbf{a}_t \\ Q_t &= \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t. \end{aligned}$$

> Posteriori em t :

$$(\theta_t | D_t) \sim N(\mathbf{m}_t, \mathbf{C}_t). \quad (2.2.7)$$

com

$$\begin{aligned} \mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t \\ \mathbf{C}_t &= \mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t' Q_t, \\ \mathbf{A}_t &= \mathbf{R}_t \mathbf{F}_t Q_t^{-1} \\ e_t &= Y_t - f_t. \end{aligned}$$

No modelo dado pela definição 2.2.1 as matrizes de regressão e evolução são definidas de acordo com a natureza do problema. Na prática, a variância \mathbf{W}_t é geralmente desconhecida e deve ser estimada e West e Harrison (1997) sugerem a sua avaliação através do conceito de fator de desconto.

Considere agora o modelo com a variância observacional constante e desconhecida.

Para a análise conjugada, geralmente trabalha-se com VW_t no lugar de W_t .

Definição 2.2.2. *Para todo t , o modelo é definido por:*

$$\begin{aligned} \text{(Equação de Observação)} \quad & Y_t = \mathbf{F}_t' \theta_t + \nu_t, & \nu_t &\sim N(0, V) \\ \text{(Equação de Sistema)} \quad & \theta_t = \mathbf{G}_t \theta_{t-1} + \omega_t, & \omega_t &\sim N(0, VW_t^*) \\ \text{(Informação Inicial)} \quad & (\theta_0 | D_0, \phi) \sim N(\mathbf{m}_0, V\mathbf{C}_0^*), \\ & (\phi | D_0) \sim G\left(\frac{n_0}{2}, \frac{d_0}{2}\right), \end{aligned}$$

onde $\phi = V^{-1}$

As quantidades iniciais \mathbf{m}_0 , \mathbf{C}_0^* , n_0 e d_0 são dadas. Analogamente, as matrizes \mathbf{W}_t^* também são fornecidas. Conforme o modelo anterior ainda temos a suposição de independência, só que agora temos a independência condicionada a V . Dessa forma segue o seguinte teorema o qual se encontra demonstrado em West e Harrison (1997, pág 119-120).

Teorema 2.2.2. *Dado o modelo como na definição 2.2.2, temos os seguintes resultados distribucionais, para $t \geq 1$.*

➤ Condicionado em V ,

$$\begin{aligned}(\theta_{t-1}|D_{t-1}, \mathbf{V}) &\sim N(\mathbf{m}_{t-1}, \mathbf{V}\mathbf{C}_{t-1}^*), \\(\theta_t|D_{t-1}, \mathbf{V}) &\sim N(\mathbf{a}_t, \mathbf{V}\mathbf{R}_t^*), \\(Y_t|D_{t-1}, \mathbf{V}) &\sim N(f_t, \mathbf{V}\mathbf{Q}_t^*), \\(\theta_t|D_t, \mathbf{V}) &\sim N(\mathbf{m}_t, \mathbf{V}\mathbf{C}_t^*).\end{aligned}$$

onde

$$\begin{aligned}f_t &= \mathbf{F}_t' \mathbf{a}_t \\ \mathbf{a}_t &= \mathbf{G}_t \mathbf{m}_{t-1} \\ \mathbf{Q}_t^* &= \mathbf{F}_t' \mathbf{R}_t^* \mathbf{F}_t + 1 \\ \mathbf{R}_t^* &= \mathbf{G}_t \mathbf{C}_{t-1}^* \mathbf{G}_t' + \mathbf{W}_t^*.\end{aligned}$$

a atualização das componentes é dada por

$$\begin{aligned}\mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t \\ \mathbf{C}_t^* &= \mathbf{R}_t^* - \mathbf{A}_t \mathbf{A}_t' \mathbf{Q}_t^*, \\ \mathbf{A}_t &= \mathbf{R}_t^* \mathbf{F}_t / \mathbf{Q}_t^* \\ e_t &= Y_t - f_t.\end{aligned}$$

➤ Para $\phi = V^{-1}$ teremos:

$$\begin{aligned}(\phi|D_{t-1}) &\sim G(n_{t-1}/2, d_{t-1}/2), \\(\phi|D_t) &\sim G(n_t/2, d_t/2).\end{aligned}$$

onde

$$\begin{aligned}n_t &= n_{t-1} + 1 \\ d_t &= d_{t-1} + e_t^2 / \mathbf{Q}_t^*.\end{aligned}$$

➤ Não condicionado em V teremos,

$$\begin{aligned}(\theta_{t-1}|D_{t-1}) &\sim T_{n_{t-1}}(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}), \\(\theta_t|D_{t-1}) &\sim T_{n_{t-1}}(\mathbf{a}_t, \mathbf{R}_t), \\(Y_t|D_{t-1}) &\sim T_{n_{t-1}}(f_t, Q_t), \\(\theta_t|D_t) &\sim T_{n_t}(\mathbf{m}_t, \mathbf{C}_t).\end{aligned}$$

onde

$$\begin{aligned}R_t &= S_{t-1}R_t^* \\Q_t &= S_{t-1}Q_t^* \\C_{t-1} &= S_{t-1}C_{t-1}^* \text{ e } C_t = S_tC_t^* \\S_{t-1} &= d_{t-1}/n_{t-1} \text{ e } S_t = d_t/n_t\end{aligned}$$

Muitos desses conceitos e teoremas serão demonstrados durante sua aplicação para modelos de volatilidade estocástica no capítulo 5 deste trabalho.

2.2.2 Modelos Dinâmicos Generalizados

Teorias envolvendo estes modelos dinâmicos generalizados foram inicialmente introduzidas por Migon (1984) e West, Harrison e Migon (1985). Essa classe dos modelos dinâmicos visa aperfeiçoar a modelagem quando a suposição de normalidade não é adequada. West, Harrison e Migon consideraram que as observações, Y_t , possui distribuição amostral pertencente a família exponencial. Dessa forma, supõe-se que o processo gerador de dados foi obtido de uma distribuição com a seguinte caracterização,

$$p(Y_t|\eta_t, V_t) = \exp\{V_t^{-1}[y_t(Y_t)\eta_t - a(\eta_t)]\}b(Y_t, V_t) \quad (2.2.8)$$

Definição 2.2.3. *O modelo linear dinâmico generalizado (DGLM) para uma série Y_t , ($t = 1, 2, \dots$) é definido por*

$$\begin{aligned}(\text{Modelo Observacional}) \quad & p(Y_t|\eta_t), & g(\eta_t) &= \lambda_t = F_t'\theta_t, \\(\text{Equação de Sistema}) \quad & \theta_t = \mathbf{G}_t\theta_{t-1} + \omega_t, & \omega_t &\sim (\mathbf{0}, \mathbf{W}_t)\end{aligned}$$

Pode-se verificar que temos nessa definição uma generalização do **MLD**. Aqui, não é suposto linearidade e normalidade na equação observacional, mas a equação de sistema continua análoga ao **MLD** exceto que as perturbações dessa equação não são consideradas Gaussianas. Considerando-se o modelo definido por 2.2.3 ganha-se em ajustes e perde-se no fato que não se tem mais o tratamento analítico como no caso onde tem-se a suposição de distribuição Gaussiana. Assim, para proceder as análises é necessário o uso de aproximações, seja numericamente, ou como sugerido por West, Harrison e Migon (1985), considerando aproximações Gaussiana para distribuição a priori e distribuição a posteriori.

Como não se tem a suposição de normalidade para **DGLM** tem-se que a posteriori para o vetor de estados será especificada por

$$(\theta_{t-1}|D_{t-1}) \sim (\mathbf{m}_{t-1}, \mathbf{C}_{t-1}), \quad (2.2.9)$$

Assim, segue que os momentos a priori de θ_t serão dados por

$$(\theta_t|D_{t-1}) \sim (\mathbf{a}_t, \mathbf{R}_t), \quad (2.2.10)$$

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

2.3 Forma de Espaço de Estado

O desenvolvimento inicial da metodologia de espaço foi dado em engenharia de controle, e não em estatística, iniciando com o artigo de Kalman (1960) em uma revista de engenharia. Neste artigo Kalman mostrou dois resultados de maior relevância. Primeiro, ele mostrou que uma extensa classe de problemas pode ser formulada como

um simples modelo linear, o modelo de espaço de estado. Depois ele mostrou que, devido a natureza Markoviana do modelo, os cálculos necessários para aplicação prática do modelo podem ser colocadas de forma recursiva, de maneira bastante conveniente para uso computacional.

As contribuições de estatísticos na metodologia de espaço de estado foi muito tímida até os anos 80. Desde então muitos nomes, como Harvey, Kitagawa, Stevens, Harrison, dentre outros, tem trabalhado intensamente sobre esta metodologia.

A principal vantagem da representação em **FEE** é que essa baseia-se na análise estrutural do problema. Isto permite, por exemplo, que vários componentes de uma série, tais como tendência, sazonalidade, ciclos e variações calendárias, juntamente com efeitos de variáveis explicativas e intervenções, sejam modeladas separadamente no modelo de espaço de estado **MEE**. *Trading-day* e efeito calendário também podem ser incorporados facilmente.

Adicionalmente, devido a natureza Markoviana dos **MEE**, os cálculos necessários podem ser colocados de forma recursiva. Esta natureza recursiva do modelo e das técnicas computacionais utilizadas na sua análise permite modelar diretamente mudanças na estrutura do sistema ao longo do tempo.

Outra vantagem dos modelos de espaço de estado é que eles são muito flexíveis e muito gerais. Eles cobrem uma grande variedades de modelos, incluindo os modelos da família Box-Jenkins (BJ). Além disso, a extensão para os casos multivariados é direta a partir da teoria univariada.

2.4 Forma de Espaço de Estado Linear e Gaussiana

Em geral, o modelo representado pelas duas equações a seguir é chamado modelo de espaço de estados.

$$\text{(Equação de Medida)} \quad y_t = Z_t \alpha_t + c_t + G_t \varepsilon_t \quad (2.4.1)$$

$$\text{(Equação de Transição)} \quad \alpha_t = T_t \alpha_{t-1} + d_t + R_t \eta_t \quad (2.4.2)$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right),$$

$$y_t : q \times 1, \quad Z_t : q \times k, \quad d_t : k \times 1, \quad G_t : q \times q, \quad \varepsilon_t : q \times 1,$$

$$\alpha_t : k \times 1, \quad T_t : k \times k, \quad c_t : q \times 1, \quad R_t : k \times k, \quad \eta_t : k \times 1,$$

onde, na literatura é suposto que ε_t é uma perturbação serialmente não correlacionada com $E(\varepsilon_t) = 0$ e $Var(\varepsilon_t) = H_t$, e α_t é um vetor $k \times 1$ não observável e assume-se que ele é gerado por um processo de Markov de primeira ordem, como visto na equação (2.4.2). Z_t e T_t são matrizes que relacionam os estados com a série y_t e com estados anterior, respectivamente. A variável η_t é um vetor aleatório de perturbações serialmente não-correlacionadas com $E(\eta_t) = 0$ e $Var(\eta_t) = Q_t$. Aos vetores c_t e d_t pode-se incorporar o efeito de covariáveis na série e estados respectivamente. Temos que Z_t, d_t, G_t, T_t, c_t e R_t podem depender de informação obtida no tempo $t - 1$ se assumimos normalidade para os erros ε_t e η (ver Harvey, 1989, pág. 100). A equação (2.4.1) é conhecida como equação de medida ou de observação e a equação (2.4.2) é chamada de equação de transição.

Tem-se adicionalmente mais duas suposições para os modelos na **FEE**:

- (i) O vetor inicial α_0 tem $E(\alpha_0) = a_0$ e $Var(\alpha_0) = \Sigma_0$.
- (ii) As perturbações ε_t e η_t são geralmente não-correlacionadas entre eles, para todo período de tempo, e não-correlacionadas com a variável de estado inicial, i.e., $E(\varepsilon_t \eta'_s) = 0$ para todo t e todo s , e $E(\varepsilon_t \alpha'_0) = E(\eta_t \alpha'_0) = 0$ para $t = 1, \dots, T$.

Dessas suposições pode-se verificar alguns pontos como segue:

- (1) A suposição (ii) assegura não correlação entre ε_t e α_t , e entre η_t e α_{t-1} , i.e., $E(\varepsilon_t \alpha'_t) = 0$ e $E(\eta_t \alpha'_{t-1}) = 0 \forall t$.
- (2) Z_t, d_t, G_t, T_t, c_t e H_t podem depender de um vetor de parâmetros desconhecidos ψ . Neste caso, deve-se estimar ψ .
- (3) Os erros aqui são assumidos normais, mas esta suposição pode ser relaxada.

2.4.1 Filtro de Kalman

O filtro de Kalman é um procedimento recursivo para encontrar estimadores ótimos² para o vetor de estados no tempo t , baseado em informações no tempo t . Essas informações consistem de observações y_1, \dots, y_t .

Quando a perturbações e o vetor de estado inicial são normalmente distribuído, o filtro de Kalman, através da decomposição do erro de predição, permite o cálculo da função de verossimilhança (ver Schweppe, 1965). Isso possibilita, junto com um

²Estimadores com mínimo erro quadrático médio

algoritmo adequado, encontrar EMV dos parâmetros do modelo e testar os parâmetros e especificação do modelo.

Equações do Filtro de Kalman

As equações do algoritmo do Filtro de Kalman são dadas por 2 estágios: Predição e Atualização.

Equações de predição :

Seja a_t o estimador ótimo de estado α_t baseado em informação até o tempo t , e seja P_t matriz de covariância do erro de estimação. Conhecendo a_{t-1} e P_{t-1} , o preditor de α_t é dado por

$$a_{t|t-1} = T_t a_{t-1} + c_t \quad (2.4.3)$$

enquanto a matriz de covariância da estimação do erro é

$$P_{t|t-1} = T_t P_{t-1} T_t' + R_t Q_t R_t', \quad t = 1, \dots, T \quad (2.4.4)$$

Depois de obter a nova informação, y_t , o estimador de α_t , $a_{t|t-1}$, pode ser atualizado.

Equações de atualização :

$$P_t = P_{t|t-1} - K_t F_{t|t-1} K_t' \quad (2.4.5)$$

$$a_t = a_{t|t-1} + K_t (y_t - y_{t|t-1}) \quad (2.4.6)$$

onde

$$y_{t|t-1} = Z_t a_{t|t-1} + d_t \quad (2.4.7)$$

$$F_{t|t-1} = Z_t P_{t|t-1} Z_t' + G_t H_t G_t', \quad (2.4.8)$$

$$K_t = P_{t|t-1} Z_t' F_{t|t-1}^{-1} \quad (2.4.9)$$

onde K_t é a matriz de ganho. As equações de predição e de atualização, juntas formam o conjunto recursivo chamado Filtro de Kalman.

Os valores iniciais do filtro de Kalman, no caso de modelo de séries temporais estacionário, podem ser inicializado em termos da média e matriz de variância não-condicionais de a_0 e P_0 ou $a_{1|0}$ e $P_{1|0}$.

Dada as condições iniciais o filtro de Kalman fornece o estimador ótimo do vetor de estado conforme cada nova observação é obtida.

Após todas as T observações tenham sido processada, o filtro retorna o estimador ótimo do vetor de estado corrente, e/ou o vetor de estado no próximo período de tempo, baseado em todo conjunto de informação. Este estimador contém todas as informações necessárias para fazer predição ótima de futuros valores de estados e observações.

2.4.2 Filtro de Kalman Estendido

Quando as suposições de linearidade e normalidade são satisfeitas procede-se conforme algoritmos que foram dados nas seções anteriores. Anderson e Moore (1979), Harvey (1989) dentre outros, propuseram um método para o caso de não linearidade das equações de medida e/ou estado. Esse método é conhecido como filtro de Kalman estendido. Sua base é aproximar as equações de medida e/ou estado por uma série de

Taylor de primeira ordem em torno das médias condicionais $a_{t|t-1}$ e $a_{t-1|t-1}$. Assim pode-se utilizar o filtro de Kalman para obtenção das estimativas.

Em geral essas aproximações produzem estimativas tendenciosas. Gelb (1974) propôs utilizar expansão de ordem mais altas para diminuir este vício. Essas aproximações não serão abordadas pois elas não serão utilizadas no presente trabalho.

Linearização

As equações de medida e transição são aproximadas como o seguinte **MEE**:

Equação de Medida

$$y_t = f_t(\alpha_t, \varepsilon_t) \approx f_{t|t-1} + Z_{t|t-1}(\alpha_{t-1} - a_{t-1|t-1}) + G_{t|t-1}\eta_t, \quad (2.4.10)$$

Equação de Transição

$$\alpha_t = h_t(\alpha_{t-1}, \eta_t) \approx h_{t|t-1} + T_{t|t-1}(\alpha_{t-1} - a_{t-1|t-1}) + H_{t|t-1}\eta_t, \quad (2.4.11)$$

onde tem-se,

$$\begin{aligned} f_{t|t-1} &= f_t(a_{t|t-1}, 0), \\ Z_{t|t-1} &= \left. \frac{\partial f_t(\alpha_t, \varepsilon_t)}{\partial \alpha'_t} \right|_{(\alpha_t, \varepsilon_t) = (a_{t|t-1}, 0)} \\ G_{t|t-1} &= \left. \frac{\partial f_t(\alpha_t, \varepsilon_t)}{\partial \varepsilon'_t} \right|_{(\alpha_t, \varepsilon_t) = (a_{t|t-1}, 0)} \\ h_{t|t-1} &= h_t(a_{t-1|t-1}, 0), \\ T_{t|t-1} &= \left. \frac{\partial h_t(\alpha_{t-1}, \varepsilon_t)}{\partial \alpha'_{t-1}} \right|_{(\alpha_{t-1}, \eta_t) = (a_{t-1|t-1}, 0)} \\ H_{t|t-1} &= \left. \frac{\partial h_t(\alpha_{t-1}, \eta_t)}{\partial \eta'_t} \right|_{(\alpha_{t-1}, \eta_t) = (a_{t-1|t-1}, 0)} \end{aligned}$$

Após a linearização aplica-se o algoritmo de **FK** para obter as estimativas desejadas.

2.4.3 Estimativa Suavizada Através do Filtro de Kalman

Um estimador suavizado é baseado em toda a informação contida na amostra; ou seja; toma-se a média condicionada a toda amostra, i.e., $E(\alpha_t|\mathbf{y})$. Como essa estimativa é baseada em um número maior de informação que a estimativa filtrada o (EQM) da estimativa suavizada vem a ser menor. Em um modelo gaussiano o estimador suavizado é dado pela equação:

$$a_{t|T} = E(\alpha_t) = E(\alpha_t|\mathbf{y}) \quad (2.4.12)$$

e a matriz de covariância de α_T condicional à todas as T observações é

$$P_{t|T} = E[(\alpha_t - a_{t|T})(\alpha_t - a_{t|T})'] \quad (2.4.13)$$

Isto é classicamente chamado de suavizamento do estado. Um método recursivo utilizado para suavizamento do estado consiste de três passos:

- O **FK** obtém F_t^{-1} , ν_t e K_t para $t = 1, \dots, n$,
- As estimativas suavizadas das perturbações são obtidas através das equações:

$$\begin{aligned} e_t &= F_t^{-1}\nu_t - K_t'r_t, & r_{t-1} &= Z_t'F_t^{-1}\nu_t + L_t'r_t, \\ d_t &= F_t^{-1} + K_t'N_tK_t, & N_t &= Z_t'F_t^{-1}Z_t + L_t'N_tL_t, \\ L_t &= T_t - K_tZ_t, & G_t &= R_tQ_tR_t'. \end{aligned} \quad (2.4.14)$$

tem-se os resultados para os erros tomando

$$\begin{aligned}
\varepsilon_t &= G_t G_t' e_t, & P_{t|T}^\varepsilon &= H_t H_t' - H_t H_t' d_t H_t H_t', \\
\tilde{\eta}_t &= H_t H_t' r_t, & P_{t|T}^\eta &= G_t G_t' - G_t G_t' N_t G_t G_t',
\end{aligned}
\tag{2.4.15}$$

- o estimador suavizado para o estado é obtido pela seguinte recursão:

$$a_{t+1|T} = T_t a_{t|T} + d_t + \tilde{\eta}_t \tag{2.4.16}$$

2.4.4 Suavizador de de Jong e Shephard

O algoritmo do **FK** fornece estimativas do vetor de estado utilizando as observações passadas. Tradicionalmente a densidade posteriori de α , os estados, dado Y_T é chamada de densidade suavizada. No caso Gaussiano existem algoritmos que obtém a média e matriz de covariância desta distribuição .

O algoritmo desenvolvido por de Jong e Shephard (1995), produz amostras aleatórias da densidade condicional multivariada de $u = (u'_1, \dots, u'_n)$ onde $u \sim (\Gamma \eta | Y_T)$, onde Γ é uma matriz diagonal de seleção com valores zero e um. Esta é introduzida para contornar problemas de amostrar de distribuições degeneradas e, pode-se gerar subconjuntos de η_t , o que é mais eficiente, principalmente quando o vetor de estado é de alta dimensão e é necessário apenas um subconjunto de amostra. A matriz F é escolhida de forma a evitar singularidades ³. O algoritmo é recursivo e requer a saída

³Ver de Jong e Shephard (1995) para maiores detalhes.

do **FK**, i.e., F_t^{-1} , ν_t e K_t . Seja $L_t = T_t - K_t Z_t$ e $J_t = H_t - K_t G_t$ para $t = T, \dots, 1$. Inicializando $N_T = 0$ e $r_T = 0$ tem-se as equações como segue:

$$\begin{aligned} C_t &= M_t(I - G_t' F_t^{-1} G_t - J_t' N_t J_t) M_t' \\ V_t &= M_t(G_t' F_t^{-1} Z_t + J_t' N_t J_t), \quad \xi_t \sim N(0, C_t) \\ r_{t-1} &= Z_t' F_t^{-1} \nu_t + L_t' r_t - V_t' C_t^{-1} \xi_t \\ N_{t-1} &= Z_t' F_t^{-1} Z_t + V_t' N_t V_t + L_t' N_t L_t \end{aligned}$$

2.5 Forma de Espaço de Estado Não-Gaussiano

Existem nas literaturas várias propostas para lidar com modelos representados na forma de espaço de estados. Com a evolução computacional e a introdução de novas técnicas de simulação, tem-se ampliado os estudos dos modelos na forma de espaço de estado não-Gaussiana (**FEENG**). Esses modelos serão de maior interesse nesta tese devido ao enfoque dado a um caso particular de modelos que podem ser representados na forma de espaço de estado não-Gaussiano **MEENG**, os modelos de **VE**. Nos capítulos posteriores serão dados maiores explicações sobre o modelo. Como o interesse é na aplicação ao modelo de **VE** serão consideradas perturbações não-Gaussianas somente na equação de observação.

2.5.1 Formas de Representação

Existem muitas formas de representação dos modelos na forma de espaço de estados. Durbin e Koopman (1997a) e Shephard e Pitt (1997) consideraram a equação de estados da forma linear e gaussiana como dada pela equação (2.4.2) e a equação de

medida de uma forma mais geral. Assim o modelo considerado foi escrito da seguinte forma:

$$y_t \sim f(y_t|\theta_t) \quad \theta_t = Z_t\alpha_t + c_t \quad (2.5.1)$$

$$\alpha_t = T_t\alpha_{t-1} + d_t + H_t\eta_t \quad \eta_t \sim NID(0, I) \quad (2.5.2)$$

onde y_t são as observações, $f(y_t|\theta_t)$ é uma densidade conhecida, $\theta_t = Z_t\alpha_t + c_t$ é o sinal no tempo t , que está deterministicamente relacionado linearmente com o estado correspondente. Em muitos trabalhos, tais como Durbin e Koopman (1997a) e Shephard e Pitt (1997), assume-se $f(y_t|\theta_t)$ log-côncava em θ_t . No caso específico do modelo de **VE** enfatizado neste trabalho, a suposição de log-concavidade é satisfeita. De maneira mais geral autores do grupo ligado a Mariano que trabalham com desenvolvimento de filtros não-lineares consideram o modelo representado da seguinte forma:

$$y_t = f_t(\alpha_t, \varepsilon_t) \quad (2.5.3)$$

$$\alpha_t = h_t(\alpha_{t-1}, \eta_t) \quad (2.5.4)$$

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right),$$

$$y_t : g \times 1, \quad \varepsilon_t : g \times 1, \quad \alpha_t : k \times 1, \quad \eta_t : k \times 1,$$

Assumindo que as perturbações ε_t e η_t são independentemente distribuídas, e não necessariamente Gaussianas. As funções f_t e h_t tem suas formas funcionais supostamente conhecidas e são inversíveis, i.e.,

$$y_t = f_t(\alpha_t, \varepsilon_t) \Rightarrow q_t(y_t, \alpha_t) = \varepsilon_t,$$

$$\alpha_t = h_t(\alpha_{t-1}, \eta_t) \Rightarrow g_t(\alpha_t, \alpha_{t-1}) = \eta_t,$$

2.5.2 Estimação, Filtragem e Suavizamento

Como no caso de **MEEG** o objetivo é estimar os parâmetros do modelo e as variáveis de estado. No caso dos modelos lineares e Gaussianos foi visto que, dado os parâmetros o **FK** fornece as estimativas dos estados a partir das densidades condicionais. A verossimilhança também é dada pelo filtro e assim, utilizando um algoritmo para maximizá-la pode-se encontrar as estimativas dos parâmetros desconhecidos no modelo.

Para o caso não Gaussiano as densidades condicionais muitas vezes não tem uma forma analítica, e quando existe pode não ser simples trabalhar com elas. Atualmente muitos pesquisadores se dedicam ao desenvolvimento dos modelos de espaço de estado não-Gaussiano **MEENG**.

Para os **MEENG** as densidades condicionais são dadas da seguinte forma:

Equação de Predição:

$$p(\alpha_t | Y_{t-1}) = \int p(\alpha_t, \alpha_{t-1} | Y_{t-1}) d\alpha_{t-1}$$

$$= \int p(\alpha_t | \alpha_{t-1}, Y_{t-1}) p(\alpha_{t-1} | Y_{t-1}) d\alpha_{t-1}$$

sabendo-se que a densidade conjunta de $\alpha_t, \dots, \alpha_1, y_{t-1}, \dots, y_1$ é dada por:

$$p(\alpha_t, \dots, \alpha_1, Y_{t-1}) = p(\alpha_t | \alpha_{t-1}) \cdots p(\alpha_1 | \alpha_0) p(y_{t-1} | \alpha_{t-1}) \cdots p(y_1 | \alpha_1), \quad (2.5.5)$$

onde

$$\begin{aligned}
p(\alpha_t, \alpha_{t-1}, Y_{t-1}) &= p(\alpha_t, \alpha_{t-1}, y_{t-1}, Y_{t-2}) \\
&= p(\alpha_{t-1}, Y_{t-2})p(y_{t-1}|\alpha_{t-1}, Y_{t-2})p(\alpha_t|\alpha_{t-1}, y_{t-1}, Y_{t-2}) \\
&= p(\alpha_{t-1}, Y_{t-2})p(y_{t-1}|\alpha_{t-1})p(\alpha_t|\alpha_{t-1}),
\end{aligned}$$

onde a densidade conjunta $p(\alpha_{t-1}, Y_{t-2})$ é dada por:

$$\begin{aligned}
p(\alpha_{t-1}, Y_{t-2}) &= \\
&= \int \cdots \int p(\alpha_t|\alpha_{t-1}) \cdots p(\alpha_1|\alpha_0)p(y_{t-1}|\alpha_{t-1}) \cdots p(y_1|\alpha_1) d\alpha_{t-2}, \cdots, d\alpha_1.
\end{aligned}$$

Agora, para encontrar a densidade marginal conjunta de Y_{t-1} tem-se,

$$\begin{aligned}
p(Y_{t-1}) &= \iint p(\alpha_t|\alpha_{t-1})p(y_{t-1}|\alpha_{t-1})p(\alpha_{t-1}, Y_{t-2}) d\alpha_t d\alpha_{t-1} \\
&= \int p(y_{t-1}|\alpha_{t-1})p(\alpha_{t-1}, Y_{t-2}) d\alpha_{t-1}.
\end{aligned}$$

Portanto a densidade condicional será dada como segue,

$$\begin{aligned}
p(\alpha_t, \alpha_{t-1}|Y_{t-1}) &= p(\alpha_{t-1}|Y_{t-1})p(\alpha_t|\alpha_{t-1}, Y_{t-1}) \\
&= \frac{p(\alpha_t|\alpha_{t-1})p(y_{t-1}|\alpha_{t-1})p(\alpha_{t-1}, Y_{t-2})}{\int p(y_{t-1}|\alpha_{t-1})p(\alpha_{t-1}, Y_{t-2}) d\alpha_{t-1}} \\
&= p(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1})
\end{aligned}$$

Assim, a **equação de predição** é dada por:

$$p(\alpha_t|Y_{t-1}) = \int p(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1}) d\alpha_{t-1}$$

Equação de Atualização:

$$\begin{aligned}
 p(\alpha_t|Y_t) &= \frac{p(Y_{t-1})p(\alpha_t|Y_{t-1})p(y_t|\alpha_t, Y_{t-1})}{p(Y_{t-1})p(y_t|Y_{t-1})} \\
 &= \frac{p(y_t|\alpha_t)p(\alpha_t|Y_{t-1})}{p(y_t|Y_{t-1})}, \\
 p(y_t|Y_{t-1}) &= \int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1}) d\alpha_t
 \end{aligned} \tag{2.5.6}$$

2.5.3 Filtro de Partículas

Tanto na equação de predição quanto na de atualização é necessário calcular uma integral, que pode ser vista como a esperança de uma densidade. Por exemplo, no caso de atualização, deseja-se calcular a esperança de $p(y_t|\alpha_t)$ na população com densidade $p(\alpha_t|Y_{t-1})$. Esta integral pode ser calculada através de simulação Monte Carlo como pode ser visto no apêndice A. Um método de simulação que pode ser utilizado é o chamado filtro de partículas.

O nome filtro de partículas foi proposto por Carpenter *et al.*(1998) apesar de Kitagawa (1996) já ter utilizado o termo partículas. Filtro de partículas é uma classe de filtros de simulação que recursivamente aproxima a distribuição da variável aleatória pela distribuição de "partículas" $\alpha_t^1, \dots, \alpha_t^M$, com distribuição de probabilidade discreta π_t^1, \dots, π_t^M . Assim, a variável contínua é aproximada por uma discreta com suporte aleatório. Os pontos discretos são considerados como sendo uma amostra da população. A maioria dos pesquisadores envolvidos com esse tema consideram que os π_t^j são todos iguais a $1/M$, onde considera-se M suficientemente grande.

Os filtros de partículas consideram o suporte discreto gerado por partículas, como a densidade verdadeira. Isso nos permite produzir uma aproximação para esta densidade. Por exemplo, no caso de densidade de predição, utilizando o suporte discreto

de partículas, obtém-se

$$\hat{p}(\alpha_t|Y_{t-1}) = \sum_{j=1}^M p(\alpha_t|\alpha_{t-1}^j)\pi_{t-1}^j, \quad (2.5.7)$$

uma "densidade de predição empírica" como uma estimativa da densidade procurada. Analogamente a estimativa da densidade de filtragem da equação (2.5.6), é obtida da seguinte maneira

$$\hat{p}(\alpha_t|Y_t) \propto p(y_t|\alpha_t) \sum_{j=1}^M p(\alpha_t|\alpha_{t-1}^j)\pi_{t-1}^j \quad (2.5.8)$$

a "densidade de filtragem empírica". Assim, o filtro de partículas amostra desta densidade para obter novas partículas $\alpha_{t+1}^1, \dots, \alpha_{t+1}^M$ com pesos $\pi_{t+1}^1, \dots, \pi_{t+1}^M$.

Uma forma de amostrar da densidade filtragem é considerar a função dada por $\sum_{j=1}^M f(\alpha_t|\alpha_{t-1}^j)\pi_{t-1}^j$ como uma densidade a priori (2.5.7) que combinada com a verossimilhança $\hat{p}(y_t|\alpha_t)$ fornece a posteriori (2.5.8). Pode-se amostrar de (2.5.7) tomando α_{t-1}^j com probabilidade π_{t-1}^j e amostrando de $p(\alpha_t|\alpha_{t-1}^j)$. Consegue-se obter $p(y_t|\alpha_t)$ a menos de uma constante, e pode-se utilizar métodos tradicionais, tais como MCMC, amostragem/reamostragem importante, para amostragem de tal distribuição.

A fim de tornar o filtro de partículas mais robusto, Pitt e Shephard (1999a) propuseram um filtro de partículas com variável auxiliar. A idéia foi amostrar de uma densidade conjunta $p(\alpha_t, k|Y_t)$ onde k é o índice da mistura em (2.5.7). Assim, eles definiram

$$p(\alpha_t, k|Y_{t-1}) = p(y_t|\alpha_t)p(\alpha_t|\alpha_{t-1}^k)\pi_{t-1}^k, \quad k = 1, \dots, M. \quad (2.5.9)$$

Amostrando-se dessa densidade conjunta e descartando o índice, obtém-se amostras da densidade empírica de filtragem (2.5.8). Como anteriormente usa-se os métodos tradicionais para amostrar da densidade conjunta ⁴. Kim *et al.*(1998) utilizaram desta técnica para filtragem na análise dos modelos **VE**.

2.6 Conclusões

Resumidamente, pode-se dizer que o objetivo principal é encontrar a estimativa condicional do vetor de estado, i.e., $E(\alpha_t|Y_s)$ onde a estimativa é chamada de suavizada quando $s > t$ (geralmente $s = T$), filtrada quando $s = t$ e de predição quando $s < t$. O algoritmo para a densidade de predição L passos à frente é dado por:

$$p(\alpha_{t+L}|Y_t) = \int p(\alpha_{t+L}|\alpha_{t+L-1})p(\alpha_{t+L-1}|Y_t) d\alpha_{t+L-1} \quad (2.6.1)$$

onde $p(\alpha_{t+L}|\alpha_{t+L-1})$ pode ser obtida da equação (2.5.4). Ou seja, dada a função densidade $p(\alpha_{t+L}|\alpha_{t+L-1})$, a densidade de predição, $p(\alpha_{t+L}|Y_t)$, para $L = 1, \dots$, pode ser obtida e recursivamente obtém-se $p(\alpha_{t+L+1}|Y_t)$.

Para filtragem o algoritmo para densidade recursiva é dado por:

$$p(\alpha_t|Y_{t-1}) = \int p(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1}) d\alpha_{t-1} \quad (2.6.2)$$

⁴Para maiores detalhes ver Pitt e Shephard (1999a)

$$p(\alpha_t|Y_t) = \frac{p(y_t|\alpha_t)p(\alpha_t|Y_{t-1})}{\int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1}) d\alpha_{t-1}} \quad (2.6.3)$$

onde as densidades $p(y_t|\alpha_t)$ e $p(\alpha_t|\alpha_{t-1})$ são obtidas através da equação de observação e equação de transição respectivamente. Baseado em duas densidades, Equação (2.6.2) produz $p(\alpha_t|Y_{t-1})$ dado $p(\alpha_{t-1}|Y_{t-1})$ e Equação (2.6.3) produz $p(\alpha_t|Y_t)$ dado $p(\alpha_t|Y_{t-1})$.

O algoritmo de suavização é dado da seguinte forma

$$p(\alpha_t|Y_T) = p(\alpha_t|Y_t) \int \frac{p(\alpha_{t+1}|Y_T)p(\alpha_{t+1}|\alpha_t)}{p(\alpha_{t+1}|Y_t)} d\alpha_{t+1}, \quad (2.6.4)$$

onde $p(\alpha_t|Y_t)$ e $p(\alpha_{t+1}|Y_t)$ são obtidas do algoritmo de filtragem em (2.6.2) e (2.6.3), enquanto $p(\alpha_{t+1}|\alpha_t)$ é obtida da equação de transição (2.5.4).

Obtidas as densidades, o interesse está em estimar algumas quantidades, como a esperança condicional e variância de certas funções de α_r :

$$g_{r|s} \equiv E(g(\alpha_r)|Y_s) = \int g(\alpha_r)p(\alpha_r|Y_s) d\alpha_r, \quad (2.6.5)$$

Como pode-se verificar existem parâmetros desconhecidos no modelos, e para obtermos estimativas desses parâmetros, calcula-se a verossimilhança por:

$$\begin{aligned} p(Y_T) &= \prod_{t=1}^T p(y_t|Y_{t-1}) \\ &= \prod_{t=1}^T \left(\int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1}) d\alpha_t \right). \end{aligned} \quad (2.6.6)$$

Note que $p(y_t|Y_{t-1})$ é dado pelo denominador da Eq. (2.6.3).

CAPÍTULO 3

Modelo de Volatilidade Estocástica

O modelo de volatilidade estocástica foi proposto por Taylor (1986, pág. 30) para modelar retornos de ativos, onde a volatilidade aparece como um processo latente a ser estimado. Uma estimativa da volatilidade é de vital importância no mercado financeiro, pois a partir dessa pode-se, por exemplo, precificar opções e medir o risco de um ativo. Nesse capítulo será apresentado inicialmente alguns conceitos e características de retornos financeiros, apresentado o modelo de **VE** e mostrado como este modelo pode gerar séries com algumas das características encontradas nas séries de retornos.

3.1 Retornos Diários das Séries Financeiras

Geralmente os investidores não estão interessados no preço de um ativo, e sim no retorno que eles irão obter pelo investimento. Se o preço de fechamento (abertura) de um ativo é definido como p_t então o retorno, ao qual chamamos de retorno simples, é definido como

$$y_t^{(s)} = \frac{p_t - p_{t-1}}{p_{t-1}}$$

Um dos primeiros benefícios verificados é a perda das unidades de medida. Ele é invariante por escala, o que facilita a comparação da performance de vários ativos. Essa não é a única forma de definição de retorno. Uma outra forma de definição que é chamada de retorno composto é dada por

$$y_t = \log \frac{p_t}{p_{t-1}}$$

Enquanto o retorno simples também é conhecido como retorno discreto, o composto é contínuo em sua natureza. Esses retornos estão relacionados da seguinte maneira:

$$\begin{aligned} \log\left(\frac{p_t}{p_{t-1}}\right) &= \log\left(1 + \frac{p_t - p_{t-1}}{p_{t-1}}\right) \\ &= \log\left(1 + y_t^{(s)}\right) \approx y_t^{(s)} \quad \text{para } y_t^{(s)} \text{ pequeno.} \end{aligned}$$

3.2 Séries Financeiras

A sequência de valores de preços de ativos tais como, ações, taxas de câmbio, etc., formam uma série temporal financeira. Pesquisas nessa área tiveram um grande impulso após o artigo de Mandelbrot (1963). Neste trabalho ele fixou suas atenções nos fatos estilizados da distribuição não condicional dos retornos das séries temporais financeiras. Ele constatou características como, caudas pesadas, a mudança de variância com o tempo, i.e., os retornos são heteroscedástico e que altos (baixos) retornos (em módulo) são geralmente seguidos por altos (baixos) retornos. Essas características peculiares conhecidas como fatos estilizados da variância, motivaram muitos pesquisadores para trabalhos teóricos e aplicados, dado que, os primeiros modelos teóricos assumiam que os retornos das séries temporais financeiras seguiam um processo Gaussiano não correlacionados com média zero, ou seja, não poderiam exibir caudas pesadas e eram homoscedástico. A seguir são apresentados resumidamente alguns fatos estilizados.

Fatos Estilizados

(a) Caudas Pesadas

Mandelbrot (1963) e Fama (1963, 1965), dentre outros, observaram que os retornos tem distribuição leptocúrtica. Como resultado vários pesquisadores propuseram modelar os retornos de ações com amostras *i.i.d* de distribuições com caudas pesadas, tais como mistura de normais, Lévy e Pareto.

(b) Conglomerados de Volatilidade

As séries financeiras apresentam períodos de alta e baixa variabilidade. Assim, pode-se observar que períodos de baixa (alta) volatilidade tende a ser seguido por

período de baixa (alta) volatilidade. Vale ressaltar que os modelos lineares tradicionais de séries temporais não são capazes de reproduzir tais características. Os fenômenos de conglomerados de volatilidade e caudas pesadas dos retornos das ações estão intimamente relacionados.

(c) *Efeito Alavanca (Leverage Effect)*

O efeito alavanca citado por Black (1976), sugere que o movimento dos preços das ações estão negativamente correlacionados com a volatilidade. O motivo é que retornos negativos (preços em baixa) implicam em maior incerteza, i.e., maior volatilidade.

(d) *Memória Longa e Persistência*

Observa-se que a volatilidade em um certo instante de tempo t geralmente é próxima a volatilidade no tempo $t - 1$, principalmente em dados de altas frequências. Esta alta persistência, também chamada de memória longa, implicará em raízes próximas a unitária no ajuste de modelos do tipo ARCH e VE.

(e) *Efeito Calendário*

A volatilidade em séries financeiras pode ser influenciada pelo calendário. Por exemplo, espera-se que na segunda-feira a volatilidade seja maior, devido ao fato dos mercados terem permanecido fechados durante o fim de semana. Nos finais de semana e feriados, quando os mercados se encontram fechados, muitas informações que afetam o mercado são obtidas e influenciarão o mercado na sua abertura.

O período do mês ou do ano, quando o governo anuncia os dados macroeconômicos mensais ou anuais, ou mesmo os anúncios de novas políticas econômicas, também tendem a afetar a volatilidade.

(f) Volatilidade com Movimentos Semelhantes

Existe uma extensa literatura sobre movimentos comuns em mercados internacionais especulativos. As correlações entre os retornos de ativos e de suas volatilidades tem sido objeto de recentes estudos. Geralmente usam-se modelos fatoriais para modelar a comunalidade das volatilidades internacionais e explorar os padrões e as tendências comuns das volatilidades.

3.2.1 Modelo Básico para Retornos de Séries Financeiras

Na especificação inicial geralmente considera-se que o retorno é composto de um nível mais uma perturbação na forma:

$$r_t^* = \mu_t + \sigma_t \varepsilon_t$$

O termo μ_t descreve o nível médio esperado do retorno r_t^* no tempo t . Efeitos calendários e outras variáveis explicativas podem ser incorporados na modelagem de μ_t , por exemplo efeito autoregressivo e dummies para dias da semana. Por outro lado, $\sigma_t \varepsilon_t$ governa a perturbação em torno da média no tempo t . ε_t é uma variável aleatória estacionária independente e identicamente distribuída com média zero e σ_t é chamada volatilidade dos retornos. Dada a formulação do modelo, resta modelar μ_t e σ_t e escolher a distribuição de ε_t . A suposição usualmente adotada para o processo $\{\varepsilon_t\}$ é que são *i.i.d.* Algumas distribuições sugeridas na literatura são distribuição normal padrão, distribuição *t*-student ou distribuição de erro generalizado (*GED*).

3.2.2 Volatilidade Estocástica

Existem diferentes modelos propostos para estimar a volatilidade. Engle (1982) propôs os modelos ARCH, os quais foram modificados por vários autores dando origem a uma família de modelos denotados por XARCH. A segunda maneira é modelar através do modelo de volatilidade estocástica, proposto por Taylor (1986), que será apresentado a seguir.

3.3 O Modelo de Volatilidade Estocástica

O modelo de volatilidade estocástica é uma alternativa aos modelos ARCH propostos por Engle (1982) e suas várias extensões. A diferença básica entre esses modelos e o modelo de volatilidade estocástica **VE** é que os modelos ARCH consideram a volatilidade como uma variável observável pois ela é conhecida quando os parâmetros do modelo são conhecidos.

O modelo geral de volatilidade estocástica formulado inicialmente por Taylor (1980) considera, $\mathbf{y} = y_1, \dots, y_T$, como um vetor gerado por um modelo de probabilidade, $p(\mathbf{y}|\mathbf{h})$, onde \mathbf{h} é um vetor de volatilidades, sendo que a variância h_t , evolui temporalmente, independente dos valores passados de y_t . As volatilidades são não observáveis e assume-se que são geradas por um mecanismo de probabilidade, $p(\mathbf{h}|\theta)$. A definição do modelo é dada a seguir.

Definição 3.3.1. *Um processo y_t é chamado processo de volatilidade estocástica se*

$$y_t = \sigma_t \varepsilon_t \quad (3.3.1)$$

onde ε_t é um processo estacionário com média zero e variância um, e σ_t é a volatilidade de y_t no tempo t o qual é descrito pela seguinte relação.

$$\sigma_t = e^{\frac{1}{2}h_t} \quad (3.3.2)$$

onde h_t segue um processo que pode ou não ser estacionário¹. O termo h_t é dito ser a equação de volatilidade de y_t e independe dos valores passados de y_t .

Na formulação mais simples h_t é modelado como um processo autoregressivo de primeira ordem

$$h_t = \alpha + \phi h_{t-1} + \eta_t \quad (3.3.3)$$

onde α e ϕ são parâmetros invariantes no tempo e η_t é um processo estacionário Gaussiano com média zero e variância σ_η^2 .

1. Assume-se que as perturbações ε_t e η_t são ruídos branco Gaussianos independentes com variâncias 1 e σ_η^2 respectivamente.
2. O termo h_t é estacionário, então devemos impor que $|\phi| < 1$.

Na literatura pode-se encontrar o modelo de **VE** representado de muitas maneiras, no artigo de Kim *et al.* (1998) os pesquisadores consideraram o modelo da forma canônica, ou seja, o modelo das equações (3.3.1), (3.3.2) e (3.3.3) é dado por

$$\begin{aligned} y_t &= \beta e^{h_t'/2} \varepsilon_t, & t \geq 1, \\ h'_{t+1} &= \mu + \phi(h'_t - \mu) + \sigma_\eta \eta_t, \\ h'_t &\sim N\left(\mu, \frac{\sigma_\eta^2}{(1 - \phi^2)}\right), \end{aligned} \quad (3.3.4)$$

¹Em nosso caso trabalharemos com processo estacionário

Onde ε_t e η_t tem distribuição normal padrão e fixa-se $\beta = 1$ ou $\mu = 0$ por problemas de identificabilidade, as outras suposições seguem como anteriormente. Outra forma de representação é a que foi considerada por Jacquier *et al.*(1994), Sandmann e Koopman (1998), Fridman e Harris (1998), dentre outros, que consideram que a variância condicional segue um processo log-AR(1). Assim temos

$$\begin{aligned} y_t &= h_t^{1/2} \varepsilon_t \quad t \geq 1, \\ \log h_t &= \alpha + \phi \log h_{t-1} + \sigma_\eta \eta_t, \end{aligned} \tag{3.3.5}$$

Transformação Logaritmo do Quadrado

Para obtermos linearidade na equação (3.3.1), tomamos uma transformação. Assim, elevando-se ao quadrado e tomando o logaritmo neperiano temos:

$$\log(y_t^2) = \log(\sigma_t^2) + \log(\varepsilon_t^2) \tag{3.3.6}$$

e modelamos $\log(\sigma_t^2)$ como o processo autoregressivo em (3.3.3), onde $h_t = \log(\sigma_t^2)$.

Dado que ε_t é normalmente distribuído com $E[\varepsilon_t] = 0$ e $Var[\varepsilon_t] = 1$ então $\log(\varepsilon_t^2)$ segue a distribuição *log-qui-quadrado* com:

$$\begin{aligned} E[\log(\varepsilon_t^2)] &\cong -1.27, \\ V[\log(\varepsilon_t^2)] &= \frac{\pi^2}{2}, \end{aligned} \tag{3.3.7}$$

e fazendo a transformação tal que:

$$\xi_t = \log(\varepsilon_t^2) - E[\log(\varepsilon_t^2)] = \log(\varepsilon_t^2) + 1.27, \tag{3.3.8}$$

as equações (3.3.6) e (3.3.3) serão modificadas para:

$$\begin{aligned} \log(y_t^2) &= -1.27 + h_t + \xi_t & \xi_t &\sim ID(0, \frac{\pi^2}{2}), \\ h_t &= \alpha + \phi h_{t-1} + \eta_t & \eta_t &\sim NID(0, \sigma_\eta^2), \\ h_t &= \log(\sigma_t^2) & \xi_t &\perp \eta_t. \end{aligned} \quad (3.3.9)$$

Pela definição de ξ_t temos que $E[\xi_t] = 0$ e $Var[\xi_t] = \frac{\pi^2}{2}$ e os ξ_t são independentes.

Outra vantagem desta transformação é que, em geral, para as séries financeiras empíricas da distribuição de $\log(y_t^2)$ é mais simétrica do que a distribuição de y_t . Porém o coeficiente de simetria é ligeiramente negativo.

3.3.1 Propriedades do Modelo de Volatilidade Estocástica

Esta seção é destinada as propriedades básicas do modelo de VE definido na seção anterior. Aqui assume-se que o processo estocástico h_t é um processo AR(1) e todas as suposições consideradas anteriormente.

Média

Como ε_t e η_t são independentes, σ_t e ε_t também o são. E a média do processo é igual a zero pois

$$E(y_t) = E(\sigma_t \varepsilon_t) = E(\sigma_t)E(\varepsilon_t) = 0 \quad (3.3.10)$$

Muitos pesquisadores assumem que a média da série dos retornos, r_t , é zero ou desprezível.

Variância

A variância não condicional do modelo **VE** é dada por

$$\begin{aligned} \text{Var}(y_t) &= E(y_t - E(y_t))^2 \stackrel{(3.3.10)}{=} E(y_t^2) \\ &= E(\sigma_t^2)E(\varepsilon_t^2) = E(\sigma_t^2) \end{aligned}$$

pois $E(\varepsilon_t^2) = 1$. Como σ_t^2 é dada por um termo estocástico h_t , sabendo-se a distribuição de h_t temos a distribuição de σ_t^2 . Como a perturbação η_t é normalmente distribuída, o processo autoregressivo, h_t , tem distribuição não condicional normal com os parâmetros dados por

$$h_t \sim N\left(\frac{\alpha}{1-\phi}, \frac{\sigma_\eta^2}{1-\phi^2}\right)$$

onde $\alpha, \phi, \sigma_\eta$ são os parâmetros dados em (3.3.3). Desde que $\sigma_t^2 = e^{h_t}$, σ_t^2 tem distribuição lognormal. Assim o k -ésimo momento central de σ_t^2 é dado por $e^{(k\mu_h + \frac{1}{2}k^2\sigma_h^2)}$, onde μ_h e σ_h^2 são a média e variância de h_t respectivamente. Logo a variância condicional do modelo **VE** é como segue

$$\begin{aligned} E(y_t^2) &= E(\sigma_t^2) \\ &= \exp\left(\mu_h + \frac{1}{2}\sigma_h^2\right) \\ &= \exp\left(\frac{\alpha}{1-\phi} + \frac{\sigma_\eta^2}{2(1-\phi^2)}\right) \end{aligned} \tag{3.3.11}$$

a variância será sempre positiva, devido os valores possíveis dos parâmetros presentes na equação (3.3.3). As únicas restrições necessárias para esses parâmetros no momento da estimação é que $|\phi| < 1$, porque o processo, h_t será não estacionário quando $|\phi| = 1$ ou explosivo quando $|\phi| > 1$.

Momentos de Ordem Superior

Considerando ε_t uma variável aleatória simétrica temos que os momentos de ordem ímpar, serão todos iguais a zero, dado que a distribuição de y_t é simétrica em torno de zero. O quarto momento do modelo será dado por

$$\begin{aligned} E(y_t^4) &= E(\sigma_t^4)E(\varepsilon_t^4) \\ &= E(\varepsilon_t^4)E(e^{2h_t}) \\ &= 3e^{(2\mu_h + 2\sigma_h^2)} \end{aligned}$$

Portanto a estatística de curtose será dada por

$$\begin{aligned} \frac{E(y_t^4)}{E(y_t^2)^2} &= \frac{3e^{(2\mu_h + 2\sigma_h^2)}}{(\exp(\mu_h + \frac{1}{2}\sigma_h^2))^2} \\ &= 3 \exp(\sigma_h^2) \end{aligned}$$

Como a variância de h_t é estritamente positiva, temos que $e^{\sigma_h^2} > 1$ e a curtose para o modelo sempre é maior que 3 (estatística de curtose da normal), que é uma característica desejada, já que esse é um dos fatos estilizados das séries financeiras (caudas pesadas).

Autocorrelação

A estrutura de correlação de um processo estocástico é de importância fundamental, pois o conhecimento da estrutura da correlação de tais processos auxilia no entendimento do processo subjacente. No modelo proposto as duas variáveis aleatórias ε_t e η_t são serialmente não correlacionadas. Assim, combinando esta propriedade com o fato de que a média do processo **VE** é zero, a estrutura de auto-covariância é dada por

$$\begin{aligned}
E(y_t y_{t+\tau}) &= E((\sigma_t \varepsilon_t)(\sigma_{t+\tau} \varepsilon_{t+\tau})) \\
&= E((\sigma_t \sigma_{t+\tau})(\varepsilon_t \varepsilon_{t+\tau})) \\
&= 0
\end{aligned}$$

onde assume-se que as perturbações ε_t e η_t são serialmente não correlacionadas. Então para todo $\tau > 0$ o processo **VE** é serialmente não correlacionado.

O processo **VE** é serialmente não correlacionado, mas não independente, pois existe correlação, por exemplo, nas transformações $\log(y_t^2)$ do processo como visto em seção anterior. Das equações em (3.3.9) a função de autocorrelação para $\log y_t^2$ é dada por

$$\gamma(\tau) = E(\log y_t^2 - E \log y_t^2)(\log y_{t+\tau}^2 - E \log y_{t+\tau}^2)$$

Dado que $E(\log y_t^2) = -1.27 + E(h_t)$, então $\log y_t^2 - E(\log y_t^2) = h_t - E h_t + \xi_t$. Portanto, usando o fato de que h_t é independente de ξ_t tem-se,

$$\begin{aligned}
\gamma(\tau) &= E(h_t - E h_t + \xi_t)(h_{t+\tau} - E h_{t+\tau} + \xi_{t+\tau}) \\
&= E(h_t - E h_t)(h_{t+\tau} - E h_{t+\tau}) + E(\xi_t \xi_{t+\tau})
\end{aligned} \tag{3.3.12}$$

Além disso tem-se que:

$$\gamma(\tau) = \gamma_h(\tau) + \gamma_\xi(\tau)$$

utilizando-se do fato que h_t é um processo autoregressivo de ordem 1, a função de auto-covariância de h_t será dada por

$$\gamma_h(\tau) = \phi^\tau \frac{\sigma_\eta^2}{1 - \phi^2}$$

Para $\tau > 0$ e $\gamma_\xi(\tau) = 0$. E temos $\gamma_h(\tau) = \gamma(\tau)$ para τ diferente de zero. Assim, a função de autocorrelação é dada como segue

$$\begin{aligned} \rho_{\log y_t^2}(\tau) &= \frac{\gamma(\tau)}{\gamma(0)} = \frac{\phi^\tau \frac{\sigma_\eta^2}{1 - \phi^2}}{\gamma_h(0) + \gamma_\xi(0)} \\ &= \frac{\phi^\tau}{1 + \sigma_h^{-2} \frac{\pi^2}{2}} \end{aligned}$$

ou seja, a ACF decai exponencialmente a partir do segundo lag; que é uma característica de um processo ARMA(1,1). Nos ajustes de séries financeiras, em geral, são encontradas estimativas de ρ_1 baixa (dependendo de $\sigma_h^{-2} \frac{\pi^2}{2}$) e de ϕ próximo de 1. Esta alta correlação da volatilidade é que gerará séries com características de conglomerados de volatilidade.

3.3.2 Formulação na Forma de Espaço de Estado

Após a transformação do modelo na forma da equação (3.3.9), pode-se facilmente representá-lo na forma de espaço de estados descrita no capítulo 2, onde as equações de medida e transição serão dadas, respectivamente, por

$$y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ h_t \end{bmatrix} + \xi_t$$

$$\begin{bmatrix} \alpha \\ h_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} \alpha \\ h_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \end{bmatrix}$$

onde as matrizes do sistema representado pelas equações (2.4.1) e (2.4.2) são dadas por

$$Z_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad T_t = \begin{bmatrix} 1 & 0 \\ 0 & \phi \end{bmatrix} \quad H_t = 4.93 \quad Q_t = \sigma_\eta^2$$

Algumas representações das extensões do modelo, incluindo efeitos calendários e *outliers* podem ser encontrados em Motta (1998) e Fukui (2000) respectivamente. Como serão feitas algumas análises considerando *outliers* do tipo AO (*aditive outlier*) e IO (*innovation outlier*)² serão dadas as representações desses casos.

A ocorrência de um AO na t_{AO} -ésima observação e de um IO na t_{IO} -ésima observação, são incorporadas ao modelo da seguinte forma

$$\log(y_t^2) = -1.27 + h_t + \xi_t + \Delta_t \quad (3.3.13)$$

onde $\Delta_t = 0$ p/ $t \neq t_{AO}$ e $\Delta_t \neq 0$ p/ $t = t_{AO}$

$$h_t = \alpha + \phi h_{t-1} + \eta_t + \Lambda_t \quad (3.3.14)$$

onde $\Lambda_t = 0$ p/ $t \neq t_{IO}$ e $\Lambda_t \neq 0$ p/ $t = t_{IO}$

$$(3.3.15)$$

²Para maiores detalhes ver Fukui (2000)

$$y_t = \begin{bmatrix} 1 & 1 & I_{AO} & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ h_t \\ \Delta_t \\ \Lambda_t \end{bmatrix} + \xi_t \quad (3.3.16)$$

$$\begin{bmatrix} \alpha \\ h_t \\ \Delta_t \\ \Lambda_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \phi & 0 & I_{IO} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ h_{t-1} \\ \Delta_{t-1} \\ \Lambda_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ \eta_t \\ 0 \\ 0 \end{bmatrix} \quad (3.3.17)$$

$$I_{IO} = \begin{cases} 0 & \text{se } t \neq t_{IO} \\ 1 & \text{se } t = t_{IO} \end{cases} \quad \text{e } I_{AO} = \begin{cases} 0 & \text{se } t \neq t_{AO} \\ 1 & \text{se } t = t_{AO} \end{cases}$$

Pode-se notar que essa representação, como foi dito, é similar a representação dos modelos dinâmicos apresentados por West e Harrison (1997). Entretanto, no enfoque bayesiano do modelo como um modelo hierárquico, teria que ser considerado as distribuições a prioris. No capítulo 5 serão apresentadas algumas prioris adotadas no modelo de **VE**.

3.4 Conclusões

Apesar de parsimonioso, o modelo de **VE** é capaz de reproduzir os fatos estilizados das séries temporais financeiras. Assim como os modelos ARCH/GARCH, o modelo de **VE** reproduz os conglomerados de volatilidade e é uma família de modelos muito flexível e que podem ser facilmente generalizados sob vários aspectos, incluindo casos multivariados.

CAPÍTULO 4

Análise Clássica do Modelo de Volatilidade Estocástica

O capítulo a seguir é destinado a abordagem clássica dos modelos que podem ser representados na forma de espaço de estado, em particular o modelo de volatilidade estocástica que é objeto em estudo neste trabalho. Serão apresentados os métodos propostos por Durbin e Koopman, Shephard e Pitt, Tanizaki, Mariano e Geweke. Os quatro primeiros pesquisadores dedicaram a atenção em estimar os parâmetros do modelo através da estimativa direta da verossimilhança. Os três últimos propuseram algoritmos para estimar as densidades condicionais de filtragem e suavização, que podem ser utilizadas para estimar a verossimilhança e conseqüentemente estimar os parâmetros do modelo e também os estados.

4.1 Introdução

Nesta seção serão apresentados alguns métodos numéricos desenvolvidas para o tratamento de modelos de espaço de estados sob a abordagem frequentista. Na seção 4.2 será desenvolvido a metodologia proposta por Durbin e Koopman (1997a, 2000), onde eles utilizaram um modelo aproximado que torna possível a utilização do filtro de Kalman. O **FK** é utilizado para calcular uma verossimilhança aproximada e também para gerar amostras dos estados de uma densidade importante, que permite corrigir a função de verossimilhança aproximada. Depois de obtida amostras da densidade importante Durbin e Koopman utilizaram duas variáveis antitéticas e uma variável controle para tornar mais eficiente a simulação.

Na seção 4.3 será apresentado o método proposto por Shephard e Pitt (1997). Apesar desse trabalho ter um enfoque voltado para a análise bayesiana, eles consideram também a estimação da verossimilhança e dos parâmetros do modelo através da maximização da verossimilhança. A seção 4.4 apresenta os resultados do modelo aproximado utilizados por Durbin e Koopman e também por Shephard e Pitt. O apêndice B apresenta a análise frequentista proposta por Tanizaki, Mariano e Geweke, onde a idéia principal não está em estimar a verossimilhança apenas, e sim metodologias mais eficazes para amostrar os estados a fim de estimar as densidades condicionais de filtragem e suavização.

4.2 Correção da quasi-Verossimilhança Através da Amostragem dos Estados

Nesta seção será apresentada a abordagem proposta por Durbin e Koopman (1992, 1997a, 1997b, 2000) **DK** que é estimar a verossimilhança através da correção da quasi-verossimilhança. Tanto a verossimilhança quanto a correção são obtidas utilizando o **FK**. Nos dois trabalhos iniciais eles calcularam a verossimilhança com o objetivo de estimar os parâmetros, já no quarto trabalho foram apresentados resultados também sob a abordagem bayesiana. Na apresentação da abordagem nem sempre será destacado a origem do resultado. Considere o modelo na forma de espaço de estados da forma dada pelas equações (2.5.1) e (2.5.2), i.e., por

$$\begin{aligned} y_t &\sim p(y_t|\theta_t) & \theta_t &= Z_t\alpha_t + c_t \\ \alpha_t &= T_t\alpha_{t-1} + d_t + H_t\eta_t & \eta_t &\sim NID(0, I) \end{aligned}$$

Seja $\mathbf{y} = (y'_1, \dots, y'_n)'$, $\alpha = (\alpha'_1, \dots, \alpha'_n)'$ e denote o sinal por θ_t . Assim pode-se denotar as densidades como segue:

$p(\mathbf{y})$ - Densidade marginal de \mathbf{y} .

$p(\theta)$ - Densidade marginal de θ .

$p(\mathbf{y}, \theta)$ - Densidade conjunta de \mathbf{y} e θ .

$p(\mathbf{y}|\theta)$ - Densidade de \mathbf{y} dado θ .

onde todas essas densidades são para um dado conjunto de parâmetros Ψ . Assim a função de verossimilhança é dada da seguinte forma,

$$\begin{aligned}
L(\Psi) &= p(\mathbf{y}) \\
&= \int p(\mathbf{y}, \theta) d\theta \\
&= \int p(\mathbf{y}|\theta)p(\theta) d\theta
\end{aligned} \tag{4.2.1}$$

$L(\Psi)$ pode ser estimada através de simulação, escolhido N retiradas independentes de $\theta^{(1)}, \dots, \theta^{(N)}$ da distribuição com densidade $p(\theta)$ e tomando-se a média dos valores de $p(\mathbf{y}|\theta^{(i)})$ para $i = 1, \dots, N$. A eficiência deste procedimento pode ser melhorada através de técnicas tais como amostragem por importância, variável antitética e variável controle (ver apêndice A). Entretanto, este método esbarra na dificuldade de se amostrar da densidade $p(\theta)$

DK propuseram o método de amostragem por importância considerando uma densidade, a qual será chamada de densidade importante, $\tilde{p}(\theta|\mathbf{y})$ uma densidade condicional, de θ dado \mathbf{y} e para um certo Ψ , da qual podem ser retiradas amostras mais facilmente. Assim

$$\begin{aligned}
L(\Psi) &= \int p(\mathbf{y}|\theta)p(\theta) d\theta \\
&= \int p(\mathbf{y}|\theta) \frac{p(\theta)}{\tilde{p}(\theta|\mathbf{y})} \tilde{p}(\theta|\mathbf{y}) d\theta \\
&= \tilde{E} \left\{ p(\mathbf{y}|\theta) \frac{p(\theta)}{\tilde{p}(\theta|\mathbf{y})} \right\}
\end{aligned} \tag{4.2.2}$$

Fazendo-se N retiradas de $\theta^{(i)}$ da densidade $\tilde{p}(\theta|\mathbf{y})$ pode-se estimar $L(\Psi)$ por :

$$\hat{L}_1(\Psi) = \frac{1}{N} \sum_{i=1}^N p(\mathbf{y}|\theta^{(i)}) \frac{p(\theta^{(i)})}{\tilde{p}(\theta^{(i)}|\mathbf{y})} \tag{4.2.3}$$

Para aumentar a eficiência do método é necessário escolher $\tilde{p}(\theta|\mathbf{y})$ próximo de $p(\theta|\mathbf{y})$ dentro da classe de densidades que seja possível amostrar facilmente. Assim, **DK** sugeriram como densidade importante um modelo aproximado onde considera-se que as observações são geradas por uma equação linear, com maior flexibilidade para a distribuição dos erros ε_t da equação (2.4.1), i.e., $\varepsilon_t \sim N(\mu_t, A_t)$ para facilitar a amostragem.

Desta forma, lembrando que a perturbação da equação de transição é Gaussiana, o modelo aproximado passa a estar na **FEEG** e a verossimilhança associada a esse modelo, $L_g(\Psi)$ pode ser calculada utilizando o filtro de Kalman. Após obtermos $L_g(\Psi)$ pode-se mostrar que a verdadeira verossimilhança pode ser obtida através de uma correção, usando amostragem por importância.

Considere $p(\cdot)$ e $g(\cdot)$ as densidades para o modelo original e modelo Gaussiano aproximado respectivamente. A verossimilhança verdadeira para o modelo não Gaussiano é dada por

$$L(\Psi) = \int p(y, \theta) d\theta = \int p(y|\theta)p(\theta) d\theta \quad (4.2.4)$$

A verossimilhança para o modelo Gaussiano aproximado será considerada como

$$L_g(\Psi) = \frac{g(\mathbf{y}, \theta)}{g(\theta|\mathbf{y})} = \frac{g(\mathbf{y}|\theta)p(\theta)}{g(\theta|\mathbf{y})} \quad (4.2.5)$$

logo temos que

$$p(\theta) = L_g(\Psi) \frac{g(\theta|y)}{g(\mathbf{y}|\theta)} \quad (4.2.6)$$

então fazendo-se as devidas substituições e sabendo-se que $p(\theta)$ é o mesmo para ambos os modelos, a verossimilhança é apresentada como segue

$$\begin{aligned}
L(\Psi) &= \int p(y|\theta) \frac{L_g(\Psi)g(\theta|y)}{g(y|\theta)} d\theta \\
&= L_g(\Psi) \int \frac{p(y|\theta)}{g(y|\theta)} g(\theta|y) d\theta \\
&= L_g(\Psi) E_g \left\{ \frac{p(y|\theta)}{g(y|\theta)} \right\}
\end{aligned} \tag{4.2.7}$$

É fácil observar a simplicidade da função em relação a dada pela equação (4.2.2). A vantagem essencial é que não é necessário simular para estimar toda a verossimilhança e sim apenas o desvio da verossimilhança verdadeira em relação a verossimilhança considerando densidades Gaussianas.

Durbin e Koopman (1992, 1997a, 1997b) e Shephard e Pitt (1997) sugerem que $g(\theta|y)$ seja a densidade importante para a simulação. Assim amostra-se desta densidade utilizando-se o algoritmo de suavização proposto por de Jong e Shephard (1995).

Como comentado anteriormente a eficiência do método depende da escolha do modelo linear Gaussiano apropriado e a forma eficiente de estimar a correção. O primeiro problema será destacado na subseção 4.2.1 enquanto na subseção 4.2.2 serão discutidas formas eficientes de estimar a correção por simulação através de variáveis antitéticas e variáveis controle. Finalmente a subseção 4.2.3 apresenta resumidamente os passos necessários para estimar a verossimilhança através deste método.

4.2.1 A Seleção do Modelo Aproximado

A escolha de um modelo aproximado está intimamente relacionado a escolha da média μ_t e do fator de escala A_t dos distúrbios ε_t . Como a densidade importante deve ser escolhida tão próximo quanto possível à densidade $p(\theta|y)$, **DK** propuseram escolher uma densidade tal que a moda e a curvatura fossem aproximadamente iguais.

A partir dessa idéia **DK** tomaram $\hat{\theta} = E_g(\theta)$, o valor de θ obtido quando aplica-se o filtro de Kalman e o algoritmo de suavização de de Jong e Shephard (1995), considerando o modelo Gaussiano correto. Assim, eles verificaram que como os θ_t , gerado pelo algoritmo de suavização, são normalmente distribuídos em torno de $\hat{\theta}_t$, tomaram a escolha de μ_t e A_t de forma que as densidades dos modelos fossem bem próximas em torno de $\hat{\theta}$. **DK** sugeriram a nulidade da primeira e segunda derivadas do log-quociente entre a densidade verdadeira e o modelo aproximado de maneira que tem-se,

$$l(\theta) = \log \frac{p(\theta|y)}{g(\theta|y)} = \log \frac{p(y|\theta)p(\theta)}{g(y|\theta)g(\theta)} \propto \log p(y|\theta) - \log g(y|\theta) \quad (4.2.8)$$

a sugestão de **DK** é escolher μ_t e A_t tal que $\frac{\partial l(\theta)}{\partial \theta} = 0$ e $\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} = 0$ em $\theta = \hat{\theta}$. Dado que

$$l(\theta) = \sum_{t=1}^n \{\log p(y_t|\theta_t) - \log g(y_t|\theta_t)\}, \quad (4.2.9)$$

e sabendo-se que $g(y_t|\theta_t) \sim N(\mu_t + \theta_t, A_t)$, tem-se

$$\log g(y_t|\theta_t) = \text{cte} - \frac{1}{2}(y_t - \theta_t - \mu_t)'A_t^{-1}(y_t - \theta_t - \mu_t)$$

então as seguintes condições são suficientes para que as duas primeiras derivadas de $l(\theta)$ seja igual a zero em $\hat{\theta}$.

$$\left. \frac{\partial \log p(y_t|\theta_t)}{\partial \theta_t} \right|_{\theta_t=\hat{\theta}_t} - A_t^{-1}(y_t - \hat{\theta}_t - \mu_t) = 0 \quad (4.2.10)$$

$$\left. \frac{\partial^2 \log p(y_t|\theta_t)}{\partial \theta \partial \theta'} \right|_{\theta_t=\hat{\theta}_t} + A_t^{-1} = 0 \quad (4.2.11)$$

Essas equações são resolvidas iterativamente iniciando-se com $\mu_t = 0$ e A_t qualquer. Para cada novo par μ_t e A_t obtém-se um novo vetor de $\hat{\theta}_t$ utilizando-se o algoritmo de filtro de Kalman e suavizador de de Jong e Shephard (1995). Este novo valor é utilizado nas equações (4.2.10) e (4.2.11) para se obter novos valores de A_t e μ_t .

A restrição necessária para que (4.2.10) e (4.2.11) tenham solução é que A_t seja positiva definida, o que é verdadeiro para o caso do modelo de VE.

4.2.2 Estimação da Logverossimilhança e Correção de Vício

Considere uma amostra $\theta^{(i)}$ obtido pelo algoritmo de suavizamento, e seja $w_i = w(\theta^{(i)})$, para $i = 1, \dots, N$ onde temos

$$w(\theta) = \frac{p(y|\theta)}{g(y|\theta)} \quad (4.2.12)$$

A estimativa de $L(\psi)$ por Monte Carlo baseada em amostragem por importância será dada como segue,

$$\hat{L}_2(\Psi) = L_g(\Psi)\bar{w} \quad (4.2.13)$$

onde $\bar{w} = N^{-1} \sum_{i=1}^N w_i$. Portanto, pode-se verificar facilmente que $\hat{L}_2(\Psi)$ é uma estimativa não viciada para $L(\Psi)$. Como é mais conveniente trabalhar com $\log L(\Psi)$ temos o seguinte resultado:

$$\log \hat{L}_2(\Psi) = \log L_g(\Psi) + \log \bar{w} \quad (4.2.14)$$

mas, pela desigualdade de Jensen, $E \log \bar{w} \leq \log \mu_w$ onde $\mu_w = E_g(w_i)$. Logo, (4.2.13) é viciada para $\log L(\Psi)$. A fim de corrigir o vício tem-se que:

$$\log \bar{w} - \log \mu_w = \log \left(1 + \frac{\bar{w} - \mu_w}{\mu_w} \right) = \frac{\bar{w} - \mu_w}{\mu_w} - \frac{1}{2} \left(\frac{\bar{w} - \mu_w}{\mu_w} \right)^2 + O_p(N^{-\frac{3}{2}}) \quad (4.2.15)$$

então

$$E_g(\log \bar{w}) = \log \mu_w + \frac{\sigma_w^2}{2N\mu_w^2} + O(N^{-\frac{3}{2}}) \quad (4.2.16)$$

onde $\sigma_w^2 = \text{var}(w_i)$. Substituindo μ_w por \bar{w} e σ_w^2 por s_w^2 onde $s_w^2 = \frac{\sum_{i=1}^N (w_i - \bar{w})^2}{N-1}$ temos que um estimador não viciado aproximado para $\log L(\Psi)$ é dado por

$$\log \hat{L}_3(\Psi) = \log L_g(\Psi) + \log \bar{w} + \frac{s_w^2}{2N\bar{w}^2} \quad (4.2.17)$$

Variáveis Antitéticas

Uma breve explicação sobre variáveis antitéticas é dado no apêndice A. Para um estudo mais detalhado ver Ripley (1987, p.122-123). Suponha o sinal θ amostrado da densidade $g(\theta|y)$, usando o algoritmo de suavização com μ_t e A_t escolhidos como na seção anterior. A primeira variável antitética, denominada por **DK** de variável antitética balanceada por locação, é dada por

$$\tilde{\theta}_t = 2\hat{\theta}_t - \theta_t, \quad t = 1, \dots, n$$

Portanto, o ganho de eficiência do uso desta variável antitética é a obtenção de 2 amostras $\theta^{(i)}$ e $\tilde{\theta}^{(i)}$ negativamente correlacionadas, o que induz uma correlação negativa entre w_i e \tilde{w}_i .

Para a segunda variável antitética assuma y_t e θ_t univariado. Seja θ retirado da densidade $g(\theta|y)$ obtido pelo algoritmo de suavização, então θ é dada como uma transformação linear não singular de u da forma $\theta = \hat{\theta} + Bu$, onde a suavização é derivada do vetor u que é retirado de uma $N(0, I_n)$ e $\text{var}(\theta) = V = BB'$.

Defina

$$c = (\theta - \hat{\theta})'V^{-1}(\theta - \hat{\theta}) = u'u \sim \chi_n^2 \quad (4.2.18)$$

Para um dado c , seja $q = Pr(\chi_n^2 < c) = F(c)$, e seja $\bar{c} = F^{-1}(1 - q)$, portanto como c varia, $\bar{c} \sim \chi_{(n)}^2$. Segue que se definirmos $\bar{\theta} = \hat{\theta} + \left(\frac{\bar{c}}{c}\right)^{\frac{1}{2}}(\theta - \hat{\theta})$ então $\bar{\theta}$ tem a mesma distribuição que θ . Além disso, quando $|\theta_t - \hat{\theta}_t|$ é grande, $|\bar{\theta}_t - \hat{\theta}_t|$ é pequeno e vice versa. Os pares de amostras θ e $\bar{\theta}$ com essa propriedade são chamados de balanceado por escala.

DK verificaram que a redução da variância da estimativa de $\log L(\psi)$ com o uso de variáveis antitéticas balanceadas por escala é sempre maior que aquelas com o uso de variável antitética balanceada por locação. As variáveis serão utilizadas conjuntamente, portanto teremos 4 valores para $w(\theta)$ para cada valor obtido pelo algoritmo de suavização. Seja $\check{\theta}_{(i)}$ a variável balanceada por escala a qual é antitética para $\tilde{\theta}^{(i)}$ e seja $\check{w}_{(i)} = w(\check{\theta}^{(i)})$ então

$$w_i^* = \frac{1}{4}(w_i + \tilde{w}_i + \bar{w}_i + \check{w}_i) \quad (4.2.19)$$

e estima-se $\log L(\Psi)$ por $L_4(\Psi)$ a qual é dada por

$$\log \hat{L}(\psi) = \log L_g(\psi) + \log \bar{w}^* + \frac{s_{w^*}^2}{2N(\bar{w}^*)^2} \quad (4.2.20)$$

Variáveis Controle Baseadas em Série de Taylor

Durbin e Koopman (1997a) propuseram também a utilização de variáveis controle para melhorar a eficiência nas simulações. Uma explicação sucinta da teoria de variáveis controle encontra-se no apêndice A. Como variável controle, **DK** utilizaram a expansão de Taylor do quociente $w(\theta)$ em torno de $\hat{\theta}$. Considerando a notação:

$$l = l(\theta), \quad l'_t = \frac{\partial l(\theta)}{\partial \theta_t}, \quad l''_t = \frac{\partial^2 l(\theta)}{\partial \theta_t^2}, \dots \quad (4.2.21)$$

$$w = w(\theta), \quad w'_t = \frac{\partial w(\theta)}{\partial \theta_t}, \quad w''_t = \frac{\partial^2 w(\theta)}{\partial \theta_t^2}, \dots \quad (4.2.22)$$

As derivadas de w serão dadas por:

$$\begin{aligned} w'_t &= w l'_t, & w''_t &= w \{l''_t + (l'_t)^2\}, & w'''_t &= w \{l'''_t + 3l'_t l''_t + (l'_t)^3\} \\ w''''_t &= w \{l''''_t + 4l'_t l'''_t + 3(l''_t)^2 + 6l''_t (l'_t)^2 + (l'_t)^4\} \end{aligned} \quad (4.2.23)$$

Como μ_t e A_t foram tomados de forma que a primeira e segunda derivadas fossem zero, l'_t e l''_t são iguais a zero em $\theta = \hat{\theta}$ para $t = 1, \dots, n$.

Para a facilidade de notação, denote as funções definidas em $\hat{\theta}$ por, $w|_{\hat{\theta}} = \hat{w}$, $l'''_t|_{\hat{\theta}} = \hat{l}'''_t$ e $l''''_t|_{\hat{\theta}} = \hat{l}''''_t$. Assim a expansão em série de Taylor é da seguinte forma:

$$w(\theta) = \hat{w} \left\{ 1 + \frac{1}{6} \sum_{t=1}^n \hat{l}'''_t (\theta_t - \hat{\theta}_t)^3 + \frac{1}{24} \sum_{t=1}^n \hat{l}''''_t (\theta_t - \hat{\theta}_t)^4 + \dots \right\} \quad (4.2.24)$$

DK consideraram $\hat{w}\bar{a}$ como a variável controle utilizada para estimar $\log L(\Psi)$, onde $\bar{a} = \frac{\sum_{i=1}^N a_i}{N}$ e a_i dado como segue.

$$a_i = \frac{1}{6} \sum_{t=1}^n \hat{l}_t''' (\theta_t^{(i)} - \hat{\theta}_t)^3 + \frac{1}{24} \sum_{t=1}^n \hat{l}_t'''' (\theta_t^{(i)} - \hat{\theta}_t)^4 \quad (4.2.25)$$

O primeiro termo some por causa do resultado da variável antitética balanceada por locação. Seja $\nu_t = \text{var}(\epsilon_t|y_t)$ para $t = 1, \dots, n$ que pode ser obtido pelo algoritmo de suavização. Dado que $\theta_t = y_t - \epsilon_t$, $\text{var}(\theta_t|y_t) = \nu_t$ tem-se:

$$E_g(a_i) = \frac{1}{8} \sum_{t=1}^n \hat{l}_t'''' \nu_t^2 \quad (4.2.26)$$

O valor de a_i é obtido calculando-se a média dos 4 valores de $\frac{1}{24} \sum_{t=1}^n \hat{l}_t'''' (\theta_t^{(i)} - \hat{\theta}_t)$ com $\theta_t^{(\cdot)} = \theta_t^{(\cdot)}, \tilde{\theta}_t^{(\cdot)}, \bar{\theta}_t^{(\cdot)}, \check{\theta}_t^{(\cdot)}$

$$w_i^\dagger = w_i - \hat{w} \left(a_i - \frac{1}{8} \sum_{t=1}^n \hat{l}_t'''' \nu_t^2 \right) \quad (4.2.27)$$

Portanto a estimativa de $\log L(\Psi)$ usando a variável controle a_i é $\log \hat{L}_5(\Psi)$ dada por (4.2.17), com w_i^\dagger no lugar de w_i . De (4.2.24) e (4.2.25) tem-se que

$$E_g\{w(\theta)\} \simeq \hat{w} \{1 + E_g(a_i)\} \quad (4.2.28)$$

Portanto obtém-se a log-verossimilhança aproximada dada por:

$$\log \hat{L}_0 = \log L_g + \log \hat{w} + \log \left(1 + \frac{1}{8} \sum_{t=1}^n \hat{l}_t'''' \nu_t^2 \right) \quad (4.2.29)$$

4.2.3 Algoritmo de Estimação da Verossimilhança

O algoritmo para estimação, que utiliza as variáveis controle e antitéticas para estimar a verossimilhança pode ser implementado através dos seguintes passos.

- ① Identifique a densidade gaussiana $g(y|\theta)$, para aproximar $p(y|\theta)$ baseada no fato de que $y_t \sim N(\theta_t + \mu_t, A_t)$, onde μ_t e A_t são soluções iterativas das equações (4.2.10) e (4.2.11).
- ② Calcule $\log L_g(\psi)$ usando o filtro de Kalman.
- ③ Tome $g(y|\theta)$ como a densidade importante e amostre $\theta^{(i)}$ de $g(\theta|y)$ usando o algoritmo de suavização de de Jong e Shephard (1995) para $i = 1, \dots, N$.
- ④ Para cada i calcule w_i^\dagger por (4.2.27) e calcule $\log \hat{L}_5(\Psi)$ pela equação (4.2.17) com w_i^\dagger no lugar de w_i .

4.3 Correção da Quasi-Verossimilhança Através da Amostragem das Perturbações

Shephard e Pitt (1997) (**SP**) utilizaram a abordagem bayesiana para os modelos de séries temporais não Gaussianos. Apesar do trabalho ser quase totalmente voltado para a estimação bayesiana das distribuições a posterioris, **SP** propuseram o uso de amostragem por importância para estimar a função de verossimilhança.

A proposta de **SP** está estritamente ligada a proposta de Durbin e Koopman (1997a). Assim como Durbin e Koopman (2000), **SP** amostram dos distúrbios η a fim de contornar o problema de amostrar uma densidade degenerada e minimizar problemas de grandes dimensões do vetor de observação. Portanto, além de não utilizarem variável controle, eles utilizaram apenas uma variável antitética e sugeriram um

distribuição diferente na amostragem por importância. O método proposto também é eficiente e utiliza o fato de existir uma relação determinística entre os sinais θ e η . Considerando-se a densidade importante por $\tilde{p}(\eta|\mathbf{y}, \Psi)$ tem-se

$$\begin{aligned} L(\Psi) &= p(\mathbf{y}) \\ &= \int \frac{p(\mathbf{y}|\eta)p(\eta)}{\tilde{p}(\eta|\mathbf{y})} \tilde{p}(\eta|\mathbf{y}) d\eta \end{aligned}$$

Logo, pode-se estimar essa integral através da esperança em relação a densidade importante e obtendo-se

$$\hat{L}(\Psi) = \hat{p}(\mathbf{y})_M = \frac{1}{M} \sum_{j=1}^M \frac{p(\mathbf{y}|\eta^j)p(\eta^j)}{\tilde{p}(\eta^j|\mathbf{y})} \quad (4.3.1)$$

Assim, como na seção 4.2, este estimador é não viciado para a verossimilhança. No entanto, como o estimador da log-verossimilhança é viciado, **SP** fazem a correção do vício de forma análoga a **DK** e obtém o seguinte estimador para a log-verossimilhança

$$\log \hat{p}(y)_M + \frac{1}{2M} \frac{1}{M-1} \sum_{j=1}^M \{p(y)^j - p(y)_M\}^2 \quad (4.3.2)$$

Outra divergência entre os métodos que vale ser ressaltada é a utilização de apenas uma variável antitética balanceada por locação para a redução de variância.

4.4 Modelo Aproximado para o Modelo de Volatilidade Estocástica

Os trabalhos de Durbin e Koopman (1997a, 2000), Shephard e Pitt (1997) consideram a mesma forma de aproximação para o modelo. Como nesta dissertação o enfoque é

dado ao caso particular do modelo de volatilidade estocástica esta seção apresenta a forma de se retirar amostra da densidade importante. No modelo de **VE** a densidade das observações é dada por:

$$\log f(y_t|\theta_t) = -\frac{1}{2} \log 2\pi\sigma^2 - \frac{1}{2}\theta_t - \frac{y_t^2}{2\sigma^2} \exp(-\theta_t). \quad (4.4.1)$$

Assim, a linearização para o modelo aproximado segue como na seção 4.2.1 com

$$\left. \frac{\partial \log f(y_t|\theta_t)}{\partial \theta_t} \right|_{\theta_t=\hat{\theta}_t} - A_t^{-1}(y_t - \hat{\theta}_t - \mu_t) = 0, \quad (4.4.2)$$

$$\left. \frac{\partial^2 \log f(y_t|\theta_t)}{\partial \theta_t \partial \theta_t'} \right|_{\theta_t=\hat{\theta}_t} + A_t^{-1} = 0, \quad (4.4.3)$$

De (4.2.10) e (4.2.11) obtém-se

$$A_t = 2\sigma^2 \frac{\exp(\hat{\theta}_t)}{y_t^2} \quad (4.4.4)$$

$$\tilde{y}_t = \hat{\theta}_t - \frac{1}{2}A_t + 1 \quad (4.4.5)$$

para o qual A_t é sempre positiva. A solução iterativa pode ser obtida, aplicando-se o filtro de Kalman suavizado considerando:

$$\tilde{y}_t = \theta_t + \varepsilon_t \quad \varepsilon_t \sim N(0, A_t) \quad \tilde{y}_t = y_t - \mu_t \quad (4.4.6)$$

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t \quad \eta_t \sim N(0, Q_t) \quad \theta_t = Z_t \alpha_t \quad (4.4.7)$$

onde o algoritmo de suavizamento nos produz um novo valor para θ . O processo repete-se até a convergência de $\hat{\theta}$. Após a verossimilhança estimada pode-se encontrar os estimadores de máxima verossimilhança do modelo utilizando-se um algoritmo de otimização apropriado.

CAPÍTULO 5

Análise Bayesiana do Modelo de Volatilidade Estocástica

Neste capítulo será apresentada a análise do modelo de volatilidade estocástica sob a abordagem bayesiana são considerados os trabalhos de Jacquier *et al.*(1994), Shephard e Pitt (1997) e Kim *et al.*(1998), sendo o artigo de Jacquier *et al.*(1994) responsável pela difusão dessa abordagem ao modelo de **VE**.

5.1 Introdução

Considere o modelo de **VE** dado como na equação (3.3.5), i.e., da forma

$$\begin{aligned} y_t &= h_t^{1/2} \varepsilon_t & t \geq 1, \\ \log h_t &= \alpha + \phi \log h_{t-1} + \sigma_\eta \eta_t, \end{aligned} \tag{5.1.1}$$

e com distribuição a priori para os parâmetros $(\alpha, \phi, \sigma_\eta^2)$, e seja $\mathbf{y} = y_1, \dots, y_T$ o vetor t -dimensional de retornos, $\mathbf{h} = (h_1, \dots, h_T)$ vetor de volatilidades e $\theta = (\alpha, \phi, \sigma_\eta^2)$ o vetor de parâmetros da volatilidade. Como pode ser observado, o modelo de **VE** é um exemplo de modelo hierárquico. Os retornos são determinados pelas volatilidades que são determinadas pelos parâmetros da volatilidade.

Jacquier *et al.* (1994) propuseram a análise do modelo de **VE** com enfoque bayesiano, explorando esta estrutura hierárquica do modelo. A partir deste trabalho, muitas modificações foram propostas, dentre essas Shephard e Pitt (1997) consideraram amostragem por blocos na tentativa de diminuir os efeitos de dependência. Esses pesquisadores também estimaram os parâmetros considerando um conjunto de distribuições a priori diferente do proposto por Jacquier *et al.*

Kim *et al.* (1998) utilizaram mistura de distribuições normais e, condicionada a distribuição da mistura, utilizaram o **FK** para amostrar todo o vetor de volatilidades de uma única vez. Além desse aperfeiçoamento eles propuseram o uso de amostragem conjunta dos parâmetros para reduzir problemas de dependência.

Na seção 5.2 serão desenvolvidas as diversas densidades condicionais que serão utilizadas nas demais seções. Na subseção 5.2.1 será apresentado o algoritmo inicial utilizando amostragem simples. Na subseção 5.2.2 será feito o desenvolvimento da aproximação necessária para a amostragem da densidade multivariada e apresentado

o algoritmo de amostragem por blocos. Finalizando a seção de amostragem das volatilidades, é apresentada a metodologia utilizada por Kim *et al.* onde é utilizada a amostragem conjunta. Na seção 5.3 serão apresentadas as distribuições a priori e posterioris para os parâmetros do modelo de volatilidade estocástica.

5.2 Amostragem da Variável Latente

Jacquier *et al.*(1994) propuseram um método para estimar a densidade conjunta de \mathbf{h}, θ dado \mathbf{y} . Uma vez obtida essa densidade pode-se encontrar as densidades marginais de \mathbf{h} e θ dado \mathbf{y} . A densidade conjunta de $\mathbf{h}, \theta|\mathbf{y}$ é dada da seguinte forma:

$$p(\mathbf{h}, \theta|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{h}, \theta)p(\mathbf{h}|\theta)p(\theta)}{p(\mathbf{y})} \quad (5.2.1)$$

onde $p(\theta)$ é uma densidade a priori para o vetor de parâmetros, θ e a densidade conjunta $p(\mathbf{h}, \theta|\mathbf{y})$ a densidade à posteriori. A densidade $p(\mathbf{y})$ é dada como segue.

$$p(\mathbf{y}) = \iint p(\mathbf{y}|\mathbf{h}, \theta)p(\mathbf{h}|\theta)p(\theta) d\theta d\mathbf{h} \quad (5.2.2)$$

A marginal de $\mathbf{h}|\mathbf{y}$ é obtida integrando a conjunta em relação a θ . Por outro lado, a marginal de $\theta|\mathbf{y}$ pode ser encontrada integrando a posteriori conjunta em relação a \mathbf{h} da seguinte forma

$$p(\theta|\mathbf{y}) = \int p(\mathbf{h}, \theta|\mathbf{y}) d\mathbf{h} \quad (5.2.3)$$

Como não existe uma forma fechada para $p(\mathbf{h}, \theta|\mathbf{y})$ Jacquier *et al.*(1994) utilizaram o método de aumentar os dados observados \mathbf{y} pelas variáveis latentes \mathbf{h} , como proposto

anteriormente por Carlin *et al.* (1992) para o caso geral de modelos FEENG. A relação entre a densidade aumentada $p(\theta|\mathbf{h}, \mathbf{y})$ e a conjunta $p(\mathbf{h}, \theta|\mathbf{y})$ é dada por

$$p(\mathbf{h}, \theta|\mathbf{y}) = p(\theta|\mathbf{h}, \mathbf{y})p(\mathbf{h}|\mathbf{y}) \quad (5.2.4)$$

como o interesse não é a densidade aumentada e sim em $p(\theta|\mathbf{y})$. Utiliza-se a identidade acima, e encontra-se a densidade de θ condicionado a \mathbf{y} como segue

$$p(\theta|\mathbf{y}) = \int p(\mathbf{h}, \theta|\mathbf{y}) d\mathbf{h} = \int p(\theta|\mathbf{h}, \mathbf{y})p(\mathbf{h}|\mathbf{y}) d\mathbf{h} \quad (5.2.5)$$

Agora, conhecendo-se $p(\mathbf{h}|\mathbf{y})$ esta integral pode ser calculada usando técnicas de simulação de Monte Carlo (ver apêndice A). Analogamente, tomando-se

$$p(\mathbf{h}, \theta|\mathbf{y}) = p(\mathbf{h}|\theta, \mathbf{y})p(\theta|\mathbf{y}) \quad (5.2.6)$$

A distribuição marginal de $\mathbf{h}|\mathbf{y}$ pode ser obtida através de

$$p(\mathbf{h}|\mathbf{y}) = \int p(\mathbf{h}, \theta|\mathbf{y}) d\theta = \int p(\mathbf{h}|\theta, \mathbf{y})p(\theta|\mathbf{y}) d\theta \quad (5.2.7)$$

Portanto, é necessário saber amostrar de $\theta|\mathbf{y}$ para estimar $p(\mathbf{h}|\mathbf{y})$. Logo tem-se um problema recursivo como mostrado na Figura 5.1, onde é necessário a densidade de $\mathbf{h}|\mathbf{y}$ para estimar $p(\theta|\mathbf{y})$ e vice-versa.

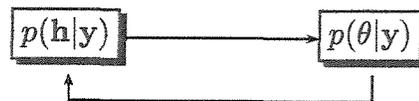


Figura 5.1: Diagrama de Recursão para Estimar $p(\mathbf{h}|\mathbf{y})$

5.2.1 Amostragem Simples da Variável Latente

O problema do modelo de volatilidade estocástica é que não conseguimos obter a forma analítica da densidade $p(\mathbf{y}|\theta) = \int p(\mathbf{y}|\mathbf{h}, \theta)p(\mathbf{h}|\theta)d\mathbf{h}$, isso impede a análise direta da posteriori $p(\theta|\mathbf{y})$ através de métodos de MCMC. Pode-se contornar esse problema, trabalhando-se com a densidade $p(\theta, \mathbf{h}|\mathbf{y})$ e através dos métodos tradicionais de MCMC pode-se amostrar da posteriori conjunta e calcular quantidades de interesse sem obter a verossimilhança.

O algoritmo inicial pode ser implementado utilizando os seguintes passos no procedimento recursivo.

- ① Dado o vetor inicial $\mathbf{h}^{(i)}$ e $\theta^{(i)}$, amostre $\mathbf{h}^{(i+1)}$ da densidade $p(\mathbf{h}|\theta^{(i)}, \mathbf{y})$.
- ② Dado $\mathbf{h}^{(i+1)}$, amostre $\theta^{(i+1)}$ de $p(\theta|\mathbf{h}^{(i+1)}, \mathbf{y})$

Para amostrar das densidades acima, é necessário ter a forma analítica de $p(\mathbf{h}|\theta^{(i)}, \mathbf{y})$ e $p(\theta|\mathbf{h}^{(i+1)}, \mathbf{y})$ e que seja fácil amostrar dessas densidades. Em princípio, considere $p(\theta|\mathbf{h}^{(i+1)}, \mathbf{y})$ representada como

$$p(\theta|\mathbf{h}, \mathbf{y}) = \frac{p(\theta, \mathbf{h}, \mathbf{y})}{\int p(\theta, \mathbf{h}, \mathbf{y}) d\theta} \quad (5.2.8)$$

onde

$$p(\theta, \mathbf{h}, \mathbf{y}) = p(\mathbf{y}|\theta, \mathbf{h})p(\mathbf{h}|\theta)p(\theta) \quad (5.2.9)$$

A densidade de $\mathbf{y}|\mathbf{h}$ pode ser encontrada diretamente, pois dado que $\varepsilon_t \sim N(0, 1)$ e $\eta_t \sim N(0, \sigma_\eta^2)$, segue que $y_t|h_t \sim N(0, h_t)$. Portanto,

$$\begin{aligned}
p(\mathbf{y}|\mathbf{h}) &= \prod_{t=1}^T p(y_t|h_t) \\
&= \prod_{t=1}^T \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{y_t^2}{2h_t}\right)
\end{aligned} \tag{5.2.10}$$

Dado que $\eta_t \sim N(0, \sigma_\eta^2)$ então, $E(\log h_t|\theta, \Psi_t) = \alpha + \phi \log h_{t-1}$ e $V(\log h_t|\theta, \Psi_t) = \sigma_\eta^2$.
Suponha $\Psi_t = (h_{t-1}, h_{t-1}, \dots)$, então

$$\log h_t|\theta, \Psi_t \sim N(\alpha + \phi \log h_{t-1}, \sigma_\eta^2)$$

Então

$$p(\log h_t|\theta, \Psi_t) = \frac{1}{\sqrt{2\pi\sigma_\eta}} \exp\left(-\frac{(\log h_t - \alpha - \phi \log h_{t-1})^2}{2\sigma_\eta^2}\right) \tag{5.2.11}$$

Assim, a densidade $p(h_t|\theta, \Psi_t)$ é dada por,

$$p(h_t|\theta, \Psi_t) = \left(\frac{\partial \log h_t}{\partial h_t}\right) p(\log h_t|\theta, \Psi_t) = \frac{1}{h_t} p(\log h_t|\theta, \Psi_t) \tag{5.2.12}$$

logo, h_t tem densidade lognormal e a conjunta é dada como segue

$$p(\mathbf{h}|\theta) = p(h_1|\theta) \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_\eta}} \exp\left(-\frac{(\log h_t - \alpha - \phi \log h_{t-1})^2}{2\sigma_\eta^2}\right) \tag{5.2.13}$$

como h_0 não é conhecido e $\log h_1|\theta \sim N\left(\frac{\alpha}{1-\phi}, \frac{\sigma_\eta^2}{1-\phi^2}\right)$, a densidade de $h_1|\theta$ será a distribuição não condicional dada por

$$p(h_1|\theta) = \frac{1}{h_1} p(\log h_1|\theta) = \frac{1}{h_1} \frac{\sqrt{1-\phi^2}}{\sqrt{2\pi}\sigma_\eta} \exp\left\{-\frac{1-\phi^2}{2\sigma_\eta^2} \left(\log h_1 - \frac{\alpha}{1-\phi}\right)^2\right\} \quad (5.2.14)$$

e

$$p(h_t|\theta, \Psi_t) = \frac{1}{h_t} p(\log h_t|\theta)$$

Deve-se agora obter uma forma de amostrar de $p(\mathbf{h}|\theta, \mathbf{y})$, pois não é possível amostrar diretamente, dado que é uma densidade multivariada. Portanto, utiliza-se o amostrador de Gibbs (ver apêndice A) para obter amostras da densidade conjunta a partir de suas condicionais univariadas. Isto pode ser conseguido através do seguinte algoritmo:

① Dado $\theta^{(i)}$, $\mathbf{h}^{(i+1)}$ é amostrado como segue:

② Amostre $h_1^{(i+1)}$ de $h_1|\mathbf{h}_{\setminus 1}^{(i)}, \mathbf{y}\theta^{(i)}$

Amostre $h_2^{(i+1)}$ de $h_2|\mathbf{h}_{\setminus 2}^{(i)}, \mathbf{y}\theta^{(i)}$

⋮

Amostre $h_t^{(i+1)}$ de $h_t|\mathbf{h}_{\setminus t}^{(i)}, \mathbf{y}\theta^{(i)}$

③ Dado $\mathbf{h}^{(i+1)}$, amostre $\theta^{(i+1)}$ de $\theta|\mathbf{h}^{(i+1)}, \mathbf{y}$

Para utilizar o algoritmo de amostrador de Gibbs, é necessária a forma analítica de $p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y})$. Portanto, considere a densidade condicional representada como.

$$p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y}) = \frac{p(h_t, \mathbf{h}_{\setminus t}, \theta, \mathbf{y})}{\int p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y}) dh_t} \quad (5.2.15)$$

A integral em relação a h_t torna o denominador constante relativo a variável h_t . O denominador é conhecido na literatura como constante normalisadora, não sendo necessária para determinar a densidade, podendo ser desconsiderada. A forma funcional da densidade é obtida da seguinte maneira,

$$\begin{aligned} p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y}) &= \frac{p(\mathbf{h}, \theta, \mathbf{y})}{\int p(\mathbf{h}, \theta, \mathbf{y}) dh_t} \\ &= \frac{p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\theta)p(\theta)}{\int p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\theta)p(\theta) dh_t} \\ &\propto p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\theta)p(\theta) \\ &\propto p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\theta) \end{aligned}$$

A priori para os parâmetros também pode ser omitida por motivo análogo à omissão da integral. De $p(\mathbf{y}|\mathbf{h})$ e $p(\mathbf{h}|\theta)$, é necessário somente os termos onde existe dependência em h_t . Assim, pode-se simplificar a equação da maneira como segue

$$\begin{aligned} p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y}) &\propto p(\mathbf{y}|\mathbf{h})p(\mathbf{h}|\theta) \\ &\propto p(y_t|h_t)p(h_t|h_{t-1}, \theta)p(h_{t+1}|h_t, \theta) \\ &\propto \frac{1}{\sqrt{h_t}} \exp\left(-\frac{y_t^2}{2h_t}\right) \frac{1}{h_t} \exp\left(-\frac{(\log h_t - \alpha - \phi \log h_{t-1})^2}{2\sigma_\eta^2}\right) \exp\left(-\frac{(\log h_{t+1} - \alpha - \phi \log h_t)^2}{2\sigma_\eta^2}\right) \\ &\propto \frac{1}{h_t^{1.5}} \exp\left(-\frac{y_t^2}{2h_t}\right) \underbrace{\exp\left(-\frac{1}{2\sigma_\eta^2}(\log h_t - \alpha - \phi \log h_{t-1})^2 + (\log h_{t+1} - \alpha - \phi \log h_t)^2\right)}_{\textcircled{2}} \end{aligned}$$

Temos que,

$$\begin{aligned}
\textcircled{2} &= -\frac{1}{2\sigma_\eta^2}(\log h_t - \alpha - \phi \log h_{t-1})^2 + (\log h_{t+1} - \alpha - \phi \log h_t)^2 \\
&= -\frac{1}{2\sigma_\eta^2}((\log h_t)^2 + (\alpha + \phi \log h_{t-1})^2 - 2(\log h_t)(\alpha + \phi \log h_{t-1}) + \phi^2(\log h_t)^2 \\
&\quad + (\log h_{t+1} - \alpha)^2 - 2(\log h_t)(\log h_{t+1} - \alpha)) \\
&= -\frac{1}{2\sigma_\eta^2}((1 + \phi^2)(\log h_t)^2 - 2(\log h_t)(\phi \log h_{t+1} + \phi \log h_{t+1} + \alpha - \alpha\phi) + \text{cte}) \\
&= -\frac{1 + \phi^2}{2\sigma_\eta^2}((\log h_t)^2 - 2(\log h_t)\gamma_t + \text{cte}) \\
&= -\frac{1 + \phi^2}{2\sigma_\eta^2}((\log h_t - \gamma_t)^2 + \text{cte} - \gamma_t^2) \\
&\propto -\frac{1}{2\sigma^2}(\log h_t - \gamma_t)^2
\end{aligned}$$

onde

$$\gamma_t = \frac{\alpha(1 - \phi) + \phi(\log h_{t+1} + \log h_{t-1})}{1 + \phi^2} \qquad \sigma^2 = \frac{\sigma_\eta^2}{1 + \phi^2}$$

Portanto, a forma analítica de $p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y})$ é dada por:

$$p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y}) \propto \frac{1}{h_t^{1.5}} \exp\left(-\frac{y_t^2}{2h_t}\right) \exp\left(-\frac{(\log h_t - \gamma_t)^2}{2\sigma^2}\right) \quad (5.2.16)$$

Para $t = 1$ e $t = T$ teremos:

$$\begin{aligned}
p(h_1|\mathbf{h}_{\setminus 1}, \theta, \mathbf{y}) &\propto \frac{1}{h_1^{1.5}} \exp\left(-\frac{y_1^2}{2h_1}\right) \exp\left(-\frac{(\log h_1 - \frac{\alpha}{1-\phi})^2}{2\sigma_\eta^2}\right) \exp\left(-\frac{(\log h_2 - \alpha - \phi \log h_1)^2}{2\sigma_\eta^2}\right) \\
p(h_T|\mathbf{h}_{\setminus T}, \theta, \mathbf{y}) &\propto \frac{1}{h_T^{1.5}} \exp\left(-\frac{y_T^2}{2h_T}\right) \exp\left(-\frac{(\log h_T - \alpha - \phi \log h_{T-1})^2}{2\sigma_\eta^2}\right)
\end{aligned}$$

Obtida a forma analítica da densidade condicional a menos de uma constante, deve-se amostrar de tal distribuição. Em princípio uma forma de se obter amostra dessa função, seria utilizar o algoritmo de aceitação-rejeição (ver apêndice A). Nesse caso, é necessário uma densidade $q(x)$, da qual sabe-se amostrar, tal que $cq(x)$ envelope $p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y})$ para todo h_t . Dois motivos básicos fazem com que não seja indicado tal procedimento. Primeiro, seria necessário um processo de otimização a cada passo para encontrar uma nova constante c ; e por outro lado, caso tivéssemos facilmente essa constante, mesmo que ela fosse encontrada existe o problema da probabilidade de aceitação, que pode ser muito baixa.

A proposta de Jacquier *et al.*(1994) foi utilizar o algoritmo de Metropolis-Hastings (ver apêndice A) para amostrar da densidade desejada. O diagrama do algoritmo é apresentado na Figura 5.2

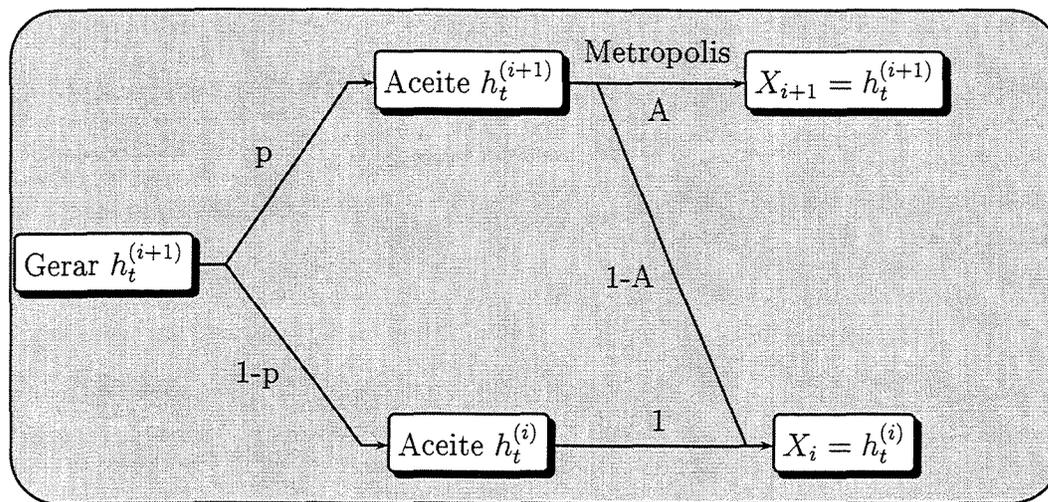


Figura 5.2: Diagrama do Algoritmo de Amostragem dos Estados

O algoritmo proposto para amostragem de $p(\theta|\mathbf{y})$ e $p(\mathbf{h}|\mathbf{y})$ é dado como segue: Seja $X_i = h_t^{(i)}$, onde $t = 1, \dots, T$.

① Dado $\mathbf{h}^{(i)}, \theta^{(i)}$, gere candidatos para $h_t^{(i+1)}$ amostrando de $q(h_t^{(i+1)}|\mathbf{h}_{\setminus t}, \theta, \mathbf{y})$

② (i) O candidato $X_{i+1} = h_t^{(i+1)}$ gerado é aceito com probabilidade

$$\min \left\{ \frac{p(h_t^{(i+1)}|\mathbf{h}_{\setminus t}, \theta^{(i)}, \mathbf{y})}{c q(h_t^{(i+1)}|\mathbf{h}_{\setminus t}, \theta^{(i)}, \mathbf{y})}, 1 \right\}$$

e vá para ③.

(ii) Caso contrário rejeite $h_t^{(i+1)}$, faça $X_{i+1} = h_t^{(i)}$ e pare.

③ (i) Aceite o candidato $X_{i+1} = h_t^{(i+1)}$ com probabilidade

$$\alpha(h_t^{(i)}, h_t^{(i+1)}) = \min \left\{ \frac{\pi(h_t^{(i+1)}) \min\{\pi(h_t^{(i)}, 1)\}}{\pi(h_t^{(i)}) \min\{\pi(h_t^{(i+1)}, 1)\}}, 1 \right\}$$

e pare.

(ii) Caso contrário rejeita $h_t^{(i+1)}$, faça $X_{i+1} = h_t^{(i)}$ e pare.

④ Dado $p(\mathbf{h}|\mathbf{y})$, amostre θ de $\theta|\mathbf{h}, \mathbf{y}$

⑤ Volte a ②.

Para implementação do algoritmo é necessário saber a forma funcional das densidades $q(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y})$ e $p(\theta|\mathbf{h}, \mathbf{y})$. A forma analítica de $p(\theta|\mathbf{h}, \mathbf{y})$ já foi discutida. Seja a forma funcional de $p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y})$ dada como na equação (5.2.16).

$$p(h_t|\mathbf{h}_{\setminus t}, \theta, \mathbf{y}) \propto \frac{1}{h_t^{1.5}} \exp\left(-\frac{y_t^2}{2h_t}\right) \exp\left(-\frac{(\log h_t - \gamma_t)^2}{2\sigma^2}\right)$$

Pode-se ver que o segundo termo em $p(h_t | \mathbf{h}_{\setminus t}, \theta, \mathbf{y})$ é uma densidade log-Normal. Tal densidade foi aproximada por Jacquier *et al.* por uma variável aleatória inversa de Gama. Essa aproximação foi efetuada igualando os dois primeiros momentos da log-Normal e Inversa de Gama. Para a densidade log-Normal os dois primeiros momentos são dados por $\exp(\gamma_t + \frac{1}{2}\sigma^2)$ e $\exp(2\gamma_t + 2\sigma^2)$, e para a Inversa de Gama são $\frac{\lambda}{\nu-1}$ e $\frac{\lambda^2}{(\nu-1)(\nu-2)}$. Assim,

$$\begin{cases} \exp(\gamma_t + \frac{1}{2}\sigma^2) &= \frac{\lambda}{\nu-1}, \\ \exp(2\gamma_t + 2\sigma^2) &= \frac{\lambda^2}{(\nu-1)(\nu-2)}. \end{cases}$$

Dessas Equações tem-se

$$\lambda = (\nu - 1) \exp(\gamma_t + \frac{1}{2}\sigma^2),$$

$$\nu - 1 = \frac{1 - 2 \exp(\sigma^2)}{1 - \exp(\sigma^2)}$$

Desta forma pode-se aproximar o termo log-Normal por uma Inversa de Gamma com parâmetros $\nu_1 = \frac{(1-\exp \sigma^2)}{(1-\exp \sigma^2)}$ e $\lambda_1 = (\nu - 1) \exp(\gamma_t + \frac{1}{2}\sigma^2)$. Portanto, aproxima-se $q(h_t | \mathbf{h}_{\setminus t}, \theta, \mathbf{y})$ pela seguinte densidade

$$\frac{1}{\sqrt{h_t}} \exp\left(\frac{y_t^2}{2h_t}\right) \frac{1}{h_t^{(\nu+1)}} \exp\left(-\frac{\lambda}{h_t}\right) \quad (5.2.17)$$

Esses termos podem ser combinados para obter uma densidade Inversa de Gamma com parâmetros $\nu = \nu_1 + \frac{1}{2}$ e $\lambda = \lambda_1 + \frac{1}{2}y_t^2$. Assim, a densidade é dada como segue

$$q(h_t | \mathbf{h}_{\setminus t}, \theta, \mathbf{y}) \propto \frac{1}{h_t^{(\nu+1)}} \exp\left(-\frac{\lambda}{h_t}\right)$$

onde

$$\nu = \frac{1 - 2 \exp(\sigma^2)}{1 - \exp(\sigma^2)} + \frac{1}{2} \quad \lambda = (\nu - 1) \exp(\gamma_t + \frac{1}{2}\sigma^2) + \frac{1}{2}y_t^2$$

onde γ_t e σ^2 são dados na equação (5.2.16).

5.2.2 Amostragem da Variável Latente por Blocos

Como o algoritmo de simulação via cadeia de Markov atualiza uma variável por vez, tem-se a vantagem de que densidade condicional são geralmente simples, mas por outro lado a correlação entre as variáveis entre os sucessivos "sweeps" do Metropolis é uma desvantagem, particularmente para problemas de modelos de séries temporais com altas correlações. Shephard e Pitt (1997) utilizaram o método de amostragem por blocos sugerido por Liu *et al.*(1994) extendendo esse método para os modelos que podem ser representados como nas equações (2.5.1) e (2.5.3), assumindo que $\log f(y_t | \theta_t)$ é côncava em θ_t .

Esse método tem a mesma função que o filtro de Kalman e o algoritmo de suavização nos modelos em **FEEG**, que estimam os estados \mathbf{h} condicionada às observações \mathbf{y} . Considerando o modelo de **VE** como na equação (3.3.4) a menos de reparametrização, tem-se

$$\begin{aligned} y_t &= \beta e^{h_t/2} \varepsilon_t, & t \geq 1, \\ h_{t+1} &= \phi h_t + \sigma_\eta \eta_t, \end{aligned} \tag{5.2.18}$$

onde ε_t e η_t são processos independentes Gaussianos com variância 1 e σ_η^2 respectivamente.

Como Shephard e Pitt utilizaram uma outra parametrização, diferente de Jacquier *et al.*, para o modelo VE, serão apresentados apenas os métodos para a parametrização utilizada. Análogo à seção 5.2.1 e beneficiando-se da relação direta entre as densidades log-Normal e Normal tem-se

$$\log f(y_t|h_t) = -\frac{1}{2} - \frac{y_t^2}{2\beta^2} \exp(-h_t) \quad \Rightarrow \quad \frac{\partial^2 \log f(y_t|h_t)}{\partial h_t^2} = -\frac{y_t^2}{2\beta^2} \exp(-h_t) < 0.$$

Como $h_t|\mathbf{h}_{\setminus t} \sim N(\mu_t, \sigma_t^2)$, onde $\mu_t = \frac{\phi(h_{t+1}+h_{t-1})}{(1+\phi^2)}$ e $\sigma_t^2 = \frac{\sigma_\eta^2}{(1+\phi^2)}$, tem-se

$$\begin{aligned} \log f(h_t|y_t, \mathbf{h}_{\setminus t}) &= \log f(h_t|h_{t-1}, h_{t+1}) + \log f(y_t|h_t) \\ &= \frac{(h_t - \mu_t)^2}{2\sigma_t^2} - \frac{h_t}{2} - \frac{y_t^2}{2\beta^2} \exp(-h_t) \\ &\simeq \frac{(h_t - \mu_t)^2}{2\sigma_t^2} - \frac{h_t}{2} - \frac{y_t^2}{2\beta^2} \exp(-\mu_t) \left\{ 1 - (h_t - \mu_t) + \frac{1}{2}(h_t - \mu_t)^2 \right\} \\ &= \log g \end{aligned} \tag{5.2.19}$$

O termo quadrático em $\log g$ implica que não é possível utilizar g como uma função envelope a fim de obter amostras de f através do algoritmo de aceitação/rejeição. A partir das seguintes notações $l'(h_t) = \frac{\partial \log f(y_t|h_t)}{\partial h_t}$ e $l''(h_t) = \frac{\partial^2 \log f(y_t|h_t)}{\partial h_t^2}$, a densidade sugerida para amostragem é dada, sem demonstração, pelo seguinte teorema:

Teorema 5.2.1. *Suponha σ_t não singular e que $\frac{\partial^2 \log f(y_t|h_t)}{\partial h_t^2}$ é negativa semi-definida para todos valores de h_t . Então os seguintes resultados são válidos.*

➤ *Pode-se amostrar de $h_t|h_{t-1}, h_{t+1}, y_t$ propondo h_t de*

$$N(\mu_t + \sigma_t^2 l'(m_t), \sigma_t^2),$$

a qual é aceita com probabilidade

$$\exp\{l(h_t) - l(m_t) - l'(m_t)(h_t - m_t)\},$$

➤ A probabilidade de se rejeitar a sugestão é minimizada tomando-se m_t como a moda de $f(h_t|h_{t-1}, h_{t+1}, y_t)$.

A sugestão dada para uma proposta pseudo-dominada h_t para $h_t|y_t, \mathbf{h}_{\setminus t}$ é retirada de

$$N\left[\frac{\sigma_t^{*2}}{\sigma_t}\mu_t + \frac{\sigma_t^{*2}}{2}\left\{\frac{y_t^2}{\beta^2}\exp(-\mu_t)(1 + \mu_t) - 1\right\}, \sigma_t^{*2}\right], \quad (5.2.20)$$

onde

$$\sigma_t^{*-2} = \sigma_t^{-2} + \frac{y_t^2}{2\beta\exp(\mu_t)}. \quad (5.2.21)$$

Na aceitação/rejeição, essas propostas são feitas até a aceitação, que é dada com probabilidade $\min(f/g, 1)$, enquanto a probabilidade do Metropolis aceitar a proposta é

$$\min\left[\frac{f(z|y_t, \mathbf{h}_{\setminus t}^{(k)}) \min\{f(h_t^{(k)}|y_t, \mathbf{h}_{\setminus t}^{(k)}), g(h_t^{(k)})\}}{f(h_t^{(k)}|y_t, \mathbf{h}_{\setminus t}^{(k)}) \min\{f(z|y_t, \mathbf{h}_{\setminus t}^{(k)}), g(z)\}}, 1\right] \quad (5.2.22)$$

SP generalizaram esses resultados a fim de amostrar de uma densidade g , multivariada. Não é razoável amostrar todos os estados de uma única vez, pois a probabilidade de aceitação no passo de M-H seria muito baixa. A amostragem por blocos diminui o problema de dependência e contorna o problema de alto índice de rejeição. Seja, $h_{t,k} = (h_{t-1}, \dots, h_{t+k-1})$, $\eta_{t,k} = (\eta_{t-1}, \dots, \eta_{t+k-1})$ blocos de estados e perturbações, **SP** utilizaram uma expansão de Taylor da densidade f para obter uma função que pudesse ser amostrada. Considerando a relação direta entre h e η tem-se que

$$\begin{aligned}
\log f &= \log f(h_{t,k}|h_{t-1}, \dots, h_{t+k+1}, y_t, \dots, y_{t+k}) \\
&= -\frac{1}{2}\eta'_{t,k}\eta_{t,k} + \sum_{i=t}^{t+k} l(\theta_i) \\
&\approx -\frac{1}{2}\eta'_{t,k}\eta_{t,k} + \sum_{i=t}^{t+k} l(\hat{\theta}_i) + (\theta_i - \hat{\theta}_i)^T l'(\hat{\theta}_i) + \frac{1}{2}(\theta_i - \hat{\theta}_i)^T l''(\hat{\theta}_i) + (\theta_i - \hat{\theta}_i) \quad (5.2.23)
\end{aligned}$$

$$= cte - \frac{1}{2}\eta'_{t,k}\eta_{t,k} - \frac{1}{2} \sum_{i=t}^{t+k} (\tilde{y}_i - \theta_i)^T v_i^{-1} (\tilde{y}_i - \theta_i) \quad (5.2.24)$$

$$= \log g$$

onde $l(\theta_t) = \log f(y_t|\theta_t)$ e

$$\tilde{y}_i = \hat{\theta}_i + v_i l'(\hat{\theta}_i) \quad e \quad v_i^{-1} = -l''(\hat{\theta}_i)$$

Agora a função g pode ser considerada Gaussiana com observações \tilde{y} e pode-se utilizar o filtro de Kalman e suavizador de de Jong ao modelo dado por

$$\begin{aligned}
\tilde{y}_i &= \theta_i + \varepsilon_i \quad \varepsilon_i \sim N(0, v_i) \quad i = t, \dots, t+k \\
\alpha_{i+1} &= T_i \alpha_i + H_i \eta_i \quad \eta_i \sim NID(0, I)
\end{aligned}$$

Como pode ser visto o procedimento é análogo ao que foi visto na seção 4.2. Os valores obtidos pelo suavizador de de Jong serão o bloco de variáveis proposto. Assim

a probabilidade ¹ de aceitação, i.e., passar de um estado $\alpha_{t,k}^0$ para $\alpha_{t,k}^n$ será

$$\mathbb{P}(\alpha_{t,k}^0 \rightarrow \alpha_{t,k}^n) = \min \left\{ 1, \frac{\omega(\theta_{t,k}^n)}{\omega(\theta_{t,k}^0)} \right\}$$

onde

$$\omega(\theta_{t,k}) = \exp\{l(\theta_{t,k}) - \tilde{l}(\theta_{t,k})\} = \exp \left[\sum_{i=t}^{t+k} \{l(\theta_i) - \tilde{l}(\theta_{t,k})\} \right]$$

Para utilizar o algoritmo A-R conjuntamente com M-H deve-se amostrar blocos propostos utilizando o suavizador e aceitar esse bloco, como proposta para o M-H, com probabilidade igual a $\min[\omega(\theta_{t,k}), 1]$. No passo M-H a probabilidade de aceitação da proposta $\alpha_{t,k}^n$ é dada por

$$\mathbb{P}(\alpha_{t,k}^0 \rightarrow \alpha_{t,k}^n) = \min \left\{ 1, \frac{\omega_{ar}(\theta_{t,k}^n)}{\omega_{ar}(\theta_{t,k}^0)} \right\}$$

onde

$$\omega_{ar}(\theta_{t,k}) = \exp[l(\theta_{t,k}) - \min\{\tilde{l}(\theta_{t,k}), l(\theta_{t,k})\}] = \max\{1, \omega(\theta_{t,k})\}$$

Após amostrado os estados, procede-se a amostragem dos parâmetros de acordo com a seção 5.3.

¹Para maiores detalhes ver Shephard e Pitt (1997)

5.2.3 Amostragem Conjunta da Variável Latente

Nesta subseção a representação do modelo de volatilidade estocástica considerada será a apresentada na equação (3.3.4), i.e., da forma .

$$y_t = \beta e^{\frac{h_t}{2}} \varepsilon_t \quad t \geq 1$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t \quad (5.2.25)$$

$$(5.2.26)$$

O parâmetro $\beta = e^{\frac{\mu}{2}}$ é chamado o fator de escala constante, ϕ é a persistência na volatilidade e σ_η é a volatilidade da logvolatilidade. Por questão de identificabilidade toma-se $\beta = 1$ ou $\mu = 0$, mas para facilidade de interpretação tomou-se $\beta = 1$.

Melhorando o Algoritmo MCMC

Kim *et al.*(1998) utilizaram uma mistura de distribuições normais para aproximar a verossimilhança exata. Tal aproximação possibilita a implementação do que é chamado na literatura de amostrador de Gibbs adaptado, o qual permite amostrar todas as log volatilidades de uma vez. Objetivando-se melhorar o erro de aproximação Kim *et al.*(1998) propuseram a reponderação das amostras dos parâmetros e volatilidades obtidas.

Aproximação por Mistura

O modelo de volatilidade estocástica linearizado como em (3.3.9), é aproximado através de misturas de densidades normais considerando a seguinte transformação

$$y_t^* = h_t + z_t \quad (5.2.27)$$

onde $y_t^* = \log(y_t^2 + c)$ e

$$f(z_t) = \sum_{i=1}^K q_i f_N(z_t | m_i - 1.2704, v_i^2), \quad (5.2.28)$$

onde $f_N(z_t | m_i - 1.2704, v_i^2)$ é a suposição de normalidade para a variável z_t , com média $m_i - 1.2704$ e variância v_i e as constantes $\{q_i, m_i, v_i^2\}$ são escolhidas de forma a aproximar bem a densidade exata de $\log \varepsilon_t^2$. Fuller (1996) propôs o compensador c na literatura dos modelos de **VE**. Ele tinha como objetivo tornar o estimador de quasi máxima verossimilhança do modelo de **VE** mais robusto quando y_t^2 fosse próximo a zero. O valor de c pode ser obtido através dos valores de y_t^2 .

Deve-se notar que a mistura de densidade pode ser representada em termos de uma variável indicadora s_t tal que

$$z_t | s_t = i \sim N(m_i - 1.2704, v_i^2), \quad (5.2.29)$$

$$\Pr(s_t = i) = q_i$$

Para a escolha de K e $\{q_i, m_i, v_i^2\}$ ($i \leq K$), Kim *et al.*(1998) utilizaram a sugestão de Titterington *et al.*(1985, p. 133), onde utilizou-se mínimos quadrados não linear a fim de se obter valores tais que os primeiros quatro momentos da densidade $f_{\exp(z)}(r)$ e $f(z_t)$ fossem o mais próximo possível dos momentos das variáveis aleatórias χ_1^2 e $\log \chi_1^2$. Eles encontraram que $K = 7$ produz uma aproximação satisfatória. Os pesos, médias e variâncias desta aproximação são dados na tabela 5.1

s	$\Pr(s = i)$	m_i	v_i^2
1	0.00730	-10.12999	5.79596
2	0.10556	-3.97281	2.61369
3	0.00002	-8.56686	5.17950
4	0.04395	2.77786	0.16735
5	0.34001	0.61942	0.64009
6	0.24566	1.79518	0.34023
7	0.25750	-1.08819	1.26261

Tabela 5.1: Seleção das misturas de normais para aproximar a distribuição $\log \chi_1^2$

Simulando da Mistura

Na abordagem de MCMC, os modelos de misturas são melhor estimados explorando a representação em (5.2.29). A densidade à posteriori de interesse é $\pi(s, h, \phi, \sigma_\eta^2, \mu | y^*)$, onde $\mathbf{s} = (s_1, \dots, s_n)$. Neste caso, \mathbf{h} e \mathbf{s} podem ser amostrados separadamente em um bloco.

- ① Inicialize $\mathbf{s}, \phi, \sigma_\eta^2$ e μ .
- ② Amostre h de $\mathbf{h} | \mathbf{y}^*, \mathbf{s}, \phi, \sigma_\eta^2, \mu$.
- ③ Amostre \mathbf{s} de $\mathbf{s} | \mathbf{y}^*, \mathbf{h}$.
- ④ Atualize ϕ, σ_η^2, μ conforme a seção de amostragem de parâmetros
- ⑤ Volte a 2.

Agora pode-se amostrar da distribuição Gaussiana multivariada $\mathbf{h} | \mathbf{y}^*, s, \phi, \sigma_\eta, \mu$, pois $\mathbf{y}^* | \mathbf{s}, \phi, \sigma_\eta, \mu$ é uma série Gaussiana, que pode ser colocada na **FEEG**, possibilitando a utilização do filtro de Kalman para amostrar de $\mathbf{h} | \mathbf{y}^*, s, \sigma_\eta, \mu$ através do algoritmo

de suavização de de Jong e Shephard (1995). A amostragem de s de $\mathbf{s}|\mathbf{y}^*, \mathbf{h}$ é feita amostrando independentemente cada s_t da seguinte função de probabilidade

$$\begin{aligned} f(s_t|y_t^*, h_t) &\propto f(s_t)f(y_t^*|\mathbf{s}, \phi, \sigma_\eta, \mu) \\ \Pr(s_t = i|y_t^*, h_t) &\propto q_i f_N(y_t^*|h_t + m_i - 1.2704, v_i^2) \quad i \leq K. \end{aligned} \quad (5.2.30)$$

como $y_t^* = h_t + z_t$ e z_t é mistura de normais então, $y_t^*|s_t, h_t, m_i, v_i \sim N(h_t + m_i - 1.2704, v_i^2) \quad i \leq K$

Integrando as log-volatilidades

Com o algoritmo anterior obtém-se uma significativa melhora no comportamento da correlação das simulações. Kim *et al.*(1998) sugeriram amostrar da densidade conjunta $\pi(\phi, \sigma_\eta^2, \mathbf{h}, \mu|\mathbf{y}^*, \mathbf{s})$ amostrando-se (ϕ, σ_η^2) de $\pi(\phi, \sigma_\eta^2|\mathbf{y}^*, \mathbf{s}) \propto f(\mathbf{y}^*|\mathbf{s}, \phi, \sigma_\eta^2)\pi(\phi, \sigma_\eta^2)$ e depois amostrar (\mathbf{h}, μ) de $\pi(\mathbf{h}, \mu|\mathbf{y}^*, \mathbf{s}, \phi, \sigma_\eta^2)$ utilizando o filtro de Kalman. Então, escrevendo $\mu|\mathbf{y}^*, \mathbf{s}, \phi, \sigma_\eta^2 \sim N(\tilde{\mu}, \sigma_\mu^2)$ tem-se

$$\begin{aligned} \pi(\phi, \sigma_\eta^2|\mathbf{y}^*, \mathbf{s}) &\propto \pi(\phi)\pi(\sigma_\eta^2)f(\mathbf{y}^*|\mathbf{s}, \phi, \sigma_\eta^2) = \pi(\phi)\pi(\sigma_\eta^2) \frac{f(\mathbf{y}^*|\mathbf{s}, \phi, \sigma_\eta^2, \mu = 0)\pi(\mu = 0)}{\pi(\mu = 0|\mathbf{y}^*, \mathbf{s}, \phi, \sigma_\eta^2)} \\ &\propto \pi(\phi)\pi(\sigma_\eta^2) \prod_{t=1}^n F_t^{-1/2} \exp\left(-\frac{1}{2} \sum_{t=1}^n v_t^2/F_t\right) \exp\left(\frac{1}{2\sigma_\eta^2 \tilde{\mu}^2} \tilde{\mu}^2\right) \sigma_\mu, \end{aligned} \quad (5.2.31)$$

onde v_t é o erro de previsão um passo-à-frente e F_t é seu erro quadrático médio. As quantidades $v_t, F_t, \tilde{\mu}, \sigma_\mu^2$ são obtidas pelo filtro de Kalman, condicionados em \mathbf{s} . Isto implica que dado o valor atual $\{\phi^{(i-1)}, \sigma_\eta^{2(i-1)}\}$, pode-se amostrar de $\phi, \sigma_\eta^2|\mathbf{y}^*, \mathbf{s}$ diretamente propondo-se $\{\phi^{(i)}, \sigma_\eta^{2(i)}\}$, através da amostragem de uma densidade $g(\phi, \sigma_\eta^2)$, com probabilidade de aceitação

$$\min \left\{ 1, \frac{\pi(\phi^{(i)}, \sigma_\eta^{2(i)} | \mathbf{y}^*, \mathbf{s})}{\pi(\phi^{(i-1)}, \sigma_\eta^{2(i-1)} | \mathbf{y}^*, \mathbf{s})} \frac{g(\phi^{(i-1)}, \sigma_\eta^{2(i-1)})}{g(\phi^{(i)}, \sigma_\eta^{2(i)})} \right\} \quad (5.2.32)$$

O algoritmo para integração é dado da seguinte forma

- ① Inicializa $(\mathbf{s}, \phi, \sigma_\eta, \mu)$.
- ② Amostra (ϕ, σ_η) de $\pi(\phi, \sigma_\eta^2 | \mathbf{y}^*, \mathbf{s})$ usando uma sugestão de M-H baseada em $g(\sigma_\eta^2, \phi)$, aceitando com probabilidade (5.2.32).
- ③ Amostra $\mathbf{h}, \mu | \mathbf{y}^*, \mathbf{s}, \phi, \sigma_\eta^2$ usando o algoritmo de suavização aumentado.
- ④ Amostra $\mathbf{s} | \mathbf{y}^*, \mathbf{h}$ como no algoritmo anterior.
- ⑤ Volte a 2.

Uma vantagem desse amostrador é que pode-se amostrar de (\mathbf{h}, μ) a partir do algoritmo de suavização. A amostragem de $\pi(\phi, \sigma_\eta^2 | \mathbf{y})$ é feita utilizando-se o algoritmo de Metropolis-Hastings de Gilks *et al.*(1995).

Reponderação

Após obtidas amostras das distribuições posterioris dos parâmetros Kim *et al.*(1998) propuseram fazer uma reponderação das amostras para melhorar as estimativas obtidas. Supondo que as amostras são retiradas de $k(\theta, \mathbf{h} | \mathbf{y}^*)$, onde $\theta = (\phi, \sigma_\eta, \mu)$, defina

$$w(\theta, \mathbf{h}) = \log \frac{f(\theta, \mathbf{h} | \mathbf{y})}{k(\theta, \mathbf{h} | \mathbf{y}^*)} = cte + \log f(\mathbf{y} | \mathbf{h}) - \log k(\mathbf{y}^* | \mathbf{h}),$$

onde temos,

$$f(\mathbf{y}|\mathbf{h}) = \prod_{t=1}^n f_N(y_t|0, e^{h_t})$$

$$k(\mathbf{y}^*|\mathbf{h}) = \prod_{t=1}^n \sum_{i=1}^k q_i f_N(y_t|h_t + m_i - 1.2704, v_i^2).$$

dessa forma temos para qualquer função de θ a seguinte expressão

$$E(g(\theta)|\mathbf{y}) = \int g(\theta) f(\theta|\mathbf{y}) d\theta = \frac{\int g(\theta) e^{w(\theta, \mathbf{h})} k(\theta, \mathbf{h}|\mathbf{y}^*) d\theta d\mathbf{h}}{\int e^{w(\theta, \mathbf{h})} k(\theta, \mathbf{h}|\mathbf{y}^*) d\theta d\mathbf{h}} \quad (5.2.33)$$

e fazemos as estimativas por reponderação como segue

$$E(g(\hat{\theta})|\mathbf{y}) = \sum_j g(\theta^j) \frac{e^{w(\theta^j, \mathbf{h}^j)}}{\sum_i e^{w(\theta^i, \mathbf{h}^i)}} \quad (5.2.34)$$

5.3 Amostragem dos Coeficientes do Modelo

Jacquier *et al.*, Shephard e Pitt e Kim *et al.* para facilitar os cálculos e aumentar a eficiência nos resultados propuseram prioris conjugadas para todos os coeficientes do modelo.

Jacquier *et al.* (1994) consideraram um vetor θ dado como $\theta = (\Delta, \sigma_\eta)$, onde $\Delta = (\alpha, \phi)$, então $p(\theta) = p(\Delta, \sigma_\eta) = p(\Delta|\sigma_\eta)p(\sigma_\eta)$. Eles sugeriram modelar $\Delta|\sigma_\eta$ como uma normal multivariada e σ_η como uma inversa de gama. Assim tem-se,

$$\begin{aligned}\Delta|\sigma_\eta &\sim NM(\Delta, \sigma_\eta^2 A^{-1}) \\ p(\Delta|\sigma_\eta) &= \frac{|A|^{\frac{1}{2}}}{2\pi\sigma_\eta^2} \exp\left(-\frac{(\Delta - \bar{\beta})^T A(\Delta - \bar{\beta})}{2\sigma_\eta^2}\right) \\ \sigma_\eta &\sim IG\left(\nu, \frac{1}{\lambda}\right) \\ p(\sigma_\eta) &= \frac{\lambda^\nu}{\Gamma(\nu)\sigma_\eta^{(\nu+1)}} \exp\left(-\frac{\lambda}{\sigma_\eta}\right)\end{aligned}$$

onde $\alpha, \phi, \sigma_\eta, A, \nu, \lambda$ são parâmetros a priori. Com as prioris conjugadas escolhidas por Jacquier *et al.*(1994) a posteriori resultante é um produto de uma normal multivariada e uma inversa de gama. Assim, é necessário amostrar de uma normal multivariada e uma inversa de gama apropriadas para amostrar de $p(\theta|\mathbf{h}, \mathbf{y})$.

Amostrando σ_η^2, ϕ . A amostragem de σ_η^2 e ϕ feita por Kim *et al.*(1998) é direta. Eles assumem, assim como Shephard e Pitt (1997), uma priori conjugada $\sigma_\eta^2|\phi, \mu \sim IG(\sigma_r/2, S_\sigma/2)$, Assim, amostra-se de uma Gama-inversa:

$$\sigma_\eta^2|\mathbf{y}, \mathbf{h}, \phi, \mu \sim IG\left\{\frac{n + \sigma_r}{2}, \frac{S_\sigma + (h_1 - \mu)^2(1 - \phi^2) + \sum_{t=1}^{n-1} ((h_{t+1} - \mu) - \phi(h_t - \mu))^2}{2}\right\}; \quad (5.3.1)$$

Kim *et al.*(1998) fixaram $\sigma_r = 5$ e $S_\sigma = 0.01 \times \sigma_r$. A partir da relação entre a distribuição gama e χ^2 , Shephard e Pitt, partindo de uma priori $\sigma_\eta^2|\phi \sim \chi_p^{-2}S_0$ encontraram a posteriori

$$\sigma_\eta^2|\mathbf{y}, \alpha, \phi \sim \chi_{n+p}^{-2}\left\{S_0 + \alpha_1^2(1 - \phi^2) + \sum_{t=1}^{n-1} (\alpha_{t+1} - \phi\alpha_t)^2\right\}, \quad (5.3.2)$$

onde a parametrização de **SP** considera $\alpha_t = h_t - \mu$ relativo a parametrização do modelo que foi utilizada por Kim *et al.*

O parâmetro ϕ é amostrado da densidade condicional completa através do método A-R. Assim, considerando a reparametrização, $\phi = 2\phi^* - 1$, onde a distribuição de ϕ é uma beta com parâmetros $(\phi^{(1)}, \phi^{(2)})$, a priori de ϕ é dada por

$$\pi(\phi) \propto \left\{ \frac{(1+\phi)}{2} \right\}^{\phi^{(1)}-1} \left\{ \frac{(1-\phi)}{2} \right\}^{\phi^{(2)}-1}, \quad \phi^{(1)}, \phi^{(2)} > \frac{1}{2}, \quad (5.3.3)$$

através dessa priori Kim *et al.* obtiveram suporte em $(-1, 1)$. Foi fixado no artigo deles $\phi^{(1)} = 20$ e $\phi^{(2)} = 1.5$ obtendo-se uma média a priori de 0.86.

Dado a priori (5.3.3), a densidade condicional completa de ϕ é proporcional a

$$\pi(\phi) f(h|\mu, \phi, \sigma_\eta^2),$$

onde

$$\log f(h|\mu, \phi, \sigma_\eta^2) \propto -\frac{(h_t - \mu)^2(1 - \phi)}{2\sigma_\eta^2} + \frac{1}{2} \log(1 - \phi^2) - \frac{\sum_{t=1}^{n-1} \{(h_{t+1} - \mu) - \phi(h_t - \mu)\}^2}{2\sigma_\eta^2}, \quad (5.3.4)$$

Esta função é côncava em ϕ para quaisquer valores de $\phi^{(1)}$ e $\phi^{(2)}$. Isto nos fornece que pode-se amostrar ϕ usando um algoritmo A-R. Kim *et al.* (1998) tomaram a expansão por série de Taylor da priori em torno de

$$\hat{\phi} = \frac{\sum_{t=1}^{n-1} (h_{t+1} - \mu)(h_t - \mu)}{\sum_{t=1}^{n-1} (h_t - \mu)^2}, \quad (5.3.5)$$

Logo, dado um valor proposto ϕ^* de $N(\hat{\phi}, V_\phi)$ onde $V_\phi = \sigma_\eta^2 \{\sum_{t=1}^{n-1} (h_t - \mu)^2\}^{-1}$ este valor é aceito com probabilidade $\exp\{g(\phi^*) - g(\phi^{(i-1)})\}$, onde

$$g(\phi) = \log \pi(\phi) - \frac{(h_t - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} + \frac{1}{2} \log(1 - \phi^2). \quad (5.3.6)$$

Se o valor proposto é rejeitado, tomamos $\phi^{(i)} = \phi^{(i-1)}$.

Amostrando μ . Supondo uma priori difusa² para μ , ele é amostrado de uma distribuição condicional completa dada por

$$\mu | \mathbf{h}, \phi, \sigma_\eta^2 \sim N(\hat{\mu}, \sigma_\mu^2), \quad (5.3.7)$$

onde

$$\hat{\mu} = \sigma_\mu^2 \left\{ \frac{(1 - \phi^2)}{\sigma_\eta^2} h_1 + \frac{(1 - \phi)}{\sigma_\eta^2} \sum_{t=1}^{n-1} (h_{t+1} - \phi h_t) \right\}, \quad (5.3.8)$$

e

$$\sigma_\mu^2 = \sigma_\eta^2 \{ (n - 1)(1 - \phi^2) + (1 - \phi^2) \}^{-1}. \quad (5.3.9)$$

Como existe uma relação direta entre μ e β foi amostrado o parâmetro μ e feita a transformação $\beta = \exp(\mu/2)$. Shephard e Pitt não usaram essa parametrização, eles amostraram de $\beta^2 | \mathbf{y}, \alpha \sim \chi_n^{-2} \sum y_t^2 \exp(-\alpha_t)$

²No trabalho foi usada $N(0, 10)$ por exemplo

5.4 Conclusões

As diversas formas de amostrar as volatilidades tem aumentado a eficiência e diminuindo os problemas de dependência no passo de M-H. A amostragem das posteriores dos parâmetros consideradas pelos autores são análogas, podem ser feitas, fora amostragem de ϕ , diretamente. Vale ressaltar a melhora ao amostrarmos os parâmetros conjuntamente, coonforme o artigo de Pitt e Shephard (1999b).

CAPÍTULO 6

Aplicação dos Métodos

6.1 Introdução

Neste capítulo serão aplicadas as metodologias apresentadas anteriormente à séries financeiras simuladas e a série de retornos compostos do IBOVESPA. Nas séries simuladas utilizou-se amostra de tamanho 1000 e também foi feita análise considerando a presença de *outliers* do tipo AO e IO, verificando seus efeitos nas estimativas de vários valores de interesse, como verossimilhança, estimativas dos parâmetros e estados. A série de retornos IBOVESPA é considerada no período de 02/jan/1995 a 27/dez/2000 totalizando 1500 observações, sendo que as principais crises nesse período foram: fevereiro/março de 1995 (México), outubro de 1997 (Ásia), agosto de 1998 (Rússia), janeiro 1999 (Brasil) e abril de 2000 (Nasdaq).

6.2 Análise das Séries Simuladas

Esta seção é dedicada a aplicação dos métodos estudados as séries simuladas. O objetivo é verificar o comportamento dos métodos nas estimativas dos parâmetros do modelo **VE**. Foram simuladas diversas séries de tamanho 1000, considerando diversos conjuntos de parâmetros. Todas as séries foram analisadas em quatro situações diferentes. Na primeira, as séries foram geradas sem a presença de valores aberrantes, no segundo caso foi inserido no processo gerador dos dados, um AO na posição 250 e um IO na posição 600. Nos dois últimos casos considerou-se a presença de apenas um AO e um IO respectivamente. O tamanho do AO foi tomado como sendo igual a Δ desvios padrões da série de retornos simulada sem *outlier* e o IO como sendo igual a Λ vezes o desvio padrão do processo autoregressivo, i.e., $\Lambda[\sigma_\eta^2(1 - \phi^2)^{-1}]^{0.5}$. Em todas as séries foram mantidos os mesmos conjuntos de perturbações nas equações de observação e estado. As séries analisadas foram geradas com os seguintes conjuntos de parâmetros.

Casos	Parâmetros				
	ϕ	σ_η	β	Δ_{250}	Λ_{600}
I	0.98	0.12	1.0	4	4
II	0.95	0.10	1.0	4	4
III	0.90	0.50	1.0	4	4
IV	0.98	0.12	1.0	7	9

Tabela 6.1: Valores para o conjunto de parâmetros para o processo gerador das séries, com Δ = tamanho do AO e Λ = tamanho do IO

Os valores dos parâmetros foram escolhidos baseados nos valores presentes na literatura dos modelos de **VE**. Para cada caso foi considerada a presença conjunta dos *outliers* do tipo AO e IO e a presença marginal dos valores espúrios.

A Tabela 6.2 e 6.3 apresentam algumas estatísticas descritivas do caso I e IV respectivamente. Pode-se notar que a presença de um *outlier* do tipo AO na série tem influência maior na média dos retornos, enquanto que a presença de um IO aumenta sensivelmente o desvio padrão da série e altera o sinal do coeficiente de assimetria.

Tipo Outlier	Média	D. Padrão	Assimetria	E. Curtose	Mínimo	Máximo
Sem Outlier	0.002340	1.052716	0.024955	1.124290	-4.252678	4.068046
AO e IO	0.006303	1.203698	0.027712	1.858323	-5.846982	5.017068
AO	0.006340	1.062854	0.098231	1.344327	-4.252678	4.685400
IO	0.002303	1.194756	-0.021503	1.769920	-5.846982	5.017068

Tabela 6.2: Estatísticas descritivas das Séries Simuladas para o Caso I

Tipo Outlier	Média	D. Padrão	Assimetria	E. Curtose	Mínimo	Máximo
Sem Outlier	0.002340	1.052716	0.024955	1.124290	-4.252678	4.068046
AO e IO	0.007969	1.445329	-0.02493	10.39526	-11.60306	8.599363
AO	0.008890	1.082088	0.394731	3.540255	-4.252678	7.908236
IO	0.007390	1.423694	-0.41662	10.27050	-11.60306	8.599363

Tabela 6.3: Estatísticas descritivas das Séries Simuladas para o Caso IV

Através das estatísticas descritivas, a primeira diferença que pode ser destacada é a maior influência dos *outliers* na série do Caso IV em relação a série do Caso I, o que era esperado devido a magnitude dos valores inseridos no processo gerador de dados. Os gráficos das figuras 6.1 e 6.2 apresentados de forma ilustrativa, são das séries simuladas para o Caso I e Caso IV respectivamente, como as estatísticas descritivas e gráficos dos casos II e III são similares eles não serão apresentados.

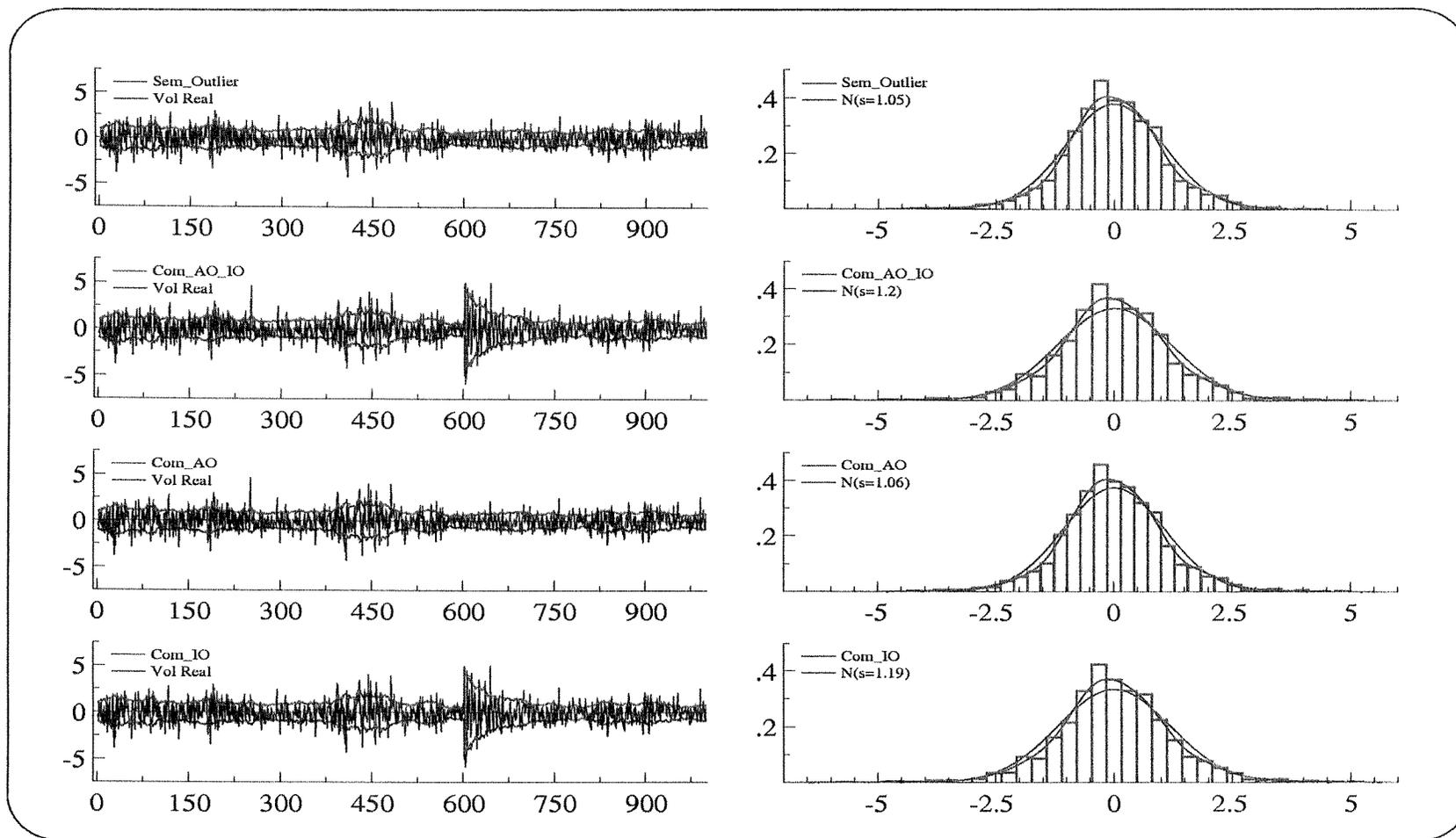


Figura 6.1: Gráficos das séries simuladas com o conjunto de parâmetro para modelo para Caso I.

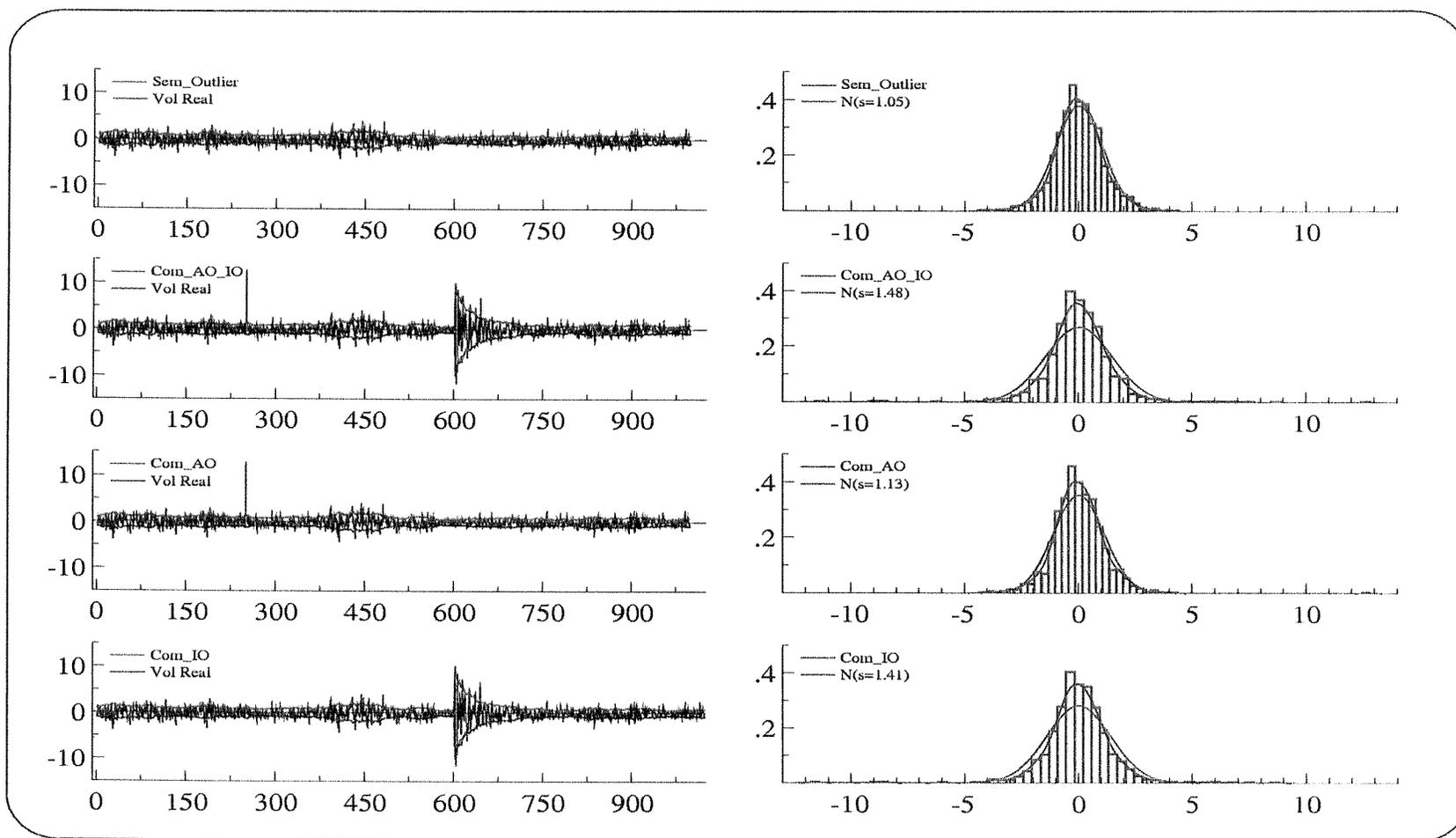


Figura 6.2: Gráficos das séries simuladas com o conjunto de parâmetro para modelo para Caso IV.

6.2.1 Cálculo da Verossimilhança Corrigida

Nesta subsecção são apresentados os valores das log-verossimilhanças corrigidas, que foram calculadas utilizando-se os algoritmos de Durbin e Koopman (1997a) e Shephard e Pitt (1997). Os valores foram obtidos considerando os quatro casos da tabela 6.1.

Tipo Outlier	Log-Verossimilhança	
	Durbin e Koopman	Shephard e Pitt
Sem Outlier	-1.410778597	-1.410696019
AO e IO	-1.533930887	-1.532933350
AO	-1.420669052	-1.420407575
IO	-1.525432642	-1.524605183

(a) Valores da log-verossimilhança para série simuladas com o conjunto de parâmetros do Caso I

Tipo Outlier	Log-Verossimilhança	
	Durbin e Koopman	Shephard e Pitt
Sem Outlier	-1.450987713	-1.450216403
AO e IO	-1.628560330	-1.625334055
AO	-1.470792787	-1.464106513
IO	-1.620536332	-1.620386259

(b) Valores da log-verossimilhança para série simuladas com o conjunto de parâmetros do Caso II

Tabela 6.4: Valores das Log-verossimilhanças corrigidas para os Casos I e II

Tipo Outlier	Log-Verossimilhança	
	Durbin e Koopman	Shephard e Pitt
Sem Outlier	-1.549062804	-1.548716542
AO e IO	-1.588619407	-1.587920780
AO	-1.554558655	-1.554037803
IO	-1.583365717	-1.582837744

(a) Valores da log-verossimilhança para série simuladas com o conjunto de parâmetros do Caso III

Tipo Outlier	Log-Verossimilhança	
	Durbin e Koopman	Shephard e Pitt
Sem Outlier	-1.410778597	-1.410696019
AO e IO	-1.533535657	-1.532768688
AO	-1.420767139	-1.420747324
IO	-1.525393043	-1.525386254

(b) Valores da log-verossimilhança para série simuladas com o conjunto de parâmetros do Caso IV

Tabela 6.5: Valores das Log-verossimilhanças corrigidas para os Casos III e IV

Nota-se pelas tabelas 6.4 e 6.5 que os métodos propostos produzem valores da verossimilhança próximos, sendo que a técnica proposta por Durbin e Koopman produz valores menores para todos os casos. O comportamento dos métodos na presença de *outliers* são semelhantes, havendo decréscimo mais acentuado quando da presença dos dois tipos de valores aberrantes. No entanto, vale ressaltar que o IO tem maior contribuição no decréscimo da log-verossimilhança.

6.2.2 Estudo da Convergência na Análise Bayesiana

O algoritmo de MCMC foi inicializado tomando-se as volatilidades $h_t = 0$, o parâmetro de persistência $\phi = 0$, a variância da volatilidade $\sigma_\eta = 0.02$ e $\mu = 0$. Esses valores foram tomados de acordo com os resultados fornecidos por Kim *et al.*(1998). Eles também verificaram que não se consegue estimativas diferentes ao rodar múltiplas cadeias com valores iniciais diferentes.

Foram feitas 1000 iterações do algoritmo de estimação da volatilidade e depois 50000 iterações de amostragem dos parâmetros e volatilidades. Após o *burn-in* foram computadas 150000 amostras das densidades posterioris que se encontram no gráfico da figura 6.3. O número de iterações para o *burn-in* foi sugerido no artigo de Kim *et al.*(1998) onde, com esse número de iterações eles asseguram que o efeito do valor inicial é insignificante.

Pode-se notar pelo gráfico 6.3 que a função de autocorrelação, i.e., o correlograma para amostragem simples indica uma grande correlação nas amostras para os parâmetros ϕ e σ_η , portanto necessita-se de um número bastante grande de simulações para perder a influência do ponto inicial.

A comparação entre os gráficos das figuras 6.3 da amostragem simples e 6.4 da amostragem por mistura deixa claro a redução da correlação para os três parâmetros. Dessa forma, pode-se fazer inferência para os parâmetros a partir de um número menor de iterações e obter bons resultados.

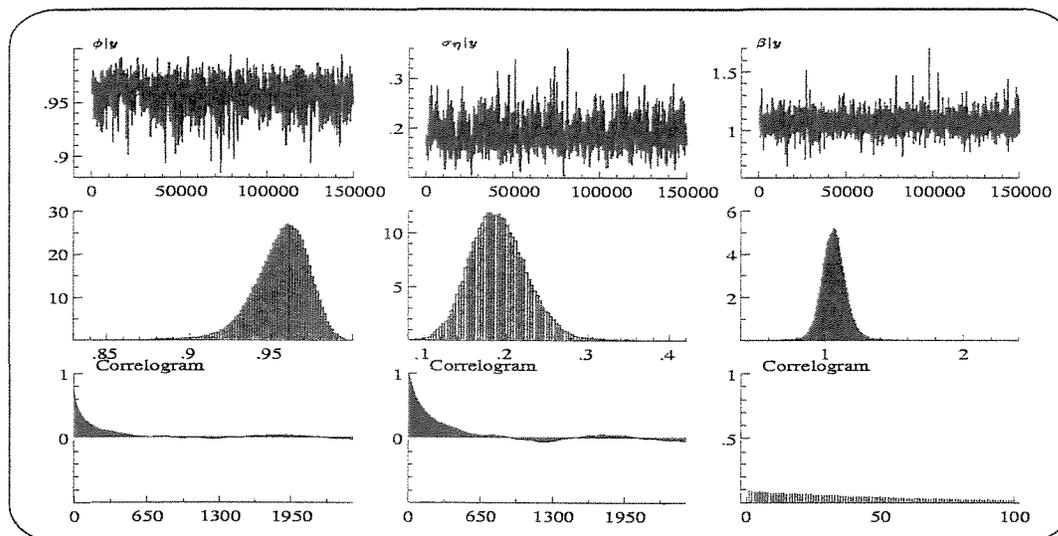


Figura 6.3: Gráficos dos valores dos parâmetros, histogramas das densidades posteriores e correlograma para amostragem simples com a presença de *outliers* do tipo AO e IO. Tempo para estimativa dos parâmetros em um Pentium III 550 foi de 399.42 segundos

Verifica-se pela figura 6.3 que o tempo necessário para as estimativas é de 399.42 segundos (aproximadamente 7 minutos) utilizando amostragem simples. Por outro lado, através da figura 6.4 verifica-se que 2176.4 segundos (36 minutos aproximadamente) é o tempo necessário para obter as estimativas através do método de integração, mas vale ressaltar que esse tempo foi obtido utilizando-se o mesmo número de iterações que na amostragem simples, o que pode-se verificar através do correlograma que não é necessário.

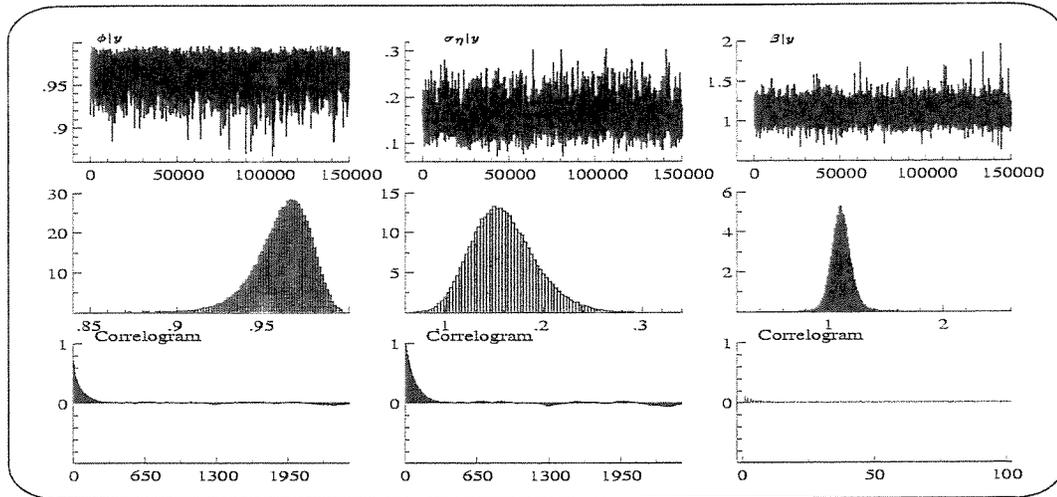


Figura 6.4: Gráficos dos valores dos parâmetros, histogramas das densidades posteriores e correlograma para amostragem utilizando o algoritmo de mistura com a presença de *outliers* do tipo AO e IO. Tempo para estimativa dos parâmetros em um Pentium III 550 foi de 2176.4 segundos

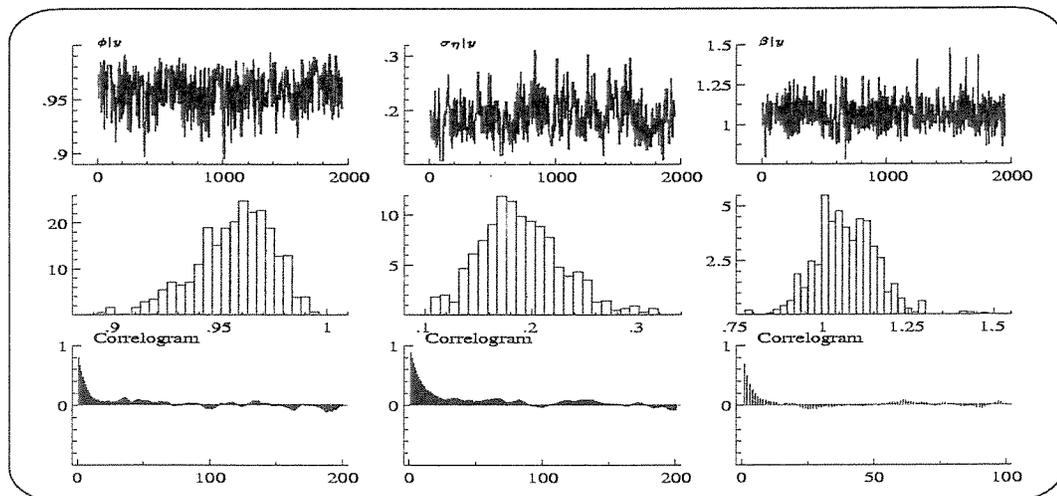


Figura 6.5: Gráficos dos valores dos parâmetros, histogramas das densidades posteriores e correlograma para amostragem utilizando o algoritmo de integração com a presença de *outliers* do tipo AO e IO. Tempo para estimativa dos parâmetros em um Pentium III 550 foi de 38.840 segundos

A figura 6.5 mostra que com apenas 2000 iterações do algoritmo de integração, atingiu-se rapidamente uma baixa correlação entre as iterações para os parâmetros. Assim, pode-se diminuir o tempo consideravelmente (aproximadamente 39 segundos) para obter estimativas razoáveis para os parâmetros do modelo.

6.2.3 Estimativas dos Parâmetros

As séries foram simuladas para os conjuntos de parâmetros dados na tabela 6.1 e foram obtidas estimativas para os parâmetros do modelo de VE utilizando-se os métodos discutidos. As estimativas foram calculadas considerando o enfoque clássico e o enfoque bayesiano. Para a análise frequentista foi calculada a logverossimilhança pelos métodos de Durbin e Koopman (1997a) e Shephard e Pitt (1997) e esta foi maximizada utilizando o algoritmo de otimização BFGS (Broyden-Fletcher-Goldfarb-Shanno), que é uma das rotinas da linguagem OX, para se obter as estimativas dos parâmetros que se encontram nas tabelas 6.6, 6.7, 6.8 e 6.9. O erro padrão para as metodologias clássicas foi calculado pela segunda derivada numérica da logverossimilhança.

As estimativas bayesianas dos parâmetros do modelo encontradas nas tabelas 6.6, 6.7, 6.8 e 6.9, foram obtidas calculando-se a média amostral dos valores amostrados da distribuição posteriori, pois Durbin e Koopman (1992) mostraram que as estimativas obtidas pela média e mediana são análogas. O erro padrão foi calculado através da raiz quadrada da variância amostral dos valores para cada um dos coeficientes. As principais conclusões são:

Θ	Tipo Outlier	Clássica		Bayesiana			
		Durbin Koopman	Shephard Pitt	Amostragem Simples	Mistura	Integração	Reponderação
$\phi = 0.98$	Sem Outlier	$\hat{\phi} =$ (0.00814)	0.98176 (0.00823)	$\phi y = 0.98061$ (0.00859)	0.98506 (0.01286)	0.98048 (0.00892)	0.98056 (0.00892)
		$\hat{\sigma}_\eta =$ (0.02605)	0.11073 (0.02389)	$\sigma_\eta y = 0.11525$ (0.02194)	0.10719 (0.01253)	0.11402 (0.02271)	0.11426 (0.02253)
		$\hat{\beta} =$ (0.08943)	0.95690 (0.09154)	$\beta y = 0.98133$ (0.11067)	1.13367 (0.08090)	0.98348 (0.11341)	0.98328 (0.11417)
$\sigma_\eta = 0.12$	AO IO	$\hat{\phi} =$ (0.00493)	0.96156 (0.00487)	$\phi y = 0.96219$ (0.01094)	0.97833 (0.00779)	0.96475 (0.01023)	0.96458 (0.01028)
		$\hat{\sigma}_\eta =$ (0.01925)	0.23744 (0.02032)	$\sigma_\eta y = 0.233779$ (0.04791)	0.15868 (0.02633)	0.22138 (0.04252)	0.22587 (0.04229)
		$\hat{\beta} =$ (0.22398)	1.13291 (0.23106)	$\beta y = 1.08018$ (0.28080)	1.28382 (0.12369)	1.08332 (0.30331)	1.08203 (0.29738)
$\beta = 1.0$	AO	$\hat{\phi} =$ (0.00573)	0.97132 (0.00539)	$\phi y = 0.93981$ (0.01759)	0.97340 (0.01118)	0.94421 (0.01676)	0.94121 (0.01675)
		$\hat{\sigma}_\eta =$ (0.01533)	0.15830 (0.01665)	$\sigma_\eta y = 0.12227$ (0.04752)	0.10422 (0.02071)	0.109045 (0.04725)	0.118556 (0.04561)
		$\hat{\beta} =$ (0.20026)	0.91718 (0.20160)	$\beta y = 0.98213$ (0.10011)	1.16155 (0.12018)	0.984865 (0.10397)	0.98395 (0.10198)
	IO	$\hat{\phi} =$ (0.00578)	0.96808 (0.00533)	$\phi y = 0.97092$ (0.00901)	0.98187 (0.00667)	0.97107 (0.00917)	0.97077 (0.00935)
		$\hat{\sigma}_\eta =$ (0.01684)	0.18967 (0.01699)	$\sigma_\eta y = 0.17845$ (0.03923)	0.14805 (0.02269)	0.15658 (0.04073)	0.17748 (0.04128)
		$\hat{\beta} =$ (0.24432)	1.12257 (0.23783)	$\beta y = 1.07745$ (0.32515)	1.27843 (0.37520)	1.08392 (0.32388)	1.08416 (0.32811)

Tabela 6.6: Estimativas dos parâmetros para as séries simuladas e abaixo seus respectivos erros padrão Caso I.

- Pode-se observar através das tabelas 6.6, 6.7, 6.8 e 6.9 que as técnicas utilizadas para estimação da logverossimilhança e conseqüentemente as estimativas dos parâmetros do modelo, produzem resultados semelhantes. Embora não sabendo o valor verdadeiro para a EMV, nota-se que as estimativas para o parâmetro de persistência ϕ foi estimado próximo ao valor do coeficiente do processo gerador de dados, quando não existe a influência de valores aberrantes.

- A presença de *outliers* do tipo AO e do tipo IO na série tende a subestimação do parâmetro de persistência, por outro lado as estimativas de σ_η aumentam. Essa influência parece ser causada principalmente pela presença do IO, porque pode-se observar que na presença apenas do AO são obtidas estimativas mais próximas àquelas obtidas sem presença de valores espúrios.
- Através dos resultados pode-se notar que o fator de escala constante β é pouco influenciado pela presença de *outliers* do tipo AO, por outro lado valores aberrantes com essa característica tendem a aumentar as estimativas obtidas para σ_η .

Θ	Tipo Outlier	Clássica		Bayesiana			
		Durbin Koopman	Shephard Pitt	Amostragem Simples	Mistura	Integração	Reponderação
$\phi = 0.95$	Sem Outlier	$\hat{\phi} = 0.95170$ (0.01678)	0.95189 (0.01714)	$\phi y = 0.94416$ (0.01621)	0.96725 (0.01231)	0.94400 (0.01627)	0.94383 (0.01652)
		$\hat{\sigma}_\eta = 0.085612$ (0.01341)	0.08513 (0.01532)	$\sigma_\eta y = 0.13176$ (0.04563)	0.12612 (0.02420)	0.12937 (0.04472)	0.13241 (0.04611)
		$\hat{\beta} = 0.908878$ (0.08882)	0.90895 (0.09002)	$\beta y = 1.02835$ (0.12296)	1.20656 (0.14036)	1.02747 (0.11879)	1.02766 (0.11799)
	$\sigma_\eta = 0.1$	AO IO	$\hat{\phi} = 0.92839$ (0.01522)	0.92386 (0.01544)	$\phi y = 0.93760$ (0.01665)	0.95804 (0.01337)	0.93699 (0.01648)
$\hat{\sigma}_\eta = 0.15692$ (0.01617)			0.16083 (0.01800)	$\sigma_\eta y = 0.17849$ (0.04807)	0.14556 (0.02687)	0.17669 (0.04633)	0.17774 (0.04665)
$\hat{\beta} = 1.00069$ (0.08629)			0.99890 (0.08518)	$\beta y = 1.05252$ (0.13199)	1.23012 (0.14250)	1.05209 (0.12854)	1.05199 (0.12865)
$\beta = 1.0$	AO	$\hat{\phi} = 0.94003$ (0.01505)	0.93943 (0.01545)	$\phi y = 0.93824$ (0.01761)	0.96432 (0.01337)	0.93812 (0.01722)	0.93857 (0.01716)
		$\hat{\sigma}_\eta = 0.11574$ (0.01363)	0.11714 (0.01485)	$\sigma_\eta y = 0.15141$ (0.04920)	0.12955 (0.02648)	0.14852 (0.04699)	0.15055 (0.04698)
		$\hat{\beta} = 0.91080$ (0.08235)	0.91013 (0.08399)	$\beta y = 1.02951$ (0.11935)	1.21012 (0.13770)	1.02854 (0.11687)	1.02859 (0.11627)
	IO	$\hat{\phi} = 0.93502$ (0.01438)	0.93441 (0.01472)	$\phi y = 0.9417$ (0.01559)	0.96144 (0.01259)	0.94144 (0.01573)	0.94336 (0.01586)
	$\hat{\sigma}_\eta = 0.12869$ (0.01385)	0.13018 (0.01483)	$\sigma_\eta y = 0.15756$ (0.04554)	0.14106 (0.02549)	0.15836 (0.04473)	0.15634 (0.04497)	
	$\hat{\beta} = 0.995245$ (0.08737)	0.99453 (0.08769)	$\beta y = 0.97117$ (0.13595)	1.22731 (0.14689)	1.05238 (0.13267)	1.05281 (0.13282)	

Tabela 6.7: Estimativas dos parâmetros para as séries simuladas e seus respectivos erros padrão Caso II.

- O *outlier* do tipo IO, quando inserido no processo gerador de dados, tende a diminuir as estimativas do parâmetro de persistência ϕ e como pode-se notar, contribui com um aumento significativo nas estimativas de σ_η e β .

Θ	Tipo Outlier	Clássica		Bayesiana			
		Durbin Koopman	Shephard Pitt	Amostragem Simples	Mistura	Integração	Reponderação
$\phi = 0.9$	Sem Outlier	$\hat{\phi} =$ 0.90750 (0.01686)	0.90407 (0.01699)	$\phi y =$ 0.89896 (0.02673)	0.93734 (0.01601)	0.90913 (0.01917)	0.90861 (0.01923)
		$\hat{\sigma}_\eta =$ 0.37808 (0.03187)	0.39686 (0.03373)	$\sigma_\eta y =$ 0.65953 (0.08952)	0.38028 (0.03843)	0.61444 (0.05572)	0.62028 (0.05609)
		$\hat{\beta} =$ 0.925072 (0.09913)	0.91940 (0.09725)	$\beta y =$ 0.97191 (0.10984)	1.17370 (0.12538)	0.97592 (0.11256)	0.97389 (0.11234)
$\sigma_\eta = 0.5$	AO IO	$\hat{\phi} =$ 0.89520 (0.01763)	0.89260 (0.01642)	$\phi y =$ 0.88991 (0.03043)	0.92333 (0.01765)	0.89672 (0.02054)	0.89681 (0.02062)
		$\hat{\sigma}_\eta =$ 0.45141 (0.04413)	0.46801 (0.03779)	$\sigma_\eta y =$ 0.70829 (0.10536)	0.43415 (0.03964)	0.67932 (0.05745)	0.68276 (0.05775)
		$\hat{\beta} =$ 0.97178 (0.10200)	0.96641 (0.10068)	$\beta y =$ 1.00010 (0.11201)	1.20270 (0.11819)	0.99705 (0.11206)	0.99608 (0.11112)
$\beta = 1.0$	AO	$\hat{\phi} =$ 0.90369 (0.01749)	0.89996 (0.01667)	$\phi y =$ 0.89433 (0.03012)	0.93273 (0.01660)	0.90498 (0.01955)	0.90418 (0.01966)
		$\hat{\sigma}_\eta =$ 0.40152 (0.03379)	0.42225 (0.03481)	$\sigma_\eta y =$ 0.68352 (0.09590)	0.39984 (0.03892)	0.63773 (0.05630)	0.64353 (0.05672)
		$\hat{\beta} =$ 0.92579 (0.09893)	0.91982 (0.09655)	$\beta y =$ 0.97232 (0.10931)	1.17550 (0.12229)	0.97564 (0.11233)	0.97435 (0.11203)
	IO	$\hat{\phi} =$ 0.89994 (0.01929)	0.89652 (0.01594)	$\phi y =$ 0.89217 (0.017101)	0.92771 (0.01714)	0.90083 (0.01981)	0.90032 (0.01981)
		$\hat{\sigma}_\eta =$ 0.42285 (0.03708)	0.44287 (0.03583)	$\sigma_\eta y =$ 0.69419 (0.099368)	0.41600 (0.03933)	0.65664 (0.05742)	0.66242 (0.05812)
		$\hat{\beta} =$ 0.97175 (0.10297)	0.96571 (0.10134)	$\beta y =$ 0.99798 (0.11154)	1.20040 (0.12023)	0.99999 (0.11363)	0.99768 (0.11365)

Tabela 6.8: Estimativas dos parâmetros para as séries simuladas e seus respectivos erros padrão Caso III.

- Como na análise clássica do modelo, as estimativas bayesianas também produziram estimativas pontuais próximas aos valores teóricos. Pode-se verificar pelas tabelas 6.6, 6.7, 6.8 e 6.9 que a técnica de mistura, teve tendência de produzir estimativas para a persistência e para β um pouco maior relativamente aos outros métodos. Por outro lado, as estimativas para σ_η foram menores para esta técnica.
- Ao inserir os valores aberrantes no processo gerador de dados pode-se verificar que as estimativas bayesianas tiveram comportamento análogo aos valores obtidos através da análise frequentista, mas nota-se que as estimativas de σ_η pela amostragem simples foram mais próximas aos valores da análise frequentista. No entanto, para os outros métodos as estimativas foram maiores que as obtidas pelo método de **DK** e **SP**.
- Pode-se verificar que a presença de um IO na série influencia o erro padrão do parâmetro β , aumentando-o em uma proporção considerável. Por outro lado, apenas dois tipos de *outliers* influenciam nas estimativas do erro padrão para a volatilidade da log-volatilidade σ_η .

Θ	Tipo Outlier	Clássica		Bayesiana			
		Durbin Koopman	Shephard Pitt	Amostragem Simples	Mistura	Integração	Reponderação
$\phi = 0.98$	Sem Outlier	$\hat{\phi} =$ 0.98161 (0.00814)	0.98176 (0.00823)	$\phi y = 0.98061$ (0.00859)	0.98506 (0.01286)	0.98048 (0.00892)	0.98056 (0.00892)
		$\hat{\sigma}_\eta =$ 0.11062 (0.02605)	0.11073 (0.02389)	$\sigma_\eta y = 0.11525$ (0.02194)	0.10719 (0.01253)	0.11402 (0.02271)	0.11426 (0.02253)
		$\hat{\beta} =$ 0.95679 (0.08943)	0.95690 (0.09154)	$\beta y = 0.98133$ (0.11067)	1.13367 (0.08090)	0.98348 (0.11341)	0.98328 (0.11417)
$\sigma_\eta = 0.12$	AO IO	$\hat{\phi} =$ 0.94845 (0.00683)	0.94923 (0.00681)	$\phi y = 0.95286$ (0.01549)	0.97340 (0.01086)	0.95251 (0.01583)	0.95259 (0.01541)
		$\hat{\sigma}_\eta =$ 0.21263 (0.03923)	0.21296 (0.03778)	$\sigma_\eta y = 0.25415$ (0.03756)	0.12224 (0.01974)	0.25310 (0.03981)	0.25553 (0.03776)
		$\hat{\beta} =$ 1.05762 (0.07235)	1.05737 (0.07273)	$\beta y = 1.04986$ (0.10855)	1.19499 (0.12887)	1.05166 (0.10931)	1.05152 (0.11019)
$\beta = 1.0$	AO	$\hat{\phi} =$ 0.97803 (0.00657)	0.97819 (0.00632)	$\phi y = 0.97089$ (0.01352)	0.97320 (0.01429)	0.97560 (0.01177)	0.97129 (0.01349)
		$\hat{\sigma}_\eta =$ 0.12454 (0.02338)	0.12376 (0.02468)	$\sigma_\eta y = 0.15132$ (0.03332)	0.10755 (0.14524)	0.13224 (0.03098)	0.14991 (0.03626)
		$\hat{\beta} =$ 0.96283 (0.08611)	0.96320 (0.08687)	$\beta y = 0.98649$ (0.10081)	1.13903 (0.08075)	0.98756 (0.10575)	0.98701 (0.10176)
	IO	$\hat{\phi} =$ 0.95568 (0.00715)	0.95559 (0.00692)	$\phi y = 0.96311$ (0.01289)	0.97690 (0.00957)	0.96289 (0.01319)	0.96317 (0.01294)
		$\hat{\sigma}_\eta =$ 0.19342 (0.03563)	0.19300 (0.03628)	$\sigma_\eta y = 0.21838$ (0.03338)	0.13000 (0.01782)	0.21660 (0.03512)	0.21807 (0.03426)
		$\hat{\beta} =$ 1.05196 (0.07569)	1.05214 (0.07677)	$\beta y = 1.04551$ (0.12069)	1.13910 (0.13619)	1.04838 (0.12077)	1.04914 (0.12167)

Tabela 6.9: Estimativas dos parâmetros para as séries simuladas e seus respectivos erros padrão Caso IV.

6.2.4 Estimativas da Volatilidade

Nesta seção será realizada a análise das estimativas da volatilidade de duas séries propostas, a série do Caso I e Caso IV. O objetivo é verificar o comportamento das estimativas com a presença de valores aberrantes bem distantes da média dos retornos. As estimativa suavizadas são obtidas através do algoritmo proposto por Jong e Shephard (1995), que encontra-se implementado no pacote SsfPack 2.2 de Koopman *et al.*(1999) que é utilizado através da linguagem OX. As estimativas filtradas foram obtidas utilizando-se o algoritmo do filtro de partículas utilizado por Kim *et al.* (1998), onde é feita a amostragem da densidade de filtragem condicionado aos parâmetros, i.e., assume-se que os parâmetros sejam conhecidos; nesse caso é tomado como estimador de máxima verossimilhança o estimador da média posteriori.

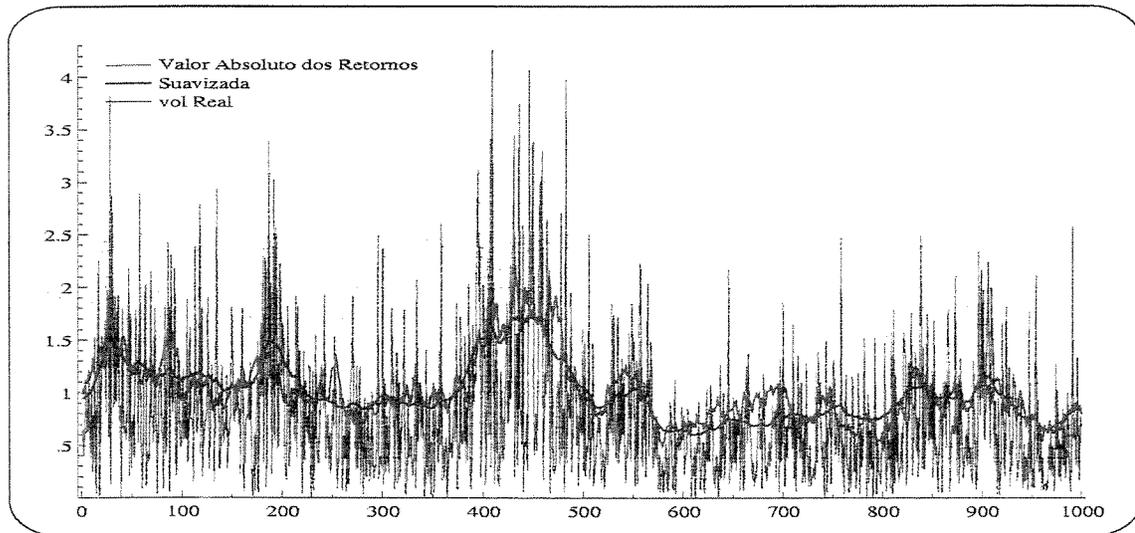


Figura 6.6: Gráfico do valor absoluto dos retornos e da estimativa suavizada da sem outliers para o caso I utilizando a metodologia de Durbin e Koopman.

A figura 6.6 apresenta a volatilidade verdadeira e estimativa suavizada pelo método proposto por Durbin e Koopman (1997a) para as séries sem a presença de *outliers*. As estimativas pelo método proposto por Shephard e Pitt (1997) foram omitidas porque os métodos de estimativas da volatilidade são análogos.

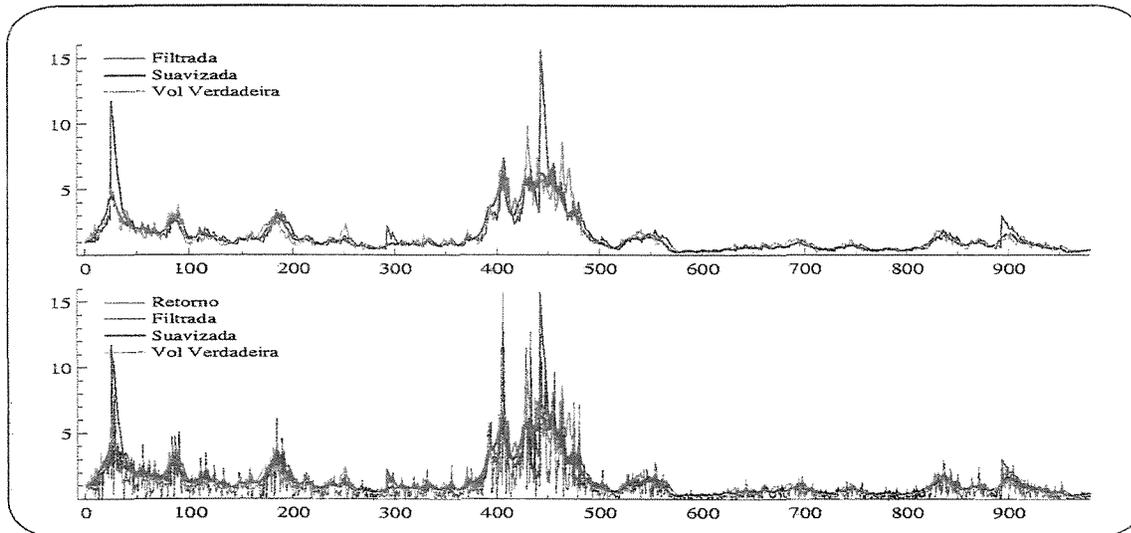


Figura 6.7: Gráfico do valor absoluto dos retornos e da volatilidade verdadeira e estimativa, filtrada e suavizada sem outliers para o caso I utilizando a metodologia de Kim *et al.*.

O gráfico da figura 6.7 apresenta as estimativas da volatilidade através dos métodos propostos por Kim *et al.*. Na figura 6.8 temos a presença de um AO e um IO. Como o comportamento das séries na presença de apenas um AO ou um IO foram iguais quando temos a presença dos dois *outliers*, os gráficos para os casos da ocorrência de um IO e um AO separados foram omitidos.

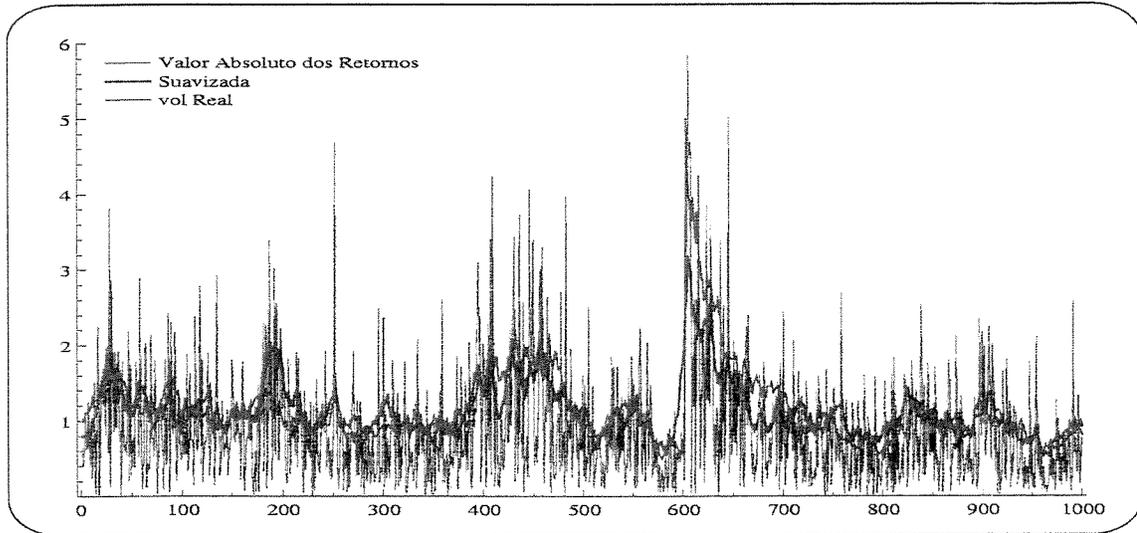


Figura 6.8: Gráfico do valor absoluto dos retornos, da volatilidade verdadeira e suavizada considerando AO na posição $t = 250$ e IO em $t = 600$ para o caso I utilizando o método de Durbin e Koopman.

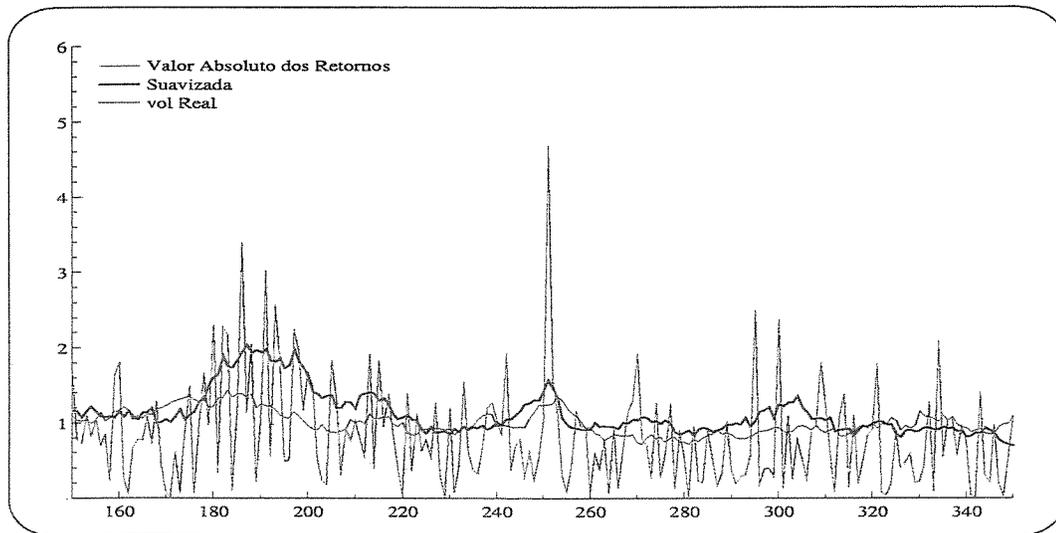


Figura 6.9: Gráfico dos retornos e volatilidades verdadeira e estimativa suavizada próximo ao AO em $t = 250$ para o caso I utilizando o método de Durbin e Koopman.

Pode-se notar através das figuras 6.8, 6.9 e 6.10 que mesmo com a presença dos dois tipos de *outliers* as estimativas da volatilidade acompanham o processo de ascensão e decrescimento do valor absoluto dos retornos. Na figura 6.10 verifica-se que o comportamento da volatilidade estimada permanece quase o mesmo de antes da inclusão do *outlier* do tipo IO. Entretanto, a volatilidade estimada próximo ao IO na figura 6.10 fica abaixo da volatilidade real após a ocorrência do IO e acima da volatilidade real depois da ocorrência do *outlier*.

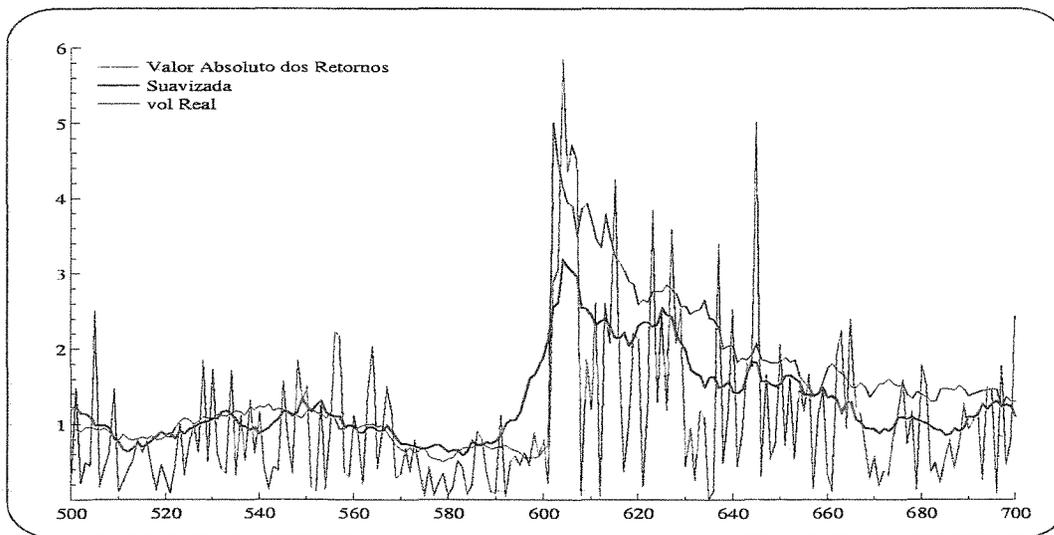


Figura 6.10: Gráfico dos retornos e volatilidades verdadeira e estimativa suavizada próximo ao IO em $t = 600$ para o caso I utilizando o método de Durbin e Koopman.

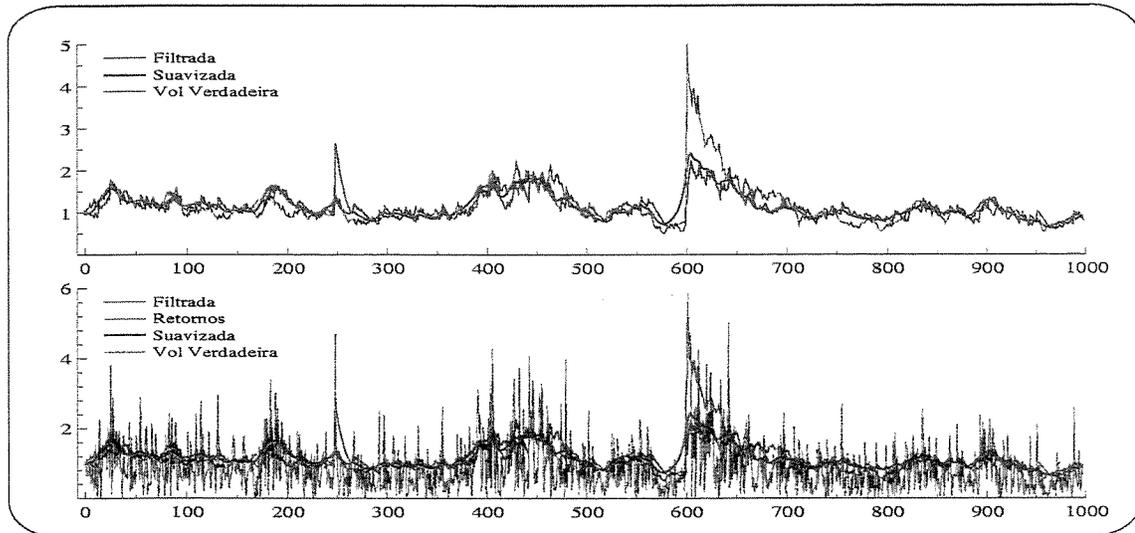


Figura 6.11: Volatilidade verdadeira e estimativas, filtrada e suavizada com *outliers* do tipo AO e IO utilizando o método de Kim *et al.*.

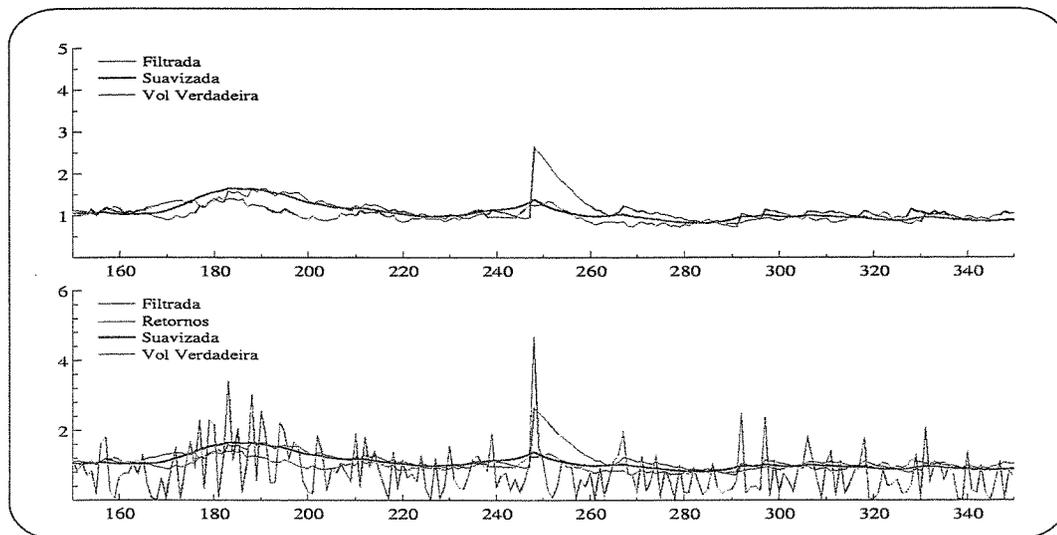


Figura 6.12: Gráfico dos retornos e volatilidades verdadeira e estimativas, filtrada e suavizada próximo ao AO em $t = 250$ utilizando o método de Kim *et al.*.

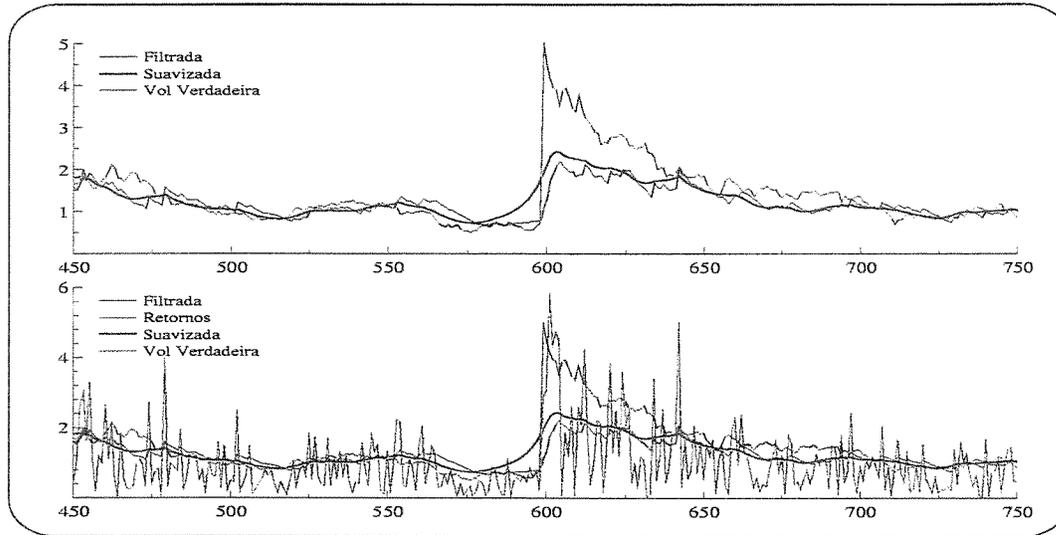


Figura 6.13: Gráfico dos retornos e volatilidades verdadeira e estimativas, filtrada e suavizada próximo ao IO em $t = 600$ para o caso I utilizando o método de Kim *et al.*

Os gráficos da figura 6.11 apresentam as volatilidades filtrada, suavizada e a volatilidade verdadeira para o Caso I utilizando os métodos propostos por Kim *et al.*, onde pode-se verificar o comportamento das estimativas obtidas.

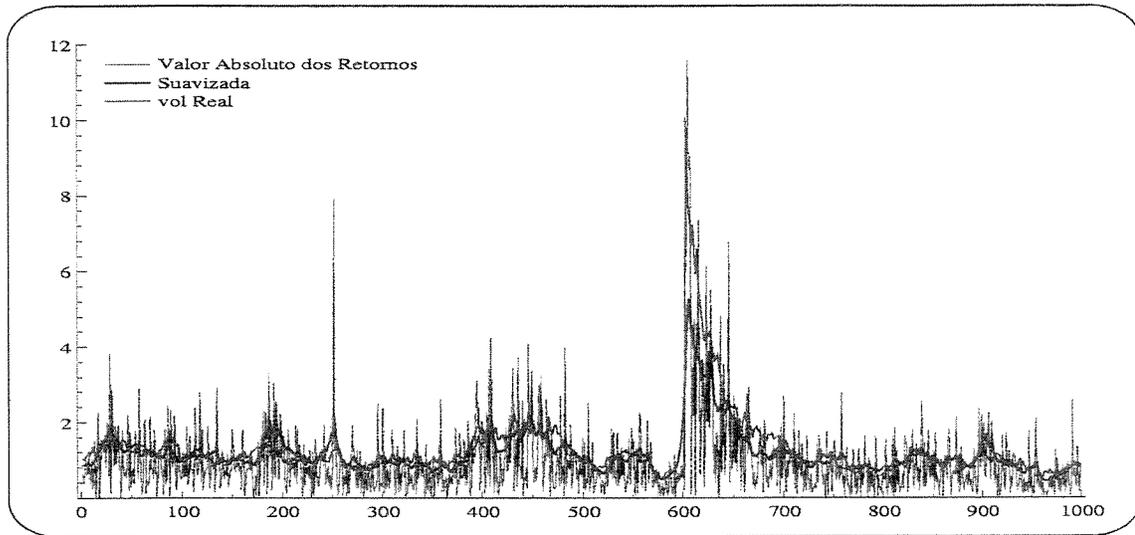


Figura 6.14: Volatilidade verdadeira e estimativa suavizada com *outliers* do tipo AO e IO para o caso IV utilizando método proposto por Durbin e Koopman.

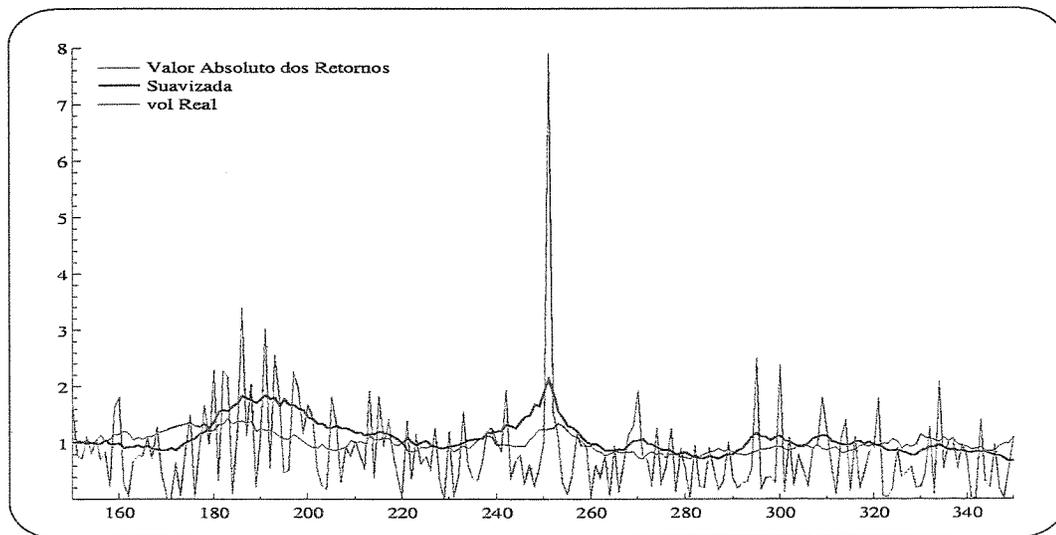


Figura 6.15: Gráfico dos retornos e volatilidades verdadeiras e estimativas suavizadas próximo ao AO em $t = 250$ para o caso IV utilizando método proposto por Durbin e Koopman.

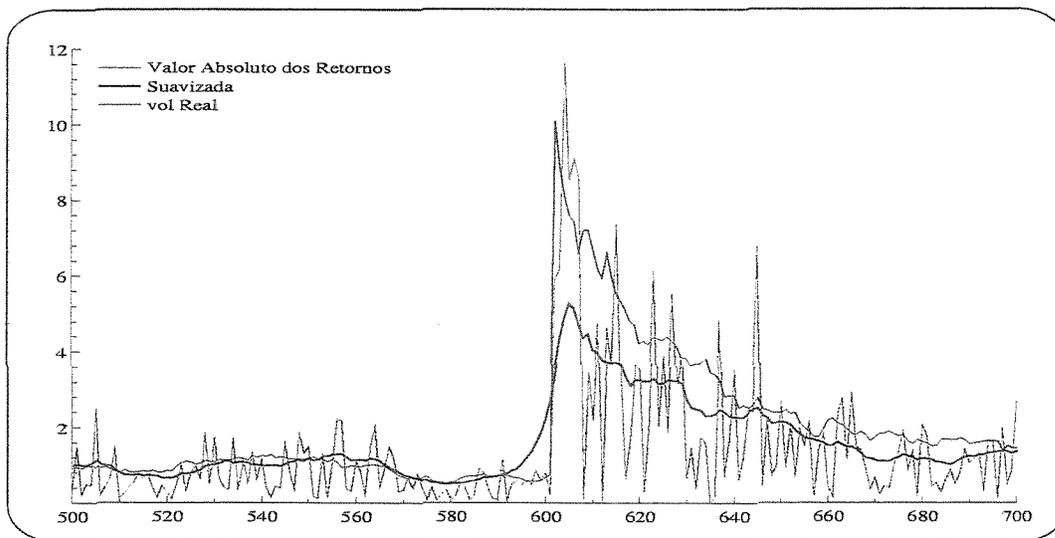


Figura 6.16: Gráfico dos retornos e volatilidades verdadeira e estimativa suavizada próximo ao IO em $t = 600$ para o caso IV utilizando método proposto por Durbin e Koopman.

A série da figura 6.14 foi simulada com o conjunto de parâmetros do Caso IV, com objetivo de verificar o comportamento dos métodos quando tem-se valores espúrios bastante distante da média.

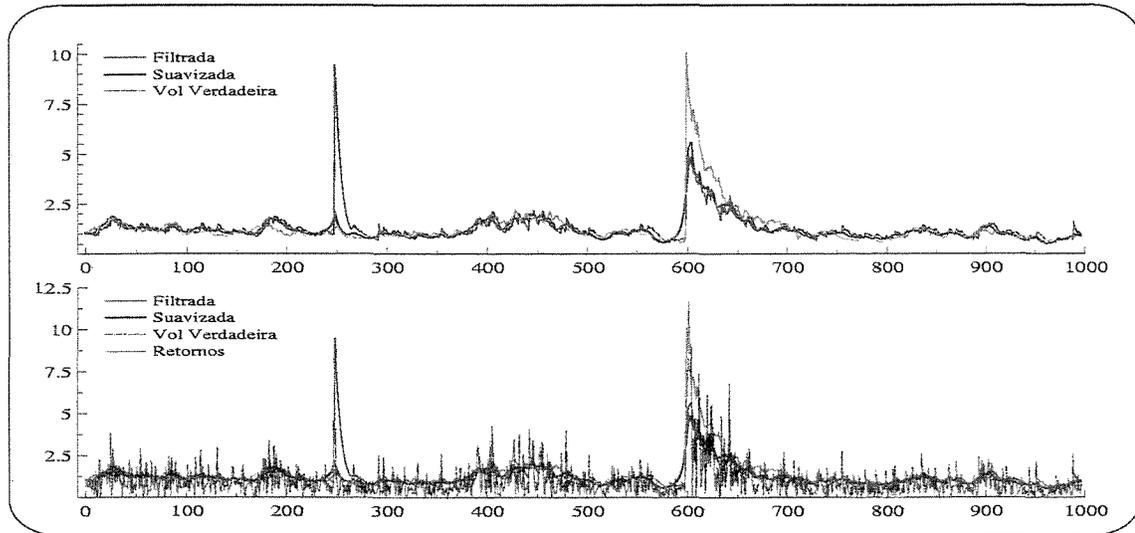


Figura 6.17: Volatilidade verdadeira e estimativas, filtrada e suavizada com *outliers* do tipo AO e IO para o caso IV utilizando método proposto por Kim *et al.*.

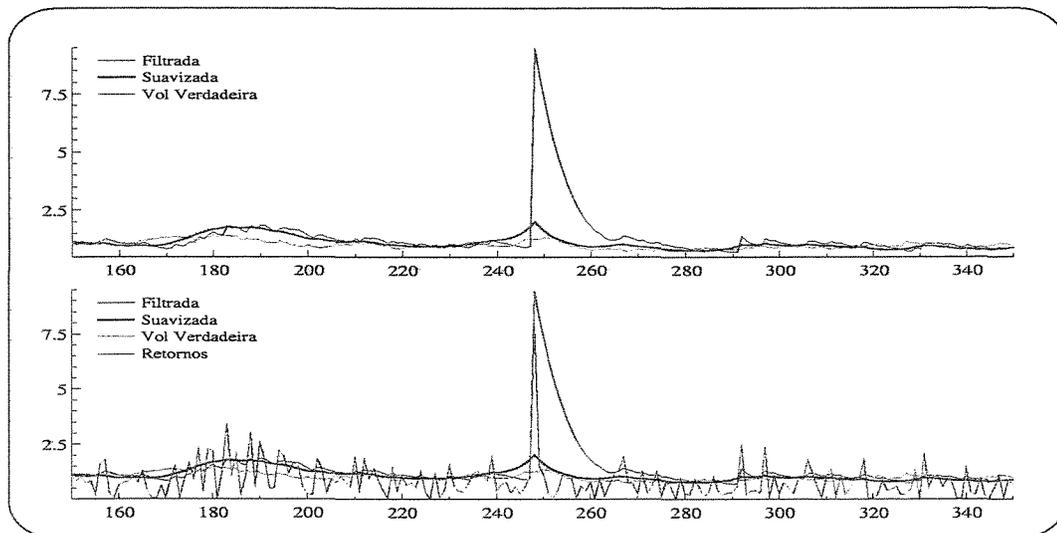


Figura 6.18: Gráfico dos retornos e volatilidades verdadeira e estimativas, filtrada e suavizada próximo ao AO em $t = 250$ para o caso IV utilizando método proposto por Kim *et al.*.

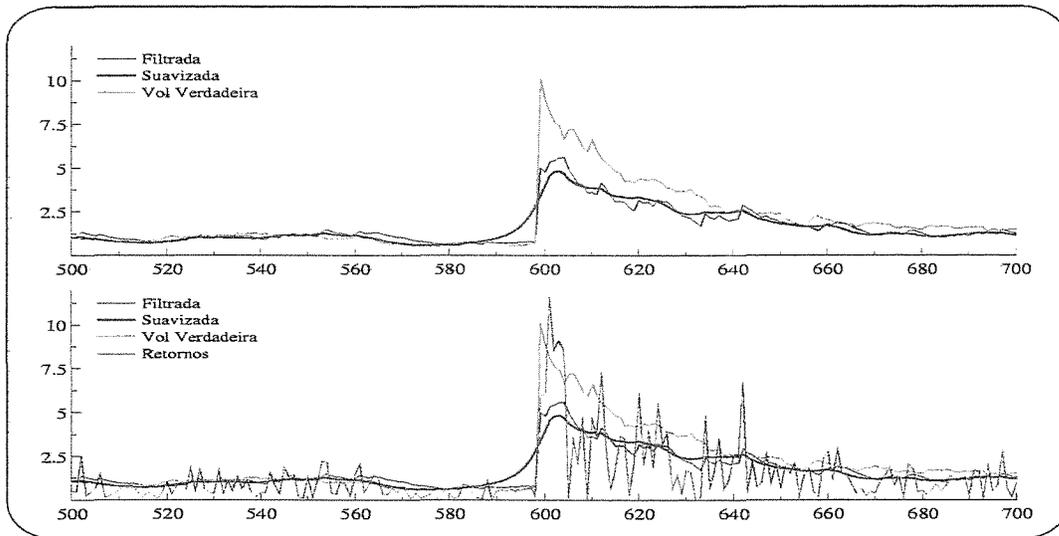


Figura 6.19: Gráfico dos retornos e volatilidades verdadeira e estimativas, filtrada e suavizada próximo ao IO em $t = 600$ para o caso IV utilizando método proposto por Kim *et al.*.

Pode-se observar que as estimativas suavizadas conseguem acompanhar o movimento da volatilidade verdadeira, mesmo próximo dos valores aberrantes. No entanto, quando na presença do IO a volatilidade estimada é superestimada antes da ocorrência do IO e subestimada após a ocorrência.

6.3 Análise da Série IBOVESPA

Nesta seção é apresentada a análise da série do Índice da Bolsa de Valores do Estado de São Paulo (IBOVESPA). Os retornos são do período de 03 de Janeiro de 1995 a 27 de Dezembro de 2000, num total de 1500 observações. Durante este período ocorreram várias crises em outros países que influenciaram o mercado financeiro brasileiro. As principais crises foram: crise do México em Fevereiro e Março de 1995, crise na Ásia em Outubro de 1997, moratória da Rússia em Agosto de 1998, crise do Brasil (desvalorização do real) em Janeiro de 1999 e Nasdaq em Abril de 2000. Os retornos corrigidos da série foram calculados da seguinte forma¹.

$$y_t = 100 \times \{(\log r_t - \log r_{t-1}) - n^{-1} \sum_{i=1}^n (\log r_i - \log r_{i-1})\}$$

onde n é o número de observações. A tabela 6.10 apresenta os valores das estatísticas descritivas dos retornos e pelo gráfico da figura 6.20 pode-se observar alguns dos fatos estilizados citados no capítulo 3. A série corrigida tem média igual a zero.

Série	IBOVESPA
D. Padrão	2.859335
Assimetria	0.743485
E. Curtose	12.944355
Mínimo	-17.316949
Máximo	28.729927

Tabela 6.10: Estatísticas descritivas dos retornos do IBOVESPA

Através do histograma da figura 6.20 pode-se verificar que a distribuição dos retornos possuem caudas mais pesadas que a da distribuição normal e alguns valores que se

¹Seguindo sugestão dada por Kim *et al.*(1998)

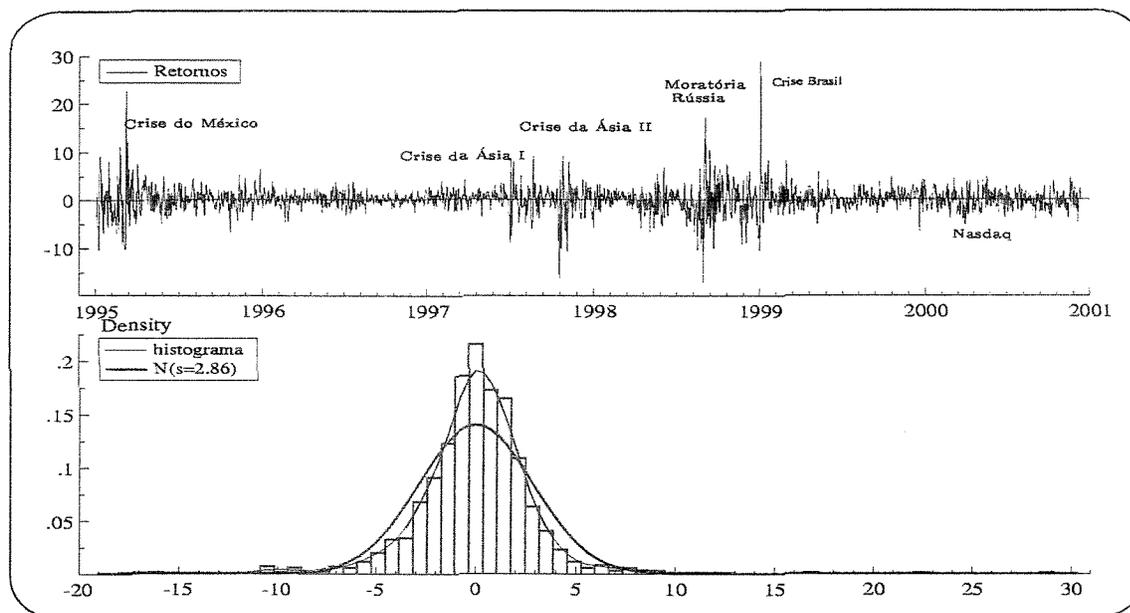


Figura 6.20: Gráfico dos retornos e histograma do IBOVESPA

encontram distantes da média. Os conglomerados de volatilidade podem ser verificados através do gráfico de retornos, onde os períodos de alta volatilidade coincide com os períodos onde ocorreram as crises citadas no início da seção.

Série	Clássica		Bayesiana			
	Durbin Koopman	Shephard Pitt	Amostragem Simples	Mistura	Integração	Reponderação
IBOVESPA	$\hat{\phi} =$ 0.97329 (0.00754)	0.97357 (0.00756)	$\phi y =$ 0.97539 (0.00812)	0.97856 (0.00745)	0.97515 (0.00802)	0.97539 (0.00885)
	$\hat{\sigma}_\eta =$ 0.052320 (0.02631)	0.05179 (0.02624)	$\sigma_\eta y =$ 0.04923 (0.02647)	0.03846 (0.02432)	0.04931 (0.02567)	0.048325 (0.02521)
	$\hat{\beta} =$ 2.12143 (0.23446)	2.12132 (0.23587)	$\beta y =$ 2.15320 (0.29552)	2.23360 (0.31471)	2.14450 (0.27148)	2.16880 (0.25195)

Tabela 6.11: Estimativas dos parâmetros para a série IBOVESPA e seus respectivos erros padrão

As estimativas dos parâmetros dadas na tabela 6.11 fornece resultados análogos, com a persistência do modelo proposto em torno de 0.975.

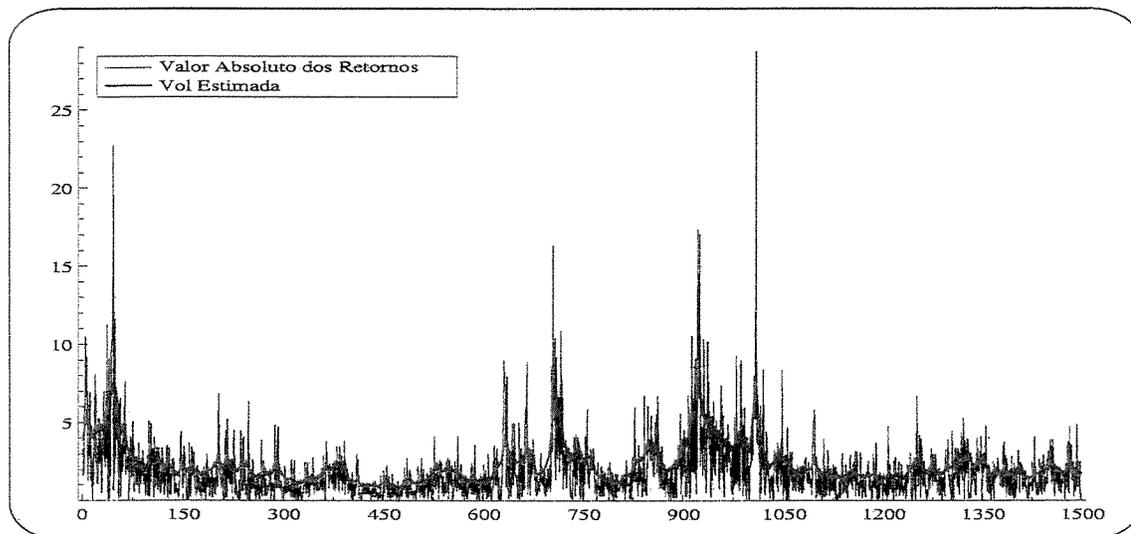


Figura 6.21: Gráfico do valor absoluto dos retornos e da estimativa Suavizada pelo método de Durbin e Koopman.

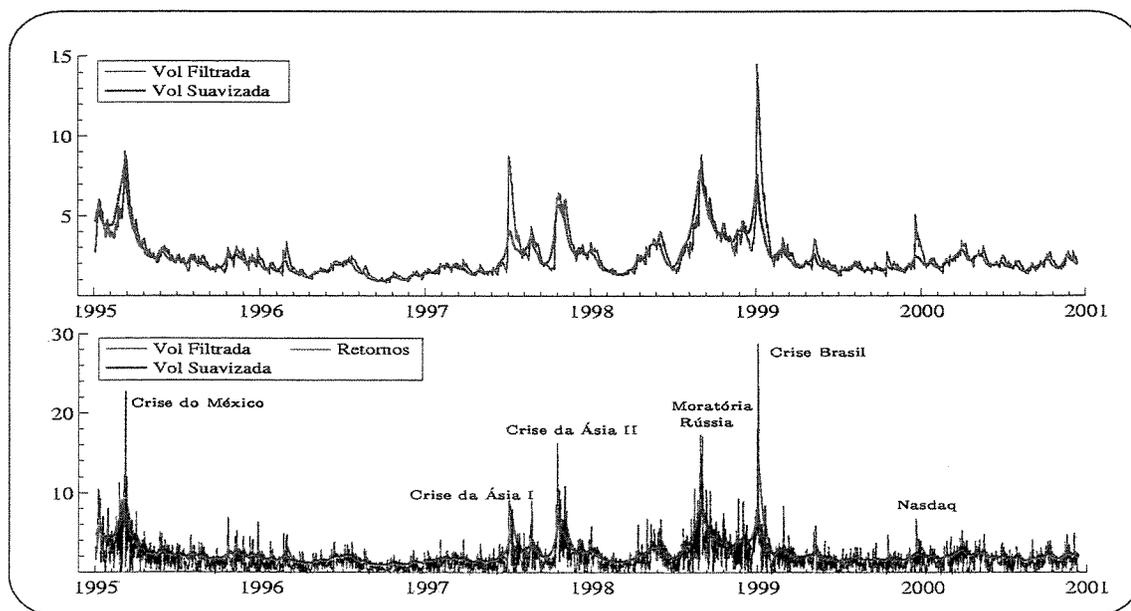


Figura 6.22: Gráfico dos retornos e das estimativas filtradas e suavizadas da volatilidade sobreposto ao do valor absoluto dos retornos utilizando o método de Kim *et al.*

Pode-se verificar pelas figuras 6.21 e 6.22 que boas estimativas da volatilidade foram obtidas pelo método de Durbin e Koopman e Kim *et al.* respectivamente, sendo que essas conseguem captar os períodos de alta e baixa volatilidade nos retornos da série analisada. Os métodos se mostraram equivalentes a respeito das estimativas dos parâmetros e volatilidade.

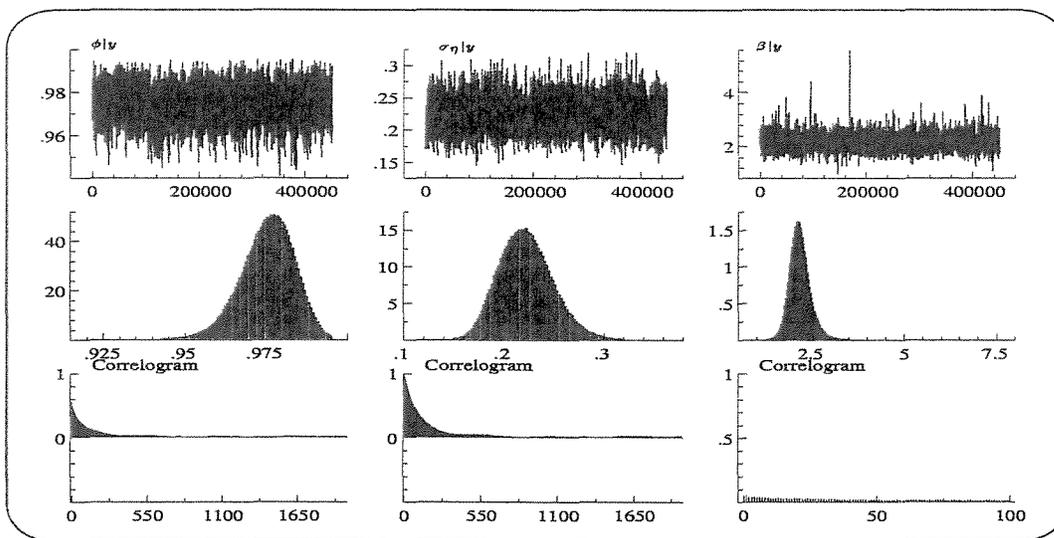


Figura 6.23: Gráficos resultante das simulações usando amostragem simples. Na primeira linha as iterações, na segunda os histogramas das distribuições a posterioris e por último o correlograma da amostra.

Nota-se pelos gráficos das figuras 6.23, 6.24 e 6.25 que o comportamento das estimativas dos parâmetros foi o mesmo observado nas séries simuladas, ou seja, a amostragem simples possui a desvantagem de taxa de convergência baixa, necessitando-se assim um número de iterações muito grande para obter estimativas razoáveis. Por outro lado, utilizando-se o algoritmo de integração consegue-se bons resultados e baixa correlação das iterações com apenas 2000 iterações.

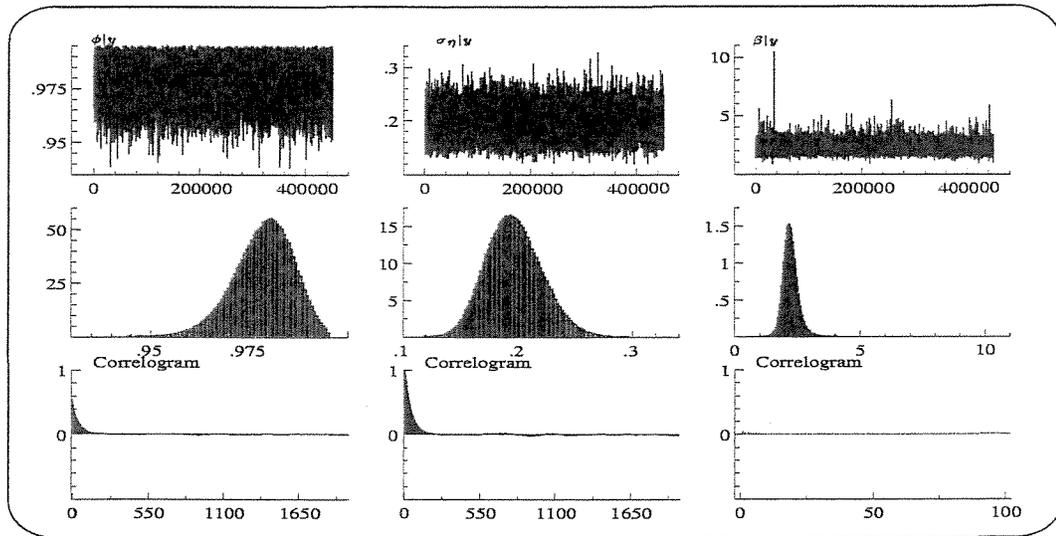


Figura 6.24: Gráficos resultante das simulações usando algoritmo de mistura. Na primeira linha as iterações, na segunda os histogramas das distribuições a posterioris e por último o correlograma da amostra.

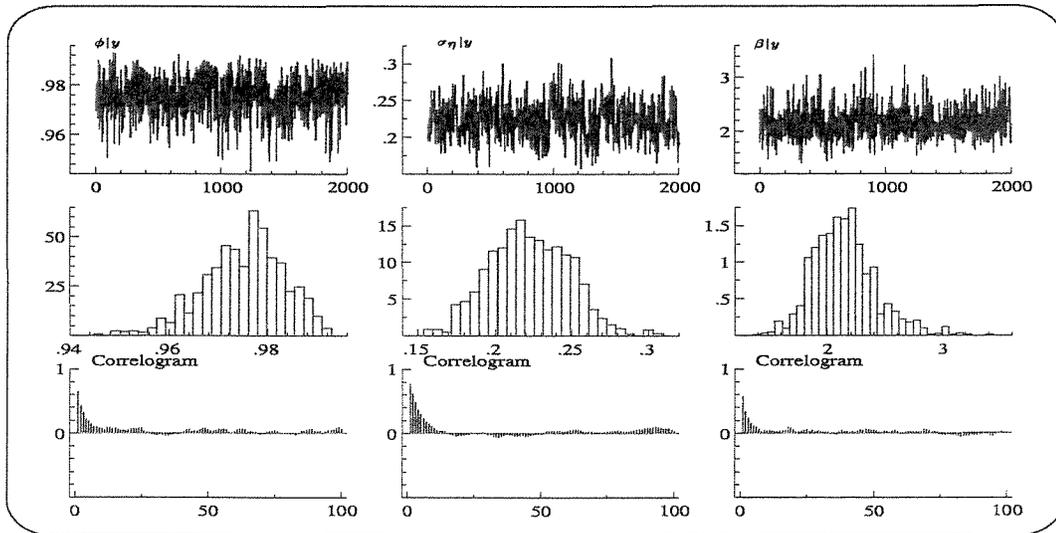


Figura 6.25: Gráficos resultante das simulações usando algoritmo de integração. Na primeira linha as iterações, na segunda os histogramas das distribuições a posterioris e por último o correlograma da amostra.

6.4 Conclusão

Neste trabalho foi analisado o modelo na forma de espaço de estado não Gaussiano, em particular o modelo de volatilidade estocástica, com enfoque bayesiano e clássico. Dentro da abordagem frequentista foram implementados os métodos propostos por Durbin e Koopman (1997a, 2000) e Shephard e Pitt (1997), entre esses métodos não houve divergência na estimação de parâmetros, como pode ser observado nas tabelas 6.6, 6.7, 6.8 e 6.9 a diferença foi por volta da terceira casa decimal. O comportamento na presença de *outliers* foi análogo considerando os dois métodos de estimação. A diferença entre os métodos se deu no tempo computacional onde o método de Durbin e Koopman exigiu maior tempo computacional nas simulações, 4 minutos aproximadamente para estimativas e o método proposto por Shephard e Pitt obteve as estimativas em aproximadamente 3 minutos.

A abordagem bayesiana foi apresentada através dos métodos proposto por Jacquier *et al.*(1994), Shephard e Pitt (1997) e Kim *et al.*(1998). dentre esses métodos, o artigo de Kim *et al.*, onde temos três métodos para tornar mais eficiente a amostragem, melhora significativamente os resultados obtidos pelos métodos até então propostos no sentido de tempo de convergência e correlação das amostras.

O método proposto por Jacquier *et al.*, apesar de eficiente, tem convergência lenta, Kim *et al.* através da utilização de mistura, integração e amostragem conjunta dos parâmetros, diminuíram a correlação entre as iterações na amostragem dos parâmetros e assim pode-se diminuir o número de amostras necessárias para fazermos inferências. As duas abordagens produzem boas estimativas dos parâmetros e volatilidades para o modelo de volatilidade estocástica. Computacionalmete a abordagem clássica é de menor custo e de maior facilidade para implementação, pois utiliza-se de técnicas

tradicionais de simulação. Sendo o custo computacional crucial na escolha de uma técnica a ser utilizada em problemas práticos, principalmente tratando-se de mercado financeiro, onde muitas vezes é necessário respostas rápidas e precisas, a abordagem clássica seria preferível a abordagem bayesiana.

O tempo necessário para as estimativas torna muito caro computacionalmente a aplicação de algumas metodologias encontradas na literatura que melhoram as estimativas obtidas, por exemplo, o método de detecção de outlier através de varredura discutido em Fukui (2000). Neste método é necessário a estimação dos parâmetros do modelo de VE a cada *outlier* detectado.

As técnicas que utilizam MCMC para estimação dos parâmetros tem evoluído rapidamente, tornando o enfoque bayesiano cada vez mais "competitivo" para aplicação em problemas onde existe a necessidade de respostas rápidas. Junto com o desenvolvimento das técnicas existentes, o avanço tecnológico com computadores cada vez mais velozes torna promissor a utilização de tais técnicas.

APÊNDICE A

Métodos de Simulação

Neste apêndice serão apresentadas algumas técnicas de simulação utilizadas neste trabalho. Nas seções A.1 e A.2 serão apresentados alguns métodos para gerar amostras de uma distribuição $\pi(x)$, quando não existe método eficiente para amostrar destas distribuições diretamente. Na seção A.3 são apresentadas algumas técnicas de redução de variância presentes na literatura, que tem com objetivo tornar os métodos de simulação mais eficientes.

A.1 Método de Aceitação-Rejeição

O objetivo é gerar amostras de uma densidade contínua $\pi(x) = f(x)/K$, onde $x \in \mathfrak{R}^d$, $f(x)$ é uma densidade não normalizada, e K é uma constante normalizadora. Seja $h(x)$ uma densidade da qual pode-se amostrar mais facilmente, e suponha que exista algum c tal que $f(x) \leq c h(x)$ para todo x . Então, para obter uma amostra de $\pi(\cdot)$,

(i) Faça:

Gere um candidato y de $h(\cdot)$

Gere um valor u de $\mathcal{U}(0, 1)$

(ii) Se $u \leq \frac{f(y)}{c h(y)}$

Devolve $x = y$.

(iii) Caso contrário

Volte a (i)

O valor retornado x é da densidade $\pi(x) = f(x)/K$. Essa proposição assegurada é pelo seguinte teorema.

Teorema A.1.1. *Seja X uma variável aleatória distribuída com densidade $\pi(x)$, $x \in I$, a qual é representada como*

$$\pi(x) = c h(x) \quad (\text{A.1.1})$$

Prova: Pela fórmula de Bayes, temos

$$\pi(y \leq x | Y \text{ Aceito}) = P(Y \leq x | U \leq \frac{\pi(Y)}{c h(Y)}) = \frac{P(U \leq \frac{\pi(Y)}{c h(Y)} | Y = x) h(x)}{P(U \leq \frac{\pi(Y)}{c h(Y)})} \quad (\text{A.1.2})$$

$$P(U \leq \frac{\pi(Y)}{c h(Y)} | Y = x) = P(U \leq \frac{\pi(x)}{c h(x)}) = \frac{\pi(x)}{c h(x)} \quad (\text{A.1.3})$$

$$P(U \leq \frac{\pi(Y)}{c h(Y)}) = \int P(U \leq \frac{\pi(x)}{c h(x)}) h(x) dx = \frac{1}{c} \int \pi(x) dx = \frac{1}{c} \quad (\text{A.1.4})$$

Substituindo (A.1.3) e (A.1.4) em (A.1.2) temos o resultado desejado. \square

Para tornar o método mais eficaz, toma-se a constante $c = \sup_x \frac{f(x)}{h(x)}$. Uma técnica onde, não é necessário encontrar um valor para c e uma escolha eficiente para $h(\cdot)$ será visto mais adiante.

A.2 Método Monte Carlo via Cadeia de Markov

Monte Carlo via Cadeia de Markov é, genericamente, integração de Monte Carlo usando cadeias de Markov. A necessidade de integrar densidades de alta dimensão, tanto na abordagem bayesiana como frequentista a necessidade existe, de um lado integrando densidades posterioris dos parâmetros do modelo condicionados aos dados e na análise clássica integrando a distribuição das observações condicionadas aos parâmetros. Integração por Monte Carlo retira amostras de uma distribuição e estima a esperança através da média amostral. MCMC retira tais amostras iterando uma cadeia de Markov por um tamanho de iterações suficientemente grande. Nesta seção serão apresentadas alguns métodos de se construir as cadeias de Markov¹.

A.2.1 O Algoritmo de Metropolis-Hastings

O algoritmo proposto por Metropolis *et al.*(1953), e generalizado mais tarde por Hastings (1970), é um poderoso método de cadeia de Markov para simular distribuições multivariadas. Para o algoritmo de Metropolis-Hastings, em cada tempo t ,

¹Para maiores detalhes ver (Geman e Geman, 1984) e (Hastings, 1970)

o próximo estado x_{t+1} é escolhido amostrando um candidato y de uma densidade proposta $q(\cdot|x_t)$, i.e., a densidade proposta pode depender do valor de x_t . Assim, o candidato y é aceito com probabilidade

$$\alpha(x, y) = \min \left\{ \frac{\pi(y)q(x|y)}{\pi(x)q(y|x)} \right\}.$$

se o candidato é aceito, o próximo estado será $x_{t+1} = y$, caso contrário $x_{t+1} = x_t$. Assim o algoritmo é dado como segue

(i) Inicialize x_0 .

REPITA {

(ii) Gere um candidato y usando a densidade de transição $q(\cdot, x_t)$ e faça $x_{t+1} = y$ com probabilidade

$$\alpha(x, y) = \min \left\{ \frac{\pi(y)q(x_t|y)}{\pi(x_t)q(y|x_t)}, 1 \right\}. \quad (\text{A.2.1})$$

(iii) Incremente t .

}

Para o algoritmo ser funcional temos de sugerir uma densidade proposta de transição $q(x, y)$. Chib e Greenberg (1995) resumiram cinco propostas.

1. Metropolis *et al.*(1953) sugeriram a família de densidades chamada *random walk*, que é dada por $q(x, y) = q_1(y - x)$. Note que quando q_1 é simétrica $q_1(z) = q_1(-z)$ e a probabilidade de mover é dada por

$$\alpha(x, y) = \min \left\{ \frac{\pi(y)}{\pi(x)}, 1 \right\}.$$

2. Uma segunda família introduzida por Hastings (1970), chamada de *cadeia independente* é dada por $q(x, y) = q_2(y)$.
3. A terceira densidade proposta por Chib e Greenberg (1995) é obtida explorando explorar a forma, se conhecida, de $\pi(\cdot)$. Por exemplo, se $\pi(z)$ pode ser escrita como $\pi(z) \propto g(z)h(z)$, onde $h(z)$ é uma densidade da qual pode-se amostrar e $g(z)$ é uniformemente limitada, então faça $q(x, y) = h(y)$ (como em 2), como densidade proposta. Assim a probabilidade de mover será dada por

$$\alpha(x, y) = \min \left\{ \frac{g(y)}{g(x)}, 1 \right\} \quad (\text{A.2.2})$$

4. O quarto método desenvolvido por Tierney (1994) utiliza o método A-R com uma densidade pseudo-dominada².
5. A última família de densidades, que também foi sugerida por Tierney (1994) é a *cadeia autoregressiva*. Essa é formada fazendo $q(x, y) = q_3\{y - a - B(x - a)\}$, onde a é um vetor e B é uma matriz. Assim tem-se que $y = a + B(x - a) + \varepsilon$, onde ε tem q_3 como densidade.

O quarto método foi utilizado por Jacquier *et al.*(1994) em sua análise do modelo de volatilidade estocástica **VE**. Uma razão para o uso de tal procedimento, no lugar de utilizar Metropolis, é para a redução da autocorrelação integrada (ineficiência devido as amostras dependentes).

²Não consegue-se função que a domine por completo

A.2.2 O Amostrador de Gibbs

O amostrador de Gibbs sugerido por Geman e Geman (1984), é um algoritmo que gera uma cadeia de Markov tal que a distribuição estacionária é a distribuição conjunta $\pi(x_1, \dots, x_N)$. Para aplicar o algoritmo, supõe-se que consiga gerar amostras das condicionais $\pi(x_i | x_{\setminus i})$, i.e, onde $x_{\setminus i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_N)$.

Por exemplo, considere o caso bivariado com densidade conjunta $\pi(x, y)$. Suponha as condicionais $\pi(x|y)$ e $\pi(y|x)$ sejam conhecidas e se possa gerar amostras aleatórias de tais distribuições. O algoritmo é dado com segue

- (i) Inicialize $x = x^{(0)}$,
- (ii) Dado $x^{(i-1)}$, gere $y^{(i)}$ de $\pi(y|x^{(i-1)})$
- (iii) Dado $y^{(i)}$, gere $x^{(i)}$ de $\pi(x|y^{(i)})$
- (iv) Repita o procedimento para $i = 1, \dots, N$

Pode-se provar que sob certas condições de regularidades, quando $N \rightarrow \infty$, $(x^{(N)}, y^{(N)})$ é uma amostra aleatória da densidade conjunta $\pi(x, y)$.

O amostrador de Gibbs pode convergir muito vagarosamente caso exista alta correlação entre as variáveis. Isto ocorre, por exemplo, no modelo de volatilidade estocástica quando existe alta persistência, o que ocorre normalmente em séries empíricas.

A.2.3 Amostragem por Rejeição Adaptativa com Metropolis Hastings Amostrador de Gibbs

A amostragem por rejeição adaptativa é implementada construindo uma função envelope do log da densidade alvo, o qual é então usada no algoritmo A-R Ripley (1987).

Sempre que um ponto é rejeitado, o envelope é atualizado para ficar mais próximo da verdadeira log-densidade reduzindo-se a chance de rejeitar pontos subsequentes. Na formulação original deste método, o envelope é construído a partir de um conjunto de retas tangentes à log-densidade, Gilks e Wild (1992).

Como esses métodos estavam reritos a log concavidade da densidade, Gilks *et al.*(1995) propuseram um novo algoritmo denominado **ARMS** . O **ARMS** lida com a não log concavidade fazendo um passo do algoritmo de M-H em cada ponto aceito no passo de rejeição ARS. No passo de M-H, o novo ponto pesa novamente os pontos amostrados anteriormente. Se o novo ponto é rejeitado, o ponto anterior é o novo ponto. O procedimento é exato, de modo que retorna valores da densidade exata, independentemente do grau de convexidade da log densidade. Quando a densidade é log-côncava o passo de M-H nunca é rejeitado. O algoritmo é dado como segue:

- (i) Inicialize n e X_n independentemente de X_{cur} ,
- (ii) Amostre X de $g_n(x)$,
- (iii) Amostre U da $\mathcal{U} \sim (0, 1)$,
- (iv) Se $U > \frac{f(X)}{\exp h_n(X)}$ então { Passo de rejeição ARS:
 - Faça $X_{n+1} = X_n \cup \{X\}$;
 - $n = n + 1$ e volte a (i) }
 caso contrário { Passo de aceitação de ARS:
 - Faça $X_A = X$ };
- (v) Amostre $U \sim \mathcal{U}(0, 1)$;
- (vi) Se $U > \min \left[\frac{f(X_A) \min\{f(X_{cur}), \exp h_n(X_{cur})\}}{f(X_{cur}) \min\{f(X_A), \exp h_n(X_A)\}}, 1 \right]$ então { Passo de rejeição do M-H:

Faça $X_M = X_{cur}$ }

caso contrário { Passo de aceitação do M-H:

Faça $X_M = X_A$ } ;

(vii) Devolve X_M

A.3 Integração por Monte Carlo

Muitas técnicas eficientes de integração são utilizadas atualmente. Um dos métodos inicialmente representam integrais que são intratáveis analiticamente por meio de valores esperados. Assim, o problema de integração passa a ser um problema de estimação de um parâmetro desconhecido por simulação de Monte Carlo.

Considere o cálculo da seguinte integral

$$\theta = \int_a^b g(x) dx \quad (\text{A.3.1})$$

Assumindo uma densidade $\pi(x)$, tem-se

$$\theta = \int_a^b \frac{g(x)}{\pi(x)} \pi(x) dx = E \left[\frac{g(X)}{\pi(X)} \right] \quad (\text{A.3.2})$$

Assim podemos estimar a integral, amostrando x_1, \dots, x_N de $\pi(x)$ e calculando

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \frac{g(x_i)}{\pi(x_i)} \quad (\text{A.3.3})$$

Observe que cálculos de distribuições marginais também podem ser considerados como cálculo de esperanças. Algumas técnicas para reduzir a variância desse estimador serão apresentadas a seguir.

A.3.1 Amostragem por Importância

Consideramos aqui o problema de estimar uma integral.

$$\theta = \int g(x) dx, \quad x \in D \subset \mathbb{R}^d \quad (\text{A.3.4})$$

A idéia básica de Amostragem por Importância consiste de concentrar a distribuição dos pontos amostrais nas partes da região D que são as áreas de mais importância.

Dessa forma evitamos amostrar pontos com probabilidade baixa de ocorrência, diminuindo assim a variabilidade da amostra. Assim, pode-se representar a integral como

$$\theta = \int \frac{g(x)}{p_I(x)} p_I(x) dx \quad (\text{A.3.5})$$

$$= E \left\{ \frac{g(X)}{p_I(X)} \right\}. \quad (\text{A.3.6})$$

a função $p_I(\cdot)$ é a densidade importante, da qual retiramos amostras x_1, \dots, x_N e estimamos a integral pela média amostral dada por

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^N \frac{g(x_i)}{p_I(x_i)} \quad (\text{A.3.7})$$

A.3.2 Variáveis Antitéticas

Nesta técnica utiliza-se dois estimadores não viciados $\hat{\theta}$ e $\tilde{\theta}$ com correlação negativa para um parâmetro desconhecido. Note que $\frac{1}{2}(\hat{\theta} + \tilde{\theta})$ será um estimador não viciado para a integral θ com varância dada por,

$$\text{var}\left\{\frac{1}{2}(\hat{\theta} + \tilde{\theta})\right\} = \frac{1}{4}\text{var}\{\hat{\theta}\} + \frac{1}{4}\text{var}\{\tilde{\theta}\} + \frac{1}{2}\text{cov}\{\hat{\theta}, \tilde{\theta}\}, \quad (\text{A.3.8})$$

Assim, se a covariância for negativa temos redução na variância do estimador obtido. Uma forma padrão de se obter a redução de variância, é gerar $\hat{\theta}$ e $\tilde{\theta}$ por inversão, i.e., $\hat{\theta} = F^{-1}(U)$ e $\tilde{\theta} = F^{-1}(1 - U)$ onde $U \sim \mathcal{U}(0, 1)$. Pode-se provar que $\text{corr}(\hat{\theta}, \tilde{\theta}) < 0$ (ver Ripley, 1987 p.122-123).

A.3.3 Variáveis Controle

Uma variável $\check{\theta}$ é uma variável controle para o estimador não viciado $\hat{\theta}$ se ela é correlacionada com $\hat{\theta}$ e sua esperança é conhecida. A variável controle é usada para encontrar um estimador para θ que tenha variância menor que o estimador dado por $\hat{\theta}$. Então para algum β

$$\hat{\theta}(\beta) = \hat{\theta} - \beta[\check{\theta} - E(\check{\theta})]$$

é um estimador não viciado para θ , tal que

$$\text{var}[\hat{\theta}(\beta)] = \text{var}[\hat{\theta}] - 2\beta\text{cov}[\hat{\theta}, \check{\theta}] + \beta^2\text{var}[\check{\theta}]$$

desta forma a variância poderá ser reduzida se $\text{cov}[\hat{\theta}, \check{\theta}] > 0$.

Outro tipo de variável controle é uma tal que, sua média é desconhecida mas igual ao parâmetro, i.e., $E(\check{\theta}) = E(\hat{\theta}) = \theta$.

Para aplicação prática de variáveis controle existe o problema de encontrar tais variáveis altamente correlacionadas com os estimadores de interesse.

APÊNDICE B

Amostragem das Densidades Condicionais

B.1 Introdução

Um grupo de pesquisadores ligados a Tanizaki tem contribuído bastante na área de filtragem, suavizamento e estimação de parâmetros em modelos de espaço de estados não linear e não Gaussiano. Como visto na seção 2.5 o interesse está em estimar esperanças condicionais (ver equação (2.6.5)).

Em particular, existe interesse em obter boas estimativas para as densidades de filtragem e suavizamento e para a verossimilhança, de forma que possibilite obter boas estimativas dos parâmetros do modelo e da volatilidade.

Os algoritmos propostos proporciona a flexibilidade de trabalhar com as duas formas do modelo de volatilidade estocástica encontradas na literatura: forma não linear como na equação (3.3.4) e a forma não Gaussiana como na equação (3.3.6). Como visto anteriormente no modelo de volatilidade estocástica as perturbações possuem distribuição *log - qui - quadrado*.

Tanizaki e Mariano (1998) propuseram dois novos algoritmos baseados em técnicas de simulação de Monte Carlo com o objetivo de diminuir as dificuldades de simulação e problemas de tempo computacional. O primeiro método proposto, apresentado com o nome de aproximação de Monte Carlo tipo 1 (MA1), tem como base a amostragem direta de distribuições condicionais apropriadas. No segundo método, aproximação de Monte Carlo tipo 2 (MA2), usa-se amostragem por rejeição.

B.1.1 Aproximação Monte Carlo: Tipo 1 (MA1)

Predição MA1 (MA1P): O algoritmo de predição (MA1P) é obtido da seguinte forma. Suponha $\alpha_{i,t+L-1|t}$, $i = 1, \dots, n$. i.e., a i -ésima amostra da densidade $p(\alpha_{t+L-1}|Y_t)$, disponível. O algoritmo de predição L passos à frente é dado pela equação (2.6.1) aproximada por:

$$p(\alpha_{t+L}|Y_t) \approx \frac{1}{n} \sum_{i=1}^n p(\alpha_{t+L}|\alpha_{i,t+L-1|t}), \quad (\text{B.1.1})$$

ou seja, $p(\alpha_{t+L}|Y_t)$ é aproximada por uma mistura de distribuição.

Logo, o algoritmo para gerar números aleatórios de α_{t+L} dado Y_t é dado como segue:

- ① Escolha $\alpha_{i,t+L-1|t}$ aleatoriamente (i.e., escolha i com probabilidade $1/n$).
- ② Dado $\alpha_{i,t+L-1|t}$, gere amostras aleatórias de α_{t+L} (i.e., $\alpha_{j,t+L|t}$) da equação de transição $\alpha_{j,t+L|t} = f_{t+L}(\alpha_{i,t+L-1|t}, \eta_{j,t+L})$ após ter gerado um número aleatório para η_{t+L} , i.e., $\eta_{j,t+L}$ para $j = 1, \dots, n$.

Assim, dado $\alpha_{i,t|t}$, $\alpha_{i,t+L|t}$ é recursivamente obtido para $L = 1, 2, \dots$ e $\alpha_{i,t|t}$ é gerado pelo algoritmo de filtragem representado a seguir.

Filtragem MA1 (MA1F) : Suponha que seja possível retirar amostra aleatória de $\alpha_{i,t-1|t-1}$. No algoritmo de (MA1F) a equação de filtragem dada por (2.6.3) é aproximada por

$$\begin{aligned}
 p(\alpha_t|Y_t) &= \frac{p(y_t|\alpha_t)p(\alpha_t|Y_{t-1})}{\int p(y_t|\alpha_t)p(\alpha_t|Y_{t-1}) d\alpha_{t-1}} \\
 &\propto \int p(y_t|\alpha_t)p(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1}) d\alpha_{t-1} \\
 &\approx \frac{1}{n} \sum_{i=1}^n p(y_t|\alpha_t)p(\alpha_t|\alpha_{i,t-1|t-1}) \\
 &\propto \frac{1}{n} \sum_{i=1}^n w_1(\alpha_t; y_t)p(\alpha_t|\alpha_{i,t-1|t-1}) \tag{B.1.2}
 \end{aligned}$$

onde $w_1(\alpha_t; y_t) \propto p(y_t|\alpha_t)$, $0 \leq w_1(\alpha_t; y_t) \leq 1$.

Quando não se tem a forma funcional para $p(\alpha_t|Y_t)$, utiliza-se amostragem por rejeição para obter amostras aleatórias de $P(\alpha_t|Y_t)$, sabendo-se que esta é dominada pela densidade $p(\alpha_t|\alpha_{j,t-1|t-1})$. Portanto, para $t = 1, \dots, T$, escolha $\alpha_{j,t-1|t-1}$ com probabilidade $\frac{1}{n}$, amostre da distribuição $p(\alpha_t|\alpha_{j,t-1|t-1})$ e aceite com probabilidade $w_1(\alpha_t; y_t)$.

O algoritmo para obter estimativa de (MA1F) será dado como segue:

- ① Escolha $\alpha_{j,t-1|t-1}$ para $j = 1, \dots, n$, com probabilidade $\frac{1}{n}$.
- ② Dado j , gere $\alpha_{i,t} = f_t(\alpha_{j,t-1|t-1}, \eta_{i,t})$, onde $\eta_{i,t}$ é o i -ésimo número aleatório de η_t .
- ③ (i) Se $\alpha_{i,t}$, é aceito com probabilidade $w_1(\alpha_{i,t}; y_t)$, ele é tomado como uma retirada aleatória para a densidade de filtragem no tempo t , i.e., $\alpha_{i,t|t}$.
(ii) Se $\alpha_{i,t}$ não é aceito com probabilidade $w_1(\alpha_{i,t}; y_t)$, volte para ②.

④ Repita ① a ③ para $i = 1, \dots, n$.

Se a probabilidade de aceitação for baixa, deve-se voltar a ①, tomar outro α_j e repetir o procedimento, caso contrário os passos 2 e 3(i) se repetirão indefinidamente.

Suavizamento MA1 (MA1S) : Seja $\alpha_{i,t-1|t-1}$ retirada de $P(\alpha_{t-1}|Y_{t-1})$ e $\alpha_{j,t+1|T}$ de $P(\alpha_{t+1}|\mathbf{y})$. Suponha $\alpha_{i,t|t}$ para $t = 1, \dots, T$ e $i = 1, \dots, n$ obtida de (MA1F) .

A fim de obter a estimativa (MA1S) da equação (2.6.3) tem-se

$$\begin{aligned} p(\alpha_t|Y_t) &= \frac{1}{p(y_t|Y_{t-1})} \int p(y_t|\alpha_t)p(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1}) d\alpha_{t-1} \\ &\approx \frac{1}{p(y_t|Y_{t-1})} \frac{1}{n} \sum_{i=1}^n p(y_t|\alpha_t)p(\alpha_t|\alpha_{i,t-1|t-1}) \end{aligned} \quad (\text{B.1.3})$$

$$\int \frac{p(\alpha_{t+1}|Y_t)p(\alpha_{t+1}|\alpha_t)}{p(\alpha_{t+1}|Y_t)} d\alpha_{t+1} \approx \frac{1}{n} \sum_{j=1}^n \frac{p(\alpha_{j,t+1|T}|\alpha_t)}{p(\alpha_{j,t+1|T}|Y_t)} \quad (\text{B.1.4})$$

o denominador em (2.6.3) pode ser escrito como

$$p(\alpha_{t+1}|Y_t) = \int p(\alpha_{t+1}|\alpha_t)p(\alpha_t|Y_t) d\alpha_t \approx \frac{1}{n} \sum_{m=1}^n p(\alpha_{t+1}|\alpha_{m,t|t}) \quad (\text{B.1.5})$$

Portanto a densidade de suavizamento (2.6.3) é aproximada por

$$\begin{aligned}
p(\alpha_t | \mathbf{y}) &\approx \frac{1}{p(y_t | Y_{t-1})} \left\{ \frac{1}{n} \sum_{i=1}^n p(y_t | \alpha_t) p(\alpha_{t+1} | \alpha_{m,t|t}) \right\} \frac{1}{n} \sum_{j=1}^n \frac{p(\alpha_{j,t+1|T} | \alpha_t)}{p(\alpha_{j,t+1|T} | Y_t)} \\
&\propto \left\{ \frac{1}{n} \sum_{i=1}^n p(y_t | \alpha_t) p(\alpha_t | \alpha_{i,t-1|t-1}) \right\} \left\{ \frac{1}{n} \sum_{j=1}^n \frac{p(\alpha_{j,t+1|T} | \alpha_t)}{p(\alpha_{j,t+1|T} | Y_t)} \right\} \\
&\propto \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{q_{j,t}} p(y_t | \alpha_t) p(\alpha_{j,t+1|T} | \alpha_t) p(\alpha_t | \alpha_{i,t-1|t-1}) \\
&\propto \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{q_{j,t}} w_1(\alpha_t; y_t) w_2(\alpha_t; \alpha_{j,t+1|T}) p(\alpha_t | \alpha_{i,t-1|t-1}) \tag{B.1.6}
\end{aligned}$$

Para $t = T - 1, T - 2, \dots, 1$, e $q_{j,t}$ satisfaz as seguintes condições

1.

$$\begin{aligned}
q_{j,t} &\propto \frac{1}{n} \sum_{m=1}^n p(\alpha_{j,t+1|T} | \alpha_{m,t|t}) \approx \int p(\alpha_{j,t+1|T} | \alpha_t) p(\alpha_t | Y_t) d\alpha_t \\
&= p(\alpha_{j,t+1|T} | Y_t)
\end{aligned}$$

2.

$$\sum_{j=1}^n \frac{1}{q_{j,t}} = 1$$

e $w_1(\alpha_t; y_t)$ e $w_2(\alpha_t; \alpha_{j,t+1|T})$ satisfazem

$$w_1(\alpha_t; y_t) \propto p(y_t | \alpha_t),$$

$$w_2(\alpha_t; \alpha_{j,t+1|T}) \propto p(y_t | \alpha_t),$$

$$0 \leq w_1(\alpha_t; y_t) w_2(\alpha_t; \alpha_{j,t+1|T}) \leq 1.$$

Como no algoritmo de filtragem, a forma funcional da densidade a posteriori (suavizada) muitas vezes não é obtida analiticamente e pode-se utilizar amostragem

por rejeição para gerar amostras de $p(\alpha_t|Y_T)$ para $t = T - 1, T - 2, \dots, 1$. Para isto escolhe-se $\alpha_{i,t-1|t-1}$ com probabilidade $1/n$ e $\alpha_{j,t+1|T}$ com probabilidade $1/q_{j,t}$ (i.e., escolha i e j aleatoriamente), amostre de $p(\alpha_{i,t-1|t-1})$ e aceite com probabilidade $w_1(\alpha_t; y_t)w_2(\alpha_t; \alpha_{j,t+1|T})$, onde $\alpha_{i,T|T}$ é retirada aleatoriamente da densidade de filtragem.

Com objetivo de diminuir o tempo computacional, Tanizaki e Mariano aproximaram $q_{j,t} = n, \forall j$. Portanto, a densidade à posteriori é dada por

$$p(\alpha_t|Y_T) \propto \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n w_1(\alpha_t; y_t)w_2(\alpha_t; \alpha_{j,t+1|T})p(\alpha_t|\alpha_{i,t-1|t-1}). \quad (\text{B.1.7})$$

Tal procedimento fornece as amostras aleatórias de α_t dado $\alpha_{i,t-1|t-1}$ e $\alpha_{j,t+1|T}$, onde i e j são ambos escolhidos com probabilidade $1/n$.

Resumindo-se, as funções densidades requeridas serão obtidas numericamente através das seguintes estimativas:

☞ Densidade de predição,

$$p(\alpha_{t+L}|Y_t) \approx \frac{1}{n} \sum_{i=1}^n p(\alpha_{t+L}|\alpha_{i,t+L-1|t}), \quad L = 1, 2, \dots,$$

☞ Densidade de filtragem,

$$p(\alpha_t|Y_t) \propto \frac{1}{n} \sum_{i=1}^n w_1(\alpha_t; y_t)p(\alpha_t|\alpha_{i,t-1|t-1}), \quad i = 1, 2, \dots, T.$$

☞ Densidade de suavização,

$$p(\alpha_t|\mathbf{y}) \propto \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n w_1(\alpha_t; y_t)w_2(\alpha_t; \alpha_{j,t+1|T})p(\alpha_t|\alpha_{i,t-1|t-1}), \quad t = T, \dots, 1.$$

Após conseguir amostrar das densidades de $\alpha_{r|s}$ para $(r, s) = (t+L, t), (t, t), (t, T)$, predição, filtragem e suavizamento respectivamente; pode-se calcular o estimador **MA1** de $g_{r|s}$ dado por

$$g_{r|s}^* = \frac{1}{n} \sum_{i=1}^n g(\alpha_{i,r|s})$$

A função de verossimilhança estimada será obtida da seguinte forma:

$$p(\mathbf{y}) = \prod_{t=1}^T \left(\int p(y_t|\alpha_t) p(\alpha_t|Y_{t-1}) d\alpha_t \right) \approx \prod_{t=1}^T \left(\frac{1}{n} \sum_{i=1}^n p(y_t|\alpha_{i,t|t-1}) \right). \quad (\text{B.1.8})$$

B.1.2 Aproximação Monte Carlo: Tipo 2 (MA2)

O segundo algoritmo proposto por Tanizaki e Mariano (1998), segundo os experimentos de Monte Carlo feitos pelos pesquisadores, mostrou-se menos eficiente que o algoritmo (**MA1**) na maioria dos casos, principalmente quando o número de observações é grande e quando a equação de estado (2.5.4) evolui segundo um processo não estacionário. Dessa forma, será elaborada apenas uma breve apresentação do algoritmo proposto pelos autores. Pode-se escrever a função de densidade conjunta de (\mathbf{y}, α) como

$$p(\mathbf{y}, \alpha) = p(\mathbf{y}|\alpha)p(\alpha) \quad (\text{B.1.9})$$

onde

$$p(\mathbf{y}|\alpha) = \prod_{t=1}^s p(y_t|\alpha_t) \quad (\text{B.1.10})$$

$$p(\alpha) = \begin{cases} p(\alpha_0) \prod_{t=1}^T p(\alpha_s|\alpha_{t-1}), & \text{se } \alpha_0 \text{ é estocástico,} \\ \prod_{t=1}^T p(\alpha_t|\alpha_{t-1}), & \text{caso contrário,} \end{cases} \quad (\text{B.1.11})$$

A distribuição condicional de α_r dada a informação Y_s , é dada por:

$$p(\alpha_r|y_s) = \frac{\int p(y_s|\alpha_s)p(\alpha) d\alpha_{T,\setminus r}}{\int p(y_s|\alpha_s)p(\alpha) d\alpha} \quad (\text{B.1.12})$$

onde $\alpha_{T,\setminus r} = \{\alpha_T, \dots, \alpha_{r+1}, \alpha_{r-1}, \dots, \alpha_0\}$, um subconjunto de α excluindo α_r . Pode-se calcular a esperança de $(g(\alpha_r)|y_s)$ por:

$$g_{r|s} \equiv E(g(\alpha_r)|Y_s) = \frac{\int g(\alpha_r)p(y_s|\alpha_s)p(\alpha) d\alpha}{\int p(y_s|\alpha_s)p(\alpha) d\alpha} \quad (\text{B.1.13})$$

Definindo $A_{i,T} = \{\alpha_{iT}, \dots, \alpha_{i0}\}$, $i = 1, \dots, n$ retiradas aleatórias de $p(\alpha)$. Então temos que $A_{i,s}$ é um conjunto de amostras aleatórias da equação de transição da forma $\alpha_{i,t} = f_t(\alpha_{i,t-1}, \eta_{it})$, $t = 1, \dots, s$.

Portanto, dada n amostras de α o estimador (MA2) de $g_{r|s}$ será dado por

$$\bar{g}_{r|s} = \frac{\frac{1}{n} \sum_{i=1}^n g(\alpha_{i,r})p(y_s|A_{i,s})}{\frac{1}{n} \sum_{i=1}^n p(y_s|A_{i,s})}$$

Com objetivo de simplificar as notações pode-se escrever:

$$\bar{g}_{r|s} = \sum_{i=1}^n g(\alpha_{i,r})\omega_{i,s} \quad (\text{B.1.14})$$

onde os pesos $\omega_{i,s}$ são da seguinte forma

$$\omega_{i,s} = \frac{\prod_{t=1}^s p(y_t|\alpha_{i,t})}{\sum_{j=1}^n \prod_{t=1}^s p(y_t|\alpha_{j,t})} = \frac{p(y_s|\alpha_{i,s})\omega_{i,s-1}}{\sum_{j=1}^n p(y_s|\alpha_{j,s})\omega_{j,s-1}} \quad (\text{B.1.15})$$

com condicional inicial $\omega_{i,0} = \frac{1}{n}$ para $i = 1, \dots, n$ e temos que $\sum_{i=1}^n \omega_{i,s} = 1$.

As características principais do (MA2) é a necessidade da forma funcional das densidades $p(y_t|\alpha_t)$ e os α_t são gerados da equação de transição para suavização, filtragem e predição.

O Algoritmo é dado da seguinte forma:

- ① Gere $\alpha_{i,0}$ para $i = 1, \dots, n$ de $p(\alpha_0)$,
- ② Dado $\alpha_{i,0}$, faça $\alpha_{i,t} = f_t(\alpha_{i,t-1}, \eta_{it})$ com $i = 1, \dots, n$ e $t = 1, \dots, T$ onde η_{it} é retirada de sua distribuição base.
- ③ Dado peso inicial, $\omega_{i,0} = \frac{1}{n}$, use a equação (B.1.15) para obter funções pesos para $i = 1, \dots, n$ e $t = 1, \dots, T$,
- ④ $\bar{g}_{r|s}$ obtida de (B.1.14)

Assim como os métodos propostos por **DK** e **SP**, os parâmetros do modelo são estimados por procedimentos numéricos, onde a verossimilhança maximizada é aproximada por

$$p(\mathbf{y}) = \int p(\mathbf{y}|\alpha)p(\alpha) d\alpha \approx \sum_{i=1}^n p(\mathbf{y}|A_{i,T}) \quad (\text{B.1.16})$$

B.1.3 Suavizamento Intervalo-Fixo

Geweke e Tanizaki (1999) propuseram um algoritmo de suavizamento, baseado em MCMC, para modelos de espaço de estados não linear e não Gaussiano. Eles utilizaram Metropolis-Hastings e amostrador de Gibbs (ver apêndice A) para gerar amostras aleatórias da densidade de interesse.

As densidades $p(y_t|\alpha_t)$ e $p(\alpha_t|\alpha_{t-1})$ são obtidas das equações (2.5.3) e (2.5.4). Para obter a média suavizada, i.e., $E(\alpha_t|\mathbf{y}) = \alpha_{t|T}$ é necessário obter a densidade condicional $p(\alpha|\mathbf{y})$, onde $\alpha = (\alpha_0, \dots, \alpha_T)$ e $\mathbf{y} = (y_1, \dots, y_T)$. Pode-se escrever essa densidade como equação (B.1.9) e assim encontrar a densidade posteriori $p(\alpha|\mathbf{y})$ pelo teorema de Bayes, da seguinte forma

$$\begin{aligned} p(\alpha|\mathbf{y}) &= \frac{p(\alpha, \mathbf{y})}{\int p(\alpha, \mathbf{y}) d\alpha} \\ &= \frac{p(\alpha)p(\mathbf{y}|\alpha)}{\int p(\alpha)p(\mathbf{y}|\alpha) d\alpha} \end{aligned} \quad (\text{B.1.17})$$

$$\begin{aligned} p(\alpha_t|\mathbf{y}) &= \int p(\alpha|\mathbf{y}) d\alpha_{\setminus t} \\ &= \frac{\int p(\alpha|\mathbf{y}) d\alpha_{\setminus t}}{\int p(\mathbf{y}, \alpha) d\alpha} \\ &= \frac{\int p(\alpha)p(\mathbf{y}|\alpha) d\alpha_{\setminus t}}{\int p(\alpha)p(\mathbf{y}|\alpha) d\alpha} \end{aligned} \quad (\text{B.1.18})$$

Como é necessário amostrar da densidade T dimensional $p(\alpha|\mathbf{y})$, pode-se utilizar o amostrador de Gibbs. Mas para utilizar-se tal técnica são necessárias as densidades condicionais. Então tem-se

$$\begin{aligned}
p(\alpha_s | \alpha_{\setminus s}, \mathbf{y}) &= \frac{p(\alpha | \mathbf{y})}{p(\alpha_{\setminus s} | \mathbf{y})} \\
&= \frac{p(\alpha)p(\mathbf{y} | \alpha)}{\int p(\alpha)p(\mathbf{y} | \alpha) d\alpha_s} \\
&= \begin{cases} \frac{p(\alpha_s | \alpha_{s-1})p(\alpha_{s+1} | \alpha_s)p(y_s | \alpha_s)}{\int p(\alpha_s | \alpha_{s-1})p(\alpha_{s+1} | \alpha_s)p(y_s | \alpha_s) d\alpha_s}, & \text{se } s = 1, 2, \dots, T-1, \\ \frac{p(\alpha_s | \alpha_{s-1})p(y_s | \alpha_s)}{\int p(\alpha_s | \alpha_{s-1})p(y_s | \alpha_s) d\alpha_s}, & \text{se } s = T, \end{cases} \\
&\propto \begin{cases} p(\alpha_s | \alpha_{s-1})p(\alpha_{s+1} | \alpha_s)p(y_s | \alpha_s), & \text{se } s = 1, 2, \dots, T-1, \\ p(\alpha_s | \alpha_{s-1})p(y_s | \alpha_s) & \text{se } s = T, \end{cases} \quad (\text{B.1.19})
\end{aligned}$$

Assim, pode-se utilizar a densidade em (B.1.19) para a implementação do algoritmo. Segue agora um novo problema, já que muitas vezes não se tem a forma analítica para a densidade marginal. Nesse passo Geweke e Tanizaki (1999) propuseram o algoritmo de Metropolis-Hastings a fim de obter amostras da densidade condicional necessária.

Como pode-se verificar no apêndice A, para implementar o algoritmo de M-H é necessário uma densidade da qual pode-se amostrar, a densidade proposta. Através dos estudos de simulação Monte Carlo eles concluíram que a densidade proposta mais eficiente no caso de modelos de espaço de estado não linear e não Gaussiano é a obtida através da equação de transição (2.5.4). Considerando $A_{i,s} = (\alpha_{i,0}, \dots, \alpha_{i,s})$ e $A_{i,s}^* = (\alpha_{i,s}, \dots, \alpha_{i,T})$, a probabilidade de aceitação é dada por

$$\omega(\alpha, z) = \begin{cases} \min\left(\frac{p(z|A_{i,s-1}, A_{i-1,s+1}^*, \mathbf{y})p_*(\alpha|z)}{p(\alpha|A_{i,s-1}, A_{i-1,s+1}^*, \mathbf{y})p_*(z|\alpha)}, 1\right), & \text{se } p(\alpha|A_{i,s-1}, A_{i-1,s+1}^*, \mathbf{y})p_*(z|\alpha) > 0, \\ 1, & \text{caso contrário} \end{cases}$$

Assim o algoritmo para obtenção de amostras α de $p(\alpha|\mathbf{y})$ é dado como segue

- ① Tome valores apropriados para α_1 e $\alpha_{0,s}$, $s = 1, 2, \dots, T$. Geralmente, as estimativas suavizadas baseadas no filtro de Kalman estendido (KFE) são tomados para α_s , $s = 1, \dots, T$. O estado $\alpha_{1,0}$ dependerá da suposição inicial da densidade $P(\alpha_0)$.
- ② Dado $A_{i,s-1}$ e $A_{i-1,s+1}^*$, utiliza-se o algoritmo de Metropolis-Hastings para gerar uma amostra de α_s , da densidade condicional $p(\alpha_s|A_{i,s-1}A_{i-1,s+1}^*, \mathbf{y})$, denotada por $\alpha_{i,s}$, com o seguinte procedimento
 - (i) Gere z de $p_*(\cdot|\alpha_{i-1,s})$ e u de uma densidade uniforme $(0, 1)$.
 - (ii) Calcule $\omega(\alpha_{i-1,s}, z)$ e
 - (1) Se $u \leq \omega(\alpha_{i-1,s}, z)$, faça $\alpha_{i,s} = z$.
 - (2) caso contrário, faça $\alpha_{i,s} = \alpha_{i-1,s}$.
- ③ Repita ② para $s = 1, \dots, T$.
- ④ Repita ② e ③ para $i = 1, \dots, N$.

B.1.4 Amostragem por Importância para Filtro Não Linear

Tanizaki (1999a) propôs um algoritmo de filtragem para modelos na forma de espaço de estado **FEE** não linear e não Gaussiano onde utilizou amostragem por importância e o conceito de antiteticidade. O método é um aperfeiçoamento de um algoritmo que usa amostragem por importância utilizado anteriormente por Tanizaki e Mariano (1994), Mariano e Tanizaki (1995) e Tanizaki (1996).

O algoritmo de densidade recursiva de filtragem é dado pelas equações (2.6.2) e (2.6.3) onde a condição inicial é dada por:

$$p(\alpha_1|Y_0) = \begin{cases} \int p(\alpha_1|\alpha_0)p(\alpha_0)d\alpha_0, & \text{se } \alpha_0 \text{ é estocástico,} \\ p(\alpha_1|\alpha_0), & \text{se } \alpha_0 \text{ não é estocástico} \end{cases} \quad (\text{B.1.20})$$

as densidades $p(\alpha_t|\alpha_{t-1})$ e $p(y_t|\alpha_t)$ são obtidas das equações de observação e de transição respectivamente.

Mariano e Tanizaki (1995), Tanizaki (1996) e Tanizaki e Mariano (1994) propuseram algoritmos de filtragem usando amostragem por importância onde as funções densidades são convertidas por pesos, por exemplo, para calcular $a_{t|s} = E(\alpha_t|Y_s)$ tem-se

$$\begin{aligned} a_{t|s} &= \int \alpha_t p(\alpha_t|Y_s) d\alpha_t \\ &= \int \alpha_t \frac{p(\alpha_t|Y_s)}{p_I(\alpha_t)} p_I(\alpha_t) d\alpha_t \\ &= \int \alpha_t \omega(\alpha_t|Y_s) p_I(\alpha_t) d\alpha_t \end{aligned} \quad (\text{B.1.21})$$

onde $p_I(\cdot)$ é uma densidade importante. Dessa forma as funções pesos são representadas por:

$$\omega(\alpha_t|Y_s) = \frac{p(\alpha_t|Y_s)}{p_I(\alpha_t)},$$

para $s = t - 1, t$.

Usando as funções pesos as densidades de filtragem podem ser estimadas da seguinte forma:

$$\begin{aligned} \omega(\alpha_t|Y_{t-1}) &= \int \frac{p(\alpha_t|\alpha_{t-1})}{p_I(\alpha_t)} \omega(\alpha_{t-1}|Y_{t-1}) p_I(\alpha_{t-1}) d\alpha_{t-1} \\ &\approx \frac{1}{n} \sum_{j=1}^n \frac{p(\alpha_t|\alpha_{j,t-1})}{p_I(\alpha_t)} \omega(\alpha_{j,t-1}|Y_{t-1}) \end{aligned} \quad (\text{B.1.22})$$

$$\begin{aligned} \omega(\alpha_t|Y_t) &= \frac{p(y_t|\alpha_t) \omega(\alpha_t|Y_{t-1})}{\int p(y_t|\alpha_t) \omega(\alpha_t|Y_{t-1}) p_I(\alpha_t) d\alpha_t} \\ &\approx \frac{p(y_t|\alpha_t) \omega(\alpha_t|Y_{t-1})}{\frac{1}{n} \sum_{j=1}^n p(y_t|\alpha_{j,t}) \omega(\alpha_{j,t}|Y_{t-1})} \end{aligned} \quad (\text{B.1.23})$$

Vale ressaltar que os valores de $\alpha_{i,t}$ para as equações (B.1.22) e (B.1.23), são obtidos através da mesma densidade importante $p_I(\alpha_t)$. Deve-se também encontrar as funções pesos através de um algoritmo recursivo, dado que não se consegue obter a forma analítica de $p(\alpha_t|Y_t)$ e $P(\alpha_t|Y_{t-1})$. Outro fato importante é a utilização da mesma retirada aleatória para estimar os pesos.

$$\omega(\alpha_{i,t}|Y_{t-1}) \approx \frac{1}{n} \sum_{j=1}^n \frac{p(\alpha_{i,t}|\alpha_{j,t-1})}{p_I(\alpha_{i,t})} \omega(\alpha_{j,t-1}|Y_{t-1}), \quad (\text{B.1.24})$$

$$\omega(\alpha_{i,t}|Y_t) \approx \frac{p(y_t|\alpha_{i,t}) \omega(\alpha_{i,t}|Y_{t-1})}{\frac{1}{n} \sum_{j=1}^n p(y_t|\alpha_{j,t}) \omega(\alpha_{j,t}|Y_{t-1})}, \quad (\text{B.1.25})$$

$$\omega(\alpha_{i,1}|Y_0) \approx \frac{1}{n} \sum_{j=1}^n \frac{p(\alpha_{i,1}|\alpha_{j,0})}{p_I(\alpha_{i,1})} \omega(\alpha_{j,0}|Y_0), \quad \text{se } \alpha_0 \text{ é estocástico} \quad (\text{B.1.26})$$

$$\omega(\alpha_{i,1}|Y_0) = \frac{p(\alpha_{i,1}|\alpha_0)}{p_I(\alpha_{i,1})}, \quad \text{se } \alpha_0 \text{ é não estocástico} \quad (\text{B.1.27})$$

A média, variância condicional e densidade de filtragem são estimadas como segue

$$a_{t|s} \approx \frac{1}{n} \sum_{i=1}^n \alpha_{i,t} \omega(\alpha_{i,t}|Y_s), \quad (\text{B.1.28})$$

$$\Sigma_{t|s} \approx \frac{1}{n} \sum_{i=1}^n (\alpha_{i,t} - a_{t|s})(\alpha_{i,t} - a_{t|s})' \omega(\alpha_{i,t}|Y_s), \quad (\text{B.1.29})$$

$$p(\alpha_{i,t}|Y_s) = \omega(\alpha_{i,t}|Y_s) p_I(\alpha_{i,t}), \quad \text{para } s = t - 1, t. \quad (\text{B.1.30})$$

Método de Redução de Variância em Integração por Monte Carlo

Da mesma forma que foi descrito na seção 4.2 tem-se o problema da escolha da densidade importante, $p_I(\cdot)$, que produza estimativas precisas para as quantidades de interesse. Portanto, é desejado que essa seja o mais próximo possível de $p(\alpha_t|Y_{t-1})$ e $p(\alpha_t|Y_t)$.

Mariano e Tanizaki (1995), Tanizaki (1996) e Tanizaki e Mariano (1994) sugeriram como densidade importante uma distribuição bimodal dada por:

$$p_I(\alpha_t) = \frac{1}{2} N(a_{t|t-1}^*, c\Sigma_{t|t-1}^*) + \frac{1}{2} N(a_{t|t}^*, c\Sigma_{t|t}^*) \quad (\text{B.1.31})$$

onde $a_{t|s}^*$ e $\Sigma_{t|s}^*$ para $s = t - 1, t$ são, respectivamente, a média e variância obtidas do filtro de Kalman estendido e c é uma constante maior ou igual a 1.

A densidade importante deve ser tomada de modo que seu domínio e imagem contêm os mesmos das densidades de predição e de atualização. Portanto, como essas

densidades possuem diferentes domínios e modas, Tanizaki tomou uma densidade bimodal com objetivo de solucionar esse problema. Baseado na distribuição bimodal acima, as retiradas aleatórias de α_t de $P_I(\alpha_t)$ são geradas por:

$$\alpha_{i,t} = a + \Sigma^{1/2} z_{i,t}, \quad (\text{B.1.32})$$

onde $z_{i,t} \sim N(0, 1)$ para $i = 1, \dots, T$. (a, Σ) é dada pela seguinte variável discreta:

$$(a, \Sigma) = \begin{cases} (a_{t|t-1}^*, c\Sigma_{t|t-1}^*), & \text{com probabilidade } 1/2 \\ (a_{t|t}^*, c\Sigma_{t|t}^*), & \text{com probabilidade } 1/2 \end{cases} \quad (\text{B.1.33})$$

Na prática para escolher uma das distribuições acima, gera-se um número aleatório uniforme $u \in (0, 1)$. Assim, escolha $(a_{t|t-1}^*, c\Sigma_{t|t-1}^*)$ quando $u \leq 0.5$ e $(a_{t|t}^*, c\Sigma_{t|t}^*)$ quando $u > 0.5$.

Para melhorar as estimativas Tanizaki propôs o uso do conceito de antiteticidade, onde considerou dois métodos para amostrar α . O primeiro método considera $\alpha_{i,t}$ da seguinte forma

$$\alpha_{i,t} = a + \Sigma^{1/2} z_{i,t}, \quad \alpha_{i+m,t} = a + \Sigma^{1/2} (-z_{i,t}), \quad (\text{B.1.34})$$

onde $z_{i,t} \sim N(0, 1)$ para $i = 1, \dots, m$, $n = 2m$ e $t = 1, \dots, T$. Note que (a, Σ) é dada por (B.1.33).

De maneira análoga o segundo procedimento é dado por:

$$\alpha_{i,t} = a + \Sigma^{1/2} z_{i,t}, \quad \alpha_{i+m,t} = \tilde{a} + \tilde{\Sigma}^{1/2} (-z_{i,t}), \quad (\text{B.1.35})$$

onde $(\tilde{a}, \tilde{\Sigma})$ é obtido da seguinte forma:

$$(\tilde{a}, \tilde{\Sigma}) = \begin{cases} (a_{t|t-1}^*, c\Sigma_{t|t-1}^*), & \text{se } (a, \Sigma) = (a_{t|t}^*, c\Sigma_{t|t}^*) \\ (a_{t|t}^*, c\Sigma_{t|t}^*), & \text{se } (a, \Sigma) = (a_{t|t-1}^*, c\Sigma_{t|t-1}^*) \end{cases}$$

e (a, Σ) é dado por (B.1.33).

A escolha por um dos métodos dependerá do problema tratado. Por exemplo, no caso onde $P(\alpha_t|Y_{t-1})$ e $P(\alpha_t|Y_t)$ são distantes, i.e., $N(a_{t|t-1}^*, c\Sigma_{t|t-1}^*)$ está distante de $N(a_{t|t}^*, c\Sigma_{t|t}^*)$ escolhe-se (B.1.35), pois $\alpha_{i,t}$ e $\alpha_{i+m,t}$ são mais negativamente correlacionadas. Por outro lado, quando $P(\alpha_t|Y_{t-1})$ e $P(\alpha_t|Y_t)$ estão próximas, (B.1.34) pode ser preferível no lugar de (B.1.35).

APÊNDICE C

Resultados em Inferência Bayesiana

C.1 Distribuições Conjugadas

Definição C.1.1. *Seja \mathcal{F} a classe de f.d.p $f(x|\theta)$ (indexada por θ). A classe Π de distribuições a priori é uma família conjugada para \mathcal{F} se a distribuição a posteriori está na classe Π para todo $f \in \mathcal{F}$, toda priori em Π e para todo valor amostral \mathbf{x} .*

Exemplo C.1.1. *Seja $\mathbf{X} \sim N(\mu, \theta)$ onde μ é parâmetro conhecido, e seja a priori $\pi(\theta) \sim \mathcal{IG}(r, \lambda)$. Assim teremos*

$$\begin{aligned}\pi(\theta) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \frac{1}{\theta^{\alpha+1}} e^{-\frac{1}{\beta\theta}} \quad \theta > 0 \\ \pi(\theta|\mathbf{x}) &= \left(\frac{1}{2\pi\theta}\right)^{n/2} \exp\left\{-\frac{\sum(x_i - \mu)^2}{2\theta} - \frac{1}{\beta\theta}\right\} \frac{1}{\theta^{\alpha+1}} \\ &\propto \frac{1}{\theta^{\alpha+(n/2)+1}} \exp\left\{-\frac{1}{\theta}\left(\frac{\sum(x_i - \mu)^2}{2} + \frac{1}{\beta}\right)\right\}\end{aligned}$$

que é $\mathcal{IG}(\alpha + n/2; (\sum(x_i - \mu)^2/2 + 1/\beta)^{-1})$.

C.1.1 Famílias Conjugadas para Famílias Exponenciais

Teorema C.1.1. *Considere um amostra cuja distribuição pertence à família exponencial, i.e., sua f.d.p. conjunta é dada por:*

$$f(\mathbf{x}|\theta) = h^*(\mathbf{x}) \exp\left\{ \sum_{j=1}^k w_j(\theta) t_j^*(\mathbf{x}) + nd(\theta) \right\} \quad (\text{C.1.1})$$

onde θ tem dimensão k . Então uma família conjugada para esta família é dada pela família exponencial com hiper parâmetro β de dimensão $k + 1$ dada por:

$$\pi(\theta|\beta_1, \dots, \beta_{k+1}) = b(\beta_1, \dots, \beta_{k+1}) \exp\left\{ \sum_{j=1}^k \beta_j w_j(\theta) + \beta_{k+1} d(\theta) \right\} \quad (\text{C.1.2})$$

REFERÊNCIAS

- [1] Abramovitz, M. e Stegun, N. C. (1970). *Handbook of Mathematical Functions*. Nova Iorque : Dover Publications Inc.
- [2] Anderson, B. D. O. e Moore, J. B. (1979). *Optimal Filtering*. Englewood Cliffs: Prentice Hall.
- [3] Aoki, M. (1987). *State Space Modelling of Time Series*. Berlim : Springer-Verlag.
- [4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.
- [5] Bollerslev, T., Chou, R. Y., e Kroner, K. F. (1992). ARCH modelling in finance: A review of theory and empirical evidence. *Journal of Econometrics*, 52:5–59.
- [6] Box, G. E. P. e Tiao, G. C. (1968). A bayesian approach to some outlier problems. *Biometrika*, 55:119–129.
- [7] Box, G. E. P. e Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Massachussets:Addison-Wesley.

- [8] Brockwell, P. J. e Davis, R. A. (1990). *Time Series: Theory and Methods*. Nova Iorque : Springer-Verlag.
- [9] Brown, T. D. (1990). *C for Fortran Programmers*. Prentice Hall.
- [10] Carlin, B. P., Polson, N. G., e Stoffer, D. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *Journal of the American Statistical Association*, 87:493–500.
- [11] Carpenter, J. R., Clifford, P., e Fearnhead, P. (1998). An improved particle filter for non-linear problems. Artigo em Trabalho, Universidade de Oxford.
- [12] Carter, C. K. e Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81:541–553.
- [13] Carter, C. K. e Kohn, R. (1996). Markov Chain Monte Carlo in conditionally gaussian state space models. *Biometrika*, 83:589–601.
- [14] Casella, G. e George, E. I. (1992). Explaining the Gibbs sampling. *American Statistician*, 46:167–174.
- [15] Chib, S. e Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *American Statistician*, 49:327–335.
- [16] Chib, S. e Nardari, F. (1998). Markov Chain Monte Carlo methods for generalized stochastic volatility models. Documento de Trabalho.
- [17] Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative process. *Econometrica*, 41:135–156.

- [18] Danielson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics*, 64:375–400.
- [19] de Jong, P. (1988). The likelihood for a state space models. *Biometrika*, 75:165–169.
- [20] de Jong, P. (1989). Smoothing and interpolation with the state space models. *Journal of the American Statistical Association*, 84:1085–1088.
- [21] de Jong, P. (1991). The diffuse Kalman filter. *The Annals of Statistics*, 19(2):1073–1083.
- [22] de Jong, P. e Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, 82:339–350.
- [23] Doornik, J. A. (1999). *OX an Object-Oriented Matrix Programming Language*. London: Timberlake Consults Press.
- [24] Durbin, J. (2000). The state space approach to time series analysis and its potential for official statistics (with discussion). *Australian and New Zealand J. Statist.*, 42(1):1–23.
- [25] Durbin, J. e Koopman, S. J. (1992). Filtering, smoothing and estimation for time series models when the observations come from exponential family distributions. Documento de Trabalho: London School of Economics.
- [26] Durbin, J. e Koopman, S. J. (1997a). Monte Carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika*, 84:669–684.

- [27] Durbin, J. e Koopman, S. J. (1997b). Time series analysis of non-gaussian observations based on state space models. Documento de Trabalho: London School of Economics.
- [28] Durbin, J. e Koopman, S. J. (2000). Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society (Série B)*, 62:3–56.
- [29] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of uk inflation. *Econometrica*, 50:987–1008.
- [30] Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *Journal of the American Statistical Association*, 87:501–509.
- [31] Fama, E. F. (1963). Mandelbrot and the stable paretian distribution. *Journal of Business*, 36:420–429.
- [32] Fama, E. F. (1965). The behavior of stock market prices. *Journal of Business*, 38:34–105.
- [33] Fridman, M. e Harris, L. (1998). A maximum likelihood approach for non-gaussian stochastic volatility models. *Journal of Business and Economic Statistics*, 3(16):284–291.
- [34] Fukui, P. (2000). Detecção de valores aberrantes em modelos de volatilidade estocástica. Tese de mestrado, Departamento de Estatística, UNICAMP.
- [35] Fuller, W. A. (1996). *Introduction to Time Series*. Nova Iorque : John Wiley, segunda edição edition.

- [36] Gamerman, D. (1996). *Simulação Estocástica via Cadeia de Markov*. XII SINAPPE. ABE.
- [37] Gamerman, D. (1998). Markov Chain Monte Carlo for dynamic generalised linear models. *Biometrika*, 85:215–227.
- [38] Gamerman, D. e Migon, H. S. (1993). Dynamic hierarchical models. *Journal of the Royal Statistical Society (Série B)*, 55:629–642.
- [39] Gelb, A. (1974). *Applied Optimal Estimation*. MIT Press.
- [40] Gelfand, A. E. e Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.
- [41] Geman, S. e Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.
- [42] Geweke, J. e Tanizaki, H. (1999). On Markov Chain Monte Carlo methods for nonlinear and non-gaussian state-space models. *Communications in Statistics, Simulation and Computation*, 28(4):867–894.
- [43] Gilks, W., Best, N., e Tan, K. (1995). Adaptative rejection Metropolis sampling within Gibbs sampling. *Appl. Statist.*, 44(4):455–472.
- [44] Gilks, W. e Wild, P. (1992). Adaptative rejection sampling for Gibbs sampling. *Appl. Statist.*, 41(2):337–348.
- [45] Gilks, W. K., Richardson, S., e Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

- [46] Harrison, P. J. e Stevens, C. F. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society (Série B)*, 38:205–247.
- [47] Harvey, A. C. (1981). *Time Series Models*. N. J: Humanities Press, Oxford: Philip Allan.
- [48] Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge Univ. Press., Cambridge.
- [49] Harvey, A. C. e Fernandes, C. (1989). Time series models for count data or qualitative observations. *Journal of Business and Economic Statistics*, 7:407–422.
- [50] Harvey, A. C., Ruiz, E., e Shephard, N. (1994). Multivariate stochastic variance models. *Reviews of Economic Studies*, 61:247–264.
- [51] Hastings, W. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57:97–109.
- [52] Hull, J. e White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300.
- [53] Jacquier, E., Polson, N., e Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models” (with discussion). *Journal of Business and Economic Statistics*, 12:371–417.
- [54] Jacquier, E., Polson, N., e Rossi, P. E. (1995). Priors and models for multivariate stochastic volatility. unpublished manuscript, Universidade de Chicago, Graduate School of Business.

- [55] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASMA, Series D*, 82:35–45.
- [56] Kim, S., Shephard, N., e Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *Review of Economic Studies*, 85:361–393.
- [57] Kitagawa, G. (1987). Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82:1032–1063.
- [58] Kitagawa, G. (1996). Monte Carlo filter and smoother for non-gaussian nonlinear state-space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25.
- [59] Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika*, 80:117–126.
- [60] Koopman, S. J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association*, 92:1630–1638.
- [61] Koopman, S. J., Shephard, N., e Doornik, J. A. (1999). Statistical algorithms for models in state space using ssfpack 2.2. *Econometrics Journal*.
- [62] Liu, J., Wong, W. H., e Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparison of estimator and augmentation schemes. *Biometrika*, 81:27–40.
- [63] Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 36:394–419.

- [64] Mariano, R. S. e Tanizaki, H. (1995). Prediction of final data with use of preliminary and/or revised data. *Journal of Forecasting*, 14(4):351–380.
- [65] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., e Teller, E. (1953). Equations of states calculations by fast computing machines. *Journal of Chemical Physics.*, 21:1087–1092.
- [66] Migon, H. S. (1984). *An Approach to Non-Linear Bayesian Forecasting Problems with Applications*. Ph.d. thesis, Departamento de Estatística, Universidade de Warwick.
- [67] Motta, A. C. (1998). Utilização do filtro de Kalman em modelos estatístico. Iniciação científica, Departamento de Estatística, UNICAMP.
- [68] Pitt, M. K. e Shephard, N. (1999a). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94:590–599.
- [69] Pitt, M. K. e Shephard, N. (1999b). Joint MCMC methods for non-gaussian measurement time series. Artigo em Trabalho.
- [70] Pitt, M. K. e Walker, S. G. (1999). Construction of stationary time series with application to volatility models. Artigo em Trabalho.
- [71] Ripley, B. (1987). *Stochastic Simulation*. Nova Iorque : Wiley.
- [72] Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic variance models. *Journal of Econometrics*, 63:284–306.

- [73] Sandmann, G. e Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, 87:271–301.
- [74] Schweppe, F. (1965). Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory*, 11:61–70.
- [75] Shephard, N. (1994). Partial non-gaussian state space. *Biometrika*, 81:115–131.
- [76] Shephard, N. e Pitt, M. K. (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika*, 84:653–667.
- [77] Smith, A. F. M. e Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46(2):84–88.
- [78] Stroustrup, B. (1991). *The C++ Programming Language*. Addison Wesley, segunda edição edition.
- [79] Tanizaki, H. (1996). *Nonlinear Filters: Estimation and Applications*. Nova Iorque : Springer-Verlag, segunda edição edition.
- [80] Tanizaki, H. (1997). Nonlinear and nonnormal filters using Monte Carlo methods. *Comput. Stat. and Data Analysis*, 25:417–439.
- [81] Tanizaki, H. (1999a). Nonlinear and nonnormal filter using importance sampling: Antithetic Monte Carlo integration. *Communications in Statistics, Simulation and Computation*, 28(2):463–486.
- [82] Tanizaki, H. (1999b). On nonlinear and nonnormal filter using rejection sampling. *IEE Transactions on Automatic Control*, 44(2):314–319.

- [83] Tanizaki, H. (2000a). Nonlinear and non-gaussian state-space estimation: A quasi-optimal estimator. Artigo em Trabalho.
- [84] Tanizaki, H. (2000b). *Nonlinear and Non-Gaussian State-Space modeling with Monte Carlo Techniques: A Survey and Comparative Study*, chapter Stochastic Processes: Modeling and Simulation. Handbook of Statistics. Rao, C.R. and Shanbhag, D.N., North-Holland.
- [85] Tanizaki, H. e Mariano, R. S. (1994). Prediction, filtering and smoothing in nonlinear and nonnormal cases using Monte Carlo integration. *Journal of Applied Econometrics*, 9(2):163–179.
- [86] Tanizaki, H. e Mariano, R. S. (1998). Nonlinear and non-gaussian state-space modeling with Monte Carlo stochastic simulations. *Journal of Applied Econometrics*, 83(1):263–290.
- [87] Taylor, S. J. (1980). Conjectural models for trend in financial prices tests as forecasts. *Journal of the Royal Statistical Society (Série B)*, 143:338–362.
- [88] Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester: John Wiley.
- [89] Taylor, S. J. (1994). Modelling stochastic volatility: A review and comparative study. *Mathematical Finance*, 4:183–204.
- [90] Tierney, L. (1994). Markov Chains for exploring posterior distributions (with discussion). *Annals of Statistics*, 21:1701–1762.
- [91] Titterington, D. M., Smith, A. F. M., e Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Chinchester : John Wiley.

- [92] West, M. (1993). Approximating posterior distribution by mixtures. *Journal of the Royal Statistical Society (Série B)*, 55:409–442.
- [93] West, M. e Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*. Nova Iorque : Springer.
- [94] West, M., Harrison, P. J., e Migon, H. S. (1985). Dynamic generalised models and bayesian forecasting (with discussion). *Journal of the American Statistical Association.*, 80:73–97.
- [95] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. Nova Iorque : John Wiley.

ÍNDICE REMISSIVO

- amostrador
 - de Gibbs, 142
- amostragem
 - por importância, 52, 57, 145
 - por rejeição, 134, 135
- aproximação
 - de Monte Carlo
 - tipo 1 (MA1), 134
 - tipo 2 (MA2), 134
- conglomerados de volatilidade, 116
- densidade, 25
 - a posteriori, 95, 138
 - de filtragem, 26
 - filtragem, 1
 - importante, 50, 58, 145, 147
 - log-côncava, 22, 129
 - posteriori, 142
 - proposta, 143
 - suavizamento, 1
- distribuição, 33
 - log – qui – quadrado*, 133
 - t*-student, 33
 - a priori, 12
 - bimodal, 147
 - de erro generalizado, 33
 - degenerada, 20, 57
 - família exponencial, 11
 - log-qui-quadrado, 2
 - multivariada, 20
 - normal, 115
 - normal padrão, 33
 - posteriori, 12, 57
- equações, 17
 - de atualização, 16, 17
 - de medida, 14, 17
 - de observação, 14, 21, 145
 - de predição, 16, 17
 - de transição, 14, 145
 - observacional, 12
- espaço
 - de estado, 1, 12
 - Gaussiano, 1
 - não Gaussiano, 3, 143, 145
 - não linear, 143, 145
 - de estados não-Gaussiano, 21
- filtro
 - de Kalman, 1, 15
 - estendido, 17
 - suavizado, 59
 - de partículas, 25
 - com variável auxiliar, 26
- função

- de quasi-verossimilhança, 2, 47
- de verossimilhança, 50, 57, 60, 89, 139
- log-verossimilhança, 1
- modelo
 - aproximado, 49, 51, 59
 - ARCH, 2
 - da família Box-Jenkins (BJ), 13
 - de regressão, 7
 - de volatilidade estocástica, 1, 59, 133
 - dinâmicos, 6
 - dinâmicos generalizados, 11
 - Gaussiano, 50
 - linear Gaussiano, 50
- perturbações, 1, 12
 - Gaussianas, 12
 - não-Gaussianas, 21
- retornos, 31, 90, 115, 118
 - IBOVESPA, 89, 115
- série
 - simulada, 89–91
- teorema
 - de Bayes, 142
- variável
 - antitética balanceada
 - por escala, 54
 - por locação, 53–55, 58
 - controle, 50, 54, 56
- volatilidade
 - estimada, 110
 - suavizada, 105
 - verdadeira, 105, 110