SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 17.06.2005

Assinatura: Www. Walle wu

Modelos estocásticos contínuos e discretos aplicados em finanças

Juliana Cobre

Orientador: Prof. Dr. Marinho Gomes de Andrade Filho

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Área: Ciências de Computação e Matemática Computacional.

USP - São Carlos Junho/2005

A meus queridos pais.

Agradecimentos

Agradeço a Deus e ao meu anjo da guarda pela saúde e por todos os insights. Por terem colocado em meu caminho pessoas maravilhosas, que diretamente ou indiretamente me ajudaram muito.

Agradeço especialmente aos meus pais, João e Edua, que sempre me deram incentivo em seguir meus estudos e muito conforto para cumpri-los. São fontes de forças na minha luta em sempre dar o melhor de mim. Exemplos de honestidade e respeito.

Agradeço às amigas que moraram comigo, Carla Bossu, Juliana B. Garrido, Karina K. de Lima, Priscila Ap. de Moraes e Viviane S. Tonaki, que me suportaram nos momentos de crise, que me deram apoio e que me divertiram muito.

Agradeço aos funcionários do ICMC, aos professores da matemática, em especial ao Prof. Dr. Wagner Vieira Leite Nunes que me orientou durante a graduação, é um dos primeiros responsáveis pelo meu início de carreira, me auxilia e me atura até hoje! Sem dúvida um grande amigo.

Agradeço ao Prof. Dr. Mário de Castro. Mais que um professor do instituto, ele foi para mím um colega de laboratório, um solucionador de muitos dos meus problemas com o computador e com os gráficos, um incentivador da pesquisa. Sempre aprendi muito com ele, mesmo nas conversas mais descontraídas, fora dos projetos científicos. Uma pessoa admirável.

À amiga Sandra C. de Oliveira que me animava a participar dos eventos científicos, pelas dicas de sites, materiais e da abordagem dos resultados. Por tanto ter ouvido meus desabafos nos momentos difíceis. E também pelos bons momentos de descontração. Agradeço à Vera L. D. Tomazzella pelo material que me forneceu e por ser tão prestativa.

Também agradeço muito a todo o pessoal do laboratório de estatística, que sempre colaborou nas minhas dúvidas.

E finalmente agradeço a men orientador, Prof. Dr. Marinho G. de Andrade Filho, pelo projeto e pelos momentos dedicados a ele.

Resumo

Os modelos de volatilidade estocástica (MVE) são bastante utilizados pela sua semelhança com os modelos habitualmente usados na Teoria Financeira. Nos MVE a volatilidade
independe dos retornos passados e é modelada como uma variável latente não observada,
através de uma componente preditível e outra aleatória. A função de verossimilhança desses
modelos é difícil de ser obtida e maximizada. Neste trabalho descrevemos as suposições
em que os modelos de difusão para séries de retornos se baseiam, assim como as suposições
tomadas pela modelagem discreta. Apresentamos os MVE e alguns de seus métodos de estimação. Tratamos de dois modelos contínuos, de algunas de suas propriedades e também de
dois MVE discretos que convergem para tais contínuos. Trabalhamos com uma aproximação
linear de um deles, apresentando o filtro de Kahnan, e sua verossimilhança obtida depois da
filtragem. O algoritmo de Metropolis-Hastings foi empregado na abordagem da verossimilhança, assim como na bayesiana do caso linear. Utilizamos o filtro estendido de Kalman
combinado com a aproximação de Laplace na construção da função de verossimilhança dos
dois MVE abordados neste trabalho.

Abstract

The stochastic volatility models (SVM) are quite habitually used by their similarity with the models used in the Financial Theory. In them the volatility is described through their last values and it does not depend on the last returns. The likelihood function of SV is difficult of being obtained and maximized. In this paper, we have described the hypothesis in which the diffusion models for series of returns are based, as well as the suppositions taken by the discrete modelling. We presented SVM and some of their estimate methods. We treated of two continuous models, of some of their properties and also of two discrete SVM that converge for the continuous ones. We worked with a linear approach of one of them, presenting the Kalman filter, and its likelihood obtained after the filtration. The Metropolis-Hastings algorithm was used in the approach of the likelihood, as well as in the Bayesian of the linear case. We used the extended Kalman filter combined with the Laplace approximation in the construction of the likelihood function of the two SVM approached in this work.

Sumário

1	Inti	rodução 1
	1.1	Terminologia em Finanças
		1.1.1 Retorno
		1.1.2 Opção
		1.1.3 Volatilidade em Mercados Financeiros
	1.2	Dinâmica do Preço do Ativo
		1.2.1 Modelos Discretos
	1.3	Modelos de Volatilidade Estocástica
	1.4	Métodos de Estimação
2	Mo	delos I e II
	2.1	Preliminares
	2.2	Modelo discreto por modelo de difusão
		2.2.1 Modelo I
		2.2.2 Modelo II
	2.3	Conhecendo os Modelos
		2.3.1 Modelo I
		2.3.2 Modelo II
	2.1	Couclusão

X	SUMÁRIO

3	Mo	delo Linear	27
	3.1	Filtro de Kalman	27
		3.1.1 Obtendo o Filtro de Kalman	28
	3.2	Modelo Invariante no Tempo	31
		3.2.1 Verossimilhança	33
		3.2.2 Abordagem Bayesiana	36
	3.3	Conclusão	39
4	Infe	erência para o Modelo I	41
	1.1	Filtro Estendido de Kalman	41
	1.2	Modelos Condicionalmente Gaussianos	43
	1.3	MVEM	44
		4.3.1 Verossimilhança	14
	4.4	Abordagem Bayesiana	47
	1.5	Modelo I	48
	1.6	Modelo II	50
	4.7	Conclusão	52
5	Res	ultados	53
	5.1	Abordagem da Verossimilhança	53
	5.2	Abordagem Bayesiana	57
	5.3	Conclusão	60
A	Mét	codos Numéricos	61
	A.1	MCMC	61
		A.4.1 Amostrador de Gibbs	62
		A.1.2 Metropolis-Hastings	62
	$\Lambda.2$	Aproximação de Laplace	63
	A.3	Algoritmo de Newton-Raphson	64
	A.4	Critério de Geweke	65
В	Mod	delos Discretos	67
	B.1	Modelos ARCH(p)	67

SUMÁRIO		
B.2 Modelos GARCH(p,q)	71	
B.3 Modelos EGARCII(p,q)	73	
B.4 Modelos ARCH-M	73	
B.5 Modelos IGARCII	74	
Referências Bibliográficas	79	

xii SUMÁRIO

Lista de Figuras

1.1	Série de retorno e série de log-retorno da Telebras $(02/01/92 \text{ a } 05/01/96)$,
5.1	Histogramas das amostras selecionadas: MIT A	55
5.2	Gráficos das amostras selecionadas: MIT A	58
5.3	Histogramas das amostras selecionadas: MIT B	50
5.4	Gráficos das amostras selecionadas: MIT B	56
5.5	Histogramas das amostras selecionadas: MIT A bayesiano	58
5.6	Gráficos das amostras selecionadas: MIT A bayesiano	58
5.7	Histogramas das amostras selecionadas: MIT B bayesiano.	59
5.8	Gráficos das amostras selecionadas: MIT B bayesiano	59

Lista de Tabelas

5.1	Resultados para o MIT A	5
5.2	Resultados para o MIT B	5
5.3	Resultados para o MIT A bayesiano	5
5.4	Resultados para o MIT B bayesiano.	5

Capítulo

1

Introdução

Os primeiros estudos dedicados a descrever o comportamento do retorno ou log-retorno de uma ação, (veja, na seção 1.1, a definição dos termos financeiros utilizados neste trabalho), enfocavam apenas o primeiro momento do modelo. Momentos de ordem mais alta, como por exemplo a variância, eram ignorados pois eram considerados constantes. No entanto, na prática, isto não ocorre. Um exemplo simples desse fato é notar que em dias com maior números de transações o segundo momento é, em geral, maior do que em dias não comerciais.

A confirmação da presença de heterocedasticidade nas séries financeiras incentivou as pesquisas na modelagem da volatilidade, nome dado em finanças a variância condicional de uma variável. Uma das técnicas capaz de caracterizar o comportamento da volatilidade do retorno foi introduzida por Engle (1982) e é chamada de processo auto-regressivo com heterocedasticidade, ou simplesmente ARCH. São modelos a tempo discreto fundamentados na idéia de que a volatilidade é caracterizada por valores passados dos retornos. Outros modelos discretos foram desenvolvidos a partir do ARCH, como por exemplo GARCH proposto por Bollerslev (1986), EGARCH apresentado por Nelson (1991), IGARCH por Engle & Bollerslev (1986), entre outros (veja p.ex. Hamilton, 1994). Uma ampla revisão das pro-

priedades desses modelos pode ser encontrada em Bollerslev, Chou & Kroner (1992) e uma breve descrição desses modelos é encontrada no apêndice B.

A importância dos modelos ARCH veio de certa forma da sua direta associação entre variância e risco e da fundamental relação entre risco e retorno (no comércio de compra). Uma aplicação desse modelo em séries financeiras brasileiras pode ser vista em Issler (1999). Três das mais proeminentes teorias de precificação de ação são fundamentadas nos modelos ARCH: Capital Asset Pricing Model (CAPM), Consumption-Based CAPM e Arbitrage Pricing Theory (APT).

Existem também os modelos de volatilidade estocástica (MVE) propostos originalmente por Taylor (1982). A fundamental diferença entre esses modelos e os modelos do tipo ARCH é que, nos MVE, a volatilidade é modelada como uma variável latente não observada, através de uma componente preditível e outra aleatória.

Há ainda os modelos a tempo contínuo que se propõem a estimar a volatilidade. Esses modelos são baseados em sistemas de equações diferenciais estocásticas e também são MVE, ou seja, modelam a volatilidade por um processo não observado. Como as observações podem ser feitas apenas a tempo discreto, existe um grande interesse em aproximar um modelo de difusão por um modelo a tempo discreto e vice-versa, (veja p.ex. Nelson, 1990a).

A verossimilhança de sistemas de equações diferencias estocásticas não lineares observados em tempos discretos pode ser muito difícil de ser calculada, especialmente no caso dos MVE que são descritos por variáveis de estado não observadas. Mais detalhes a respeito das dificuldades e das vantagens deste tipo de modelo serão dados nas seções 1.3 e 1.4.

Este trabalho está organizado da seguinte forma: nas seções seguintes introduziremos alguns termos financeiros necessários para seu entendimento, apresentaremos as suposições em que os modelos de difusão para séries de retornos se baseiam, assim como as suposições tomadas pela modelagem discreta. Na seção 1.3 introduziremos os modelos de volatifidade estocástica e, em seguida, alguns de seus métodos de estimação. No capítulo 2 apresentaremos os modelos contínuos tratados neste trabalho, algumas de suas propriedades e também aproximações discretas que convergem para os contínuos. Para facilitar a compreensão da metodologia aplicada aos modelos abordados neste trabalho, trabalharemos primeiramente com uma aproximação linear no capítulo 3, apresentando o filtro de Kalman, e a verossimilhança obtida depois da filtragem. No capítulo 4 consta uma extensão do filtro de Kalman,

e uma forma de usá-lo na construção da função de verossimilhança dos modelos aqui tratados. Finalmente no capítulo 5 apresentaremos os resultados empíricos referentes ao terceiro capítulo. No apêndice A consta a descrição dos métodos numéricos utilizados neste trabalho. Já no apêndice B há uma breve apresentação dos modelos discretos aqui citados.

1.1 Terminologia em Finanças

Nesta seção, apresentamos uma breve explicação de certos termos comuns em finanças que são usados neste trabalho. Para mais detalhes sugerimos (p.ex. Hull, 1996).

1.1.1 Retorno

Seja S_t o preço de uma ação no instante t, o retorno da ação no intervalo (t-1,t), $\Delta S_t,$ é

$$\Delta S_t = S_t - S_{t-1}, \tag{1.1}$$

ou seja, retorno é a variação do preço de um ativo (no caso consideramos uma ação). Já variação relativa de preços de um ativo é chamada de retorno líquido simples e é dada por

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}} - \frac{S_t}{S_{t-1}} - 1. \tag{1.2}$$

Normalmente R_t é expresso em porcentagem sendo também chamado de taxa de retorno. À fração S_t/S_{t-1} damos o nome de retorno bruto simples.

É muito comum enfocarmos o retorno composto continuamente ou simplesmente logretorno de uma ação definido por

$$H_t = \log \frac{S_t}{S_{t-1}} = \log S_t - \log S_{t-1}, \tag{1.3}$$

A série de retornos tem características mais tratáveis do que as séries financeiras, pois os retornos raramente apresentam tendências ou sazonalidades; são em geral não-correlacionados; os quadrados dos retornos são auto-correlacionados; as séries de retornos apresentam agrupamento de volatilidade ao longo do tempo; além disso a distribuição não condicional dos retornos tem caudas mais pesadas comparadas com as da normal e é leptocúrtica, embora quase simétrica. Algumas dessas considerações podem ser observadas comparando os gráficos das séries de retorno e log-retorno do índice de fechamento da Telebras dados na figura 1.1.1.

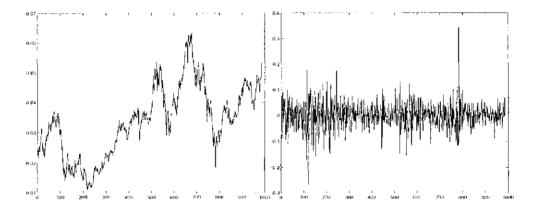


Figura 1.1: Série de retorno e série de log-retorno da Telebras (02/01/92 a 05/01/96).

Uma das explicações a respeito do comportamento leptocúrtico das distribuições dos retornos dada em finanças é baseada na maneira como as informações chegam ao mercado financeiro. Com freqüência várias informações ou "notícias" chegam em um mesmo intervalo de tempo. A chegada de novas informações é que impulsiona o comércio de ações fazendo com que os preços sofram muitas mudanças. Do contrário, ou seja, quando poucas informações chegam ao mercado financeiro, o valor das ações varia pouco. É devido aos extremos pouca ou muita mudança que a distribuição tem candas mais pesadas e picos mais agudos em relação aos da distribuição normal.

1.1.2 Opção

Suponha que tenhamos um título, chamado opção, no tempo t=0 que nos dá o direito de comprar ou vender uma ação antes do ou no tempo T, tempo de maturidade ou tempo de vencimento da opção. Se pudermos exercer essa opção, ou seja, exercer o direito de compra ou de venda da ação a um determinado preço K, chamado de preço de exercício da opção, somente no tempo de vencimento T, teremos uma opção curopéia (de compra ou de venda). Caso possamos exercer a opção antes do tempo de exercício, temos uma opção americana. No mercado financeiro existem outros tipos de opções, no entanto o descrito acima é suficiente para que se entenda o conteúdo deste trabalho.

O titular de uma opção não é obrigado a exercê-la. Por exemplo, se no tempo T o preço da ação S_T mais o preço da opção é menor do que K (no caso de opção de compra), a opção não será exercida, pois pode-se obter a ação pelo preço S_T no mercado financeiro.

Saber qual será o preço de uma ação é muito importante para se estudar o preço de exercício e mesmo o preço da opção. Um dos primeiros modelos de precificação de opção surgiu no trabalho de Black & Scholes (1973) e se restringe a modelar o valor das opções de compra européia. Já as publicações Engle (1982) e Bollerslev (1986) modelam o retorno de uma ação e se basciam em premissas diferentes das de Black & Scholes.

1.1.3 Volatilidade em Mercados Financeiros

Volatilidade é o nome genérico dado à variância condicional de uma variável aleatória. Em finanças, o termo é habitualmente usado para denotar a variância condicional dos retornos de uma ação. Portanto, a volatilidade do preço de uma ação é a medida da incerteza quanto às oscilações futuras em seu preço. Quanto maior a volatilidade, maior a possibilidade de a ação ter um desempenho tanto muito bom quanto muito ruim.

Se o comportamento dos retornos fosse homocedástico e normalmente distribuído, uma estimativa para a volatilidade seria

$$\sigma_t^2 = \frac{\sum_{j=1}^{N-1} \Delta S_j^{-2}}{N}, \qquad t = N, N+1, \dots,$$
 (1.4)

em que ΔS_j é o retorno de uma ação no tempo j, veja (Issler, 1999, veja).

Neste trabalho tratamos por volatilidade a chamada volatilidade estatística. Existe também a volatilidade implícita que está associada a mudança que o preço de uma opção (não o de uma ação) sofre no mercado. A volatilidade implícita, que também considera homocedástico o comportamento dos retornos, é dada pela solução da fórmula de Black & Scholes. Nessa é assumido que o preço de uma ação tem distribuição log-normal, apesar das evidências empíricas mostrarem o contrário, (veja p.ex. Fama, 1965; Hsich, 1989; Lee & Tse, 1991). O trabalho Black & Scholes (1973) é um ponto de referência para o estudo das noções de volatilidade, mesmo porque esse modelo se baseia em um processo estocástico a tempo contínuo, e tal tipo de processo é amplamente utilizado para descrever o comportamento dos preços de ações.

Nas séries de retornos são observados aglomerados de volatilidade (volatility clustering). São episódios de alta ou baixa volatilidade que estão intimamente ligados com as caudas densas da distribuição dos retornos. A modelagem dessas séries é essencialmente construída para descrever esse comportamento. Além disso pode-se notar o chamado "efeito alavança",

ou seja, a resposta da volatilidade é mais rápida a retornos negativos do que a retornos positivos.

1.2 Dinâmica do Preço do Ativo

Para modelar o preço das ações por processos de difusão são necessárias algumas suposições iniciais. Vamos considerar um ativo financeiro, por exemplo uma ação, com cotação diária denotada por S_t . Seja Ω_t o conjunto de informações obtidas até o instante t e considere a distribuição condicional do retorno S_{t+h}/S_t da ação sobre o período [t, t+h] dado Ω_t . Vamos assumir que os retornos do ativo têm esperança condicional finita, ou seja,

$$\mathbb{E}(S_{t+h}/S_t|\Omega_t) = S_t^{-1} \mathbb{E}(S_{t+h}|\Omega_t) < +\infty$$
(1.5)

e também, variància condicional finita

$$\operatorname{Var}(S_{t+h}/S_t|\Omega_t) = S_t^{-2} \operatorname{Var}(S_{t+h}|\Omega_t) < +\infty.$$
(1.6)

A taxa de retorno continuamente ajustada é caracterizada por

$$h^{-1}\log \mathbb{E}(S_{t+h}/S_t|\Omega_t). \tag{1.7}$$

Então podemos formular as hipóteses acima como segue.

Hipótese 1.1. A taxa de retorno esperada continuamente ajustada converge quase certamente para valores finitos, $\mu_S(\Omega_t)$, quando h tende a zero pela direita. Temos então que

$$E(S_{t+h}|\Omega_t) - S_t \sim h\mu_s(\Omega_t)S_t, \tag{1.8}$$

o que na notação diferencial é o mesmo que

$$\frac{d}{d\tau} \mathbf{E}(S_{\tau}|\Omega_t) \bigg|_{\tau=t} = \mu_s(\Omega_t) S_t, \quad \text{quase certamente}.$$
 (1.9)

Hipótese 1.2. A variância condicional do retorno $h^{-1}\text{Var}(S_{t+h}/S_t|\Omega_t)$ converge quase certamente para um valor finito $\sigma_S^2(\Omega_t)$ quando h tende a zero pela direita. Na notação diferencial temos

$$\frac{d}{d\tau} \operatorname{Var}(S_{\tau}|\Omega_t) \bigg|_{\tau=t} = \sigma_S^2(\Omega_t) S_t^2, \quad \text{quase certainente}.$$
 (1.10)

As hipóteses 1.1 e 1.2 nos levam a representar a dinâmica do preço do ativo da seguinte forma

$$dS_t = \mu_S(\Omega_t) S_t dt + \sigma_S(\Omega_t) S_t dW_t, \qquad (1.11)$$

em que W_t é um movimento browniano. Logo, temos definido o chamado processo de volatilidade instantânea $\sigma_S(\Omega_t)$ que pode ser escrito, seguindo a notação acima, na forma

$$\sigma_S(\Omega_t) = \left(\lim_{h\downarrow 0} h^{-1} \text{Var}(S_{t+h}/S_t|\Omega_t)\right)^{1/2}.$$
 (1.12)

O uso do movimento browniano na precificação de ação se justifica devido aos retornos dos ativos não serem totalmente previsíveis. Isto se dá pela eficiente e quase que instantânea chegada de informações ao mercado financeiro provenientes, atualmente, do mundo inteiro. A demanda de informações que chegam ao mercado justifica também a suposição de que os retornos em períodos regulares de tempo, [t+k,t+k+1], k=0,2,...h-1, são independentes e identicamente distribuídos.

Vale notar que a fórmula de Black & Scholes foi baseada num processo com $\mu_S(\Omega_t) = \mu_S$ e $\sigma_S(\Omega_t) = \sigma_S$ constantes para todo t, ou seja, baseada na idéia de que o preço da ação segue um movimento browniano geométrico.

Agora consideremos que a dinâmica do preço do ativo é governada por variáveis de estado não observadas, como feito, de forma simplificada, no modelo de Hull & White (1987). Assim, assumimos que um processo de difusão U_t descreve a variável de estado como

$$\begin{cases} dS_t = \mu_t S_t dt + \sigma_t S_t dW_t \\ dU_t = \gamma_t dt + \delta_t dW_t^U \\ \operatorname{Cov}(dW_t, dW_t^U) = \rho_t dt \end{cases}$$

$$(1.13)$$

em que $\mu_t, \sigma_t, \gamma_t$ e ρ_t são como na hipótese a seguir.

Hipótese 1.3. $\mu_t, \sigma_t, \gamma_t$ e ρ_t são $I_t^U = [U_\tau, \tau \le t]$ adaptados.

A hipótese 1.3 nos dá que o processo U identifica a dinâmica do preço do ativo S. E implica que, dada uma trajetória das variáveis de estado $(U_{\tau})_{0 \leq \tau} \leq T$, os retornos consecutivos $S_{t_{k+1}}/S_{t_k}, 0 \leq t_1 < t_2 < \ldots < t_k \leq T$ são estocasticamente independentes e com distribuição log-normal.

1.2.1 Modelos Discretos

Nesta seção trataremos de modelar a dinâmica de um ativo por um processo discreto no tempo análogo ao processo descrito em (1.13). Isto também justifica a utilização da aproximação de Euler neste trabalho. Para maiores detalhes veja capítulo 2.

Um modelo discreto baseado no modelo (1.13) é dado por

$$\log(S_{t+1}/S_t) = \mu(U_t) + \sigma(U_t)\varepsilon_{t+1}. \tag{1.14}$$

Como feito na modelagem contínua, devemos impor algumas hipóteses plausíveis capazes de assegurar a descrição das características das séries financeiras.

Hipótese 1.4. O processo ε_t é independente e identicamente distribuído (i.i.d.) e também independente da variável de estado do processo.

Considerando esta hipótese, podemos interpretar estatisticamente as funções $\mu(U_t)$ e $\sigma(U_t)$ como coeficiente de tendência e coeficiente de volatilidade, respectivamente. De fato,

$$\mathbb{E}[\log(S_{t+1}/S_t)|\Omega_t] = \mathbb{E}[\mathbb{E}[\log(S_{t+1}/S_t)|(U_\tau, \varepsilon_\tau, \tau \le t)]|\Omega_t]$$

$$= \mathbb{E}[\mu(U_t)|\Omega_t], \qquad (1.15)$$

já que $\mathrm{E}[\varepsilon_{t+1}|(U_{\tau},\varepsilon_{\tau},\tau\leq t)]=\mathrm{E}[\varepsilon_{t+1}|\varepsilon_{t},\tau\leq t]=0$ devido a não influência de U_{t} em ε_{t} de acordo com a hipótese acima. Da mesma forma mostramos que

$$\operatorname{Var}[\log(S_{t+1}/S_t) - \mu(U_t)|\Omega_t] = \operatorname{E}[[\log(S_{t+1}/S_t) - \mu(U_t)|\Omega_t]^2] - \underbrace{\left[\operatorname{E}[\log(S_{t+1}/S_t) - \mu(U_t)|\Omega_t]\right]^2}_{=0}$$

$$= \operatorname{E}[[\log(S_{t+1}/S_t) - \mu(U_t)|\Omega_t]^2] = \operatorname{E}[[\sigma(U_t)\varepsilon_{t+1}]^2]$$

$$= \operatorname{E}[\sigma^2(U_t)|\Omega_t]\operatorname{E}[\varepsilon_{t+1}^2|\Omega_t] = \operatorname{E}[\sigma^2(U_t)|\Omega_t],$$
(1.16)

em que a penúltima passagem se dá pela hipótese acima.

Hipótese 1.5. Seja $\Omega_t^R \equiv \sigma[S_\tau/S_{\tau-1}: \tau=0,1,\ldots,t-1,t]$ a σ - álgebra gerada pelo conjunto de retornos passados. Suponhamos, então, $\mu(U_t)$ seja uma função Ω_t^R - mensurável.

Incluindo essa última hipótese nas equações (1.15) e (1.16) concluímos que

$$E[\log(S_{t+1}/S_t)|\Omega_t^R] = \mu(U_t)$$
(1.17)

e

$$\operatorname{Var}[\log(S_{t+1}/S_t) - \mu(U_t)|\Omega_t^R] = \operatorname{E}[\sigma^2(U_t)|\Omega_t^R]. \tag{1.18}$$

O trabalho Ghysels, Harvey & Renault (1995) aborda a questão da dinâmica dos preços de ativos, destacando o caso de o ativo ser uma opção.

1.3 Modelos de Volatilidade Estocástica

Os modelos de volatilidade estocástica (MVE) são bastante usados pela sua semelhança com os modelos habitualmente usados na teoria financeira. Comparados com os modelos do tipo ARCH, os MVE são capazes de descrever de forma mais adequada as propriedades observadas nas séries diárias de ativos financeiros, (veja p.ex. Camero, Peña & Ruiz, 2004). Além disso, na classe ARCH é suposto que a variância condicional depende dos retornos passados. Já no modelo proposto inicialmente por Taylor (1982), a volatilidade é descrita através de seus valores passados e independe dos retornos passados.

Consideremos agora um processo descrito por

$$H_{t+1} = \mu_t + \sigma_t \varepsilon_{t+1} \,, \tag{1.19}$$

em que μ_t é uma função mensurável pelo conjunto de observações $H_{\tau} \in \Omega_{\tau}^R, \tau \leq t$. Pela equação (1.18) temos que

$$\operatorname{Var}[H_{t+1}|\Omega_t^R] = \operatorname{E}[\sigma_t^2|\Omega_t^R], \qquad (1.20)$$

sugerindo que

- 1. os aglomerados de volatilidade podem ser capturados por um processo com esperança condicional descrita por um modelo auto-regressivo; e
- 2. as caudas densas podem ser obtidas através de erros ε_t com distribuição de caudas pesadas ou por características estocásticas da $\mathrm{E}[\sigma_t^2|\Omega_t^R]$.

Geralmente um processo que detém as características dadas em (1) e (2) é um processo auto-regressivo de ordem 1, AR(1), para alguma função não-linear de σ_t . Em suma, assume-se que a volatilidade segue um processo de Markov de ordem um, não necessariamente linear em σ_t . Foi isso que motivou o trabalho de Andersen (1994) a introduzir a classe de modelos auto-regressivo de volatilidade estocástica (MARVE).

Essa classe engloba vários modelos já conhecidos na literatura, como é o caso do proposto por Taylor (1982), um modelo auto-regressivo de variância aleatória descrito por

$$y_{\ell} = \sigma_{\ell} \varepsilon_{\ell} \tag{1.21}$$

$$\log \sigma_t = \alpha + \phi \log \sigma_{t-1} + \eta_t \,, \tag{1.22}$$

sendo η_t um ruído branco.

A forma canônica de Kim, Kim, Shephard & Chib (1998), também faz parte da classe MARVE e é descrita por

$$y_t = \sigma^* e^{\nu_t/2} \varepsilon_t \tag{1.23}$$

$$\nu_{t+1} - \alpha = \beta(\nu_t - \alpha) + \sigma_\eta \eta_t \,, \tag{1.24}$$

em que ε_t e η_t são normalmente distribuídos com média zero e variância unitária.

O MVE mais popular na literatura dentro desse contexto é também um MARVE. E é dado por

$$y_t = \mu_t + \sigma^* e^{\nu_t/2} \varepsilon_t \tag{1.25}$$

$$\nu_t = \phi \nu_{t-1} + \sigma_\eta \eta_t \,, \tag{1.26}$$

em que $\nu_t = \ln(\sigma_t^2/\sigma^{*2})$. Mais detalhes a respeito deste tipo de modelo serão dados na seção 2.3.1.

As propriedades dos MARVE têm sido estudadas em vários trabalhos (veja p.ex. Taylor, 1994; Barndorff-Nielsen & Shephard, 2001). Há também trabalhos que, ao invés de tomarem erros ε_t com distribuição Normal, assumem erros com distribuições de caudas pesadas, (veja p.ex. Harvey, Ruiz & Shephard, 1994; Sandmann & Koopman, 1998; Watanabe & Asai, 2001). No entanto, vale ressaltar que os trabalhos Andersen, Bollerslev, Diebold & Labys (2001) e Andersen, Bollerslev, Diebold & Labys (2003) mostram que a log-volatilidade pode ser aproximada por uma distribuição normal, ou seja, considerar erros gaussianos é adequado.

A importância dos MARVE também se dá por poder incorporar o efeito alavanca (comentado na seção 1.1.3). Para isso, o trabalho Harvey & Shephard (1996) propõe que os erros ε_t e η_{t-1} do MARVE sejam negativamente correlacionados. Para o mesmo fim, o trabalho de Jacquier, Polson & Rossi (2002) sugere que os erros ε_t e η_t sejam correlacionados.

1.4 Métodos de Estimação

Ao contrário do que ocorre com os modelos ARCH, a função de verossimilhança dos MVE é difícil de ser obtida e maximizada. Nos últimos anos, várias técnicas de estimação foram estudadas a fim de suprir estas dificuldades e permitir as aplicações empíricas dos MVE. Basicamente, tais técnicas se dividem em dois grupos: métodos que visam a construção da verossimilhança exata e métodos que buscam aproximá-la on simplesmente a evitam. Trabalhos como Kim et al. (1998) e Sandmann & Koopman (1998) se encaixam no primeiro grupo. O método da quase-verossimilhança usado por Harvey et al. (1994), assim como o método dos momentos originalmente proposto por Taylor (1986) são exemplos de técnicas do segundo grupo.

Um dos principais objetivos da modelagem de uma série financeira é a estimação de sua volatilidade. Nos MVE a volatilidade é descrita por um processo não observado, o que requer o uso de filtros para sua estimação. Nos modelos lineares com relação a variável de estado em que a variável observada é descrita por equações cujo ruído não depende da variável de estado, o uso do filtro de Kalman (FK) cumpre o objetivo em questão. Os MVE não possuem essa característica, pois caso tivessem, pela equação (1.14), deveríamos ter $\sigma(U_t)$ uma função constante, ou ao menos, independente da variável de estado. Isto implicaria, pela equação (1.16), em um modelo homocedástico, ou condicionalmente homocedástico (com relação ao conjunto de observações passadas). No entanto, neste trabalho, propomos uma aproximação linear em que o FK se adequa, veja seção 3.2.

Quando não temos modelos lineares mas ainda que tenham ruído independente da variável de estado, podemos aplicar o filtro estendido de Kalman (FEK). Sua construção está feita no capítulo 4. No entanto, nos modelos tratados neste trabalho, além da linearidade não ser atendida, o ruído está multiplicado por uma função da variável não-observada. Nesse caso, não é possível o uso completo do FEK, usaremos apenas seu primeiro estágio. O

inconveniente dessa abordagem é que ela não nos fornece um filtro capaz de estimar os estados, e sim apenas nos permite estimar os parâmetros do modelo, (veja seção 4.3).

O desenvolvimento de métodos de integração numérica como por exemplo os métodos de integração de Monte Carlo via cadeias de Markov (MCMC) e os métodos de re-amostragem por importância, veja apêndice A.1, permitem a obtenção das estimativas de máxima verossimilhança para os parâmetros dos MVE. A re-amostragem por importância, comparada com os algoritmos MCMC, é menos custosa computacionalmente e evita problemas de convergência. Além disso, suas estimativas podem ficar mais exatas aumentando o número de iterações. Já os algoritmos MCMC são mais flexíveis e capazes de dividir problemas de dimensões altas em casos de menor dimensão. O trabalho de Sandmann & Koopman (1998) é um exemplo do emprego do MCMC dentro desse contexto. Esse utiliza uma trans formação na equação das observações de um MVE básico, fazendo com que esta seja linear na dependência dos estados e tenha ruído log qui-quadrado. No caso de modelos com média estocástica, essa linearização não é possível. Mas há outras formas de abordagens desses modelos. Uma delas é dada por Koopman & Uspensky (2002) que faz uso do MCMC para se obterem as estimativas de máxima verossimilhança.

Para conhecer outros métodos de estimação recomendamos Pelegrín (2004) assim como os trabalhos nele citados e Ghysels *et al.* (1995).

Modelos I e II

Apresentaremos aqui os dois modelos enfocados neste trabalho, juntamente com suas aproximações discretas que seguem primeiramente, pois algumas de suas propriedades dependem delas. Antes disso, introduzimos alguns conceitos básicos para o entendimento do capítulo, baseados em Morettin & Toloi (2004), Çinlar (1975) e Øksendal (1998).

2.1 Preliminares

Definição 2.1 (Processo Estocástico). Seja T um conjunto arbitrário. Um processo estocástico é uma família $Z = \{Z(t), t \in T\}$, tal que, para cada $t \in T$, Z(t) é uma variável aleatória.

Informalmente podemos dizer que um processo estocástico é o conjunto de todas as possíveis trajetórias de um certo processo. Da definição formal concluímos que processo estocástico é uma família de variáveis aleatórias.

Definição 2.2 (σ -álgebra). Seja Ω um conjunto qualquer. Então a σ -álgebra \mathcal{F} com relação a Ω é uma família F de subconjuntos de Ω com as seguintes propriedades

- (i) $\emptyset \in \mathcal{F}$;
- (ii) $F \in \mathcal{F} \Rightarrow F^C \in \mathcal{F}$, em que F^C denota o conjunto complementar de F em Ω ;
- (iii) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow A \doteq \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$

Definição 2.3 (Processo Mensurável). Considere o espaço de probabilidades (Ω, \mathcal{F}, P) . Dizemos que $X : \Omega \to \mathbb{R}$ é um processo mensurável com relação a \mathcal{F} , ou simplesmente \mathcal{F} -mensurável, se e somente se $\{\omega : X(\omega) < a\} \in \mathcal{F}$, para todo $a \in \mathbb{R}$.

Definição 2.4 (Processo Adaptado). Seja \mathcal{M}_t uma família crescente de σ-álgebras de subconjuntos de Ω . Um processo $B_t(\omega) \in \mathbb{R}$, para $t \geq 0$ e $\omega \in \Omega$, é um processo adaptado com relação a \mathcal{M}_t , chamado \mathcal{M}_t -adaptado, se para cada $t \geq 0$ a função $\omega \to B_t(\omega)$ é \mathcal{M}_t -mensurável.

De maneira simples, podemos dizer que um processo $B_t(\omega)$ é \mathcal{M}_t -adaptado quando as informações "contidas" na σ -álgebra referente são suficientes para "descrevê-lo".

Definição 2.5 (Processo de Markov). Um processo estocástico $\{X_n, n \in \mathbb{N}\}$ é um processo de Markov com espaço de estado E enumerável e finito se

$$P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = P(X_{n+1} = j | X_n = i_n), \tag{2.1}$$

para todo $n \in \mathbb{N}$ e $i_0, i_1, \ldots, i_n \in E$.

De maneira informal, podemos dizer que um processo estocástico é um processo de Markov se a probabilidade de movimento para j depende apenas do estágio imediatamente anterior, i_n , ou seja, não importa a trajetória, e sim o "tamanho" do passo.

Definição 2.6 (Processo de Wiener). O processo de Wiener ou movimento browniano $W = \{W_t, t \ge 0\}$ é um processo estocástico com as seguintes propriedades

(i)
$$W_0 = 0$$
;

2.1. PRELIMINARES 15

- (ii) os incrementos $W_t W_s$ são estacionários e independentes;
- (iii) para todo t > 0, $W_t \sim \mathcal{N}(0, t)$;
- (iv) as trajetórias são contínuas quase certamente (q.c.).

Observação 2.7. A condição de estacionariedade junto com o fato de W_t ter distribuição normal implicam que $W_t - W_s$, para t > s tem distribuição normal com média zero e variância t - s.

Definição 2.8 (Equação Diferencial Estocástica). Uma equação diferencial estocástica (EDE) dada por

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t , \qquad (2.2)$$

em que $a(X_t,t)$ e $b(X_t,t)$ são funções conhecidas, é apenas uma forma abreviada de representar a equação integral estocástica

$$X_{t} = X_{0} + \int_{0}^{t} a(X_{s}, s)ds + \int_{0}^{t} b(X_{s}, s)dW_{s}.$$
(2.3)

A equação (2.2) não faz sentido matemático, já que ela envolve a derivada de um processo de Wiener que não tem derivada em quase todo ponto (exceto como processo generalizado). Representando a EDE na forma (2.3) damos sentido a ela, pois sua solução pode ser dada explicitamente. Sobre certas condições impostas às equações $a(X_t, t)$ e $b(X_t, t)$ é possível mostrar que a EDE tem uma única solução X_t , veja Øksendal (1998). Tal solução é um processo estocástico e não uma solução determinística.

Definição 2.9 (Aproximação de Euler). Seja X_t , $t_0 \le t \le T$ um processo de Itô tal como em (2.42) e com condição inicial $X_{t_0} = X_0$. Tomemos uma discretização $t_0 < t_1 < \ldots < t_n \ldots < t_N = T$ do intervalo de tempo [0,T]. Seja Y = Y(t) para $t_0 \le t \le T$ um processo estocástico contínuo no tempo com condição inicial $Y_0 = X_0$, e tal que

$$Y_{n+1} = Y_n + a(t_n, Y_n)(t_{n+1} - t_n) + b(t_n, Y_n)(W_{t_{n+1}} - W_n),$$
(2.4)

para n = 0, 1, ..., N - 1, em que $Y_n = Y(t_n)$. O processo Y é o que chamamos de **aproximação de Euler**, ou seja, a aproximação de Euler é uma técnica básica de discretização de um processo de Itô.

2.2 Modelo discreto por modelo de difusão

A modelagem contínua é capaz de representar as premissas teóricas da precificação de ativos. No entanto, a coleta dos dados somente pode ser feita em intervalos discretos de tempo. Temos então de um lado um modelo teórico contínuo e de outro dados reais discretos. Uma possibilidade de unir a teoria à prática é procurar um processo limite para um processo discreto, ou seja, mostrar que a prática pode se aproximar da teoria. Nesta seção apresentaremos condições gerais sob as quais uma seqüência de processos de Markov discretos no tempo converge fracamente para um processo de Itô.

Primeiramente vamos construir um processo contínuo no tempo a partir de um processo de Markov discreto no tempo. Dado h>0 arbitrário, considere o processo de Markov discreto no tempo $X_0, X_h, X_{2h}, \ldots, X_{kh}$ denotado por $\{X_{kh}\}$, em que X_{kh} toma valores no \mathbb{R}^n para todo k. Assuma que sejam conhecidas as probabilidades de transição de $\{X_{kh}\}$ e a distribuição da variável inicial X_0 . O processo contínuo no tempo $\{X_t^{(h)}\}$ é construído através do processo discreto $\{X_{kh}\}$ fazendo $X_t^{(h)}$ uma função step com saltos nos tempos $h, 2h, \ldots$, e tal que $X_t^{(h)} = X_{kh}$ quase certamente para $kh \leq t < (k+1)h$. Sejam

$$\mu_h(x) \equiv h^{-1} \mathbb{E}[(X_{k+1}^{(h)} - X_k^{(h)}) | X_k^{(h)} = x]$$
(2.5)

e

$$\Sigma_h(x) \equiv h^{-1} \text{Cov}[(X_{k+1}^{(h)} - X_k^{(h)}) | X_k^{(h)} = x].$$
 (2.6)

Considere também que $F_h(X_0^{(h)})$ denota a função distribuição acumulada de $X_0^{(h)}$.

Teorema 2.10 (Stroock e Varadhan (1979)). Seja $\{X_t\}$ um processo governado pela seguinte equação integral estocástica

$$X_t = X_0 + \int_0^t \mu(X_s)ds + \int_0^t \Sigma^{1/2}(X_s)dW_s, \qquad (2.7)$$

em que W_t é um movimento browniano padrão, $\mu(\cdot)$ é uma função contínua do espaço das matrizes reais $N \times N$. Suponha que a integral em (2.7) tenha solução fraca única. Se

- 1. $F_h(\cdot) \longrightarrow F(\cdot)$ quando $h \downarrow 0$ para todo ponto de continuidade de $F(\cdot)$;
- 2. $\mu_h(\cdot) \to \mu(\cdot)$ uniformemente para todo conjunto limitado de x quando $h \downarrow 0$;

- 3. $\Sigma_h(\cdot) \to \Sigma(\cdot)$ uniformemente para todo conjunto limitado de x quando $h \downarrow 0$;
- 4. $\exists \ \delta > 0$ tal que $h^{-1}\mathbb{E}[\|X_{k+1}^{(h)} X_k^{(h)}\|^{2+\delta}|X_k^{(h)} = x] \to 0$ uniformemente para todo conjunto limitado de x quando $h \downarrow 0$. $(\|A\| = [trac(AA')]^{1/2})$.

Então $\{X_t^{(h)}\}$ converge fracamente para $\{X_t\}$ quando $h\downarrow 0$.

As hipóteses acima equivalem a dizer que

- 1. é necessário que as medidas de probabilidades F_h das variáveis iniciais $X_0^{(h)}$ convirjam para uma medida limite F quando $h\downarrow 0$;
- 2. a tendência do processo $\{X_t\}$, μ_h , e a matriz de difusão, Σ_h , devem convergir uniformemente em conjuntos compactos para funções contínuas e bem comportadas μ e Σ , respectivamente;
- 3. As diferenças do processo $X_k^{(h)}$ devem ter pelo menos um dos momentos absolutos de ordem maior que dois, por unidade de tempo, convergiudo para zero a uma taxa apropriada quando $h\downarrow 0$.

2.2.1 Modelo I

O modelo I é um processo de difusão descrito pelas equações (2.8)-(2.10) dadas a seguir.

$$dH_t = \left(\mu - \frac{e^{\nu_t}}{2}\right)dt + e^{\nu_t/2}dW_t \tag{2.8}$$

$$d\nu_{t+h} = \beta(\alpha - \nu_t)dt + \gamma d\widetilde{W}_t, \quad \text{com} \quad \nu_t = \ln \sigma_t^2, \tag{2.9}$$

em que W_t e \widetilde{W}_t são processos de Wiener com matriz de variâncias

$$\operatorname{Var}\left(\frac{W_t}{\widetilde{W}_t}\right) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{2.10}$$

Agora considere o processo a tempo discreto dado por

$$H_{t+h} = H_t + h\left(\mu - \frac{e^{\nu_t}}{2}\right) + h^{1/2}e^{\nu_t/2}Z_{t+h}$$
 (2.11)

$$\nu_{t+h} = \nu_t + h\beta(\alpha - \nu_t) + h^{1/2}\gamma \tilde{Z}_{t+h},$$
(2.12)

para $t=h,2h,3h,\ldots$, com H_0 e ν_0 valores iniciais fixados e conhecidos, e (Z_t,\widetilde{Z}_t) tem distribuição normal bivariada independente e identicamente distribuída (i.i.d.) com vetor de médias (0,0) e matriz de variância

$$\operatorname{Var}\left(\begin{array}{c} Z_t \\ \widetilde{Z}_t \end{array}\right) = \left(\begin{array}{c} 1 & \rho \\ \rho & 1 \end{array}\right). \tag{2.13}$$

Definamos os processos contínuos no tempo $H_{\tau}^{(h)}$ e $\nu_{\tau}^{(h)}$ por

$$H_{\tau}^{(h)} = H_t \quad \text{e} \quad \nu_{\tau}^{(h)} = \nu_t, \quad \text{para} \quad t \le \tau < t + h.$$
 (2.14)

Seja \mathcal{M}_t a σ -álgebra gerada por $\{H_\tau, \nu_\tau, 0 \leq \tau \leq t\}$. Desta forma

$$h^{-1} \mathbf{E} \left[\left(\begin{array}{c} H_{t+h} - H_t \\ \nu_{t+h} - \nu_t \end{array} \right) \middle| \mathcal{M}_t \right] = h^{-1} \mathbf{E} \left[\left(\begin{array}{c} h \left(\mu - \frac{e^{\nu_t}}{2} \right) + h^{1/2} e^{\nu_t/2} Z_{t+h} \\ h \beta (\alpha - \nu_t) + h^{1/2} \gamma \widetilde{Z}_{t+h} \end{array} \right) \middle| \mathcal{M}_t \right]$$

$$= h^{-1} \left[\left(\begin{array}{c} h \left(\mu - \frac{e^{\nu_t}}{2} \right) \\ h \beta (\alpha - \nu_t) \end{array} \right) \middle| \mathcal{M}_t \right] + \left(\begin{array}{c} \mu - \frac{e^{\nu_t}}{2} \\ \beta (\alpha - \nu_t) \end{array} \right) \right]$$

$$(2.15)$$

Além disso, a matriz de variância por unidade de tempo é obtida como a seguir.

$$h^{-1}\operatorname{Var}\left[\begin{pmatrix} H_{t+h} - H_t \\ \nu_{t\cdots h} - \nu_t \end{pmatrix} \middle| \mathcal{M}_t \right] = h^{-1}\operatorname{Var}\left[\begin{pmatrix} h^{1/2}e^{\nu_t/2}Z_{t+h} \\ h^{1/2}\gamma\widetilde{Z}_{t+h} \end{pmatrix} \middle| \mathcal{M}_t \right]$$

$$\operatorname{Var}\left[\begin{pmatrix} e^{\nu_t/2} & 0 \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} Z_t \\ \widetilde{Z}_t \end{pmatrix} \middle| \mathcal{M}_t \right] = \begin{pmatrix} e^{\nu_t/2} & 0 \\ 0 & \gamma \end{pmatrix} \operatorname{Var}\left(\frac{Z_t}{\widetilde{Z}_t}\right) \begin{pmatrix} e^{\nu_t/2} & 0 \\ 0 & \gamma \end{pmatrix}'$$

$$= \begin{pmatrix} e^{\nu_t} & \gamma \rho e^{\nu_t} \\ \gamma \rho e^{\nu_t} & \gamma^2 \end{pmatrix}.$$

$$(2.16)$$

Como H_0 e ν_0 são fixos e conhecidos então a hipótese 1 do teorema 2.10 é satisfeita. Os cálculos em (2.14) e (2.16) mostram que

$$\mu_h((H_t, \nu_t)) = \begin{pmatrix} \mu - \frac{e^{\nu_t}}{2} \\ \beta(\alpha - \nu_t) \end{pmatrix}$$
 (2.17)

0

$$\Sigma_h((H_t, \nu_t)) = \begin{pmatrix} e^{\nu_t} & \gamma \rho e^{\nu_t} \\ \gamma \rho e^{\nu_t} & \gamma^2 \end{pmatrix}. \tag{2.18}$$

E isto nos dá que $\mu_h((H_t, \nu_t))$ e $\Sigma_h((H_t, \nu_t))$ não dependem de h, portanto convergem uniformemente em qualquer conjunto limitado de (H_t, ν_t) quando $h \downarrow 0$. Logo são satisfeitas as condições 2 e 3 do teorema 2.10. É fácil, embora trabalhoso mostrar que

$$E[h^{-1}(H_{t+h} - H_t)^4 | \mathcal{M}_t] = h^3 \left(\mu - \frac{e^{\nu_t}}{2}\right)^4 + 4h^2 \left(\mu - \frac{e^{\nu_t}}{2}\right) e^{\nu_t} + 3he^{2\nu_t} + 2h^2 \left(\mu - \frac{e^{\nu_t}}{2}\right)^2 e^{\nu_t} \xrightarrow{h10} 0$$
(2.19)

 \mathbf{e}

$$E[h^{-1}(\nu_{t+h} - \nu_t)^4 | \mathcal{M}_t] = h^3 \beta^4 (\alpha - \nu_t) + h^2 \beta^2 \gamma^2 (\alpha - \nu_t)^2 + 4h^2 \beta^2 (\alpha - \nu_t)^2 + h^2 \beta^2 \gamma^2 (\alpha - \nu_t)^2 + 3h \gamma^4 \xrightarrow{h_{10}} 0.$$
(2.20)

A hipótese 4 é então satisfeita com $\delta=2$

Ainda resta provar que o processo $\{X_t\}$ tem solução fraca única. O trabalho Nelson (1991) resume algumas condições suficientes para a existência e unicidade de um processo de difusão limite. Tomemos como exemplo o caso enunciado a seguir.

Proposição 2.11. Consideremos a notação apresentada no teorema 2.10. Assegurando que $\mu_h(x)$ e $\Sigma_h(x)$ sejam funções contínuas tanto em x quanto em h e também garantindo que as derivadas parciais de primeira e segunda ordens com relação a x sejam contínuas, teremos definido unicamente um processo de difusão limite.

Pela equação (2.14) temos que μ_h é contínua tanto em H_t como em ν_t , as derivadas de primeira e segunda ordens são nulas e, portanto, contínuas. As derivadas parciais com relação a ν_t também são contínuas como podemos notar

$$\frac{\partial \mu_h}{\partial \nu_t}((H_t, \nu_t)) = \begin{pmatrix} --e^{\nu_t}/2 \\ -\beta \end{pmatrix}$$
 (2.21)

e

$$\frac{\partial^2 \mu_h}{\partial \nu_t^2} ((H_t, \nu_t)) = \begin{pmatrix} -e^{\nu_t}/2 \\ 0 \end{pmatrix} . \tag{2.22}$$

A continuidade da função Σ_h nas duas variáveis é facilmente notada pela equação (2.18). As derivadas parciais com relação a H_t e a ν_t são iguais a zero. Já as derivadas parciais com relação a ν_t são dadas por

$$\frac{\partial \Sigma_h}{\partial \nu_t}((H_t, \nu_t)) = \begin{pmatrix} e^{\nu_t} & \frac{\gamma \rho}{2} e^{\nu_t/2} \\ \frac{\gamma \rho}{2} e^{\nu_t/2} & 0 \end{pmatrix}$$
(2.23)

20

 \mathbf{e}

$$\frac{\partial^2 \Sigma_h}{\partial \nu_t^2}((H_t, \nu_t)) = \begin{pmatrix} e^{\nu_t} & \frac{\gamma \rho}{4} e^{\nu_t/2} \\ \frac{\gamma \rho}{4} e^{\nu_t/2} & 0 \end{pmatrix} , \qquad (2.24)$$

e portanto são contínuas. Isto implica a unicidade da solução.

Finalmente podemos concluir que (2.11)-(2.13), conhecida como aproximação de Euler do modelo I, converge fracamente para o processo de difusão definido por (2.8)-(2.10), que nada mais é do que o modelo I na sua forma matricial.

No entanto essa aproximação de Euler não é um processo ARCH. Para termos uma aproximação ARCH para o processo de difusão em questão, substituímos a equação (2.12) do sistema discreto por

$$\nu_{t+h} = \nu_t + h\beta(\alpha - \nu_t) + h^{1/2}g(Z_{t+h}), \qquad (2.25)$$

em que $g(\cdot)$ é uma função mensurável com $\mathbb{E}[|g(Z_t)|^{2+\delta}] < \infty$ para algum $\delta > 0$ e

$$\operatorname{Var}\left(\frac{Z_t}{g(Z_t)}\right) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{2.26}$$

Para que o sistema definido pelas equações (2.11), (2.25) e (2.26) corresponda a um processo EGARCH, como feito em Bollerslev, Engle & Nelson (1994) e em Nelson (1990a), consideramos a função $g(\cdot)$ como

$$g(Z_t) = \rho \gamma Z_t + \gamma \left(\frac{1 - \rho^2}{1 - 2/\pi}\right)^{1/2} \left[|Z_t| - \left(\frac{2}{\pi}\right)^{1/2} \right], \qquad (2.27)$$

já que $E(|Z_t|) = (2/\pi)^{1/2}$, $E(Z_t|Z_t|) = 0$ e $Var(|Z_t|) = 1 - 2/\pi$. Este modelo discreto é valorizado pois incorpora as propriedades dos modelos ARCH, como a fácil obtenção da verossimilhança e sua inferência. Neste trabalho enfocaremos no capítulo 4 uma outra forma de inferência, que não utiliza a aproximação ARCH, e sim a aproximação de Euler.

2.2.2 Modelo II

Chamaremos de modelo II o processo de difusão descrito a seguir.

$$dH_t = \left(\mu - \frac{\sigma_t^2}{2}\right)dt + \sigma_t dW_t \tag{2.28}$$

$$d\sigma_t^2 = \delta(\lambda - \sigma_t^2)dt + \xi \sigma_t d\widetilde{W}_t, \qquad (2.29)$$

em que W_t e \widetilde{W}_t são processos de Wiener com matriz de variâncias

$$\operatorname{Var}\left(\frac{W_t}{\widetilde{W}_t}\right) = \left(\frac{1-\rho}{\rho-1}\right). \tag{2.30}$$

Seja o processo a tempo discreto dado por

$$H_{t+h} = H_t + h\left(\mu - \frac{\sigma_t^2}{2}\right) + h^{1/2}\sigma_t Z_{t+h}$$
 (2.31)

$$\sigma_{t+h}^2 = \sigma_t^2 + h\delta(\lambda - \sigma_t^2) + h^{1/2}\xi\sigma_t\widetilde{Z}_{t+h}, \qquad (2.32)$$

para t = h, 2h, 3h, ..., com H_0 e ν_0 valores iniciais fixados e conhecidos, e (Z_t, \widetilde{Z}_t) tem distribuição normal bivariada i.i.d. com vetor de médias (0,0) e matriz de variância

$$\operatorname{Var}\left(\begin{array}{c} Z_t \\ \widetilde{Z}_t \end{array}\right) = \left(\begin{array}{c} 1 & \rho \\ \rho & 1 \end{array}\right) \,. \tag{2.33}$$

Sejam os processos contínuos no tempo $H_{\tau}^{(h)}$ e $\sigma_{\tau}^{2^{(h)}}$ definidos por

$$H_{\tau}^{(h)} = H_t \quad \text{e} \quad \sigma_{\tau}^{2(h)} = \sigma_t^2, \quad \text{para} \quad t \le \tau < t + h.$$
 (2.34)

Análogo à seção anterior, \mathcal{M}_t denota a σ -álgebra gerada por $\{H_\tau, \sigma_\tau^2, 0 \le \tau \le t\}$. Desta forms

$$h^{-1} \mathbf{E} \left[\begin{pmatrix} H_{t+h} - H_t \\ \sigma_{t+h}^2 - \sigma_t^2 \end{pmatrix} \middle| \mathcal{M}_t \right] = h^{-1} \mathbf{E} \left[\begin{pmatrix} h \left(\mu - \frac{\sigma_t^2}{2} \right) + h^{1/2} \sigma_t^2 Z_{t+h} \\ h \delta(\lambda - \sigma_t^2) + h^{1/2} \xi \sigma_t \widetilde{Z}_{t+h} \end{pmatrix} \middle| \mathcal{M}_t \right]$$

$$= h^{-1} \left[\begin{pmatrix} h \left(\mu - \frac{\sigma_t^2}{2} \right) \\ h \delta(\lambda - \sigma_t^2) \end{pmatrix} \middle| \mathcal{M}_t \right] = \begin{pmatrix} \mu - \frac{\sigma_t^2}{2} \\ \delta(\lambda - \sigma_t^2) \end{pmatrix}.$$

$$(2.35)$$

Vamos agora calcular a matriz de variância por unidade de tempo,

$$h^{-1}\operatorname{Var}\left[\begin{pmatrix} H_{t+h} - H_t \\ \sigma_{t+h}^2 - \sigma_t^2 \end{pmatrix} \middle| \mathcal{M}_t\right] = h^{-1}\operatorname{Var}\left[\begin{pmatrix} h^{1/2}\sigma_t Z_{t+h} \\ h^{1/2}\xi\sigma_t \widetilde{Z}_{t+h} \end{pmatrix} \middle| \mathcal{M}_t\right]$$

$$\operatorname{Var}\left[\begin{pmatrix} \sigma_t & 0 \\ 0 & \xi\sigma_t \end{pmatrix} \begin{pmatrix} Z_t \\ \widetilde{Z}_t \end{pmatrix} \middle| \mathcal{M}_t\right] = \begin{pmatrix} \sigma_t & 0 \\ 0 & \xi\sigma_t \end{pmatrix} \operatorname{Var}\begin{pmatrix} Z_t \\ \widetilde{Z}_t \end{pmatrix} \begin{pmatrix} \sigma_t & 0 \\ 0 & \xi\sigma_t \end{pmatrix}'$$

$$= \begin{pmatrix} \sigma_t^2 & \xi\rho\sigma_t^2 \\ \xi\rho\sigma_t^2 & \xi^2\sigma_t^2 \end{pmatrix}.$$
(2.36)

As equações (2.35) e (2.36) nos dão respectivamente as funções $\mu_h((H_t, \sigma_t^2))$ e $\Sigma_h((H_t, \nu_t))$. Novamente temos que a tendência e a matriz de difusão não dependem de h e então temos satisfeitas as condições 2 e 3 do teorema 2.10. O mesmo ocorre com a hipótese 1, pois consideramos as condições iniciais, H_0 e σ_0^2 conhecidas e fixas. Pela equação (2.19) concluímos diretamente que

$$E_{t}^{r}h^{-1}(H_{t+h} - H_{t})^{4}[\mathcal{M}_{t}] = h^{3}\left(\mu - \frac{\sigma_{t}^{2}}{2}\right)^{4} + 4h^{2}\left(\mu - \frac{\sigma_{t}^{2}}{2}\right)\sigma_{t}^{2} + 3h\sigma_{t}^{4} + 2h^{2}\left(\mu - \frac{\sigma_{t}^{2}}{2}\right)^{2}\sigma_{t}^{2} \xrightarrow{h10} 0,$$
(2.37)

e não é difícil mostrar que

$$E[h^{-1}(\sigma_{t+h}^2 - \sigma_t^2)^4 | \mathcal{M}_t] = h^3 \delta^4 (\lambda - \sigma_t^2) + 4h^2 \delta^2 \xi^2 (\lambda - \sigma_t^2)^2 + 3h \xi^4 \sigma_t^4 + 2h^2 \delta^2 \xi^2 (\lambda - \sigma_t^2)^2 \xrightarrow[h \downarrow 0]{} 0.$$
(2.38)

Logo, para $\delta = 2$, a hipótese 4 é satisfeita.

Como feito para o modelo I, analisaremos a continuidade das funções e das derivadas parciais de primeira e segunda ordem de μ_h e Σ_h . Pelas equações (2.35) e (2.36) são contínuas tanto em H_t como em σ_t^2 . Ambas as funções têm derivadas parciais de primeira e segunda ordens com relação a H_t nulas. As derivadas parciais com relação a σ_t^2 são dadas por

$$\frac{\partial \mu_h}{\partial \sigma_t^2}((H_t, \sigma_t^2)) = \begin{pmatrix} -1/2 \\ -\delta \end{pmatrix}, \qquad \frac{\partial^2 \mu_h}{\partial (\sigma_t^2)^2}((H_t, \sigma_t^2)) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad (2.39)$$

$$\frac{\partial \Sigma_h}{\partial \sigma_t^2}((H_t, \sigma_t^2)) = \begin{pmatrix} 1 & \xi \rho \\ \xi \rho & \xi^2 \end{pmatrix} \quad \text{e} \qquad \frac{\partial^2 \Sigma_h}{\partial (\sigma_t^2)^2}((H_t, \sigma_t^2)) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.40}$$

De acordo com a proposição 2.11 fica provada a unicidade da solução.

Concluínos que a aproximação de Fuler dadas pelas equações (2.31)-(2.33) converge fracamente para o processo descrito por (2.28)-(2.30), que é o modelo II descrito matricialmente.

2.3 Conhecendo os Modelos

O sistema descrito em (1.13) atende às hipóteses em que se basciam um modelo de difusão para precificação de ativos. Nos casos a serem tratados neste trabalho, consideramos

a função μ_S constante e igual a μ , e a função σ_S igual a volatilidade no tempo $t,\,\sigma_t$ obtendo a seguinte equação

$$dS_t = \mu S_t dt + \sigma_t S_t dW_t \,, \tag{2.41}$$

em que W_t é um movimento browniano. Isso significa que a taxa de retorno esperada continuamente ajustada é considerada proporcional ao valor passado do ativo, e que a variância condicional do retorno é dada pela volatilidade no tempo a que o retorno é condicionado, veja hipóteses (1.1) e (1.2).

Ao invés de trabalharmos com série des retornos, trataremos de séries de log-retornos, pois, como foi dito na seção 1.1.1, estas últimas apresentam propriedades vantajosas comparadas às séries de retornos. Para chegarmos ao processo descrito em (2.8) e também em (2.28) aplicamos o lema de Itô à equação (2.41) enunciado após a definição dada a seguir.

Definição 2.12 (Processo de Itô). Chamamos de processo de Itô um processo X_t definido no espaço de probabilidade (Ω, \mathcal{B}, P) que satisfaça a equação diferencial

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \qquad \forall \omega \in \Omega,$$
(2.42)

em que $\sqrt{|a|}$ e b perteucem ao espaço L^2 .

Lema 2.13 (Lema de Itô). Sejam X_t um processo de Itô e $f:[0,T]\times\mathbb{R}\to\mathbb{R}$ uma função cujas derivadas parciais $\frac{\partial f}{\partial t}, \frac{\partial f}{\partial x}, \frac{\partial^2 f}{\partial x^2}$ são contínuas. Considere o processo $Y_t = f(t,X_t)$ para $0 \le t \le T$. Então

$$df(t, X_t) = \left[\frac{\partial f}{\partial t}(t, X_t) + a(t, X_t)\frac{\partial f}{\partial x}(t, X_t) + \frac{1}{2}b^2(t, X_t)\frac{\partial^2 f}{\partial x^2}(t, X_t)\right]dt + b(t, X_t)\frac{\partial f}{\partial x}(t, X_t)dW_t \quad (2.43)$$

com probabilidade 1 para $0 \le t \le T$. (veja p.ex. Øksendal, 1998, p. 44).

O processo S_t descrito em (2.41) é um processo de Itô pois é representado na forma (2.42) com

$$a(t, S_t) = \mu S_t \quad \text{e} \quad b(t, S_t) = \sigma_t S_t. \tag{2.44}$$

Seja $H \equiv \ln S$. Tomando $f(t, x) = \ln x$ temos

$$\frac{\partial f}{\partial x}(t,x) = 0, \qquad \frac{\partial f}{\partial x}(t,x) = \frac{1}{x}, \qquad \frac{\partial^2 f}{\partial x^2}(t,x) = -\frac{1}{x^2}. \tag{2.45}$$

Podemos aplicar o lema de Itô ao processo $H=\ln S$ e obtemos

$$dH_t = \left[\mu S_t \frac{1}{S_t} + \frac{1}{2} (\sigma_t S_t)^2 \left(-\frac{1}{S_t} \right) \right] dt + \sigma_t S_t \frac{1}{S_t} dW_t.$$
 (2.46)

E finalmente

$$dH_t = \left(\mu - \frac{\sigma_t^2}{2}\right)dt + \sigma_t dW_t. \tag{2.47}$$

Usamos dois tipos de modelos, um que descreve a log-volatilidade e outro que descreve a volatilidade. Ambos atendem às hipóteses base e são descritos a seguir.

2.3.1 Modelo I

Vamos relembrar que o modelo I é descrito como a seguir

$$dH_t = \left(\mu - \frac{e^{\nu_t}}{2}\right) dt + e^{\nu_t/2} dW_t$$
$$d\nu_{t+h} = \beta(\alpha - \nu_t) dt + \gamma d\widetilde{W}_t,$$

em que W_t e \widetilde{W}_t são processos de Wiener com matriz de variâncias dada por (2.10). E sua aproximação de Euler é dada por

$$H_{t+h} = H_t + h\left(\mu - \frac{e^{\nu_t}}{2}\right) + h^{1/2}e^{\nu_t/2}Z_{t+h}$$
$$\nu_{t+h} = \nu_t + h\beta(\alpha + \nu_t) + h^{1/2}\gamma \tilde{Z}_{t+h},$$

para $t = h, 2h, 3h, \ldots$, em que (Z_t, \widetilde{Z}_t) tem distribuição normal bivariada i.i.d. com vetor de médias (0,0) e matriz de variâncias descrita em (2.13).

É conveniente usarmos uma transformação nos parâmetros para descrevermos algumas propriedades deste processo. Esta transformação é dada por

$$\sigma^* = \exp\{\alpha/2\} \quad \text{e} \quad \phi = 1 - \beta. \tag{2.48}$$

Denotemos por $\tilde{\nu}_t = \ln\left(\frac{\sigma_t^2}{\sigma^{*2}}\right)$. Considerando intervalos de tempo unitário, ou seja, h=1, temos que o modelo descrito por (2.11) e (2.12) equivale ao modelo a seguir

$$H_{t+1} = H_t + \left(\mu - \sigma^{*2} \frac{e^{\tilde{\nu}_t}}{2}\right) + \sigma^* e^{\tilde{\nu}_t/2} Z_{t+1}$$
 (2.49)

$$\widetilde{\nu}_{t+1} = \phi \widetilde{\nu}_t + \gamma \widetilde{Z}_{t+1}. \tag{2.50}$$

As equações (2.49) e (2.50) definem um MARVE e tratam de uma aproximação discreta no tempo do processo de Ornstein-Uhlenbeck, um processo de difusão contínuo no tempo usado na precificação de opção. Este MARVE é chamado de modelo de volatilidade estocástica na média (MVEM), pois a média do processo depende da componente estocástica do modelo, a volatilidade, enquanto a variância é descrita pelos passados de seus próprios valores. Vale ressaltar que o MVEM é para a classe MVE o que o ARCII-M, proposto por Engle, Lilien & Robins (1987), é para a classe ARCII.

O parâmetro de escala σ^* climina a necessidade da inclusão de um termo constante na equação da log-volatilidade, anteriormente dado por $\beta\alpha$. A persistência na volatilidade passada é descrita pelo parâmetro ϕ . Para termos um processo estacionário ϕ deve ser positivo e menor que um. E sendo assim, a variância não-condicional deste modelo é dada por

$$\sigma^{*2} \exp\left(0.5 \frac{\gamma^2}{1 - \phi^2}\right) . \tag{2.51}$$

Já a variação da log-volatifidade é capturada pelo parâmetro γ .

O trabalho Koopman & Uspensky (2002) compara um MARVE mais geral do que o descrito nesta subseção com um ARCII-M, e conclui que o primeiro pode ser considerado uma alternativa mais competitiva frente ao ARCII, não somente nas questões teóricas como também nas pesquisas empíricas.

2.3.2 Modelo II

Relembrando que o modelo II e sua aproximação de Euler são dados respectivamente por

$$dH_t = \left(\mu - \frac{\sigma_t^2}{2}\right)dt + \sigma_t dW_t$$
$$d\sigma_t^2 = \delta(\lambda - \sigma_t^2)dt + \xi \sigma_t d\widetilde{W}_t,$$

em que W_t e \widetilde{W}_t são processos de Wiener com matriz de variâncias dada em (2.30). E

$$H_{t+1} = H_t + \left(\mu - \frac{\sigma_t^2}{2}\right) + \sigma_t Z_{t+1}$$
$$\sigma_{t+1}^2 = \sigma_t^2 + \delta(\lambda - \sigma_t^2) + \xi \sigma_t \widetilde{Z}_{t+1},$$

em que os intervalos de tempo foram tomados unitários, e (Z_t, \widetilde{Z}_t) tem distribuição normal bivariada i.i.d. com vetor de médias (0,0) e matriz de variância descrita em (2.33).

O modelo II é um modelo de difusão raiz quadrática e como o modelo I também é um MVEM. Nele a persistência na volatilidade passada é descrita pelo parâmetro $\varphi \doteq 1-\delta$. Para termos um processo estacionário φ deve ser positivo e menor que um.

A fundamental diferença entre esse modelo e o modelo I é que nesse a variação na volatilidade é heterocedástica, igual a $\xi^2 \sigma_t^2$.

2.4 Conclusão

Os teoremas que provam a convergência da aproximação de Euler têm como hipótese as condições globais de Lipschitz, (veja Nelson, 1990a). Essas condições não são satisfeitas pelos modelos aqui tratados. No entanto, apresentamos hipóteses que garantem a convergência fraca de certos processos, e mostramos que estas são satisfeitas pelas aproximações de Euler dos modelos 1 e II. Portanto ao tratarmos dos modelos discretizados, não estamos, por completo, abandonando as premissas estabelecidas pela modelagem contínua. Além disso, algumas características de ambos os modelos foram dadas.

Capítulo

3

Modelo Linear

A seguir apresentaremos considerações preliminares e depois daremos a idéia básica do filtro de Kalman (FK). Além disso deduziremos a forma univariada da filtragem, ou seja, as equações recursivas para um modelo univariado. Para maiores detalhes e para o caso multivariado recomendamos Harvey (1989) ou Davis & Vinter (1985).

3.1 Filtro de Kalman

Seja x_t um processo não-observado descrito por combinações de seus valores passados acrescido de uma componente aleatória. E seja y_t um processo observado descrito por x_t mais um ruído. Um problema de filtragem tem como objetivo estimar x_t através de y_t .

O FK é um algoritmo recursivo que objetiva calcular o estimador ótimo do estado x_t no tempo t, baseado nas informações obtidas até o tempo t. Essas informações consistem no

conjunto de observações $y_t = (y_0, y_1, \dots, y_t)$. O modelo é descrito por

$$y_t = a_t x_t + b_t + \varepsilon_t \qquad \text{(observação)} \tag{3.1}$$

$$x_t - c_t x_{t-1} + d_t + r_t \eta_t \qquad (estado), \qquad (3.2)$$

em que os coeficientes a_t , c_t e r_t e os termos b_t e d_t , para $t=1,2,\ldots,T$, podem ser variantes no tempo como indica a notação. Além disso, ε_t e η_t são não-correlacionados, com médias zero e variâncias iguais a σ_{ε}^2 e σ_{η}^2 , respectivamente, ou seja,

$$E(\varepsilon_t \eta_s) = 0 \quad \forall \ s, t = 0, 1 \dots, T \tag{3.3}$$

e

$$\operatorname{Var}(\varepsilon_t) = \sigma_{\varepsilon}^2 \quad \text{e} \quad \operatorname{Var}(\eta_t) = \sigma_{\eta}^2.$$
 (3.4)

A equação (3.2) é chamada de equação de transição. Para termos a descrição completa do modelo, devernos conhecer as condições iniciais do sistema e assumir que as perturbações ε_{t} e η_{t} sejam não-correlacionadas com o valor inicial. Podemos resumir estas últimas considerações da seguinte forma

$$E(x_0) = \overline{x}_0 \quad \text{e} \quad \text{Var}(x_0) = P_0, \quad \overline{x}_0 \in P_0 \text{ conhecidos},$$
 (3.5)

e

$$E(\eta_t x_0) = 0$$
 e $E(\varepsilon_t x_0) = 0$ $\forall t = 0, 1, ..., T$. (3.6)

3.1.1 Obtendo o Filtro de Kalman

O FK provem de uma propriedade da distribuição normal, tal propriedade permite calcular de forma recursiva a distribuição de x_t condicionada ao conjunto de informações obtidas até o tempo t, V $1 \le t \le T$. Essas distribuições condicionais são gaussianas e portanto são especificadas por sua média e variância. São estas duas informações que o FK obtém como veremos a seguir.

O estado para t = 1 fica descrito como

$$x_1 = c_1 x_0 + d_1 + r_1 \eta_1. (3.7)$$

Logo x_1 é uma combinação linear de duas variáveis normais independentes. Então é também uma variável normal com média condicional

$$\overline{x}_{1|0} = c_1 \overline{x}_0 + d_1 \tag{3.8}$$

e variância condicional

$$P_{1|0} = c_1^2 P_0 + r_1^2 \sigma_n^2. (3.9)$$

A notação $\vec{x}_{1|0}$ e $P_{1|0}$ se refere a média e a variância da distribuição de x_t condicionada às informações em t=0.

Queremos a distribuição de $x_1|y_1$. Para isso escrevemos

$$x_1 = \bar{x}_{1|0} + (x_1 - \bar{x}_{1|0}) \tag{3.10}$$

$$y_1 = a_1 \tilde{x}_{1|0} + a_1 (x_1 - \bar{x}_{1|0}) + b_1 + \varepsilon_1. \tag{3.11}$$

E então temos que o vetor $(x_1 - y_1)$ tem distribuição conjunta normal bivariada com vetor de médias $(\overline{x}_{1|0} - a_1\overline{x}_{1|0} + b_1)$, e matriz de variância

$$\left(\begin{array}{cc}
P_{1|0} & a_1 P_{1|0} \\
a_1 P_{1|0} & a_1^2 P_{1|0} + \sigma_{\varepsilon}^2
\end{array}\right).$$
(3.12)

Lema 3.1. Se (x y) tem distribuição normal conjunta bivariada com vetor de médias $\mu = (\mu_x - \mu_y)$ e matriz de variância dada por

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix} . \tag{3.13}$$

Então a distribuição de x condicionada a y é também normal com média

$$\mu_{x|y} = \mu_x + \sigma_{xy}^2 (\sigma_y^2)^{-1} (y - \mu_y) \tag{3.14}$$

e variància

$$\sigma_{x|y}^2 = \sigma_x^2 - (\sigma_{xy}^2)^2 (\sigma_y^2)^{-1}. \tag{3.15}$$

Prova. Por hipótese temos que

$$p(x,y) = \frac{1}{2\pi |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (|\mu_x - \mu_y|) \Sigma^{-1} (|\mu_x - \mu_y|)' \right]. \tag{3.16}$$

Vamos calcular a distribuição de x condicionada a y por

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{\frac{1}{2\pi} \frac{1}{|\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (\mu_x - \mu_y) \Sigma^{-1} (\mu_x - \mu_y)'\right]}{\frac{1}{(2\pi)^{1/2} \sigma_y} \exp\left[-\frac{1}{2\sigma_y^2} (y - \mu_y)^2\right]}.$$
 (3.17)

É conveniente utilizar a seguinte fórmula

$$\begin{pmatrix}
1 & -\sigma_{xy}^2(\sigma_y^2)^{-1} \\
0 & 1
\end{pmatrix} \Sigma \begin{pmatrix}
1 & 0 \\
-\sigma_{xy}^2(\sigma_y^2)^{-1} & 1
\end{pmatrix} = \begin{pmatrix}
\sigma_x^2 - (\sigma_{xy}^2)^2(\sigma_y^2)^{-1} & 0 \\
0 & \sigma_y^2
\end{pmatrix}$$
(3.18)

pois facilita concluir que

$$\Sigma^{-1} = \begin{pmatrix} 1 & 0 \\ -\sigma_{xy}^2(\sigma_y^2)^{-1} & 1 \end{pmatrix} \begin{pmatrix} \sigma_x^2 - (\sigma_{xy}^2)^2(\sigma_y^2)^{-1} & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \begin{pmatrix} 1 & -\sigma_{xy}^2(\sigma_y^2)^{-1} \\ 0 & 1 \end{pmatrix}. \tag{3.19}$$

Substituindo esta última informação em

$$(\mu_x - \mu_y)\Sigma^{-1}(\mu_x - \mu_y)'$$
 (3.20)

e denotando por $A = \sigma_x^2 - (\sigma_{xy}^2)^2 (\sigma_y^2)^{-1}$ chegamos a

$$(x - \mu_x)^2 A - 2(x - \mu_x)(y - \mu_y)\sigma_{xy}^2(\sigma_y^{-1})A + (y - \mu_y)\sigma_{xy}^2(\sigma_y^{-1})^2 A + (y - \mu_y)^2(\sigma_y^{-1})$$

$$= A[x - (\mu_x + (y - \mu_y)\sigma_{xy}^2(\sigma_y^{-1}))]^2 + (y - \mu_y)^2(\sigma_y^{-1}).$$

Finalmente

$$p(x|y) = \frac{1}{\sqrt{2\pi\sigma_{x|y}^2}} \exp\left[-\frac{1}{2\sigma_{x|y}^2} (x - \mu_{x|y})^2\right]$$
(3.21)

em que $\mu_{x|y}$ e $\sigma_{x|y}^2$ são como nas equações (3.14) e (3.15), respectivamente.

Pelo lema 3.1 concluímos então que

$$x_1|y_1 \sim \mathcal{N}(\overline{x}_1, P_1) \tag{3.22}$$

em que

$$\overline{x}_1 = \overline{x}_{1|0} + a_1 P_{1|0} f_1^{-1} (y_1 - a_1 \overline{x}_{1|0} - b_1). \tag{3.23}$$

$$P_1 = P_{1|0} - a_1^2 P_{1|0}^2 f_1^{-1} (3.24)$$

e

$$f_1 := a_1^2 P_{1|0} + \sigma_c^2 \,. \tag{3.25}$$

Seguindo o mesmo raciocínio para $t=2,3\ldots$, obtemos que

$$\overline{x}_{t|t-1} = c_t \overline{x}_{t-1} + d_t \tag{3.26}$$

 \mathbf{G}

$$P_{t|t-1} = c_t^2 P_{t-1} + r_t^2 \sigma_n^2 \tag{3.27}$$

e concluímos que $x_t|y_t$ tem distribuição normal com média e variância dadas respectivamente por

$$\overline{x}_t = \overline{x}_{t|t-1} + a_t^2 P_{t|t-1} f_t^{-1} (y_t - a_t \overline{x}_{t|t-1} - b_t)$$
(3.28)

e

$$P_t = P_{t|t-1} - a_t^2 P_{t|t-1}^2 f_t^{-1}, (3.29)$$

em que

$$f_t = a_t^2 P_{t|t-1} + \sigma_\varepsilon^2 \,. \tag{3.30}$$

As equações (3.28) e (3.29) são chamadas de equações de atualização. E as equações (3.26) e (3.27) são as equações de predição do FK.

Além disso, foi mostrado que a média da distribuição condicional de x_t é um estimador ótimo para x_t , no sentido que ele minimiza o quadrado do erro médio, (veja p.ex. Harvey, 1989).

3.2 Modelo Invariante no Tempo

Consideremos o modelo discreto dado por (2.11)-(2.13) com h=1 e $\rho=0$. Aproximemos este modelo por um modelo linear utilizando a expansão em série de Taylor de primeira ordem em torno da origem. A equação (2.11) fica aproximada por

$$H_{t+1} = H_t + \mu - \frac{1}{2} + \left(\frac{1}{2} + \frac{1}{2}Z_{t+1}\right)\nu_t + \frac{1}{2}Z_{t+1}. \tag{3.31}$$

Este processo é condicionalmente gaussiano. Para simplificá-lo, consideramos que o termo $1/2 - Z_{t+1}/2$ é constante igual a a, obtendo que

$$H_{t+1} = H_t + \mu - \frac{1}{2} + a\nu_t + \varepsilon_t,$$
 (3.32)

em que $\varepsilon_t \sim \mathcal{N}(0, 1/4)$.

Sejam

$$b = \mu - \frac{1}{2}, \quad c = 1 - \beta, \quad d = \beta \alpha \quad e \quad r = \gamma^2.$$
 (3.33)

Além disso, tomando

$$y_{t+1} = H_{t+1} - H_t$$
, $\sigma_{\eta}^2 = 1$ e $\sigma_{\varepsilon}^2 = 1/4$, (3.34)

para $t=1,2,\ldots,$ ficamos com um modelo linear invariante no tempo da forma

$$y_t = a\nu_t + b + \varepsilon_t$$
 (observação) (3.35)

$$\nu_t = c\nu_{t+1} + d + r\eta_t \tag{estado}.$$

E também, ε_t e η_t são independentes para todo t, e independentes do valor inicial ν_0 . Assumimos que

$$\nu_0 \sim \mathcal{N}(\overline{\nu}_0, P_0), \quad \overline{\nu}_0 \in P_0 \quad \text{conhecidos.}$$
(3.37)

Como feito na obtenção do FK, o estado para t=1 fica descrito como

$$\nu_1 = c\nu_0 + d + r\eta_1 \,. \tag{3.38}$$

Diretamente concluímos que

$$\nu_{\rm t}|\nu_0 \sim \mathcal{N}(\overline{\nu}_{1|0}, P_{\rm P0}) \tag{3.39}$$

em que

$$\overline{\nu}_{1|0} = c\overline{\nu}_0 + d \tag{3.40}$$

 \mathbf{e}

$$P_{1|0} = c^2 P_0 + r^2 \sigma_\eta^2. (3.41)$$

Como na seção anterior, $\nu_{1'0}$ se refere a média da distribuição de ν_1 condicionada às informações até t=0. Escrevendo

$$\nu_1 = \overline{\nu}_{1|0} + (\nu_1 - \overline{\nu}_{1|0}) \tag{3.42}$$

$$y_1 = a\overline{\nu}_{1|0} + a(\nu_1 - \overline{\nu}_{1|0}) + b + \varepsilon_1$$
 (3.43)

concluímos que o vetor (ν_1-y_1) tem distribuição conjunta normal bivariada, com vetor de médias $(\overline{\nu}_{1|0}-a\overline{\nu}_{1|0}+b)$ e matriz de variâncias

$$\begin{pmatrix} P_{1|0} & aP_{1|0} \\ aP_{1|0} & a^2P_{1|0} + \sigma_{\varepsilon}^2 \end{pmatrix} . \tag{3.44}$$

Do lema 3.1 segue que

$$\nu_1|y_1 \sim \mathcal{N}(\overline{\nu}_1, P_1) \tag{3.45}$$

em que

$$\overline{\nu}_1 = \overline{\nu}_{1|0} + aP_{1|0}f_1^{-1}(y_1 - a\overline{\nu}_{1|0} - b),$$
(3.46)

$$P_{\rm I} = P_{\rm I|0} - a^2 P_{\rm I|0}^2 f_{\rm I}^{-1} \tag{3.47}$$

e

$$f_1 = a^2 P_{I|0} + \sigma_e^2 \tag{3.48}$$

Analogamente para $t=1,2,\ldots,T$, obtemos que $\nu_t|y_t$ tem distribuição normal com média e variância dadas respectivamente por

$$\nu_t = \overline{\nu}_{t|t-1} + aP_{t|t-1}f_t^{-1}(y_t - a\nu_{t|t-1} + b), \qquad (3.49)$$

$$P_t = P_{t|t-1} - a^2 P_{t|t-1}^2 f_t^{-1} \tag{3.50}$$

 \mathbf{e}

$$f_t = a^2 P_{t|t-1} + \sigma_{\varepsilon}^2 \tag{3.51}$$

em que $\overline{\nu}_{t+1|t}$ e $P_{t+1|t}$ são obtidos como nas equações (3.8) e (3.9), ou seja,

$$\overline{\nu}_{t;t-1} = \epsilon \overline{\nu}_{t-1} + d \tag{3.52}$$

e

$$P_{tjt-1} = c^2 P_t + r^2 \sigma_\eta^2. (3.53)$$

3.2.1 Verossimilhança

Agora que encontramos a distribuição de $\nu_t|y_t$, para $t=1,2\ldots,n$, podemos calcular a verossimilhança do vetor $y|\theta$, sendo $y=(y_0,y_1,\ldots,y_n)$ o vetor de observações e $\theta=(a,b,c,d,\gamma^2)$ o vetor de parâmetros. Levando em consideração a estrutura temporal das observações, podemos fatorar a verossimilhança em um produtório tomando sucessivas condicionais

$$L(\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta}) = p(y_0|\boldsymbol{\theta}) \prod_{t=1}^{n} p(y_t|\boldsymbol{y}_{t+1}, \boldsymbol{\theta}), \qquad (3.54)$$

em que $y_{t+1}=(y_0,y_1,\ldots,y_{t+1}),$ ou seja, denota o conjunto de informações obtidas até o instante t+1.

Pela equação (3.35) temos que

$$y_0 = a\nu_0 + b + \varepsilon_0 \,, \tag{3.55}$$

logo y_0 é combinação linear de uma variável normal, ν_0 , mais uma constante, b, adicionado um ruído normal, ε_0 . Portanto y_0 é normalmente distribuído com média e variância dados respectivamente por

$$\bar{y_0} = a\bar{\nu}_0 + b \tag{3.56}$$

 \mathbf{c}

$$\sigma_0^2 = a^2 P_0 + \sigma_z^2 \tag{3.57}$$

Para $t=1,2,\ldots,$ escrevemos a equação (3.35) na forma

$$y_t = a\overline{x}_{t|t-1} + a(x_t - \overline{x}_{t|t-1}) + b + \varepsilon_t.$$
(3.58)

Isto nos permite concluir diretamente que

$$y_t | \boldsymbol{y}_{t-1} \sim \mathcal{N}(y_t, \sigma_t^2) \tag{3.59}$$

com

$$\overline{y}_t = a\overline{\nu}_{t|t-1} + b \tag{3.60}$$

e

$$\sigma_t^2 \sim f_t. \tag{3.61}$$

Substituindo as informações acima na equação (3.54) obtemos

$$L(\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta}) \propto \prod_{t=0}^{n} \left(\frac{1}{\sigma_t^2}\right)^{1/2} \exp\left\{-\frac{(y_t - \overline{y}_t)^2}{2\sigma_t^2}\right\}.$$
(3.62)

Devemos maximizar a equação (3.62) ou, equivalentemente, maximizar seu logaritmo, a log-verossimilhança

$$l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) \propto -\frac{1}{2} \sum_{t=0}^{n} \ln(\sigma_t^2) - \frac{1}{2} \sum_{t=0}^{n} \frac{(y_t - \bar{y}_t)^2}{\sigma_t^2}$$
(3.63)

que é uma função não-linear em θ . Uma solução possível é utilizar o algoritmo de Newton-Raphson successivas vezes, até que o máximo seja encontrado. O esquema do processo de estimação é dado a seguir.

- 1. inicia com valores $\theta^{(0)}$ e contador de iteração j=0;
- 2. utiliza o FK e o vetor inicial; $\theta^{(j)}$, para obter o conjunto de esperanças e variâncias, $\{\overline{y}_t^{(j)}, \sigma_t^{2(j)}\};$
- 3. executa uma iteração do algoritmo de Newton-Raphson utilizando os valores obtidos no passo 2 e obtendo um novo conjunto de estimativas, $\theta^{(j+1)}$;
- 4. faz j=j+1e volta ao passo 2 até que a estimativa ou a log-verossimilhança estabilize, ou seja, até que

$$\|\boldsymbol{\theta}^{(j+1)} - \boldsymbol{\theta}^{(j)}\| < \xi \quad \text{ou} \quad |l(\boldsymbol{\theta}^{(j+1)}) - l(\boldsymbol{\theta}^{(j)})| < \xi,$$
 (3.64)

para $\xi > 0$ dado.

A distribuição assintótica do estimador de máxima verossimilhança é dada pelo resultado enunciado a seguir.

Teorema 3.2. Sob condições gerais, seja $\widehat{\theta}_N$ o estimador de máxima verossimilhança de θ , obtido maximizando a expressão (3.63). Então, quando $N \to \infty$,

$$\sqrt{N}(\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I(\boldsymbol{\theta})^{-1}),$$
 (3.65)

em que $I(\boldsymbol{\theta})$ é a matriz de informação assintótica dada por

$$I(\boldsymbol{\theta}) = \lim_{N \to \infty} N^{-1} E \left[-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]. \tag{3.66}$$

Prova. (Veja p.ex. Caines, 1988).

No passo 3 do esquema, apenas uma iteração do processo de Newton-Raphson é utilizada pois, ao menos teoricamente, a cada iteração, a estimativa se aproxima do valor a ser estimado. Tendo um vetor de parâmetros mais próximo do exato, as informações fornecidas pelo FK devem ser atualizadas. Feita a atualização, novamente buscamos a proximidade do valor exato.

Esse tipo de solução requer o uso de árdua programação computacional, pois os termos a serem derivados no algoritmo de Newton-Raphson são recursivos. Muitos softwares, usados

primeiramente na área de engenharia, propõem-se a solucionar numericamente o problema da diferenciação.

Uma mancira aproximada de encontrar o conjunto de parâmetros que maximiza a função de verossimilhança, evitando o cálculo das derivadas, é através da amostragem por importância. Trata-se do mesmo método usado na abordagem bayesiana relatada na próxima subseção. A idéia é construir uma amostra de estimativas dos parâmetros que represente candidatos a maximizar a verossimilhança. As amostras são geradas a partir de uma distribuição que represente as características de cada parâmetro. Por exemplo, se um deles está restrito a ser positivo e menor que um, devemos usar uma distribuição beta, ou então, utilizar uma transformação cujo domínio permita a amostragem por uma distribuição normal. Então é usado um critério de seleção e, a partir desses candidatos pré-selecionados, estimamos os parâmetros.

3.2.2 Abordagem Bayesiana

A abordagem bayesiana se adequa a situações em que, por exemplo, modelos complexos são necessários. Geralmente nesses casos, a verossimilhança é intratável analiticamente. Outro exemplo, é a possibilidade de inserir conhecimentos prévios a respeito do comportamento dos parâmetros do modelo. Em ambas situações os métodos analíticos de aproximação não são convenientes, sendo necessários métodos de aproximação numérica. Duas abordagens bastante difundidas nesse contexto são o método de Laplace, Tierney, Kass & Kadane (1989), e a integração de Monte Carlo via amostragem por importância, (veja p. ex. Ripley, 1987). Nesse trabalho utilizaremos a segunda técnica, mais especificamente, os métodos de Monte Carlo via cadeias de Markov (MCMC). É um método de simples implementação e está descrito com mais detalhes no apêndice A.I.

Vamos relembrar que nosso objetivo é estimar o vetor de parâmetros $\theta = (a, b, c, d, \gamma^2)$ dadas as observações até o tempo t, ou seja, $\mathbf{y} = (y_0, y_1, \dots, y_n)$. Notemos que a função densidade de probabilidade (fdp) conjunta pode ser fatorada no produto da fdp dos parâmetros, $p(\theta)$, chamada de priori na inferência bayesiana, e na fdp condicional das observações dados os parâmetros, $p(\mathbf{y}|\theta)$, chamada de verossimilhança. Em resumo

$$p(\boldsymbol{\theta}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{3.67}$$

Depois de coletar os dados, atualizamos o conhecimento a priori pelo teorema de Bayes, o que resulta na posteriori, $p(\theta|y)$,

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})},$$
(3.68)

em que $p(y) = \int p(y|\theta)p(\theta)d\theta$ é a fdp marginal de y.

Como alguns parâmetros têm algunas restrições de domínio, é conveniente utilizar uma transformação destes no processo de estimação. O parâmetro c está restrito a assumir valores entre 0 e 1, já γ^{2*} a assumir valores positivos, por isso trabalharemos com

$$c^* = \log \frac{c}{1 - c}$$
, on seja, $c = \frac{\exp(c^*)}{1 + \exp(c^*)}$ (3.69)

e com

$$\gamma^{2^*} = \exp(\gamma^2), \quad \text{ou seja} \quad \gamma^2 = \log(\gamma^{2^*}). \tag{3.70}$$

Consideremos que os parâmetros são independentes e que tenham fdp a priori normais tais que

$$a \sim \mathcal{N}(\mu_a, \sigma_a^2), \quad b \sim \mathcal{N}(\mu_b, \sigma_b^2),$$
 (3.71)

$$c^* \sim \mathcal{N}(\mu_{c^*}, \sigma_{c^*}^2), \quad d \sim \mathcal{N}(\mu_d, \sigma_d^2) \quad \text{e} \quad \gamma^{2^*} \sim \mathcal{N}(\mu_{\gamma^{2^*}}, \sigma_{\gamma^{2^*}}^2).$$
 (3.72)

Então

$$p(\boldsymbol{\theta}) = p(a)p(b)p(c^*)p(d)p(\gamma^{2^*})$$

$$= \prod_{r=a,b,c^*,d,\gamma^{2^*}} \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left\{-\frac{1}{2\sigma_r^2}(r-\mu_r)^2\right\}$$

$$= \left(\frac{1}{2\pi}\right)^{5/2} \frac{1}{\sigma_a\sigma_b\sigma_{c^*}\sigma_d\sigma_{\gamma^{2^*}}} \exp\left\{-\sum_r \frac{1}{2\sigma_r^2}(r-\mu_r)^2\right\}.$$
(3.73)

Multiplicando as equações (3.62) e (3.73) obtemos que a posteriori é proporcional a

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \prod_{t=0}^{n} \left(\frac{1}{\sigma_t^2}\right)^{1/2} \exp\left\{-\frac{(y_t - \widetilde{y}_t)^2}{2\sigma_t^2}\right\} \times \frac{1}{\sigma_a \sigma_b \sigma_{e^*} \sigma_d \sigma_{\gamma^{2^*}}} \exp\left\{-\sum_{r=a,b,e^*,d,\gamma^{2^*}} \frac{1}{2\sigma_r^2} (r - \mu_r)^2\right\},$$
(3.74)

a mesma expressão pode ser escrita como

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto \exp\left\{-\sum_{t=0}^{n} \frac{(y_t - \overline{y}_t)^2}{2\sigma_t^2} - \sum_{r=a,b,c^*,d,\gamma^{2^*}} \frac{(r - \mu_r)^2}{2\sigma_r^2}\right\} \prod_{t=0}^{n} \left(\frac{1}{\sigma_t^2}\right)^{1/2}$$
(3.75)

já que \bar{y}_t e $\sigma_t^2,\,t=0,1,\ldots,n,$ dependem do vetor de parâmetros $oldsymbol{ heta}.$

Denotemos por θ_{-r} o vetor θ sem o parâmetro r, para $r=a,b,c^*,d,\gamma^{2^*}$. As densidades condicionais a posteriori são proporcionais a

$$p(r|\mathbf{y}, \boldsymbol{\theta}_{-r}) \propto \exp\left\{-\sum_{t=0}^{n} \frac{(y_t - y_t)^2}{2\sigma_t^2} - \frac{(r - \mu_r)^2}{2\sigma_r^2}\right\} \prod_{t=0}^{n} \left(\frac{1}{\sigma_t^2}\right)^{1/2}, \quad r = a, b, e^*, d, \gamma^{2^*}.$$
(3.76)

Portanto não são densidades fechadas, ou seja, não são densidades conhecidas como a Normal, a Gama, a Beta, entre outras. Isto justifica o uso dos métodos MCMC com o algoritmo de Metropolis-Hastings desenvolvido por Metropolis, Rosenbluth, Teller & Teller (1953) e generalizado por Hastings (1970). Por estarmos considerando independência entre os parâmetros, vamos gerar candidatos para cada um deles separadamente, no lugar de usar uma distribuição multivariada. O esquema do processo usado segue.

- 1. inicia com valor $\boldsymbol{\theta}^{(0)}$ e contador de iteração j=0;
- 2. toma r = a e gera um candidato r^* de acordo com o núcleo de transição $q(r^{(j)}, r^*)$;
- 3. atualiza $r^{(j)}$ por $r^{(j+1)} = r^*$ com probabilidade

$$p = \min\left\{1, \frac{p(r^*)q(r^{(j)}, r^*)}{p(r^{(j)})q(r^*, r^{(j)})}\right\},\tag{3.77}$$

ou seja, permanece com $r^{(j)}$ com probabilidade $1-p_{-j}$;

- 4. toma r = b e repete os passos 2 e 3;
- 5. toma $r = c^*$ e repete os passos 2 e 3;
- 6. toma r = d e repete os passos 2 e 3;
- 7. toma $r = \gamma^{2*}$ e repete os passos 2 e 3;
- 8. faz j = j + 1 e repete os passos de 2 a 6 até conseguir uma distribuição estacionária.

Esse processo nos fornece um conjunto de estimativas usadas na inferência dos parâmetros. Algumas aplicações podem ser vistas no capítulo 5.

3.3. CONCLUSÃO 39

3.3 Conclusão

Neste capítulo apresentamos a construção do FK, assim como a função de verossimilhança de um processo de estado. Tratamos de um modelo invariante no tempo, apresentando sua verossimilhança. Devido a sua não-linearidade e a presença de termos recursivos, a obtenção de seu máximo é intratável analiticamente. Sugerimos uma técnica numérica, bastante usada na inferência bayesiana, para estimá-lo. A abordagem bayesiana foi apontada como uma maneira de suprir a dificuldade da abordagem clássica.

Capítulo

4

Inferência para o Modelo I

No capítulo anterior vimos que casos lineares podem ser solucionados com o uso do FK. No entanto, a aproximação discreta do modelo I é não-linear e apresenta dependência do estado na variância. Para a inferência deste tipo de modelo sugerimos o uso do filtro estendido de Kalman (FEK) descrito na primeira seção deste capítulo.

4.1 Filtro Estendido de Kalman

Considere o seguinte modelo não-linear em estado de espaço

$$y_t = A_t(x_t) + \varepsilon_t$$
 (observação) (4.1)

$$x_t = C_t(x_{t-1}) + R_t(x_{t-1})\eta_t$$
 (estado) (4.2)

em que $A_t(x_t)$ e $C_t(x_{t-1})$ não são necessariamente funções lineares dos elementos do estado, como são na equação (3.1), e $R_t(x_{t-1})$ pode depender do vetor de estado enquanto r_t em (3.2) não pode. Aqui também consideraremos a hipórese de que ε_t e η_t são não-correlacionados, com médias zero e variâncias ignais a $\sigma_{\varepsilon_t}^2$ e $\sigma_{\eta_t}^2$, respectivamente.

Uma das formas de obter um filtro ótimo, que na verdade é um filtro aproximado, para o caso não linear é considerar uma aproximação linear do modelo, e então fazer algumas modificações nas equações do FK original. Para isso devemos ter funções $A_t(x_t)$, $C_t(x_{t-1})$ e $R_t(x_{t+1})$ suficientemente suaves, tais que sejam possíveis de serem expandidas em séries de Taylor em torno de suas médias condicionais e médias, $\tilde{x}_{t|t-1}$ e \tilde{x}_{t-1}

$$A_t(x_t) \simeq A_t(\overline{x}_{t|t-1}) + \frac{\partial A_t}{\partial x_t}(\overline{x}_{t|t-1})(x_t - \overline{x}_{t|t-1}), \qquad (4.3)$$

$$C_t(x_{t-1}) \simeq C_t(\overline{x}_{t-1}) + \frac{\partial C_t}{\partial x_t}(\overline{x}_{t-1})(x_{t-1} - \overline{x}_{t-1})$$

$$\tag{4.4}$$

e

$$R_t(x_{t-1}) \simeq R_t(\overline{x}_{t-1}). \tag{4.5}$$

Desta forma, aproximamos o modelo original não-linear por

$$y_t \simeq \frac{\partial A_t}{\partial x_t} (\overline{x}_{t|t-1}) x_t + A_t (\overline{x}_{t|t-1}) - \frac{\partial A_t}{\partial x_t} (\widetilde{x}_{t|t-1}) \overline{x}_{t|t-1} + \varepsilon_t$$

$$(4.6)$$

$$x_t \simeq \frac{\partial C_t}{\partial x_t} (\overline{x}_{t-1}) x_{t-1} + C_t (\overline{x}_{t-1}) - \frac{\partial C_t}{\partial x_t} (\overline{x}_{t-1}) \overline{x}_{t-1} + R_t (\overline{x}_{t-1}) \eta_t, \qquad (4.7)$$

que ainda pode ser reescrito considerando

$$a_t = \frac{\partial A_t}{\partial x_t} (\dot{x}_{t|t-1}), \tag{4.8}$$

$$c_t = \frac{\partial C_t}{\partial x_t} (\overline{x}_{t-1}), \qquad (4.9)$$

$$b_t = A_t(\overline{x}_{t|t-1}) - \frac{\partial A_t}{\partial x_t}(\overline{x}_{t|t-1})\overline{x}_{t|t-1}, \qquad (4.10)$$

$$d_t = C_t(\overline{x}_{t-1}) - \frac{\partial C_t}{\partial x_t}(\overline{x}_{t-1})\overline{x}_{t-1}$$
(4.11)

6

$$r_t = R(\overline{x}_{t-1}) \tag{4.12}$$

na forma

$$y_t \simeq a_t x_t + b_t + \varepsilon_t \tag{4.13}$$

$$x_t \simeq c_t x_{t-1} + d_t + r_t \eta_t \,. \tag{4.14}$$

Isto resulta num modelo análogo ao modelo descrito pelas equações (3.2) e (3.1). Então aplicamos o FK ao modelo aproximado, obtendo como equações de predição

$$\overline{x}_{t|t-1} = C_t(\overline{x}_{t-1}) \tag{4.15}$$

e

$$P_{t|t-1} = c_t^2 P_{t-1} + r_t^2 \sigma_{n_t}^2, (4.16)$$

e como equações de atualização

$$\overline{x}_t = \overline{x}_{t|t-1} + a_t P_{t|t-1} f_t^{-1} (y_t - A_t(\overline{x}_{t|t-1})), \qquad (4.17)$$

$$P_t = P_{t|t-1} - a_t^2 P_{t|t-1}^2 f_t^{-1}$$
(4.18)

е

$$f_t = a_t^2 P_{t|t-1} + \sigma_{\varepsilon_t}^2 \,. \tag{4.19}$$

Em síntese, aplicamos o FK a uma aproximação linear do modelo original, dada pelas equações (4.13) e (4.14), e obtemos o mesmo filtro descrito no capítulo anterior utilizando, no lugar das equações (3.26) e (3.28), as equações (4.15) e (4.17). Essas duas últimas equações servem apenas para evitar a soma e subtração de termos que se cancelam quando utilizamos as equações do FK obtidas no capítulo anterior.

O objetivo do FEK é análogo ao objetivo do FK, ou seja, é estimar a distribuição do estado no instante t condicionado às observações obtidas até t. Nesse caso ela é aproximada por uma normal com média (4.17) e variância (4.18), a qual denotaremos por $\tilde{p}(x_t|y_t,\theta)$.

Observação 4.1. Em geral não há uma trajetória de referência em torno da qual o sistema evolui. O FK adota como trajetória de referência os valores estimados. Este é o motivo de tomar as expansões em Taylor em torno das médias condicionais.

Como agora estamos tratando de um caso linear, mesmo sendo este uma aproximação a um outro, o procedimento para a obtenção da verossimilhança é análogo ao do capítulo anterior. Se o processo é invariante no tempo, sua verossimilhança pode ser descrita pela equação (3.62), considerando as mudanças nas equações de predição e de atualização dadas acima.

4.2 Modelos Condicionalmente Gaussianos

Aqui apenas mostramos que os modelos podem ser um pouco mais abrangentes. Muitas vezes os coeficientes do processo não dependem somente do tempo, mas também das observações obtidas até o instante anterior, t-1. Condicionando o modelo ao conjunto de

informações obtidas até t-1, ficamos com um modelo condicionalmete gaussiano. Esse pode ser escrito explicitamente por

$$y_t = a_t(\boldsymbol{Y}_{t-1})x_t + b_t(\boldsymbol{Y}_{t-1}) + \varepsilon_t$$
 (observação) (4.20)

$$x_{t} = c_{t}(\boldsymbol{Y}_{t-1})x_{t-1} + d_{t}(\boldsymbol{Y}_{t-1}) + r_{t}(\boldsymbol{Y}_{t-1})\eta_{t}$$
 (estado), (4.21)

em que $\varepsilon_t | \mathbf{Y}_{t-1} \sim \mathcal{N}(0, \sigma_{\varepsilon_t}(\mathbf{Y}_{t-1})), \, \eta_t | \mathbf{Y}_{t-1} \sim \mathcal{N}(0, \sigma_{\eta_t}(\mathbf{Y}_{t-1})) \, e \, x_0 \sim \mathcal{N}(\overline{x}_0, P_0)$. O sistema deve ser considerado fixo, já que estamos no tempo t. Por isso, as equações do FK e as do FEK são obtidas como anteriormente.

4.3 MVEM

Lembremos que os MVEM tratados neste trabalho têm equação de observação dada por

$$H_{t+h} = H_t + h\left(\mu - \frac{\sigma_t^2}{2}\right) + h^{1/2}\sigma_t Z_{t+h}. \tag{4.22}$$

Notemos que o ruído presente na equação das observações está multiplicado por uma função dos estados. Isto impossibilita a aplicação completa do FEK, já que não temos mais a condição de normalidade do ruído satisfeita. Vale ressaltar que, mesmo condicionando esse modelo ao conjunto de observações passadas, a hipótese de normalidade do erro é corrompida.

Usaremos então apenas o primeiro estágio do FEK. Esse estágio nos fornece a fdp da variável latente no tempo t condicionada ao conjunto de informações até t-1. Como já foi dito anteriormente, o inconveniente disso é que apenas podemos estimar os parâmetros do modelo, não tendo um filtro capaz de estimar o estado. Na próxima seção obteremos a verossimilhança de um processo geral, que atende às condições dos MVEM.

4.3.1 Verossimilhança

Obter a verossimilhança condicionada às variáveis de estado de um MVEM não é tão simples como no caso linear ou como no caso aproximadamente linear. Nos MVEM a verossimilhança não é dada por uma forma fechada e requer uma integração n-dimensional sobre o vetor de estados $\boldsymbol{x}=(x_1,x_2,\ldots,x_n)$ como vemos

$$L(\boldsymbol{\theta}) = \int p(\boldsymbol{y}, \boldsymbol{x} | \boldsymbol{\theta}) d\boldsymbol{x} = \int p(\boldsymbol{y} | \boldsymbol{x}, \boldsymbol{\theta}) p(\boldsymbol{x} | \boldsymbol{\theta}) d\boldsymbol{x}. \tag{4.23}$$

4.3. MVEM 45

Levando em consideração a estrutura temporal das observações fatoramos a verossimilhança através de sucessivas condicionais obtendo

$$L(\boldsymbol{\theta}) = p(y_0|\boldsymbol{\theta}) \prod_{t=1}^{n} p(y_t|\boldsymbol{y}_{t-1}, \boldsymbol{\theta}), \qquad (4.24)$$

em que $y_{t-1} = (y_0, y_1, \dots, y_{t-1})$ é o vetor de observações coletadas até o tempo t-1. Cada fator do produtório acima é dado pelas integrais

$$p(y_0|\boldsymbol{\theta}) = \int p(y_0|x_0,\boldsymbol{\theta})p(x_0|\boldsymbol{\theta})dx_0, \qquad (4.25)$$

e, para t = 1, 2, ..., n,

$$p(y_t|\boldsymbol{y}_{t-1},\boldsymbol{\theta}) = \int p(y_t|x_t,\boldsymbol{\theta})p(x_t|\boldsymbol{y}_{t-1},\boldsymbol{\theta})dx_t.$$
 (4.26)

Isto reduz a integral n-dimensional em (4.23) a um produto de n integrais unidimensionais.

Analisemos a integral em (4.25). Como y_0 não é combinação linear de x_0 , tal integral não fica fácil de ser obtida. O trabalho de Meyer, Fournier & Berg (2003) usou a aproximação de Laplace, uma técnica de aproximação de integrais que data do século XVIII, (veja p.ex. Laplace, 1986) e veja apêndice A.2. Através dessa técnica temos

$$p(y_0|\boldsymbol{\theta}) \approx \sqrt{2\pi} e^{\psi_0(y_0,\widehat{x}_0,\boldsymbol{\theta})} |D^2 \psi_0(y_0,\widehat{x}_0,\boldsymbol{\theta})|^{-1/2}, \tag{4.27}$$

em que

$$\psi_0(y_0, x_0, \theta) = -\log(p(y_0|x_0, \theta)p(x_0|\theta)), \qquad (4.28)$$

 \hat{x}_0 é o mínimo da função ψ_0 com relação a x_0 , e $D^2\psi_0(y_0, x_0, \theta)$ denota a derivada de segunda ordem da função $\psi_0(y_0, x_0, \theta)$ com relação a x_0 .

Tendo apenas a observação inicial, podemos, através do teorema de Bayes, atualizar a densidade $p(x_0|\theta)$ do estado desconhecido pela densidade de filtragem $p(x_0|y_0,\theta)$ como mostra a equação

$$p(x_0|y_0,\boldsymbol{\theta}) = \frac{p(y_0|x_0,\boldsymbol{\theta})p(x_0|\boldsymbol{\theta})}{p(y_0|\boldsymbol{\theta})}.$$
(4.29)

Sua distribuição pode ser aproximada por uma distribuição normal dada por

$$\mathcal{N}\left(\widehat{x}_0, I^{-1}(\widehat{x}_0)\right),$$
 (4.30)

em que

$$I(\widehat{x}_0) = -\frac{\partial^2}{\partial x_0^2} \log(p(y_0|x_0, \boldsymbol{\theta})p(x_0|\boldsymbol{\theta})) \bigg|_{x_0 = \widehat{x}_0}, \tag{4.31}$$

ou seja, igual a informação de Fisher observada da função $\log(p(y_0|x_0, \boldsymbol{\theta})p(x_0|\boldsymbol{\theta}))$, (veja p.ex. Seu & Singer, 1993). Ainda vale notar que

$$I(\widehat{x}_0) = D^2 \psi_0(y_0, \widehat{x}_0, \boldsymbol{\theta}).$$
 (4.32)

Expandindo a equação dos estados em série de Taylor de primeira ordem em torno de \widehat{x}_0 , a média de $x_0|\boldsymbol{\theta}$, podemos aplicar o primeiro estágio do FEK ao modelo. Essa etapa consiste em estimar a média e a variância do estado no próximo instante de tempo, condicionado ao conjunto de informações obtidas até então, no caso, $x_1|y_0,\boldsymbol{\theta}$. Note que nessa abordagem, a variância da filtragem, denotada anteriormente por P_0 , é dada por $I^{-1}(\widehat{x}_0)$, e sua média, ao invés de ser o valor estimado do estado, é dada por \widehat{x}_0 . Então, pelas equações (4.15) e (4.16) obtemos que $x_1|y_0,\boldsymbol{\theta}$ tem distribuição aproximadamente normal com média

$$\overline{x}_{1|0} = C_1(\widehat{x}_0) \tag{4.33}$$

e variância

$$P_{10} = c_1^2(\widehat{x}_0)I^{-1}(\widehat{x}_0) + r_1\sigma_n^2. \tag{4.34}$$

Analogamente para $t=1,2,\ldots,n$, a fdp preditiva de $x_t|\boldsymbol{y}_{t-1},\boldsymbol{\theta}$ é aproximada por uma normal, $\widehat{p}(x_t|\boldsymbol{y}_{t-1},\boldsymbol{\theta})$, com média e variância dadas, respectivamente, por

$$\overline{x}_{t|t-1} = C_t(\widehat{x}_{t-1}) \tag{4.35}$$

е

$$P_{t|t-1} = [c_t(\widehat{x}_{t-1})]^2 I^{-1}(\widehat{x}_{t-1}) + r_t^2 \sigma_{\eta}^2.$$
(4.36)

Lembrando que, depois de observar y_t , a fdp $p(x_t|y_t, \theta)$ é atualizada através do teorema de Bayes por

$$p(x_t|y_t, \boldsymbol{\theta}) \propto p(y_t|x_t, \boldsymbol{\theta})p(x_t|\boldsymbol{y}_{t-1}, \boldsymbol{\theta}) \approx p(y_t|x_t, \boldsymbol{\theta})\widetilde{p}(x_t|\boldsymbol{y}_{t-1}, \boldsymbol{\theta}), \qquad (4.37)$$

implicando que $x_t|y_t, \boldsymbol{\theta}$ tem fdp $p(x_t|y_t, \boldsymbol{\theta})$ aproximadamente normal com média \hat{x}_t e variância $I^{-1}(\hat{x}_t)$.

Finalmente podemos aproximar os termos em (4.26) através de

$$p(y_t|\mathbf{y}_{t-1},\boldsymbol{\theta}) \approx \int p(y_t|x_t,\boldsymbol{\theta})\widetilde{p}(x_t|\mathbf{y}_{t-1},\boldsymbol{\theta})dx_t.$$
 (4.38)

Ainda assim, essas integrais e $p(x_0|\boldsymbol{\theta})$ requerem ser aproximadas já que $p(y_t|x_t,\boldsymbol{\theta})$ e $p(x_0|\boldsymbol{\theta})$ não são combinações lineares de uma variável normal adicionado um ruído normal, como no

caso linear apresentado anteriormente. Novamente usaremos para seu cálculo a aproximação de Laplace. Isso nos dá que

$$p(y_t|\boldsymbol{y}_{t-1},\boldsymbol{\theta}) \simeq \sqrt{2\pi} e^{\psi_t(y_t,\widehat{\boldsymbol{x}}_t,\boldsymbol{\theta})} |D^2\psi_t(y_t,\widehat{\boldsymbol{x}}_t,\boldsymbol{\theta})|^{-1/2}, \tag{4.39}$$

em que

$$\psi_t(y_t, x_t, \boldsymbol{\theta}) = -\log(p(y_t|x_t, \boldsymbol{\theta})\widetilde{p}(x_t|\boldsymbol{y}_{t-1}, \boldsymbol{\theta})), \qquad (4.40)$$

 \widehat{x}_t é o mínimo da função $\psi_t(y_t, x_t, \boldsymbol{\theta})$ com relação a x_t , e $D^2\psi_t(y_t, x_t, \boldsymbol{\theta})$ denota a derivada de segunda ordem da função $\psi_t(y_t, x_t, \boldsymbol{\theta})$ com relação a x_t , para t = 1, 2, ..., n. Na prática, encontrar \widehat{x}_t requer alguma técnica de estimação numérica. Sugerimos o uso do algoritmo de Newton-Raphson, veja apêndice A.3.

Completadas as etapas de aproximação dos fatores da equação (4.23), obtemos uma expressão para a verossimilhança que não mais depende do conjunto de variáveis latentes \boldsymbol{x} . Mais precisamente, uma aproximação para a verossimilhança é dada por

$$\widetilde{p}(\boldsymbol{y}|\boldsymbol{\theta}) = (2\pi)^{\frac{n+1}{2}} \exp\left\{ \sum_{t=0}^{n} \psi_t(y_t, \widehat{x}_t, \boldsymbol{\theta}) \right\} \prod_{t=0}^{n} |D^2 \psi_t(y_t, \widehat{x}_t, \boldsymbol{\theta})|^{-1/2}.$$
(4.41)

4.4 Abordagem Bayesiana

Agora que obtemos uma aproximação para a função de verossimilhança, podemos inserir conhecimentos a respeito do comportamento de cada parâmetro, conhecimentos a priori. Por exemplo, no caso dos MVEM, um dos parâmetros deve ser positivo e menor que um. Além disso pesquisas mostram que seu valor fica próximo de um. Novamente temos situações em que os métodos analíticos não são convenientes, sendo necessários métodos de aproximação numérica como o método de Laplace (Tierney et al., 1989) e a integração de Monte Carlo via amostragem por importância, (veja p. ex. Ripley, 1987).

Pelo teorema de Bayes a densidade a posteriori é dada por

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})}, \tag{4.42}$$

em que $p(y) = \int p(y|\theta)p(\theta)d\theta$ é a densidade marginal de θ . No nosso caso, ela é aproximadamente proporcional a

$$\widetilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \widetilde{p}(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$$
 (4.43)

em que $\widetilde{p}(\boldsymbol{y}|\boldsymbol{\theta})$ é dada pela expressão em (4.41). Devemos então maximizar a posteriori aproximada com relação ao vetor de parâmetros $\boldsymbol{\theta}$, ou seja, encontrar sua moda. Para essa tarefa temos, em geral, a necessidade do uso de métodos de estimação numérica devido a complexidade da função.

O trabalho de Meyer et al. (2003) trata da forma canônica de Kim (veja equações (1.23) e (1.24)) e emprega a técnica automatic differentiation (AD), um conjunto de algoritmos computacionais para a obtenção numérica das derivadas de funções não-lineares intratáveis explicitamente, no uso do algoritmo quase-Newton. Além disso, sugere o uso do software AD Model Builder (Fournier, 2000) (http://otter-rsch.com/admodel.htm), um pacote baseado na linguagem C++ que integra a técnica AD com o algoritmo quase-Newton no cálculo de mínimos de funções, para minimizar $-\log(\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}))$.

O uso do algoritmo de Metropolis-Hastings (Hastings, 1970) também é uma alternativa para encontrar a moda a posteriori. No entanto, sabemos que a eficiência desse algoritmo pode ser comprometida por uma má escolha da densidade geradora. Quanto mais próxima essa está da densidade a posteriori, melhor o desempenho do algoritmo. Uma alternativa para tal escolha, sugerida no trabalho de Meyer et al. (2003), é utilizar os conhecimentos obtidos através da técnica AD-quase-Newton na média e na variância da densidade geradora. O autor afirma que essa combinação de recursos resulta numa técnica MCMC bastante flexível, eficaz e adequada aos MVE.

Vale destacar que o uso da AD não é indispensável. Uma vez que a posteriori foi aproximada com o uso do EKF e do teorema de Bayes, podemos usar diretamente o MCMC para obter o vetor de parâmetros estimado.

4.5 Modelo I

Vamos relembrar que a aproximação de Euler que converge para o modelo I, descrito em (2.11) e (2.12), tomando h - 1 e ρ = 0, tem como equação de estado

$$\nu_{t+1} = \nu_t + \beta(\alpha - \nu_t) + \gamma \widetilde{Z}_t, \qquad (4.44)$$

para $t=0,1,\ldots,n$. Estabelecendo condição inicial ν_0 tal que

$$\nu_0 \sim \mathcal{N}\left(\alpha, \frac{\gamma^2}{1 - (1 - \beta)^2}\right)$$
 (4.45)

1.5. MODELO 1

Pela equação (4.27), uma aproximação para $p(y_0|\boldsymbol{\theta})$ é dada por

$$p(y_0|\boldsymbol{\theta}) \approx \sqrt{2\pi} e^{\psi_0(y_0,\widehat{\nu}_0,\boldsymbol{\theta})} [D^2 \psi_0(y_0,\widehat{\nu}_0,\boldsymbol{\theta})]^{-1/2}, \qquad (4.46)$$

com

$$\psi_0(y_0, \nu_0, \boldsymbol{\theta}) = \frac{1}{2} \log(2\pi e^{\nu_0}) + \frac{1}{2} \frac{(y_0 - \mu + e^{\nu_0}/2)^2}{e^{\nu_0}} + \frac{1}{2} \log\left[\frac{2\pi\gamma^2}{1 - (1 - \beta)^2}\right] + \frac{1}{2\gamma^2} \frac{(1 - \beta)^2}{2\gamma^2} (\nu_0 - \alpha)^2, \quad (4.47)$$

 $\widehat{\nu}_0$ é o mínimo da função $\psi_0(y_0, \nu_0, \boldsymbol{\theta})$ com relação a ν_0 e $D^2\psi_0(y_0, \nu_0, \boldsymbol{\theta})$, a derivada segunda da função $\psi_0(y_0, \widehat{\nu}_0, \boldsymbol{\theta})$ com relação a ν_0 . Por (4.30), podemos concluir que $\nu_0|y_0, \boldsymbol{\theta}$ tem distribuição normal com média $\widehat{\nu}_0$, e variância $I^{-1}(\widehat{\nu}_0) = (D^2\psi_0(y_0, \widehat{\nu}_0, \boldsymbol{\theta}))^{-1}$.

Seguindo a mesma notação dada no início deste capítulo temos

$$C_t(\nu_{t-1}) = \nu_{t-1} + \beta(\alpha - \nu_{t-1}) \tag{4.48}$$

G

$$c_t(\nu_{t-1}) = 1 - \beta. (4.49)$$

Então, o primeiro estágio do FEK nos dá que $\nu_t|y_{t-1}$, para $t=1,2,\ldots,n$, tem distribuição aproximada por uma normal com média e variância dadas, respectivamente, pelas seguintes equações

$$\overline{\nu}_{ttt+1} = \widehat{\nu}_{t+1} + \beta(\alpha - \widehat{\nu}_{t+1}) \tag{4.50}$$

e

$$P_{l|t-1} = (1 - \beta)^2 t^{-1}(\widehat{\nu}_{t-1}) + \gamma^2. \tag{4.51}$$

 $\widehat{\nu}_t$ é o mínimo de $\psi_t(y_t, \nu_t, \boldsymbol{\theta})$ com relação a ν_t e $D^2\psi_t(y_t, \nu_t, \boldsymbol{\theta})$, a derivada segunda, com relação a ν_t , da função $\psi_t(y_t, \nu_t, \boldsymbol{\theta})$, cuja expressão para t = 1, 2, ..., n é obtida a seguir. Autes disso, novamente vamos atualizar a densidade do estado através do teorema de Bayes, veja (4.37). E então temos que $p(\nu_t|y_t, \boldsymbol{\theta})$ se aproxima de uma normal com média $\widehat{\nu}_t$ e variância $I^{-t}(\widehat{\nu}_t)$.

Como em (4.39) temos

$$p(y_t|\boldsymbol{y}_{t-1},\boldsymbol{\theta}) \simeq \sqrt{2\pi} e^{\psi_t(y_t,\nu_t,\boldsymbol{\theta})} |D^2\psi_t(y_t,\widehat{\nu}_t,\boldsymbol{\theta})|^{-1/2}, \qquad (4.52)$$

em que

$$\psi_{t}(y_{t}, \nu_{t}, \boldsymbol{\theta}) = \frac{1}{2} \log(2\pi e^{\nu_{t}}) + \frac{1}{2} \frac{(y_{t} - \mu + e^{\nu_{t}}/2)^{2}}{e^{\nu_{t}}} + \frac{1}{2} \log(2\pi P_{t|t-1}) + \frac{1}{2P_{t|t-1}} (\nu_{t} - \widehat{\nu}_{t|t-1})^{2},$$
(4.53)

 $\widehat{\nu}_t$ e $D^2\psi_t(y_t,\nu_t,\pmb{\theta})$ são análogos aos termos em t=0.

Finalmente podemos expressar a verossimilhança por

$$\widetilde{p}(\boldsymbol{y}|\boldsymbol{\theta}) = \exp\left\{-\frac{1}{2}\sum_{t=0}^{n}\log(2\pi e^{\widetilde{\nu}_{t}}) + \frac{1}{2}\sum_{t=0}^{n}\frac{1}{2e^{\widetilde{\nu}_{t}}}\left(y_{t} - \mu + \frac{e^{\widetilde{\nu}_{t}}}{2}\right)^{2} - \frac{1}{2}\sum_{t=0}^{n}\log(2\pi P_{t|t-1}) + \frac{1}{2}\sum_{t=0}^{n}\frac{\widehat{\nu}_{t} - \overline{\nu}_{t|t-1}}{P_{t_{t}t-1}} - \frac{1}{2}\sum_{t=0}^{n}\log(I(\widehat{\nu}_{t}))\right\}. \quad (4.54)$$

4.6 Modelo II

O processo para obter a verossimilhança da aproximação de Euler do modelo II, dada em (2.11) e (2.12), é semelhante ao caso da seção auterior, pois os processos apenas diferem na equação de estado, que neste caso, considerando h=1 e $\rho=0$, é dada por

$$\sigma_{t+1}^2 = \sigma_t^2 + \delta(\lambda - \sigma_t^2) + \xi \sigma_t \widetilde{Z}_t, \qquad (4.55)$$

para t = 0, 1, ..., n. Assumimos condição inicial σ_0^2 de forma que

$$\sigma_0^2 \sim \mathcal{N}\left(\lambda, \frac{\xi^2}{1 - (1 - \delta)^2}\right)$$
 (4.56)

Um aproximação para $p(y_0|\boldsymbol{\theta})$, como mostra (4.27), é dada por

$$p(y_0|\boldsymbol{\theta}) \approx \sqrt{2\pi} e^{\psi_0(y_0,\widehat{\sigma_0^2},\boldsymbol{\theta})} |D^2 \psi_0(y_0,\widehat{\sigma_0^2},\boldsymbol{\theta})|^{-1/2}, \qquad (4.57)$$

com

$$\psi_0(y_0, \sigma_0^2, \boldsymbol{\theta}) = \frac{1}{2} \log(2\pi\sigma_0^2) + \frac{1}{2} \frac{(y_0 - \mu + \sigma_0^2/2)^2}{\sigma_0^2} + \frac{1}{2} \log\left[\frac{2\pi\xi^2}{1 - (1 - \delta)^2}\right] + \frac{1 - (1 - \delta)^2}{2\xi^2} (\sigma_0^2 - \lambda)^2, \quad (4.58)$$

 $\widehat{\sigma_0^2}$ é o mínimo da função $\psi_0(y_0, \sigma_0^2, \boldsymbol{\theta})$ com relação a σ_0^2 e $D^2\psi_0(y_0, \sigma_0^2, \boldsymbol{\theta})$, a derivada segunda da função $\psi_0(y_0, \widehat{\sigma_0^2}, \boldsymbol{\theta})$ com relação a σ_0^2 . Concluímos por (4.30) que $\sigma_0^2|y_0, \boldsymbol{\theta}$ tem distribuição normal com média $\widehat{\sigma_0^2}$, e variância $I^{-1}(\widehat{\sigma_0^2}) = (D^2\psi_0(y_0, \widehat{\sigma_0^2}, \boldsymbol{\theta}))^{-1}$.

4.6. MODELO II 51

Com a mesma notação usada na construção do FEK temos

$$C_t(\sigma_{t-1}^2) = \sigma_{t-1}^2 + \delta(\lambda - \sigma_{t-1}^2), \qquad (4.59)$$

$$R_t(\sigma_{t-1}^2) = \sqrt{\sigma_{t-1}^2},$$
 (4.60)

$$c_t = 1 - \delta \tag{4.61}$$

 \mathbf{c}

$$r_t = R_t(\overline{\sigma^2}_{t-1}) = \sqrt{\overline{\sigma^2}_{t-1}}. \tag{4.62}$$

A partir disso, o primeiro estágio do FEK nos dá que $\sigma_t^2 | \boldsymbol{y}_{t-1}$, para $t=1,2,\ldots,n$, tem distribuição aproximada por uma normal com média e variância dadas, respectivamente, por

$$\overline{\sigma^2}_{t|t-1} = \widehat{\sigma^2}_{t-1} + \delta(\lambda - \widehat{\sigma^2}_{t-1}) \tag{4.63}$$

е

$$P_{t_i t-1} = (1-\delta)^2 I^{-1}(\widehat{\sigma}_{t-1}^2) + r_t^2 \xi^2. \tag{4.64}$$

 $\widehat{\sigma}_t^2$ é o mínimo de $\psi_t(y_t, \sigma_t^2, \boldsymbol{\theta})$ com relação a σ_t^2 e $D^2\psi_t(y_t, \sigma_t^2, \boldsymbol{\theta})$, a derivada segunda, com relação a σ_t^2 , da função $\psi_t(y_t, \sigma_t^2, \boldsymbol{\theta})$, cuja expressão para t = 1, 2, ..., n é obtida a seguir. Mas precisamos antes, atualizar a densidade do estado, veja (4.37). Obtemos que $p(\sigma_t^2|y_t, \boldsymbol{\theta})$ se aproxima de uma normal com média $\widehat{\sigma}_t^2$ e variância $I^{-1}(\widehat{\sigma}_t^2)$.

E pela aproximação de Laplace, como em (4.39), temos

$$p(y_t|\mathbf{y}_{t-1},\boldsymbol{\theta}) \simeq \sqrt{2\pi} e^{\psi_t(\mathbf{y}_t,\widehat{\sigma_t^2},\boldsymbol{\theta})} |D^2\psi_t(y_t,\widehat{\sigma_t^2},\boldsymbol{\theta})|^{-1/2}, \tag{4.65}$$

em que

$$\psi_t(y_t, \sigma_t^2, \boldsymbol{\theta}) = \frac{1}{2} \log(2\pi\sigma_t^2) + \frac{1}{2} \frac{(y_t - \mu + \sigma_t^2/2)^2}{\sigma_t^2} + \frac{1}{2} \log(2\pi P_{t|t-1}) + \frac{1}{2P_{t|t-1}} (\sigma_t^2 - \widehat{\sigma}_{t|t-1}^2)^2,$$
(4.66)

 $\widehat{\sigma_t^2}$ é o mínimo e $D^2\psi_t(y_t, \sigma_t^2, \boldsymbol{\theta})$ a derivada segunda da função $\psi_t(y_t, \sigma_t^2, \boldsymbol{\theta})$, ambos com relação a σ_t^2 .

A verossimilhança para o modelo II discretizado fica expressa por

$$\widetilde{p}(\boldsymbol{y}|\boldsymbol{\theta}) = \exp\left\{-\frac{1}{2}\sum_{t=0}^{n}\log(2\pi\widehat{\sigma_{t}^{2}}) - \frac{1}{2}\sum_{t=0}^{n}\frac{1}{2\widehat{\sigma_{t}^{2}}}\left(y_{t} - \mu + \frac{\widehat{\sigma_{t}^{2}}}{2}\right)^{2} - \frac{1}{2}\sum_{t=0}^{n}\log(2\pi P_{t|t-1}) + \frac{1}{2}\sum_{t=0}^{n}\frac{\widehat{\sigma_{t}^{2}} - \overline{\sigma_{t}^{2}}_{t|t-1}}{P_{t|t-1}} - \frac{1}{2}\sum_{t=0}^{n}\log\left(I(\widehat{\sigma_{t}^{2}})\right)\right\}. \quad (4.67)$$

4.7 Conclusão

Uma extensão do FK foi apresentada de forma a abranger os casos não lineares e permitindo que o ruído da equação do estado dependa de seus valores passados. No entanto, essa abordagem não despreza a hipótese de normalidade do erro da outra equação. Isso nos permitiu aplicar apenas o primeiro estágio do FEK e impossibilitou a estimação do vetor de estados.

Nitidamente obter o vetor de parâmetros $\hat{\boldsymbol{\theta}}$ que maximiza as funções (4.54) e (4.67) não é uma tarefa simples. Mesmo antes desse passo, há a necessidade de se encontrar o mínimo de uma função não-linear, usado na aproximação de Laplace. O algoritmo de Newton-Raphson, descrito no apêndice A.3 e sugerido, na seção (4.3.1), como método numérico para obter tal mínimo, requer o uso de derivadas, que muitas vezes precisam também ser estimadas numericamente. Portanto o uso de técnicas numéricas de estimação é primordial, assim como sua implementação computacional. Sugerimos para suprir a parte final do problema, o algoritmo de Metropolis-Hastings e o uso da inferência bayesiana.

CAPÍTULO

5

Resultados

Neste capítulo apresentaremos resultados empíricos para diferentes conjuntos de parâmetros. Utilizaremos para isso o modelo descrito em (3.35) e (3.36), o qual chamaremos de modelo invariante no tempo (MIT) e seguiremos os procedimentos descritos nas subseções 3.2.1 e 3.2.2.

5.1 Abordagem da Verossimilhança

Utilizando um vetor de parâmetros "verdadeiro" θ , foram gerados 800 dados e descartada a primeira metade deles, para evitar dados enviesados pelo palpite inicial, servindo-nos como amostra os 400 restantes. Depois tratamos o vetor de parâmetros como desconhecido e utilizamos o método descrito na seção 3.2.1 para estimá-lo. No algoritmo de Metropolis-Hastings tomamos como geradora para os parâmetros a, b, c^*, d e γ^{2*} distribuições normais. Através do software MATLAB, geramos 1 cadeia com 60000 iterações para os 5 parâmetros. As primeiras 30000 foram desprezadas. A partir das restantes selecionamos de 30 em 30, resultando numa amostra de 1000 valores para a inferência, dada pela média dos valores

Tabela 5.1: Resultados para o MIT A.

Parâmetro	Verdadeiro	Estimativa	DP	TA (%)	CG
a	2.5	2.4012	0.0872	36.9683	1.2071
b	-0.8	-0.7714	0.3183	77.1067	-0.6782
c	0.9	0.8874	0.0121	6.1233	0.6162
d	0.5	0.5628	0.0643	13.7250	0.1789
γ^2	1.0	1.0635	0.0903	9.7117	-0.8350

Tabela 5.2: Resultados para o MIT B.

Parâmetro	Verdadeiro	Estimativa	DΡ	TA (%)	CG
\overline{a}	3.0000	2.6743	0.1307	30.4817	-1.0728
b	0	0.0944	0.1038	80.1117	0.4750
c	0.7	0.6812	0.0403	15.6833	-0.5081
d	0	-0.0132	0.0523	34.5700	1.1984
γ^2	1.0	1.2191	0.1071	9.4500	1.3848

selecionados. A convergência da cadeia foi avaliada usando-se o critério de Geweke (Geweke, 1992). Seu valor deve ser, em módulo, menor do que 2 para indicar que há convergência da cadeia, veja apêndice A.4.

Os resultados estão sumarizados para dois casos, A e B, nas tabelas 5.1 e 5.2, juntamente com os valores verdadeiros de cada parâmetro, o desvio padrão (DP) da seleção, a taxa de aceitação (TA) do algoritmo Metropolis-Hastings e o valor obtido pelo critério de Geweke (CG). Nas figuras 5.1 e 5.3 são apresentados os histogramas das densidades das amostras selecionadas, respectivamente do MIT A e B, para cada parâmetro. E os gráficos dos 50% últimos valores amostrados são apresentados nas figuras 5.2 e 5.4, em que podemos analisar a convergência de cada parâmetro.

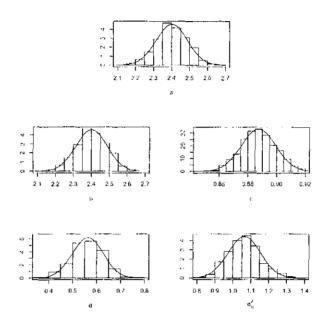


Figura 5.1: Histogramas das amostras selecionadas: MIT A.

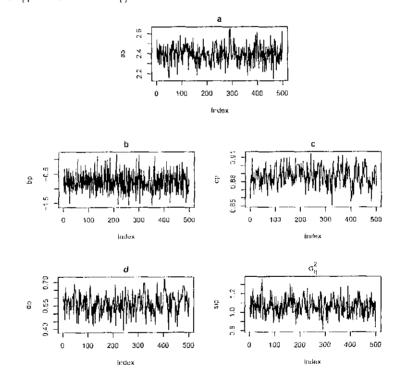


Figura 5.2: Gráficos das amostras selecionadas: MIT $\Lambda.$

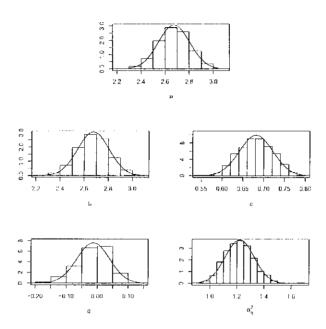


Figura 5.3: Histogramas das amostras selecionadas: MIT B.

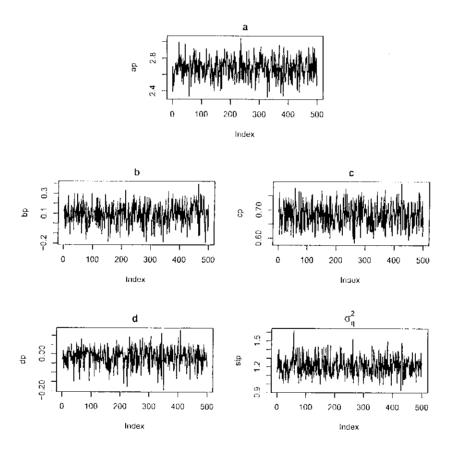


Figura 5.4: Gráficos das amostras selecionadas: MIT B.

Tabela 5.3: Resultados para o MIT A bayesiano.

Parâmetro	Verdadeiro	Estimativa	DP	TA (%)	CG
a	2.5	2.4448	0.1543	31.8933	-1.5390
b	-0.8	-0.7417	0.2294	90.7133	0.0050
c	0.9	0.8859	0.0084	59.7883	1.7412
d	0.5	0.5875	0.0586	41.9217	-0.4240
γ^2	1.0	1.0585	0.1402	29.3733	0.6164

Tabela 5.4: Resultados para o MIT B bayesiano.

Parâmetro	Verdadeiro	Estimativa	DP	TA (%)	CG
а	-1.0	2.4012	0.0872	36.9683	1.2071
b	0.4	-0.7714	0.3183x	77.1067	-0.6782
c	0.7	0.8874	0.0121	6.1233	0.6162
d	0	0.5628	0.0643	13.7250	0.1789
γ^2	0.25	1.0635	0.0903	9.7117	-0.8350

5.2 Abordagem Bayesiana

Foram usados os mesmos dados gerados para os casos A e B da seção anterior com o procedimento descrito na seção 3.2.2. As médias das prioris foram consideradas como sendo a estimativa do parâmetro dada pela abordagem de verossimilhança. Para os parâmetros a, b, c^*, d e γ^{2^*} , tomamos como geradora do algoritmo de Metropolis-Hastings distribuições normais. Também nessa abordagem, geramos 1 cadeia com 60000 iterações. Descartamos as primeiras 30000 e, das restantes fizemos uma seleção com passo de 30 em 30. Ficamos com uma amostra de 1000 valores para a inferência que se deu pela média da amostra selecionada para cada parâmetro. Utilizamos o critério de Geweke para avaliar a convergência. Os resultados numéricos estão sumarizados para os casos A e B nas tabelas 5.3 e 5.4, respectivamente. Já os resultados gráficos do MIT A estão nas figuras 5.5 e 5.6, e os do MIT B, nas figuras 5.7 e 5.8.

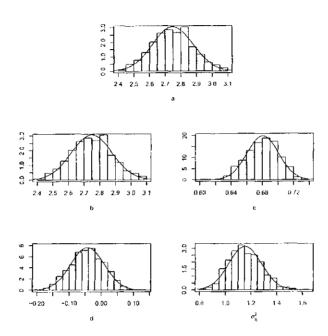


Figura 5.5: Histogramas das amostras selecionadas: MIT A bayesiano.

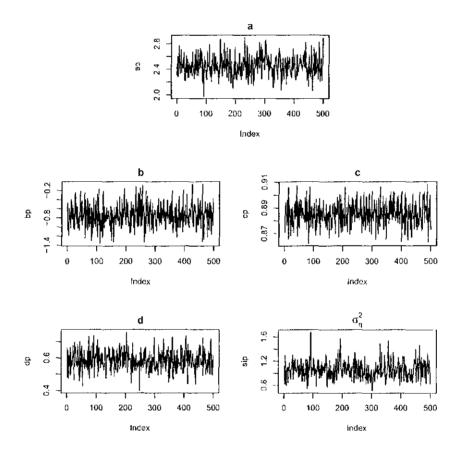


Figura 5.6: Gráficos das amostras selecionadas: MIT A bayesiano.

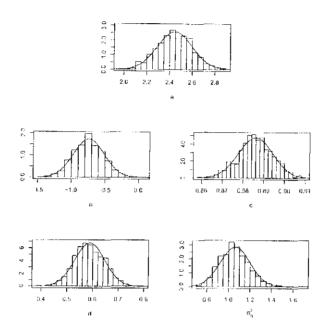


Figura 5.7: Histogramas das amostras selecionadas: MIT B bayesiano.

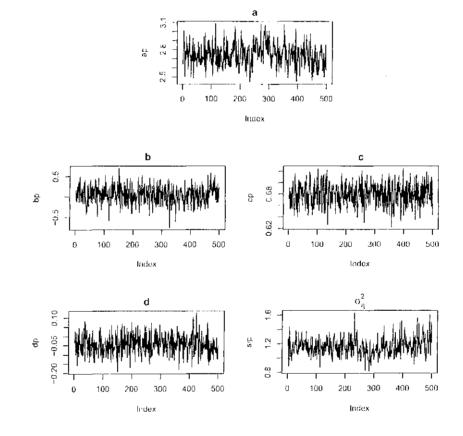


Figura 5.8: Gráficos das amostras selecionadas: MIT B bayesiano.

5.3 Conclusão

Os resultados mostram que a metodologia utilizada é eficaz na estimação dos parâmetros, tanto na abordagem da verossimilhança quanto na bayesiana. Conhecimentos a priori mais específicos podem diminuir o tempo computacional.

APÊNDICE

A

Métodos Numéricos

Neste apêndice apresentaremos alguns métodos numéricos citados neste trabalho. Para mais detalhes a respeito das duas primeiras seções sugerimos Press (1989) e Gamerman (1997), e Milne (1953), da última.

A.1 MCMC

O objetivo dos métodos de Monte Carlo com cadeia de Markov (MCMC) é gerar uma amostra de uma distribuição $p(\theta)$, para $\theta = (\theta_1, \theta_2, \dots, \theta_k) \in \mathbb{R}^k$, que não pode ser gerada diretamente. A idéia é simular uma cadeia de Markov irredutível aperiódica cuja distribuição de equilíbrio é a distribuição de interesse $p(\theta)$. Se o número de simulações é grande, os valores simulados da cadeia podem ser usados como uma base para identificar características da distribuição de interesse, $p(\theta)$.

A seguir damos dois métodos para gerar cadeia de Markov, comumente utilizados nos diversos trabalhos publicados dentro desse contexto.

A.1.1 Amostrador de Gibbs

O amostrador de Gibbs é um método MCMC essencialmente iterativo, cujo núcleo de transição é formado pelas condicionais completas da distribuição $p(\theta)$, dadas por $p(\theta_1|\theta_2,\theta_3,\ldots,\theta_k),\ p(\theta_2|\theta_1,\theta_3,\ldots,\theta_k),\ldots,p(\theta_k|\theta_1,\theta_2,\ldots,\theta_{k-1})$. O algoritmo do processo é dado a seguir.

- 1. inicia com valor $\boldsymbol{\theta}^{(0)}=(\theta_1^{(0)},\dots,\theta_k^{(0)})$ e contador de iteração j=0;
- 2. gera
 - (i) $\theta_1^{(j+1)}$ de $p(\theta_1|\theta_2^{(j)},\theta_3^{(j)},\ldots,\theta_k^{(j)})$;
 - (ii) $\theta_2^{(j+1)}$ de $p(\theta_1|\theta_2^{(j)},\theta_3^{(j)},\ldots,\theta_k^{(j)})$;

:

- (**k**) $\theta_k^{(j+1)}$ de $p(\theta_2|\theta_1^{(j)},\theta_3^{(j)},\ldots,\theta_k^{(j)})$;
- 3. faz j=j+1 e volta ao passo 2 até obter a convergência.

Conforme o número de iterações aumenta, a cadeia se aproxima de sua condição de equilíbrio. Desta forma, assume-se que a convergência foi obtida quando a distribuição da amostra estiver arbitrariamente próxima da distribuição de equilíbrio. Para verificar a convergência do algoritmo existem algumas técnicas. Gelfand & Smith (1995) sugere o uso de técnicas gráficas. Já o critério de Gelman-Rubin (Gelman & Rubin, 1992) e o critério de Geweke (Geweke, 1992) utilizam uma análise estatística dos dados da amostra gerada.

A.1.2 Metropolis-Hastings

Quando as distribuições condicionais não são facilmente identificadas, ou seja, não possuem formas padrões (normal, beta, gama, entre outras), podemos utilizar, para gerar as amostras das cadeias de Markov, o algoritmo **Metropolis-Hastings**, ou métodos de amostragem por importância, Metropolis *et al.* (1953).

Suponhamos que queremos gerar amostras de uma densidade não-regular $p(\theta_i|\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_k) \doteq p(\theta_i|\theta_{(-i)})$. Devemos definir um núcleo de transição $q(\theta,\theta^*)$ da distribuição $p(\theta)$ que represente $p(\theta_i|\theta_{(-i)})$ transformando θ em θ^* . O algoritmo de Metropolis-Hastings pode ser esquematizado como segue

63

- 1. inicia com valor $\theta^{(0)}$ e contador de iteração j=0;
- 2. θ^* de acordo com o núcleo de transição $q(\theta^{(j)}, \theta^*)$;
- 3. atualiza $\theta^{(j)}$ por $\theta^{(j+1)} = \theta^*$ com probabilidade

$$p = \min \left\{ 1, \frac{p(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{(j)}, \boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(j)})q(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(j)})} \right\}, \tag{A.1}$$

ou seja, permanece com $\theta^{(j)}$ com probabilidade 1-p ;

4. faz j=j+1 e repete os passos de 2 e 3 até conseguir uma distribuição estacionária.

Observação A.1. Algumas considerações podem ser feitas, como seguem.

- (i) O algoritmo de Metropolis-Hastings é especificado pela sua densidade candidata para geração q(x,y);
- (ii) se um valor candidato é rejeitado, o valor atual é considerado na próxima etapa;
- (iii) o cálculo da probabilidade de transição p, cm (A.1), não depende da constante normalizadora;
- (iv) se a densidade candidata para geração das amostras é simétrica, ou seja, q(x,y) = q(y,x), a probabilidade de movimento se reduz a $p(\theta^*)/p(\theta^{(j)})$. Assim, se $p(\theta^*) > p(\theta^{(j)})$, a cadeia se move para θ^* ; caso contrário, move-se para θ^* com probabilidade $p(\theta^*)/p(\theta^{(j)})$. Em outras palavras podemos dizer que um salto na direção "ascendente" é sempre aceito, enquanto que um salto na direção "descendente" é aceito com uma certa probabilidade.

A.2 Aproximação de Laplace

A aproximação de Laplace é um método numérico que aproxima assintoticamente a distribuição a posteriori de uma variável aleatória y. Consideremos x uma variável aleatória de densidade p(x) e com p(y|x) conhecida. A densidade a posteriori de y com relação a x é dada por

$$p(y) = \int_{-\infty}^{\infty} p(y|x)p(x)dx.$$
 (A.2)

Sejam $\psi(y,x) = -\log(p(y|x)p(x))$ e $\widehat{x} = \min_x \{\psi(y,x)\}$ o mínimo da função ψ em relação a x. Tomemos a expansão em série de Taylor de segunda ordem de ψ , apenas em relação a x, em torno de \widehat{x} . Assim

$$\psi(y,x) \approx \psi(y,\widehat{x}) - \frac{1}{2}D^2\psi(y,\widehat{x})(x-\widehat{x})^2, \qquad (A.3)$$

em que $D^2\psi(y,\widehat{x})=\partial^2\psi(y,x)/\partial x^2$ calculada em \widehat{x} . Então

$$p(y) \approx e^{-\psi(y,\widehat{x})} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \frac{(x-\widehat{x})^2}{(-D^2 \psi(y,\widehat{x}))^{-1}}\right\} dx. \tag{A.4}$$

Fazendo uma mudança de variáveis obtemos que

$$p(y) \approx \sqrt{2\pi} e^{-\psi(y,\widehat{x})} |D^2 \psi(y,\widehat{x})|^{-1/2}.$$
 (A.5)

A.3 Algoritmo de Newton-Raphson

O método de Newton-Raphson é um algoritmo recursivo para encontrar máximos de funções que possuam derivadas de, ao menos, segunda ordem contínuas. A idéia do algoritmo é utilizar a expansão em séries de Taylor até segunda ordem da função em questão.

Seja f uma função real, com derivadas de primeira e segunda ordens contínuas. O algoritmo de Newton-Raphson para f é descrito a seguir.

- 1. inicia com valor x_0 e contador de iteração j=0;
- 2. enquanto $|f'(x_j)| > \varepsilon$
 - (i) $x_{j+1} = x_j \frac{1}{f''(x_j)} f'(x_j),$
 - (ii) j = j + 1,
 - (iii) $x_{\max} = x_j$.
- 3. o valor que maximiza f localmente é x_{max} .

A derivada primeira da função deve ser nula no ponto de mínimo. O valor dado a ε expressa o quanto é permitido que a derivada difira de zero.

O algoritmo de Newton-Raphson pode ser facilmente estendido para funções multivariadas.

A.4 Critério de Geweke

O critério de Geweke (Geweke, 1992) é um método numérico para avaliar a convergência da quantidade amostrada, quando esta é gerada utilizando-se apenas uma cadeia.

A idéia é desprezar uma parte da cadeia, e dividir a restante em duas partes não necessariamente iguais. Para cada parte, estimam-se sua média e variância. Considera-se que a cadeia completa convergiu, quando as médias das duas seqüências selecionadas estão próximas. E isso se avalia com um teste de comparação das médias. Com um erro de 5 %, valor do teste deve ser, em módulo, menor do que 1,96.

APÊNDICE

В

Modelos Discretos

B.1 Modelos ARCH(p)

O processo ARCH pode ser resumido por

$$Y_t = X_{t-1}\beta + Z_t \tag{B.1}$$

$$Z_t | \Omega_{t-1} \sim P(0, \sigma_t^2) \tag{B.2}$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2,$$
 (B.3)

em que $P(\cdot)$ é uma distribuição paramétrica, usualmente a Normal ou a t-Student, X_t denota um vetor de variáveis exógena, β um vetor de parâmetros desconhecidos e Ω_{t-1} é o conjunto de informações obtidas até o instante t-1, ou seja, $\Omega_{t-1} = \{Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots\}$.

Dada uma série financeira S_t (preço de uma ação, de uma opção de compra ou opção de venda ...) observada em tempo discreto t_1, t_2, t_3, \ldots , queremos modelar a volatilidade do retorno ou do log-retorno dessa série, ou seja, modelar a volatilidade da série Z_t dada por

$$Z_t = \log \frac{S_t}{S_{t-1}} \,. \tag{B.4}$$

Para representar esta série, consideramos em (B.1) $\beta = 0$, e o processo Z_t satisfazendo o modelo

$$Z_t = \sigma_t \varepsilon_t \tag{B.5}$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2 \,, \tag{B.6}$$

em que $\{\varepsilon_t, t \geq 0\}$ é uma seqüência i.i.d. com $\mathbb{E}(\varepsilon_t) = 0$ e $\mathrm{Var}(\varepsilon_t) = 1$.

Por definição Z_t é serialmente não correlacionado com média zero e variância condicional σ_t^2 , uma função do tempo que pode não ser constante. Assumindo que $\varepsilon_t \sim \mathcal{N}(0,1)$ então

$$Z_t | \Omega_{t+1} \sim \mathcal{N}(0, \sigma_t^2)$$
 (B.7)

A equação (B.6) pode ser expressada por

$$\sigma_t^2 = \sigma(Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}, \boldsymbol{\alpha}),$$
(B.8)

em que $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_p)$ é o vetor de parâmetros desconhecidos.

Como σ_t^2 é a variância condicional, devemos assegurar que σ_t^2 seja estritamente positiva para qualquer realização de Z_t . Para isso, impomos que $\alpha_0 \ge 0$ e $\alpha_j > 0, \forall j = 1, 2, ..., p$ em (B.6).

O cálculo da esperança não condicional é simples

$$E(Z_t) = E[E(Z_t|\Omega_{t-1})] = E[(\sigma_t \varepsilon_t | \Omega_{t-1})]$$

$$= E\{E[(\alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2) \varepsilon_t | \Omega_{t-1}]\}$$

$$= E[(\alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2) E(\varepsilon_t | \Omega_{t-1})] = 0.$$
(B.9)

Já para calcularmos a variância não condicional de um modelo ARCH vamos utilizar a Lei das Esperanças Iterativas. Esta lei sugere que as esperanças das observações correntes ou função delas, $g(Z_t)$, com respeito as informações disponíveis no tempo t-J, podem ser encontradas tomando a esperança condicional da informação no período anterior, t-1, e então tomar a esperança condicional da informação no período anterior a este, e assim por diante até chegar em t-J. Portanto a esperança de $g(Z_t)$ no tempo t-J pode ser obtida

por

$$E(Z_t^2) = \underbrace{E}_{Z_1}(\underbrace{E}_{Z_2}(\dots \underbrace{E}_{Z_t}(Z_t^2|\Omega_{t-1})|\Omega_{t-2}|\dots \Omega_1|\Omega_0))$$

$$= \alpha_0 + \alpha_0 \left(\sum_{j=1}^p \alpha_j\right) + \alpha_0 \left(\sum_{j=1}^p \alpha_j\right)^2 + \dots + \alpha_0 \left(\sum_{j=1}^p \alpha_j\right)^t,$$
(B.10)

em que $\mathbb{E}(\cdot)$ é o valor esperado com relação a v.a. Z_t . E então temos que

$$Var(Z_t) = E(Z_t^2) = \frac{\alpha_0}{1 - \sum_{j=1}^p \alpha_j}$$
 (B.11)

Outra forma de demonstrar (B.11) é usando a propriedade descrita na equação (B.6). Como devemos ter $Var(Z_t) > 0$ é necessário que $\alpha_0 > 0$ e $\sum_{j=1}^{p} \alpha_j < 1$. Esta condição é necessária e suficiente para que o processo ARCH(p) tenha covariância estacionária, veja Engle (1982).

ARCH(1)

Quando p = 1 temos

$$\sigma_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2, \quad \alpha_0 > 0, \alpha_1 \ge 0. \tag{B.12}$$

Se $0 \le \alpha_1 < 1$ o processo é estacionário de segunda ordem. Pela equação (B.11) a variância não condicional é dada por

$$Var(Z_t) = \frac{\alpha_0}{1 - \alpha_1}. (B.13)$$

Também temos que Z_t é uma seqüência de variáveis não-correlacionadas (ruído branco) com média zero e variância $\frac{\alpha_0}{1-\alpha_1}$. De fato,

$$Cov(Z_t, Z_{t+k}) = E[Z_t Z_{t+k}] = EE[Z_t Z_{t+k} | \Omega_{t+k-1}] = E[Z_t E(\sigma_{t+k} \varepsilon_{t+k} | \Omega_{t+k-1})] = 0, \quad (B.14)$$

pois $\Omega_{t+k+1} = \{Z_{t+k-1}, Z_{t+k-2}, \ldots\}$ e σ_{t+k} é explicado por Z_{t+k-1} . Além disso, $\mathbb{E}(\varepsilon_{t+k}|\Omega_{t+k-1}) = 0$

Supondo $\varepsilon_t \sim \mathcal{N}(0,1)$ temos

$$E(Z_t^4|\Omega_{t-1}) = E(h_t^4 \varepsilon_t^4 | \Omega_{t-1}) = (\alpha_0 + \alpha_1 Z_{t-1}^2)^2 E(\varepsilon_t^2) | \Omega_{t-1} = 3(\alpha_0 + \alpha_1 Z_{t-1}^2)^2.$$
(B.15)

Assim

$$E(Z_t^4) = E[E(Z_t^4 | \Omega_{t-1})] = 3E[(\alpha_0 + \alpha_1 Z_{t-1}^2)^2] = 3E(\alpha_0^2 + 2\alpha_0 \alpha_1 Z_{t-1}^2 + \alpha_1 Z_{t-1}^4).$$
 (B.16)

Considerando que o processo é estacionário de quarta ordem, temos

$$E(Z_t^4) = 3[\alpha_0^2 + 2\alpha_0\alpha_1 \text{Var}(Z_t) + \alpha_1^2 E(Z_{t-1}^4)] = 3\alpha_0^2 + 6\alpha_0\alpha_1 \frac{\alpha_0}{1 - \alpha_1} + 3\alpha_1^2 E(Z_t^4)$$
(B.17)

$$\Rightarrow (1 - 3\alpha_i^2) E(Z_t^4) = 3\alpha_0^2 + 6\alpha_0^2 \frac{\alpha_1}{1 - \alpha_1}$$
 (B.18)

$$E(Z_t^4) = \frac{3\alpha_0^2}{1 - 3\alpha_1^2} + \frac{6\alpha_0^2}{(1 - 3\alpha_1^2)} \frac{\alpha_1}{(1 - \alpha_1)} = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - 3\alpha_1^2)(1 - \alpha_1)}.$$
 (B.19)

Então, para que momento de quarta ordem seja finito devemos ter $0 \le \alpha_1^2 < 1/3$.

Analisando a curtose de Z_t

$$K = \frac{\mathbb{E}(Z_t^4)}{(\text{Var}(Z_t))^2} = \frac{3\alpha_0^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)} \frac{(1-\alpha_1)^2}{\alpha_0^2} = \frac{3(1-\alpha_1^2)}{(1-3\alpha_1^2)} > 3$$
(B.20)

notamos que embora o modelo seja gaussiano, a distribuição não condicional tem caudas mais pesadas do que as da normal, o que significa uma vantagem ao modelo. Já uma desvantagem do modelo é tratar os retornos positivos e negativos de forma similar uma vez que para o cálculo da volatilidade são usados os quadrados dos retornos.

Muitas representações alternativas para o modelo ARCII foram criadas. Uma delas pode ser obtida fazendo

$$Z_t^2 = \sigma_t^2 + (Z_t^2 - \sigma_t^2), (B.21)$$

em que $\sigma_t = \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2$ e $Z_t = \sigma_t \varepsilon_t$. Assim

$$Z_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t+j}^2 + ((\sigma_t \varepsilon_t)^2 - \sigma_t) = \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2 + \nu_t,$$
 (B.22)

em que $\nu_t = \sigma_t^2(\varepsilon_t^2 - 1)$. Portanto temos um modelo AR(p) para Z_t^2 com ruído ν_t . Através da função de auto correlação parcial do processo Z_t^2 podemos obter a ordem do modelo ARCH(p). Apesar de $\{\nu_t\}$ ser uma sequência de variáveis aleatórias com média zero, não-correlacionada, sua variância é não-constante e sua distribuição não é normal, o que deixa inviável a inferência dos parâmetros do modelo ARCH através da representação em (B.22).

Vale ressaltar que a distribuição leptocúrtica das séries de retornos é melhor caracterizada quando, ao invés de considerarmos a seqüência ε_t com distribuição normal, tomarmos uma distribuição condicional não normal a ε_t . Como por exemplo de algumas distribuições sugeridas a $\varepsilon_t | \Omega_{t-1}$ são: t-Student, mistura normal-loguormal e potência exponencial, veja p.e. Bollerslev (1987), Hsieh (1989) e Baillie & Bollerslev (1989), respectivamente.

B.2 Modelos GARCH(p,q)

Várias das aplicações do modelo ARCH necessitam de muitos valores passados da série. Uma alternativa mais flexível é dada pelo ARCH generalizado, GARCH(p,q), modelado por Bollerslev (1986)

$$Z_t = \sigma_t \varepsilon_t \tag{B.23}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_1 Z_{t-i} + \sum_{i=1}^q \beta_j \sigma_{t-j}^2$$
(B.24)

em que $\{\varepsilon_t, t \geq 0\}$ é uma seqüència i.i.d com $\mathbf{E}(\varepsilon_t) = 0$ e $\mathrm{Var}(\varepsilon_t) = 1$.

O modelo (B.23) pode ser escrito como

$$\sigma_t^2 = \alpha_0 + \alpha(B)\varepsilon_t^2 + \beta(B)\sigma_t^2, \qquad (B.25)$$

em que $\alpha(B)=\alpha_1B+\cdots+\alpha_pB^p$, $\beta(B)=\beta_1+\cdots+\beta_qB^q$ e B é o operador retardo. Para assegurar que a variància seja estritamente positiva devemos ter $\alpha_0>0, \alpha_i\geq 0$ e $\beta_j\geq 0$ para $i=1,\ldots,p,j-1,\ldots,q$.

Em Bollerslev (1986) foi mostrado que o modelo GARCH(p,q) equivale a um ARCH(∞), isto é, um ARCH de ordem infinita, se as raízes de $1 - \beta(B) = 0$ estiverem fora do círculo unitário.

Como no modelo ARCII, considerando $\varepsilon_t \sim \mathcal{N}(0,1)$ temos

$$Z_t | \Omega_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$$
 (B.26)

Uma condição necessária e suficiente para o GARCH(p,q) definido em (B.23) ser estacionário é

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \tag{B.27}$$

Teremos então

$$E(Z_t) = 0, (B.28)$$

$$Var(Z_t) = \frac{\alpha_0}{1 - (\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j)}$$
 (B.29)

e

$$Cov(Z_t, Z_{t+k}) = 0, \quad k \ge 1.$$
 (B.30)

Uma outra representação para (B.23) é obtida considerando

$$\nu_t = Z_t^2 - \sigma_t^2 \,. \tag{B.31}$$

Substituindo a equação (B.31) em (B.23) obtemos

$$Z_t^2 = \alpha_0 + \sum_{i=1}^{p^*} (\alpha_i + \beta_i) Z_{t-i}^2 + \nu_t - \sum_{j=1}^{q} \beta_j \nu_{t-j},$$
 (B.32)

em que $p^* = \max(p,q)$. Logo temos um modelo ARMA(p*, q) para Z_t^2 . No entanto ν_t uão é, em geral, um processo i.i.d. A representação dada por (B.32) é usada para a identificação da ordem do modelo.

GARCH(1,1)

Quando p = q - 1 temos

$$\sigma_t^2 = \alpha_0 + \alpha_1 Z_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \qquad \alpha_0 > 0, \alpha_1 \ge 0, \beta_1 \ge 0.$$
 (B.33)

Para que o processo seja estacionário no sentido amplo devemos ter $\alpha_4+\beta_1<1$. E então teremos

$$E(Z_t) = 0$$
 e $Var(Z_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$ (B.34)

Vamos também analisar a curtose deste processo

$$K = \frac{\mathbb{E}(Z_t^4)}{[\mathbb{E}(Z_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3.$$
 (B.35)

Como no modelo ARCH, temos que as caudas de Z_t são mais pesadas do que as da normal.

Em geral, identificar a ordem de un modelo GARCH a ser ajustado é difícil. Em geral são utilizados modelos de ordem baixa como (1,1), (1,2) ou (2,1).

Nelson (1990a) mostrou que o modelo a tempo discreto GARCII(1,1) converge para um modelo de difusão a tempo contínuo quando os intervalos de tempo ficam arbitrariamente pequenos. Já em Nelson (1992) foi mostrado que se o verdadeiro modelo é um modelo de difusão sem saltos, então as variâncias a tempo discreto são estimadas com consistência por uma média ponderada dos resíduos passados como na formulação GARCII(1,1). As duas pesquisas citadas anteriormente dão exemplos da importância da classe GARCII(p,q), em particular do GARCII(1,1).

B.3 Modelos EGARCH(p,q)

Os modelos GARCH(p,q) são capazes de expressar a volatilidade das séries de retorno das ações. No entanto, os modelos GARCH levam em conta apenas o tamanho dos retornos e não o seu sinal , ou seja, ignoram o fato deles serem positivos ou negativos. Isto sugeriu que se procurasse um modelo no qual σ_t^2 respondesse assimetricamente ao resíduo positivo e negativo.

O trabalho de Nelson (1991) introduziu o modelo exponencial GARCH ou EGARCH dado por

$$\ln(\sigma_t^2) = \alpha_t + \sum_{k=1}^{\infty} \beta_k g(\varepsilon_{t-k}), \quad \beta_1 \equiv 1,$$
(B.36)

em que $\alpha_t, t \geq 1$ e $\beta_k, k \geq 1$ são seqüências não estocásticas reais.

Para descrever a relação entre a mudança de sinal do retorno e a volatilidade existente, a função $g(\varepsilon_t)$ deve ser capaz de expressar a magnitude e o sinal de Z_t .

Um exemplo importante para a função q é:

$$g(\varepsilon_t) = \theta \varepsilon_t + \gamma(|\varepsilon_t| - \mathbb{E}(|\varepsilon_t|)). \tag{B.37}$$

Por construção $g(\varepsilon_t)$ é uma sequência i.i.d. com média zero. Podemos escrever (B.37) na forma

$$g(\varepsilon_t) = \begin{cases} (\theta + \gamma)\varepsilon_t - \gamma \mathbb{E}(|\varepsilon_t|), & \text{se } \varepsilon_t \ge 0\\ (\theta - \gamma)\varepsilon_t - \gamma \mathbb{E}(|\varepsilon_t|), & \text{se } \varepsilon_t < 0. \end{cases}$$
(B.38)

E isto nos permite notar na volatilidade o chamado "efeito alavanca", ou seja, a resposta da volatilidade é mais rápida a retornos negativos do que a retornos positivos. Logo $g(\varepsilon_t)$ permite que o processo da variância condicional $\{\sigma_t^2\}$ responda assimetricamente a quedas e aumentos no preço da ação, ou seja, é capaz de expressar o "efeito alavanca".

B.4 Modelos ARCH-M

Muitas teorias em finanças são baseadas na existência de uma relação entre risco e retorno esperado. Vale ressaltar que o risco é muitas vezes aproximada pela volatilidade. No modelo ARCH-M introduzido por Engle et al. (1987) a esperança condicional é uma função explícita da variância condicional

$$Y_t = X_{t-1}\beta + g(\sigma_t^2) + Z_t$$
 (B.39)

$$Z_t = \sigma_t \varepsilon_t \tag{B.40}$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^p \alpha_j Z_{t-j}^2 \,. \tag{B.41}$$

Neste modelo um aumento na variância condicional será associado com um aumento ou uma diminuição na esperança condicional de Y_t dependendo do sinal da derivada de g com relação a σ_t^2 . As escolhas mais comuns para g são funções lineares ou logarítmicas de σ_t^2 ou σ_t .

B.5 Modelos IGARCH

Chamamos de processo IGARCH um processo GARCH tal que, ao invés da equação (B.27) ser satisfeita, temos

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1.$$
 (B.42)

Se Z_t segue um processo IGARCH então a variância não condicional de Z_t é infinita, e nem o processo nem seu quadrado Z_t^2 têm covariância estacionária. No entanto é possível que Z_t seja originário de um processo estritamente estacionário no sentido da densidade não condicional de Z_t ser a mesma para todo t, veja Nelson (1990b).

Referências Bibliográficas

- Andersen, T. G. (1994). Stochastic autoregressive volatility: A framework for volatility modelling. *Mathematical Finance*, **4**, 75–102.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of The American Statistical Association*, **96**, 42–55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003). Modelling and forecastic realized volatility. *Econometrica*, **2**(71), 579–626.
- Baillie, R. T. & Bollerslev, T. (1989). The message in daily exchange rates: a conditional variance tale. *Journal of Business and Economic Statistics*, 7, 297-305.
- Barndorff-Nielsen, O. E. & Shephard, N. G. (2001). Non-gaussian ornstein-uhlenbeck based models and some of their uses in financial econometrics. *Journal of The Royal Statistical Society*, Series B(63), 167–241.
- Black, F. & Scholes, M. S. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 7, 637–54.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–327.

- Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics, 69, 542–547.
- Bollerslev, T., Chou, R. Y. & Kroner, K. F. (1992). ARCH modelling in finance. *Journal os Econometrics*, **52**, 5–59.
- Bollerslev, T., Engle, R. F. & Nelson, D. B. (1994). ARCH models. *Handbook of Econometrics*, **IV**, 2959–3038. chapter 49.
- Caines, P. E. (1988). Linear Stochastic Systems. Wiley, New York.
- Carnero, A., Peña, D. & Ruiz, E. (2004). Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics*, **2**, 319–342.
- Çinlar, E. (1975). Introduction to stochastic processes. Prentice-Hall, London.
- Davis, M. H. A. & Vinter, R. B. (1985). Stochastic Modelling and Control. Chapman & Hall, London New York.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance os United Kingdom inflation. *Econometrica*, **50**(4), 987–1007.
- Engle, R. F. & Bollerslev, T. (1986). Modelling the persistence of conditional variances.

 Econometric Reviews, 5, 1–50.
- Engle, R. F., Lilien, D. M. & Robins, R. P. (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica*, **55**(2), 391–407.
- Fama, E. F. (1965). The behavior of stock market prices. Journal of Business, 38, 34-105.
- Fournier, D. (2000). AD Model Builder Version 5.0.1. Otter Research Ltd, Canada.
- Gamerman, D. (1997). Markov Chain Monte Carlo, Stochastic Simulation for Bayesian Inference. Chapman & Hall, London.
- Gelfand, A. E. G. & Smith, A. F. M. (1995). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

- Gelman, A. & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7, 457–511.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics*, **4**, 169-193.
- Ghysels, E., Harvey, A. & Renault, E. (1995). Stochastic volatility. *Handbook of Statistics*, 14.
- Griewank, A. & Corliss, G. F. (1991). Automatic Differentiation of Algorithms: Theory, Implementation and Application. SIAM, Philadelphia.
- Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press, Princeton.
- Harvey, A. C. (1989). Forecasting Structural Time Series Models and the Kalman Filter.

 Cambridge University Press, New York.
- Harvey, A. C. & Shephard, N. G. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, **14**, 429-434.
- Harvey, A. C., Ruiz, E. & Shephard, N. (1994). Multivariate stochastic variance models. Review of Economic Studies, 61, 247–264.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57, 247–264.
- Hsich, D. A. (1989). Modelling heteroscedasticity in daily foreing exchange rates. *Journal of Business and Economic Statistics*, **7**, 307–317.
- Hull, J. (1996). Introdução aos mercados futuros e de opções. Bolsa de Mercados & Futuros, second edition.
- Hull, J. & White, A. (1987). The princing of options on assets with stochastic volatilities.

 Journal of Finance, 42, 281–300.
- Issler, J. V. (1999). Estimating and forecasting the volatility of brasilian finance. Revista de Econometria, 19(1).

- Jacquier, E., Polson, N. G. & Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models with fat-tails and correleted errors. *Journal of Econometrics*.
- Kim, S., Shephard, N. & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Econometrics Studies*, **85**, 361–393.
- Koopman, S. J. & Uspensky, E. H. (2002). The stochastic volatility in mean model: Empirical evidence from international stock markets. *Journal of Applied Econometrics*, 17(6), 667–689.
- Laplace, P. S. (1986). Memoir on the probability of the causes of events. *Statistical Science*, 1, 364–378.
- Lee, T. K. Y. & Tse, Y. K. (1991). Term structure of interest rates in singapore asian dollar market. *Journal of Applied Econometries*, **6**, 143–152.
- Metropolis, N., Rosenbluth, A. W., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computation machine. *Journal of Chemical Physics*, 21, 1087–1091.
- Meyer, R., Fournier, D. A. & Berg, A. (2003). Stochastic volatility: Bayesian computation using automatic differentiation and the extended Kalman filter. *Ecometrics Journal*, 6, 408–420.
- Milne, W. E. (1953). Numerical Solution of Differential Equations. Wiley, New York.
- Morettin, P. A. & Toloi, C. M. C. (2004). Análise de Séries Temporais. Projeto Fisher.
- Nelson, D. B. (1990a). ARCH models as diffusion approximation. Journal of Econometric, 45, 7-38.
- Nelson, D. B. (1990b). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, **6**, 318–334.
- Nelson, D. B. (1991). Conditional heteroskecedasticity in asset returns: A new approach. *Econometrica*, **59**(2), 347–370.
- Nelson, D. B. (1992). Filtering and forecasting with misspecified ARCH models i: Getting the right variance with the wrong model. *Journal of Econometrics*, **52**, 61–90.

- Øksendal, B. (1998). Stochastic Differential Equations: An Introduction with Aplications.

 Springer, fifth edition.
- Pelegrín, C. B. (2004). Estimación de Modelos de Volatilidad Estocástica y Modelos de Componentes Inobservadas Condicionalmente Hetrocedásticos. Ph.D. thesis, Universidad Carlos III de Madrid.
- Press, S. J. (1989). Bayesian Statistics: Principles, Models, and Applications. John Wiley Sons.
- Ripley, B. D. (1987). Stochastic Simulation. Wiley, New York.
- Sandmann, G. & Koopman, S. J. (1998). Estimation of stochastic volatility via Monte Carlo maximum likelihood. *Journal of Econometrics*, 87, 271–301.
- Sen, P. K. & Singer, J. M. (1993). Large Sample Methods in Statistics: An Introduction with Applications. Chapman & Hall, New York.
- Taylor, J. B. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices. *Time Series Analysis: Theory and Practice*, 1, 203–226.
- Taylor, S. (1994). Modelling stochastic volatility: A review and comparative study. *Mathematical Finance*, **4**, 183-204.
- Taylor, S. J. (1986). Modelling Financial Time Series. Chichester.
- Tierney, L., Kass, E. E. & Kadane, J. B. (1989). Approximate marginal densities for nonlinear functions. *Biometrika*, 76, 425–433.
- Watanabe, T. & Asai, M. (2001). Stochastic volatility models with heavy-tailed distributions:

 A bayesian approach. *IMES Discussion Papers Series 2001-E-17*.