

Marcos Maia
Thiago da Silva Alves

***Métodos Bayesianos de Estimación e
Comparação de Modelos na Família GARCH***

Curitiba – PR

2008

Marcos Maia
Thiago da Silva Alves

***Métodos Bayesianos de Estimación e
Comparação de Modelos na Família GARCH***

Trabalho de Conclusão de Curso apresentado à Universidade Federal do Paraná para obtenção do título de Bacharel em Estatística.

Orientador:
Prof. Ph.D Ricardo Sandes Ehlers

UNIVERSIDADE FEDERAL DO PARANÁ
DEPARTAMENTO DE ESTATÍSTICA

Curitiba – PR

2008

Ficha Catalográfica

MAIA, M. e ALVES, T. S.

Métodos Bayesianos de estimação e comparação de
modelos na família GARCH

Curitiba, 2008

76 f.

Trabalho de conclusão de curso - Estatística -
UFPR, 2008

Trabalho de Conclusão de Curso sob o título “*Métodos Bayesianos de Estimação e Comparação de Modelos na Família GARCH*”, defendido por Marcos Maia e Thiago da Silva Alves e aprovado em XX de junho de 2008, em Curitiba, Paraná, pela banca examinadora constituída pelos doutores:

Prof. Ph.D Ricardo Sandes Ehlers
Departamento de Estatística - UFPR
Orientador

Prof. Dr. Joel Maurício Correa da Rosa
Departamento de Estatística - UFPR

À Deus, a nossa família e a todos que nos ajudaram.

Agradecimentos

Primeiramente agradecemos ao nosso orientador Prof. Ph.D Ricardo Sandes Ehlers, por todo conhecimento transmitido, seja os obtidos através da matéria Análise de Séries Temporais, quanto aos obtidos em explicações informais sobre Inferência Bayesiana e Econometria Financeira. Agradecemos também por todos os conselhos pertinentes ao desenvolvimento de um verdadeiro e excelente trabalho acadêmico.

Agradecemos também a nossas famílias por todo seu empenho desde o começo do curso, e principalmente nessa fase final, onde sua ajuda foi essencial para construção deste.

E, não poderíamos nos esquecer de nossos colegas que também foram de grande importância na execução deste, seja com simples opiniões, como também com troca de informações e conhecimentos.

Resumo

A maioria das séries econômicas e financeiras caracterizam-se pela não-estacionariedade na média, principalmente, por apresentarem movimentos de baixa e elevada volatilidade, o que dificulta a previsão do seu comportamento futuro. A volatilidade constitui uma característica fundamental nas aplicações em Economia e Finanças, como, por exemplo, a análise das taxas de retorno dos ativos financeiros, a gestão de risco dos ativos financeiros e a análise das taxas de câmbio. Têm-se então, o interesse dos pesquisadores em encontrar modelos para prever a volatilidade de séries financeiras, e esses modelos são conhecidos como modelos não-lineares em séries temporais e fazem parte da família GARCH (*Generalized Autoregressive Conditional Heteroscedastic*). Porém, a estimação dos parâmetros dos modelos da família GARCH via o método clássico da máxima verossimilhança apresentam dificuldades, tais quais, problemas numéricos e restrições de positividade impostas aos parâmetros, por exemplo. Estas são as principais razões que levam à proposição do método bayesiano com a utilização de simulação de Monte Carlo em Cadeias de Markov (MCMC), a ser desenvolvido neste trabalho, haja vista a necessidade em técnicas de otimização numérica. Portanto, neste trabalho o método clássico de estimação de parâmetros via máxima verossimilhança é comparado ao método bayesiano utilizando o algoritmo de Metropolis-Hastings para estimação de um modelo GARCH(1, 1) na análise de três séries conhecidas: Ibovespa, valor do petróleo tipo Brent e tipo WTI utilizando diversos modelos da família GARCH.

Palavras-chave: Volatilidade, Inferência Bayesiana, MCMC, Modelos GARCH.

Abstract

Most of the economic and financial time series are characterized by non-stationarity in the mean, especially because they show low and high volatility movements, which makes prediction of their future behaviour a difficult task. Volatility constitutes a fundamental feature in applications in Economics and Finance, for example in the analysis of the asset returns, in the risk management of financial assets and in the analysis of exchange rates. This leads to a growing interest by researcher in finding models to predict volatility in financial time series. Such models are inherently non-linear and many of them are in the so called GARCH family (Generalized Autoregressive Conditional Heteroscedastic). However, parameter estimation via maximum likelihood in such models present some difficulties such as numerical problems mainly due to restrictions impose on the parameters. These are the main reasons that lead us to propose the Bayesian approach using Markov chain Monte Carlo (MCMC) methods which will be studied in this work. Thus, in this work the classical maximum likelihood method is compared to the Bayesian method using the Metropolis-Hasting algorithm to estimate a GARCH(1, 1) model in the analysis of three well known time series: the Brazilian BOVESPA index, the Brent and WTI type values of crude oil using models of the GARCH family.

Key-words: Volatility, Bayesian Inference, MCMC, GARCH Models

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 1
1.1	Processos Estocásticos	p. 2
1.2	Estacionaridade	p. 2
1.3	Função de Autocovariância	p. 3
1.4	Operador de Retardo	p. 4
2	Modelos Lineares	p. 5
2.1	Ruído Branco	p. 5
2.2	Seqüência Aleatória	p. 5
2.3	Passeio Aleatório	p. 6
2.4	Modelo Auto-Regressivo (AR)	p. 7
2.5	Modelo Médias Móveis (MA)	p. 7
2.6	Modelo Auto-Regressivo e de Médias Móveis (ARMA)	p. 8
2.7	Modelo ARMA Integrado (ARIMA)	p. 8
2.8	Modelo ARIMA Sazonal (SARIMA)	p. 9
3	Séries Financeiras	p. 11

3.1	Alguns Exemplos de Séries Financeiras	p. 12
3.2	Modelos Não-Lineares	p. 13
3.2.1	Modelo Auto-Regressivo com Heteroscedasticidade Condi- cional (ARCH)	p. 14
3.2.2	Modelo ARCH Generalizado (GARCH)	p. 17
3.2.3	Extensões dos modelos ARCH-GARCH	p. 18
4	Estimação em Modelos Não-Lineares	p. 22
4.1	Método da Máxima Verossimilhança	p. 22
4.2	Método Bayesiano	p. 26
4.2.1	Teorema de Bayes	p. 27
4.2.2	Priori	p. 29
4.2.3	Posteriori	p. 31
4.3	Simulação de Monte Carlo via Cadeias de Markov	p. 32
4.3.1	Cadeias de Markov	p. 32
4.3.2	Algoritmo de Metropolis-Hastings	p. 33
4.3.3	Amostrador de Gibbs	p. 35
4.4	Abordagem Bayesiana em Modelos ARCH	p. 36
4.4.1	MCMC em Modelos ARCH	p. 37
4.5	Estendendo para Modelos GARCH	p. 39
4.5.1	MCMC em modelos GARCH	p. 40
5	Aplicações	p. 42
5.1	Descrição dos Dados	p. 42
5.2	Pacote Estatístico	p. 45

5.3 Resultados	p. 47
6 Conclusões	p. 54
Referências Bibliográficas	p. 56
Apêndice A – Funções do R	p. 58

Lista de Figuras

3.1	Exemplo de Série Financeira (Série do Índice BOVESPA)	p. 13
5.1	Série do Índice BOVESPA	p. 43
5.2	Série do Preço do Petróleo tipo BRENT	p. 44
5.3	Série do Preço do Petróleo tipo WTI	p. 45
5.4	Estimação Bayesiana da série IBOVESPA para o parâmetro α_0	p. 49
5.5	Estimação Bayesiana da série IBOVESPA para o parâmetro α_1 e β_1	p. 50
5.6	Estimação Bayesiana da série Petróleo - Brent para o parâmetro α_0	p. 51
5.7	Estimação Bayesiana da série Petróleo - Brent para o parâmetro α_1 e β_1	p. 51
5.8	Estimação Bayesiana da série Petróleo - WTI para o parâmetro α_0	p. 52
5.9	Estimação Bayesiana da série Petróleo - WTI para o parâmetro α_1 e β_1	p. 53

Lista de Tabelas

5.1	Estimativas dos Parâmetros de um GARCH(1,1) utilizando o pacote tseries	p. 47
5.2	Estimativas dos Parâmetros de um GARCH(1,1) utilizando a função lik.garch	p. 48
5.3	Estimativas dos Parâmetros de um GARCH(1,1) utilizando o algo- ritmo Metropolis - Hastings	p. 48

1 *Introdução*

Este trabalho tem por objetivo desenvolver e aplicar a dados reais, métodos Bayesianos para estimação e comparação de modelos conhecidos na literatura econométrica como *Generalized Autoregressive Conditional Heteroscedastic models* ou modelos GARCH. Porém, para que haja o pleno entendimento do mesmo faz-se necessário o conhecimento de alguns conceitos que serão revisados a seguir.

Começamos então pela definição de série temporal, que nada mais é do que uma coleção de observações feitas sequencialmente ao longo do tempo, sendo sua característica mais importante a de que as observações vizinhas são dependentes e se está interessado justamente em analisar e modelar esta dependência. Enquanto em modelos de regressão, por exemplo, a ordem das observações é irrelevante, para a análise em séries temporais a ordem dos dados é crucial. Quando as observações são obtidas continuamente, isto é, a todo instante ao longo do tempo, diz-se que a série temporal é contínua, cuja representação é $X(t)$. Contrariamente, uma série temporal discreta é aquela em que as observações são tomadas em um conjunto discreto, ou seja, em intervalos fixos (e geralmente equiespaçados) de tempo, cuja representação é dada por X_t .

Outros conceitos básicos, porém muito importantes, tais como processos estocásticos, estacionaridade, função de autocovariância e operador de retardo, serão apresentados nas seções subsequentes. Cabe salientar que estes conceitos serão introduzidos resumidamente e que maiores detalhes podem ser obtidos em [8] e [9].

1.1 Processos Estocásticos

Os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas. Então, seja T um conjunto arbitrário; um processo estocástico é uma família $\{X_t, t \in T\}$, tal que para cada $t \in T$, X_t é uma variável aleatória.

Nessas condições, um processo estocástico é uma família de variáveis aleatórias, que supomos definidas num espaço de probabilidades (Ω, A, P) . Normalmente supõe-se que as variáveis aleatórias envolvidas sejam reais, porém, elas podem ser complexas. O conjunto T é normalmente tomado como o conjunto dos inteiros $\mathbb{Z} = \{0, 1, 2, \dots\}$, ou como conjunto dos números reais \mathbb{R} .

1.2 Estacionaridade

Uma das suposições mais freqüentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, com uma variância também constante, refletindo alguma forma de equilíbrio estável.

Todavia, a maior parte das séries que encontramos na prática apresentam alguma forma de não estacionaridade. Assim, as séries econômicas e financeiras apresentam em geral tendências, sendo o caso mais simples aquele em que a série flutua ao redor de uma reta, com inclinação positiva ou negativa (tendência linear).

Intuitivamente, um processo X_t é estacionário se ele se desenvolve no tempo de modo que a escolha de uma origem dos tempos não é importante. Em outras palavras, as características de $X(t + \tau)$, para todo $\tau > 0$, são as mesmas de $X(t)$. Tecnicamente, há duas formas de estacionaridade: a fraca e a estrita.

Um processo estocástico $X = \{X(t), t \in T\}$ diz-se fracamente estacionário se e somente se:

1. $E\{X(t)\} = \mu(t) = \mu$, constante, para todo $t \in T$;

2. $E\{X^2(t)\} < \infty$, para todo $t \in T$;
3. $\gamma(t_1, t_2) = Cov\{X(t_1), X(t_2)\}$ é uma função de $|t_1 - t_2|$.

Já um processo estocástico $X = \{X(t), t \in T\}$ diz-se estritamente estacionário se todas as distribuições conjuntas finito-dimensionais, dadas por $F(x_1, \dots, x_n; t_1, \dots, t_n) = P\{X(t_1) \leq x_1, \dots, X(t_n) \leq x_n\}$, permanecem as mesmas sob translações no tempo, ou seja, $F(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau) = F(x_1, \dots, x_n; t_1, \dots, t_n)$ para quaisquer $t_1, \dots, t_n, \tau \in T$.

Isto significa, em particular, que todas as distribuições unidimensionais são invariantes sob translações do tempo, logo a média $\mu(t)$ e a variância $V(t)$ são constantes, isto é, $\mu(t) = \mu$ e $V(t) = \sigma^2$, para todo $t \in T$. Do mesmo modo, todas as distribuições bidimensionais dependem de $t_2 - t_1$.

1.3 Função de Autocovariância

Define-se a função de autocovariância de uma série temporal estacionária X_t como $\gamma_x(h) = cov(X_{t+h}, X_t)$. A função de autocorrelação de uma série temporal estacionária X_t é definida como $\rho_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)} = Corr(X_{t+h}, X_t)$. As funções de autocovariância e de autocorrelação fornecem uma medida útil do grau de dependência entre os valores de uma série temporal em diferentes períodos. As autocorrelações medem ainda o tamanho e a força da "memória" do processo.

O gráfico das autocorrelações amostrais versus h é chamado de correlograma. Tal gráfico apresenta valores que serão utilizados para caracterizar as propriedades lineares ou não do mecanismo gerador do processo. Porém, não é simples examinar um correlograma e extrair dele as correspondentes propriedades populacionais. O que se faz necessário é averiguar alguns modelos plausíveis que projetam correlogramas de formas reconhecidas.

As funções de autocovariância e autocorrelação amostrais podem ser calculadas para qualquer conjunto de dados e não estão restritas a observações de séries temporais estacionárias. Para dados contendo tendência, a função de autocorrelação exibirá um

decaimento lento na medida que h aumenta, enquanto para dados com um componente periódico determinístico, como sazonalidade, a função exibirá um comportamento similar ao período. Assim, o correlograma pode ser utilizado como um indicador de não-estacionaridade da série temporal.

1.4 Operador de Retardo

Considere a série temporal x_t com $t = 0, \dots, n$. A primeira diferença da série é definida como:

$$\Delta x_t = x_t - x_{t-1}, \quad t = 1, 2, 3, \dots$$

O operador Δ é denominado operador de diferença, e generalizando para n observações temos que o operador de retardo B é definido como:

$$B^n x_t = x_{t-n}, \quad n = 0, 1, 2, \dots$$

Após a introdução de conceitos básicos sobre séries temporais, faz-se necessário a apreciação do leitor a respeito de modelos lineares, modelos estes que são muito importantes e que servirão de base para o entendimento de modelos mais complexos, como os modelos da família GARCH, que são alguns dos modelos não-lineares. Na seqüência veremos uma não exaustiva introdução sobre o mercado financeiro, bem como os modelos utilizados para se modelar a volatilidade financeira. E, por fim, uma explanação do método Bayesiano com a utilização de simulação de Monte Carlo em Cadeias de Markov (MCMC) utilizando o algoritmo de Metropolis-Hastings para estimação de um modelo GARCH(1, 1) em três séries financeiras.

2 Modelos Lineares

Nesta seção serão apresentados alguns processos estocásticos que são utilizados com frequência na especificação de modelos para séries temporais. Porém, como nosso foco principal não são os modelos lineares, e sim os não-lineares, haja vista sua funcionalidade em séries financeiras, apresentaremos resumidamente alguns modelos probabilísticos e maiores detalhes podem ser obtidos em [17], [6], [9] e [8].

2.1 Ruído Branco

Dizemos que $\{\varepsilon_t, t \in \mathbb{Z}\}$ é um ruído branco se as variáveis aleatórias ε_t são não correlacionadas, isto é, $Cov\{\varepsilon_t, \varepsilon_s\} = 0, t \neq s$. Tal processo será estacionário se $E\{\varepsilon_t\} = \mu$ e $Var\{\varepsilon_t\} = \sigma_\varepsilon^2$, para todo t .

Supõe-se então que $\mu_\varepsilon = 0$, logo $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$, ou seja, o erro é um ruído branco. Para maior definição sobre o processo ruído branco consultar [16].

2.2 Sequência Aleatória

Consideremos $\{X_n, n = 1, 2, \dots\}$ uma sequência de variáveis aleatórias definidas no mesmo espaço amostral Ω . Para todo $n \geq 1$, podemos escrever

$$P\{X_1 = \alpha_1, \dots, X_n = \alpha_n\} = P\{X_n = \alpha_n | X_1 = \alpha_1, \dots, X_{n-1} = \alpha_{n-1}\}$$

Onde, os α_j 's representam estados do processo e o espaço pode ser tomado como o conjunto dos reais, onde o caso mais simples é aquele em que temos uma sequência

$\{X_n, n \geq 1\}$ de variáveis aleatórias mutuamente independentes, e neste caso temos

$$P\{X_1 = \alpha_1, \dots, X_n = \alpha_n\} = P\{X_1 = \alpha_1\} \cdots P\{X_n = \alpha_n\}$$

Se as variáveis aleatórias X_1, X_2, \dots tiverem todas a mesma distribuição, teremos, então, uma seqüência de variáveis aleatórias independentes e identicamente distribuídas (*i.i.d.*). Neste caso, o processo X_n é estacionário. Se $E\{X_n\} = \mu$, $Var\{X_n\} = \sigma^2$, para todo $n \geq 1$, então:

$$\lambda_\tau = Cov\{X_n, X_{n+\tau}\} = \begin{cases} \sigma^2 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases}$$

e, conseqüentemente

$$\rho_\tau = \begin{cases} 1 & \tau = 0, \\ 0 & \text{caso contrário.} \end{cases}$$

2.3 Passeio Aleatório

Consideremos uma seqüência aleatória $\{\varepsilon_t, t = 1, 2, \dots\}$, de variáveis aleatórias *i.i.d.* tais que $E(\varepsilon_t) = \mu_\varepsilon$ e $Var(\varepsilon_t) = \sigma_\varepsilon^2$. Definamos agora uma nova seqüência dada por $X_t = \varepsilon_1 + \dots + \varepsilon_t$. Segue-se que $E(X_t) = t\mu_\varepsilon$ e $Var(X_t) = t\sigma_\varepsilon^2$, ou seja, ambas dependem de t . Então, não é difícil mostrar que $\lambda_x(t_1, t_2) = \sigma_\varepsilon^2 \min(t_1, t_2)$ e portanto a autocovariância de X_t depende de t_1 e t_2 . Este processo é chamado de passeio aleatório ou casual. À medida que o tempo passa, X_t tende a oscilar ao redor de $t\mu_\varepsilon$ com amplitude crescente. O processo é claramente não estacionário.

Passeios aleatórios têm grande importância em econometria e finanças, foco deste trabalho, onde uma suposição usual é de que os preços de ativos financeiros sigam um passeio aleatório Gaussiano, ou seja, $\varepsilon_t \sim N(0, h_t)$.

2.4 Modelo Auto-Regressivo (AR)

Um modelo Auto-Regressivo é definido de maneira que os valores da séries no tempo t dependem dos valores passados. Mais especificamente, dizemos que $\{X_t, t \in Z\}$ é um processo auto-regressivo de ordem p , e escrevemos $X_t \sim AR(p)$, se satisfizer à seguinte equação de diferenças,

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \dots + \alpha_p(X_{t-p} - \mu) + \varepsilon_t$$

sendo que $\mu, \alpha_1, \dots, \alpha_p$ são parâmetros definidos nos reais e ε_t é um ruído branco, ou seja, ε_t é serialmente não correlacionado.

O comportamento da função de autocorrelação de um processo auto-regressivo é uma mistura de decaimento exponencial e/ou decaimento senoidal. Se as raízes da equação auto-regressiva forem reais, então as autocorrelações decairão exponencialmente. Caso as raízes sejam complexas, o decaimento será de forma senoidal.

2.5 Modelo Médias Móveis (MA)

Dizemos que $\{X_t, t \in Z\}$ é um processo de médias móveis de ordem q , e escrevemos $X_t \sim MA(q)$, se satisfizer à seguinte equação de diferenças,

$$X_t = \mu + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$$

onde $\mu, \theta_1, \dots, \theta_q$ são constantes reais e ε_t é um ruído branco. Segue-se que X_t é estacionário e de média μ .

Para um processo MA(q) utilizando o operador de retardo temos:

$$B^j X_t = x_{t-j}, \quad \forall j$$

Então, podemos escrever um modelo médias móveis da seguinte forma:

$$X_t = (1 + \beta_1 B + \beta_2 B^2 + \dots + \beta_q B^q) \varepsilon_t = \theta(B) \varepsilon_t$$

onde $\theta(B)$ é um polinômio de ordem q em B .

2.6 Modelo Auto-Regressivo e de Médias Móveis (ARMA)

Combinando modelos auto-regressivos e de médias móveis pode-se obter uma representação adequada com um número menor de parâmetros. Processos auto-regressivos médias móveis (ARMA) formam uma classe de modelos muito úteis e parcimoniosos para descrever dados de séries temporais. Estes modelos são bastante populares em algumas áreas, como Economia, onde é natural pensar no valor de alguma variável no instante t com função de valores defasados da mesma variável.

Então, um processo auto-regressivo e de médias móveis, de ordem (p, q) , denotado por $ARMA(p, q)$ é definido por:

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q}$$

sendo que ε_t é um ruído branco e a média do processo é μ .

Cabe salientar que as funções de autocorrelação e autocorrelação parcial ficam consideravelmente mais complicadas em processos ARMA. De uma forma geral, para um processo $ARMA(p, q)$ estacionário a função de autocorrelação tem um decaimento exponencial ou oscilatório após a defasagem q , enquanto que a função de autocorrelação parcial tem o mesmo comportamento após a defasagem p [5]. A princípio este resultado pode ser utilizado para auxiliar na determinação da ordem (p, q) do processo, mas na prática é bastante difícil distinguir entre os decaimentos exponenciais e oscilatórios através das estimativas destas funções.

2.7 Modelo ARMA Integrado (ARIMA)

Os modelos discutidos até agora são apropriados para séries temporais estacionárias. Assim, para ajustar estes modelos a uma série temporal observada é necessário remover as fontes de variação não estacionárias. Por exemplo, se a série observada for não estacionária na média pode-se tentar remover a tendência tomando-se uma ou

mais diferenças, que é uma abordagem muito utilizada em séries financeiras. Se d for um número inteiro não negativo, então uma série temporal X_t é dita ser um processo ARIMA(p, d, q) ou um processo ARMA integrado de ordem d se

$$W_t = \alpha_1 W_{t-1} + \dots + \alpha_p W_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$$

for um processo ARMA(p, q) causal [6]. Um processo integrado é utilizado para séries não-estacionárias homogêneas.

2.8 Modelo ARIMA Sazonal (SARIMA)

Muitas séries temporais contêm uma componente periódica sazonal que se repete a cada s observações ($s > 1$). Por exemplo, com dados mensais e $s = 12$ espera-se que X_t dependa de X_{t-12} e talvez de X_{t-24} além de X_{t-1}, X_{t-2}, \dots

Neste caso, tomar a primeira diferença ($x_t - x_{t-1}$) não é suficiente para tornar a série estacionária. A forma apropriada de diferenciar dados com padrão sazonal acentuado é tomar diferenças no período sazonal.

Para dados mensais, por exemplo, a primeira diferença sazonal é

$$\nabla_{12} x_t = (1 - B^{12})x_t = x_t - x_{t-12}$$

e terá variabilidade menor que a primeira diferença não sazonal $\nabla x_t = x_t - x_{t-1}$, sendo portanto mais difícil de identificar e estimar.

Em geral, uma diferença sazonal é denotada por ∇_s onde s é o período sazonal. A D -ésima diferença sazonal é então denotada por ∇_s^D . Combinando-se os dois tipos de diferenciação obtém-se o operador $\nabla^d \nabla_s^D$.

[5] generalizaram o modelo ARIMA para lidar com sazonalidade e definiram o modelo ARIMA sazonal, denominado SARIMA, como

$$\phi(B)\phi(B^s)W_t = \theta(B)\Theta(B^s)\varepsilon_t.$$

Este modelo é chamado de SARIMA multiplicativo de ordem $(p, d, q)(P, D, Q)$ e à

primeira vista parece muito complicado, porém, na prática os valores de d e D em geral não são maiores do que 1 e um número pequeno de coeficientes será suficiente.

Agora que já vimos alguns conceitos primordiais para o bom entendimento do presente trabalho, e haja vista que seu enfoque é modelos aplicados ao mercado financeiro, no próximo capítulo será apresentado ao leitor uma introdução sobre o mercado financeiro, bem como os modelos utilizados para se modelar a volatilidade financeira, objetivo do presente trabalho.

3 *Séries Financeiras*

Em finanças, um dos principais objetivos é a avaliação de riscos de uma carteira de ativos financeiros, sendo estes riscos freqüentemente medidos em termos de variação de preço de ativos. As séries financeiras apresentam uma alta volatilidade além de geralmente não serem estacionárias na média, fazendo-se necessário o uso do chamado *retorno* cuja definição é dada a seguir.

Denotando por P_t o preço de um ativo no instante t , normalmente um dia de negociação, o retorno é definido como:

$$Y_t = \ln(P_t/P_{t-1}). \quad (3.1)$$

As variações de um dia para o outro costumam ser pequenas e por isso a razão P_t/P_{t-1} oscila em torno de 1, e tomando-se o logaritmo desta razão, em geral, obtemos uma distribuição simétrica em torno de zero. Se considerarmos que $\ln(1 + \Delta) \approx \Delta$ para Δ pequeno, então o retorno definido desta forma é aproximadamente igual a variação relativa do preço dos ativos analisados. Na prática é preferível trabalhar com retornos, que são livres de escala, do que com os preços, pois os primeiros têm propriedades estatísticas mais interessantes (como estacionaridade e ergodicidade). Então, um dos objetivos da análise é modelar os retornos.

A característica de interesse nestas séries de retornos é a sua volatilidade, a qual está diretamente associada à variabilidade dos preços de um determinado ativo. Então, se os preços variam muito, diz-se que o ativo é muito volátil. Usualmente utiliza-se a variância ou o desvio padrão como uma medida da volatilidade, onde sua estimação e previsão são fundamentais tanto para quantificar o risco de um determinado ativo

quanto para a precificação de produtos financeiros. Outra característica interessante é que os retornos financeiros raramente apresentam tendências e sazonalidades, além de, em geral, serem não correlacionados, apresentando agrupamentos de volatilidades ao longo do tempo, bem como geralmente não possuindo uma distribuição incondicional normal.

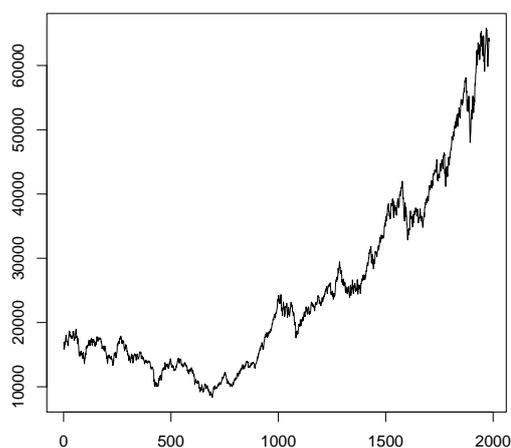
Diversas classes de modelos podem ser utilizadas para modelar esta volatilidade. Em particular os modelos não-lineares, tais quais: ARCH, GARCH e suas várias extensões são utilizados com maior frequência, ver [20], modelos estes que serão descritos na seção 3.2. Na seção 3.1 mostraremos um exemplo de série financeira.

3.1 Alguns Exemplos de Séries Financeiras

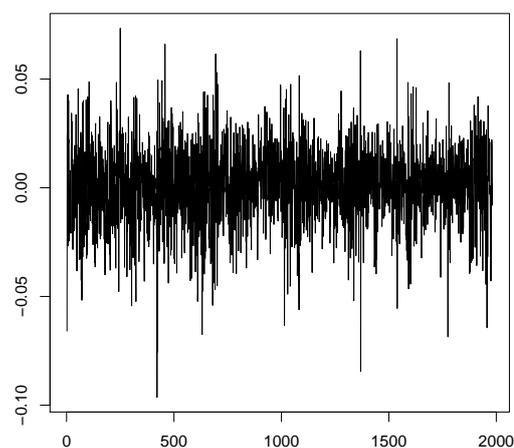
Após a introdução de séries financeiras, faz-se necessário alguns exemplos para o entendimento do leitor. Na seqüência veremos uma breve análise descritiva sobre uma série financeira conhecida: Índice Bovespa (IBV). Cabe salientar que os resultados obtidos foram frutos das funções anexas ao apêndice (A).

Na figura 3.1 (a) temos a série de índices diários do IBOVESPA (Índice da bolsa de valores de São Paulo) no período de 03 de fevereiro de 2000 à 27 de dezembro de 2007, num total de $N= 1982$ observações. Na figura 3.1 (b) temos o retorno desta série, no qual notamos os fatos introduzidos anteriormente. Na figura 3.1 (c) temos o histograma dos retornos, com densidade ajustada, e na figura 3.1 (d) temos um gráfico QxQ, onde os quartis empíricos dos dados são plotados contra os quartis da normal padrão. Se os dados fossem normalmente distribuídos, os pontos estariam sobre a reta, o que não ocorre, fazendo-se necessário o uso de modelos não-gaussianos.

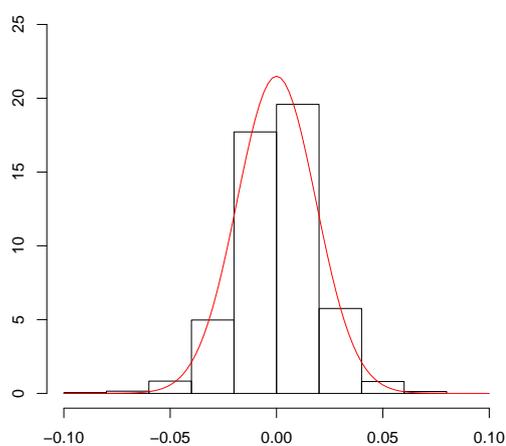
Figura 3.1: Exemplo de Série Financeira (Série do Índice BOVESPA)



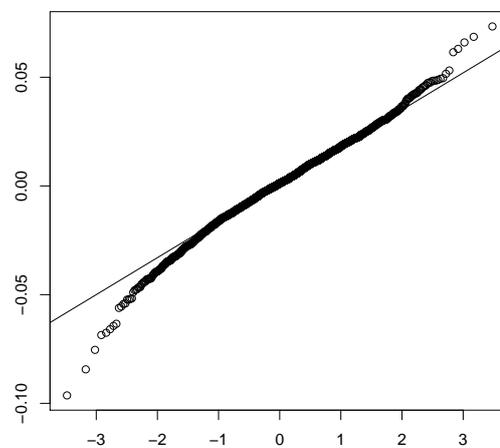
(a) gráfico da série



(b) retorno da série



(c) histograma dos retornos



(d) QxQ plot

3.2 Modelos Não-Lineares

Nesta seção iremos estudar alguns modelos apropriados para séries financeiras que apresentam a variância condicional evoluindo no tempo. Os modelos lineares do tipo

ARIMA, como visto anteriormente, não são adequados para descrever esse tipo de comportamento. Maiores detalhes podem ser obtidos em [20], [11] e [27].

Há uma variedade muito grande de modelos não-lineares disponíveis na literatura, mas vamos nos concentrar na classe de modelos ARCH (*autoregressive conditional heterocedastic*), GARCH (*generalized ARCH*) e algumas de suas extensões. Estes modelos são não-lineares no que se refere a variância, como veremos nas próximas seções.

Como dito anteriormente, o objetivo será modelar o que se chama de volatilidade, que nada mais é do que a variância condicional de uma variável, comumente chamada de retorno, definido em (3.1). Embora não seja medida diretamente, a volatilidade manifesta-se de várias maneiras numa série financeira:

1. a volatilidade aparece em grupo, de maior ou menor variabilidade;
2. a volatilidade evolui continuamente no tempo, podendo ser considerada estacionária;
3. ela reage de modo diferente a valores positivos ou negativos da série.

3.2.1 Modelo Auto-Regressivo com Heteroscedasticidade Condicional (ARCH)

Os modelos ARCH, ou modelos auto-regressivos com heteroscedasticidade condicional foram introduzidos por [10], com o objetivo de estimar a variância da inflação. A idéia básica é que o retorno X_t é não-correlacionado serialmente, mas a volatilidade depende de retornos passados por meio de uma função quadrática.

Um modelo ARCH de ordem r , ou ARCH(r), é definido como,

$$\begin{aligned} X_t &= \sqrt{h_t} \varepsilon_t \\ h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_r X_{t-r}^2 \end{aligned}$$

sendo que os ε_t são assumidos independentes e identicamente distribuídos. Como $h_t > 0, \forall t$ precisamos impor as restrições de positividade, $\alpha_0 > 0$ e $\alpha_i \geq 0, i = 1, \dots, r$.

Na prática, usualmente supomos que $\varepsilon_t \sim N(0, 1)$, $\varepsilon_t \sim t_v$ (t de Student com v graus de liberdade) ou alguma outra distribuição que descreva melhor as caudas pesadas das séries financeiras. Mais recentemente, a distribuição GED (*generalized error distribution*) bem como as versões assimétricas destas distribuições têm sido postuladas também como possíveis candidatas. Os coeficientes α_i devem satisfazer certas condições, dependendo do tipo de imposição que colocamos sobre o processo X_t .

Pela própria definição, valores grandes de X_t são provavelmente seguidos por outros valores grandes da série. Em um modelo ARCH os retornos apresentam geralmente caudas mais pesadas que a normal, de modo que a curtose é maior do que 3, o que é uma propriedade vantajosa do modelo. Por outro lado, uma desvantagem do modelo é que trata retornos positivos e negativos de forma similar, já que os quadrados dos retornos entram na fórmula da volatilidade. Na prática, sabe-se que a volatilidade reage de modo diferente a retornos positivos e negativos. Também, devido ao fato de termos retornos ao quadrado, alguns retornos grandes e isolados podem conduzir a super-previsões.

Identificação

Um primeiro passo na construção de modelos ARCH é tentar ajustar modelos ARMA, para remover a correlação serial na série, se esta existir. Sendo assim, tem-se

$$\phi(B)X_t = \theta_0 + \theta(B)a_t$$

sendo que $a_t \sim \text{ARCH}(r)$. Então, quando refere-se a X_t , supõe-se que ou a série é não correlacionada ou ela é o resíduo da aplicação de um modelo ARMA à série original.

Para verificarmos se a série apresenta heterocedasticidade condicional, pode-se utilizar dois testes, o Teste de Box-Pierce Ljung e/ou o Teste de multiplicadores de Lagrange.

Dada a forma h_t de modelarmos a volatilidade e dado que X_t^2 é um estimador não-viesado de h_t , o valor atual do quadrado do retorno depende dos quadrados dos retornos passados, comportamento similar a de um modelo auto-regressivo. Segue-se então que a função de autocorrelação parcial de X_t^2 pode ser usada para sugerir a ordem

r de um modelo ARCH(r).

Estimação

Os estimadores dos parâmetros do modelo são obtidos pelo método de máxima verossimilhança condicional. A função de verossimilhança é dada por

$$L(x_1, \dots, x_T | \alpha) = f(x_T | F_{T-1}) f(x_{T-1} | F_{T-2}) \cdots f(x_{r+1} | F_r) f(x_1, \dots, x_r | \alpha)$$

e supondo normalidade dos ε_t podemos escrever

$$L(x_1 \cdots x_T | \alpha) = \left[\prod_{t=r+1}^T (2\pi\sigma_t^2)^{-1/2} \exp\left(\frac{-x_t^2}{2\sigma_t^2}\right) \right] f(x_1, \dots, x_r | \alpha).$$

Para T grande, que é em geral o caso das aplicações em Finanças, o termo $f(x_1, \dots, x_r | \alpha)$ pode ser desprezado sem grande perda de informação. Logo, temos que maximizar a função de verossimilhança condicional

$$L(x_1 \cdots x_T | \alpha) = \prod_{t=r+1}^T (2\pi\sigma_t^2)^{-1/2} \exp\left(\frac{-x_t^2}{2\sigma_t^2}\right)$$

sendo que a volatilidade σ_t^2 é obtida recursivamente.

Verificação

Para um modelo ARCH(r), com ε_t normal ou t -Student os resíduos padronizados

$$\tilde{X}_t = \frac{X_t}{\sqrt{h_t}}$$

são variáveis aleatórias i.i.d com distribuição normal padrão ou t -Student. Logo, uma maneira de verificar se o modelo é adequado é calcular a estatística Q de Ljung-Box para a seqüência \tilde{X}_t . Além disso, podemos calcular os coeficientes de assimetria e curtose estimados e fazer um gráfico quantil-quantil para avaliar a suposição de normalidade. Para verificar se ainda existe heterocedasticidade condicional nos resíduos, pode-se aplicar o teste do multiplicador de Lagrange para a seqüência X_t^2 .

Previsão

As previsões para a volatilidade utilizando o modelo ARCH(r) são obtidas recursivamente. Assim,

$$\hat{h}_t(1) = \alpha_0 + \alpha_1 X_t^2 + \cdots + \alpha_r X_{t-r+1}^2$$

é a previsão de h_{t+1} , com origem fixada no instante t . As previsões l passos a frente, com origem em t , são dadas por

$$\hat{h}_t(l) = \alpha_0 + \sum_{i=1}^r \alpha_i \hat{h}_t(l-i),$$

em que $\hat{h}_t(l-i) = X_{t+l-i}^2$, se $l-i \leq 0$.

3.2.2 Modelo ARCH Generalizado (GARCH)

Uma generalização dos modelos ARCH foi sugerida por [4], o chamado modelo GARCH ("generalized ARCH"). Vimos que um modelo ARMA pode ser mais parcimonioso, no sentido de apresentar menos parâmetros do que um modelo AR ou MA puro. Do mesmo modo, um modelo GARCH pode ser usado para descrever a volatilidade com menos parâmetros do que um modelo ARCH. Um modelo GARCH(r, s) é definido por:

$$\begin{aligned} X_t &= \sqrt{h_t} \varepsilon_t \\ h_t &= \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j} \end{aligned}$$

sendo h_t a variância condicional e os ε_t independentes e identicamente distribuídos. As restrições de positividade e estacionariedade são dadas por,

$$\begin{aligned} \alpha_0 &> 0 \\ \alpha_i &\geq 0, \quad i = 1, \dots, r \\ \beta_j &\geq 0, \quad j = 1, \dots, s \\ \sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j &< 1. \end{aligned}$$

Coeficientes positivos dão uma condição suficiente, mas não necessária para $h_t > 0$. Assim como no caso dos modelos ARCH, usualmente supõe-se que os ε_t são normais ou seguem uma distribuição leptocúrtica como a t de Student, ou ainda, uma distribuição de erro generalizada.

Para os modelos GARCH temos as mesmas vantagens e desvantagens dos modelos ARCH. Volatilidades altas são precedidas de retornos ou volatilidades grandes, observando-se os grupos de volatilidades presentes em séries financeiras.

3.2.3 Extensões dos modelos ARCH-GARCH

Há uma literatura vasta sobre extensões dos modelos ARCH-GARCH. Nesta seção mostraremos alguns destes modelos para conhecimento do leitor.

Os modelos da família GARCH assumem que os termos dos erros positivos e negativos têm um efeito simétrico sobre a volatilidade. Ou seja, boas ou más notícias não influenciam em nada os modelos, pois as mesmas notícias têm o mesmo efeito sobre a volatilidade. Na prática, essa hipótese é freqüentemente violada pelo retorno das ações, em que a volatilidade aumenta mais depois uma má notícia do que de uma boa notícia. De um ponto de vista empírico a volatilidade reage assimetricamente ao sinal dos choques e, por conseguinte, uma série de extensões do modelo padrão GARCH têm sido sugeridas recentemente.

Modelo Exponencial GARCH (EGARCH)

Vimos que os modelos ARCH e GARCH tratam simetricamente os retornos, pois a volatilidade é uma função quadrática destes retornos. Por outro lado, também é sabido que a volatilidade reage de forma assimétrica aos retornos, tendendo a ser maior para retornos negativos do que para retornos positivos. Tendo em vista o exposto, [21] introduziu os modelos EGARCH (GARCH exponencial). Um modelo EGARCH(r, s)

é definido como,

$$X_t = \sqrt{h_t} \varepsilon_t$$

$$\log(h_t) = c + \sum_{i=1}^r \alpha_i \left(\left| \frac{\varepsilon_{t-i}}{\sqrt{h_{t-i}}} \right| + \gamma \frac{\varepsilon_{t-i}}{\sqrt{h_{t-i}}} \right) + \sum_{j=1}^s \beta_j \log(h_{t-j}).$$

Note que neste modelo não há restrições de positividade sobre os parâmetros já que estamos modelando o logaritmo da volatilidade. A única restrição é $\gamma < 0$ pois assim a volatilidade aumenta.

Modelo TARARCH

O modelo TARARCH (*Threshold ARCH*) é um caso particular do modelo ARCH não-linear, e a volatilidade segue a seguinte fórmula:

$$h_t^\gamma = \alpha_0 + \alpha_1 g^{(\gamma)}(\varepsilon_{t-1}) + \beta_1 h_{t-1}^\gamma$$

onde

$$g^{(\gamma)}(\varepsilon_t) = \theta I_{\{\varepsilon_t > 0\}} |\varepsilon_t|^\gamma + (1 - \theta) I_{\{\varepsilon_t \leq 0\}} |\varepsilon_t|^\gamma$$

Para $\gamma = 1$ temos o modelo de Zakoian e para $\gamma = 2$ o modelo GJR.

Modelo GARCH Integrado (IGARCH)

Em um modelo GARCH(1,1), se a soma dos parâmetros $\alpha_1 + \beta_1$ for igual a um, têm-se o modelo IGARCH ("integrated GARCH"). Um modelo IGARCH é definido por:

$$X_t = \sqrt{h_t} \varepsilon_t$$

$$h_t = \alpha_0 + \beta_1 h_{t-1} + (1 - \beta_1) X_{t-1}^2$$

com $0 < \beta_1 < 1$. Porém vale notar que nesse caso a variância incondicional de X_t não está definida.

GARCH com potência assimétrica (APARCH)

O modelo *Asymmetric Power ARCH*, ou APARCH, desenvolvido por Ding, Granger e Engle em 1993 foi elaborado a partir do GARCH, incluindo um parâmetro de assimetria por rotação de impacto. O modelo APARCH(r, s) é formalmente definido por:

$$h_t^\delta = \alpha_0 + \sum_{i=1}^r \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^s \beta_j h_{t-j}.$$

onde, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\delta \geq 0$ e $-1 \leq \gamma \leq 1$ para $i = 1, \dots, r$, $j = 1, \dots, s$.

Vários casos particulares deste modelo são listados a seguir. Maiores detalhes podem ser vistos em [19].

Modelo de Taylor & Schwert (TS-GARCH (p, q))

O modelo TS-GARCH desenvolvido por Taylor em 1986 [26] e Schwert em 1990 [25] é outro modelo popular usado para capturar a informação contida na cauda espessa, que é comum em distribuição do retorno do preço das especulações. O modelo é um caso particular do modelo APARCH quando $\delta=1$ e $\gamma=0$, sendo sua especificação baseada no desvio padrão. O modelo é definido como:

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i |\varepsilon_{t-i}| + \sum_{j=1}^q \beta_j h_{t-j}$$

Modelo de Glosten, Jagannathan e Runkle (GJR-GARCH (p, q))

Glosten, Jagannathan e Runkle [15] apresentaram um modelo denominado como GJR-GARCH, que é um caso particular do APARCH quando $\delta=2$. O modelo é dado pela seguinte equação:

$$h_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^2 + \sum_{j=1}^q \beta_j h_{t-j}^2$$

A assimetria é capturada pelo coeficiente γ , o qual indica a influência com que os choques negativos apresentam impactos maiores do que os positivos sobre a volatili-

dade, logo se $\gamma = 0$, não haverá efeito assimétrico. O coeficiente β mede a persistência dos choques nas variâncias futuras.

Modelo de Zakoian (T-GARCH (p, q))

O modelo TGARCH, ou Threshold GARCH [28], é uma caso particular do modelo APARCH, onde $\delta=1$, além de ser um modelo similar ao GJR-GARCH, sendo que aquele aplica-se ao desvio padrão condicional, ao contrário deste que se aplica a variância condicional. O modelo é dado pela seguinte equação:

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}) + \sum_{j=1}^q \beta_j h_{t-j}$$

Modelo de Higgens e Bera (N-ARCH (p, q))

O modelo N-ARCH, ou *Nonlinear ARCH*, é um caso particular do modelo APARCH quando $\gamma=0$. Este modelo, introduzido por Higgens e Bera [18], faz uma transformação de Box-Cox sem, entretanto admitir expoentes diferentes para desvios padrão e choques. O modelo N-ARCH não admite assimetria e foi especificado originalmente como um modelo ARCH, mas pode ser facilmente estendido para a especificação GARCH. O modelo é dado pela seguinte equação:

$$h_t^\delta = \alpha_0 + \sum_{i=1}^p \alpha_i (|\varepsilon_{t-i}|)^\delta$$

Agora que já temos conhecimento sobre os modelos utilizado em séries econométricas, torna-se necessário o uso de métodos para estimação dos mesmos modelos, e isto será assunto do próximo capítulo, onde descreveremos o método de máxima verossimilhança e o método bayesiano. Ao método bayesiano será dado maior enfoque, pois o mesmo é objeto do presente trabalho, além das técnicas de simulação de Monte Carlo via cadeias de Markov, amostrador de Gibbs e algoritmo de Metropolis- Hastings.

4 Estimação em Modelos Não-Lineares

A determinação de estimadores por máxima verossimilhança dos parâmetros dos modelos da família GARCH exige a maximização de uma função de verossimilhança não linear, ou seja, as estimativas por máxima verossimilhança só podem ser obtidas numericamente, onde muitos procedimentos para identificação, ajuste e diagnóstico desses modelos, assim como previsões de valores futuros de séries econométricas usam propriedades da teoria assintótica. Como estes modelos são muito distantes da linearidade, quaisquer uma dessas abordagens tornam-se analiticamente inviáveis, onde uma alternativa para contornar essas dificuldades é considerar a abordagem Bayesiana dos modelos com heterocedasticidade.

Nas primeiras seções deste capítulo veremos brevemente a filosofia do método de máxima verossimilhança, assim como algumas definições de Inferência Bayesiana. Posteriormente descrever-se-á sobre o método de simulação via MCMC. Por fim, será mostrado sua aplicação a modelos da família ARCH-GARCH.

4.1 Método da Máxima Verossimilhança

Embora bastante antigo, foi a partir dos anos oitenta que o método de máxima verossimilhança começou a ser utilizado extensivamente em econometria, onde seu grande obstáculo à utilização prática do método consiste na freqüente incapacidade de obter-se uma solução explícita para a maioria dos problemas em questão. Neste sentido, existe a necessidade de utilizar-se algum método de otimização numérica para

a obtenção dos parâmetros de interesse.

A grande importância do método de máxima verossimilhança consiste nas boas propriedades assintóticas dos estimadores, que são consistentes e assintoticamente eficientes.

Seja uma amostra aleatória y_1, y_2, \dots, y_n , retirada de uma população com uma função densidade de probabilidade $f(y|\theta)$, a qual depende do vetor de parâmetros θ , tem uma função de densidade de probabilidade conjunta dada por:

$$\prod_{i=1}^n f(y_i|\theta)$$

Isto é, a função de densidade de probabilidade conjunta é simplesmente o produto das densidades de cada uma das observações,

$$f(y_1|\theta)f(y_2|\theta)\dots f(y_n|\theta)$$

Note que, antes da retirada da amostra, cada valor observado é uma variável aleatória cuja função de densidade de probabilidade é igual a função de densidade de probabilidade da população. A média e a variância de cada observação a ser retirada são iguais à média e variância da população em questão. É neste sentido que dizemos que na função de densidade conjunta, antes de retirada a amostra, θ é fixo e y_i é variável.

Contudo, uma vez que tenha sido obtida uma amostra específica, y_i torna-se fixo e a função de densidade de probabilidade conjunta pode então ser interpretada como sendo uma função do vetor de parâmetros θ , que se tornam variáveis. Para uma dada amostra y_1, y_2, \dots, y_n a função de densidade de probabilidade conjunta vista como função do vetor de parâmetros desconhecidos θ , é denominada de função de verossimilhança.

Em Econometria, o problema que se coloca é o de: dada uma amostra, obter-se uma estimativa dos valores dos parâmetros populacionais desconhecidos. Uma possibilidade para a resolução do problema de estimação é escolher o vetor $\hat{\theta}$ que maximize a probabilidade de obtenção da amostra específica y_1, y_2, \dots, y_n que se tem em

mãos. Em outras palavras, queremos o vetor $\hat{\theta}$ que torna a probabilidade de obter-se a amostra já obtida a maior possível, ou seja, temos que achar o $\hat{\theta}$ que maximize a função de verossimilhança.

Têm-se então, a função de verossimilhança $L(\theta, y)$, onde y é fixo e θ é variável, e o problema consiste em obter-se o vetor $\hat{\theta}$ que maximiza esta função. O estimador de máxima verossimilhança $\hat{\theta}$ é o vetor que faz

$$L(\hat{\theta}, y) > L(\hat{\theta}, y)$$

onde $\hat{\theta}$ é qualquer outro estimador de θ em um espaço paramétrico.

Do ponto de vista matemático a implementação deste procedimento parece ser simples, pois tudo que temos a fazer é maximizar a função de verossimilhança com respeito a $\hat{\theta}$. Para tanto, basta igualar a zero as derivadas parciais da função de verossimilhança e achar o vetor θ que resolve este conjunto de equações. Na maioria dos casos trabalha-se com o logaritmo natural da função de verossimilhança $\ln(L)$, pois maximizar o logaritmo natural de uma função é, em geral, mais simples e produz os mesmos resultados da maximização da função original.

Considere as seguintes definições:

1. escore eficiente (*efficient score*): $\frac{\partial \ln L}{\partial \theta}$;
2. matriz de informação: $E \left(-\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \right)$.

Note que o estimador de máxima verossimilhança ($\hat{\theta}$) será a solução do conjunto de equações $S(\theta) = 0$. Mais ainda, dadas algumas condições bem gerais, é possível mostrar-se que ($\hat{\theta}$) é consistente, assintoticamente normalmente distribuído e tem variância $[I(\theta)]^{-1}$. Este valor é conhecido como o limite inferior de Cramer-Rao, pois não existe outro estimador consistente do vetor θ que tenha variância menor. Neste sentido, o estimador de máxima verossimilhança ($\hat{\theta}$) também é assintoticamente eficiente.

Apresentaremos agora um exemplo do método de máxima verossimilhança usado em séries financeiras, mas especificamente para o modelo ARCH.

Dada uma série temporal $Z = \{z_t\}$, $t = 1, 2, \dots, T$, do processo estocástico z_t assume-se que a probabilidade que melhor explica este conjunto de observações tem forma conhecida (Gaussiana, t-Student, etc.) e depende de um vetor de parâmetros fixos, α , porém desconhecidos.

Esta densidade, chamada de função de verossimilhança de Z dado α , é denotado por $L(Z|\alpha)$, sendo escrita para o modelo ARCH como:

$$L(Z|\alpha) \propto \prod_{t=p+1}^T \frac{1}{\sqrt{h_t}} \exp \left\{ -\frac{z_t^2}{2h_t} \right\}$$

na qual $\alpha = [\alpha_0 \ \alpha_1 \ \dots \ \alpha_p]'$, e o sobrescrito ($'$) indica transposição de vetores. Como dito anteriormente, para evitarmos o trabalho com produtos, é comum calcular o logaritmo da função de verossimilhança, ou seja:

$$l(Z|\alpha) = \log(L(Z|\alpha)) = \frac{1}{T} \sum_{t=p+1}^T l_t(Z|\alpha)$$

obtendo assim a função log-verossimilhança da t -ésima observação:

$$l_t(Z|\alpha) = -\frac{1}{2} \log(h_t) - \frac{z_t^2}{2h_t}$$

Para um dado conjunto de observações Z , a estimativa de máxima verossimilhança do vetor de parâmetros é definida como aquele valor de α que maximiza $L(Z|\alpha)$ ou $l(Z|\alpha)$. Como as equações são não-lineares, [10] propôs o uso de técnicas de otimização numérica para calcular as estimativas de máxima verossimilhança para os parâmetros α_j , $j = 1, 2, \dots, p$, tais como o método de quasi-Newton descrito pela seguinte equação recursiva:

$$\alpha^{(k+1)} = \alpha^{(k)} + [\hat{I}]^{-1} \left[\frac{\partial l^{(k)}}{\partial \alpha} \right] \quad (4.1)$$

onde $[\hat{I}^{(k)}]$ é a matriz de informação e $[\partial l^{(k)} / \partial \alpha]$ é o gradiente da função log-verossimilhança, ambos avaliados para a estimativa $\alpha^{(k)}$.

Pode-se mostrar que o gradiente da função $l(Z|\alpha)$ é dado por:

$$\frac{\partial l}{\partial \alpha} = \frac{1}{T} \sum_{t=p+1}^T \frac{1}{2h_t} v_t \left(\frac{z_t^2}{h_t} - 1 \right)$$

em que $v_t = [1z_{t-1}^2 \dots z_{t-p}^2]$. Já a matriz de informação $I\{I_{ij}\}, i, j = 0, 1, \dots, p$ é definida como:

$$I_{ij} = -E \left[\frac{1}{T} \sum_{t=p+1}^T \frac{\partial^2 l_t}{\partial \alpha_i \partial \alpha_j} \right]$$

onde, usando o fato de que $E(2z_t^2/h_t - 1) = 2E(z_t^2)/h_t - 1 = 1$, os elementos da matriz de informação $[\hat{I}_{ij}]$ podem ser consistentemente estimados por:

$$I_{ij} = \frac{1}{T} \sum_{t=p+1}^T \frac{1}{2h_t^2} \left\{ \frac{\partial h_t}{\partial \alpha_i} \frac{\partial h_t}{\partial \alpha_j} \right\}$$

Vale enfatizar que é comum a ocorrência de problemas numéricos no método de máxima verossimilhança, causados principalmente pelo mal-condicionamento da matriz de informação. Esta situação causa imprecisões na inversão matricial requerida em 4.1, particularmente para valores mais altos da ordem do modelo ARCH, por exemplo. Outra dificuldade no método de máxima verossimilhança é considerar as restrições de positividade e estacionariedade impostas ao parâmetros. Neste trabalho estas restrições são relaxadas e somente são verificadas após o cálculo das estimativas. Estas são as principais razões que levam à proposição do método bayesiano a ser mostrado a seguir, haja vista que não há necessidade de técnicas de otimização numérica.

4.2 Método Bayesiano

A informação que se têm sobre um parâmetro de interesse é fundamental em qualquer série que se deseja estudar. Porém, o verdadeiro valor deste parâmetro é desconhecido e a idéia que se tem é tentar reduzir este desconhecimento, além da intensidade desta incerteza que pode assumir diferentes graus. Do ponto de vista Bayesiano, estes diferentes graus de incerteza são representados através de modelos probabilísticos para o parâmetro em estudo. Neste contexto, é natural que diferentes pesquisadores pos-

sam ter diferentes graus de incerteza sobre o parâmetro, especificando modelos distintos. Sendo assim, não existe nenhuma distinção entre quantidades observáveis e os parâmetros de um modelo estatístico, todos são considerados quantidades aleatórias.

4.2.1 Teorema de Bayes

Considere uma quantidade de interesse desconhecida θ (tipicamente não observável). A informação de que dispomos sobre θ resumida probabilisticamente através de $p(\theta)$, pode ser aumentada observando-se uma quantidade aleatória X relacionada com θ . A distribuição amostral $p(x|\theta)$ define esta relação. A idéia de que após observar $X = x$ a quantidade de informação sobre θ aumenta é bastante intuitiva e o teorema de Bayes é a regra de atualização utilizada para quantificar este aumento de informação,

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(\theta, x)d\theta} \quad (4.2)$$

Note que $1/p(x)$ não depende de θ e funciona como uma constante normalizadora de $p(\theta|x)$.

Para um valor fixo de x , a função $l(\theta;x) = p(x|\theta)$ fornece a plausibilidade ou verossimilhança de cada um dos possíveis valores de θ , enquanto $p(\theta)$ é chamada distribuição a priori de θ . Estas duas fontes de informação priori e verossimilhança são combinadas levando à distribuição a posteriori de θ , $p(\theta|x)$. Assim, a forma usual do teorema de Bayes é

$$p(\theta|x) \propto l(\theta;x)p(\theta)$$

Em outras palavras temos que

$$\text{distribuição a posteriori} \propto \text{verossimilhança} \times \text{distribuição a priori.}$$

Note que, ao omitir o termo $p(x)$ a igualdade em (4.2) foi substituída por uma proporcionalidade. Esta forma simplificada do teorema de Bayes será útil em problemas que envolvam estimação de parâmetros já que o denominador é apenas uma constante normalizadora. Em outras situações, como seleção de modelos, este termo têm papel

crucial.

É intuitivo também que a probabilidade a posteriori de um particular conjunto de valores de θ será pequena se $p(\theta)$ ou $l(\theta;x)$ for pequena para este conjunto. Em particular, se atribuirmos probabilidade a priori igual a zero para um conjunto de valores de θ então a probabilidade a posteriori será zero, para qualquer que seja a amostra observada.

A constante normalizadora da posteriori pode ser facilmente recuperada, pois $p(\theta|x) = kp(x|\theta)p(\theta)$, onde

$$k^{-1} = \int p(x|\theta)p(\theta)d\theta = E_{\theta}p[X|\theta] = p(x)$$

chamada de distribuição preditiva. Esta é a distribuição esperada para a observação x dado θ . Assim, antes de observar X podemos checar a adequação da priori fazendo predições via $p(x)$. Se X observado recebia pouca probabilidade preditiva então o modelo deve ser questionado.

Em séries econométricas o maior interesse é na previsão do processo em pontos não observados do tempo. Suponha então que, $X = x$, estamos interessados na previsão de uma quantidade Y , também relacionada com θ , e descrita probabilisticamente por $p(y|x, \theta)$. A distribuição preditiva Y dado por x é obtida por integração, como

$$p(y|x) = \int p(y, \theta|x)d\theta = \int p(y, \theta|x)p(\theta|x)d\theta \quad (4.3)$$

Em muitos problemas estatísticos a hipótese de independência condicional entre X e Y dado θ está presente e a distribuição preditiva fica:

$$p(y|x) = \int p(y|\theta)p(\theta|x)d\theta$$

Em muitas aplicações práticas a integral em (4.3) não têm solução analítica e precisará ser obtida por algum método de aproximação.

Então, fica claro que os conceitos de priori e posteriori são relativos àquela observação que está sendo considerada no momento. Assim, $p(\theta|x)$ é a posteriori de θ em relação a X (que já foi observado) mas é a priori de θ em relação a Y (que ainda

não foi observado). Após observar $Y = y$ uma nova posteriori é obtida aplicando-se novamente o Teorema de Bayes. Então, observando-se as quantidades x_1, x_2, \dots, x_n , independentes dado θ e relacionadas a θ através de $p_i(x_i|\theta)$.

Ou seja, a ordem em que as observações são processadas pelo Teorema de Bayes é irrelevante. Na verdade, elas até podem ser processadas em subgrupos.

4.2.2 Priori

A utilização de informação a priori em Inferência Bayesiana requer a especificação de uma distribuição a priori para a quantidade de interesse. Esta distribuição deve representar (probabilisticamente) o conhecimento que se tem sobre o parâmetro antes da realização do experimento. Nesta subseção serão discutidas diferentes formas de especificação da distribuição a priori.

Prioris conjugadas

A partir do conhecimento que se tem sobre o parâmetro de interesse, pode-se definir uma família paramétrica de densidades. Neste caso, a distribuição a priori é representada por uma forma funcional, cujos parâmetros devem ser especificados de acordo com este conhecimento. Estes parâmetros indexadores da família de distribuições a priori são chamados de hiperparâmetros para distingui-los dos parâmetros de interesse.

Esta abordagem em geral facilita a análise e o caso mais importante é o de prioris conjugadas. A idéia é que as distribuições a priori e a posteriori pertençam a mesma classe de distribuições e assim a atualização do conhecimento que se tem sobre o parâmetro de interesse envolve apenas uma mudança nos hiperparâmetros. Neste caso, o aspecto sequencial do método Bayesiano pode ser explorado definindo-se apenas a regra de atualização dos hiperparâmetros, já que as distribuições permanecem as mesmas.

Seja $F = \{p(x|\theta), \theta \in \Theta\}$ uma classe de distribuições amostrais, então, uma classe

de distribuições P é conjugada a F se

$$\forall p(x|\theta) \in F \text{ e } p(\theta) \in P \Rightarrow p(\theta|x) \in P$$

[12] alerta sobre o cuidado com a utilização indiscriminada de prioris conjugadas. Essencialmente, o problema é que a priori conjugada nem sempre é uma representação adequada da incerteza a priori. Sua utilização está muitas vezes associada à tratabilidade analítica decorrente.

Uma vez entendidas suas vantagens e desvantagens a questão que se coloca agora é como obter uma família de distribuições conjugadas.

1. Identifique a classe P de distribuições para θ tal que $l(\theta;x)$ seja proporcional a um membro desta classe;
2. Verifique se P é fechado por amostragem, isto é, se $\forall p_1, p_2 \in P \exists k$ tal que $kp_1p_2 \in P$.

Se, além disso, existe uma constante k tal que $k^{-1} = \int l(\theta;x)d\theta < \infty$ e todo $p \in P$ é definido como $p(\theta) = kl(\theta;x)$ então P é a família conjugada natural ao modelo amostral gerador de $l(\theta;x)$.

Priori não Informativa

Priori não Informativa refere-se a especificação de distribuições a priori quando se espera que a informação dos dados seja dominante, no sentido de que a informação a priori é vaga. Os conceitos de conhecimento vago, não informação, ou ignorância a priori claramente não são únicos e o problema de caracterizar prioris com tais características pode se tornar bastante complexo.

Por outro lado, reconhece-se a necessidade de alguma forma de análise que, em algum sentido, consiga captar esta noção de uma priori que tenha um efeito mínimo, relativamente aos dados, na inferência final. Tal análise pode ser pensada como um ponto de partida quando não se consegue fazer uma elicitacão detalhada do verdadeiro

conhecimento a priori. Neste sentido, serão apresentadas aqui algumas formas de como fazer enquanto discussões mais detalhadas são encontradas em [2] e [3].

A primeira idéia de não informação a priori que se pode ter é pensar em todos os possíveis valores do parâmetro como igualmente prováveis, isto é com uma distribuição a priori uniforme. Neste caso, fazendo $p(\theta) \propto k$ para θ variando em um subconjunto da reta significa que nenhum valor particular tem preferência. Porém, esta escolha de priori pode trazer algumas dificuldades técnicas,

1. Se o intervalo de variação de θ for ilimitado então a distribuição a priori é imprópria, isto significa que

$$\int p(\theta)d\theta = \infty$$

2. Se $\phi = g(\theta)$ é uma reparametrização não linear monótona de θ então $p(\phi)$ é não uniforme já que pelo teorema de transformação de variáveis

$$p(\phi) = p(\theta(\phi)) \left| \frac{d\theta}{d\phi} \right| \propto \left| \frac{d\theta}{d\phi} \right|.$$

Na prática, como estaremos interessados na distribuição a posteriori, não daremos muita importância à impropriedade da distribuição a priori. No entanto, devemos sempre nos certificar de que a posteriori é própria antes de fazer qualquer inferência.

4.2.3 Posteriores

A distribuição a posteriori de um parâmetro contém toda a informação probabilística a respeito deste parâmetro e um gráfico da sua função de densidade a posteriori é a melhor descrição do processo de inferência. No entanto, algumas vezes é necessário resumir a informação contida na posteriori através de alguns poucos valores numéricos. O caso mais simples é a estimação pontual de onde se resume a distribuição a posteriori através de um único número, por exemplo a moda a posteriori. Neste caso pode-se fazer comparações com a estimação de máxima verossimilhança já que, se a priori for pouco informativa, as estimativas pontuais de ambos os métodos devem ser similares. O ideal no entanto seria obter toda a distribuição a posteriori, ou ao menos

uma estimativa desta distribuição, já que em geral ela não pode ser obtida analiticamente. Algumas ferramentas para atacar este problema serão descritas nas próximas seções.

4.3 Simulação de Monte Carlo via Cadeias de Markov

Nesta seção será descrito o método de simulação de Monte Carlo via cadeias de Markov (MCMC), bem como o amostrador de Gibbs e o algoritmo de Metropolis-Hastings. Mais detalhes podem ser obtidos em [12], [24] e [13].

Existem várias maneiras de resumir a informação descrita na distribuição a posteriori, e esta etapa envolve a avaliação de probabilidades ou esperanças. Em vários métodos de simulação a amostra da distribuição a posteriori é obtida em um único passo, onde os valores são gerados de forma independente e não há a preocupação com a convergência do algoritmo, bastando apenas que o tamanho da amostra seja suficientemente grande. É por isso que esses métodos são chamados de iterativos. Porém, em muitos problemas pode ser bastante difícil, ou até mesmo impossível, encontrar uma densidade de importância que seja simultaneamente uma boa aproximação da posteriori e fácil de ser amostrada.

Os métodos de Monte Carlo via cadeias de Markov (MCMC) são uma alternativa aos métodos não iterativos em problemas complexos, onde o princípio básico é obter uma amostra da distribuição a posteriori e calcular estimativas amostrais de características desta distribuição.

4.3.1 Cadeias de Markov

Uma cadeia de Markov é um processo estocástico $\{X_0, X_1, \dots\}$ tal que a distribuição de X_t dados todos os valores anteriores X_0, \dots, X_{t-1} depende apenas de X_{t-1} . Matematicamente,

$$P(x_t \in A | X_0, \dots, X_{t-1}) = P(x_t \in A | X_{t-1})$$

para qualquer subconjunto de A . Os métodos MCMC requerem ainda que a cadeia seja,

1. homogênea, isto é, as probabilidades de transição de um estado para outro são invariantes;
2. irredutível, isto é, cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações;
3. aperiódica, ou seja, não haja estados absorventes.

e os algoritmos que serão vistos aqui satisfazem a estas condições.

Suponha que uma distribuição $\pi(x), x \in \mathbb{R}^d$ seja conhecida a menos de uma constante multiplicativa porém complexa o bastante para não ser possível obter uma amostra diretamente. Dadas as realizações $\{X^{(t)}, t = 0, 1, \dots\}$ de uma cadeia de markov que tenha π como distribuição de equilíbrio. Então, sob esta condição têm-se:

$$\begin{aligned} X^{(t)} &\xrightarrow{t \rightarrow \infty} \pi(x) \\ \frac{1}{n} \sum_{t=1}^n g(X_i^{(t)}) &\xrightarrow{t \rightarrow \infty} \mathbb{E}_{\pi}(g(X_i)) \end{aligned}$$

Ou seja, embora a cadeia seja por definição dependente, a média aritmética dos valores da cadeia é um estimador consistente da média teórica.

Uma questão importante e de ordem prática é como os valores iniciais influenciam no comportamento da cadeia. A idéia é que conforme o número de iterações aumenta, a cadeia gradualmente esquece os valores iniciais e eventualmente converge para uma distribuição de equilíbrio. Assim, em aplicações práticas é comum que as iterações iniciais sejam descartadas, como se formassem uma amostra de aquecimento.

4.3.2 Algoritmo de Metropolis-Hastings

Os algoritmos de Metropolis-Hastings usam a idéia de que um valor é gerado de uma distribuição auxiliar e aceito com uma dada probabilidade. Este mecanismo de

correção garante a convergência da cadeia para a distribuição de equilíbrio, que neste caso é a distribuição a posteriori.

Suponha que a cadeia esteja no estado θ e que o valor de θ' seja gerado de uma distribuição proposta $q(\cdot|\theta)$. Esta distribuição pode depender do estado atual da cadeia, como por exemplo, $q(\cdot|\theta)$ poderia ser uma distribuição normal centrada em θ . Então, o novo valor θ' é aceito com probabilidade

$$\alpha(\theta, \theta') = \min \left(1, \frac{\pi(\theta')q(\theta|\theta')}{\pi(\theta)q(\theta'|\theta)} \right) \quad (4.4)$$

onde π é um distribuição de interesse.

Uma característica importante é que só precisamos conhecer π parcialmente, isto é, a menos de uma constante já que neste caso a probabilidade (4.3.2) não se altera. Isto é fundamental em aplicações Bayesianas onde não conhecemos completamente a posteriori. Nota-se também que a cadeia pode permanecer no mesmo estado por muitas iterações e que na prática costuma-se monitorar isto calculando a porcentagem média de iterações para as quais novos valores são aceitos.

Em termos práticos, o algoritmo de Metropolis-Hastings pode ser especificado pelos seguintes passos:

1. Inicialize o contador de iterações $t = 0$ e especifique um valor inicial $\theta^{(0)}$;
2. Gere um novo valor θ' da distribuição $q(\Delta|\theta)$;
3. Calcule a probabilidade de aceitação $\alpha(\theta, \theta')$ e gere $U \sim U(0, 1)$;
4. Se $u \leq \alpha$; então aceite o novo valor e faça $\theta^{(t+1)} = \theta'$, caso contrário rejeite e faça $\theta^{(t+1)} = \theta$
5. Incremente o contador de t para $t + 1$ e volte ao passo 2.

Embora a distribuição proposta possa ser escolhida arbitrariamente, na prática deve-se tomar alguns cuidados para garantir a eficiência do algoritmo. Em aplicações Bayesianas a distribuição de interesse é a própria posteriori, ou seja, $\pi = p(\theta|x)$ e a

probabilidade de aceitação assume a seguinte forma:

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

4.3.3 Amostrador de Gibbs

No amostrador de Gibbs a cadeia irá sempre se mover para um novo valor, isto é, não existe mecanismo de aceitação-rejeição. As transições de um estado para outro são feitas de acordo com as distribuições condicionais completas $\pi(\theta_i, \theta_{-i})$, onde $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)'$.

Em geral, cada uma das componentes θ_i pode ser unidimensional ou multidimensional. Portanto, a distribuição condicional completa é a distribuição da i -ésima componente de θ condicionada em todas as outras componentes. Ela é obtida a partir da seguinte distribuição conjunta

$$\pi(\theta_i, \theta_{-i}) = \frac{\pi(\theta)}{\int \pi(\theta) d\theta_i}$$

Assim, para obter a distribuição condicional completa de x_i basta pegar os termos da distribuição conjunta que não dependem de x_i .

O amostrador de Gibbs é definido pelo seguinte esquema:

1. inicialize o contador de iterações da cadeia com $t = 0$;
2. especifique valores iniciais $\theta^0 = (\theta_1^0, \dots, \theta_d^0)'$;
3. obtenha um novo valor de $\theta^{(t)}$ a partir de $\theta^{(t-1)}$ através da geração sucessiva de valores

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

4. incremente o contador de t para $t + 1$ e retorne ao passo 2 até obter convergência.

Assim, cada iteração se completa após d movimentos ao longo dos eixos coordenados das componentes de θ . Após a convergência, os valores resultantes formam uma amostra de $\pi(\theta)$. Cabe salientar que mesmo em problemas de grandes dimensões todas simulações podem ser univariadas, o que geralmente é uma vantagem computacional.

Vale notar também, que o amostrador de Gibbs é um caso especial do algoritmo de Metropolis-Hastings, no qual os elementos de θ são utilizados um de cada vez, ou em blocos, tomando a distribuição condicional completa como proposta e a probabilidade de aceitação igua a 1.

4.4 Abordagem Bayesiana em Modelos ARCH

Considerando uma trajetória observada dos retornos $Z = (z_1, \dots, z_T)$ a abordagem bayesiana do problema de inferência dos parâmetros do modelo ARCH(p) inicia-se combinando-se a função de verossimilhança dessa trajetória $L(\alpha|Z)$ com a densidade a priori para os parâmetros $\pi_0(\alpha)$ resultando na densidade a posteriori dada por:

$$\pi(\alpha|Z) \propto L(\alpha|Z)\pi_0(\alpha)$$

No caso do modelo ARCH(p), os parâmetros estão restritos às condições: $\alpha_0 > 0$, $\alpha_j \geq 0$, $j = 1, \dots, p$ e $\sum_{j=1}^p \alpha_j < 1$ para assegurar covariância estacionária. Estas restrições nos permite, na abordagem bayesiana, propor densidades a priori uniformes para esses parâmetros, por exemplo, podemos propor densidades a priori para $\alpha_j \sim U(0, 1)$, $j = 1, \dots, p$ uma vez que todos os α_j 's devem ser menores que um. Também podemos considerar $\log(\alpha_0) \sim N(0, 1)$.

Uma transformação que mapeia os domínios das densidades a priori uniformes $U(a_j, b_j)$ em um domínio $(-\infty, +\infty)$ pode apresentar vantagens quando se pretende usar algoritmos de MCMC, onde em nosso caso temos: $a_j = 0$ e $b_j = 1$ para todo j .

Assim, optamos por uma reparametrização dada por:

$$\phi_j = \log \left(\frac{\alpha_j - a_j}{b_j - \alpha_j} \right), \quad j = 0, 1, \dots, p$$

Com essa reparametrização a função de verossimilhança é escrita com a volatilidade h_t reescrita com os parâmetros α_j devidamente transformados em ϕ_j , ou seja,

$$L(\varphi|Z) \propto \prod_{t=p+1}^T \left(\frac{1}{h_t} \right)^2 \exp \left\{ \frac{-z_t^2}{2h_t} \right\}$$

onde $\varphi = (\phi_0, \phi_1, \dots, \phi_p)'$, e as densidades a priori são definidas como $(\phi_j) \sim \text{Normal}(0, \sigma_j^2)$, $j = 0, 1, \dots, p$. Com isto, podemos escrever a densidade a posteriori como:

$$\pi(\varphi|Z) \propto L(\varphi|Z) \prod_{j=0}^p \pi_j(\phi_j)$$

A avaliação das densidades condicionais a posteriori pode ser feita por meio de métodos de simulação de Monte Carlo em Cadeia de Markov usando algoritmo Metropolis-Hastings [7], onde considerando a posteriori conjunta obtemos as seguintes densidades condicionais a posteriori para φ_j , $j = 0, 1, \dots, p$:

$$\Phi_j(\phi_j|\varphi_{-j}, Z) \propto \psi(\varphi|Z) \pi_j(\phi_j)$$

O uso de prioris informativas como núcleo do algoritmo Metropolis-Hastings é apropriado sempre que tais densidades forem adequadamente escolhidas e o domínio dessas densidades envolverem o domínio da densidade a posteriori, como é o caso dos modelos ARCH(p). Desta forma, o algoritmo M-H pode ser implementado com os seguintes passos conforme [1].

4.4.1 MCMC em Modelos ARCH

1. faça $i = 1$ e forneça um valor inicial arbitrário $\varphi^{(0)} = (\phi_j^{(0)})$, $j = 0, 1, \dots, p$;
2. gere um novo valor β a partir da densidade $\pi(\cdot)$

3. calcule a probabilidade de aceitação do novo valor gerado β :

$$\lambda(\phi_j^{(i-1)}, \beta) = \min \left\{ 1, \frac{\psi(\beta)}{\psi(\phi_j^{(i-1)})} \right\} \quad \psi(\phi_j^{(i-1)}) \neq 0$$

4. Gere uma variável aleatória uniforme $u \sim U(0, 1)$

5. Faça $\phi_j^{(i)} = \beta$ se $u < \lambda$ e $\phi_j^{(i)} = \phi_j^{(i-1)}$ caso contrário.

6. Repetir 2 a 5 para $j = 0, \dots, p$

7. Faça $i \leftarrow i + 1$ e volte ao passo 2 até a convergência.

Após a convergência os parâmetros originais α_j podem ser recuperados pela transformação inversa:

$$\alpha_j = \frac{b_j e^{\phi_j} + a_j}{1 + e^{\phi_j}}.$$

No algoritmo Metropolis-Hastings descrito acima utilizamos $\alpha_j = 0$ e $b_j = 1$.

O problema de inferência bayesiana pode ser expresso como a avaliação da esperança de uma função de interesse $g(\alpha)$ sob a densidade a posteriori $\pi(\alpha|Z)$:

$$E[g(\alpha)] = \int_{\alpha} g(\alpha) \pi(\alpha|Z) d\alpha$$

A solução da integral múltipla acima não pode ser calculada analiticamente no presente caso. Uma aproximação dada pelo Método de Monte Carlo pode ser obtida sempre que $E[g(\alpha)]$ existe, cuja precisão numérica pode ser avaliada. Para detalhes desse método ver [14].

Assim, seja $\{\alpha\}$ uma sequência *i.i.d.* de um vetor aleatório $(p + 1)$ dimensional com densidade conjunta $\pi(\alpha|Z)$. Numericamente aproximamos $\bar{g} = E[g(\alpha)]$ por:

$$\bar{g}_M = \frac{1}{M} \sum_{i=1}^M g(\alpha)$$

Sob condições regulares para a aplicação do Teorema Central do Limite, temos

que $\bar{g}_M \rightarrow E[g(\alpha)]$ e pode ser mostrado que

$$M^{1/2}(\bar{g}_M - \bar{g}) \rightarrow N(0, \sigma^2)$$

onde $\sigma^2 = E[g(\alpha) - \bar{g}]$.

Se $\hat{\sigma}_M^2 = \frac{1}{M-1} \sum_{i=1}^M (g(\alpha^{(i)}) - \bar{g}_M)^2$, então $\hat{\sigma}_M^2 \rightarrow \sigma^2$. Nos referimos a $\hat{\sigma}_M$ como o erro padrão numérico. Então, a precisão numérica pode ser assegurada sob a suposição $\bar{g}_M \sim N(\bar{g}, \hat{\sigma}_M^2)$ quando M é grande.

Um resumo das principais estatísticas dos parâmetros α_j , $j = 0, \dots, p$ podem ser obtidas pelo método de Monte Carlo dado por:

$$E[g(\alpha)] = \frac{1}{M} \sum_{i=1}^M g(\alpha)$$

onde M representa o tamanho da cadeia gerada com o algoritmo Metropolis-Hastings.

4.5 Estendendo para Modelos GARCH

Para a análise Bayesiana no modelo GARCH(r,s) a inclusão de um novo vetor de parâmetros correspondente a ordem s implica em adicionar novas densidades a priori para esta ordem, logicamente como já visto, respeitando as restrições para o vetor de parâmetros. Seja $\theta = (\alpha_0; \alpha_1, \dots, \alpha_r; \beta_1, \dots, \beta_s)$ o vetor de parâmetros restritos à:

$$\begin{aligned} \alpha_0 &> 0 \\ \alpha_i &\geq 0, \quad i = 1, \dots, r \\ \beta_j &\geq 0, \quad j = 1, \dots, s \\ \sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j &< 1. \end{aligned}$$

Levando em conta as restrições dos parâmetros α_0 , α_i e β_j definimos a densidade à priori para $\log(\alpha_0) \sim N(0, 1)$ e para α_i e $\beta_j \sim U(0, 1)$, onde a densidade à posteriori

pode ser calculada da seguinte forma:

$$\pi(\theta|Z) \propto L(\theta|Z)\pi_0(\theta)$$

A inclusão desse novo vetor de parâmetros é a única diferença com relação às definições apresentadas para os modelos ARCH e dessa forma podemos implementar o algoritmo M-H para os modelos GARCH(r, s) com os seguintes passos:

4.5.1 MCMC em modelos GARCH

Assumindo $r = 1$ e $s = 1$ como ordens de um modelo GARCH, θ uma matriz ($n \times p + q + 1$) e n sendo o número de iterações o algoritmo fica da seguinte forma:

1. Faça $i = 1$ e forneça valores arbitrários para $\theta_{i,j} = [\theta_{i,1}, \theta_{i,2}, \theta_{i,3}]$.
2. Gere novos valores $\theta'_{i,j=1,2,3}$ a partir das densidades a priori $\pi_0(\theta)$;
3. Verifique as restrições do modelo para os valores gerados;

Se não satisfeitas as restrições atribua à $\theta'_{i,j=1,2,3}$ os valores escolhidos em (1), faça $i = i + 1$ e volte ao passo (2);

4. Calcule a probabilidade de aceitação dos novos valores gerados:

$$\lambda(\theta_{i,j}^{(i-1)}, \theta'_{i,j}) = \min(1, p)$$

onde,

$$p = \left(\frac{L(\theta'|Z)\pi_0(\theta')}{L(\theta^{(i-1)}|Z)\pi_0(\theta^{(i-1)})} \right)$$

$$p = \left(\frac{\pi(\theta')}{\pi(\theta^{(i-1)})} \right)$$

5. Gere uma amostra $u \sim U(0, 1)$
6. Verifique a condição $u < \lambda(\theta_{i,j}^{(i-1)}, \theta'_{i,j})$

Se não satisfeita a condição atribua à $\theta'_{i,j=1,2,3}$ os últimos valores gerados em (2) e aceitos em (3), faça $i = i + 1$ e volte ao passo (2);

7. Atribua à $\theta'_{i,j=1,2,3}$ os valores gerados em (2);
8. Faça $i = i + 1$;
9. Repetir (2) a (7) para $i = 1, \dots, n$.

Utilizamos para a aplicação desse algoritmo $n = 10.000$, sendo que as primeiras 5.000 simulações geradas foram descartadas, aproveitando-se apenas as últimas 5.000, fato este que se deve para não levar em consideração a aleatoriedade gerada pelas primeiras 5.000, levando à convergência do algoritmo às distribuições à posteriori.

5 *Aplicações*

Neste capítulo mostraremos as aplicações dos métodos bayesianos e a comparação de modelos da família GARCH para as seguintes séries inanceiras: Índice da bolsa de valores de São Paulo (Ibovespa), Preço do Petróleo tipo Brent e Preço do Petróleo tipo WTI.

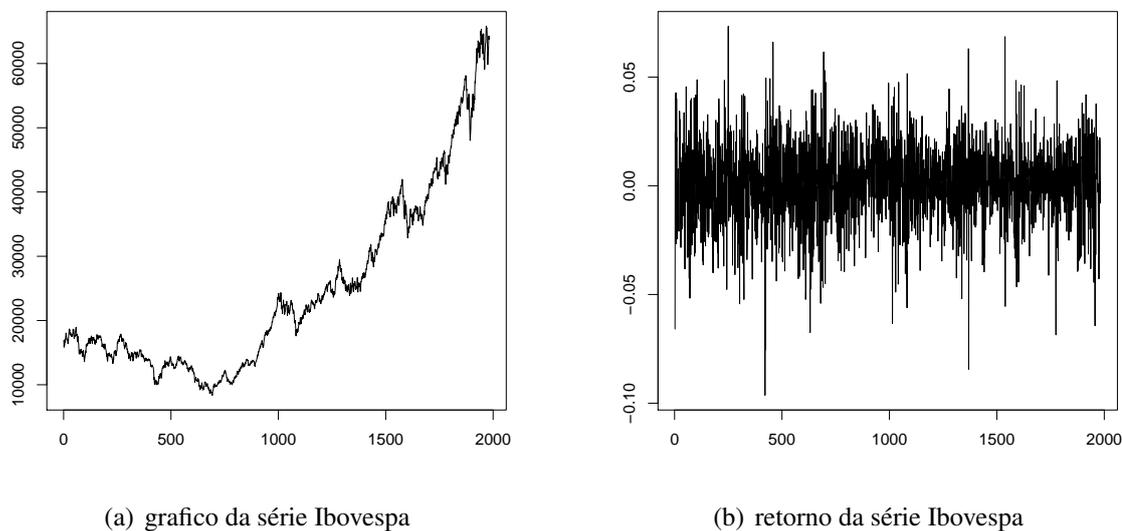
5.1 **Descrição dos Dados**

Três séries financeiras foram utilizadas no presente trabalho, as quais serão descritas nas subseções seguintes.

Índice Bovespa

O Índice Bovespa (Ibovespa) é o mais importante indicador do desempenho médio das cotações das ações negociadas na Bolsa de Valores de São Paulo. É formado pelas ações com maior volume negociado nos últimos meses. Os dados utilizados no presente são referentes ao valor do fechamento diário em pontos do Índice da bolsa de valores de São Paulo (IBOVESPA) no período de 03/02/2000 à 27/12/2007, os quais foram publicados pela Gazeta Mercantil. Estes dados podem ser obtidos através do site www.ipeadata.gov.br.

Figura 5.1: Série do Índice BOVESPA

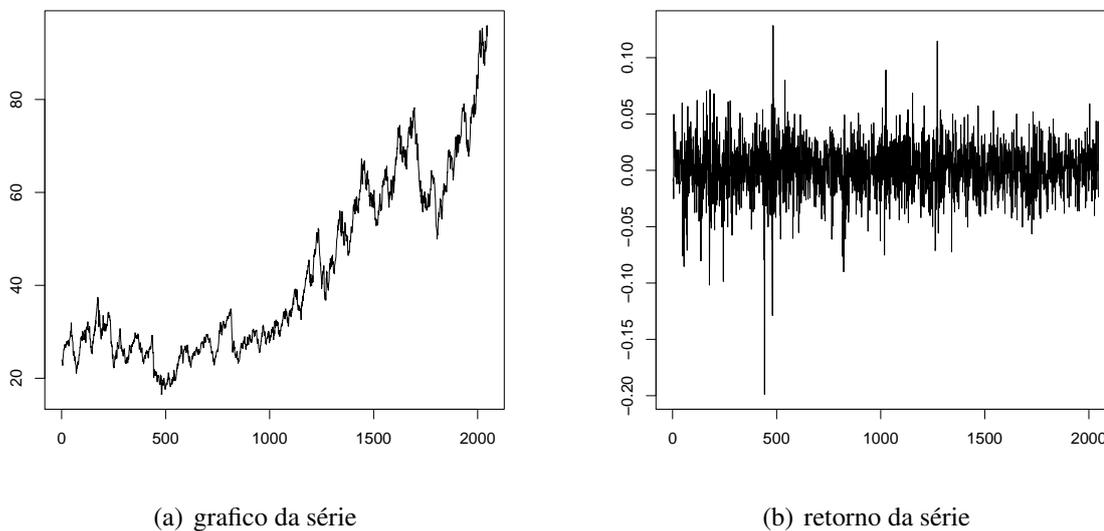


Na figura 5.1 (a) temos a série econômica do Índice Bovespa da Bolsa de São Paulo, e em 5.1 (b) os respectivos retornos da série.

Valor do Petróleo tipo Brent

O barril tipo Brent se refere ao óleo produzido no mar do Norte (Europa), sendo negociado em Londres e servindo de referência para os mercados de derivados da Europa e Ásia. Os dados utilizados são referentes ao valor diário em US\$ do barril de petróleo do tipo Brent no período de 04/01/00 à 31/12/07. Estes dados podem ser obtidos através do site www.ipeadata.gov.br.

Figura 5.2: Série do Preço do Petróleo tipo BRENT

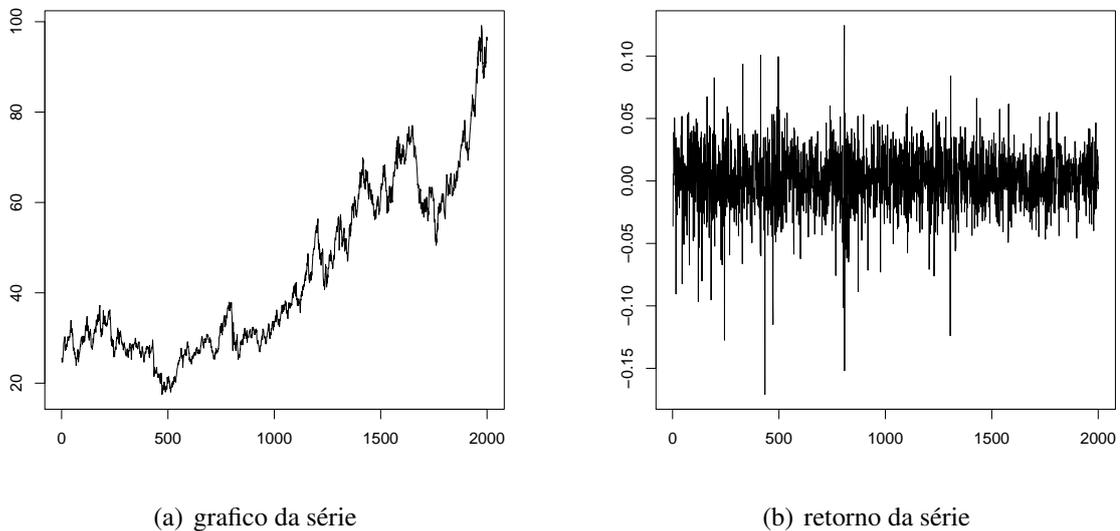


A figura 5.2 (a) apresenta a série do valor diário em US\$ do barril de petróleo do tipo Brent, e em 5.2 (b) seus respectivos retornos.

Valor do Petróleo tipo WTI

WTI é a sigla de West Texas Intermediate. A região do West Texas concentra a exploração de petróleo nos EUA. Este tipo de barril é negociado em Nova York e serve de referência para os mercados de derivados dos EUA. Os dados utilizados são referentes ao o valor diário em US\$ do preço do barril de petróleo do tipo WTI no período de 04/01/00 à 31/12/07, podendo ser obtido em www.ipeadata.gov.br.

Figura 5.3: Série do Preço do Petróleo tipo WTI



A figura 5.3 (a) apresenta a série do valor diário em US\$ do preço do barril de petróleo do tipo WTI, e em 5.3 (b) seus respectivos retornos.

5.2 Pacote Estatístico

Todas as análises deste estudo foram realizadas utilizando-se o pacote estatístico R [23], que é um sistema para computação estatística e gráfica.

Cabe primeiramente comentar que o R é um Software Livre (livre no sentido de liberdade) distribuído sob a Licença Pública Geral (<http://www.r-project.org>) e pode ser livremente copiado e distribuído entre usuários, bem como pode ser instalado em diversos computadores livremente. Isto contrasta com pacotes comerciais que têm licenças altamente restritivas e não permitem que sejam feitas cópias ou que seja instalado em mais de um computador sem a devida licença.

Segundo, sendo um Software Livre, os códigos fontes do R estão disponíveis e atualmente são gerenciados por um grupo chamado o *Core Development Team* (<http://r-project.org/contributors>). A vantagem de ter o código aberto é que

falhas podem ser detectadas e corrigidas rapidamente e atualizações para Softwares Livres podem ser disponibilizadas em uma questão de dias, dependendo pesadamente da participação dos usuários. O R fornece um interface de entrada por linha de comando (ELC), onde todos os comandos são digitados e o mouse é pouco usado. Pode parecer "antigo" ou até "pobre em recursos visuais", mas aí há o seu melhor recurso, a sua flexibilidade. Para os usuários, a linguagem da R se torna clara e simples e a flexibilidade da ELC permite que uns poucos comandos simples sejam juntados para criar funções poderosas. Além disso, a transparência das funções e a entrada de dados é altamente didática. O usuário é sempre consciente do que foi pedido através da ELC. Isso contraste com outros pacotes em que uma interface bonita e sofisticada pode esconder a dinâmica dos cálculos, e potencialmente pode esconder erros.

Por fim, o R é disponível para muitas plataformas incluindo Unix, Linux, Macintosh e Windows. Embora seja possível baixar e compilar os códigos fontes para instalá-lo em seu sistema, a maioria dos usuários optam para a via mais fácil de instalar através de arquivos binários ou executáveis. Ele provê, dentre outras coisas, uma linguagem de programação, ferramentas gráficas de alto nível, interface com outras linguagens de programação (como Fortran e C++), além de ferramentas para depuração. Este é o programa utilizado no curso de Estatística da Universidade Federal do Paraná - UFPR.

Vários pacotes do programa R foram utilizados no presente trabalho, sendo que os pacotes: `tseries`, `fSeries`, `fGarch` e `Rmetrics` foram utilizados para obtermos resultados obtidos via método da máxima verossimilhança. Um outro programa utilizado para a análise clássica foi o `OxMetrics`, o qual foi com sua interface aproveitado dentro do próprio R. Este programa serviu para obtermos estimativas de parâmetros cujos modelos não estão implantados nos pacotes do R, tais quais: `igarch`, `egarch`, por exemplo. O `OxMetrics` é um software utilizado para estimação e previsão de Modelos GARCH e suas extensões, onde o mesmo pode ser obtido através do site: <http://www.core.ucl.ac.be/~laurent/GARCH/site/garch42/default.htm>.

Além destes pacotes, utilizamos também o pacote CODA [22] que é utilizado para obter resumos de análises bayesianas. Agora, toda estimação bayesiana utilizada no presente trabalho foi obtida com o desenvolvimento das funções: `desc`, `lik.garch`, `H`,

dens e MH, todas elas inseridas no apêndice.

5.3 Resultados

Para compararmos a abordagem "clássica", método de máxima verossimilhança, com a abordagem bayesiana via cadeias de Markov (MCMC), para estimação dos parâmetros α_0 , α_1 e β_1 de um modelo $GARCH(1, 1)$ com $\varepsilon_t \sim N(0, h_t)$ utilizamos três séries financeiras conhecidas, tais quais, índice bovespa, preço do petróleo tipo Brent e WTI, os quais já foram descritos em 5.1.

Então, como visto anteriormente, para um modelo $GARCH(1, 1)$ com $\varepsilon_t \sim N(0, h_t)$ temos:

$$\begin{aligned} X_t &= \sqrt{h_t} \varepsilon_t \\ h_t &= \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1} \end{aligned}$$

onde iremos estimar os parâmetros α_0 , α_1 e β_1 .

Primeiramente temos as estimativas obtidas pelo método da máxima verossimilhança através do pacote `tseries` dos parâmetros α_0 , α_1 e β_1 de um modelo $GARCH(1, 1)$.

Tabela 5.1: Estimativas dos Parâmetros de um $GARCH(1,1)$ utilizando o pacote `tseries`

Coef.	Ibovespa	PetroBRENT	PetroWTI
α_0	0.00002	0.00001	0.00002
α_1	0.06034	0.03587	0.05139
β_1	0.89467	0.93904	0.91809

Em seguida apresentamos também as estimativas dos parâmetros α_0 , α_1 e β_1 de um modelo $GARCH(1, 1)$, via método de máxima verossimilhança. Porém, estas estimativas foram obtidas através de uma função de nossa autoria, chamada `lik.garch`, função esta que foi desenvolvida com a intenção de comparar a estimação de parâmetros de modelos disponíveis nos pacotes, do programa R, para que posteriormente pudesse-

mos desenvolvê-la para as extensões dos modelos GARCH, as quais ainda não estão implantadas nos pacotes.

Tabela 5.2: Estimativas dos Parâmetros de um GARCH(1,1) utilizando a função `lik.garch`

Coef.	Ibovespa	PetroBRENT	PetroWTI
α_0	0.00008	0.00014	0.00002
α_1	0.18940	0.54836	0.04983
β_1	0.59274	0.45164	0.92080

Então, comparando as duas maneiras de se estimar, percebe-se que as estimativas dos parâmetros α_0 , α_1 e β_1 são similares, porém não exatamente as mesmas, fato que se deve a possíveis problemas numéricos do algoritmo *optim* que foi utilizado para maximizar a função `lik.garch`, e também pelo fato de que um modelo GARCH não ser a alternativa mais adequada para explicar a variação destas séries, o que por exemplo, pode ser visto na figura 3.1, onde percebemos assimetria dos retornos, descaracterizando o modelo GARCH como o mais indicado para estas séries. No entanto, todos os parâmetros estimados atendem suas restrições características do modelo $GARCH(1,1)$, ou seja, $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ e $\alpha_1 + \beta_1 < 1$.

Na tabela abaixo veremos as estimativas obtidas através do método Bayesiano via Cadeias de Markov, com o algoritmo de Metropolis-Hastings.

Tabela 5.3: Estimativas dos Parâmetros de um GARCH(1,1) utilizando o algoritmo Metropolis - Hastings

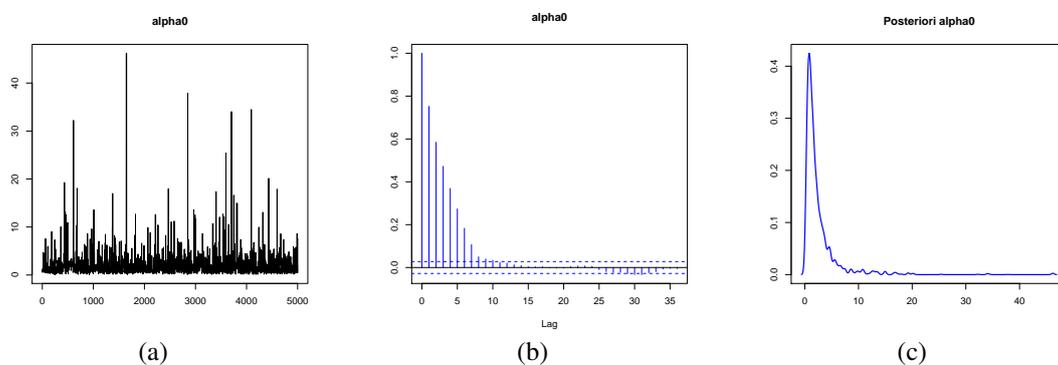
Coef.	Ibovespa	PetroBRENT	PetroWTI
α_0	0.80640	0.81390	0.77260
α_1	0.29850	0.31362	0.29225
β_1	0.40828	0.39057	0.41211

Nota-se novamente que as restrições dos parâmetros são atendidas, onde as estimativas de α_0 diferem das obtidas em métodos anteriores pelo fato da distribuição à priori do parâmetro ser $\log -N(0, 1)$. Já em consideração aos parâmetros α_1 e β_1 os valores apresentados diferem em pouco dos obtidos via método da Máxima Verossimilhança, o que se deve ao fato de o modelo GARCH não ser o modelo mais adequado a série, como já explicitado anteriormente, além de que houve uma certa dificuldade de convergência do algoritmo para as distribuições à posteriori destes parâmetros e ao fato de ter-se estabelecido como distribuição à priori uma $U(0, 1)$.

Abaixo segue resumos gráficos de cada série dos resultados obtidos via MCMC utilizando o algoritmo de Metropolis-Hastings.

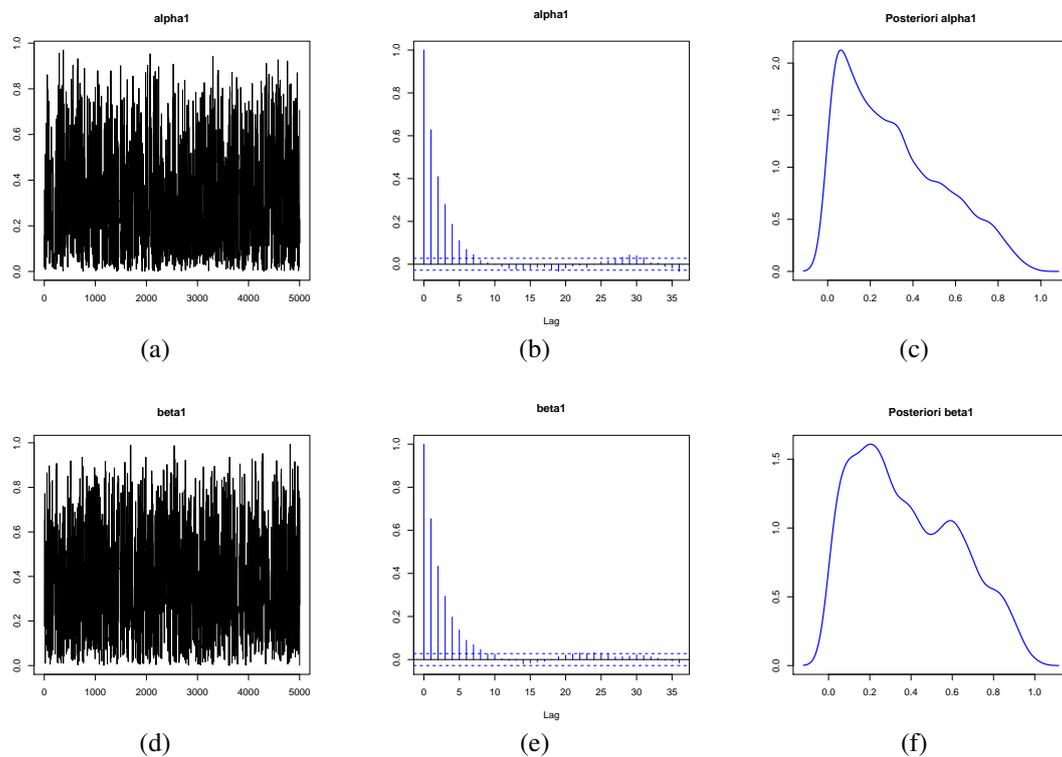
Primeiramente, vamos abordar a série do índice Bovespa, onde analisaremos separadamente cada parâmetro. Para α_0 , em 5.4 (a) podemos observar os valores simulados para o parâmetro nas 5000 iterações. Já em 5.4 (b) são apresentadas as autocorrelações dos valores simulados, sendo estes uma cadeia de Markov espera-se que as autocorrelações tornem-se não significativas em um curto número de defasagens, o que neste caso apresenta um comportamento aceitável, porém não o mais adequado pois as autocorrelações decaem lentamente até $lag = 10$. Em 5.4 (c) temos a distribuição a posteriori do parâmetro α_0 . Neste gráfico percebe-se assim como em 5.4 (a) o valor mais provável para o parâmetro, o que se deve ao fato deste valor ser a moda da distribuição à posteriori.

Figura 5.4: Estimação Bayesiana da série IBOVESPA para o parâmetro α_0



Agora, para os parâmetros α_1 e β_1 da série Ibovespa obtemos valores muito próximos, porém bem distante das estimativas clássicas, pois aqui a restrição é um pouco mais rígida que para o parâmetro α_0 devido a dependência de α_1 com β_1 cuja soma deve ser menor que 1, além de se notar dificuldade de convergência e o modelo não ser o mais indicado para a mencionada série, como dito anteriormente.

Figura 5.5: Estimação Bayesiana da série IBOVESPA para o parâmetro α_1 e β_1



Já na estimação dos parâmetros da série preço do petróleo tipo Brent, obtemos conclusões similares às já obtidas para a série do índice Bovespa, onde para o parâmetro α_0 a distribuição a posteriori mostra-se aceitável, haja vista a transformação do parâmetro, dando ao mesmo uma distribuição a priori de uma Log-Normal.

Figura 5.6: Estimação Bayesiana da série Petróleo - Brent para o parâmetro α_0

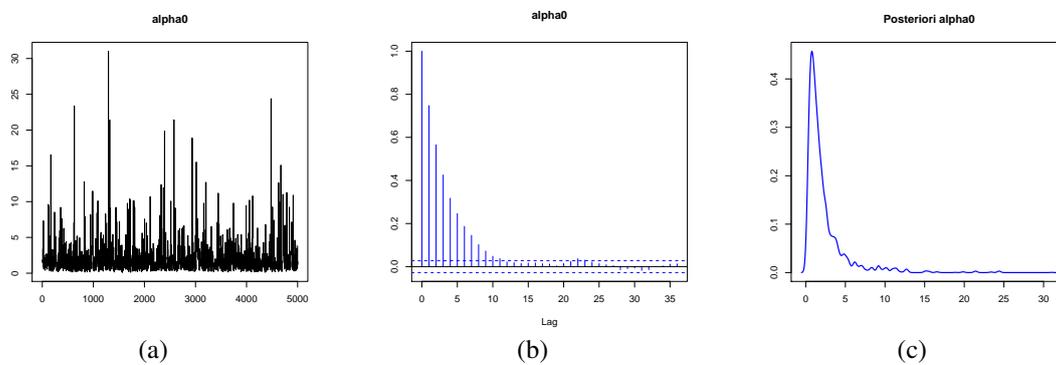
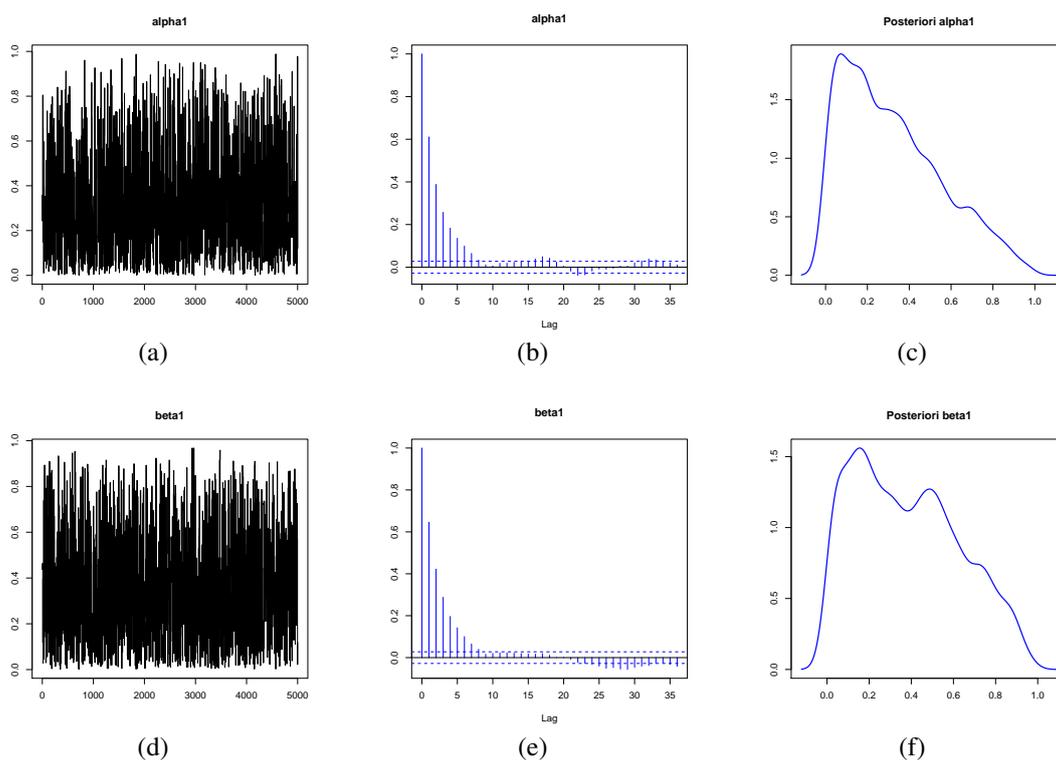


Figura 5.7: Estimação Bayesiana da série Petróleo - Brent para o parâmetro α_1 e β_1

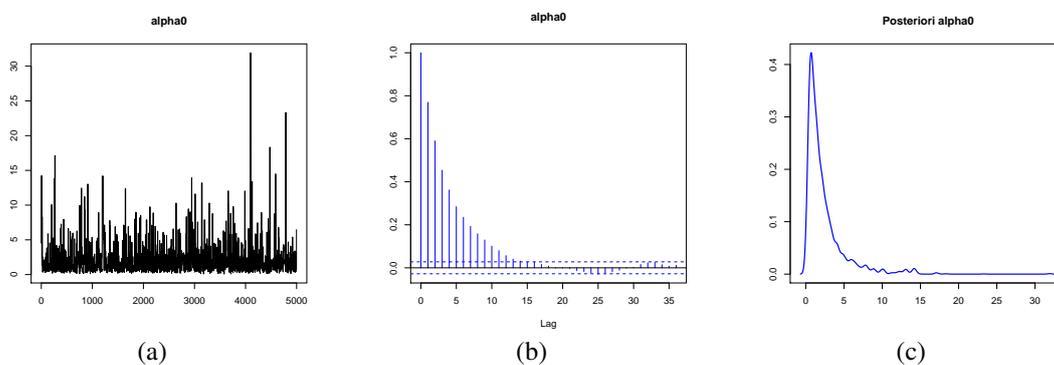


E, para os parâmetros α_1 e β_1 os gráficos apresentados, mostram-nos o caminho do algoritmo até sua convergência das últimas 5000 iterações, bem como o gráfico de

autocorrelação e as distribuições a posteriori dos respectivos parâmetros, onde podemos perceber que ambos os parâmetros atendem as restrições de positividade e estacionaridade do modelo GARCH(1, 1).

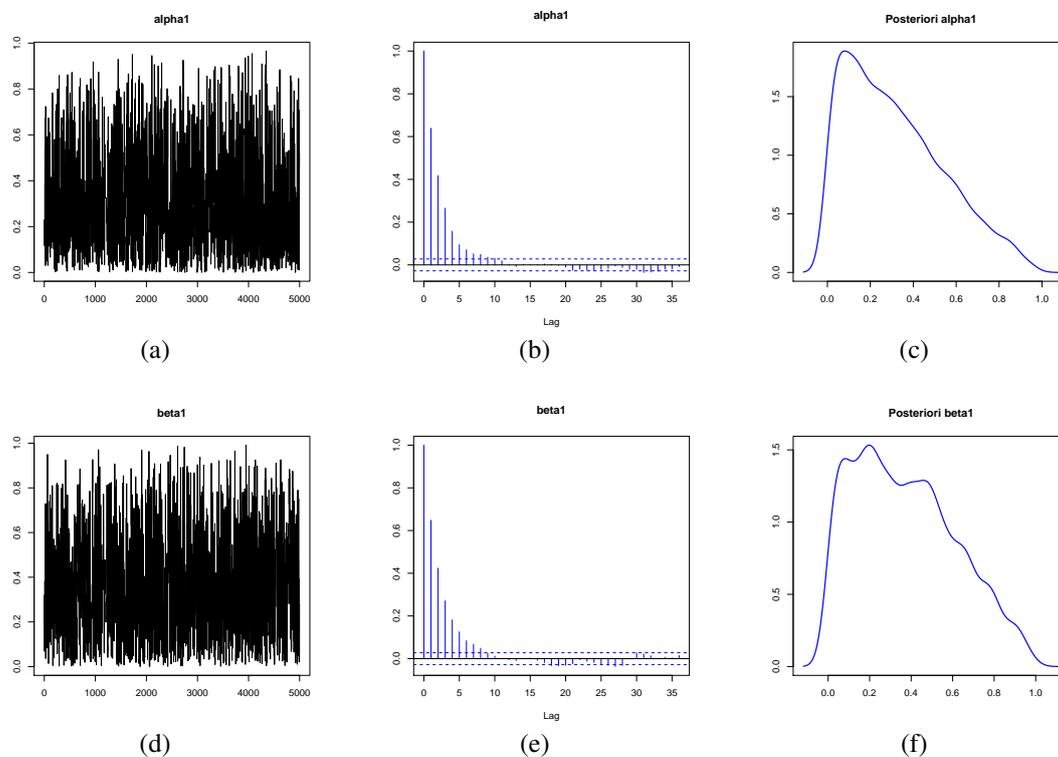
Agora, para a série preço do petróleo tipo WTI obtemos, também, conclusões análogas as já obtidas para as outras duas séries financeiras, onde todos os parâmetros atendem as restrições impostas pelo modelo, além de nos dar informações preciosíssimas a respeito das distribuições a posteriori do parâmetros, o que para inferência Bayesiana é de suma importância.

Figura 5.8: Estimação Bayesiana da série Petróleo - WTI para o parâmetro α_0



Valores similares aos já obtidos para os parâmetros da série preço do petróleo tipo Brent deve-se ao fato de que as duas séries financeiras são afetadas, basicamente, pelos mesmos motivos, o que caracterizaria um estudo multivariado, onde poderíamos analisá-las simultaneamente, haja vista que ambas as séries sofrem impactos similares através das notícias.

Figura 5.9: Estimação Bayesiana da série Petróleo - WTI para o parâmetro α_1 e β_1



Cabe salientar que o modelo GARCH(1,1) foi usado como exemplo, pois apresenta maior facilidade para escrita da função de verossimilhança e definição das priors, por exemplo, porém, o mesmo estudo está sendo estendido para os modelos APARCH, assim como para séries multivariadas e com erros assimétricos.

6 *Conclusões*

Como visto, a heterocedasticidade condicional formada por um processo autoregressivo provou ser de grande utilidade para modelos econométricos, especialmente dos que tratam de variáveis financeiras, onde constata-se que mais de uma centena de artigos já foram publicados com construções em cima do modelo básico, comparações com outros modelos, comentários ou aplicações. Pode-se observar na evolução dos modelos uma tendência de maior realismo ao longo do tempo, e a cada nova sofisticação mais uma das características das séries de finanças passa a ser levada em conta.

O trabalho claramente não é exaustivo, mas apresenta a parte mais substancial da literatura, além de explanar muito bem sobre método Bayesiano e a utilização de simulação de Monte Carlo em Cadeias de Markov (MCMC) utilizando o algoritmo de Metropolis-Hastings para estimação e comparação de um modelo GARCH(1, 1), onde tem-se uma grande vantagem devido podermos levar em consideração o conhecimento sobre o "fenômeno" para estimação dos parâmetros.

Portanto, o presente trabalho apresentou a metodologia de Inferência Bayesiana para estimação e comparação de modelos da família GARCH, como uma alternativa ao método tradicional de estimação por máxima verossimilhança. Foram utilizados os métodos de Monte Carlo via Cadeias de Markov (MCMC) para estimação de um modelo GARCH(1, 1). O mesmo método poderá ser estendido para toda família GARCH bem como para uma análise multivariada, onde várias séries financeiras serão analisadas conjuntamente. Além disso, outras distribuições para o termo de erro podem ser consideradas, tais como a t de Student ou a GED. E, recentemente também surgiram na literatura trabalhos aonde a distribuição dos erros é assimétrica, o que pode ser uma hipótese plausível em séries financeiras. Vale comentar que a estimação via

método Bayesiano está sendo desenvolvida para o modelo APARCH e suas derivações, onde têm-se muita dificuldade para escrever a própria verossimilhança.

Referências Bibliográficas

- [1] G. A. Barretos, S. C. Oliveira, and M. G. Andrade. Estimação bayesiana de parâmetros de modelos arch(p) via simulação de monte carlo em cadeias de markov. *Anais do XV Congresso Brasileiro de Automática*, 2004.
- [2] J. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York., 1985.
- [3] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley, New York., 1994.
- [4] T. Bollerslev. Generalised autoregressive conditional heteroscedasticity. *J. Econometrics*, 51:307–327, 1986.
- [5] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Englewood Cliffs NJ: Prentice-Hall, New York., 1994.
- [6] P. Brockwell and R. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York., 1991.
- [7] E. Chib, S.; Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, 49:327–335, 1995.
- [8] P. A. M. e Célia M. C. Toloí. *Análise de Séries Temporais*. Gold Blücher, São Paulo, 2ª edition, 2006.
- [9] R. S. Ehlers. *Análise de Séries Temporais*. Departamento de Estatística - UFPR, Disponível em www.leg.ufpr.br/~ehlers/notas, 4ª edition, 2008.
- [10] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1007, 1982.
- [11] P. H. Franses and D. van Dijk. *Non-linear time series models in empirical finance*. Cambridge University Press: Cambridge, 2000.
- [12] D. Gamerman. Markov chain monte carlo: Stochastic simulation for bayesian inference. *Texts in Statistical Sciences.*, 1997.
- [13] D. Gamerman and H. Lopes. Markov chain monte carlo: Stochastic simulation for bayesian inference. *Texts in Statistical Sciences.*, 2006.

- [14] J. Geweke. Exact inference in the inequality constrained normal linear regression model. *Journal of Applied Econometric*, 1:127–141, 1986.
- [15] R. e. R. D. Glosten, L. e Jagannathan. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5):1779–1801, 1993.
- [16] C. Granger and P. Newbold. *Forecasting Economic Time Series*. Academic Press, New York., 1986.
- [17] J. D. Hamilton. *Time Series Analysis*. Princeton, New Jersey, 1994.
- [18] M. Higgins and A. Bera. A class of nonlinear arch models. *International Economic Review*, pages 137–158, 1992.
- [19] S. Laurent. Analytical derivatives of the aparch model. *Forthcoming in Computational Economics*, 2003.
- [20] P. A. Morettin. *Econometria Financeira: Um Curso de Séries Temporais Financeiras*. Edgard Bluecher, São Paulo, 2008.
- [21] D. B. Nelson. Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, 59:347–370, 1991.
- [22] M. Plummer, N. Best, K. Cowles, and K. Vines. (coda): Convergence diagnosis and output analysis for (mcmc). *R News*, 6:7–11, March 2006.
- [23] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. ISBN 3-900051-07-0.
- [24] C. P. Robert and G. Casella. Monte carlo statistical methods. *Springer-Verlag, New York.*, 1999.
- [25] W. Schwert. Stock volatility and the crash of 87. *Review os Financial Studies.*, 3:77–102, 1990.
- [26] S. Taylor. Modeling financial time series. *Jornal of Economic Dynamics and Control.*, 1986.
- [27] P. Verhoeven and M. McAleer. Fat tails and asymmetry in financial volatility models. Technical Report CIRJE-F-211, Faculty of Economics, University of Tokyo, 2003.
- [28] J. M. Zakoian. Threshold heteroskedasticity models. *Jornal of Economic Dynamics and Control.*, 15:931–995, 1994.

APÊNDICE A – Funções do R

```
# FUNCAO DESCRITIVA #
## y = série
desc = function(y, fig=F)
{
  require(graphics)
  log.ret = diff(log(y))
  coord = range(quantile(log.ret))

  # JANELA 1 #
  X11()
  par(mfrow=c(2,2))
  plot(y, main="", xlab="", ylab="")
  plot(log.ret, main="", xlab="", ylab="")
  hist.ret = hist(log.ret,freq=F, main="", xlab="",
                 ylim=c(0,25), xlim=c(-0.1,0.1))
  x = hist.ret$breaks
  curve(dnorm(x,0,sd(log.ret)),add=T, col="red")
  qqnorm(log.ret, main="")
  qqline(log.ret)

  pv1 = round(shapiro.test(log.ret)$p.value,4)
  text(-1, coord[2], "Shapiro-Wilk")
  text(-1, coord[2]-0.02, paste('p-valor = ',pv1))
  pv2 = round(jarque.bera.test(log.ret)$p.value,4)
```

```

text(1, coord[1]+0.02, "Jarque - Bera")
text(1,coord[1], paste('p-valor = ',pv2))

# JANELA 2 #
X11()
par(mfrow=c(2,2))
acf.ret = acf(log.ret, xlab="", ylab="", main="")
pacf.ret = pacf(log.ret, xlab="", ylab="", main="")
acf.ret2 = acf(log.ret^2, xlab="", ylab="", main="")
pacf.ret2 = pacf(log.ret^2, xlab="", ylab="", main="")

if(fig){ # GRÁFICOS EM .pdf NO DIRETÓRIO ATUAL #
  pdf("desc1.pdf");
  plot(y, main="", xlab="", ylab=""); dev.off()
  pdf("desc2.pdf");
  plot(log.ret, main="", xlab="", ylab=""); dev.off()
  pdf("desc3.pdf"); plot(hist.ret,freq=F, main="", xlab="",
                        ylab="", ylim=c(0,25), xlim=c(-0.1,0.1))
  curve(dnorm(x,0,sd(log.ret)),add=T, col="red")
  dev.off()
  pdf("desc4.pdf"); qqnorm(log.ret, main="", xlab="", ylab="")
  qqline(log.ret); dev.off()
  pdf("desc5.pdf"); plot(acf.ret); dev.off()
  pdf("desc6.pdf"); plot(pacf.ret); dev.off()
  pdf("desc7.pdf"); plot(acf.ret2); dev.off()
  pdf("desc8.pdf"); plot(pacf.ret2); dev.off()
}
}

# MAXIMIZAÇÃO DA FUNÇÃO DE VEROSSIMILHANÇA USANDO optim #
# Garch(1,1)
# h_t = alpha0 + alpha1 * X_{t-1}^2 + beta1 * h_{t-1}

```

```

# CALCULA verossimilhança N(0,h_t)
## x = Retorno da série #
## theta = vetor de parâmetros #
lik.garch = function(x, theta){
  n = length(x)
  alpha0 = theta[1]
  alpha1 = theta[2]
  beta1 = theta[3]
  if ((alpha0 < 0) | (alpha1 < 0) | (beta1 < 0) | (alpha1 + beta1 >=1)){
    return(-Inf)
  }
  h = rep(0,n)
  h[1] = x[1]^2
  for (i in 2:n) {
    h[i] = alpha0 + alpha1 * x[i-1]^2 + beta1 * h[i-1]
  }
  sum(dnorm(x[2:n], mean=0, sd=sqrt(h[2:n]), log=T))
}

# ESTIMAÇÃO BAYESIANA USANDO MCMC #
# Garch(1,1)
#  $h_t = \alpha_0 + \alpha_1 * X_{t-1}^2 + \beta_1 * h_{t-1}$ 

# CALCULA h_t #
## x = Retorno da série #
## theta = vetor de parâmetros #
H = function(x, theta){
  n = length(x)
  h = rep(0,n)
  h[1] = x[1]^2
  a0 = theta[1]

```

```

a1 = theta[2]
b1 = theta[3]
for( t in 2:n){
  h[t] = a0 + a1 * x[t-1]^2 + b1 * h[t-1]
}
return(h)
}

# CALCULA DENSIDADE N(0,h_t) #
## x = Retorno da série #
## H = função CALCULA h_t
## theta = vetor de parâmetros #
dens = function(x, H, theta){
  x = x
  a0 = theta[1]
  a1 = theta[2]
  b1 = theta[3]
  n = length(x)
  h = rep(0,n)
  h = H(x, theta)
  d = sum(dnorm(x[2:n], 0, sd=sqrt(h[2:n]), log=T))
  return(d)
}

# MCMC - Metropolis - Hastings #
## ret = Retorno da Série
## dens = função CALCULA DENSIDADE N(0,h_t)
## init.garch = valores iniciais do vetor de parâmetros
## n.it = nº de iterações
MH = function(ret, dens, H, init.garch, n.it=10000){
  val = round(n.it/2,0)
  n = n.it

```

```

theta = matrix(NA, nr=n, nc=3)
theta.out = matrix(NA, nr=val, nc=3)
h=rep(0,n)
h[1] = ret[1]^2
theta[1,1] = init.garch[1] # alpha0
theta[1,2] = init.garch[2] # alpha1
theta[1,3] = init.garch[3] # beta1
taxa.ac = 0
taxa.rej = 0
for (i in 2:n){
  cat(i, fill=T, labels='Iteração=')
  a0 = exp(rnorm(1,0,1))
  a01 = abs(log(a0))      # valor proposto alpha0
  a1 = runif(1, 0, 1)    # valor proposto alpha1
  b1 = runif(1, 0, 1)    # valor proposto beta1
  if ( (a1 + b1 < 1) ) {
    cat('condições OK! \n')
    A = ( dens(ret, H, c(a01,a1,b1)) * dnorm(a01,0,1) ) /
          ( dens(ret, H, c(theta[i-1,1], theta[i-1,2], theta[i-1,3])) *
            dnorm(theta[i-1,1],0,1) )
    prob = min(1,A)
    u1 = runif(1)
    if ( u1 < prob ){
      theta[i,1] = a01
      theta[i,2] = a1
      theta[i,3] = b1
      taxa.ac = taxa.ac + 1
      cat('aceito \n')
    }
    next
  }
  else {
    theta[i,1] = theta[i-1,1]
  }
}

```

```
    theta[i,2] = theta[i-1,2]
    theta[i,3] = theta[i-1,3]
    cat('rejeitado prob \n')
    next
  }
}
else {
  theta[i,1] = theta[i-1,1]
  theta[i,2] = theta[i-1,2]
  theta[i,3] = theta[i-1,3]
  cat('rejeitado restrições \n')
  next
  taxa.rej = taxa.rej + 1
}
}
taxa.ac = taxa.ac/n
taxa.rej = taxa.rej/n
theta.out[,] = theta[(val+1):n,] # descartando (n.it/2) iterações
cat("\n");cat(n, fill=T, labels=" TOTAL iterações = ")
cat("\n");cat(val, fill=T, labels=" Iterações válidas = ")
return(list(theta=theta.out,taxa.ac=round(taxa.ac,2),
           taxa.rej=round(taxa.rej,2)))
}
```