



UNIVERSIDADE FEDERAL DE LAVRAS

**MODELAGEM DA VOLATILIDADE DOS ÍNDICES
FINANCEIROS IBOVESPA, DOW JONES E STANDARD
& POORS UTILIZANDO MODELOS DA CLASSE
ARCH**

WASHINGTON SANTOS DA SILVA

6328

1

2003

5851
48074

WASHINGTON SANTOS DA SILVA

**MODELAGEM DA VOLATILIDADE DOS ÍNDICES
FINANCEIROS IBOVESPA, DOW JONES e
STANDARD & POORS UTILIZANDO MODELOS DA
CLASSE ARCH**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Mestrado em Estatística e Experimentação Agropecuária, área de concentração em Teoria Matemática e Métodos Estatísticos, para obtenção do título de "Mestre".

Orientador:

Prof.^a Dr.^a Thelma Sáfyadi

Co-orientador:

Prof. Dr. Luiz Gonzaga de Castro Júnior

BIBLIOTECA CENTRAL - UFLA



55851

Lavras, Minas Gerais

BIBLIOTECA CENTRAL
UFLA
N.º CLAS. 1332.63.28
51K
mod
N.º REGISTRO 0.55851
DATA 21 10 03

2003

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Washington, Santos da Silva

Modelagem da volatilidade dos índices financeiros Ibovespa, Dow Jones e standard & Poors utilizando modelos da classe ARCH / Washington Santos da Silva. -- Lavras : UFLA, 2003.

92 p. : il.

Orientador: Thelma Sáfyadi.

Dissertação (Mestrado) – UFLA.

Bibliografia.

1. Modelo ARCH. 2. FIGARCH. 3. Volatilidade. 4. Ibovespa. 5. Dow Jones.
6. Standard & Poors. I. Universidade Federal de Lavras. II. Título.

CDD-332.6328

WASHINGTON SANTOS DA SILVA

**MODELAGEM DA VOLATILIDADE DOS ÍNDICES
FINANCEIROS IBOVESPA, DOW JONES E
STANDARD & POORS UTILIZANDO MODELOS DA
CLASSE ARCH**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Mestrado em Estatística e Experimentação Agropecuária, área de concentração em Teoria Matemática e Métodos Estatísticos, para obtenção do título de “Mestre”.

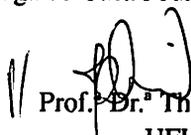
APROVADA em 4 de Fevereiro de 2003

Prof. Dr. Pedro Alberto Morettin

USP

Prof. Dr. Luiz Gonzaga de Castro Júnior

UFLA


Prof. Dr.ª Thelma Sáfyadi
UFLA
(Orientadora)

Lavras, Minas Gerais
2003

*À minha mulher Andréa, meu grande
amor, e ao meu filho Caio, meu grande
amigo,*

DEDICO

AGRADECIMENTOS

À minha mulher Andréa, pela sua dedicação e lealdade, por ter suportado bravamente minha ausência e por ser minha mulher. Ao meu filho Caio, por ser minha fonte de alegria e motivação.

À minha mãe Emília, pelo suporte indispensável, e ao meu pai Amado (*em memória*).

À Professora Thelma, por ter sido uma verdadeira orientadora e sobretudo por sua seriedade, amizade e paciência. Agradeço também ao Professor Luiz Gonzaga pela sua disponibilidade.

Ao Prof. Joel por sua dedicação constante em aperfeiçoar o Curso e à Fundação de Amparo à Pesquisa de Minas Gerais pela concessão de uma bolsa de Estudo.

Aos professores do Mestrado em Estatística, principalmente ao Professor Mário, pelos ensinamentos e pela amizade.

Aos colegas de classe, em especial ao João Marcos e a sua esposa Andréia, pela grande amizade e lealdade.

Aos funcionários do DEX e da biblioteca, pela eficiência e boa vontade.

E a Deus, pelo trabalho concluído

SUMÁRIO

RESUMO	i
ABSTRACT	ii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	4
2.1 Modelos estocásticos de séries temporais	4
2.2 Índices Acionários	5
2.3 ARCH	7
2.4 Regularidades empíricas de séries econômicas	9
2.4.1 Caudas pesadas	9
2.4.2 Agrupamentos de volatilidade	10
2.4.3 Assimetrias	10
2.4.4 Memória longa	11
2.4.5 Hipótese da eficiência do mercado	12
2.4.6 Macroeconomia e volatilidade	14
2.5 ARCH: Descrição dos modelos	14
2.5.1 Modelo ARCH	15

2.5.2	Modelo GARCH	17
2.6	ARCH: Propriedades	20
2.6.1	Estacionariedade e momentos incondicionais	20
2.6.2	Simetria e regularidade	28
2.6.3	Previsão	29
2.6.4	Persistência na variância	32
2.6.5	Relação com processos em tempo contínuo	35
2.7	ARCH: Interpretações e origens	37
2.7.1	Interpretações econômicas	38
2.7.2	Origens	38
2.7.3	ARCH como Modelo não-linear	39
2.8	ARCH: Extensões	41
2.8.1	IGARCH	41
2.8.2	FIGARCH	43
2.8.3	APARCH	48
2.9	ARCH: Estimação	48
2.9.1	Máxima verossimilhança com ε_t gaussiano	49
2.9.2	Quasi-máxima verossimilhança	54
2.9.3	Máxima verossimilhança com ε_t não-gaussiano	56
2.10	ARCH: Testes	59
2.10.1	O Teste ARCH de Engle	59
2.10.2	O Teste BDS	59

3 MATERIAL E MÉTODOS	62
3.1 Dados	62
3.2 Metodologia	64
4 RESULTADOS E DISCUSSÃO	67
4.1 Ibovespa	69
4.2 Dow Jones	72
4.3 Standard & Poors (500)	75
5 CONCLUSÕES	85
REFERÊNCIAS BIBLIOGRÁFICAS	87

RESUMO

SILVA, Washington Santos. **Modelagem da Volatilidade dos Índices Financeiros Ibovespa, Dow Jones e Standard & Poors Utilizando Modelos da Classe ARCH.** Lavras: UFLA, 2003. 95p.(Dissertação - Mestrado em Estatística e Experimentação Agropecuária) *

Examinou-se o processo de volatilidade de três índices financeiros, Ibovespa, Dow Jones e Standard & Poors (500), utilizando o modelo FIGARCH (*Fractionally Integrate GARCH*) e modelos ARCH (*Autoregressive Conditional Heteroskedasticity*) tradicionais. Utilizaram-se quatro distribuições para os retornos: Normal, t-student, t assimétrica e distribuição exponencial generalizada. Os modelos selecionados foram comparados quanto à qualidade do ajuste e acurácia preditiva fora da amostra. Os resultados empíricos mostraram que todas as séries evidenciaram a presença do efeito alavancagem na variância e as densidades com melhor desempenho foram a t-Student e a t-Student assimétrica. Os critérios de seleção de modelos evidenciaram que o modelo FIGARCH com distribuição t-Student assimétrica superou os modelos GARCH tradicionais quanto à qualidade do ajuste e acurácia preditiva no caso dos índices Ibovespa e, com a t-Student, Standard & Poors (500). Para o Dow Jones, um modelo GARCH com distribuição t-Student obteve o melhor desempenho.

* Comitê Orientador: Thelma Sáfiadi - UFLA (Orientadora), Luiz Gonzaga de Castro Júnior.

ABSTRACT

SILVA, Washington Santos. **Modelling the Volatility of the Ibovespa, Dow Jones and Standard & Poors Financial Indexes using the ARCH class of Models.** Lavras: UFLA, 2003. 95p. (Dissertation - Master Program in Statistics and Agricultural Experimentation) *

We examine the volatility process of three financial indexes Ibovespa, Dow Jones and Standard & Poors (500), using the FIGARCH (Fractionally Integrated GARCH) model and traditional ARCH (Autoregressive Conditional Heteroskedasticity) models. Four distributions was taken in to account for the returns; Normal, Student-t, skewed Student-t and the Generalized error distribution(GED). The selected models were compared regarding out-of-sample forecasting accuracy and goodness-of-fit statistics. The empirical results show that all series show signs of the leverage effect for the variance and the best densities overall were Student-t and the skewed Student-t. Regarding the several criterions used, the FIGARCH model with skewed Student-t outperform traditional ARCH models for the Ibovespa, and with a Student-t, for the Standard & Poors (500) indexes. In the case of Dow Jones, the traditional GARCH model with t-Student for the returns was the best model.

*Guidance Committee: Thelma Sáfadi - UFLA (Major Professor), Luiz Gonzaga de Castro Júnior.

Há duas décadas, em econometria aplicada e teórica, as preocupações concernentes à heterocedasticidade ocorriam quase que exclusivamente se os dados sob análise eram do tipo secção cruzada. O exemplo clássico, amplamente encontrado em livros texto, trata da relação entre consumo e renda familiar, quando o conjunto de dados contém dados de famílias com alto poder aquisitivo e de famílias com baixo poder aquisitivo. A mensagem padrão, quando há evidência de heterocedasticidade, é que apesar dos estimadores de mínimos quadrados ordinários serem ainda não viesados, os erros padrão e intervalos de confiança derivados são espúrios, no sentido que podem fornecer uma falsa impressão de precisão.

Em econometria, no que concerne a dados em secção cruzada, o desenvolvimento de métodos que fornecem ‘erros padrão robustos’ reduziu em muito as preocupações relativas à heterocedasticidade. Com um tamanho amostral adequado, pode-se obter uma boa estimativa do erro padrão ainda que verifique-se a presença de heterocedasticidade. Caso a amostra seja pequena, diversas correções têm sido propostas, correções estas que visam não afetar os coeficientes, mas somente corrigir assintoticamente os erros padrão.

A questão da heterocedasticidade é o foco principal da classe de modelos ARCH, introduzida no artigo seminal de Engle (1982). ARCH denota *Autoregressive Conditional Heteroskedasticity*. Ao invés de considerar a heterocedasticidade como um problema a ser corrigido, Robert F. Engle a considerou como um fenômeno a ser modelado, ou melhor, como uma variância a ser modelada. Isto posto, não apenas as deficiências dos estimadores de mínimos quadrados, do ponto de vista dos dados em secção cruzada, são corrigidas, como uma predição da variân-

cia de cada termo do erro é calculada. Tal previsão é de grande interesse em macroeconomia e, em particular, em Finanças. Curiosamente, os modelos da classe ARCH surgiram no contexto da análise de séries temporais em macroeconomia.

Ainda que por meio de uma inspeção visual superficial, gráficos de dados financeiros mostram claramente que há períodos mais arriscados (voláteis) que outros; além disso, estes períodos mais ou menos voláteis geralmente não estão distribuídos aleatoriamente no tempo; ao contrário, há um certo grau de dependência na sequência de períodos mais ou menos voláteis. Analistas inspecionando tais gráficos descrevem este fenômeno como ‘agrupamento da volatilidade’. Os modelos da classe ARCH tratam da modelagem deste, entre outros tipos de fenômenos.

Nesta dissertação, o foco é modelar, analisar e efetuar previsões a cerca da volatilidade de três índices financeiros, o Ibovespa, o Dow Jones e o Standard & Poors (500), utilizando modelos da classe ARCH. Destaca-se a utilização de um modelo da classe ARCH que tem recebido considerável atenção na literatura econométrica, o modelo ARCH generalizado fracionalmente integrado. Este modelo possui a capacidade de modelar separadamente as dependências de curto e longo prazos da volatilidade de uma série, implicando em uma desejável flexibilidade. Destaca-se ainda a aplicação de uma densidade que também começa a receber atenção na literatura, sobretudo em Finanças, a t-Student assimétrica.

Selecionados os melhores modelos ARCH para cada série, serão analisadas as características da volatilidade de cada série e o desempenho destes modelos e das densidades condicionais adotadas será avaliado por meio de medidas da qualidade do ajuste e acurácia preditiva. Um dos objetivos principais é verificar o desempenho dos modelos GARCH fracionalmente integrados relativamente aos modelos GARCH e IGARCH, e os efeitos da adoção da t-Student assimétrica.

Na Seção 2, referencial teórico, relatam-se os principais fatos teóricos relativos aos modelos da classe ARCH. A literatura a respeito do tema é imensa e crescente, por isso um apanhado completo é quase impossível. Na Seção 3, dedi-

cada ao material e aos métodos, descrevem-se os dados, reportam-se as fontes dos dados e relata-se a metodologia a ser seguida. Na Seção 4 estão os resultados e a discussão dos mesmos. E finalmente, na Seção 5, colocam-se as conclusões finais.

2.1 Modelos estocásticos de séries temporais

A análise da dinâmica do sistema econômico é baseada na observação dos processos relevantes. Geralmente, a série temporal resultante mostra que algumas regularidades empíricas estão presentes em componentes explosivos ou cíclicos ou co-movimentos de séries distintas. De forma a estudar estas regularidades, uma série temporal é vista como a realização de um processo estocástico. Isto posto, nesta seção serão apresentados conceitos básicos da análise de séries temporais enquanto processos estocásticos, que serão usados de forma recorrente nesta dissertação. Maiores detalhes podem ser encontrados em Morettin & Tolo (1987).

Conforme Brockwell & Davis (1997), o primeiro passo na análise de uma série temporal é a seleção de um modelo matemático adequado, ou classe de modelos, para os dados. Para permitir a possibilidade de imprevisibilidade das observações futuras seria então natural supor que cada observação y_t é um valor realizado de uma variável aleatória $\{Y_t\}$. A série temporal $\{y_t, t \in T\}$ seria então uma realização da família de variáveis aleatórias $\{Y_t, t \in T_0\}$. Tais considerações sugerem que se deve modelar os dados como se estes fossem uma realização (ou parte de uma realização) de um processo estocástico $\{Y_t, t \in T_0\}$ onde $T \supseteq T_0$. Formalizam-se a seguir estes conceitos.

Definição 1 *Um processo estocástico é uma família de variáveis aleatórias $\{Y_t, t \in T\}$ definida sobre um espaço de probabilidade (Ω, Π, P) .*

Definição 2 *As funções $\{Y(t, \omega), \omega \in \Omega\}$ sobre T são denominadas realizações*

ou trajetórias amostrais do processo $\{Y_t, t \in T\}$.

Definição 3 A série temporal $\{Y_t, t \in \mathbb{Z}\}$, com $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$, é dita ser fracamente estacionária se, e somente se,

- i. $E|Y_t|^2 < \infty, \forall t \in \mathbb{Z}$, i.e, variância finita.
- ii. $E(Y_t) = m, \forall t \in \mathbb{Z}$, i.e, média independente de t .
- iii. $\gamma_Y(r, s) = \gamma_Y(r + t, s + t), \forall t \in \mathbb{Z}$, i.e, covariância depende somente da diferença entre os índices temporais.

Definição 4 A série temporal $\{Y_t, t \in \mathbb{Z}\}$ é dita fortemente estacionária se as distribuições conjuntas de $(Y_{t_1}, \dots, Y_{t_k})'$ e $(Y_{t_1+h}, \dots, Y_{t_k+h})'$ são as mesmas para todos os inteiros positivos k e para todo $t_1, \dots, t_k, h \in \mathbb{Z}$.

2.2 Índices Acionários

Nesta seção, descreve-se brevemente o que é um índice acionário, sua função, metodologia de cálculo, etc. Focaliza-se o índice Ibovespa, contudo, dada a semelhança do princípio, as idéias fundamentais estendem-se naturalmente aos índices Dow Jones e Standard & Poors (500).

Em geral, um índice acionário é o valor atual, em moeda corrente, de uma carteira hipotética de ações, constituída a partir de uma aplicação hipotética. Este é o caso do Ibovespa, que começou a ser calculado a partir de uma carteira teórica de ações em 1968. Supõe-se não ter sido efetuado nenhum investimento adicional desde então, considerando-se somente os ajustes efetuados em decorrência da distribuição de proventos pelas empresas emissoras. Isto posto, o índice reflete não apenas as variações dos preços das ações, mas também o impacto da distribuição dos proventos, sendo considerado um indicador que avalia o retorno total de suas ações componentes.

A finalidade básica de um índice acionário é servir como um indicador do de-

sempenho médio das principais ações transacionadas em uma bolsa de valores. A composição de um índice busca obter a melhor aproximação da real configuração das negociações à vista em uma bolsa de valores. No caso do Ibovespa, as ações integrantes da carteira teórica do índice respondem por mais de 80% do número de negócios e do volume financeiro verificados no mercado à vista da BOVESPA.

A participação de cada ação na carteira tem relação direta com a representatividade desse título no mercado à vista - em termos de número de negócios e volume financeiro - ajustada ao tamanho da amostra. Essa representatividade é obtida pelo *índice de negociabilidade* da ação, calculado pela seguinte fórmula:

$$IN = \sqrt{\frac{n_i}{N} \times \frac{v_i}{V}}$$

na qual n_i é o número de negócios com a ação i no mercado à vista, N o número total de negócios no mercado à vista da BOVESPA, v_i o volume financeiro gerado pelos negócios com a ação i no mercado à vista e V o volume financeiro total do mercado à vista da Bovespa.

O Ibovespa é o somatório dos pesos (quantidade teórica da ação multiplicada pelo último preço da mesma) das ações integrantes de sua carteira teórica. Assim sendo, pode ser apurado, a qualquer momento, por meio da seguinte fórmula:

$$Ibovespa_t = \sum_{i=1}^N P_{i,t} \times Q_{i,t}$$

em que $Ibovespa_t$ é o valor do índice no instante t , n o número total de ações componentes da carteira teórica, P o último preço da ação i no instante t e Q a quantidade teórica da ação i na carteira no instante t .

A BOVESPA calcula seu índice em tempo real, considerando os preços dos últimos negócios efetuados no mercado à vista com ações componentes de sua carteira. Assim, como a maioria dos índices acionários, pode-se acompanhar 'on line' seu comportamento em qualquer parte do Brasil ou do mundo.

2.3 ARCH

Nos modelos clássicos de séries temporais do tipo ARMA, assume-se que as inovações constituem uma sequência de ruídos brancos. Se os valores dos parâmetros estruturais, isto é, dos parâmetros autorregressivos e médias-móveis, são conhecidos, então as inovações são sinônimos dos erros de predição um passo à frente.

Na análise de dados econômicos, constitui uma regularidade empírica que grandes e pequenos erros de predição ocorrem em agrupamentos. Sugerindo que, em certos períodos, as variáveis econômicas tornam-se mais ou menos voláteis que o usual. Isto posto, Engle (1982) sugeriu que este fenômeno poderia ser modelado, elaborando-se um processo estocástico para a variância das inovações.

O objetivo era obter um modelo que se ajustasse melhor aos dados dentro da amostra, bem como fornecesse uma representação adequada da precisão das previsões. Esta característica pode ser muito útil se a série em questão é uma série de taxas de juros ou índice acionário. Isto porque a atração de um ativo financeiro está relacionada diretamente à sua taxa esperada de retorno e inversamente relacionada ao seu risco, o qual é uma função da variância da taxa de retorno.

Deve-se mencionar que o uso do termo 'ARCH' adotado nesta dissertação, assim como na literatura econométrica, não se restringe ao modelo original proposto por Engle (1982). O termo 'ARCH' designará o fenômeno da heterocedasticidade condicional em geral e todos os modelos que tentam capturar tal fenômeno.

Seguindo a formalização de Bollerslev et. al. (1994), seja $\{\varepsilon_t(\theta)\}$ um processo estocástico em tempo discreto com funções de média condicional e variância parametrizadas pelo vetor de dimensão finita $\theta \in \Theta \subseteq R^m$, onde θ_0 denota o valor verdadeiro. Inicialmente, assume-se que $\varepsilon_t(\theta)$ seja um escalar. E seja $E(\varepsilon_t(\theta)|\Psi_{t-1})$ a esperança matemática do processo, condicional ao passado e a qualquer conjunto de informação Ψ_{t-1} disponível no período $t - 1$.

O processo $\{\varepsilon_t(\theta)\}$ é definido como um processo heterocedástico condicional autorregressivo se a média condicional é igual a zero,

$$E(\varepsilon_t(\theta_0)|\Psi_{t-1}) = 0 \quad t=1,2,\dots, \quad (2.1)$$

mas a variância condicional,

$$\sigma_t^2(\theta_0) \equiv Var(\varepsilon_t(\theta_0)|\Psi_{t-1}) = E(\varepsilon_t^2(\theta_0)|\Psi_{t-1}) \quad t=1,2,\dots, \quad (2.2)$$

depende não-trivialmente da σ -álgebra gerada pelas observações passadas, isto é, $\{\varepsilon_{t-1}(\theta_0), \varepsilon_{t-2}(\theta_0), \dots\}$. Quando evidente a partir do contexto, a dependência explícita da variância condicional em relação aos parâmetros θ , será suprimida.

Focalizaremos o processo $\{\varepsilon_t\}$ na descrição subsequente, mas como posto por Bollerslev et. al. (1994), os mesmos conceitos estendem-se naturalmente para o caso no qual $\{\varepsilon_t\}$ seja o resíduo de algum modelo econométrico mais elaborado. Especificamente, seja $\{Y_t(\theta_0)\}$ o processo estocástico de interesse com média condicional,

$$\mu(\theta_0) \equiv E(y_t|\Psi_{t-1}) \quad t=1,2,\dots;; \quad (2.3)$$

note-se que, pela convenção anterior, ambos, $\mu_t(\theta_0)$ e $\sigma_t^2(\theta_0)$, são mensuráveis com respeito ao conjunto de informação disponível no período $t-1$, definindo-se o processo $\{\varepsilon_t(\theta_0)\}$ como,

$$\varepsilon_t(\theta_0) \equiv y_t - \mu_t(\theta_0) \quad t=1,2,\dots, \quad (2.4)$$

a variância condicional de $\{\varepsilon_t\}$, portanto, é igual a variância condicional do processo $\{y_t\}$. Visto que muito poucas séries temporais econômicas e financeiras têm média condicional constante igual a zero, a maioria das aplicações dos modelos ARCH seguem esta estrutura.

Retornando às equações (2.1) e (2.2), segue que o processo padronizado,

$$z_t(\theta_0) \equiv \varepsilon_t(\theta_0)(\sigma_t^2(\theta_0))^{-1/2} \quad t=1,2,\dots, \quad (2.5)$$

terá média condicional zero e variância condicional invariante no tempo igual à unidade. Esta observação constitui a base para a maioria dos procedimentos de inferência subjacentes às aplicações dos modelos ARCH.

Caso assumamos que a distribuição condicional de z_t seja invariante no tempo e possua quarto momento finito, segue da desigualdade de Jensen que,

$$E(\varepsilon_t^4) = E(z_t^4)E(\sigma_t^4) \geq E(z_t^4)E(\sigma_t^2)^2 = E(z_t^4)E(\varepsilon_t^2)^2, \quad (2.6)$$

onde a igualdade é verdadeira somente para o caso em que a variância condicional seja constante. Dada uma distribuição normal para as inovações padronizadas na equação (2.5), a distribuição incondicional para ε_t é, portanto, leptocúrtica.

Bollerslev et. al. (1994) escrevem que a especificação dos modelos ARCH, conforme as equações (2.1) a (2.4), é extremamente geral e não permite uma implementação empírica sem antes imporem-se certas restrições sobre as dependências temporais nas funções da média e variância condicionais. Além disso, mesmo em sua implementação univariada, as possíveis formas funcionais permitidas pela equação (2.2) são muito vastas, e para se ter alguma esperança de selecionar um modelo ARCH apropriado, deve-se ter uma boa idéia de quais as características empíricas que os modelos devem captar. Isto posto, um bom modelo deve ser capaz de capturar e refletir estas regularidades. Assim, descrevem-se a seguir algumas das características comuns do processo de volatilidade de séries econômicas e financeiras.

2.4 Regularidades empíricas de séries econômicas

2.4.1 Caudas pesadas

É um fato bem estabelecido que a distribuição incondicional dos retornos de ativos tem caudas pesadas (i.e, leptocúrticas). Mandelbrot (1963) e Fama (1965) foram pioneiros na documentação desta regularidade empírica. As estimativas de

curtose variam entre 4 e 50, indicando não-normalidade muito extrema. A relação entre as densidades condicional e incondicional dos retornos revela parcialmente a origem das caudas pesadas. Segundo Engle (1982), se a densidade condicional é gaussiana, então a densidade incondicional terá excesso de curtose devido à mistura de densidades gaussianas com diferentes variâncias. Contudo, não há razão para assumir que a densidade condicional dos retornos seja gaussiana, e como se verá posteriormente, pode-se adotar densidades condicionais leptocúrticas como a t-Student e a distribuição generalizada do erro.

2.4.2 Agrupamentos de volatilidade

O agrupamento de grandes movimentos e pequenos movimentos, de qualquer sinal, em séries econômico-financeiras foi uma das primeiras características documentadas do processo de volatilidade de tais séries. Mandelbrot (1963) e Fama (1965) reportaram evidências de que grandes variações no preço de um ativo são frequentemente seguidas por outras grandes variações, e pequenas variações são frequentemente seguidas por pequenas variações. Este agrupamento da volatilidade torna-se imediatamente aparente quando tais séries são plotadas contra o tempo. A implicação de tal agrupamento da volatilidade é que choques no processo de volatilidade atual influenciarão a esperança da volatilidade muitos períodos no futuro.

2.4.3 Assimetrias

Vários modelos de volatilidade impõem a hipótese de que a volatilidade condicional de um ativo é afetada simetricamente por inovações positivas e negativas. O modelo GARCH, por exemplo, permite que a variância condicional seja afetada somente pelo quadrado das inovações defasadas, desconsiderando completamente o sinal da inovação.

Para séries financeiras, é particularmente improvável que choques positivos e

negativos tenham o mesmo impacto sobre a volatilidade. Esta assimetria é descrita algumas vezes como *efeito alavancagem* e outras como *efeito prêmio de risco*. Segundo a primeira teoria, relatada originalmente por Fama (1965), quando o preço da ação de uma firma cai, a sua razão dívida/ativos aumenta, aumentando a volatilidade dos retornos dos acionistas. Segundo Bollerslev et. al. (1994), custos fixos como alavancagem financeira e operacional fornecem uma explicação parcial para este fenômeno.

Segundo a última teoria, informações sobre o aumento da volatilidade de um título reduzem a demanda pelo papel por causa da *aversão ao risco*. O consequente declínio da ação é seguido pelo aumento da volatilidade, como previsto pelas informações.

Glosten et. al. (1993) e Engle & Ng (1993), encontraram evidências da volatilidade ser negativamente relacionada ao retorno de ativos. Tais evidências não têm sido encontradas para o mercado de câmbio. Para taxas de juros, uma assimetria similar aparece a partir de taxas próximas de zero.

2.4.4 Memória longa

Conforme Baillie (1996), o tópico da memória longa e persistência tem atraído considerável atenção em termos do segundo momento de um processo, sendo que muitos dos exemplos de processos de memória longa emergiram de estudos de dados econômicos. O desenvolvimento de modelos e testes teóricos para a volatilidade com memória longa foi o resultado de encontrarem-se séries que exibiam fortemente este fenômeno. Baillie (1996) escreve que a primeira contribuição neste tópico foi de Taylor (1986), que reportou um fato estilizado aparente de que os retornos absolutos de ações tendem a ter autocorrelações que decaem muito lentamente. Ding et. al. (1993) notaram o mesmo fato para as potências de retornos diários e Dacorogna et. al. (1993) reportaram que os coeficientes de autocorrelação dos re-

tornos quadráticos diários de taxas de câmbio decaem muito lentamente.

2.4.5 Hipótese da eficiência do mercado

Em modelos de apreçamento de ativos tradicionais, a hipótese martingal (ou do passeio aleatório) tem desempenhado um papel muito importante na descrição e estudo de séries de preços de ativos.

Definição 5 *Um processo estocástico $\{y_t, t \geq 0\}$ é um martingal se, e somente se,*

1. $E(|y_t|) < \infty, \forall t,$
2. $E(y_t | \Psi_s) = y_s \forall s \leq t,$ onde Ψ_s é a σ -álgebra contendo os eventos determinados pelas observações no intervalo $[0, s]$, tal que $\Psi_s \subseteq \Psi_t,$ quando $s \leq t.$

Diversas interpretações simples para esta condição são dadas usualmente. Em termos de previsão, se a série y_t é uma série de preços, então, pela Definição 5, a melhor previsão do preço futuro, na data $t,$ é o preço atual $y_s.$ O preço de mercado reflete, então, toda informação útil para se fazer a melhor previsão. A condição para um martingal pode ser reescrita em termos de retornos relativos ou absolutos, ou seja,

$$E[(y_{t+1} - y_t) | \Psi_t] = 0, \quad (2.7)$$

ou

$$E\left[\left(\frac{y_{t+1} - y_t}{y_t}\right) \middle| \Psi_t\right] = 0, \quad (2.8)$$

assim, se um agente compra uma unidade de um ativo na data t ao preço Y_t de forma a revendê-la na data $t + 1$ pelo preço $y_{t+1},$ o ganho médio é zero se a condição martingal é satisfeita. Isto significa que, na ausência de custos de transação, não há perda ou lucro sistemáticos. Sob a forma $\{y_{t+1} - y_t\},$ um processo estocástico martingal é também conhecido como uma sequência de diferenças martingais.

Neste sentido, a informação disponível, $\{\Omega_t\}$, está totalmente contida nos preços e, portanto, sem utilidade na previsão do preço ou da taxa de retorno. A hipótese que os preços refletem toda a informação disponível, tornou-se conhecida como a *hipótese da eficiência do mercado*. De fato, Fama (1970) definiu três tipos de eficiência *informacional* do mercado de capitais (não confundir com eficiência relativa à *alocação* ou eficiência de *Pareto*). Cada uma das quais é baseada em uma noção diferente do tipo de informação que é considerada relevante. Isto posto, diz-se que um mercado é eficiente na forma *fraca* se $\{\Omega_t\}$ contém preços e retornos passados somente, eficiente na forma *semi-forte* caso $\{\Omega_t\}$ possui toda informação pública disponível e eficiente na forma *forte* se $\{\Omega_t\}$ contém toda informação pública e privada. Claramente, eficiência *forte* implica eficiência *semi-forte*, que, por sua vez, implica eficiência *fraca*, mas o reverso não ocorre.

O modelo martingal pode ser escrito de forma equivalente, como,

$$y_{t+1} = y_t + \varepsilon_t, \tag{2.9}$$

onde ε_t é uma sequência de diferenças martingais. Quando escrito nesta forma, o processo martingal é idêntico ao processo passeio aleatório, o principal modelo postulado pela teoria da eficiência do mercado de capitais. Contudo, o processo martingal é menos restritivo, segundo Campbell et. al. (1997). Uma diferença martingal requer somente a independência da esperança condicional das variações de preços em relação a $\{\Omega_t\}$, enquanto o passeio aleatório requer esta condição e também a independência dos momentos condicionais de ordens mais altas (variância, assimetria e curtose) da distribuição de probabilidade das variações de preços, isto é, dos retornos.

De fato, Campbell et. al. (1997) distinguem três versões da hipótese do passeio aleatório: a versão dos retornos independentes e identicamente distribuídos, a versão dos retornos independentes e a versão dos retornos não autocorrelacionados. O modelo martingal, por não requerer a independência probabilística entre retornos sucessivos, é inteiramente consistente com a regularidade empírica que

retornos, apesar de não autocorrelacionados, não tendem a ser independentes no tempo, mas sim tendem a exibir agrupamentos de períodos de alta e baixa volatilidade, isto é, dependência em momentos de altas ordens, notadamente na variância, como descrito na seção (2.4.2).

2.4.6 Macroeconomia e volatilidade

Bollerslev et. al. (1994, p. 2967) escrevem que os valores das ações são intimamente relacionados à vitalidade da economia; tornar-se-ia, então, natural esperar que as variâncias condicionais da produção industrial, taxas de juros, crescimento da oferta de moeda, etc., ajudassem a explicar variações da volatilidade do mercado acionário. Schwert (1989) reporta que apesar de ocorrer um aumento da volatilidade durante recessões e crises financeiras e uma redução durante expansões, a relação entre incerteza macroeconômica e volatilidade do mercado acionário seria surpreendentemente fraca. Conquanto, Glosten et. al. (1993) reportam uma forte relação positiva entre a volatilidade do retorno de ativos e taxas de juros.

2.5 ARCH: Descrição dos modelos

Como posto por Bollerslev et. al. (1994, p. 2969), as regularidades empíricas citadas anteriormente restringem a busca por um modelo ARCH adequado. Entretanto, as possíveis parametrizações são ainda vastas. A diversidade da classe de modelos ARCH paramétricos tem vantagens e desvantagens. Tal diversidade torna complexa a busca pelo modelo “verdadeiro”, e torna um pouco arbitrário o processo de seleção de modelos. Por outro lado, a flexibilidade derivada de tal diversidade implica que na análise de modelos econômicos estruturais com evidências de heterocedasticidade condicional autorregressiva, existe uma grande chance que um modelo ARCH paramétrico adequado seja formulado, tornando a análise tratável.

Nesta seção, descrevem-se os modelos da classe ARCH que foram utilizados na modelagem de séries reais nas análises posteriores. Os modelos foram elaborados de forma a captar as regularidades empíricas descritas anteriormente. Destacar-se-á, sobretudo, o modelo ARCH generalizado, denominado GARCH, posto que é a formulação a partir da qual a maioria das extensões que serão analisadas foram derivadas e contém o processo ARCH como um caso particular.

2.5.1 Modelo ARCH

Um processo ARCH pode ser definido em vários contextos e seguindo Bera & Higgins (1993), um processo ARCH será definido em termos da distribuição das inovações de um modelo de regressão linear dinâmico, neste contexto é usual a denominação modelo de regressão ARCH. Isto posto, o modelo ARCH(p) proposto no artigo seminal de Engle (1982) pode ser sumarizado por:

$$Y_t = x_t' \xi + \varepsilon_t, \quad (2.10a)$$

$$\varepsilon_t | \Psi_{t-1} \sim N(0, \sigma_t^2), \quad (2.10b)$$

$$\sigma_t^2 \equiv E[\varepsilon_t^2 | \Psi_{t-1}] = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 \equiv \omega + \alpha(L) \varepsilon_t^2, \quad (2.10c)$$

$$\omega > 0, \alpha_i \geq 0, \forall i = 1, \dots, p, \quad (2.10d)$$

em que $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_p L^p$, sendo L o operador de defasagens.

As restrições representadas por (2.10d) asseguram a positividade da variância. x_t é um vetor $k \times 1$ de variáveis exógenas, o qual pode incluir valores defasados da variável dependente $\{y_t\}$, e $\{\xi\}$ é um vetor $k \times 1$ de parâmetros. O modelo ARCH, representado pelas equações (2.10b) e (2.10c), caracteriza a distribuição das inovações ε_t condicional aos valores realizados do conjunto de informação $\{\Psi_{t-1} = y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \dots\}$. Note-se que como $\{\varepsilon_{t-i} = y_{t-i} - x_{t-i}' \xi\}$, $\{\sigma_t^2\}$ é uma função dos elementos de $\{\Psi_{t-1}\}$.

Conforme Bera & Higgins (1993, p. 309), a característica marcante do modelo

ARCH não é o fato de a variância condicional $\{\sigma_t^2\}$ ser uma função do conjunto de informação $\{\Psi_{t-1}\}$, mas sim a forma funcional especificada. Episódios de volatilidade são geralmente caracterizados pelo agrupamento de grandes choques na variável dependente. O processo da variância condicional (2.10c) é formulado de forma a reproduzir este fenômeno. No modelo de regressão, um grande choque é representado por um grande desvio de $\{y_t\}$ em relação à sua média condicional $x_t'\xi$, ou de forma equivalente, um grande valor positivo ou negativo de $\{\varepsilon_t\}$. No modelo de regressão ARCH, a variância da inovação atual, condicional aos valores realizados das inovações defasadas $\{\varepsilon_{t-i}, i = 1, \dots, p\}$, é uma função crescente da magnitude das inovações defasadas, seja qual for o sinal. Portanto, em concordância com a regularidade empírica descrita na seção (2.4.2), grandes erros de qualquer sinal tendem a ser seguidos por grandes erros de qualquer sinal. E de forma similar, pequenos erros de qualquer sinal tendem a ser seguidos por pequenos erros de qualquer sinal. A ordem do lag p determina o intervalo de tempo no qual um choque persiste em condicionar a variância dos erros subsequentes. Quanto maior o valor de p , maior a duração dos episódios de volatilidade.

Para fins de simulação, uma equação geradora explícita para um processo ARCH é,

$$\varepsilon_t = z_t \sqrt{\sigma_t^2}, \quad (2.11)$$

em que $z_t \sim IID N(0, 1)$ e $\{\sigma_t^2\}$ é dado por (2.10c). Posto que $\{\sigma_t^2\}$ é uma função de $\{\Psi_{t-1}\}$ e, portanto, fixado quando condicionado a Ψ_{t-1} , pode-se verificar que $\{\varepsilon_t\}$, como dado em (2.11), será condicionalmente normal com $E(\varepsilon_t | \Psi_{t-1}) = \sqrt{\sigma_t^2} E(z_t | \Psi_{t-1}) = 0$ e $Var(\varepsilon_t | \Psi_{t-1}) = \sigma_t^2 Var(z_t | \Psi_{t-1}) = \sigma_t^2$. Portanto, o processo especificado por (2.11) é idêntico ao processo ARCH (2.10c). A equação geradora (2.11) demonstra que um processo ARCH rescalona um processo subjacente de inovação gaussiana $\{z_t\}$ pela multiplicação deste pelo desvio padrão condicional, o qual é uma função do conjunto de informação $\{\Psi_{t-1}\}$.

Conforme Issler (1999), note-se a semelhança entre o processo (2.10c) e um

modelo AR(p):

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \epsilon_t \equiv \phi_0 + \phi(L)y_t + \epsilon_t.$$

De fato, definindo $\{\nu_t \equiv \epsilon_t^2 - \sigma_t^2\}$ e $\nu_t \sim N(0, 1)$, pode-se reescrever o modelo ARCH(p) em 2.10c como,

$$\epsilon_t^2 = \omega + \alpha(L)\epsilon_t^2 + \nu_t, \quad (2.12)$$

e tal modelo corresponde diretamente a um modelo AR(p) para os erros quadráticos $\{\epsilon_t^2\}$. Este processo é fracamente estacionário se, e somente se, a soma dos parâmetros autorregressivos positivos for menor que um, conforme Bollerslev et. al. (1994).

2.5.2 Modelo GARCH

Nas primeiras aplicações empíricas do modelo ARCH à relação entre o nível e a volatilidade da taxa de inflação, Engle (1982) verificou que requeria-se um grande valor para o lag p no processo da variância condicional. Isto implicava na necessidade de estimar-se um grande número de parâmetros sujeitos a restrições de desigualdade.

Para contornar este problema, Bollerslev (1986) propôs o modelo ARCH generalizado, ou modelo GARCH(p, q). Seja $\{\epsilon_t\}$ um processo estocástico em tempo discreto com valores reais, o modelo GARCH(p, q) é definido por,

$$y_t = x_t' \xi + \epsilon_t, \quad (2.13a)$$

$$\epsilon_t | \Psi_{t-1} \sim D(0, \sigma_t^2), \quad (2.13b)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 = \omega + \alpha(L)\epsilon_t^2 + \beta(L)\sigma_t^2, \quad (2.13c)$$

no qual as restrições de desigualdade:

$$\omega > 0; \quad \alpha_i \geq 0, \forall i = 1, \dots, p; \quad \beta_i \geq 0, \forall j = 1, \dots, q, \quad (2.14)$$

são impostas para assegurar a positividade estrita de $\{\sigma_t^2\}$, $D(\cdot)$ uma distribuição paramétrica e $\alpha(L)$ e $\beta(L)$ são polinômios de ordem finita no operador de defasagens L . Para $q = 0$, o processo reduz-se a um processo ARCH(p), e para $p = q = 0$, ε_t é simplesmente um processo ruído branco. No processo ARCH(p), a variância condicional é especificada como uma função somente das variâncias amostrais passadas, enquanto o processo GARCH(p, q) permite, além destas, a inclusão de variâncias condicionais defasadas (Bollerslev (1986)).

Bollerslev (1986, p. 44) escreve que se todas as raízes da expressão $1 - \beta(z) = 0$, na qual z é um escalar, estão fora do círculo unitário, o modelo GARCH representado por (2.13c) pode ser reescrito como,

$$\begin{aligned} \sigma_t^2 &= \frac{\omega}{1 - \beta(1)} + \frac{\alpha(L)}{1 - \beta(L)} \varepsilon_t^2, \\ &= \omega \left(1 - \sum_{i=1}^q \beta_i \right)^{-1} + \sum_{i=1}^{\infty} \delta_i \varepsilon_{t-i}^2, \\ &= \alpha_0^* + \sum_{i=1}^{\infty} \delta_i \varepsilon_{t-i}^2, \end{aligned} \quad (2.15)$$

na qual $\alpha_0^* = \omega / (1 - \beta(1))$ e os coeficientes δ_i são os coeficientes da expansão em séries de potência de $\alpha(L)[1 - \beta(L)]^{-1}$. Portanto, a expressão (2.15) mostra que um processo GARCH(p, q) é um processo ARCH de dimensão (ordem) infinita com uma estrutura de defasagem racional imposta sobre os coeficientes.

Conforme Issler (1999), a formulação do processo GARCH(p, q) pode ser compreendida a partir de uma analogia direta com os modelos da classe AR(p), no qual a solução parcimoniosa é incluir termos MA(q), dando origem a um modelo ARMA(p, q). Para verificar como um modelo GARCH fornece uma flexibilidade

similar à de um modelo ARMA, considere um modelo GARCH(1, 1)

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (2.16)$$

se define como em (2.12), $\{\nu_t \equiv \varepsilon_t^2 - \sigma_t^2\}$ como o erro de predição da variância condicional, com a propriedade $\nu_t \sim N(0, 1)$; pode-se resolver (2.16) em termos de $\{\varepsilon_t^2\}$ e $\{\nu_t\}$ defasados e atuais para obter-se

$$\varepsilon_t^2 = \omega + (\alpha + \beta)\varepsilon_{t-1}^2 - \beta\nu_{t-1} + \nu_t, \quad (2.17)$$

que é um processo ARMA(1, 1) para $\{\varepsilon_t^2\}$. Conforme Bollerslev (1986), através do mesmo procedimento, pode-se mostrar que um processo GARCH(p, q) corresponde a um processo ARMA($\max(p, q), q$) para $\{\varepsilon_t^2\}$, assim,

$$\varepsilon_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \varepsilon_{t-j}^2 - \sum_{j=1}^q \beta_j \nu_{t-j} + \nu_t, \quad (2.18)$$

ou

$$\phi(L)(1 - L)\varepsilon_t^2 = \omega + |1 - \beta(L)|\nu_t, \quad (2.19)$$

onde $\phi(L) \equiv [1 - \alpha(L) - \beta(L)](1 - L)^{-1}$ tem ordem $(m - 1)$. Conforme Bollerslev et. al. (1994), esta analogia com os modelos da classe ARMA permite o uso das técnicas usuais na identificação das ordens p e q .

Apesar das restrições (2.14) serem suficientes para assegurar que a variância condicional de um processo GARCH(p, q) seja estritamente positiva, Nelson & Cao (1992) demonstraram que condições suficientes mais fracas podem ser encontradas. Eles demonstraram que a partir da representação *invertida* de $\{\sigma_t^2\}$ em (2.13c),

$$\alpha_0^* > 0 \text{ e } \delta_i \geq 0, \quad i = 1, \dots, \infty, \quad (2.20)$$

são suficientes para assegurar que a variância condicional seja estritamente positiva. Nelson & Cao (1992) mostraram, expressando α_0^* e os δ_i 's em termos dos parâmetros originais do modelo GARCH, que (2.20) não requer que todas as desigualdades em (2.14) sejam satisfeitas. Bera & Higgins (1993) citam o exemplo

de um processo GARCH(1, 2), onde $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ e $\beta_1\alpha_1 + \alpha_2 \geq 0$ são suficientes para garantir que $\sigma_t^2 > 0$. Portanto, no modelo GARCH(1, 2), α_2 pode ter sinal negativo. Nelson & Cao (1992) demonstraram resultados gerais para os processos GARCH(1, q) e GARCH(2, q) e reportaram que as demonstrações para $p \geq 3$ são complexas. Diversos estudos empíricos, entre eles Engle et. al. (1990), reportaram coeficientes negativos e ainda assim as condições para a positividade da variância condicional, baseadas em (2.20), foram satisfeitas. Isto posto, Nelson & Cao (1992) recomendaram que as restrições (2.14) não necessitam ser impostas no processo de estimação, dado que a violação destas desigualdades não implica necessariamente que o processo da variância condicional esteja erroneamente especificado.

2.6 ARCH: Propriedades

2.6.1 Estacionariedade e momentos incondicionais

Nas seções anteriores, descreveram-se informalmente algumas propriedades dos processos ARCH(p) e GARCH(p, q); as condições de finitude dos momentos e a estacionariedade das distribuições *incondicionais* destes processos caracterizam formalmente estas propriedades. Como será visto nesta seção, Engle (1982) derivou expressões para muitos dos momentos incondicionais, e estabeleceu condições necessárias e suficientes para a existência de tais momentos. Posteriormente, Bollerslev (1986) estendeu estes resultados para o processo GARCH.

Conforme Engle (1982), o modelo ARCH(p) mais simples, o modelo ARCH(1) gaussiano com $\xi = 0$, permite a discussão de algumas propriedades desta classe de modelos. Tal modelo pode ser sumarizado por:

$$\begin{aligned} \varepsilon_t | \Psi_{t-1} &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2, \end{aligned} \tag{2.21}$$

note-se que, se $\alpha_1 = 0$ e $\varepsilon_t | \Psi_{t-1} \sim N(0, \omega)$, $\{\varepsilon_t\}$ é condicionalmente um processo ruído branco gaussiano e se $\alpha_1 > 0$, as observações serão dependentes através dos momentos de alta ordem. A média incondicional do processo ARCH, como será visto, é igual a zero e, pela simetria deste processo, derivada de sua normalidade condicional, tem-se que, sendo m um inteiro ímpar, $E(\varepsilon_t^m | \Psi_{t-1}) = 0$. Vê-se que o coeficiente de assimetria do processo ARCH é igual a zero. Os momentos incondicionais pares são calculados utilizando o Teorema 1.

Teorema 1 (Engle(1982)) *Para um inteiro r , os momentos incondicionais pares de um processo ARCH(1) com $\omega > 0$, $\alpha_1 \geq 0$, existem se, e somente se,*

$$\alpha_1^r \prod_{j=1}^r (2j - 1) < 1.$$

Prova, ver Engle (1982).

Similarmente, o processo GARCH(1,1) gaussiano pode ser utilizado para descrever algumas propriedades destes modelos. Os momentos pares incondicionais de um processo GARCH(1, 1) são calculados utilizando-se o Teorema 2.

Teorema 2 (Bollerslev(1986)) *Para o processo GARCH(1,1) dado por (2.16) uma condição necessária e suficiente para a existência dos momentos pares é*

$$\mu(\alpha_1, \beta_1, m) = \sum_{j=0}^m \binom{m}{j} a_j \alpha_1^j \beta_1^{m-j} < 1,$$

no qual,

$$a_0 = 1, \quad a_j = \prod_{i=1}^j (2i - 1), \quad j = 1, \dots;$$

os momentos pares podem ser expressos pela fórmula recursiva

$$E(\varepsilon_t^{2m}) = a_m \left[\sum_{n=0}^{m-1} a_n^{-1} E(\varepsilon_t^{2n}) \omega^{m-n} \binom{m}{m-n} \mu(\alpha_1, \beta_1, n) \right],$$

$$\times [1 - \mu(\alpha_1, \beta_1, n)]^{-1}. \quad (2.22)$$

Prova, ver Bollerslev (1986).

As condições para um processo ARCH(p) ser fracamente estacionário são dadas pelo Teorema 3.

Teorema 3 (Engle(1982)) *O processo ARCH(p), com $\omega > 0, \alpha_1, \dots, \alpha_p \geq 0$, é fracamente estacionário se, e somente se, a equação característica associada tem todas as raízes fora do círculo unitário. A variância estacionária é dada por $E(\sigma_t^2) = \sigma^2 = \frac{\omega}{(1 - \sum_{j=1}^p \alpha_j)}$.*

Prova, ver Engle (1982).

O Teorema 4 fornece as condições sob as quais um processo GARCH(p, q) é fracamente estacionário.

Teorema 4 (Bollerslev(1986)) *O processo GARCH(p, q) gaussiano, como definido em (2.13b) e (2.13c), é fracamente estacionário com $E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \omega(1 - \alpha(1) - \beta(1))^{-1}$ e $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$, para $t \neq s$ se, e somente se, $\alpha(1) + \beta(1) < 1$.*

Prova, ver Bollerslev (1986).

A derivação dos momentos incondicionais de processos da classe ARCH é possível através do uso intensivo do seguinte resultado da teoria da probabilidade,

Lei das Esperanças Iteradas: Sejam $\{\Psi_1\}$ e $\{\Psi_2\}$ dois conjuntos de variáveis aleatórias tal que $\Psi_1 \subseteq \Psi_2$ e seja γ uma variável aleatória escalar. Então $E(\gamma|\Psi_1) = E[E(\gamma|\Psi_2)|\Psi_1]$.

No contexto da análise de séries temporais, $\{\Psi_1\}$ e $\{\Psi_2\}$ são conjuntos de informação (σ -álgebras) disponíveis em diferentes períodos de tempo. Como posto por Bera & Higgins (1993), um caso especial da lei é frequentemente empregado para o cálculo dos momentos incondicionais de processos da classe ARCH. Se $\Psi = \emptyset$ é o conjunto vazio, então $E(\gamma) = E[E(\gamma|\Psi_2)]$. A utilidade desta expressão está no fato de relacionar um momento condicional a um momento incondicional. Posto que processos do tipo ARCH são especificados em termos de momentos condicionais, a lei fornece um método para a derivação dos momentos incondicionais.

Por exemplo, considere-se o cálculo da média incondicional de um processo ARCH(p) para o erro $\{\varepsilon_t\}$ com variância condicional parametrizada por (2.10c). Aplicando a lei das esperanças iteradas, tem-se que $E(\varepsilon_t) = E[E(\varepsilon_t|\Psi_{t-1})]$. Entretanto, posto que o modelo ARCH especifica, por definição, que $E(\varepsilon_t|\Psi_{t-1}) = 0$ para todas as realizações de $\{\Psi_{t-1}\}$, segue imediatamente que $E(\varepsilon_t) = 0$; portanto, o processo ARCH tem média incondicional igual a zero; claramente, o mesmo verifica-se para um processo GARCH.

Considerando agora o segundo e quarto momentos incondicionais do processo ARCH(1), pelo Teorema 1, conforme Engle (1982), a condição para que a variância incondicional seja finita é $\alpha_1 < 1$, enquanto, para ter-se um quarto momento incondicional finito, requer-se $3\alpha_1^2 < 1$. Se tais condições são verificadas, pode-se mostrar que o segundo (variância) e o quarto momentos incondicionais de um processo ARCH(1) são respectivamente iguais a,

$$\begin{aligned} E(\varepsilon_t^2) &= \frac{\omega}{1 - \alpha_1}, \\ E(\varepsilon_t^4) &= \left(\frac{3\omega^2}{(1 - \alpha_1)^2} \right) \left(\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} \right). \end{aligned} \quad (2.23)$$

Isto tem duas implicações. Primeira: apesar do modelo ser condicionalmente normal, a distribuição incondicional é leptocúrtica em relação à distribuição normal. Isto ocorre porque se $\alpha_1 \neq 0$ tem-se que $\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 1$; assim, (2.23) é maior

que 3. Segunda: apesar da distribuição condicional ser heterocedástica, a distribuição incondicional é homocedástica. Isto é, apesar da variância de $\{\varepsilon_t\}$, condicional a $\{\Psi_{t-1}\}$, alterar-se com os elementos do conjunto de informação, incondicionalmente o processo ARCH é homocedástico. Portanto, o processo é fracamente estacionário apesar de exibir heterocedasticidade condicional. Tal resultado é generalizado para uma ampla classe de processos do tipo ARCH através do Teorema 3.

Quanto ao modelo GARCH(1, 1), se $\alpha_1 + \beta_1 < 1$, a variância incondicional existe. Caso $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$, o mesmo ocorre para o quarto momento incondicional. Pelo Teorema 3, as expressões destes momentos são, respectivamente,

$$E(\varepsilon_t^2) = \frac{\omega}{1 - \alpha_1 - \beta_1}, \quad (2.24)$$

$$E(\varepsilon_t^4) = \frac{3\omega^2(1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)}. \quad (2.25)$$

As condições gerais para a existência do momento de quarta ordem finito são desconhecidas. Contudo, conforme Bollerslev (1986), dada uma ordem específica para o modelo, as condições podem ser derivadas através da mesma linha de argumentação do Teorema 3. Por exemplo, considere-se um processo GARCH(p, q) condicionalmente gaussiano. O segundo e quarto momentos são relacionados por,

$$E(\varepsilon_t^4 | \varepsilon_{t-1}) = 3[E(\varepsilon_t^2 | \varepsilon_{t-1})]^2,$$

isto posto, pode-se mostrar que,

$$k = \frac{E(\varepsilon_t^4)}{[E(\varepsilon_t^2)]^2} = 3 + 3 \frac{E[E(\varepsilon_t^2 | \varepsilon_{t-1})^2] - E[E(\varepsilon_t^2 | \varepsilon_{t-1})]}{[E(\varepsilon_t^2)]^2}, \quad (2.26)$$

da qual pode-se deduzir que a distribuição incondicional de $\{\varepsilon_t\}$ de um processo GARCH(p, q) é leptocúrtico em relação à normal. Além disso, a curtose pode ser vista como uma medida natural de heterocedasticidade condicional.

Verificando-se as condições citadas, as propriedades observadas anteriormente para o processo ARCH, i.e, distribuição incondicional leptocúrtica em relação à distribuição normal e homocedasticidade incondicional, generalizam-se naturalmente ao processo GARCH, fato formalizado pelos Teoremas 3 e 4. O fato de processos da classe ARCH serem leptocúrticos em relação à distribuição normal torna estes processos atrativos no que concerne à modelagem de séries econômicas, posto que a distribuição do retorno de ativos frequentemente demonstra esta característica, como descrito na seção (2.4.1). Um resultado derivado por Nelson (1990a) indica por que distribuições leptocúrticas são tão prevalentes em séries econômicas de alta frequência. Nelson (1990a) demonstrou que sob condições regulares, quando o intervalo de tempo tende no limite a zero, o processo GARCH(1,1) aproxima-se de um processo estocástico em tempo contínuo, denominado processo de difusão, cuja distribuição incondicional é uma t-Student.

Bollerslev et. al. (1994, p. 2989) utilizam um argumento recursivo, através de lei das esperanças iteradas, na avaliação da estacionariedade de processos da classe ARCH. Para manter a discussão em um nível geral, Bollerslev et. al. (1994) consideram o caso multivariado de (2.11), onde,

$$\varepsilon_t = \Omega^{1/2} Z_t, \quad Z_t \sim iid, \quad E(Z_t) = 0_{n \times 1}, \quad E(Z_t Z_t') = I_{n \times n}, \quad (2.27)$$

e,

$$\Omega_t = \Omega(t, Z_{t-1}, Z_{t-2}, \dots). \quad (2.28)$$

Utilizando o critério de ergodicidade, segue que a estacionariedade *forte* ou estrita de $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ é equivalente à condição,

$$\Omega_t = \Omega(Z_{t-1}, Z_{t-2}, \dots), \quad (2.29)$$

com $\Omega_t(\cdot, \cdot, \dots)$ mensurável, e

$$\text{Traço}(\Omega_t \Omega_t') < \infty \quad \text{quase-certamente}; \quad (2.30)$$

a equação (2.29) elimina a dependência direta de $\{\Omega_t\}$ de t , enquanto (2.30) assegura que choques aleatórios em $\{\Omega_t\}$ decaiam rapidamente o bastante de forma a evitar que $\{\Omega_t\}$ tenha um trajetória assintoticamente explosiva.

Bollerslev et. al. (1994, p. 2989) escrevem que uma condição suficiente para (2.30) é a limitação dos momentos, i.e, $E[\text{Traço}(\Omega_t \Omega_t')^p]$ finita para algum $p > 0$ implica em $\text{Traço}(\Omega_t \Omega_t') < \infty$ quase-certamente. Por exemplo, como pode ser verificado pelo Teorema 4, Bollerslev (1986) mostrou que no modelo GARCH(p, q), $E(\sigma_t^2)$ é finita e $\{\varepsilon_t\}$ é fracamente estacionário, quando $\sum_{i=1,p} \alpha_i + \sum_{j=1,q} \beta_j < 1$. Isto é uma condição suficiente, mas não necessária para estacionariedade estrita. Bollerslev et. al. (1994, p. 2989) colocam que, posto que processos da classe ARCH são leptocúrticos, as condições para estacionariedade fraca são frequentemente mais restritivas que as condições para estacionariedade forte.

De fato, uma propriedade notável de processos da classe ARCH, inicialmente demonstrada por Nelson (1990b) para o modelo GARCH(1,1) e generalizada para o modelo GARCH(p, q) por Bougerol & Picard (1992), é que estes podem ser fortemente estacionários apesar de não serem fracamente estacionários. Para o modelo GARCH(1,1), (2.28) é expresso como,

$$\sigma_t^2 = \omega \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^k (\beta_1 + \alpha_1 \eta_{t-i}^2) \right]. \quad (2.31)$$

Nelson (1990b) mostrou que quando $\omega > 4$, $\sigma_t^2 < \infty$ quase-certamente, $\{\varepsilon_t, \sigma_t^2\}$ é fortemente estacionária se, e somente se, $E[\ln(\beta_1 + \alpha_1 z_t^2)] < 0$. Nelson (1990b) verificou que por uma aplicação da desigualdade de Jensen, esta seria uma condição consideravelmente mais fraca que $\alpha_1 + \beta_1 < 1$, a condição necessária e suficiente para estacionariedade fraca. Por exemplo, o modelo GARCH(1,0) com $\varepsilon_t \sim N(0, 1)$, $\alpha_1 = 3$ e $\beta_1 = 0$ é fortemente, mas não fracamente estacionário.

O fato da parametrização de processos da classe ARCH não impor a priori a existência dos momentos incondicionais é uma característica importante des-

tes modelos. A literatura relata há tempos, pelo menos desde Mandelbrot (1963) e Fama (1965), que tratam desta questão, que diversas séries econômicas, particularmente séries de preços dos mercados de commodities e financeiros, podem ser bem representadas por uma classe de distribuições com variância infinita. Um exemplo é a distribuição proposta pelo economista italiano Vilfredo Pareto (1848-1923), conhecida como distribuição de Pareto, com densidade $f(e) = c(e - e_0)^{-\alpha-1}$, onde c , e_0 e α são constantes, sendo a variância infinita para $\alpha < 2$.

Em aplicações empíricas do processo GARCH, frequentemente os parâmetros não satisfazem $\alpha(1) + \beta(1) < 1$. O fato de processos da classe ARCH admitirem variância infinita é desejável posto que tal comportamento pode ser uma característica do processo gerador de dados que deve estar refletida no modelo estimado. Além disso, como será visto posteriormente, mesmo para modelos GARCH com variâncias infinitas, resultados usuais relativos à consistência e normalidade assintótica podem ainda ser válidos.

Considerou-se até agora a distribuição univariada de um único $\{\varepsilon_t\}$. Os momentos da distribuição conjunta dos $\{\varepsilon_t\}$'s também revelam propriedades importantes de processos da classe ARCH. Para $k \geq 1$, as autocovariâncias do processo GARCH(p, q) são,

$$\begin{aligned} E(\varepsilon_t \varepsilon_{t-k}) &= E[E(\varepsilon_t \varepsilon_{t-k} | \Psi_{t-1})], \\ &= E[\varepsilon_{t-k} E(\varepsilon_t | \Psi_{t-1})], \\ &= 0. \end{aligned} \tag{2.32}$$

O fato de processos GARCH serem não autocorrelacionados os torna adequados para a modelagem de séries econômicas, particularmente as financeiras. A hipótese da eficiência dos mercados, informalmente, postula que a informação contida nas taxas passadas de retorno não melhora a predição de taxas futuras de retorno. Em (2.13a), supondo-se que $\{y_t\}$ seja a taxa de retorno de um ativo e que $\beta = 0$, tal que não há componente de regressão no modelo. Assim, $\{y_t\}$ é idêntico

a $\{\varepsilon_t\}$ e torna-se um processo GARCH puro. A previsão ótima do retorno $\{Y_t\}$ é a esperança do retorno condicional a toda informação disponível. Mas dado que, o modelo GARCH especifica que $E(y_t|\Psi_{t-1}) = E(y_t) = 0$, as observações passadas de $\{y_t\}$ contidas em $\{\Psi_{t-1}\}$ não alteram a previsão ótima da taxa de retorno. Portanto, a presença de efeitos ARCH não representa uma violação da hipótese de eficiência dos mercados.

2.6.2 Simetria e regularidade

As definições a seguir, postuladas por Engle (1982), estabelecem as condições sob as quais um processo da classe ARCH é simétrico e regular, respectivamente. Conforme Engle (1982, p. 993), seja ζ_t um vetor aleatório $p \times 1$ retirado de um espaço amostral Ω , o qual possui elementos $\zeta'_t = (\zeta_{t-1}, \dots, \zeta_{t-p})$. Para todo ζ_t , seja ζ_t^* idêntico, exceto que o m -ésimo elemento foi multiplicado por (-1) , onde m está entre 1 e p . Isto posto,

Definição 6 *Um processo da classe ARCH é simétrico se,*

1. $\sigma_t^2(\zeta_t) = \sigma_t^2(\zeta_t^*) \forall m \text{ e } \zeta_t \in \Omega$,
2. $\partial\sigma_t^2(\zeta_t)/\partial\alpha_i = \partial\sigma_t^2(\zeta_t^*)/\partial\alpha_i \forall m, i \text{ e } \zeta_t \in \Omega$,
3. $\partial\sigma_t^2(\zeta_t)/\partial\zeta_{t-m} = -\partial\sigma_t^2(\zeta_t^*)/\partial\zeta_{t-m} \forall m \text{ e } \zeta_t \in \Omega$.

Neste sentido, o processo GARCH(p, q) é simétrico.

Definição 7 *Um processo da classe ARCH é regular se,*

1. $\min \sigma_t^2(\zeta_t) \geq \vartheta$ para algum $\vartheta > 0$ e $\zeta_t \in \Omega$,
2. $E(\|\partial\sigma_t^2(\zeta_t)/\partial\alpha_i\| \|\partial\sigma_t^2(\zeta_t)/\partial\zeta_{t-m}\| |\Psi_{t-m-1}|) \exists \forall i, m, t$.

Segundo Engle (1982, p. 994), o item 1 da Definição 7 é importante e fácil de verificar, posto que requer que a variância seja sempre positiva. O item 2 é uma condição suficiente para a existência de algumas esperanças do hessiano utilizado no Teorema 5.

2.6.3 Previsão

Conforme Bera & Higgins (1993), uma utilização muito importante de modelos ARCH é o cálculo da acurácia de previsões. Na metodologia tradicional de séries temporais que utiliza processos ARMA condicionalmente homocedáticos, a variância do erro de previsão não depende do conjunto de informação disponível. Se a série sendo prevista exibe ARCH, $\{\Psi_{t-1}\}$ pode indicar a acurácia pela qual a série pode ser prevista. Nesta seção serão revisados os principais resultados relativos à previsão com modelos ARCH.

Para ilustrar os efeitos de erros ARCH sobre a medida da incerteza das previsões no contexto da previsão de uma série temporal, seja um processo ARMA(k, l)

$$\phi(L)Y_t = \theta(L)\varepsilon_t, \quad (2.33)$$

e $\{\varepsilon_t\}$ um processo GARCH(p, q). Considere-se a previsão do valor do processo s períodos a partir de uma origem t , a qual é dada por,

$$y_{t+s} = \sum_{i=1}^k \phi_i y_{t+s-i} + \sum_{i=1}^l \theta_i \varepsilon_{t+s-i} + \varepsilon_{t+s}, \quad (2.34)$$

o preditor ótimo é a média de y_{t+s} condicional ao conjunto de informação disponível em t , Ψ_t . Dado que $E(\varepsilon_{t+s}|\Psi_t) = 0$, o preditor ótimo é,

$$E(y_{t+s}|\Psi_t) = \sum_{i=1}^k \phi_i E(y_{t+s-i}|\Psi_t) + \sum_{i=1}^l \theta_i E(\varepsilon_{t+s-i}|\Psi_t), \quad (2.35)$$

sendo que:

a) $E(y_{t+s-i}|\Psi_t)$, para $i < s$, é dada recursivamente por (2.35),

- b) $E(y_{t+s-i}|\Psi_t) = y_{t+s-i}$, para $i \geq s$,
c) $E(\varepsilon_{t+s-i}|\Psi_t) = 0$, para $i < s$,
d) $E(\varepsilon_{t+s-i}|\Psi_t) = \varepsilon_{t+s-i}$, para $i \geq s$.

A expressão (2.35) é a relação recursiva padrão para a previsão pontual ótima do processo ARMA convencional. Isto posto, Gouriéroux (1997) escreveu que a presença de ARCH não afeta o modo pelo qual a previsão é construída. Isto ocorre porque ARCH introduz dependência em momentos de altas ordens e afeta somente a incerteza da previsão pontual.

Para avaliar o impacto de ARCH sobre a incerteza da previsão pontual, requer-se uma expressão para o erro de previsão. Assumindo-se que as raízes de $\phi(L) = 1 - \phi_1 L - \dots - \phi_k L^k$ estão fora do círculo unitário, o processo (2.33) pode ser invertido, fornecendo,

$$y_{t+s} = \sum_{i=0}^{\infty} \varpi_i \varepsilon_{t+s-i}, \quad (2.36)$$

sendo que ϖ_i é o coeficiente de L^i na expansão de $\phi(L)^{-1}\theta(L)$. Usando-se a representação em médias-móveis, o preditor ótimo é,

$$E(y_{t+s}|\Psi) = \sum_{i=s}^{\infty} \varpi_i \varepsilon_{t+s-i}, \quad (2.37)$$

e seja $\{\varepsilon_{t,s}\}$ o erro de previsão a partir da origem t com horizonte de previsão s . Subtraindo (2.37) de (2.36), o erro de previsão é,

$$\varepsilon_{t,s} = y_{t+s} - E(y_{t+s}|\Psi_t) = \sum_{i=s}^{s-1} \varpi_i \varepsilon_{t+s-i}, \quad (2.38)$$

sendo, portanto, uma combinação linear das inovações a partir de $t+1$ ao horizonte $t+s$. A incerteza em uma previsão pode então ser medida pela variância do erro de previsão condicional ao conjunto de informação, $\{\Psi_t\}$, utilizado para construir

a previsão. A partir de (2.38), a variância condicional do erro de previsão é,

$$Var(\varepsilon_{t,s}|\Psi_t) = \sum_{i=0}^{s-1} \varpi_i^2 E(\varepsilon_{t+s-i}^2|\Psi_t). \quad (2.39)$$

A expressão (2.39) mostra como erros ARCH afetam a variância condicional do erro de previsão. Conforme Bera & Higgins (1993), quando ARCH está presente, $E(\varepsilon_{t+s-i}^2|\Psi_t)$ dependerá dos elementos de $\{\Psi_t\}$ e será, em geral, variante no tempo. Entretanto, para um modelo condicionalmente homocedástico, no qual $E(\varepsilon_{t+s-i}^2|\Psi_t) = \sigma_\varepsilon^2$, a variância do erro de previsão reduz-se à,

$$Var(\varepsilon_{t,s}|\Psi_t) = \sigma_\varepsilon^2 \sum_{i=0}^{s-1} \varpi_i^2, \quad (2.40)$$

assim, a variância do erro de previsão não depende dos elementos de $\{\Psi_t\}$, mas somente da extensão do horizonte de previsão s .

De forma a operacionalizar (2.39) para, por exemplo, construir intervalos de previsão, é necessário calcular as esperanças $E(\varepsilon_{t+s-i}^2|\Psi_t)$. Isto pode ser feito utilizando-se a representação ARMA($max(p, q), q$) do processo GARCH(p, q),

$$\varepsilon_{t+s}^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t+s-i}^2 + \sum_{j=1}^q \beta_j \varepsilon_{t+s-j}^2 - \sum_{j=1}^p \beta_j \nu_{t+s-j} + \nu_{t+s},$$

fazendo-se $m = \{max(p, q), q\}$, pode-se utilizar a seguinte representação ARMA(m, q), equivalente à anterior,

$$\varepsilon_{t+s}^2 = \omega + \sum_{i=1}^m (\alpha_i + \beta_i) \varepsilon_{t+s-i}^2 - \sum_{i=1}^q \beta_i \nu_{t+s-i} + \nu_{t+s}.$$

Tomando-se a esperança condicional, têm-se.

$$E(\varepsilon_{t+s}^2|\Psi_t) = \omega + \sum_{i=1}^m (\alpha_i + \beta_i) E(\varepsilon_{t+s-i}^2|\Psi_t) - \sum_{i=1}^q \beta_i E(\nu_{t+s-i}|\Psi_t), \quad (2.41)$$

sendo:

- a) $E(\varepsilon_{t+s-i}^2|\Psi_t)$, para $i < s$, é dada recursivamente por (2.41),
 b) $E(\varepsilon_{t+s-i}^2|\Psi_t) = \varepsilon_{t+s-i}^2$, para $i \geq s$,
 c) $E(\nu_{t+s-i}|\Psi_t) = 0$, para $i < s$,
 d) $E(\nu_{t+s-i}|\Psi_t) = \nu_{t+s-i}$, para $i \geq s$.

A expressão para $E(\varepsilon_{t+s}^2|\Psi_t)$ em (2.41) é completamente análoga ao preditor ótimo $E(y_{t+s}|\Psi_t)$ em (2.37).

Como um exemplo, para a construção de estimativas da variância da previsão, considere-se um processo AR(1) estacionário,

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \text{ com } |\phi_1| < 1,$$

em que ε_t é um processo GARCH(1,1). A previsão pontual ótima segue a recursão,

$$E(y_{t+s}|\Psi_t) = \phi_1 E(y_{t+s-1}|\Psi_{t-1}),$$

sendo que a previsão para o primeiro período é $E(y_{t+1}|\Psi_t) = \phi_1 y_t$. Invertendo o processo AR(1), os coeficientes em (2.37) são dados por $\varpi_i = \phi_1^i$. Portanto, a partir de (2.39), a variância do erro de previsão é,

$$Var(\varepsilon_{t,s}|\Psi_t) = \sum_{i=0}^{s-1} \varpi_1^{2i} E(\varepsilon_{t+s-i}^2|\Psi_t), \text{ para } s \geq 1,$$

sendo que as esperanças podem ser calculadas recursivamente por,

$$E(\varepsilon_{t+s}^2|\Psi_t) = \omega + (\alpha_1 + \beta_1)E(\varepsilon_{t+s}^2|\Psi_t), \text{ para } s > 1,$$

com a esperança inicial igual a $E(\varepsilon_{t+s}^2|\Psi_t) = \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2$.

2.6.4 Persistência na variância

Como exposto na seção anterior, quando ARCH está presente, o conjunto de informação disponível $\{\Psi_t\}$ é útil para verificar a acurácia com a qual um processo pode ser previsto. Isto posto, torna-se interessante considerar como $\{\Psi_t\}$ afeta a

incerteza da previsão, conforme o horizonte de previsão s aumenta. Segundo Bera & Higgins (1993), para $s > p$, a esperança condicional (2.41) reduz-se a,

$$E(\varepsilon_{t+s}^2 | \Psi_t) = \omega + \sum_{i=1}^m (\alpha_i + \beta_i) E(\varepsilon_{t+s-i}^2 | \Psi), \quad (2.42)$$

que é uma equação em diferenças linear para a sequência $\{E(\varepsilon_{t+s}^2 | \Psi_t)\}_{s=p+1}^{\infty}$. Se as raízes de $1 - (\alpha_1 + \beta_1)L - \dots - (\alpha_m + \beta_m)L^m = 1 - \alpha(L) - \beta(L)$ estão fora do círculo unitário, a solução de (2.42) converge para,

$$\lim_{s \rightarrow \infty} E(\varepsilon_{t+s}^2 | \Psi_{t-1}) = \frac{\omega}{1 - \alpha_1 - \dots - \alpha_p - \beta_1 - \dots - \beta_p}, \quad (2.43)$$

que é a variância incondicional de $\{\varepsilon_t\}$. Neste caso, à medida que o horizonte de previsão aumenta, $\{\Psi_t\}$ não fornece informação sobre a variância de $\{\varepsilon_{t+s}\}$. Contudo, caso as raízes de $1 - \alpha(L) - \beta(L)$ estejam sobre ou dentro do círculo unitário, não ocorrerá tal convergência. Por exemplo, considere-se um processo GARCH(1,1) com $1 - \alpha(L) - \beta(L)$ possuindo uma raiz unitária, implicando que $\alpha_1 + \beta_1 = 1$; assim, (2.42) reduz-se a,

$$E(\varepsilon_{t+s}^2 | \Psi_t) = \omega + E(\varepsilon_{t+s-1}^2 | \Psi_t), \quad (2.44)$$

que tem como solução,

$$E(\varepsilon_{t+s}^2 | \Psi_t) = s\omega + E(\varepsilon_t^2 | \Psi_t), \quad (2.45)$$

assim, quando $\alpha_1 + \beta_1 = 1$, a previsão da variância condicional cresce linearmente com o horizonte de previsão s e a dependência de $\{\Psi_t\}$ persiste através de $E(\varepsilon_t^2 | \Psi_t)$.

Como posto por Bollerslev et. al. (1994), a noção de *persistência* de um choque sobre a variância ou volatilidade no contexto da classe de processos ARCH é significativamente mais complexa que o conceito equivalente de persistência sobre a média de modelos lineares como o ARMA. Isto ocorreria porque mesmo modelos da classe ARCH fortemente estacionários, como visto, frequentemente

não possuem momentos incondicionais finitos.

Bollerslev et. al. (1994) escrevem que seria bastante razoável definir persistência de choques em termos da previsão de momentos. Isto é, escolher algum $\eta > 0$ e definir que choques sobre $\{\sigma_t^2\}$ não persistem se, e somente se, para todo s , $E_s(\sigma_t^{2\eta})$ converge, quando $t \rightarrow \infty$, para um limite finito independente da informação até o tempo s . Tal definição de persistência seria particularmente apropriada quando uma teoria econômica faz da previsão de momentos, ao invés da previsão da distribuição, seu objeto de interesse. Isto posto, a ambiguidade gerada por como se define persistência sobre $\{\sigma_t^2\}$ fica exposta, dado que o momento condicional $E_s(\sigma_t^{2\eta})$ pode divergir para o infinito para algum η , mas convergir para um limite bem comportado e independente das condições iniciais para outro η , mesmo que $\{\sigma_t^2\}$ seja estacionária e ergódica.

Assim, Engle & Bollerslev (1986) consideraram um processo GARCH com $\alpha(1) + \beta(1) = 1$ como um modelo com propriedades específicas, que foi denominado processo GARCH *integrado* (IGARCH). Engle & Bollerslev (1986) viram similaridades entre o processo IGARCH e processos integrados na média. Em um processo que é integrado na média, isto é, que deve ser diferenciado para obter-se um processo estacionário, um choque no período atual afeta o nível da série infinitamente. De forma similar, em um processo IGARCH, um choque condiciona infinitamente as variâncias futuras.

Nelson (1990b) mostrou que a analogia com processos integrados na média tem sérias limitações. Nelson (1990b) mostrou que apesar de processos IGARCH não serem fracamente estacionários, dado que possuem variância infinita, estes podem ser fortemente estacionários. Entretanto, processos integrados na média não são estacionários em nenhum sentido.

Bera & Higgins (1993) escrevem que o relato consistente de grande persistência na variância de séries temporais financeiras é surpreendente, posto que nenhuma teoria econômica prediz tal comportamento. Lamoureux & Lastrapes

(1990) escrevem que esta alta persistência pode ser causada por especificação errada da variância e resultado de uma mudança estrutural na variância incondicional do processo, causada por uma mudança em ω por exemplo. Além disso, Bollerslev et. al. (1994) escrevem que a persistência aparente de choques pode dever-se à leptocurticidade das distribuições, em vez de por não-estacionariedades inerentes.

2.6.5 Relação com processos em tempo contínuo

Formalmente, processos da classe ARCH são equações em diferenças estocásticas não-lineares. Bollerslev et. al. (1994) escrevem que isto torna suas propriedades probabilísticas e estatísticas, tal como estacionariedade, finitude dos momentos, consistência e normalidade assintótica dos estimadores de máxima verossimilhança, consideravelmente mais complexas que as dos modelos lineares. Assim, uma forma de simplificar a análise destas propriedades dos processos ARCH é aproximar as equações em diferenças estocásticas por mais tratáveis equações diferenciais estocásticas, este foi o tipo de argumentação utilizado em Nelson (1990a). Por outro lado, processos discretos, como os da classe ARCH, são mais convenientes para certos propósitos que os processos representados por equações diferenciais estocásticas, notavelmente, no que concerne aos procedimentos de estimação e inferência.

Expor a teoria dos processos ARCH como aproximação de processos contínuos de difusão, e vice-versa, está muito além do escopo desta dissertação. Contudo, considerou-se relevante traçar as linhas gerais do problema da aproximação de um processo de difusão contínuo por um processo discreto como o ARCH. Ou seja, dado que os economistas elaboram modelos teóricos para o sistema econômico utilizando-se processos estocásticos em tempo contínuo e considerando que, em última instância, os dados disponíveis são discretos, haveria fundamentos para a utilização de processos discretos como aproximações adequadas destes modelos teóricos? Isto é, poder-se-ia aproximar, de forma adequada, um processo em tempo contínuo por um processo discreto como o ARCH e vice-versa? Nelson

(1990a) respondeu a esta questão utilizando um instrumental bastante sofisticado, o cálculo estocástico.

Por exemplo, segundo Bollerslev et. al. (1994), um modelo de difusão bastante aplicado na literatura econométrica sobre apreçamento de opções postula que seja $\{y_t\}$ o preço de um ativo e $\{\sigma_t\}$ a volatilidade do seu retorno. O processo em tempo contínuo para a evolução conjunta de $\{y_t, \sigma_t\}$ é dado por,

$$dy_t = \mu y_t dt + y_t \sigma_t dW_{1,t}, \quad (2.46)$$

e,

$$d[\ln(\sigma_t^2)] = -\beta[\ln(\sigma_t^2) - \alpha]dt + \psi dW_{2,t}, \quad (2.47)$$

em que μ , ψ , β e α denotam os parâmetros do processo e $\{W_{1,t}\}$ e $\{W_{2,t}\}$ são movimentos brownianos que satisfazem,

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} \begin{bmatrix} dW_{1,t} & dW_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \quad (2.48)$$

Na prática, claramente o processo de preço pode ser observado apenas em intervalos discretos de tempo. Entretanto, modelos em tempo contínuo como o das equações (2.46), (2.47) e (2.48) proporcionariam uma estrutura muito conveniente para a análise de vários aspectos teóricos relacionados ao apreçamento de ativos.

Um processo de difusão, de forma bastante geral, segundo Nelson (1990a), pode ser formalizado pela seguinte equação integral estocástica,

$$X_t = X_0 + \int_0^t \mu(X_s) ds + \int_0^t \Omega^{1/2}(X_s) dW_s, \quad (2.49)$$

na qual $\{W_t\}$ é um movimento browniano padrão $N \times 1$, $\mu(\cdot)$ e $\Omega^{1/2}(\cdot)$ são funções contínuas de \mathbb{R}^N em \mathbb{R}^N e o espaço das $N \times N$ matrizes reais, respectivamente. O valor inicial, $\{X_0\}$, pode ser fixo ou aleatório. Bollerslev et. al. (1994) escrevem que se a equação (2.49) possui uma solução única no sentido fraco, a distribuição do processo $\{X_t\}$ seria então completamente determinada pelas quatro caracterís-

ticas seguintes:

- i. a função de distribuição acumulada, $F(x_0)$, do ponto inicial X_0 ,
- ii. a tendência $\mu(x)$,
- iii. a matriz de variância condicional $\Omega(x) \equiv \Omega(x)^{1/2}[\Omega(x)^{1/2}]'$,
- iv. a continuidade, com probabilidade um, de $\{X_t\}$ como uma função do tempo.

Isto posto, a questão seria como aproximar (2.49) por um processo ARCH, ou vice-versa. Nelson (1990a) mostrou que se um modelo econômico especifica um modelo de difusão tal como o da equação (2.49), em que algumas variáveis de estado, inclusive $\{\Omega(x_t)\}$, são não observáveis, é possível formular um processo ARCH similar ao verdadeiro processo, no sentido que a distribuição das trajetórias amostrais geradas pelo processo ARCH e o modelo de difusão na equação (2.49) tornam-se arbitrariamente próximos para discretizações cada vez mais finas, isto é, para intervalos de tempo entre as realizações cada vez menores. Portanto, um processo ARCH discreto é uma aproximação adequada de um processo de difusão contínuo e que possui propriedades desejáveis, como exposto anteriormente.

2.7 ARCH: Interpretações e origens

Sem considerar sua relativa simplicidade, Bera & Higgins (1993) escrevem que a principal razão para o sucesso empírico de modelos da classe ARCH é que tais modelos conseguem captar muitas das regularidades empíricas observadas, como a leptocurticidade da distribuição dos retornos, o agrupamento da volatilidade, não-linearidades e variações na habilidade de prever-se valores futuros. Isto posto, não seria surpresa que tais modelos possam ser interpretados de diversas maneiras. Nesta seção serão apresentadas algumas destas interpretações.

2.7.1 Interpretações econômicas

Mizrach (1990) desenvolveu um modelo de apreçamento de ativos e aprendizagem no qual distúrbios ARCH são o resultado do problema de decisão dos agentes econômicos. Ele mostrou que os erros cometidos pelos agentes durante o processo de aprendizagem são altamente persistentes, e que os erros atuais são dependentes de todos os erros passados. Isto levaria a variância condicional a possuir a estrutura de um processo ARCH com longa defasagem. Isto pode ser relacionado com a conhecida *hipótese das expectativas adaptativas* em macroeconomia.

2.7.2 Origens

Stock (1988) estabeleceu uma conexão entre deformação temporal e modelos ARCH. Segundo Stock (1988), toda variável econômica, em geral, evolui em uma escala de tempo *operacional*, enquanto na prática as variáveis são medidas na escala de tempo do *calendário*. Este uso inapropriado da escala de tempo do calendário poderia levar ao agrupamento de volatilidade, dado que relativamente à escala de tempo do calendário, a variável poderia evoluir de forma mais rápida ou lenta.

Stock (1988) mostrou que um modelo com deformação temporal de uma variável aleatória $\{\varepsilon_t\}$ poderia ser aproximado por,

$$\varepsilon_t = \rho_t \varepsilon_{t-1} + \nu_t, \quad \nu_t | \Psi_{t-1} \sim N(0, \sigma_t^2), \quad (2.50)$$

em que $\sigma_t^2 = \omega + \alpha_1 \varepsilon_t^2$. Stock (1988) também mostrou que quando termina um período relativamente longo de tempo operacional durante uma unidade do tempo calendário, ρ_t é pequeno e σ_t^2 é grande, isto é, o parâmetro autorregressivo variante no tempo seria inversamente relacionado à variância condicional.

2.7.3 ARCH como Modelo não-linear

Conforme Bera & Higgins (1993), uma das características essenciais do modelo ARCH é $Cov(\varepsilon_t^2, \varepsilon_{t-j}^2) \neq 0$, mas $Cov(\varepsilon_t, \varepsilon_{t-j}) = 0$, para $j \neq 0$. Em palavras, modelos da classe ARCH postulam uma relação não-linear entre $\{\varepsilon_t\}$ e seus valores passados. Existem outros modelos não-lineares de séries temporais que também podem exibir tal propriedade, como o modelo bilinear ou o modelo autorregressivo *threshold* [ver Gouriéroux (1997)]. Por simplicidade, seguindo Bera & Higgins (1993), abordaremos o modelo bilinear e sua relação com o modelo ARCH.

Conforme Gouriéroux (1997), diz-se que uma série temporal $\{\varepsilon_t\}$ segue um modelo bilinear se esta satisfaz,

$$\varepsilon_t = \sum_{i=1}^p \phi_i \varepsilon_{t-i} + \sum_{j=1}^r \sum_{k=1}^s b_{jk} \varepsilon_{t-j} u_{t-k} + u_t, \quad (2.51)$$

onde u_t é uma sequência de variáveis aleatórias $IID(0, \sigma_u^2)$. Os primeiros dois momentos condicionais são,

$$E(\varepsilon_t | \Psi) = \sum_{i=1}^p \phi_i \varepsilon_{t-i} + \sum_{j=1}^r \sum_{k=1}^s b_{jk} \varepsilon_{t-j} u_{t-k},$$

$$Var(\varepsilon_t | \psi_{t-1}) = \sigma_u^2.$$

Tais momentos condicionais contrastam com aqueles de um processo da classe ARCH, nos quais a média condicional é, em geral, uma constante, mas a variância condicional varia no tempo. Os momentos incondicionais, no entanto, podem ser similares. Por exemplo, o modelo bilinear,

$$\varepsilon_t = b_{21} \varepsilon_{t-2} u_{t-1} + u_t,$$

tem $E(\varepsilon_t) = 0$ e $Cov(\varepsilon_t^2, \varepsilon_{t-2}^2) = b_{21}^2 \sigma_u^2$. Assim, este processo exibe o agrupamento de grandes e pequenos desvios como processos da classe ARCH. Bera

& Higgins (1993) escrevem que um modelo bilinear é bastante similar a um modelo ARCH, posto que ambos podem ser reescritos como modelos de parâmetros aleatórios para ε_t . De fato, um modelo ARCH(p) pode ser representado por um modelo MA(p) de parâmetros aleatórios para ε_t ,

$$\varepsilon_t = \omega + \alpha(L)\zeta_{t-1}\varepsilon_{t-1}, \quad (2.52)$$

no qual $\{\zeta_t\}$ é um processo estocástico escalar i.i.d com média zero e variância unitária, enquanto (2.51) pode ser reescrita como,

$$\varepsilon_t = \sum_{j=1}^m [\phi_j + A_j(t)]\varepsilon_{t-j} + u_t, \quad (2.53)$$

sendo $m = \max(p, r)$, e $A_j(t) = \sum_{k=1}^s b_{jk}u_{t-k}$, com $\phi = 0$, $i \geq p + 1$, $b_{jk} = 0$, $j \geq r + 1$.

Conforme Bera & Higgins (1993), há uma outra forma de observar-se as similaridades e diferenças entre os modelos ARCH e bilinear. Apesar de ambos os modelos captarem dependências não-lineares, um modelo ARCH representa isto de forma multiplicativa, enquanto o modelo bilinear postula uma estrutura aditiva.

Para finalizar, deve-se mencionar, conforme Bera & Higgins (1993), um problema inerente à utilização de uma especificação não-linear para a média condicional na modelagem de séries financeiras. Bera & Higgins (1993) escrevem que para um modelo de média condicional não-linear explicar o tipo de variabilidade observada na prática, a variação na média condicional deve ser expressiva, implicando grandes oportunidades de lucros não exploradas pelos agentes, o que constitui um fato pouco provável. Assim, por esta razão, modelos não-lineares na média não se tornaram tão populares na análise de séries financeiras. Os modelos ARCH não teriam esta limitação, posto que variações na volatilidade são representadas por variações na variância condicional, ligando a volatilidade a uma medida natural de risco.

2.8 ARCH: Extensões

Conforme Bera & Higgins (1993), na exposição original do processo ARCH, foi natural para Engle (1982) assumir que o processo da variância condicional fosse linear nos erros quadráticos e que a distribuição condicional fosse a normal. Entretanto, o próprio Robert F. Engle admitia que as hipóteses de linearidade e normalidade condicional poderiam não ser apropriadas em algumas aplicações. Trabalhos subsequentes foram preenchendo as deficiências do modelo ARCH original, dando origem a diversas extensões do modelo. Nesta seção, serão analisadas algumas formulações alternativas para o processo da variância condicional que têm mostrado-se úteis na pesquisa aplicada.

2.8.1 IGARCH

Conforme Gouriéroux (1997), na estrutura usual dos processos ARMA para a média condicional, a condição de estacionariedade é caracterizada pelas raízes do polinômio autorregressivo, requerendo-se que o módulo de tais raízes seja estritamente menor que um. Tornou-se uma questão extremamente importante a investigação do caso limite, em que as raízes podem ter módulo um. Desta linha de pesquisa surgiram os denominados processos não-estacionários, ou processos ARIMA (I para *integrado*), que generalizam o conceito do processo passeio aleatório. As mesmas considerações foram implementadas para a variância condicional, i.e, para a classe de modelos ARCH, produzindo, entretanto, resultados significativamente diferentes.

De forma análoga às questões concernentes à modelagem adequada das dependências de longo prazo na média condicional de séries temporais econômicas, questões similares tornaram-se relevantes na modelagem da variância condicional. Isto porque, nas aplicações empíricas com dados financeiros de alta frequência, a estimativa de $\alpha(1) + \beta(1)$ para o modelo GARCH são frequentemente muito próximas de um. Isto forneceu a motivação empírica para a elaboração do modelo

GARCH *integrado* ou IGARCH por Engle & Bollerslev (1986), o qual possui algumas características dos processos de raiz unitária, ou integrados de ordem 1, I(1), para a média. Contudo, conforme Baillie et. al. (1996), a analogia está longe de ser perfeita.

Como visto anteriormente, um processo GARCH(p, q) pode ser expresso como um processo ARMA(m, p) em $\{\varepsilon_t^2\}$,

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \omega + [1 - \beta(L)]\nu_t$$

sendo $m \equiv \max(p, q)$ e $\{\nu_t \equiv \varepsilon_t^2 - \sigma_t^2\}$. Quando o polinômio autorregressivo $[1 - \alpha(L) - \beta(L)]$ possui uma raiz unitária, o processo GARCH(p, q) foi definido por Engle & Bollerslev (1986) como sendo integrado na variância. Conforme Baillie et. al. (1996), o processo IGARCH(p, q) pode ser sucintamente expresso por,

$$\phi(L)(1 - L)\varepsilon_t^2 = \omega + [1 - \beta(L)]\nu_t, \quad (2.54)$$

em que $\phi(L) \equiv [1 - \alpha(L) - \beta(L)](1 - L)^{-1}$ tem ordem $(m - 1)$. Por simplicidade, seguindo Nelson (1990a), considera-se o processo GARCH(1,1) para analisar brevemente a propriedade relevante do processo IGARCH(1,1). Sob a hipótese de normalidade condicional, o processo é,

$$\begin{aligned} \varepsilon_t &\sim N(0, \sigma_t^2) \\ \sigma_t^2 &= \omega + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2, \quad \alpha \geq 0, \beta \geq 0, \\ \alpha_1 + \beta_1 &= 1. \end{aligned}$$

Conforme Nelson (1990a), as previsões para a variância condicional para diferentes horizontes tomam a seguinte forma,

$$E(\sigma_{t+k}^2 | \varepsilon_{t-1}) = (\alpha_1 + \beta_1)^k \sigma_t^2 + \omega \left(\sum_{i=0}^{k-1} (\alpha_1 + \beta_1)^i \right).$$

Como visto na seção (2.6.1), se $\alpha_1 + \beta_1 < 1$, o processo $\{\varepsilon_t\}$ é fracamente

estacionário, e um choque sobre a variância condicional $\{\sigma_{t+k}^2\}$ decai quando k aumenta, sendo assintoticamente negligível. Isto é, $E(\sigma_{t+k}^2|\varepsilon_{t-1})$ converge para a variância incondicional igual a $(\omega/(1 - \alpha_1 - \beta_1))$ quando $t \rightarrow \infty$ se, e somente se, $\alpha_1 + \beta_1 < 1$.

Se $\alpha_1 + \beta_1 \geq 1$, o efeito sobre $\{\sigma_{t+k}^2\}$ não decai assintoticamente. Trata-se da propriedade denominada *persistência* que foi analisada na seção (2.6.4). No caso particular do modelo IGARCH(1,1) com $\omega > 0$ e $\alpha_1 + \beta_1 = 1$, tem-se,

$$E(\sigma_{t+k}^2|\varepsilon_t) = \sigma_t^2 + k\omega.$$

Nelson (1990a) mostrou que $E(\sigma_{t+k}^2|\varepsilon_t) \rightarrow \infty$ quase-certamente quando $t \rightarrow \infty$. Ainda assim, Nelson (1990a) mostrou que o processo IGARCH(1,1) é estacionário e ergódico. Ou seja, o processo IGARCH com $\omega > 0$ impõe propriedades para as previsões de $\{\sigma_t\}$ iguais à de um passeio aleatório com tendência, entretanto o processo IGARCH é estacionário e ergódico. Este é um exemplo claro das diferenças concernentes à questão da não-estacionariedade entre modelos para a média condicional e para a variância condicional.

2.8.2 FIGARCH

Ainda que o processo GARCH seja flexível e tenham sido propostas várias extensões deste modelo de forma a incluir algumas características particulares verificadas em algumas séries econômicas como assimetrias e mudanças de regime, processos GARCH não captam adequadamente algumas regularidades encontradas concernentes à persistência ou memória longa na volatilidade do retorno de algumas séries econômicas.

Como exposto na seção anterior, a análise de questões relativas à modelagem apropriada de dependências de longo prazo na média condicional de séries econômicas foi estendida para o âmbito da variância condicional. Esta linha de pesquisa levou Engle & Bollerslev (1986) à formulação do processo GARCH in-

tegrado ou IGARCH, o qual possui algumas características dos processos de raiz unitária ou $I(1)$. Como visto, o efeito de um choque sobre a previsão ótima da variância condicional futura faz com que os correspondentes pesos de resposta à impulsos (choques) acumulados tendam à uma constante diferente de zero, tal que as previsões aumentam linearmente com o horizonte de previsão.

Conforme Baillie et. al. (1996), isto implicaria por exemplo, que o apreçamento de ativos com risco, incluindo opções de longo prazo e contratos futuros, poderia exibir uma dependência extrema das condições iniciais, ou do estado atual da economia. Contudo, Baillie et. al. (1996) escrevem que este grau extremo de dependência parece contrário ao comportamento observado. Além disso, diversos estudos como o de Crato & Lima (1994) e Ding et. al. (1993) reportaram fortes evidências de memória longa nas autocorrelações de retornos quadráticos ou absolutos de vários ativos, como descrito na seção (2.4.4).

Motivados por esses fatos, Baillie et. al. (1996) formularam o processo GARCH *fracionalmente integrado* ou processo FIGARCH. De forma a introduzir o processo FIGARCH, é importante rever brevemente os principais fatos relativos aos processos fracionalmente integrados para a média. Conforme Baillie et. al. (1996), os conceitos de memória longa e movimento browniano fracionário foram introduzidos originalmente por Hurst (1951) e Mandelbrot & Ness (1968), contudo, as idéias foram realmente operacionalizadas para as aplicações com processos em tempo discreto por Hosking (1981) e Granger (1980).

É um fato bem conhecido que, para processos ARMA estacionários, a função de autocorrelação tende a exibir um decaimento exponencial. Em contraste, um processo com memória longa deve implicar em uma função de autocorrelação que tende a exibir um "lento" decaimento. De forma a obter um processo com memória longa, Hosking (1981) e Granger (1980) incorporaram ao processo ARMA(p, q) o *operador de diferenciação fracionária* $(1 - L)^d$, sendo $0 \leq d \leq 1$, em que L denota o operador de defasagens. O processo resultante é denominado *processo autorregressivo fracionalmente integrado médias-móveis* ou pro-

cesso ARFIMA(p, d, q). Especificamente, seja $\{X_t\}$ um processo estocástico em tempo discreto, um processo ARFIMA(p, d, q) para $\{X_t\}$ é definido por,

$$\phi(L)(1 - L)^d X_t = \theta(L)\varepsilon_t, \quad (2.55)$$

em que $\phi(L)$ e $\theta(L)$ são polinômios no operador de defasagem L , com ordens p e q , respectivamente, e $\{\varepsilon_t\}$ um ruído branco. Hosking (1981) mostrou que a função de autocorrelação de um processo ARFIMA decresce hiperbolicamente para $0 < d < 1$, que é um padrão significativamente mais lento que o padrão de um processo ARMA. Baillie et. al. (1996) escrevem que o operador de diferenciação fracionária possui uma expansão binomial que é expressa de forma mais conveniente em termos da função hipergeométrica, ou seja,

$$\begin{aligned} (1 - L)^d &= F(-d, 1, 1; L), \\ &= \sum_{k=0}^{\infty} \Gamma(k - d)\Gamma(k + 1)^{-1}\Gamma(-d)^{-1}L^k, \\ &= \sum_{k=0}^{\infty} \pi_k L^k, \end{aligned} \quad (2.56)$$

$\Gamma(\cdot)$ denotando a função gama. A função hipergeométrica pode ser formalmente definido por,

$$F(m, n, s; x) = \Gamma(s)\Gamma(m)^{-1}\Gamma(n)^{-1} \sum_{j=0}^{\infty} \Gamma(m+j)\Gamma(n+j)\Gamma(s+j)^{-1}\Gamma(j+1)^{-1}x^j.$$

A estacionariedade fraca e invertibilidade do processo são obtidas restringindo d ao intervalo $-0,5 < d < 0,5$, implicando que o processo terá representações médias móveis e autorregressivas únicas. Entretanto, para $d > 0$, diz-se que o processo possui memória longa no sentido que,

$$\lim_{k \rightarrow \infty} \sum_{j=-k}^k |\rho_j|$$

não converge para um limite finito, em que ρ_j denota a autocorrelação do processo

no lag k . Essencialmente, diz-se que um modelo ARFIMA impõe uma dicotomia entre a dinâmica de curto e longo prazos, modelando o comportamento de curto prazo através da estrutura ARMA tradicional, enquanto a dinâmica de longo prazo é modelada pelo parâmetro de diferenciação fracionária, d .

Um processo ARMA estacionário e invertível tem autocorrelações que são geometricamente limitadas, isto é, $|\rho_j| \leq cm^{-k}$, para $k \rightarrow \infty$, com $0 < m < 1$, sendo portanto um processo com memória curta, enquanto, para um processo ARFIMA(p, d, q) estacionário, verifica-se a seguinte relação à medida que $k \rightarrow \infty$,

$$\rho_k \approx ck^{2d-1} \quad (2.57)$$

assim, a função de autocorrelação é aproximada por uma constante $c > 0$, que pode depender dos parâmetros do modelo, mas não de k , e por um termo hiperbólico. Portanto, a convergência para zero destas autocorrelações é significativamente mais lenta que a de um processo ARMA.

Isto posto, o processo GARCH fracionalmente integrado foi introduzido por Baillie et. al. (1996) para capturar a memória longa na variância condicional. A partir da formulação padrão de um processo GARCH(p, q),

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_{t-i}^2 + \beta(L)\sigma_{t-j}^2,$$

que, como visto, pode ser rescrita como um modelo ARMA para os erros quadráticos, ou seja,

$$\phi(L)(1 - L)\varepsilon_t^2 = \omega + [1 - \beta(L)]\nu_t,$$

em que $\phi(L) \equiv [1 - \alpha(L) - \beta(L)](1 - L)^{-1}$ tem ordem $(m - 1)$, o processo FIGARCH(p, d, q) é obtido pela substituição do operador de diferenças $(1 - L)$ pelo operador de diferenciação fracionária $(1 - L)^d$; assim,

$$\phi(L)(1 - L)^d \varepsilon_t^2 = \omega + [1 - \beta(L)]\nu_t \quad (2.58)$$

com $0 < d < 1$, e todas as raízes de $\phi(L)$ e $[1 - \beta(L)]$ estão fora do círculo

unitário.

Conforme Baillie et. al. (1996), rearranjando os termos da equação (2.58) e utilizando o fato que $\{\nu_t \equiv \varepsilon_t^2 - \sigma_t^2\}$, uma representação alternativa para o processo FIGARCH(p, d, q) é,

$$[1 - \beta(L)]\sigma_t^2 = \omega + [1 - \beta(L) - \phi(L)(1 - L)^d]\varepsilon_t^2. \quad (2.59)$$

Assim, a variância condicional de $\{\varepsilon_t\}$ é dada por,

$$\begin{aligned} \sigma_t^2 &= \omega[1 - \beta(1)]^{-1} + \{1 - [1 - \beta(L)]^{-1}\phi(L)(1 - L)^d\}\varepsilon_t^2, \\ &= \omega[1 - \beta(1)]^{-1} + \lambda(L)\varepsilon_t^2, \end{aligned} \quad (2.60)$$

em que $\lambda(L)$ é um polinômio em L . Esta equação é denominada representação ARCH infinita. Para que o processo FIGARCH(p, d, q) na equação (2.58) seja bem definido e a variância condicional seja positiva quase-certamente $\forall t$, todos os coeficientes na Equação (2.60) devem ser não-negativos, i.e, $\lambda \geq 0$ para $\{k = 1, 2, \dots\}$. Entretanto, Baillie et. al. (1996) postulam que os resultados reportados por Nelson & Cao (1992) quanto à necessidade de se imporem tais restrições na estimação do modelo GARCH extenderiam-se ao modelo FIGARCH.

Segundo Baillie et. al. (1996), para $0 < d \leq 1$, a função hipergeométrica calculada em $L = 1$ é igual à $F(-d, 1, 1; 1) = 0$, tal que $\lambda(1) = 1$. Assim o termo $\omega > 0$ teria a mesma interpretação válida para o modelo IGARCH. Consequentemente, o segundo momento da distribuição incondicional de ε_t seria infinito. Isto posto, o processo FIGARCH claramente não é fracamente estacionário, característica compartilhada pelo modelo IGARCH. Baillie et. al. (1996) estendem os resultados obtidos por Nelson (1990b) e Bougerol & Picard (1992) para o modelo IGARCH ao modelo FIGARCH. Ou seja, apesar de não ser fracamente estacionário, o processo FIGARCH, assim como o IGARCH, seria fortemente estacionário e ergódico.

2.8.3 APARCH

Uma das principais limitações dos modelos ARCH vistos anteriormente, é que o impacto de choques sobre a volatilidade é simétrico nestes modelos. Assim, choques positivos e negativos têm o mesmo impacto sobre a variância condicional (ou volatilidade). Conforme Issler (1999), dado que a maioria das aplicações dos modelos ARCH ocorre em Finanças, e para estes dados é pouco provável que choques positivos e negativos tenham o mesmo impacto, seria desejável formular modelos que permitissem a estimação e teste da assimetria.

Isto posto, diversos modelos nesta linha foram propostos, como o modelo EGARCH de Nelson (1991) e o TARARCH de Zakoian (1994). Um dos modelos mais promissores nesta linha é o modelo ARCH com potência assimétrica (APARCH) proposto por Ding et. al. (1993). O modelo APARCH (p, q) pode ser representado por,

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^p \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^q \beta_j \sigma_{t-j}^\delta, \quad (2.61)$$

em que $\alpha_0 > 0$, $\delta \geq 0$, $\alpha_i, \beta_j \geq 0$ e $-1 < \gamma_i < 1$. Este modelo concilia a flexibilidade de um expoente variante com o coeficiente de assimetria; desta forma, este modelo é capaz de detectar impactos assimétricos de choques sobre a volatilidade. Especificamente, se $\gamma > 0$, verifica-se o efeito alavancagem, isto é, choques negativos têm um impacto maior sobre a volatilidade da série que choques positivos. Se $\gamma < 0$, verifica-se o contrário.

2.9 ARCH: Estimação

A abordagem baseada na verossimilhança é o procedimento mais utilizado na estimação de modelos ARCH, e será adotada nesta dissertação. No entanto, outras abordagens como métodos não paramétricos, semi - paramétricos e bayesianos foram propostas.

2.9.1 Máxima verossimilhança com ε_t gaussiano

Suponha que deseja-se estimar os parâmetros de um modelo de regressão ARCH, com especificação dada por,

$$\begin{aligned}
 \varepsilon_t &= y_t - x_t' \xi, \\
 \varepsilon_t &= z_t \sqrt{\sigma_t^2}, \\
 \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2, \\
 &= \alpha_0 + \alpha_1 (y_{t-1} - x_{t-1}' \xi) + \dots + \alpha_p (y_{t-p} - x_{t-p}' \xi), \\
 &= Z_t(\xi)' \alpha,
 \end{aligned} \tag{2.62}$$

sendo,

$$\begin{aligned}
 \alpha &\equiv (\alpha_0, \alpha_1, \dots, \alpha_p) \\
 Z_t(\xi)' &\equiv (1, y_{t-1} - x_{t-1}' \xi, y_{t-2} - x_{t-2}' \xi, \dots, y_{t-p} - x_{t-p}' \xi),
 \end{aligned}$$

assumindo $z_t \sim i.i.d. N(0, 1)$, com z_t independente de Ψ_{t-1} , a distribuição condicional de y_t é gaussiana com média $x_t' \xi$ e variância σ_t^2 ,

$$f(y_t | \Psi_{t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right). \tag{2.63}$$

É conveniente colocar os parâmetros a serem estimados em um vetor θ de dimensão $(a \times 1)$,

$$\theta \equiv (\xi', \alpha')'.$$

Isto posto, a função de log verossimilhança condicional às primeiras m observações é,

$$L_T(\theta) = \sum_{t=1}^T \ln(f(y_t | \Psi_{t-1}; \theta)) = \sum_{t=1}^T l_t(\cdot), \tag{2.64}$$

em que $l_t(\cdot)$ é a contribuição individual da verossimilhança, especificamente,

$$L_T(\theta) = -\left(\frac{T}{2}\right)\ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^T \left(\frac{\varepsilon_t^2}{\sigma_t^2}\right). \quad (2.65)$$

Para um dado valor numérico do vetor de parâmetros θ , a sequência de variâncias condicionais pode ser calculada por (2.62), sendo então utilizadas para calcular a função de verossimilhança condicional (2.65).

Uma característica atraente da função de verossimilhança gaussiana é que a matriz de informação é bloco diagonal entre os vetores de parâmetros ξ e α . O Teorema 5 apresentado em Engle (1982) fundamenta esta propriedade.

Teorema 5 (Engle(1982)) *Se um modelo de regressão ARCH é simétrico e regular, então a matriz de informação é bloco diagonal entre ξ e α*

Prova, ver Engle (1982).

Para verificar isto, conforme Bera & Higgins (1993), note-se que o (i, j) -ésimo elemento do bloco fora da diagonal principal da matriz de informação pode ser escrito como,

$$\frac{1}{T} \sum_{t=1}^T E\left(\frac{\partial^2 L_t}{\partial \xi_i \partial \alpha_j}\right) = \frac{1}{T} \sum_{t=1}^T E\left(\frac{1}{2(\sigma_t^2)^2} \frac{\partial \sigma_t^2}{\partial \xi_i} \frac{\partial \sigma_t^2}{\partial \alpha_j}\right). \quad (2.66)$$

Se σ_t^2 é um processo simétrico no sentido da definição de Engle (1982), a última expressão entre parênteses é anti-simétrica e, portanto, tem esperança zero. Os modelos simétricos no sentido da Definição 7, como o modelo GARCH, compartilham desta propriedade e, conforme Bera & Higgins (1993), o mesmo não se verifica para os modelos assimétricos como o EGARCH de Nelson (1991). A importância desta propriedade é que, no contexto da abordagem baseada na verossimilhança, a estimação de ξ e α pode ser implementada separadamente sem perda assintótica de eficiência e suas variâncias podem ser calculadas separadamente (ver Engle (1982) e Bollerslev (1986)).

O *score* para o modelo de regressão ARCH pode ser calculado a partir de,

$$s_t(\theta) = \frac{\partial L(\theta)}{\partial \theta} = -\frac{1}{2} \frac{\partial \ln(\sigma_t^2)}{\partial \theta} - \frac{1}{2} \left(\frac{1}{\sigma_t^2} \frac{\partial \varepsilon_t^2}{\partial \theta} - \frac{\varepsilon_t^2}{(\sigma_t^2)^2} \frac{\partial \sigma_t^2}{\partial \theta} \right), \quad (2.67)$$

mas,

$$\frac{\partial \varepsilon_t^2}{\partial \theta} = \begin{bmatrix} -2x_t \varepsilon_t \\ 0 \end{bmatrix}, \quad (2.68)$$

e,

$$\begin{aligned} \frac{\partial \sigma_t^2}{\partial \theta} &= \frac{\partial(\alpha_0 + \sum_{j=1}^m \alpha_j \varepsilon_{t-j}^2)}{\partial \theta}, \\ &= \frac{\partial \alpha_0}{\partial \theta} + \sum_{j=1}^m \left(\frac{\partial \alpha_j}{\partial \theta} \right) \varepsilon_{t-j}^2 + \sum_{j=1}^m \alpha_j \left(\frac{\partial \varepsilon_{t-j}^2}{\partial \theta} \right), \\ &= \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \varepsilon_t^2 \\ \vdots \\ 0 \end{bmatrix} + \dots + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \varepsilon_{t-m}^2 \end{bmatrix} + \sum_{j=1}^m \alpha_j \begin{bmatrix} -2\varepsilon_{t-j} x_{t-j} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \\ &= \begin{bmatrix} \sum_{j=1}^m -2\alpha_j \varepsilon_{t-j} x_{t-j} \\ Z_t(\xi) \end{bmatrix}, \end{aligned} \quad (2.69)$$

substituindo (2.68) e (2.69) em (2.67), tem-se que a função *score* é dada por,

$$\frac{\partial L(\theta)}{\partial \theta} = -\left(\frac{1}{2\sigma_t^2} - \frac{\varepsilon_t^2}{2(\sigma_t^2)^2} \right) \begin{bmatrix} \sum_{j=1}^m -2\alpha_j \varepsilon_{t-j} x_{t-j} \\ Z_t(\xi) \end{bmatrix} + \begin{bmatrix} (x_t \varepsilon_t) / \sigma_t^2 \\ 0 \end{bmatrix}. \quad (2.70)$$

Engle (1982) utilizou o método de *scoring* para maximizar a função de verossimilhança condicional. Um algoritmo bastante utilizado nas pesquisas aplicadas é o proposto por Berndt et. al. (1974). A partir de valores iniciais da r -ésima iteração, o $(r + 1)$ -ésimo passo do algoritmo pode ser escrito como,

$$\xi^{(r+1)} = \xi^{(r)} + \left[\sum_{t=1}^T \left(\frac{\partial L_t}{\partial \xi} \right) \left(\frac{\partial L_t}{\partial \xi} \right)' \right]^{-1} \sum_{t=1}^T \frac{\partial L_t}{\partial \xi}, \quad (2.71)$$

e

$$\alpha^{(r+1)} = \alpha^{(r)} + \left[\sum_{t=1}^T \left(\frac{\partial L_t}{\partial \alpha} \right) \left(\frac{\partial L_t}{\partial \alpha} \right)' \right]^{-1} \sum_{t=1}^T \frac{\partial L_t}{\partial \alpha}. \quad (2.72)$$

A estimação do modelo de regressão GARCH(p, q) é bastante semelhante à do modelo de regressão ARCH(p); as diferenças surgem devido à inclusão de variâncias condicionais passadas, i.e σ_{t-i}^2 , no processo da variância condicional. Conforme Bollerslev (1986), sejam,

$$\begin{aligned} z'_t &= \{1, \varepsilon_{t-1}^2, \dots, \varepsilon_{t-p}^2, \sigma_{t-1}^2, \dots, \sigma_{t-q}^2\}, \\ \varpi' &= \{\alpha_0, \dots, \alpha_p, \beta_1, \dots, \beta_q\}, \end{aligned}$$

e $\theta \in \Theta$, onde $\{\theta = (\xi', \varpi')\}$ e Θ é um subespaço compacto de um espaço euclidiano tal que ε_t possua segundos momentos finitos. Assim, pode-se reescrever o modelo GARCH(p, q) como,

$$\begin{aligned} \varepsilon_t &= y_t - x'_t \xi, \\ \varepsilon_t | \Psi_{t-1} &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= z'_t \varpi'. \end{aligned}$$

Assumindo normalidade condicional para ε_t , a função de log verossimilhança condicional para o modelo de regressão GARCH é dada por,

$$l_T(\theta) = \frac{1}{T} \sum_{t=1}^T l_t(\theta), \quad (2.73)$$

$$l_t(\theta) = -\frac{1}{2} \ln \sigma_t^2 - \frac{1}{2} \frac{\varepsilon_t^2}{\sigma_t^2}, \quad (2.74)$$

obtém-se o *score* em relação à σ_t^2 fazendo,

$$\frac{\partial l_t}{\partial \varpi} = \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \varpi} \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right), \quad (2.75)$$

o hessiano é,

$$\frac{\partial^2 l_t}{\partial \omega \partial \omega'} = \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \frac{\partial}{\partial \omega'} \left(\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \omega} \right) - \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \omega} \frac{\partial \sigma_t^2}{\partial \omega'} \frac{\varepsilon_t^2}{\sigma_t^2}, \quad (2.76)$$

em que,

$$\frac{\partial \sigma_t^2}{\partial \omega} = z_t + \sum_{j=1}^q \beta_j \frac{\partial \sigma_{t-j}^2}{\partial \omega}, \quad (2.77)$$

a parte recursiva expressa na equação (2.77) é a única diferença em relação à estimação do modelo ARCH original de Engle (1982).

O *score* em relação a ξ é dado por,

$$\frac{\partial l_t}{\partial \xi} = \frac{\varepsilon_t \xi_t}{\sigma_t^2} + \frac{1}{2} \sigma_t^2 \frac{\partial \sigma_t^2}{\partial \xi} \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right), \quad (2.78)$$

sendo o hessiano dado por,

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \xi \partial \xi'} = & -\frac{\xi_t \xi_t'}{\sigma_t^2} - \frac{1}{2(\sigma_t^2)^2} \frac{\partial \sigma_t^2}{\partial \xi} \frac{\partial \sigma_t^2}{\partial \xi'} \left(\frac{\varepsilon_t^2}{\sigma_t^2} \right) - \frac{2\varepsilon_t \xi_t}{(\sigma_t^2)^2} \frac{\partial \sigma_t^2}{\partial \xi} \\ & + \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \frac{\partial}{\partial \xi'} \left(\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \xi} \right), \end{aligned} \quad (2.79)$$

onde,

$$\frac{\partial \sigma_t^2}{\partial \xi} = -2 \sum_{i=1}^p \alpha_i \xi_{t-i} \varepsilon_{t-i} + \sum_{j=1}^q \beta_j \frac{\partial \sigma_{t-j}^2}{\partial \xi}. \quad (2.80)$$

Conforme Bollerslev (1986), novamente, a única diferença em relação ao modelo ARCH original de Engle (1982) é a relação recursiva presente na expressão (2.80). Posto que o processo GARCH é simétrico e regular no sentido da Definição 6, os elementos no bloco fora da diagonal principal da matriz de informação são zero. Dada esta independência assintótica, α pode ser estimado sem perda de eficiência assintótica, baseado em uma estimativa consistente de ξ , e vice-versa.

Devido à presença dos termos recursivos mencionados, Bollerslev (1986) utilizou o método proposto por Berndt et. al. (1974); neste caso, seja $\theta^{(i)}$ a estimativa

do vetor de parâmetros depois da i -ésima iteração, $\theta^{(i+1)}$ é calculada a partir de,

$$\theta^{(i+1)} = \theta^{(i)} + \lambda_i \left(\sum_{t=1}^T \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'} \right)^{-1} \sum_{t=1}^T \frac{\partial l_t}{\partial \theta}, \quad (2.81)$$

em que λ_i é um *step length* variável escolhido para maximizar a função de log verossimilhança em uma dada direção. Apesar de não explicitado, as iterações para $\xi^{(i)}$ e $\varpi^{(i)}$ podem ser implementadas separadamente dada a bloco diagonalidade da matriz de informação.

Alternativamente, conforme Hamilton (1994), o gradiente da função de log verossimilhança pode ser calculado numericamente (ou analiticamente) pela diferenciação numérica da log verossimilhança. Assim, o gradiente calculado numericamente pode ser utilizado em conjunção com um procedimento de otimização.

2.9.2 Quasi-máxima verossimilhança

Para a maioria das aplicações, é difícil justificar a hipótese da normalidade condicional. Assim, a função de log verossimilhança condicional pode estar erroneamente especificada. Entretanto, pode-se ainda obter estimadores consistentes dos parâmetros, tais estimadores são usualmente denominados na literatura como estimadores de *quasi máxima verossimilhança*, EQMV. Contudo, os erros padrão destes estimadores devem ser corrigidos.

Weiss (1986) foi o primeiro a estudar as propriedades assintóticas dos estimadores de quasi máxima verossimilhança de modelos da classe ARCH. O importante resultado a que Weiss (1986) chegou foi que na medida em que os dois primeiros momentos forem corretamente especificados, ξ e ϖ podem ser consistentemente estimados, ainda que a hipótese de normalidade condicional seja violada.

Para estabelecer a distribuição assintótica dos EQMV, seja $\hat{\theta} = (\hat{\xi}', \hat{\varpi}')$ o EQMV e seja θ o verdadeiro valor do parâmetro. Então, sob certas condições de

regularidade,

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, D^{-1}SD^{-1}), \quad (2.82)$$

S representando o Hessiano e D o produto externo do gradiente, com,

$$S = \text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T [s_t(\theta)][s_t(\theta)]', \quad (2.83)$$

$s_t(\theta)$ sendo o vetor *score* expresso em (2.67).

A matriz D é dada por,

$$\begin{aligned} D &= \text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T -E \left(\frac{\partial s_t(\theta)}{\partial \theta'} \middle| \Psi_{t-1} \right), \\ &= \text{plim}_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \left([1/2(\sigma_t^2)^2] \begin{bmatrix} \sum_{j=1}^m -2\alpha_j \varepsilon_{t-j} x_{t-j} \\ z_t(\xi) \end{bmatrix} \right. \\ &\quad \left. \times \begin{bmatrix} \sum_{j=1}^m -2\alpha_j \varepsilon_{t-j} x_{t-j} [z_t(\xi)]' \\ + (1/\sigma_t^2) \begin{bmatrix} x_t x_t' & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix} \right). \end{aligned} \quad (2.84)$$

Segundo Hamilton (1994), a matriz S pode ser consistentemente estimada por,

$$\hat{S}_T = T^{-1} \sum_{t=1}^T [s_t(\hat{\theta})][s_t(\hat{\theta})]', \quad (2.85)$$

de forma similar, a matriz D pode ser consistentemente estimada por,

$$\begin{aligned} \hat{D}_T &= T^{-1} \sum_{t=1}^T \left([1/2(\hat{\sigma}_t^2)^2] \begin{bmatrix} \sum_{j=1}^m -2\hat{\alpha}_j \hat{\varepsilon}_{t-j} x_{t-j} \\ z_t(\hat{\xi}) \end{bmatrix} \right. \\ &\quad \left. \times \begin{bmatrix} \sum_{j=1}^m -2\hat{\alpha}_j \hat{\varepsilon}_{t-j} x_{t-j}' [z_t(\hat{\xi})]' \\ + (1/\hat{\sigma}_t^2) \begin{bmatrix} x_t x_t' & 0 \\ 0 & 0 \end{bmatrix} \end{bmatrix} \right). \end{aligned} \quad (2.86)$$

Isto posto, os erros padrão de $\hat{\theta}$ que são robustos à especificação errônea da família de densidades podem ser calculados a partir da raiz quadrada dos elementos da diagonal principal de,

$$T^{-1} \hat{D}_T^{-1} \hat{S}_T \hat{D}_T^{-1}.$$

Caso o processo gerador de dados seja realmente gaussiano, ocorre que $S = D$, reduzindo-se à matriz de variância assintótica usual no contexto da estimação por máxima verossimilhança. Bollerslev & Wooldridge (1992) generalizaram os resultados de Weiss (1986) para o modelo ARCH univariado ao modelo GARCH multivariado, sob um conjunto menos restritivo de condições de regularidade.

2.9.3 Máxima verossimilhança com ϵ_t não-gaussiano

Como visto na seção anterior, segundo Weiss (1986) e Bollerslev & Wooldridge (1992), caso esteja-se interessado somente nos dois primeiros momentos condicionais, a hipótese de normalidade condicional é justificável, posto que o estimador de quasi máxima verossimilhança é consistente, se a média e a variância condicionais forem corretamente especificadas.

Entretanto, Engle & Gonzales-Rivera (1991) mostraram que tal estimador é ineficiente, com o grau de ineficiência aumentando com o grau de desvio da normalidade. Isto posto, a busca por uma distribuição adequada torna-se muito relevante em termos de ganho de eficiência. Além disso, do ponto de vista das aplicações, a modelagem correta da curtose e assimetria é importante nas aplicações financeiras.

A utilização da distribuição t-Student padronizada é muito comum Bollerslev (1987), entre outros, mostrou que esta distribuição captura de forma mais adequada a curtose observada. Se z_t é distribuída como uma $t(0,1,\nu)$, em que ν denota os

graus de liberdade, a função de log verossimilhança torna-se,

$$L_T(\theta) = T \left[\ln \left(\Gamma \left(\frac{\nu + 1}{2} \right) \right) - \ln \left(\Gamma \left(\frac{\nu}{2} \right) \right) - 0,5 \ln(\pi(\nu - 2)) \right] - 0,5 \sum_{i=1}^T \left[\ln(\sigma_i^2) + (1 + \nu) \ln \left(1 + \frac{z_i^2}{\nu - 2} \right) \right], \quad (2.87)$$

o parâmetro ν controla o quão pesadas são as caudas, a leptocurticidade é decrescente com ν . A distribuição t é simétrica em torno de zero, entretanto para $4 < \nu < \infty$, a curtose condicional é igual a $3(\nu - 2)(\nu - 4)$, que excede a curtose da normal. Conforme Issler (1999), utilizando-se a distribuição t -Student, o procedimento de estimação selecionará o número de graus de liberdade que melhor ajusta-se aos dados, não eliminando a possibilidade de leptocurtose *a priori*. Assim, uma estimativa relativamente grande de ν poderia indicar que a normal seria uma aproximação adequada.

Nelson (1991) propôs o uso da distribuição exponencial generalizada (DEG) ou distribuição do erro generalizada. Sua função de densidade, para uma variável aleatória normalizada, pode ser representada por,

$$f(z_t(\theta); \nu) = \nu \lambda^{-1} 2^{-(1+1/\nu)} \Gamma(\nu^{-1})^{-1} \exp(-0,5 |z_t(\theta) \lambda^{-1}|^\nu), \quad (2.88)$$

em que $\lambda = (2^{(-2/\nu)} \Gamma(\nu^{-1}) \Gamma(3\nu^{-1}))^{1/2}$. Novamente, ν controla o peso das caudas. Para $\nu = 2$, a densidade torna-se a normal padrão, para $\nu < 2$, a densidade tem caudas mais pesadas que a da normal e vice-versa para $\nu > 2$. Se z_t possui é uma variável aleatória normalizada com distribuição DEG, sua função de log verossimilhança é dada por,

$$L_T(\theta) = \sum_{i=1}^T \left[\ln(\nu/\lambda) - \left| \frac{z_i}{\lambda} \right|^\nu - (1 + \nu^{-1}) \ln(2) - \ln(\Gamma(1/\nu)) - 0,5 \ln(\sigma_i^2) \right]. \quad (2.89)$$

Ambas as densidades condicionais expostas anteriormente são simétricas. As-

simetria e curtose são momentos relevantes nas aplicações financeiras sob vários aspectos. Recentemente, Laurent & Peters (2001) aplicaram e estenderam a distribuição t-Student assimétrica proposta por Fernandez & Stell (1998) aos modelos ARCH.

Assumindo-se que z_t tenha distribuição t-Student assimétrica, a função de log verossimilhança é dada por,

$$L_T(\theta) = T \left[\ln \left(\Gamma \left(\frac{\nu + 1}{2} \right) \right) - \ln \left(\Gamma \left(\frac{\nu}{2} \right) \right) - 0,5 \ln(\pi(\nu - 2)) + \ln \left(\frac{2}{\iota + \frac{1}{\iota}} \right) \right. \\ \left. + \ln(s) - 0,5 \sum_{t=1}^T \left(\ln(\sigma_t^2) + (1 + \nu) \ln \left(1 + \frac{(sz_t + m)^2}{\nu - 2} \iota^{-2I_t} \right) \right) \right] \quad (2.90)$$

em que,

$$I_t = \begin{cases} 1 & \text{se } z_t \geq \frac{-m}{s}, \\ -1 & \text{se } z_t < \frac{-m}{s}, \end{cases}$$

e,

$$m = \frac{\Gamma \left(\frac{\nu + 1}{2} \right) \sqrt{\nu - 2}}{\sqrt{\pi} \Gamma \left(\frac{\nu}{2} \right)} \left(\iota - \frac{1}{\iota} \right),$$

com,

$$s = \sqrt{\left(\iota^2 + \frac{1}{\iota^2} - 1 \right) - m^2},$$

sendo ι o parâmetro de assimetria, com $\iota = 1$ tem-se a t-Student simétrica tradicional. A função densidade de uma t-Student assimétrica para uma variável aleatória padronizada z_t , pode ser representada por,

$$f(z|\iota, \nu) = \frac{2}{\iota + \frac{1}{\iota}} \{ g(\iota(sz+m)|\nu) I_{(-\infty, 0)}(z+m/s) + g((sz+m)/\iota|\nu) I_{[0, \infty)}(z+m/s) \},$$

em que $g(\cdot)$ denota a distribuição de uma t-Student.

2.10 ARCH: Testes

2.10.1 O Teste ARCH de Engle

Conforme Bollerslev et. al. (1994), o teste do tipo multiplicador de Lagrange proposto por Engle (1982) é bastante simples de se calcular e derivar. Sob a hipótese nula, assume-se que o modelo seja um modelo de regressão dinâmico padrão,

$$Y_t = x_t' \xi + \varepsilon_t,$$

e $\varepsilon_t | \Psi_{t-1} \sim N(0, \sigma^2)$. A hipótese alternativa é que os erros sejam ARCH(p) como na equação (2.10c), ou seja, testa-se $H_0 : \alpha_1 = \dots = \alpha_p = 0$. De fato, este teste pode ser facilmente implementado seguindo-se os seguintes passos:

1. Estimar uma regressão de mínimos quadrados ordinários e obter os resíduos $\hat{\varepsilon}_t$. Elevá-los ao quadrado para obter $\hat{\varepsilon}_t^2$.
2. Estimar uma regressão de $\hat{\varepsilon}_t^2$ contra uma constante e p de suas próprias defasagens, obtendo R_u^2 , a estatística R^2 não centrada da última regressão.
3. Sob a nula que $\varepsilon_t | \Psi_{t-1} \sim N(0, \sigma^2)$, a estatística de teste $TR_u^2 \xrightarrow{d} \chi_p^2$. Comparando então, TR_u^2 com a entrada apropriada de uma tabela da distribuição χ^2 , pode-se testar $H_0 : \alpha_1 = \dots = \alpha_p = 0$.

2.10.2 O Teste BDS

Segundo Bollerslev et. al. (1994), o teste do tipo multiplicador de Lagrange sugerido por Engle (1982) é mais um teste para agrupamento de volatilidade do que para heterocedasticidade condicional em geral ou dependência não linear em geral.

Um teste amplamente utilizado e citado na literatura para detectar desvios da hipótese de um processo iid é o teste proposto por Brock et. al. (1996) (Teste BDS).

Este teste inspirou e tem inspirado uma literatura extensa sobre suas aplicações em Econometria (ver Scheinkman & LeBaron (1989)).

De forma a descrever o teste, seja $\{y_t\}_{t=1,T}$ uma sequência escalar a qual, sob a hipótese nula, assume-se que seja um processo iid no tempo. Definindo-se as m-histórias do processo y_t como os sendo os vetores (y_1, \dots, y_m) , (y_2, \dots, y_{m+1}) , $(y_3, \dots, y_{m+2}), \dots, (y_{T-m}, \dots, y_{T-1})$, (y_{T-m+1}, \dots, y_T) . Isto posto, existem $(T - m + 1)$ m-histórias e $(T - m + 1)(T - m)/2$ pares distintos de m-histórias. Defina-se agora a integral de correlação como a fração dos pares distintos de m-histórias que estão a uma distância ϵ na sup norma, ou seja,

$$C_{m,T}(\epsilon) = \left(\frac{(T-m+1)(T-m)}{2} \right)^{-1} \sum_{t=m}^s \sum_{s=m}^T I(\max_{j=0}^{m-1} |y_{t-j} - y_{s-j}| < \epsilon). \quad (2.91)$$

Brock et. al. (1996) mostraram que sob condições fracas de dependência tem-se que $C_{m,T}(\epsilon) \rightarrow C_m(\epsilon)$ quase certamente. Bollerslev et. al. (1994) escrevem que pelas propriedades das estatísticas de ordem tem-se $C_m(\epsilon) = C_m(\epsilon)^m$ quando $\{y_t\}$ é iid. O teste BDS basea-se na diferença $(C_{m,T}(\epsilon) - C_{1,T}(\epsilon)^m)$. Informalmente, $(C_{m,T}(\epsilon) > C_{1,T}(\epsilon)^m)$ significa que quando $(\max_{j=0}^{m-1} |y_{t-j} - y_{s-j}| < \epsilon)$, então a chance de y_t e y_s estarem 'próximos' é maior que em média. Isto posto, conforme Bollerslev et. al. (1994), um método baseado no vizinho mais próximo funciona na previsão de $\{y_t\}$, fato inconsistente com um processo iid.

Brock et. al. (1996) demonstraram que para m e ϵ fixados $T^{1/2}(C_{m,T}(\epsilon) - C_{1,T}(\epsilon)^m)$, é assintoticamente normal com média zero e variância $Var(m, \epsilon)$ dada por,

$$Var(m, \epsilon) = 4 \left(K(\epsilon)^m + 2 \sum_{j=1}^{m-1} K(\epsilon)^{m-j} C_1(\epsilon)^{2j} + (m-1)^2 C_1(\epsilon)^{2m} - m^2 K(\epsilon) C_1(\epsilon)^{2m-2} \right),$$

sendo $K(\epsilon) = E\{[F(y_t + \epsilon) - F(y_t - \epsilon)]^2\}$ e $F(\cdot)$ a função de distribuição

acumulada de y_t . Isto posto, a estatística BDS pode ser calculada por,

$$T^{1/2}(C_{m,T}(\epsilon) - C_{1,T}(\epsilon)^m)/\widehat{Var}(T, m, \epsilon).$$

Brock et. al. (1996) mostraram que para $m \geq 2$ e $\epsilon > 0$ fixados, verifica-se que,

$$T^{1/2}(C_{m,T}(\epsilon) - C_{1,T}(\epsilon)^m)/\widehat{Var}(T, m, \epsilon) \xrightarrow{d} N(0, 1),$$

em que $\widehat{Var}(T, m, \epsilon)$ é um estimador consistente de $Var(m, \epsilon)$ derivado por Brock et. al. (1996).

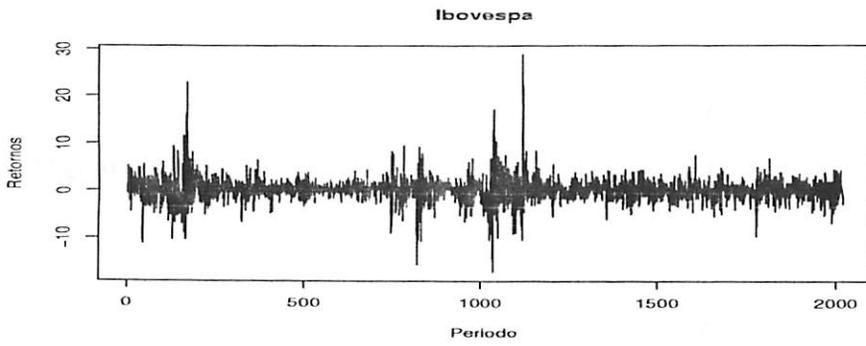
Segundo Bollerslev et. al. (1994), o teste BDS tem poder considerável contra muitos, não todos, desvios do comportamento iid. O teste BDS tem sido amplamente utilizado como um teste de especificação de modelos ARCH, sendo uma escolha natural, posto que modelos da classe ARCH postulam que os resíduos padronizados são iid, ou seja, $z_t \equiv \epsilon_t \sigma_t^{-1}$ são iid.

3.1 Dados

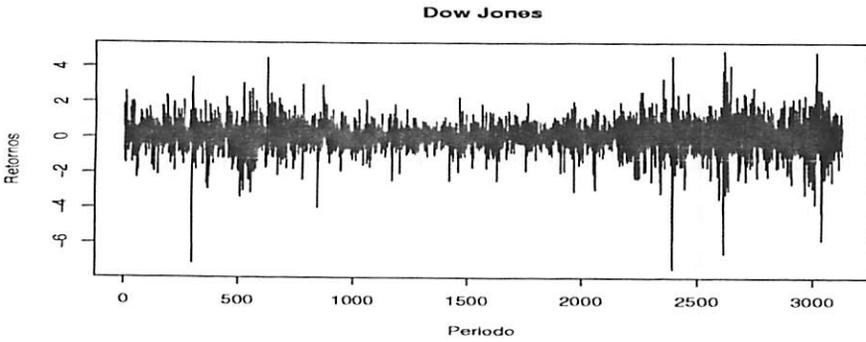
Os dados utilizados foram a série de preços à vista em dólares, dados de fechamento diário do índice Ibovespa, abrangendo o período amostral de 07/04/1994 a 09/04/2002, com um total de 2020 observações diárias; a série de preços à vista em dólares, dados de fechamento diário do índice Dow Jones, abrangendo o período de 23/08/1988 a 22/02/2000, totalizando 3.030 observações diárias, além da série de preços à vista em dólares, dados de fechamento diário do índice Standard & Poors (S&P(500)) que engloba as 500 maiores empresas norte-americanas, correspondentes ao período de 01/09/1971 a 02/10/1994, série que totaliza 8.432 observações.

As séries de preços dos índices Ibovespa e Dow Jones foram obtidas na base de dados mantida pelo Instituto de Pesquisa Economica Aplicada na internet. A série de preços do índice S&P(500) foi obtida na base de dados, acessada via internet, *Time Series Data Library*.

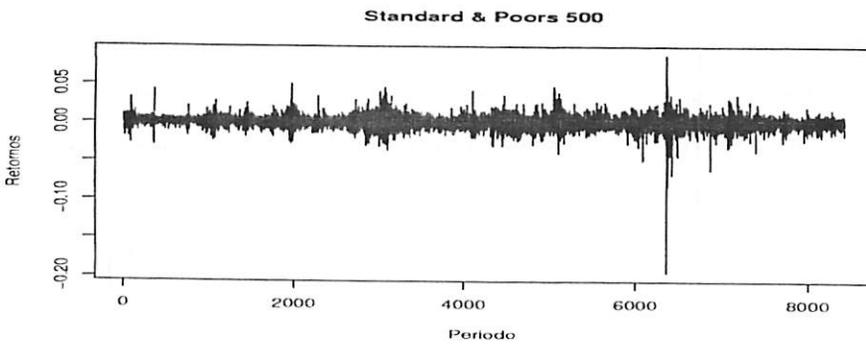
Os retornos compostos continuamente, ou retornos instantâneos percentuais, foram obtidos calculando-se a log-diferença das séries de preços, isto é, para cada série de preços y_t , efetuou-se $100 \times \Delta \ln(y_t)$. Gráficos das séries são apresentados na Figura 1. Confirmando as regularidades empíricas descritas anteriormente, todas as séries exibem fortes sinais de heterocedasticidade e agrupamento da volatilidade.



(a) Ibovespa: Retornos instantâneos - 07/04/1994:09/04/2002



(b) Dow Jones: Retornos instantâneos - 23/08/1988:22/02/2000



(c) S&P(500): Retornos instantâneos - 01/09/1971:02/10/1994

FIGURA 1: Ibovespa, Dow Jones e S&P(500): Retornos compostos continuamente

3.2 Metodologia

1. Inicialmente, modelam-se as dependências temporais, se existirem, na média condicional das séries analisadas. Dada a hipótese da eficiência dos mercados, espera-se que tais dependências sejam inexistentes ou insignificantes.
2. Modeladas as dependências temporais das médias das séries, serão analisados os resíduos gerados pelo modelo selecionado. O teste ARCH de Engle e o teste BDS serão utilizados para detectar a presença de efeitos ARCH e dependências temporais em geral.
3. Detectando-se efeitos ARCH e outras dependências, efetua-se um esforço de modelagem levando em consideração estas características. Nesta fase da modelagem, modelos da classe ARCH serão utilizados na modelagem da variância condicional. Especificamente, os modelos GARCH, IGARCH e FIGARCH utilizando-se quatro distribuições para os resíduos (normal, t-Student, distribuição generalizada do erro e t-Student assimétrica) serão estimados. O modelo APARCH será estimado para verificar se choques negativos e positivos tem impactos diferenciados sobre a volatilidade das séries.
4. Para cada série, o melhor modelo de cada tipo (i.e, GARCH, IGARCH e FIGARCH) será selecionado pelos critérios de informação de Hannan-Quinn, Schwarz e Shibata, todos baseados na verossimilhança. As funções objetivo a minimizar, para os respectivos critérios, são,

$$\text{Hannan-Quinn} = -2 \left(\frac{\log(L_T(\Theta))}{n} \right) - 2 \left(\frac{k \log(\log(n))}{n} \right),$$

$$\text{Schwarz} = -2 \left(\frac{\log(L_T(\Theta))}{n} \right) + 2 \left(\frac{\log(k)}{n} \right),$$

$$\text{Shibata} = -2 \left(\frac{L_T(\Theta)}{n} \right) + \log \left(\frac{n + 2k}{n} \right).$$

em que $\log(L_T(\Theta))$ denota o valor da log verossimilhança, n é o número de

observações e k o número de parâmetros.

5. Selecionados os melhores modelos GARCH, IGARCH e FIGARCH para cada série, serão analisadas as características da volatilidade de cada série e o desempenho destes modelos será avaliado por meio de medidas da qualidade do ajuste e acurácia preditiva. Um dos objetivos principais, é verificar o desempenho dos modelos GARCH fracionalmente integrados relativamente aos modelos GARCH e IGARCH, e os efeitos da adoção da t-Student assimétrica.
6. Os modelos da classe ARCH e os testes relacionados, com exceção do teste BDS, que foi implementado utilizando-se a linguagem R, foram todos estimados com o auxílio de programas sugeridos por Laurent & Peters (2002), elaborados na linguagem de programação matricial orientada a objetos Ox (ver Doornik (2001)). Isto posto, o método de otimização Quasi-Newton proposto independentemente por Broyden, Fletcher, Goldfarb e Shanno ou método BFGS (ver Gill et. al. (1981)) é o algoritmo de otimização. Utilizam-se gradientes numéricos no processo de otimização.

Como um teste gráfico para verificar a adequação, em termos de acurácia preditiva, das densidades condicionais assumidas para os retornos, implementou-se o teste de previsão da densidade proposto por Diebold et. al. (1998).

Conforme Diebold et. al. (1998), seja $f_i(y_i|\Omega_i)_{i=1}^m$ uma sequência de previsões da densidade *um passo à frente* produzidas por um dado modelo em que Ω_i é o conjunto de informação condicionante e $p_i(y_i|\Omega_i)_{i=1}^m$ a sequência de densidades definindo o processo gerador de dados y_i (não-observado). Diebold, Gunther & Tai (1998) mostraram que os usuários das previsões preferem a densidade correta em termos da perda esperada, sem considerar sua função perda.

Segundo Laurent & Peters (2002), a questão principal é então: a densidade é

adequada? Se a resposta é afirmativa, significa que

$$H_0 : f_i(y_i|\Omega_i)_{i=1}^m = p_i(y_i|\Omega_i)_{i=1}^m \quad (3.1)$$

Diebold, Gunther & Tai (1998) utilizam o fato que, sob a hipótese nula expressa em (3.1), a transformada integral da probabilidade $\zeta_i = \int_{-\infty}^{y_i} f_i(t)dt$ é *i.i.d* $U(0, 1)$. Para checar H_0 , eles propuseram a utilização de testes da qualidade do ajuste e testes de independência para *i.i.d* $U(0, 1)$.

A propriedade *i.i.d* de ζ_i pode ser verificada pela inspeção visual dos correlogramas de $(\zeta - \bar{\zeta})^j$, para $j = 1, 2, 3, 4, \dots$, que pode revelar dependências na média condicional, variância, assimetria, etc.

Inconsistências com a uniformidade podem também ser detectadas pela inspeção visual do histograma, que foi implementada, de ζ_i . Por exemplo, Laurent & Peters (2002) demonstraram que caso o histograma tenha a forma de um *S invertido*, tal fato poderia indicar que os erros seriam assimétricos, assim, a verdadeira densidade seria provavelmente assimétrica, o que justificaria a utilização de uma densidade assimétrica como t-student *assimétrica*. Por outro lado, um histograma em forma de *U* pode sugerir que o modelo produz previsões que sub(super)estimam sistematicamente. Além disso, um histograma em forma de 'montanha' poderia indicar que a curtose da verdadeira densidade não foi captada de forma adequada pela densidade assumida.

As funções de autocorrelação amostrais para os retornos, retornos quadráticos e módulo dos retornos para os índices Ibovespa, Dow Jones e S&P(500) estão ilustradas nas Figuras¹ 2, 3 e 4, respectivamente. Como esperado, o padrão de previsibilidade do nível dos retornos para todos os índices não é óbvio e claramente fraco. Conforme Issler (1999), se os retornos fossem previsíveis, haveria oportunidades de arbitragem para o investidor médio. Dado que todos podem ser investidores médios, não pode haver autocorrelação para os retornos, em conformidade com a hipótese da eficiência dos mercados.

Isto posto, o modelo selecionado para a média condicional dos retornos dos índices Ibovespa e Dow Jones contém apenas uma constante, dado que ao testar-se a significância de termos autorregressivos e médias móveis, utilizando-se o procedimento sugerido por Newey & West (1987) que é robusto à autocorrelação e à heterocedasticidade, a hipótese nula de um coeficiente igual a zero foi consistente com os dados em ambos os casos. Entretanto, para os retornos do índice S&P(500), além da constante acrescentou-se um termo AR(1), pois o mesmo mostrou-se significativo aos níveis usuais.

As funções de autocorrelação amostrais dos retornos quadráticos de todos os índices, por outro lado, exibem um padrão não negligível de previsibilidade, sugerindo que a variância condicional (ou volatilidade) dos retornos é previsível em certo grau. Este fato é corroborado pelo teste ARCH de Engle, efetuado sobre os resíduos quadráticos, e pelo teste BDS, efetuado sobre os resíduos padronizados dos modelos selecionados para a média condicional.

¹Todas as figuras citadas encontram-se no final da seção.

Os resultados dos testes ARCH e BDS, além de algumas estatísticas dos retornos dos índices, estão reportadas na Tabela 1. Pode-se observar que os coeficientes de curtose são típicos de séries financeiras e que os coeficientes de assimetrias são relativamente moderados. O passo seguinte foi modelar os retornos condicionais levando em consideração que os mesmos exibem efeitos ARCH e que, apesar de não autocorrelacionados, não exibem comportamento iid no tempo, como evidenciado pelo teste BDS.

TABELA 1: Sumário de Estatísticas dos Retornos dos índices financeiros analisados.

	Ibovespa	Dow Jones	SP(500)
Painel A: Momentos, Máximo e Mínimo			
Média	0,0508	0,0545	0,0004
desvio padrão	2,7480	0,9083	0,0085
Mínimo	-17,318	-7,4549	-0,1946
Máximo	28,818	4,8605	0,0882
Assimetria	0,5877	-0,5418	-1,3699
Curtose	14,518	9,1351	40,229
Jarque-Bera (p-valor)	0,0000	0,0000	0,0000
AR(1): Teste t robusto (P-valor)	0,3751	0,4724	0,0000
Número de Observações	2.020	3.130	8.432
Painel B: Teste ARCH (P-valor)			
ARCH(2)	0,0000	0,0000	0,0000
ARCH(50)	0,0000	0,0000	0,0000
ARCH(100)	0,0000	0,0000	0,0000
Painel C: Teste BDS (P-valor)			
m=2, $\epsilon = 0.5$	0,0000	0,0000	0,0000
m=5, $\epsilon = 0.5$	0,0000	0,0000	0,0000
m=2, $\epsilon = 1.5$	0,0000	0,0000	0,0000
m=5, $\epsilon = 1.5$	0,0000	0,0000	0,0000

Notas: Erro padrão do coeficiente AR(1) calculado usando o procedimento sugerido por Newey e West (1987). ARCH(p) denota o teste ARCH de Engle (1982) efetuado até o lag p . Para a estatística BDS, m é a dimensão incorporada e ϵ é a distância entre pares de observações consecutivas, medida em termos do número de desvios padrões de r_t . Se a série é IID, a estatística BDS é assintoticamente distribuída como uma normal padrão.

Isto foi implementado ajustando-se os modelos GARCH(p,q), IGARCH (p,q) e FIGARCH(p,d,q) para $p \in [0, 2]$ e $q \in [0, 2]$. Os critérios de informação de Hannan-Quinn, Schwarz e Shibata foram utilizados para determinar as ordens p e

q , assim como para selecionar as densidades, em conjunção com os testes ARCH e BDS. Quando assumiu-se a distribuição normal, utilizou-se a matriz de variância-covariância assintótica corrigida de Bollerslev & Wooldridge (1992). De forma a analisar a acurácia preditiva dos modelos selecionados, implementou-se previsões fora da amostra. As últimas 400 observações de cada série foram deixadas fora do processo de estimação. Além do teste de previsão de densidades proposto por Diebold et. al. (1998), quatro funções perda foram adotadas para medir a acurácia preditiva dos modelos selecionados.

4.1 Ibovespa

Na Tabela 1, verifica-se que a curtose do índice é típica de séries financeiras, entretanto, a assimetria positiva substancial é um fato interessante, posto que a maioria das séries financeiras possui assimetria negativa. O retorno médio diário percentual é bastante modesto. Uma inspeção visual da Figura 1 permite identificar períodos de alta volatilidade, como em 1994, período no qual a volatilidade acentuou-se após a implantação do plano Real e a crise do México, em seguida. A crise da Ásia (1997), a moratória russa (1998) e a maxidesvalorização cambial brasileira em 14/1/1999 também implicaram em nítidos aumentos da volatilidade do índice. O período que se estende de 1999 ao final da série é de relativa calma. O Ibovespa pode ser considerado um portfólio significativamente volátil, dado que possui volatilidade anualizada em torno de 44%.

Os resultados da estimação para o retorno do Ibovespa estão apresentados na Tabela 2. O modelo GARCH(1,1) com distribuição t-assimétrica teve um ajuste razoável, a soma dos parâmetros ($\alpha_1 + \beta_1$) ficou próxima de 1, sugerindo memória longa na volatilidade. A meia vida da volatilidade segundo este modelo ficou em torno de 30 dias. A estimativa de ν ficou um pouco acima das normalmente reportadas na literatura, conquanto seja significativa. Por sua vez, os resultados do modelo μ -IGARCH(1,1) com distribuição t-assimétrica são bastante semelhantes

aos do modelo μ -GARCH(1,1) com distribuição t-assimétrica. Assim, de acordo com Baillie et. al. (1996), qualquer pesquisador analisando sob a ótica da estrutura tradicional, poderia concluir que o modelo μ -GARCH (1,1) com distribuição t-assimétrica fornece uma descrição adequada do processo de volatilidade do Ibovespa.

TABELA 2: Retornos do Ibovespa: Resultados da estimação dos modelos ARCH.

Painel A: Média condicional			
μ	0,1344 (2,96)	0,1323 (2,93)	0,1313 (2,91)
Painel B: Variância condicional			
	GARCH(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)
α_0	0,1854 (3,01)	0,1189 (3,05)	0,2079 (2,11)
α_1	0,1457 (5,79)	0,1606 (6,14)	0,1208 (1,207)
β_1	0,8289 (28,46)	0,8393 -	0,4834 (3,243)
d	-	-	0,5411 (4,962)
ν	9,9587 (5,39)	9,041 (5,54)	9,3979 (6,00)
ι	-0,1019 (-2,84)	-0,0991 (-2,44)	-0,1079 (-2,92)
Distribuição	t(-)-assimétrica	t(-)-assimétrica	t(-)-assimétrica
Truncagem	-	-	1.000
Testes de Especificação			
Análise dos resíduos quadráticos			
Painel C: Teste ARCH (P-valor)			
ARCH(2)	0,0037*	0,0061*	0,0739
ARCH(50)	0,1605	0,2237	0,5734
ARCH(100)	0,0657	0,1536	0,2237
Análise dos resíduos padronizados			
Painel D: Teste BDS (P-valor)			
m=2, $\epsilon = 0.5$	0,2260	0,3753	0,6133
m=5, $\epsilon = 0.5$	0,0080*	0,0277*	0,1228
m=2, $\epsilon = 1.5$	0,3667	0,4870	0,9012
m=5, $\epsilon = 1.5$	0,2582	0,4681	0,8788

Notas: Estatística t entre parênteses. (*) denota que o teste é significativo ao nível $\alpha \leq 5\%$

Entretanto, analisando as estimativas e as medidas de qualidade do ajuste uti-

TABELA 3: Retornos do Ibovespa: medidas da qualidade do ajuste dos modelos ARCH.

Modelo	Hannan-Quinn	Schwarz	Shibata
MQO Homocedásticos	4,87985	4,88351	4,87772
GARCH(1,1)t-assimétrica	4,47549	4,48479*	4,47030
IGARCH(1,1)t-assimétrica	4,47564	4,48489	4,47030
FIGARCH(1,d,1)t-assimétrica	4,47547*	4,48829	4,46799*

Notas: (*) denota o melhor modelo. MQO é uma abreviação para mínimos quadrados ordinários.

lizadas, o modelo μ -FIGARCH(1, d , 1) com t-assimétrica supera os outros dois modelos analisados. O bom ajuste deste modelo implica que a volatilidade da série exibe memória longa. A estimativa de d , está próxima das reportadas na literatura e testes t para as nulas $d = 1$ e $d = 0$ claramente as rejeitam, aos níveis usuais. A Tabela 2 mostra que os modelos μ -GARCH(1,1) e μ -FIGARCH(1, d , 1) com distribuição t-assimétrica tiveram um desempenho bastante adequado, com o modelo μ -FIGARCH(1, d , 1) tendo um desempenho ligeiramente superior. Entretanto, quando analisam-se os resíduos (Tab. 2) dos modelos ajustados para os retornos do Ibovespa, vê-se que o modelo μ -FIGARCH(1, d , 1) capta melhor a dinâmica da volatilidade e o parâmetro de assimetria da t-Student foi significativo aos níveis usuais. Isto pode ser constatado pelo exame dos resultados dos testes ARCH e BDS na Tabela 2, de acordo com estes testes a especificação utilizada é adequada. O teste BDS sugere que as dependências observadas no retorno do Ibovespa deviam-se basicamente à presença de efeitos ARCH, fato comumente relatado na literatura. Curiosamente, o teste da razão de verosimilhança rejeitou a densidade t-Student assimétrica, enquanto todos os critérios de informação a selecionaram. O Teste de Nyblom para testar a estabilidade dos parâmetros sugeriu instabilidade apenas na estimativa da constante da média condicional da especificação μ -FIGARCH, o que é esperado.

Analisando-se a acurácia preditiva dos modelos selecionados, as medidas são apresentadas na Tabela 4, percebe-se claramente a relevância de se levar a dependência de longo prazo e a assimetria em consideração. A especificação μ -

S. J. ...

TABELA 4: Retornos do Ibovespa: medidas da acurácia preditiva dos modelos ARCH.

Modelo	Função perda			
	EQM	EAM	REQM	EPAM
MQO Homocedásticos	64,3433	6,3963	8,0214	271,27
GARCH(1,1)t-assimétrica	56,9471	5,3738	7,5463	195,3935
IGARCH(1,1)t-assimétrica	78,8612	7,2836	8,8804	288,4101
FIGARCH(1,d,1)t-assimétrica	56,8419*	5,2487*	7,5394*	184,5197*

Notas: (*) denota o melhor modelo. EQM=erro quadrático médio. EAM= erro absoluto médio. REQM= raiz do EQM, EPAM= erro percentual absoluto médio.

FIGARCH(1, d , 1) com t-Student assimétrica superou todas as outras especificações, segundo todas as funções perda adotadas. Para confirmar a adequação da t-Student assimétrica, a Figura 5 ilustra o teste de previsão da densidade proposto por Diebold et. al. (1998), não se verifica nenhum desvio significativo da uniformidade. Segundo este teste, a t-Student assimétrica foi bastante adequada para modelar a dinâmica da série.

Estimou-se um modelo μ -APARCH(1,1) com distribuição t-Student assimétrica de forma a testar-se o efeito alavancagem. A estimativa do parâmetro γ foi igual a 0,4480 e significativa aos níveis usuais, indicando a presença do efeito alavancagem. Assim, choques negativos têm um impacto maior sobre a volatilidade que choques positivos.

4.2 Dow Jones

O índice Dow Jones é composto pelas ações de 30 companhias industriais que são responsáveis por cerca de um quinto do valor total do mercado acionário americano. Novamente o retorno diário médio é modesto. O coeficiente de curtose também é típico de séries financeiras. O coeficiente de assimetria sugere que a distribuição dos retornos é negativamente assimétrica de forma substancial, fato comum para índices acionários. O índice Dow Jones pode ser considerado um portfólio com volatilidade moderada, tendo volatilidade anualizada de aproxima-

damente 14,40%.

Os Resultados da estimação para o retorno do Dow Jones estão na Tabela 5. O modelo μ -GARCH(1,1) com distribuição t-Student obteve um ajuste muito bom, a soma ($\alpha_1 + \beta_1$) ficou muito próxima de 1, sugerindo forte persistência de choques sobre a volatilidade, a meia vida da volatilidade segundo tal modelo é de 73 dias. Entretanto, um teste t revelou que a soma é significativamente menor 1, sugerindo que a série exibe reversão para a média.

De fato, todos os modelos selecionados tiveram um desempenho adequado em capturar a dinâmica da volatilidade dos retornos, o que pode ser constatado observando-se os testes de especificação presentes na Tabela 5. A Tabela 6 contém os critérios de informação utilizados como medidas da qualidade do ajuste. De forma interessante, os modelos μ -GARCH(1,1) e μ -IGARCH(1,1) com distribuição t-Student tiveram um desempenho significativamente superior ao modelo FIGARCH(1, d , 1) com distribuição t assimétrica. Ou seja, os modelos μ -GARCH(1,1) e μ -IGARCH(1,1) com distribuição t-Student obtiveram um balanço mais adequado entre acurácia e complexidade. Além disso, o teste da razão de verossimilhança selecionou a t-Student. O parâmetro d estimado do modelo μ -FIGARCH(1, d , 1) com distribuição t assimétrica e, apesar de significativamente diferente de zero, teve um valor relativamente baixo ($d = 0,3605$), conforme a Tabela 5. Novamente, o teste BDS sugere que as dependências evidenciadas pelos dados deviam-se de forma dominante aos efeitos ARCH. O teste de Nyblom sugere que os parâmetros α_1 , β_1 e ν são instáveis, o que foi interpretado como sugerindo que, apesar de ter sido o modelo selecionado, o modelo μ -GARCH(1,1) com t-Student pode não ser uma especificação totalmente adequada em termos de previsão.

No que concerne à acurácia preditiva dos modelos, cujos resultados estão na Tabela 7, confirmam-se as evidências que a incorporação da memória longa e de assimetrias na densidade condicional não implicaram em um melhor desempenho. Segundo as funções perda adotadas, o modelo com melhor desempenho preditivo

TABELA 5: Retornos do Dow Jones: Resultados da estimação dos modelos ARCH.

Painel A: Média condicional			
μ	0,0711 (5,68)	0,0711 (5,65)	0,6646 (4,86)
Painel B: Variância condicional			
	GARCH(1,1)	IGARCH(1,1)	FIGARCH(1, d, 1)
α_0	0,0045 (2,03)	0,0026 (2,06)	3,1744 (2,78)
α_1	0,0352 (4,43)	0,0360 (4,40)	0,3556 (6,64)
β_1	0,9602 (102,2)	0,9639 -	0,6656 (10,92)
d	-	-	0,3605 (6,02)
ν	5,0949 (10,5)	4,8137 (11,5)	5,3483 (10,46)
ι	-	-	-0,0238 (-1,049)
Distribuição	t(-)	t(-)	t(-) ^{assimétrica}
Truncagem	-	-	1.500
Testes de Especificação			
Análise dos resíduos quadráticos			
Painel C: Teste ARCH (P-valor)			
ARCH(2)	0,1758	0,1874	0,5310
ARCH(50)	0,9952	0,9974	0,9968
ARCH(100)	1,0000	1,0000	1,0000
Análise dos resíduos padronizados			
Painel D: Teste BDS (P-valor)			
m=2, $\epsilon = 0.5$	0,4953	0,4885	0,2011
m=5, $\epsilon = 0.5$	0,6229	0,5952	0,3773
m=2, $\epsilon = 1.5$	0,7962	0,7207	0,5968
m=5, $\epsilon = 1.5$	0,5850	0,4984	0,9871

Notas: Estatística t entre parênteses

foi o μ -GARCH(1,1) com distribuição t-Student, sendo superior em três das quatro funções perda analisadas. O teste de previsão da densidade reportado na Figura 6 confirma a adequação, em termos da densidade, do modelo μ -GARCH(1,1) com distribuição t-Student, dado que não há evidências de desvios significativos da uniformidade.

TABELA 6: Retornos do Dow Jones: medidas da qualidade do ajuste dos modelos ARCH.

Modelo	Hannan-Quinn	Schwarz	Shibata
MQO Homocedásticos	2,62303	2,62557	2,62160
GARCH(1,1)t(-)	2,39623	2,40258	2,39265*
IGARCH(1,1)t(-)	2,39552*	2,40061*	2,39266
FIGARCH(1,d,1)t-assimétrica	7,00192	7,00955	6,99763

Notas: (*) denota o melhor modelo. MQO é uma abreviação para mínimos quadrados ordinários.

TABELA 7: Retornos do Dow Jones: medidas da acurácia preditiva dos modelos ARCH.

Modelo	Função perda			
	EQM	EAM	REQM	EPAM
MQO Homocedásticos	13,4658	1,7898	3,6696	143,227*
GARCH(1,1)t(-)	13,2185*	1,2950*	3,6357*	368,4877
IGARCH(1,1)t(-)	14,7426	2,2539	3,8396	483,7318
FIGARCH(1,d,1)t-assimétrica	127631,1	151,0843	357,2552	275,8518

Notas: (*) denota o melhor modelo. EQM=erro quadrático médio. EAM= erro absoluto médio. REQM= raiz do EQM, EPAM= erro percentual absoluto médio.

O modelo μ -APARCH(1,1) com distribuição t-Student estimado, mostrou que existem evidências do efeito alavancagem, dado que o valor estimado $\gamma = 0,5308$ é positivo e significativo aos níveis usuais.

4.3 Standard & Poors (500)

O índice Standard & Poors (500) (S&P(500)) é composto pelas ações das 500 maiores companhias industriais norte-americanas. O retorno diário médio é significativamente baixo. A curtose é extremamente elevada, sugerindo que a distribuição condicional dos retornos é substancialmente leptocúrtica. O coeficiente de assimetria é também bastante elevado, sugerindo fortemente assimetrias na distribuição condicional dos retornos. O índice S&P(500) possui uma volatilidade anualizada extremamente baixa se comparada à das outras duas séries, cerca de 0,15%.

TABELA 8: Retornos do S&P(500): Resultados da estimação dos modelos ARCH.

Painel A: Média condicional			
μ	0,5769 (7,34)	0,5806 (7,4)	0,5931 (7,55)
ϕ_1	0,1491 (13,42)	0,1491 (13,49)	0,1499 (13,41)
Painel B: Variância condicional			
	GARCH(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)
α_0	0,4127 (4,47)	0,3186 (4,79)	0,5049 (2,07)
α_1	0,0668 (10,9)	0,0708 (11,77)	0,2636 (8,28)
β_1	0,9290 (153,3)	0,9291 -	0,7060 (16,31)
d	- -	- -	0,5263 (9,29)
ν	7,6866 (12,86)	7,3837 (13,42)	7,8127 (12,94)
ι	- -	- -	- -
Distribuição	t(·)	t(·)	t(·)
Truncagem	-	-	4.000
Testes de Especificação			
Análise dos resíduos quadráticos			
Painel C: Teste ARCH (P-valor)			
ARCH(2)	0,0021*	0,0036*	0,0612
ARCH(50)	0,4446	0,4270	0,4152
ARCH(100)	0,8068	0,7870	0,7611
Análise dos resíduos padronizados			
Painel D: Teste BDS (P-valor)			
m=2, $\epsilon=0.5$	0,9810	0,9956	0,3536
m=5, $\epsilon=0.5$	0,6830	0,8016	0,8700
m=2, $\epsilon=1.5$	0,7692	0,7921	0,4652
m=5, $\epsilon=1.5$	0,9425	0,9193	0,8141

Notas: Estatística t entre parênteses

A Tabela 8 contém os resultados da estimação para os retornos do índice S&P(500). Segundo o modelo AR(1)-GARCH(1,1) com distribuição t-Student, a meia vida da volatilidade é de 165 dias, um valor extremamente elevado, sugerindo fortemente a presença de memória longa na volatilidade. Os testes de especifica-

ção e as medidas da qualidade do ajuste mostram que levar em conta as dependências na média e as dependências de longo prazo implica em ganhos em termos de ajuste. Apesar das medidas de qualidade do ajuste, presentes na Tabela 9, mostrarem que todos os modelos selecionados tiveram um bom ajuste, com o modelo AR(1)-FIGARCH(1, d , 1) com distribuição condicional t-Student levando ligeira vantagem, os testes de especificação revelaram que o modelo FIGARCH(1, d , 1) com distribuição condicional t-Student capturou de forma mais adequada a dinâmica da volatilidade do índice. O teste da razão de verossimilhanças selecionou a t-Student contra a t-Student assimétrica. O teste de Nyblom evidencia instabilidades nos parâmetros ϕ_1 e α_0 . A instabilidade em ϕ_1 é esperada, a de α_0 poderia indicar algum tipo de mudança de regime na volatilidade.

TABELA 9: Retornos do S&P(500): medidas da qualidade do ajuste dos modelos ARCH.

Modelo	Hannan-Quinn	Schwarz	Shibata
MQO Homocedásticos	7,13798	7,13909	7,13740
GARCH(1,1)t(-)	6,77972	6,78305	6,77799
IGARCH(1,1)t(-)	6,77955	6,78232*	6,77811
FIGARCH(1,d,1)t(-)	6,77881*	6,78270	6,77679*

Notas: (*) denota o melhor modelo. MQO é uma abreviação para mínimos quadrados ordinários.

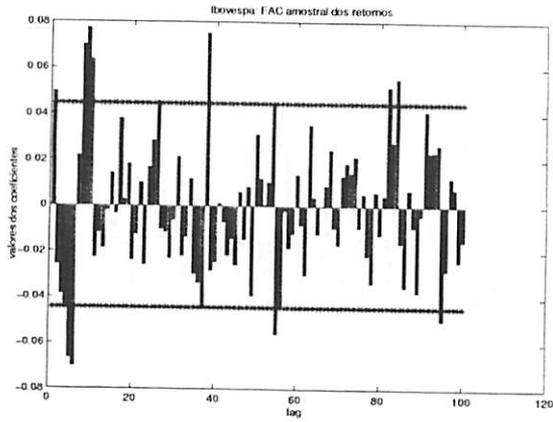
TABELA 10: Retornos do S&P(500): medidas da acurácia preditiva dos modelos ARCH.

Modelo	Função perda			
	EQM	EAM	REQM	EPAM
MQO Homocedásticos	3688,5728	55,9458	60,7336	838,0973
GARCH(1,1)t(-)	1394,7645	31,1642	37,3465	349,3261
IGARCH(1,1)t(-)	1511,2318	32,9842	38,8746	353,8245
FIGARCH(1,d,1)t(-)	1249,7201*	28,0340*	35,3514*	305,9096*

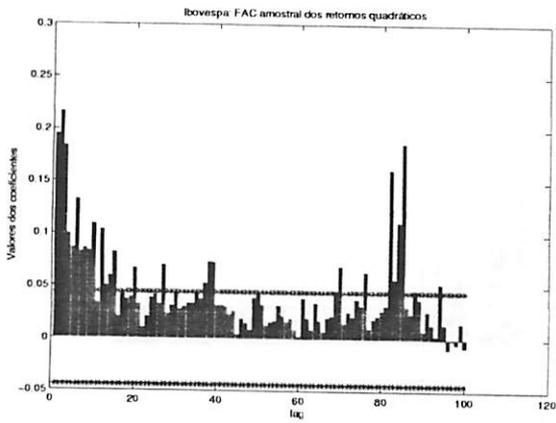
Notas: (*) denota o melhor modelo. EQM=erro quadrático médio. EAM= erro absoluto médio. REQM= raiz do EQM, EPAM= erro percentual absoluto médio.

Analisando as medidas de acurácia preditiva, apresentadas na Tabela 10, nota-se claramente que levar em consideração a memória longa evidenciada implicou em significativos ganhos em termos de acurácia preditiva. A especificação

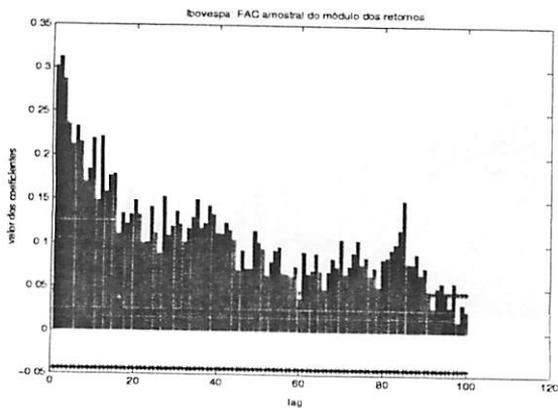
AR(1)-FIGARCH(1, d , 1) com distribuição condicional t-Student superou as outras especificações em todas as funções perda analisadas. O teste de previsão da densidade, representado na Figura 7, confirma que a especificação AR(1)-FIGARCH(1, d , 1) com distribuição condicional t-Student é bastante adequada, posto que não observam-se desvios significativos da uniformidade. O efeito alavancagem foi novamente constatado, posto que o modelo APARCH(1,1) estimado implicou em $\gamma=0,5170$, sendo significativo aos níveis usuais.



(a) FAC amostral dos retornos

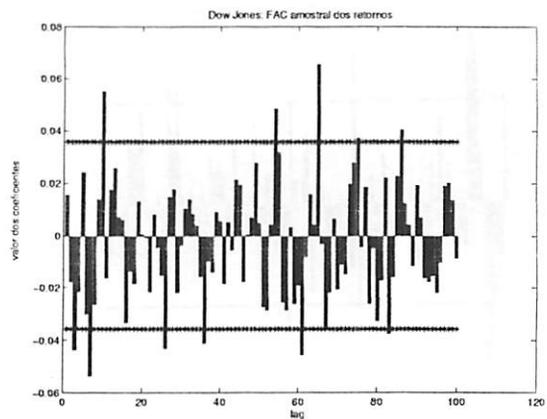


(b) FAC amostral dos retornos quadráticos

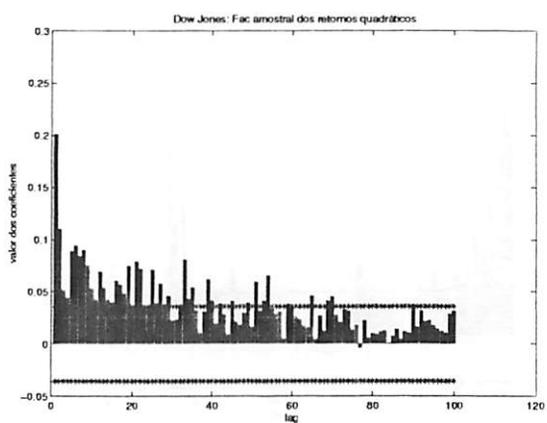


(c) FAC amostral do módulo dos retornos

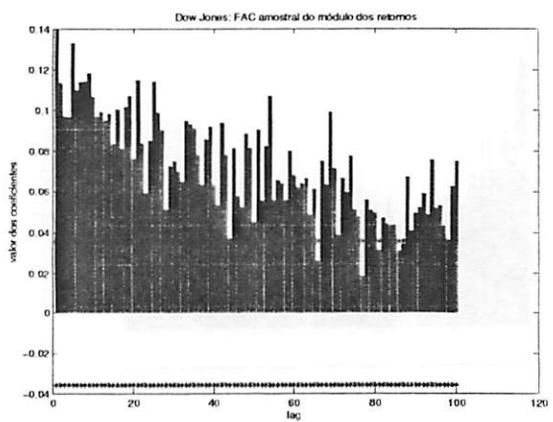
FIGURA 2: Ibovespa: Funções de autocorrelação amostral



(a) FAC amostral dos retornos

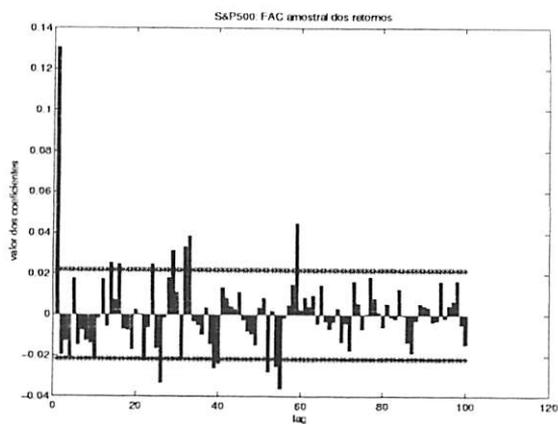


(b) FAC amostral dos retornos quadráticos

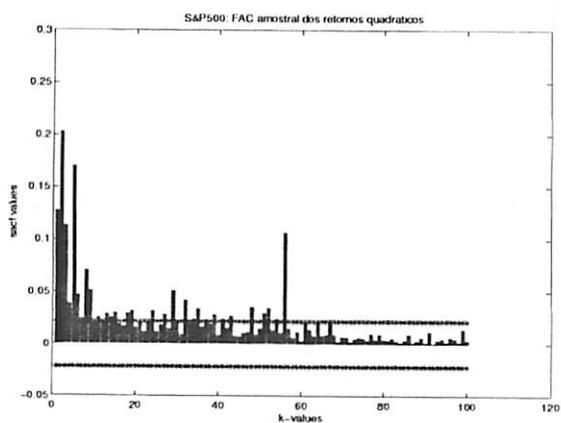


(c) FAC amostral do módulo dos retornos

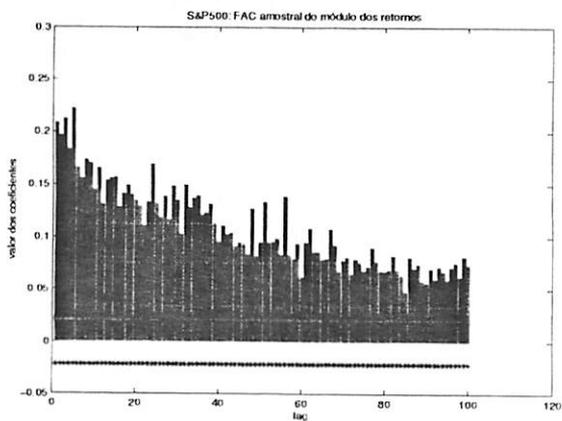
FIGURA 3: Dow Jones: Funções de autocorrelação amostral



(a) FAC amostral dos retornos



(b) FAC amostral dos retornos quadráticos



(c) FAC amostral do módulo dos retornos

FIGURA 4: S&P(500): Funções de autocorrelação amostral

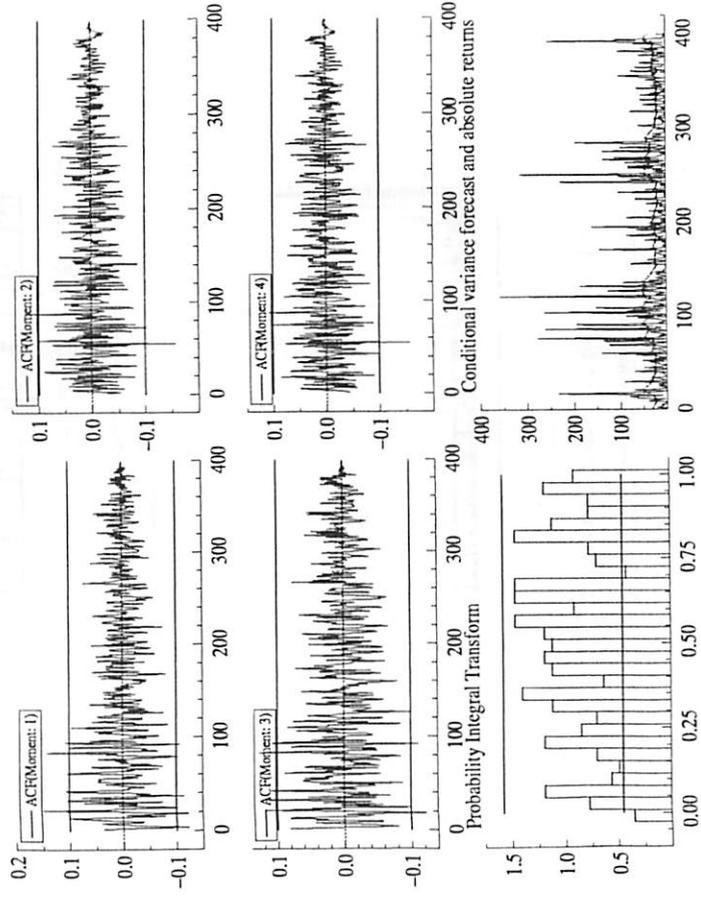


FIGURA 5: Ibovespa: Teste de previsão da densidade - μ -FIGARCH(1,d,1) \sim t-Student assimétrica

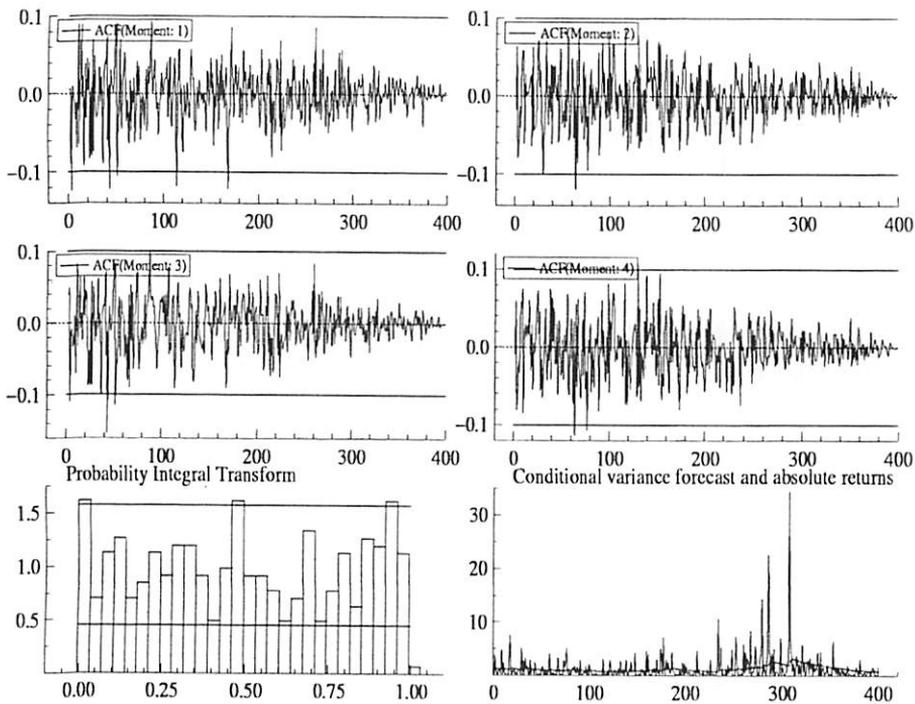


FIGURA 6: Dow Jones: Teste de previsão da densidade - μ -GARCH(1,1) \sim t-Student

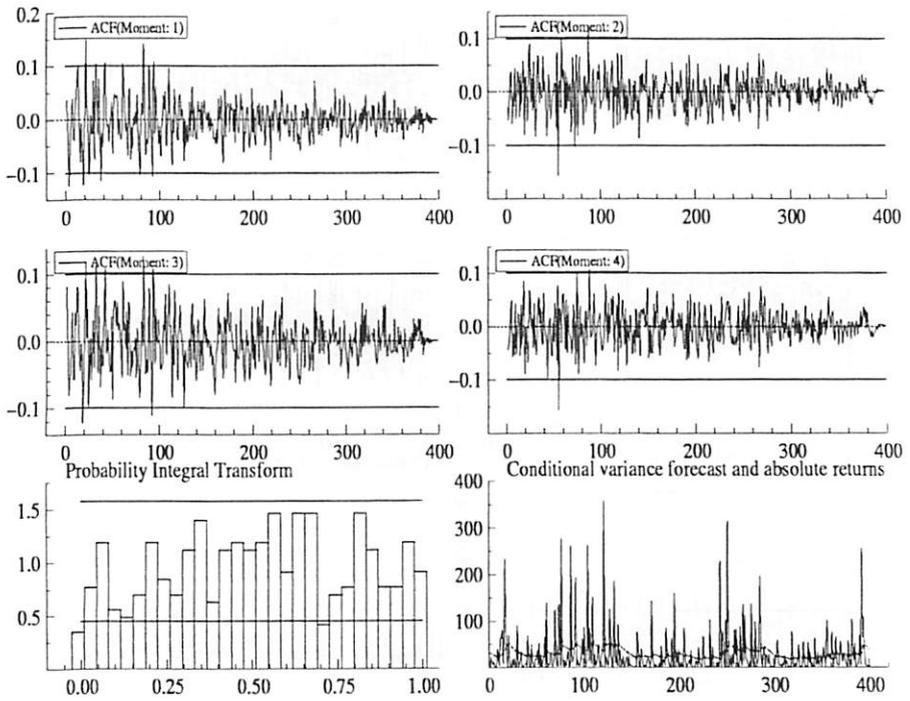


FIGURA 7: S&P(500): Teste de previsão da densidade - AR(1)-FIGARCH(1,d,1) \sim t-Student assimétrica

Nesta dissertação efetuou-se a modelagem do processo de volatilidade de três índices financeiros por meio de modelos da classe ARCH. Utilizou-se, no processo de modelagem, uma variante da classe de modelos ARCH que tem recebido grande atenção na literatura econométrica, os modelos GARCH fracionalmente integrados. Tais modelos possuem uma desejável flexibilidade na modelagem da dinâmica do processo de volatilidade captando, caso exista, a memória longa da volatilidade por meio do parâmetro de diferenciação fracionária. Utilizaram-se quatro densidades como distribuição condicional dos retornos: normal, t-Student, distribuição exponencial generalizada e t-Student assimétrica. A distribuição t-Student assimétrica mostrou-se bastante promissora, implicando em significativos ganhos na precisão da modelagem.

Os resultados evidenciaram que os processos de volatilidade dos índices analisados possuem algumas características em comum, como o efeito alavancagem, por exemplo, que foi constatado em todas as séries. Por outro lado, em termos de modelagem houve algumas diferenças, enquanto para os índices Ibovespa e Dow Jones, um modelo de média condicional constante foi suficiente para captar as dependências na média dos retornos. Para a série de retornos do índice Standard & Poors (500) foi necessário, além da constante, um termo autorregressivo de ordem 1, porém a magnitude do parâmetro estimado foi relativamente pequena.

O índice Ibovespa mostrou-se um portfólio substancialmente volátil no período analisado, com uma volatilidade anualizada em termos de 44%. O modelo μ -FIGARCH(1, d , 1) com distribuição t-Student assimétrica teve o melhor desempenho em termos da qualidade do ajuste e em termos de acurácia preditiva. Neste

caso, a flexibilidade adquirida por levar em consideração a memória longa da volatilidade e da adoção da t-Student assimétrica foi confirmada. A estimativa do parâmetro d para esta especificação foi igual a 0,5411, significativa aos níveis usuais.

O índice Dow Jones, por sua vez, mostrou-se um portfólio de volatilidade bastante moderada, a volatilidade anualizada ficou em torno de 14%. A especificação com melhor desempenho em termos de qualidade do ajuste e acurácia preditiva foi a do modelo μ -GARCH(1,1) com distribuição t-Student. No caso deste índice, as dependências temporais verificadas foram bem captadas por este modelo, não sendo necessárias maiores sofisticções. O baixo nível de dependência desta série poder ser confirmado pela estimativa do parâmetro d , que foi de 0,3605 e estatisticamente significativa.

A volatilidade do índice Standard & Poors (500) foi extremamente baixa, comparada à volatilidade dos outros índices, cerca de 0,15%. Novamente, a flexibilidade do modelo AR(1)-FIGARCH(1, d , 1) com a t-Student assimétrica foi a especificação de melhor desempenho, tanto em termos de qualidade do ajuste como em termos de acurácia preditiva. O maior grau de dependência desta série pode ser constatado pela estimativa do parâmetro d , que foi igual a (0,5263), sendo significativa aos níveis usuais.

Entre as possíveis extensões deste trabalho, destacaria como bastante interessante o desenvolvimento de um programa, utilizando a linguagem de programação Ox, para o cálculo dos coeficientes de resposta à impulsos gerados pelo modelo FIGARCH, cuja análise auxiliaria em muito o estudo do efeito de choques sobre o processo de volatilidade. A aplicação de modelos FIGARCH multivariados, envolvendo o conceito de cointegração fracionária, seria outra extensão bastante interessante. A comparação com os modelos de volatilidade estocástica com memória longa, em que se usam ondaletas para a estimação do parâmetro de diferenciação fracionária, seria outra linha interessante.

REFERÊNCIAS BIBLIOGRÁFICAS

- BAILLIE, R. T. Long memory processes and fractional integration in econometrics. **Journal of Econometrics**, Lousanne, v. 73, p. 5–59, 1996.
- BAILLIE, R. T.; BOLLERSLEV, T.; MIKKELSEN, H. O. Fractionally integrated generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, Lousanne, v. 74, n. 1, p. 3–30, 1996.
- BERA, A. K.; HIGGINS, M. L. ARCH models: Properties, estimation and testing. **Journal of Economic Surveys**, Oxford, v. 7, n. 4, p. 305–366, 1993.
- BERNDT, E. K.; HALL B.; HALL, R.; HAUSMAN, J. Estimation and inference in nonlinear structural models. **Annals of Economic and Social Measurement**, New York, v. 4, p. 653–665, 1974.
- BOLLERSLEV, T. A conditional heteroskedastic times series model for speculative prices and rates of return. **Review of Economics and Statistics**, Cambridge, v. 69, p. 542–547, 1987.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, Lousanne, v. 31, p. 307–327, 1986.
- BOLLERSLEV, T.; ENGLE, R. F.; NELSON, D. B. ARCH models. In: ENGLE, R. F.; MACFADDEN, D. L. (Ed.). **Handbook of Econometrics**. Amsterdan: North-Holland, 1994. v. 4, cap. 49, p. 2959–3038.
- BOLLERSLEV, T.; WOOLDRIDGE, J. M. Quasi-maximum likelihood estimation and inference in dynamic models with time varyng covariances. **Econometric Reviews**, New York, v. 11, n. 2, p. 143–172, 1992.
- BOUGEROL, P.; PICARD, N. Stationarity for GARCH processes and of some nonnegative time series. **Journal of Econometrics**, Lousanne, v. 52, p. 115–127, 1992.
- BROCK, W. A.; DECHERT, W. D.; SCHEINKMAN, J. A. A test for independence based on the correlation dimension. **Econometric Reviews**, New York, v. 15, p. 197–235, 1996.

- BROCKWELL, P. J.; DAVIS, R. A. **Time series: theory and methods**. New York: Springer-Verlag, 1997. 527 p.
- CAMPBELL, J.; LO, A.; MACKINLEY, A. C. **The econometrics of financial markets**. Princeton, New Jersey: Princeton University Press, 1997. 632 p.
- CRATO, N.; LIMA, P. J. F. Long range dependence in the conditional variance of stock returns. **Economics Letters**, Amsterdam, v. 45, p. 281–285, 1994.
- DACOROGNA, M. M.; MULLER, U. A.; NAGLER, R. J.; OLSEN, R. B.; PICTET, O. V. A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. **Journal of International Money and Finance**, Oxford, v. 12, p. 413–438, 1993.
- DIEBOLD, F. X.; GUNTHER, T. A.; TAY, A. S. Evaluating density forecasts with applications to financial risk management. **International Statistical Review**, Edinburg, v. 39, p. 963–883, 1998.
- DING, Z.; ENGLE, R. F.; GRANGER, C. W. J. A long memory property of stock market returns and a new model. **Journal of Empirical Finance**, v. 1, p. 83–106, 1993.
- DOORNIK, J. A. **Object-oriented matrix programming using Ox**. 4. ed. London: Timberlake Consultant Press, 2001. 538 p.
- ENGLE, R. F. Autoregressive conditional heteroskedasticity with estimates of the United Kingdom inflation. **Econometrica**, Avon, v. 50, n. 4, p. 987–1008, 1982.
- ENGLE, R. F.; BOLLERSLEV, T. Modelling the persistence of conditional variances. **Econometric Reviews**, New York, v. 5, p. 81–87, 1986.
- ENGLE, R. F.; GONZALES-RIVERA, G. Semiparametric ARCH model. **Journal of Business and Economic Statistics**, Washington, v. 9, p. 345–360, 1991.
- ENGLE, R. F.; ITO, T.; LIN, W.-L. Meteor showers or heat waves? heteroskedastic intra-daily volatility in the foreign exchange market. **Econometrica**, Avon, v. 58, p. 525–542, 1990.
- ENGLE, R. F.; NG, V. Measuring and testing the impact of news on volatility. **Journal of Finance**, New York, v. 48, n. 5, p. 1022–1082, 1993.
- FAMA, E. F. The variation of certain speculative prices. **Journal of Business**, Chicago, v. 38, n. 1, p. 34–105, 1965.

- FAMA, E. F. Efficient capital markets: a review of theory and empirical work. **Journal of Finance**, New York, v. 25, p. 383–417, 1970.
- FERNANDEZ, C.; STELL, M. On bayesian modelling of fat tails and skewness. **Journal of the American Statistical Association**, Washington, v. 93, n. 441, p. 359–371, Mar. 1998.
- GILL, P. E.; MURRAY, W.; WRIGHT, M. H. **Practical optimization**. New York: Academic Press, 1981. 412 p.
- GLOSTEN, L. R.; JAGANNATHAN, R.; RUNKLE, D. On the relation between the expected value and volatility of the nominal excess return on stocks. **Journal of Finance**, New York, v. 48, p. 1779–1801, 1993.
- GOURIEROUX, C. **ARCH models and financial applicatiuons**. Berlim: Springer - Verlag, 1997. 228 p. (Springer Series in Statistics).
- GRANGER, C. W. J. Long memory relationships and the aggregation of dynamic models. **Journal of Econometrics**, Lousanne, v. 14, p. 227–238, 1980.
- HAMILTON, J. D. **Time series analysis**. Princeton, New Jersey: Princeton University Press, 1994. 820 p.
- HOSKING, J. R. M. Fractional differencing. **Biometrika**, London, v. 68, p. 165–176, Apr. 1981.
- HURST, H. Long term storage capacity of reservoirs. **Transactions of the American Society of Civil Engineers**, New York, v. 116, p. 770–799, 1951.
- INSTITUTO DE PESQUISA ECONOMICA APLICADA. Disponível em: <<http://www.ipeadata.com.br>>. Acesso em: 13 de mar. 2001.
- ISSLER, J. V. Estimating and forecasting the volatility of brazilian finance series using ARCH models. **The Brazilian Review of Econometrics**, Rio de Janeiro, v. 19, n. 1, p. 5–56, 1999.
- LAMOUREUX, G. C.; LASTRAPES, W. D. Persistence in variance, structural change and the GARCH model. **Journal of Business and Economic Statistics**, Washington, v. 8, p. 225–234, 1990.
- LAURENT, S.; PETERS, J. P. **Modelling skewness dynamics in series of financial data**. Université Catholique de Louvain, Department of Quantitative Economics, 32p., 2001. (Discussion Paper, v.124) Disponível em: <<http://www.core.ucl.ac.be:16080/laurent>>. Acesso em: 17 de out. 2002.

- LAURENT, S.; PETERS, J. P. GARCH 2.3: An Ox package for estimating and forecasting various ARCH models. **Journal of Economic Surveys**, Oxford, v. 16, p. 447–485, 2002.
- MANDELROT, B. The variation of certain speculative prices. **Journal of Business**, Chicago, v. 36, n. 4, p. 394–419, 1963.
- MANDELROT, B.; NESS, J. W. V. Fractional brownian motions, fractional noises and applications. **SIAM Review**, Philadelphia, v. 10, p. 422–437, 1968.
- MIZRACH, B. **Learning and conditional heteroskedasticity in asset returns**. Rutgers University, Department of Economics, , 21p., 1990. (Working Papers, v. 95, n. 26) Disponível em: <<http://www.rci.rutgers.edu/mizrach>>. Acesso em: 16 de jan. 2002.
- MORETTIN, P. A.; TOLOI, C. M. **Previsão de séries temporais**. 2. ed. São Paulo: Atual, 1987. 395 p.
- NELSON, D. B. ARCH models as diffusion approximations. **Journal of Econometrics**, Lousanne, v. 45, p. 7–38, 1990a.
- NELSON, D. B. Stationarity and persistence in the GARCH(1,1) model. **Econometric Theory**, Cambridge, v. 6, p. 318–334, 1990b.
- NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. **Econometrica**, Avon, v. 59, n. 2, p. 347–370, 1991.
- NELSON, D. B.; CAO, C. Q. Inequality constraints in the univariate GARCH model. **Journal of Business & Economic Statistics**, Washington, v. 10, p. 229–235, 1992.
- NEWBY, W.; WEST, K. A simple positive semi definite heteroskedasticity and autocorrelation consistent covariance matrix. **Econometrica**, Avon, v. 55, p. 703–708, 1987.
- SCHEINKMAN, J. A.; LEBARON, B. Nonlinear dynamics and stock returns. **Journal of Business**, Chicago, v. 62, p. 311–337, 1989.
- SCHWERT, G. W. Why does stock market volatility change over time. **Journal of Finance**, New York, v. 44, p. 1115–1153, 1989.
- STOCK, J. H. Estimating continuous-time processes subject to time deformation: an application to postwar US GNP. **Journal of the American Statistical Association**, Washington, v. 83, n. 401, p. 77–85, Mar. 1988.

✓ TAYLOR, S. **Modelling financial time series**. New York: J. Wiley, 1986. 284 p.

TIME SERIES DATA LIBRARY. Disponível em: <<http://www.tsd1.com>>.
Acesso em: 15 de abr. 2001.

WEISS, A. A. Asymptotic theory for ARCH models: estimation and testing. **Econometric Theory**, Cambridge, v. 2, p. 107–131, 1986.

ZAKOIAN, J. M. Threshold heteroskedastic models. **Journal of Economic Dynamics and Control**, Amsterdam, v. 18, p. 931–955, 1994.

- 1 - indicar as características de seis Círculos
- 2 - indicar a tabela com todos os modelos
- 3 - indicar os pontos