



PAULO HENRIQUE SALES GUIMARÃES

**MODELOS DE VOLATILIDADE COM
INOVAÇÕES SKEW-T**

LAVRAS – MG

2014

PAULO HENRIQUE SALES GUIMARÃES

MODELOS DE VOLATILIDADE COM INOVAÇÕES SKEW-T

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador

Dr. Mário Javier Ferrua Vivanco

LAVRAS - MG

2014

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e
Serviços da Biblioteca Universitária da UFLA**

Guimarães, Paulo Henrique Sales.

Modelos de volatilidade com inovações skew-t / Paulo Henrique
Sales Guimarães. – Lavras : UFLA, 2014.

130 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2014.

Orientador: Mário Javier Ferrua Vivanco.

Bibliografia.

1. Série de retornos. 2. Volatilidade. 3. Skew-t. 4. Modelos
heterocedásticos. 5. Memória longa. I. Universidade Federal de
Lavras. II. Título.

CDD – 519.5

PAULO HENRIQUE SALES GUIMARÃES

MODELOS DE VOLATILIDADE COM INOVAÇÕES SKEW-T

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 31 de Julho de 2014.

Dra. Heloísa Rosa Carvalho	UFLA
Dra. Karen Luz Burgoa Rosso	UFLA
Dr. Telde Natel Custódio	UFSJ
Dra. Thelma Sáfadi	UFLA

Dr. Mário Javier Ferrua Vivanco
Orientador

LAVRAS - MG
2014

Dedico

Ao mestre dos mestres, Jesus Cristo,

Aos meus amados pais, Paulo e Nêza,

Aos meus queridos irmãos, Renato e Lívia,

Aos meus grandes amigos,

em especial, Márcio e a toda sua família,

Aos meus dedicados mestres.

AGRADECIMENTOS

A Deus, que se faz presente em minha vida, guiando meus passos e concedendo-me graças a cada dia.

A Nossa Senhora Aparecida, minha mãe celestial, por atender sempre as minhas súplicas e acalantar o meu coração.

À Universidade Federal de Lavras (UFLA), em especial ao Departamento de Ciências Exatas (DEX), pela realização deste trabalho.

Aos professores do DEX, pelos valiosos ensinamentos que foram muito importantes para minha formação.

Ao meu orientador, Mário Javier Ferrua Vivanco, pela grande atenção, competência, dedicação e arte de ensinar.

Ao professor Fortunato Silva de Menezes, pela paciência e disponibilidade de ajudar-me sempre que possível.

Às professoras Karen Luz Burgoa Rosso e Thelma Sáfadi pelas críticas e sugestões no meu trabalho.

A todos os colaboradores do DEX, pelos serviços prestados.

A todos os meus colegas do curso de Matemática e de Estatística.

Ao meu amado pai que sempre me apoiou e me ajudou nesta longa caminhada rumo ao conhecimento. À minha prezada mãe, pelo exemplo de vida, força, incentivo e dedicação.

Aos meus queridos irmãos, Renato e Lívia.

Ao Márcio Felipe e à minha segunda mãe Lucilene e a toda a sua família, que sempre me ajudaram e incentivaram.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro.

Enfim, a todos que contribuíram, direta ou indiretamente, meus eternos agradecimentos.

*“Suba o primeiro degrau com fé. Não é necessário que você veja toda a escada.
Apenas dê o primeiro passo”.*

Martin Luther King (1929-1968)

*“Refleta sobre as bênçãos do presente, que todos os homens possuem em
abundância; não sobre os infortúnios do passado, que todos os homens possuem
em alguma medida.”*

Charles Dickens (1812-1870)

RESUMO

As distribuições assimétricas tiveram um grande desenvolvimento nos últimos tempos. São utilizadas em várias áreas, inclusive, na modelagem de dados financeiros. Neste trabalho, considera-se a distribuição skew-t, que, em virtude de sua capacidade de modelar caudas pesadas de retornos financeiros, também, tem a capacidade de incorporar a assimetria presente na volatilidade. Também, são apresentadas algumas características das distribuições assimétricas, além dos principais resultados encontrados na literatura, relacionados a processos ARCH, GARCH, modelos combinados ARMA-GARCH e modelos de memória longa, em especial o FIGARCH e o FIEGARCH. Finalmente são feitas aplicações em séries reais de retornos financeiros, nos quais é feita a análise da persistência da volatilidade destas séries, além de ajustes ARMA-GARCH com inovações skew-t, comparando-se estes resultados em relação aos modelos FIGARCH e FIEGARCH, também, conseguem captar comportamentos de longa duração.

Palavras-chave: Série de retornos. Volatilidade. Skew-t. Modelos heterocedásticos. Memória longa.

ABSTRACT

The asymmetric distributions have had a great development in the last past years. They are useful in a great variety of areas, including The Financial Data Modeling. On this work, it is considered the distribution skew-t, due its capacity to model the huge tails from the financial returns, and to its capacity to incorporate the asymmetry that is seen in its volatility. Also, are presented some of the characteristics of the asymmetric distribution, in addition to its main results found in the literature related to ARCH processes , GARCH, combined models ARMA-GARCH and long memory models, specially the FIGARCH and the FIEGARCH. Finally, some applications are made to the real series of financial returns, in which is made an analysis of volatility persistence in this series, besides the adjustments ARMA-GARCH with innovations from the skew-t, comparing if the results in relation to its models, FIGARCH and FIEGARCH, also can capture the long term behavior.

Keywords: Returns series. Volatility. Skew-t. Heteroskedasticity models. Long memory.

LISTA DE FIGURAS

Figura 1	Série temporal dos retornos do índice diário RTSI da Bolsa de Valores de Moscou no período de setembro de 1995 a maio de 2014.....	34
Figura 2	(a) Histograma e (b) gráfico QxQ plot dos retornos do índice diário RTSI da Bolsa de Valores de Moscou no período de setembro de 1995 a maio de 2014.....	41
Figura 3	Formas apresentadas por uma distribuição	43
Figura 4	Gráfico da distribuição skew-t. Em (a) $\gamma = 0,8$; $\gamma = 1$ e $\gamma = 1,3$ em azul, preto e vermelho, respectivamente e ν fixo, igual a 5. Em (b) γ (fixo) é fixo e ν assume valores 5 (preto), 20 (vermelho) e 50 (verde).....	50
Figura 5	Gráfico da função distribuição acumulada da skew-t. $\gamma = 1$; $\gamma = 0,8$ e $\gamma = 1,3$ em preto, vermelho e azul, respectivamente. O parâmetro ν foi fixado em um	51
Figura 6	Simulação 10000 observações do modelo GARCH com inovações normais. Em (a) tem-se a o modelo GARCH (1,0) e em (b) um modelo GARCH (1,1).....	63
Figura 7	Simulação de 10000 observações do modelo GARCH (1,1) com inovações skew-t. Em (a) tem-se o modelo GARCH (1,1) com erros $ST(\nu=4, \gamma=0,6)$ e em (b) $ST(\nu=4, \gamma=1,5)$	71
Figura 8	Simulação de 10000 observações do modelo AR (1) - GARCH (1,1), com parâmetro autoregressivo $\phi_1 = 0,5$ em (a) e (b) MA (1) - GARCH (1,1) e parâmetro média móvel $\theta_1 = 0,5$	74

Figura 9	Simulação de 10000 observações do modelo ARMA - GARCH (1,1), com parâmetro autoregressivo $\phi_1 = 0,5$ e média móvel $\theta_1 = 0,5$ com erro $ST(\nu = 4, \gamma = 0,6)$ e parâmetros $\omega = 0,1$; $\alpha_1 = 0,3$ e $\beta_1 = 0,6$	74
Figura 10	Séries temporais simuladas segundo o processo ARFIMA.....	89
Figura 11	(a) Gráfico da série índice IBOVESPA (Brasil), (b) Série dos retornos IBOVESPA, (c) Histograma com densidade ajustada, (d) Gráfico Q x Q plot normal.....	98
Figura 12	(a) Gráfico da série índice SSE Composite (China), (b) Série dos retornos SSE Composite, (c) Histograma com densidade ajustada, (d) Gráfico Q x Q plot normal.....	99
Figura 13	FAC e FACP dos retornos diários do IBOVESPA	100
Figura 14	Resíduos padronizados, FAC dos resíduos do ajuste do modelo ARMA (0,1) para a série de retornos IBOVESPA.....	102
Figura 15	Função de correlação cruzada entre os resíduos e a série de retornos.....	103
Figura 16	Resíduos padronizados FAC e FACP dos resíduos do ajuste do modelo ARMA (10,0) para a série de retornos IBOVESPA.....	104
Figura 17	FAC e FACP dos resíduos ao quadrado do ajuste do modelo ARMA (10,0) para a série de retornos IBOVESPA.....	105
Figura 18	FAC dos resíduos padronizados e ao quadrado do ajuste do modelo AR (10)-GARCH (1,1).....	106
Figura 19	FAC dos resíduos padronizados e ao quadrado do ajuste do modelo ARMA (5,3) – GARCH (1,1).....	110

LISTA DE TABELAS

Tabela 1	Estatística descritiva IBOVESPA	96
Tabela 2	Estatística descritiva SSE	96
Tabela 3	Modelos ARMA ajustados para a série de retornos IBOVESPA...	101
Tabela 4	Critério de seleção de modelos – Retornos IBOVESPA.....	107
Tabela 5	Parâmetros do Modelo ARMA (5,3) – GARCH (1,1) – t de Student.....	108
Tabela 6	Parâmetros do Modelo ARMA (5,3) – GARCH (1,1) – skew-t.....	109
Tabela 7	Critério de seleção de modelos – Retornos SSE	110
Tabela 8	Estatística R/S e valor estimado \hat{d} das séries de retornos analisadas	112
Tabela 9	Ajuste modelo FIGARCH – Série retornos IBOVESPA	113
Tabela 10	Ajuste modelo FIEGARCH – Série retornos IBOVESPA.....	113
Tabela 11	Ajuste modelo FIGARCH – Série retornos SSE.....	114
Tabela 12	Critério de seleção de modelos – Série retornos SSE.....	114

LISTA DE SIGLAS E ABREVIATURAS

APARCH	Asymmetric Power ARCH Model
AR	Autorregressivo
ARCH	Autoregressive conditional heteroscedasticity
ARFIMA	Autoregressive fractionally integrated moving average
EGARCH	Exponencial GARCH
Fac	Função de autocorrelação
Facv	Função de autocovariância
FIGARCH	Fractionally Integrated GARCH
FIEGARCH	Fractionally Integrated Exponencial GARCH
GARCH	Generalized ARCH
IBOVESPA	Índice da Bolsa de Valores de São Paulo
IGARCH	Integrated GARCH
MA	Moving Average
ML	Memória Longa
MV	Máxima Verossimilhança
RTSI	Russian Trading System Stock Exchange
SSE	Shanghai Stock Exchange

LISTA DE SÍMBOLOS

$N(\mu, \sigma^2)$	Distribuição normal com parâmetros de posição μ e escala σ^2
t_ν	Distribuição t de Student padrão com ν graus de liberdade
$SN(\lambda)$	Distribuição skew normal com parâmetro de assimetria λ
$ST(\nu, \gamma)$	Distribuição skew-t com ν graus de liberdade e parâmetro de assimetria γ
$\Gamma(x)$	Função gama
\sim	Distribui-se como
\approx	Aproximadamente
\mathbb{N}	Conjunto dos números naturais
\mathbb{Z}	Conjunto dos números inteiros

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Objetivos específicos	19
2	REFERENCIAL TEÓRICO	20
2.1	Conceitos Preliminares	20
2.1.1	Séries Temporais	20
2.1.2	Processo estocástico	21
2.1.3	Estacionariedade	24
2.1.4	Ergodicidade	25
2.1.5	Processo Martingale	26
2.1.6	Operadores	27
2.1.7	Processos autorregressivos – AR(p)	27
2.1.8	Processos de médias móveis – MA(q)	28
2.1.9	Processo Autorregressivo de Médias Móveis – ARMA (p, q)	28
2.1.10	Séries de Retornos	31
2.1.11	Fatos estilizados sobre retornos financeiros	32
2.1.12	Volatilidade	35
2.1.13	Histograma e Gráfico QxQ Plot	38
2.1.14	Assimetria e Curtose	41
2.2	Distribuições Assimétricas	43
2.2.1	Distribuição skew-t	47
2.3	Modelos Heterocedásticos Condicionais	53
2.3.1	Modelos ARCH	56
2.3.2	Modelos GARCH	60
2.3.3	Métodos de Estimação	64
2.3.4	Estimação modelo GARCH	66
2.3.5	Estimação dos parâmetros do modelo GARCH com inovações skew-t	70
2.4	Modelos Combinados (ARMA-GARCH)	72
2.5	Testes para Detecção de Modelos GARCH	75
2.5.1	Teste FAC e FACP	75
2.5.2	Teste de Box-Pierce-Ljung	78
2.5.3	Teste Multiplicador de Lagrange	78
2.6	Previsão e adequação do modelo	79
2.7	Previsões de modelos GARCH	80
2.8	Critérios de seleção de modelos	81
2.9	Processos IGARCH e persistência na volatilidade	83
2.10	Modelos de Memória Longa	86
2.10.1	Modelos ARFIMA	86
2.10.2	Modelos GARCH com Memória Longa	90
2.10.3	Modelos FIGARCH e FIEGARCH	91

3	MATERIAL E MÉTODOS	93
3.1	Dados	93
3.2	Métodos	93
4	RESULTADOS E DISCUSÃO	96
4.1	Comportamento das séries de retornos dos índices	96
4.2	Análise da persistência da volatilidade	111
4.3	Análise de Memória Longa	112
5	CONCLUSÕES	115
	REFERÊNCIAS	117
	APÊNCICES	124

1 INTRODUÇÃO

Uma característica marcante do mercado financeiro diz respeito às condições de incerteza, em que se desenvolveu da ideia de risco e cuja importância se manifesta na sua capacidade de alterar e influenciar a tomada de decisões dos agentes econômicos.

O conceito de risco remonta ao trabalho pioneiro de Markowitz (1952), que pode ser compreendido como a probabilidade de ocorrência de impactos negativos nos resultados financeiros do mercado seja em virtude de movimentos desfavoráveis nas flutuações nos preços de ações ou derivativos, por taxas de juros ou câmbio abusivas, políticas econômicas ineficientes ou até mesmo em decorrência das condições climáticas desfavoráveis. O fato é que todos os que fazem parte do mercado financeiro estão sujeitos a condições de risco e, por isso, compreendê-las é imprescindível para que possam permanecer e auferir ganhos no mesmo.

A questão que se coloca é como definir apropriadamente a medida do grau de risco de um investimento ou ativo financeiro qualquer. Essa medida pode ser dada por meio do conhecimento da *volatilidade*, que pode ser entendida como uma medida estatística da possibilidade do preço de um dado ativo ou carteira (grupo de vários ativos) cair ou subir em um determinado período de tempo.

Segundo Morettin (2011), volatilidade é definida como a variância condicional de uma variável (no presente caso, de um retorno). É uma variável que mostra a intensidade e a frequência das oscilações das cotações de um ativo financeiro, considerando-se um determinado período de tempo, sendo esta uma das formas mais, frequentemente, utilizadas na mensuração de seu risco.

Sendo assim, um dos grandes temas de estudo na atualidade, no âmbito de séries econômico-financeiras, tem sido a volatilidade, que, em razão de sua

importância, tem motivado a criação de diversos modelos a fim de tentar compreender e até mesmo prever o risco no mercado financeiro.

Uma das primeiras formulações sobre modelos de volatilidade foi o modelo de Black-Scholes (1973), que era utilizado no apuração de opções e assumia a volatilidade como um componente constante ao longo do tempo. O fato é que a consagração do uso desse modelo pelo mercado acabou permitindo o surgimento de modelos mais sofisticados para a análise da volatilidade.

Engle (1982) propôs um modelo não linear na variância a fim de estudar a volatilidade na taxa de inflação do Reino Unido. O modelo ARCH (*Autoregressive Conditional Heteroscedasticity*), introduzido pelo autor, generaliza a suposição de variância constante para um horizonte de previsão de um período, levando em consideração que o passado recente fornece informação relevante sobre esta previsão.

Bollerslev (1986) propôs uma generalização do processo ARCH, denominado modelo GARCH (*Generalized Autoregressive Conditional Heteroscedasticity*), cuja principal contribuição foi a inclusão de volatilidades anteriores na fórmula da variância condicional, que permitiu ao modelo ser mais parcimonioso. Swaray (2002) argumenta que a inclusão de variâncias condicionais defasadas pode capturar a aprendizagem adaptativa do processo.

Bollerslev, Chou e Kroner (1992) apresentam uma vasta revisão da teoria e evidências empíricas da aplicação em finanças dos modelos ARCH/GARCH, que, por sua vez, são simétricos, isto é, tratam os retornos positivos e negativos de forma similar. Desta maneira, criou-se a necessidade de suprir estas fraquezas.

A fim de captar o efeito de assimetria dos retornos na volatilidade, Nelson (1991) apresentou o modelo EGARCH (*Exponential GARCH*). Sua ideia foi utilizar a formulação exponencial para a volatilidade, além de introduzir uma no modelo que trata de forma assimétrica os retornos positivos e negativos.

Desta forma, ao longo dos anos foram surgindo diversos modelos para o estudo da volatilidade, como o modelo TARARCH (*Threshold ARCH*) de Zakoïan (1994), o QGARCH (*Quadratic GARCH*) em função de Sentana (1995), o APARCH (*Asymmetric power ARCH*), proposto por Ding, Granger e Engle (1993), todos eles buscando capturar a assimetria dos choques, isto é, assim como o modelo EGARCH buscam, de alguma forma, tratar de maneira assimétrica os retornos negativos e positivos.

Outra característica observada em estudos acerca da modelagem da volatilidade de séries financeiras, como retornos de ações, por exemplo, em que o efeito persistente de alterações na variância condicional, isto é, observa-se que as autocorrelações amostrais dos retornos ao quadrado tendem a decair lentamente em contraste com o decaimento exponencial observado em modelos do tipo GARCH, sugerindo existência de memória longa na variância destas séries.

Dada a existência de uma série temporal que representa a volatilidade (como quadrados dos retornos ou dos retornos absolutos, por exemplo), os modelos ARFIMA (“*Autoregressivos Fracionários Integrados e de Médias Móveis*”) podem ser usados com o intuito de modelar o comportamento de memória longa observado na volatilidade de retornos de ações. Agora, mais recentemente, uma alternativa de incorporar memória longa nos modelos GARCH tem-se sido utilizados os modelos FIGARCH (“*Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity*”), introduzidos por Baillie et al. (1996), podem utilizados para modelar a dinâmica dos retornos, juntamente com a sua volatilidade.

O modelo FIEGARCH (“*Fractionally Integrated Exponencial Generalized Autoregressive Conditional Heteroskedasticity*”), introduzido por Bollerslev e Mikkelson (1996), que além da volatilidade variando com o tempo e presença de *clusters* de volatilidade, o modelo leva em consideração a longa

dependência da volatilidade (assim como o modelo FIGARCH), também considera a assimetria (a parte exponencial do modelo EGARCH).

Um ponto importante é que nem sempre a suposição de normalidade para os erros dos modelos heterocedásticos condicionais são suficientes para captar as caudas pesadas e a assimetria dos retornos financeiros. Para isso, o uso de distribuições com caudas mais pesadas, como a t de Student, têm sido empregadas. A utilização de distribuições assimétricas, também, tem sido utilizada, em especial a t de Student assimétrica (skew- t).

Objetivou-se, principalmente, neste trabalho, avaliar se a distribuição skew- t como componente do erro em modelos heterocedásticos (GARCH, FIGARCH, FIEGARCH) tem a capacidade de melhorar as estimativas e previsões em modelos de volatilidade, além do fato de discutir acerca dos modelos ARMA-GARCH combinados, verificando se os mesmos, com inovações skew- t , têm a capacidade de capturar o comportamento de memória longa em séries de volatilidade.

1.1 Objetivos específicos

- a) Considerar a estimação de modelos combinados, em especial o ARMA – GARCH, com o modelo GARCH com inovações skew- t ;
- b) Avaliar a persistência na volatilidade considerando modelos GARCH com inovações t de Student e skew- t ;
- c) Verificar se os modelos ARMA-GARCH com erros skew- t podem capturar dependência de longo alcance em séries de volatilidade

2 REFERENCIAL TEÓRICO

2.1 Conceitos Preliminares

Inicialmente, serão apresentados diversos conceitos necessários à compreensão e ao estudo dos modelos de séries temporais discutidos neste trabalho.

2.1.1 Séries Temporais

Série temporal é qualquer conjunto de observações ordenadas no tempo. Isto é, uma sequência de pontos (dados) em ordem sucessiva, geralmente ocorrendo em intervalos regulares. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes, sendo de interesse analisar e modelar esta dependência. Enquanto em modelos de regressão, por exemplo, a ordem das observações é irrelevante para a análise, em séries temporais a ordem dos dados é crucial.

A representação gráfica de uma série temporal é chamada de trajetória. Conforme Morettin e Tolo (2004), o que comumente é chamado de série temporal é *parte* de uma trajetória, dentre muitas que poderiam ter sido observadas.

As séries temporais financeiras se diferem das demais séries temporais por apresentarem características peculiares, como por exemplo, o fato de não serem serialmente correlacionadas, mas dependentes.

Conforme Morettin (2011), uma característica presente em séries de ativos financeiros é o que se convencionou chamar de *volatilidade*, que pode ser definida de várias formas, porém não é diretamente observável.

Aiube (2013) argumenta que séries temporais financeiras apresentam comportamentos típicos, tais como a fraca dependência linear e forte dependência não linear, além de apresentarem caudas pesadas e/ou excesso de curtose.

2.1.2 Processo estocástico

Um processo estocástico é uma família de variáveis aleatórias $\{X(t), t \in T\}$ definidas em um espaço de probabilidades $(\Omega, \mathcal{A}, \mathbb{P})$, de forma que $T \neq \emptyset$ seja um conjunto de índices arbitrário, podendo ser discreto ou contínuo, Ω é o espaço amostral e \mathcal{A} é a σ -álgebra de eventos aleatórios, com \mathbb{P} sendo a medida de probabilidade definida em \mathcal{A} .

O conjunto T é, geralmente, tomado como sendo o conjunto dos inteiros \mathbb{Z} ou o conjunto dos reais \mathbb{R} .

Para cada $t \in T$, $X(t)$ é uma variável aleatória definida em Ω , de forma que $X(t)$ é função de dois argumentos, $X(t, \omega)$, com $t \in T$ e $\omega \in \Omega$. Sendo assim, para cada $t \in T$, tem-se uma variável aleatória $X(t, \omega)$, com uma função densidade de probabilidades $f_t(x)$ (desde que exista). Por outro lado, para cada $\omega \in \Omega$ (fixo), ter-se-á uma função de t , isto é, uma trajetória ou realização do processo.

Sejam t_1, t_2, \dots, t_n elementos quaisquer de T e considerando:

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = P\{X(t_1) \leq x_1, \dots, X(t_n) \leq x_n\}. \quad (1)$$

Logo, o processo estocástico $\{X(t), t \in T\}$ estará especificado se forem conhecidas as distribuições finito-dimensionais (1), para todo $n \geq 1$. Por exemplo, para $n = 1$ tem-se as distribuições unidimensionais da variável aleatória $X(t_1)$.

A média de $X(t)$ é dada por:

$$\mu(1;t) = \mu(t) = E\{X(t)\} = \int_{-\infty}^{\infty} x dF(x;t). \quad (2)$$

A função de autocovariância de $X(t)$ é definida da forma:

$$\begin{aligned} & \mu(1;1;t_1,t_2) - \mu(1;t_1)\mu(1;t_2) = \tau(t_1,t_2) \\ & = E\{X(t_1)X(t_2)\} - E\{X(t_1)\}E\{X(t_2)\}, \quad t_1, t_2 \in T. \end{aligned} \quad (3)$$

No caso de $t_1 = t_2 = t$, por (3) ter-se que:

$$\tau(t,t) = \text{Var}\{X(t)\} = E\{X^2(t)\} - E^2\{X(t)\}, \quad (4)$$

que é a variância do processo $X(t)$.

Para uma amostra X_1, \dots, X_n , tem-se que o estimador de autocovariância entre variáveis defasadas por k intervalos de tempo é:

$$\hat{\tau}_k = \frac{1}{n} \sum_{t=1}^{n-k} [(X_t - \bar{X})(X_{t+k} - \bar{X})], \quad k = 0, 1, \dots, n-1, \quad (5)$$

Sendo $\bar{X} = (1/n) \sum_{t=1}^n X_t$.

Função de autocorrelação (FAC) do processo é definida por:

$$\rho_k = \frac{\tau_k}{\tau_0} = \frac{\text{Cov}(X_t, X_{t+k})}{\text{Var}(X_t)}, \quad (6)$$

Em que τ_0 é a variância da série. O estimador de ρ_k pode ser obtido como:

$$\hat{\rho}_k = \frac{\hat{\tau}_k}{\hat{\tau}_0}, \quad (7)$$

Sendo $\hat{\tau}_k$ dado por (5) e $\hat{\tau}_0 = (1/n) \sum_{t=1}^T (X_t - \bar{X})^2$.

A função de autocorrelação parcial (FACP) é a correlação parcial entre as variáveis X_t e X_{t+k} dado que são conhecidos $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$, ou seja, mede a correlação remanescente entre X_t e X_{t+k} após eliminar a influência de $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$.

Um processo estocástico importante é o ruído branco. Diz-se que $\{\varepsilon(t), t \in T\}$ é um ruído branco se:

- a) $E[\varepsilon(t)] = 0$, isto é, a média de ε_t é constante para todo $t \in T$;
- b) $\text{Var}[\varepsilon(t)] = \sigma_\varepsilon^2$, para todo $t \in T$;

- c) $\text{Cov}[\varepsilon(t), \varepsilon(s)] = 0$, para todo $t \neq s$, ou seja, $\varepsilon(t)$ é não correlacionado.

Estas considerações representam uma coleção de variáveis aleatórias não correlacionadas, cuja média é nula e variância finita e constante. Usualmente, a notação adotada é $\varepsilon(t) \sim RB(0, \sigma_\varepsilon^2)$ para denotar este particular processo estocástico.

2.1.3 Estacionariedade

Geralmente faz-se a suposição de que uma série temporal é estacionária, isto é, que ela se desenvolve ao longo do tempo, aleatoriamente, ao redor de uma média e variância constantes, evidenciando alguma forma de equilíbrio estável. Porém, nem sempre isso é verdade, uma vez que a maioria das séries encontradas na prática não exhibe tal comportamento.

Segundo Bueno (2011), o conceito de estacionariedade é a principal ideia que se deve ter para estimar uma série temporal. É fundamentalmente a constatação de estacionariedade que permitirá proceder a inferências estatísticas, acerca dos parâmetros estimados, com base na realização de um processo estocástico.

Sendo assim, se o processo estocástico que gerou a série de observações é invariante no tempo, diz-se que é estacionário. Intuitivamente, se $X(t)$ é uma série temporal estacionária, então, as características de $X(t + \delta)$, para todo $\delta > 0$, são as mesmas para $X(t)$. Há duas formas de estacionariedade: a fraca e a estrita.

Um processo $X = \{X(t), t \in T\}$ é denominado fracamente estacionário ou com estacionariedade de segunda ordem se e somente se:

- a) $E\{X(t)\} = \mu(t) = \mu$ (constante), para todo $t \in T$;
- b) $E\{X^2(t)\} < \infty$, para todo $t \in T$;
- c) $\text{Cov}\{X(t_1), X(t_2)\} = \tau(t_1, t_2)$ é função de $|t_1 - t_2|$ (defasagem).

Agora, um processo estocástico $X = \{X(t), t \in T\}$ é dito estritamente estacionário se todas as distribuições conjuntas finito-dimensionais dadas por $F(x_1, \dots, x_n; t_1, \dots, t_n) = P\{X(t_1) \leq x_1, \dots, X(t_n) \leq x_n\}$ permanecerem as mesmas sob translações no tempo, isto é, $F(x_1, \dots, x_n; t_1 + \delta, \dots, t_n + \delta) = F(x_1, \dots, x_n; t_1, \dots, t_n)$ para quaisquer $t_1, \dots, t_n, \delta \in T$.

Intuitivamente, estacionariedade estrita indica que os gráficos da função de distribuição da série em quaisquer dois intervalos de tempo de igual dimensão possuirão propriedades estatísticas similares.

2.1.4 Ergodicidade

Para a modelagem de séries temporais e sua estimação, é necessário satisfazer a propriedade de ergodicidade. Esta permite usar uma série temporal para calcular as médias em cada instante de tempo. Como as médias são todas iguais, basta uma única realização da série para possibilitar o cálculo.

Suponha que uma particular realização, s , de um processo estocástico, justamente a única série que se observa. Sendo assim, a média temporal dessa realização é dada por:

$$\bar{X}^{(s)} = \frac{1}{n} \sum_{t=1}^n X_t^{(s)}. \quad (8)$$

Se $\bar{X}^{(s)}$ convergir para $E(X_t)$, existe ergodicidade. Em outras palavras, se a média temporal convergir para a média não condicional, haverá ergodicidade. Desta forma, a série temporal pode ser estimada normalmente, mesmo com apenas uma realização do processo estocástico.

2.1.5 Processo Martingale

De acordo com Morettin (2011), um processo $X = \{X_t, F_t, t \in T\}$, sendo X_t uma variável aleatória definida em um espaço de probabilidade (Ω, F, P) ; $\{F_t, t \in T\}$ uma família crescente de sub- σ -álgebras de F , ou seja, de $F_s \subset F_t$, se $s \leq t$, e que X seja adaptado a essa família, isto é, X_t é F_t – mensurável, é um processo martingale se:

- a) X_t é integrável para todo $t \in T$;
- b) Se $s \leq t$, então, $E(X_t | F_s) = X_s$.

2.1.6 Operadores

De acordo com Morettin e Toloi (2006), alguns operadores são comumente utilizados no desenvolvimento dos modelos de Box e Jenkins. São eles:

- a) operador de retardo, defasagem ou translação para o passado, denotado por B , definido como:

$$BX_t = X_{t-1}; B^2 X_t = X_{t-2}; \dots; B^m X_t = X_{t-m}.$$

- b) operador translação para o futuro, denotado por F e definido da forma: $FX_t = X_{t+1}; F^2 X_t = X_{t+2}; \dots; F^m X_t = X_{t+m}$.

- c) operador diferença, denotado por Δ e definido por:

$$\Delta X_t = X_t - X_{t-1} = (1 - B) X_t.$$

- d) operador soma, denotado por S e definido como:

$$SX_t = \sum_{j=0}^{\infty} X_{t-j} = X_t + X_{t-1} + \dots = (1 + B + B^2 + \dots) X_t \text{ de onde}$$

$$\text{segue que: } SX_t = (1 - B)^{-1} X_t = \Delta^{-1} X_t.$$

Normalmente uma ou duas diferenças sucessivas é necessário para tornar a série estacionária.

2.1.7 Processos autorregressivos – AR(p)

Um processo $\{X_t, t \in \mathbb{Z}\}$ será dito autorregressivo de ordem p , denotado por $X_t \sim AR(p)$, se satisfizer à equação de diferenças:

$$X_t - \mu = \phi_1 (X_{t-1} - \mu) + \dots + \phi_p (X_{t-p} - \mu) + \varepsilon_t, \quad (9)$$

Em que $\mu, \phi_1, \dots, \phi_p$ são parâmetros reais e $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$.

A equação (9) pode ser reescrita de forma alternativa por:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t, \quad (10)$$

de forma que:

$$\mu = E(X_t) = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}. \quad (11)$$

2.1.8 Processos de médias móveis – MA(q)

Um processo $\{X_t, t \in \mathbb{Z}\}$ será dito de médias móveis de ordem q , denotado por $X_t \sim MA(q)$, se satisfizer à equação de diferenças:

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (12)$$

Sendo $\mu, \theta_1, \dots, \theta_q$ constantes reais e $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$.

2.1.9 Processo Autorregressivo de Médias Móveis – ARMA (p, q)

O processo autorregressivo de médias móveis é apenas a combinação dos processos AR (p) e MA (q). Assim, um ARMA (p, q) pode ser escrito da forma:

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}, \quad (13)$$

Em que $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$. Segue-se que a média do processo é μ .

Usando os operadores definidos em (2.1.6), pode-se reescrever (13) da forma:

$$\phi(B)\tilde{X}_t = \theta(B)\varepsilon_t, \quad (14)$$

Em que $\tilde{X}_t = X_t - \mu$.

De acordo com Morettin e Toloi (2006), um modelo bastante utilizado é o ARMA (1,1), de forma que:

$$p = q = 1,$$

$$\phi(B) = 1 - \phi_1 B,$$

$$\theta(B) = 1 - \theta_1 B,$$

que pode ser escrito como:

$$X_t = \phi_1 X_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}. \quad (15)$$

Na prática, os processos sobre os quais são feitas inferências, por meio de uma série temporal, são geralmente não estacionários. Neste trabalho, serão tratados os processos não estacionários homogêneos, ou seja, processos cuja diferenciação produz processos estacionários.

Seja então Z_t um processo não estacionário e X_t um processo estacionário obtido de Z_t por diferenciações sucessivas. Inversamente, pode-se dizer que Z_t é obtido com base em X_t por integração. Assim, tem-se que:

$$\Delta^d Z_t = X_t, \quad (16)$$

Em que d denota o número de diferenciações. O processo estacionário X_t pode ser representado por um modelo ARMA (p, q), isto é,

$$\phi_p(B)\Delta^d Z_t = \theta_q(B)\varepsilon_t. \quad (17)$$

Diz-se, então, que o modelo (17) é um modelo autoregressivo integrado de médias móveis, ou ARIMA (p, d, q).

A metodologia Box e Jenkins (1970) é a interpretação e análise de uma série temporal oriunda de uma realização de um processo estocástico. O objetivo é inferir sobre o processo gerador de dados, fazer previsões futuras da série, levando-se em consideração a parcimônia do modelo, ou seja, tratando o modelo com o menor número de parâmetros possíveis. A estratégia envolve a repetição do processo de identificação até encontrar o modelo que seja mais o mais acurado possível. As bases desta formulação podem ser encontradas no texto de Box e Jenkins (1970). Esta metodologia pode ser resumida nos seguintes passos:

- a) Identificar as ordens p e q do modelo;
- b) Estimar o modelo;

- c) Verificar se os resíduos estimados não rejeitam a hipótese nula de que seja um ruído branco. Caso não rejeitem, passa-se ao próximo passo, do contrário, retorna-se ao primeiro passo;
- d) Finalmente, há a verificação ou diagnóstico do modelo ajustado, por meio de uma análise de resíduos, para se saber se este é adequado ou não.

Segundo Morettin e Tolo (2006), um procedimento que, muitas vezes, é utilizado é identificar não só um único modelo, mas alguns modelos que serão então estimados e verificados. A fase crítica do procedimento é a identificação, pois é possível que vários pesquisadores identifiquem modelos diferentes para a mesma série temporal.

Geralmente, utiliza-se o princípio da parcimônia, isto é, buscam-se modelos que possuam um número reduzido de parâmetros e as previsões obtidas são precisas, comparando-se, favoravelmente, com os demais métodos de previsão.

2.1.10 Séries de Retornos

A maioria dos estudos em séries financeiras concentra-se na análise da série de retornos em vez da série de preços. Tal fato, conforme Campbell, Lo e Mckinlay (1997), estão relacionados a duas questões essenciais. A primeira é em virtude do fato de os retornos de um ativo conter informações que atendam aos interesses dos investidores. Em segundo, a série de retornos possui propriedades, estaticamente, mais atrativas que a série de preços, isto é, a estacionaridade e ergodicidade.

O conceito de ergodicidade tem a ver com a possibilidade de estimar, de forma consistente, a média ou outros momentos de conjunto com base em uma única realização que se dispõe do processo estocástico.

O retorno de um ativo R_t entre os instantes de tempo $t-1$ e t , sendo P_t o preço de um ativo no instante t é dado por:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1. \quad (18)$$

O log-retorno, denotado por y_t é definido como:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(1 + R_t) = p_t - p_{t-1}, \quad (19)$$

em que $p_t = \ln P_t$.

A diferença básica entre o uso do retorno discreto (18) e o log-retorno (19) ou o retorno contínuo, está na interpretação dada pelo desvio padrão. O desvio padrão do retorno discreto é chamado de risco e o do retorno contínuo, volatilidade. Para pequenos intervalos de tempo, os valores dos retornos de (18) e (19) serão em geral valores próximos.

2.1.11 Fatos estilizados sobre retornos financeiros

Fatos estilizados são regularidades estatísticas observadas em um grande número de séries temporais financeiras de retornos por meio de estudos empíricos em inúmeros mercados (AIUBE, 2012).

Nicolau (2012) argumenta que vários estudos empíricos têm documentado que há um conjunto de regularidades observadas que são

partilhadas por um leque grande de séries temporais financeiras. Essas regularidades têm a ver com características das distribuições marginais e condicionais que são comuns a inúmeras séries temporais econômico-financeiras.

Os retornos financeiros apresentam características peculiares que outras séries não possuem como o fato de, raramente, apresentarem tendências ou sazonalidades, com exceção de retornos intradiários e séries de taxas de câmbio ou de juros que podem apresentar tendências que variam com o tempo. Os principais fatos estilizados, citados por Morettin (2011), podem ser resumidos em:

- a) retornos são, em geral, não autocorrelacionados, isto é, a correlação entre um retorno R_t e R_{t-1} é nula;
- b) os quadrados dos retornos são autocorrelacionados, com uma correlação de lag pequena e depois uma queda lenta das demais;
- c) a distribuição (não condicional) dos retornos apresenta caudas mais pesadas do que uma distribuição normal; além disso, a distribuição, embora, aproximadamente, simétrica, é, em geral, leptocúrtica;
- d) algumas séries de retornos são não-lineares, isto é, respondem de forma diferente a choques grandes ou pequenos¹, ou a choques negativos ou positivos.

Desta forma, é possível observar que valores muito altos e muito baixos ocorrem com frequência em séries financeiras. Valores extremos aparecem de forma seguida (*volatility clustering*). Assim, se a volatilidade é alta (baixa),

¹ Choque em economia está relacionado a uma modificação súbita do curso normal dos acontecimentos, isto é, alteração do comportamento usual, ou ameaça à continuidade de certa situação. Por exemplo, desvalorização cambial ou aumento drástico da inflação.

então, é razoável esperar que a mesma se mantenha alta (baixa) durante certo intervalo de tempo. Na Figura 1 têm-se os retornos diários do índice RTSI² no período de 1º de setembro de 1995 a 5 de maio de 2014. Estão identificados alguns períodos de grande volatilidade. Estes se prolongam por até vários anos. Outro ponto importante é que períodos de intensa volatilidade coincidem com épocas nas quais ocorreram crises importantes, como neste caso, a crise da Ásia em outubro de 1997, a moratória russa em agosto de 1998 e crise mundial de 2008 e 2009.

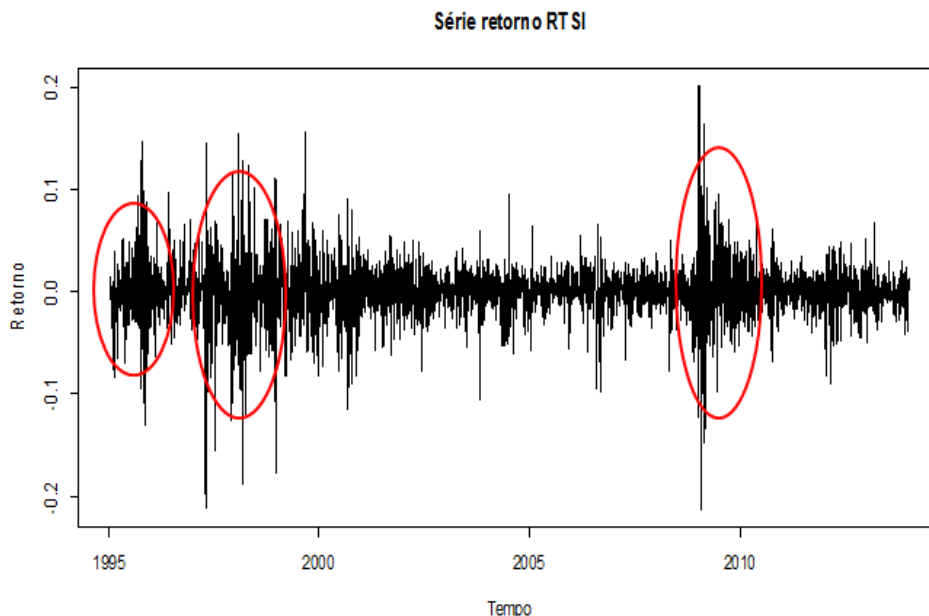


Figura 1 Série temporal dos retornos do índice diário RTSI da Bolsa de Valores de Moscou no período de setembro de 1995 a maio de 2014

² O índice RTSI é o principal índice de referência do mercado de capitais da Rússia. Calculado desde 1º de setembro de 1995, o RTSI tem em sua composição as 50 maiores ações mais negociadas na Bolsa de Valores RTS (Russian Trading System Stock Exchange).

É usual o tratamento do retorno como variáveis aleatórias contínuas e, neste caso, com função densidade conjunta da forma:

$$f(r_{i1}, r_{i2}, \dots, r_{iT}; \theta) = f(r_{i1}) f(r_{i2} | r_{i1}) \dots f(r_{iT} | r_{iT-1} \dots r_{i1}),$$

ou ainda,

$$f(r_{i1}, r_{i2}, \dots, r_{iT}; \theta) = f(r_{i1}; \theta) \prod_{t=2}^T f(r_{it}; r_{it-1} \dots, r_{i1}; \theta). \quad (20)$$

A questão importante é que a observação de como a distribuição de dado ativo evolui no tempo, isto é, a especificação da distribuição condicional é grande relevância para a proposição de modelos para o estudo da volatilidade ou a previsão do comportamento deste ativo.

2.1.12 Volatilidade

A volatilidade pode ser entendida como uma medida da variabilidade dos preços de um ativo com relação à sua média em certo período de tempo. Esta medida pode ser obtida, por meio da variância ou pelo desvio padrão de uma série temporal de retornos diário, semanal, mensal, anual, sendo expressa sob a forma de porcentagem.

Um dos temas bastante debatidos na atualidade no âmbito das Finanças tem sido a volatilidade, em razão de sua relevância, à sua eventual previsibilidade de forma a melhor compreender o comportamento do mercado e reduzir implicações ao nível de risco/rentabilidade, assessorando os investidores na tomada de decisão. Sua utilização como uma medida de risco dos ativos (conjunto de bens, valores, créditos, direitos e afins que forma o patrimônio de

uma pessoa física ou jurídica, num determinado momento) que circulam nos mercados financeiros e o conhecimento dos comportamentos que condicionam a sua evolução tem sido motivo de vários estudos ao longo dos anos por parte de vários autores.

Outrora, volatilidade era tida como movimentos incertos e irregulares numa dada série temporal, que não seguem um padrão compreendido. Kendall (1953) concluiu que o seu comportamento verificava movimentos completamente aleatórios. Porém, estudos como os de Engle (1982) e Bollerslev (1986) verificaram que a volatilidade de uma série temporal não é constante, o que levou a concluir que ela não era homocedástica (variância constante).

Desta forma, sendo a volatilidade uma variável não observável diretamente, está relacionada a algumas propriedades ou alguns fatos estilizados que são bem definidos na literatura. Pode-se dizer que a volatilidade em séries financeiras não é constante ao longo do tempo, e, portanto, responsável pelo comportamento heterocedástico. Períodos de alta volatilidade são seguidos por períodos de alta volatilidade, já períodos de baixa volatilidade são seguidos por períodos mais amenos.

A volatilidade dos mercados financeiros pode repercutir amplamente na economia como um todo. Os incidentes causados pelos ataques terroristas em 11 de setembro de 2001 e os escândalos financeiros recentes nos EUA causaram grande turbulência nos mercados financeiros internacionais e impactos negativos na economia mundial. No Brasil, o segundo semestre de 2002 mostrou o impacto da instabilidade política sobre a volatilidade nos mercados financeiros, causando e dissipando danos na economia real. Esses fatos são evidências claras da relação existente entre incerteza nos mercados financeiros e a confiança pública (POON; GRANGER, 2003).

Sendo assim, a volatilidade pode ser apresentada sob diversas formas, e é importante saber distinguir qual é a abordagem em questão quando se trata

deste tema. Dentre as abordagens no estudo da volatilidade tem-se a volatilidade histórica ou realizada, a volatilidade implícita e a volatilidade futura.

A volatilidade histórica ou realizada mede as flutuações nos preços ocorridos no passado, podendo ser utilizada como medida de risco total de dado ativo. É calculada por meio da variância ou desvio padrão do retorno de dado ativo.

A volatilidade implícita é utilizada apenas nos contratos de opções (instrumentos financeiros classificados como derivativos), no qual indica o “comportamento” do mercado acerca de contratos de compra e venda de derivativos.

Já a volatilidade futura corresponde à volatilidade que melhor descreve a distribuição futura dos preços de um determinado ativo. Caso pudesse ser determinada, o investidor conheceria a real função de distribuição de probabilidade dos preços de dado investimento financeiro, o que não ocorre na prática.

Uma das questões importantes no estudo de modelos de volatilidade tem sido a assimetria. Este é um fenômeno em que descidas inesperadas no preço de ativos aumentam a volatilidade mais do que subidas do preço de igual dimensão.

Os primeiros autores a constatar este fato foram Black (1976) e Christie (1982), que explicaram a assimetria com o efeito de alavanca, que significa que uma queda no valor de ações, por exemplo, aumentava o efeito de alavanca financeira que, por sua vez, aumentava o risco das ações assim como sua volatilidade. Posteriormente, esta teoria foi confirmada por Nelson (1991) e Glosten et al. (1993).

Assim, os modelos de volatilidade consideram dois tipos de assimetria usualmente encontrados em séries temporais financeiras: a assimetria das perturbações e o efeito de alavanca. O primeiro tipo de assimetria é utilizado para considerar um dos fatos estilizados de que as perdas têm distribuição com

cauda mais pesada do que ganhos. O segundo tipo de assimetria, efeito de alavanca, leva em consideração que as perdas têm uma influência maior na volatilidade do que ganhos (AURÉLIO, 2012).

Vários modelos têm sido propostos na literatura para modelar a volatilidade. O fato é que estes modelos devem ser capazes, além de refletir os impactos assimétricos na volatilidade, se possível fornecer previsões úteis para tomada de decisão acerca dos riscos do mercado financeiro.

2.1.13 Histograma e Gráfico QxQ Plot

O histograma é um gráfico composto por retângulos justapostos, baseado na divisão do espaço amostral em intervalos, geralmente, com mesmo comprimento. Com base nos dados x_1, \dots, x_n o histograma pode ser definido como:

$$H(x) = \sum_{i=1}^n I\{x - \tilde{x}_i; h\}, \quad (21)$$

Sendo \tilde{x}_i o centro do intervalo no qual a observação x_i está e $I\{z; h\}$ o indicador do intervalo $[-h, h]$. Deve ser feito algum tipo de escalonamento de forma que a área do histograma seja um.

Morettin (2011) argumenta que há críticas quanto ao uso do histograma. As principais dizem respeito ao fato de que o comportamento do histograma depende da escolha de h e da posição inicial da grade, e, além disso, informação é perdida, uma vez que ao substituir x_i pelo ponto médio do intervalo ao qual ele pertence. Desta forma, a fim de evitar este tipo de dificuldade, há

estimadores da densidade $f(x)$ que podem ser usados. Um dos mais utilizados substitui retângulos por uma função *núcleo* (“*kernel*”), mais suave, da forma:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n k(x - x_i; h), \quad (22)$$

em que k é o núcleo, em geral, também, é uma função densidade, cuja variância é controlada por h , denotada por largura de faixa. O comportamento de \hat{f} ainda irá depender de h , de modo que a crítica ao histograma, ainda, continua.

O gráfico QxQ plot é o gráfico dos pares ordenados:

$$\{(q_\alpha, \tilde{q}_\alpha), \alpha \in (0,1)\} \quad (23)$$

no qual q_α é o quantil de ordem α da distribuição normal padrão ($Z \sim N(0,1)$), podendo, eventualmente, ser outra distribuição; \tilde{q}_α é o quantil da distribuição empírica associada à variável padronizada da forma:

$$x_t = \frac{r_t - \bar{r}}{\bar{\sigma}} \quad (24)$$

Sendo r_t o retorno de dado ativo, \bar{r} a média do retorno e $\bar{\sigma}$ o seu desvio padrão médio. Desta forma:

$$q_\alpha : P(Z < q_\alpha) = \alpha, \quad Z \sim N(0,1) \quad \tilde{q}_\alpha : \tilde{P}(x_t < \tilde{q}_\alpha) = \alpha, \quad x_t.$$

Assim, se x_i tem a mesma distribuição de Z , então $q_\alpha \approx \tilde{q}_\alpha$ e o gráfico em (23) será uma reta de 45° . Se, pelo contrário, as distribuições de x_i e Z são diferentes, o gráfico afastar-se-á da reta de 45° (pode-se, também, comparar a distribuição de x_i com outra qualquer).

De acordo com Morettin (2011), há dois tipos de Q x Q plots: teóricos e empíricos. O primeiro tipo é usado para verificar se um conjunto de dados vem de determinada distribuição. O segundo tipo é usado para verificar se dois conjuntos de dados têm uma mesma distribuição. Para verificar se um dado conjunto de dados provém de uma distribuição especificada, considera-se o gráfico em que, no eixo horizontal, são colocados os quantis teóricos da distribuição hipotetizada para os dados e, no eixo vertical, os quantis empíricos dos dados. Se as observações, realmente, são provenientes da distribuição em questão, os pontos deverão estar distribuídos ao longo de uma reta.

Na Figura 2 tem-se em (a) o histograma dos retornos diários do RTSI da Bolsa de Valores de Moscou, com a curva normal da densidade da distribuição normal padrão. Vê-se que o histograma tem a parte central mais alta do que a normal e há valores afastados da posição central dos dados. Este fato é descrito pela chamada medida de curtose, que é descrito na seção 2.11, em que se pode dizer que os dados são leptocúrticos (caudas mais “pesadas” que a normal). Em (b) está representado o gráfico QxQ plot com respeito aos quantis da distribuição normal padrão. Se os dados fossem aproximadamente normalmente distribuídos, os pontos estariam sobre uma reta, o que não acontece neste caso.

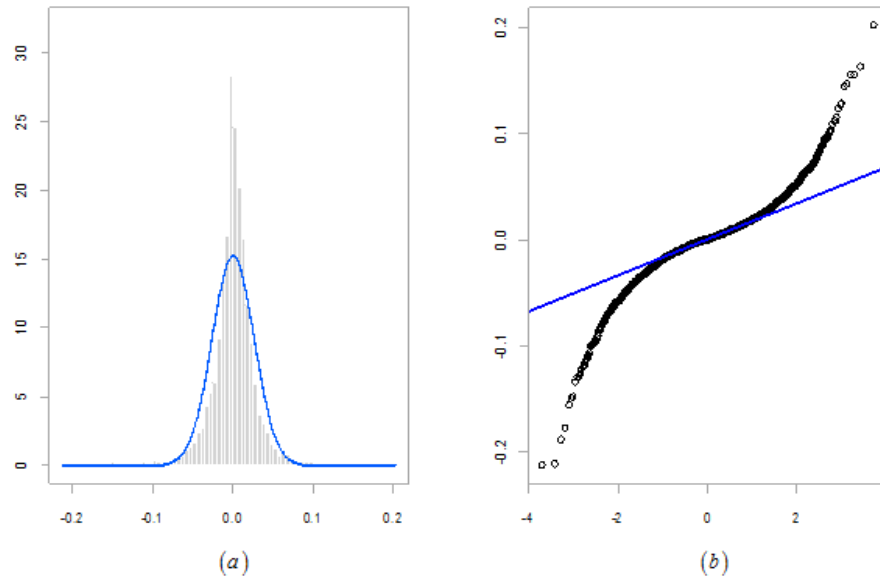


Figura 2 (a) Histograma e (b) gráfico QxQ plot dos retornos do índice diário RTSI da Bolsa de Valores de Moscou no período de setembro de 1995 a maio de 2014

2.1.14 Assimetria e Curtose

Seja X uma variável aleatória com média μ e variância σ^2 . A assimetria de X é definida como:

$$A(X) = E \left[\frac{(X - \mu)^3}{\sigma^3} \right], \quad (25)$$

Enquanto o coeficiente de curtose de X é dado por:

$$K(X) = E \left[\frac{(X - \mu)^4}{\sigma^4} \right]. \quad (26)$$

Para a distribuição normal, tem-se que $A=0$ e $K=3$. Distribuições com coeficiente de assimetria maior que zero ($A>0$) ou menor que zero ($A<0$), são assimétricas à direita e à esquerda, respectivamente. Distribuições com curtose maior que três possuem caudas pesadas ou “gordas”, isto é, atribuem maior probabilidade aos eventos que ocorrem em suas caudas.

Considerando uma amostra aleatória X_1, \dots, X_T de X . O r – ésimo momento amostral:

$$m_r = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^r,$$

Em que $\hat{\mu} = \bar{X}$. Substituindo os momentos de X pelos respectivos momentos amostrais, têm-se os estimadores:

$$\hat{A}(X) = \frac{m_3}{m_2^{3/2}} = \frac{1}{T} \sum_{t=1}^T \left(\frac{X_t - \bar{X}}{\hat{\sigma}} \right)^3, \quad (27)$$

$$\hat{K}(X) = \frac{m_4}{m_2^2} = \frac{1}{T} \sum_{t=1}^T \left(\frac{X_t - \bar{X}}{\hat{\sigma}} \right)^4, \quad (28)$$

respectivamente, sendo $\hat{\sigma}^2 = \sum_{t=1}^T (X_t - \bar{X})^2 / T$. Na prática, utiliza-se (27) e (28)

para obter informações sobre a assimetria e curtose das séries de retornos analisadas.

2.2 Distribuições Assimétricas

Na literatura existem várias propostas para se inserir assimetria em distribuições simétricas, a fim de torná-las assimétricas. Pode-se citar Azzalini (1985), Fernandez e Steel (1998), Branco e Dey (2001), Azzalini e Capitanio (2003) e Jones e Faddy (2003). Porém, em função da simplicidade e generalidade, será utilizado o método proposto por Fernandez e Steel (1998). Neste, o método dos momentos são fáceis de serem obtidos, não exigindo encontrar a função de distribuição acumulada.

Uma distribuição é dita simétrica, quando apresenta o mesmo valor para a moda, a média e a mediana. Quando esta igualdade não acontece, tem-se uma distribuição assimétrica.

A assimetria é o grau de afastamento que uma distribuição apresenta do seu eixo de simetria. Este afastamento pode acontecer do lado esquerdo ou direito da distribuição considerada. A Figura 3 apresenta três ilustrações de tipos de assimetria apresentados por uma distribuição, em (a) tem-se uma distribuição assimétrica à esquerda, (b) distribuição simétrica e em (c) representa uma distribuição assimétrica à direita.

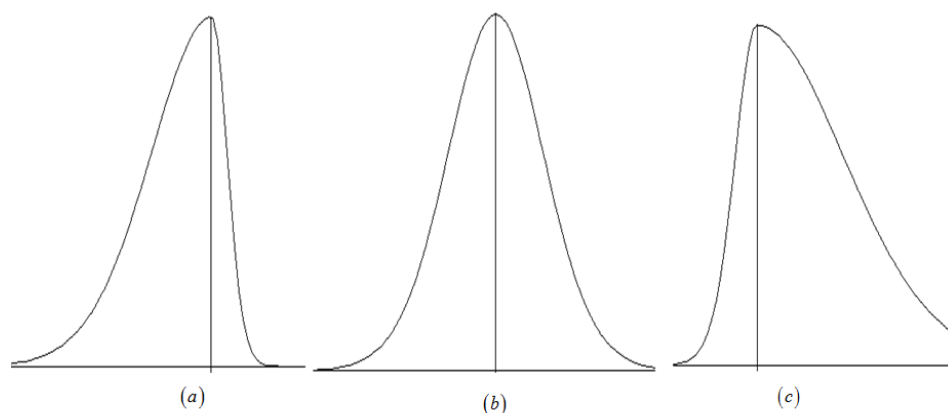


Figura 3 Formas apresentadas por uma distribuição

O método proposto em Fernandez e Steel (1998) é conhecido como método escala inversa de fatores. Neste é possível obter uma função de densidade $p(x|\gamma)$ assimétrica com grau de assimetria indexado pelo parâmetro $\gamma > 0$, com base em uma função densidade unimodal, definida na reta e simétrica em torno de zero. Quando o valor da variável aleatória é negativo tem-se $p(x|\gamma) \propto p(x\gamma)$, caso contrário, $p(x|\gamma) \propto p\left(\frac{x}{\gamma}\right)$. A função densidade de probabilidade $p(x|\gamma)$ é dada por:

$$\begin{aligned} p(x|\gamma) &= \frac{2}{\gamma + \gamma^{-1}} \left\{ p(x\gamma) I_{(-\infty, 0)}(x) + p\left(\frac{x}{\gamma}\right) I_{[0, \infty)}(x) \right\} \\ &= \frac{2}{\gamma + \gamma^{-1}} p\left(x\gamma^{-\text{sign}(x)}\right), \end{aligned} \quad (29)$$

Sendo $I_A(x)$ a função indicadora no conjunto A e $\text{sign}(x)$ a função sinal, que é igual a -1 , se $x < 0$ e 1 , se $x \geq 0$. Uma forma alternativa de escrever (29) é:

$$p(x|\gamma) = \frac{2}{\gamma + \gamma^{-1}} \begin{cases} p(x\gamma), & \text{para } x < 0, \\ p\left(\frac{x}{\gamma}\right), & \text{para } x \geq 0. \end{cases}$$

Em que $\gamma > 0$ é o parâmetro de assimetria e $p(x)$ é a densidade de uma distribuição simétrica.

As principais propriedades de $p(x|\gamma)$ são:

a) Se $\gamma = 1$, então, tem-se o caso simétrico, isto é,

$$p(x|\gamma=1) = p(x);$$

b) $p(x|\gamma)$ mantém a mesma moda de $p(x)$. Uma vez que $p(x)$ é unimodal e simétrica em torno de zero, a moda será sempre no zero;

c) A massa de probabilidade à esquerda e à direita de zero é independente de $p(x)$, ou seja,

$$P(X \geq 0) = \frac{2}{1+\gamma^2} \int_0^{\infty} p\left(\frac{x}{\gamma}\right) dx = \frac{\gamma^2}{1+\gamma^2},$$

$$P(X < 0) = 1 - P(X \geq 0) = 1 - \frac{\gamma^2}{1+\gamma^2} = \frac{1}{1+\gamma^2}, \quad \frac{P(X \geq 0)}{P(X < 0)} = \gamma^2;$$

d) A distribuição será assimétrica à direita se $\gamma > 1$ e assimétrica à esquerda se $\gamma < 1$. Para $\gamma = 1$ a distribuição é simétrica.

A existência dos momentos de $p(x|\gamma)$ depende dos momentos absolutos de $p(x)$. O r -ésimo momento é dado por:

$$E(X^r|\gamma) = \frac{\gamma^{r+1} + (-1)^r / \gamma^{r+1}}{\gamma + \gamma^{-1}} M_r, \quad (30)$$

Em que:

$$M_r = 2 \int_0^{\infty} x^r p(x) dx \quad (31)$$

O r -ésimo momento absoluto de $p(x)$.

A média e a variância de $p(x|\gamma)$ são dadas, respectivamente, por:

$$E(x|\gamma) = \mu = M_1 \left(\gamma - \frac{1}{\gamma} \right) \quad (32)$$

e

$$\text{Var}(x|\gamma) = \sigma^2 = (M_2 - M_1^2) \left(\gamma^2 + \frac{1}{\gamma^2} \right) + 2M_1^2 - M_2. \quad (33)$$

O coeficiente de assimetria de $p(x|\gamma)$ pode ser calculado por:

$$A(x|\gamma) = \left(\gamma + \frac{1}{\gamma} \right) \frac{(M_3 + 2M_1^3 - 3M_1M_2) \left(\gamma^2 + \frac{1}{\gamma^2} \right) + 3M_1M_2 - 4M_1^3}{\left[(M_2 - M_1^2) \left(\gamma^2 + \frac{1}{\gamma^2} \right) + 2M_1^2 - M_2 \right]^{3/2}}. \quad (34)$$

O coeficiente de curtose de $p(x|\gamma)$ pode ser calculado por:

$$K(x|\gamma) = \frac{E(x^4|\gamma) - 4E(x|\gamma)E(x^3|\gamma) + 6E(x^2|\gamma)E(x|\gamma)^2 - 3E(x|\gamma)^4}{\text{Var}(x|\gamma)^2}. \quad (35)$$

2.2.1 Distribuição skew-t

Na literatura há diferentes metodologias para se trabalhar com a distribuição skew-t (distribuição t de Student assimétrica). Podem-se citar Fernandes e Steel (1998) e Jones e Faddy (2003). A distribuição descrita por Fernandes e Steel (1998) preserva a moda e a de Jones e Faddy (2003) é baseada em uma transformação da distribuição beta.

Godoi (2007) trabalha a versão univariada da distribuição skew-t multivariada proposta por Branco e Dey (2001). Nesta versão, a distribuição utiliza a função distribuição acumulada da t de Student.

Antes de definir a skew-t, é importante lembrar qual é a função densidade de probabilidade da t de Student. Uma variável aleatória X tem distribuição t de Student com notação $X \sim t(\nu, \mu, \sigma^2)$ se sua densidade puder ser escrita como:

$$f_x(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi\nu}\sigma} \left(1 + \frac{1}{\nu} \frac{(x-\mu)^2}{\sigma^2}\right)^{-\frac{\nu+1}{2}} \quad (36)$$

Em que $\Gamma(\nu)$ é a função gama³, μ (parâmetro de localização), $\sigma^2 > 0$ (parâmetro de escala) e $\nu \in \mathbb{N} - \{0, 2\}$ o número de graus de liberdade da distribuição. Na prática, é comum trabalhar com $\mu = 0$ e $\sigma^2 = 1$.

A variância de X , definida para $\nu > 2$ é dada por $Var(X) = \frac{\nu}{\nu-2} \sigma^2$. A assimetria é zero, uma vez que a distribuição é simétrica e a curtose, definida para $\nu > 4$, é $K(X) = 3 + \frac{6}{\nu-4}$. Para $\nu > 30$ a distribuição t de Student, praticamente, sobrepõe-se à distribuição normal. Porém, baixos valores de ν significam excesso de curtose e caudas bem mais espessas que a normal.

Nicolau (2012) argumenta que, por vezes, tem-se o interesse de trabalhar com uma variável aleatória ε com distribuição t de Student de média zero, mas variância igual a um. Desta forma, se $X \sim t(\nu, \mu, \sigma^2)$ a solução passa pela transformação:

$$\varepsilon = X \sqrt{(\nu-2)/\nu}. \quad (37)$$

Assim, é possível verificar que $Var(\varepsilon) = 1$ e que o coeficiente de curtose continua $K(\varepsilon) = 3 + \frac{6}{\nu-4}$. A função densidade de ε , denotada por $g_\nu(x|\mu, \sigma^2)$ fica da forma:

$$g_x(x; \mu, \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)\sigma} \left[1 + \frac{1}{\nu-2} \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-\frac{\nu+1}{2}}. \quad (38)$$

³ $\Gamma(\nu) = \int_0^\infty e^{-x} x^{\nu-1} dx$.

A distribuição skew-t, utilizada neste trabalho, será aquela obtida pelo método de Fernández e Steel (1998). Neste caso, não é necessário utilizar a função de distribuição acumulada, o que facilita a sua implementação em modelos heterocedásticos de séries temporais, que, ainda, será visto. Desta forma, por meio de (36), a função densidade da distribuição skew-t será dada por:

$$f_x(x; \mu, \sigma^2, \nu, \gamma) = \frac{2}{\gamma + \gamma^{-1}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi\nu\sigma}} \begin{cases} \left(1 + \frac{1}{\nu-2} \frac{(x\gamma - \mu)^2}{\sigma^2}\right)^{-\frac{1}{2}(\nu+1)}, & \text{para } x < 0, \\ \left(1 + \frac{1}{\nu-2} \frac{\left(\frac{x}{\gamma} - \mu\right)^2}{\sigma^2}\right)^{-\frac{1}{2}(\nu+1)}, & \text{para } x \geq 0, \end{cases} \quad (39)$$

sendo $\Gamma(\nu)$ a função matemática gama, μ (parâmetro de locação), $\sigma^2 > 0$ (parâmetro de escala) e $\nu \in \mathbb{N} - \{0, 2\}$, o número de graus de liberdade e $\gamma > 0$ o parâmetro de assimetria da distribuição. Como notação, adotar-se-á $X \sim ST(\mu, \sigma^2, \nu, \gamma)$. A fim de diminuir o número excessivo de parâmetros da distribuição t de Student, neste trabalho será considerado a t de Student padrão, cujos parâmetros μ e σ^2 serão fixos e iguais a zero e a um, respectivamente, isto é, $\mu = 0$ e $\sigma^2 = 1$. Desta forma, pode-se reescrever (39) da forma:

$$f_x(x; \nu, \gamma) = \frac{2}{\gamma + \gamma^{-1}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi\nu}} \begin{cases} \left(1 + \frac{x^2 \gamma^2}{\nu-2}\right)^{-\frac{1}{2}(\nu+1)}, & \text{para } x < 0, \\ \left(1 + \frac{x^2}{\gamma^2(\nu-2)}\right)^{-\frac{1}{2}(\nu+1)}, & \text{para } x \geq 0, \end{cases} \quad (40)$$

sendo a notação agora adotada por $X \sim ST(\nu, \gamma)$.

Na Figura 4 tem-se o gráfico da função densidade da distribuição skew-t, o parâmetro de assimetria γ assumindo valores iguais a 0,8; 1 e 1,3 em (a), considerando $\nu = 5$. Em (b) tem-se γ fixo, igual a 1,3 e ν assume valores iguais a 5, 20 e 50.

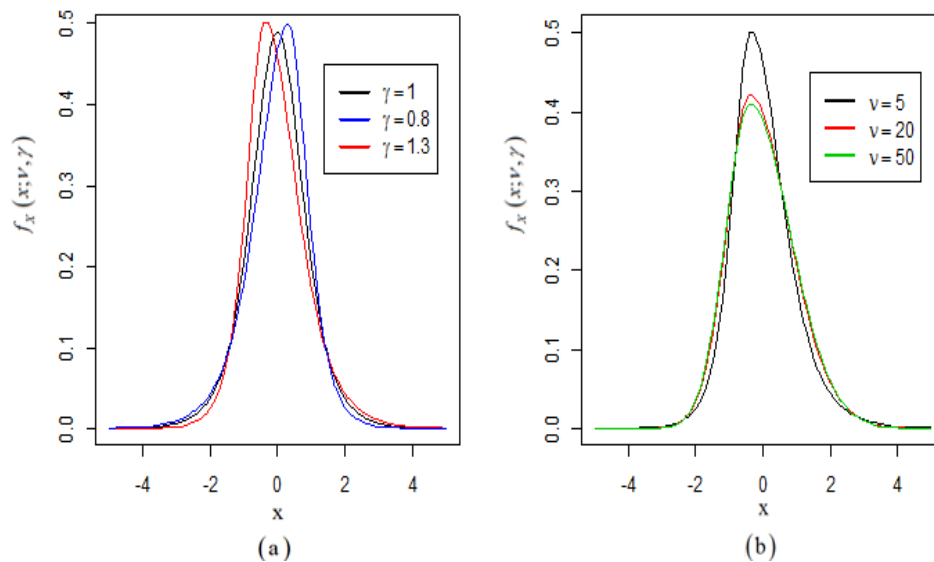


Figura 4 Gráfico da distribuição skew-t. Em (a) $\gamma = 0,8$; $\gamma = 1$ e $\gamma = 1,3$ em azul, preto e vermelho, respectivamente e ν fixo, igual a 5. Em (b) γ (fixo) é fixo e ν assume valores 5 (preto), 20 (vermelho) e 50 (verde)

Na Figura 5 tem-se a o comportamento da função de distribuição acumulada da skew-t, em função do seu parâmetro de assimetria, considerando ν fixo ($\nu=1$) e $\gamma=1$ em preto, que irá ser a função de distribuição acumulada da distribuição t de Student padrão, $\gamma=0,8$ em azul e $\gamma=1,3$ em vermelho, denotando a função de distribuição acumulada da skew-t com assimetria à esquerda e à direita, respectivamente.

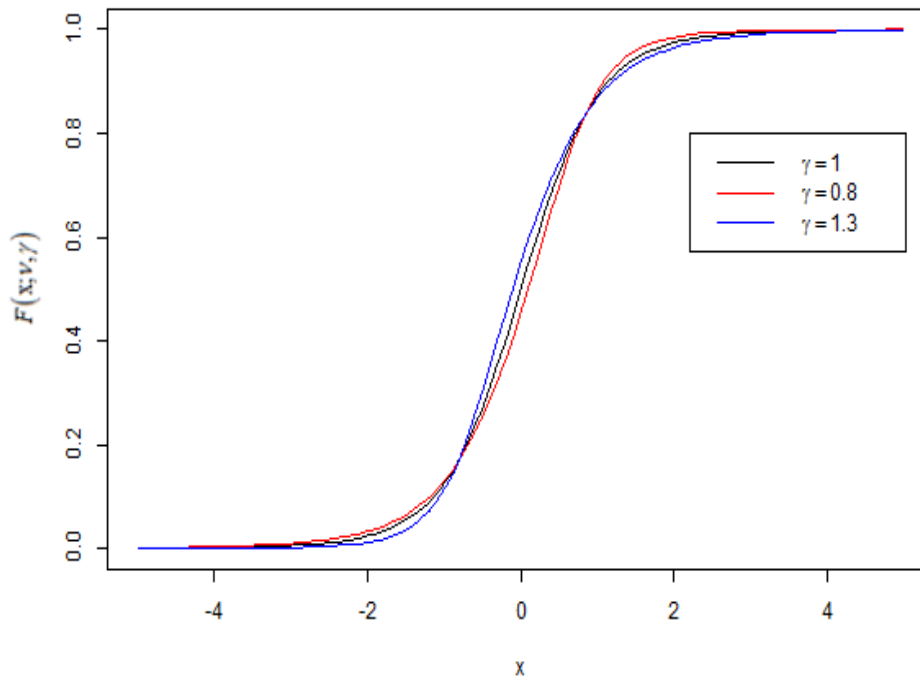


Figura 5 Gráfico da função distribuição acumulada da skew-t. $\gamma = 1$; $\gamma = 0,8$ e $\gamma = 1,3$ em preto, vermelho e azul, respectivamente. O parâmetro ν foi fixado em um

O primeiro momento de x da distribuição t de Student padrão ($\mu=0, \sigma^2=1$) truncado em valores reais positivos é:

$$\begin{aligned}
M_1 &= \frac{2\Gamma((\nu+1)/2)}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)} \int_0^\infty x \left(1 + \frac{x^2}{\nu-2}\right)^{-\frac{\nu+1}{2}} dx \\
&= \frac{2\Gamma((\nu+1)/2)}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)} \left(\frac{\nu-2}{\nu-1}\right) \\
&= \frac{\Gamma((\nu-1)/2)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\nu/2)}. \tag{41}
\end{aligned}$$

Usando as equações (32) e (40) e considerando que $\nu > 2$, tem-se que a média de $f_x(x|\nu, \gamma)$ é:

$$E(x|\nu, \gamma) = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(\gamma - \frac{1}{\gamma}\right) = m. \tag{42}$$

e variância:

$$\text{Var}(x|\nu, \gamma) = \left(\gamma^2 + \frac{1}{\gamma^2} - 1\right) - m^2 = s^2. \tag{43}$$

Para a versão padronizada da distribuição skew-t, considere a seguinte padronização $Z = (X - m)/s$, em que m é a média e s^2 a variância de (43).

Desta forma, tem-se que:

$$p(z|\gamma, \nu) = p(zs + m|\gamma, \nu) \frac{dx}{dz} = p(zs + m|\gamma) s \tag{44}$$

$$= \frac{2s}{\gamma + \gamma^{-1}} p(z^*),$$

sendo z^* dado por:

$$z^* = \begin{cases} (zs + m)\gamma, & \text{se } z < -m/s \\ (zs + m)/\gamma, & \text{se } z \geq -m/s \end{cases}. \quad (45)$$

Logo, a função densidade da distribuição skew-t na versão padronizada pode ser escrita como:

$$f_z(z|\gamma, \nu) = \frac{2s}{\sqrt{\pi}(\gamma + \gamma^{-1})} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu-2}} \left(1 + \frac{(z^*)^2}{\nu-2}\right)^{-\frac{\nu+1}{2}} \quad (46)$$

em que z^* é dado por (45).

2.3 Modelos Heterocedásticos Condicionais

Um dos fatos estilizados acerca de séries financeiras é que os retornos não têm variância constante ao longo do tempo. Sendo assim, o objetivo é modelar o que se chama de volatilidade que é a variância condicional de uma variável, comumente um retorno. Mesmo que não seja medida diretamente, a volatilidade manifesta-se de várias formas numa série temporal. Peña et al.(2001) argumentam que a volatilidade aparece em grupos (*volatility clusters*), além de que a mesma evolui continuamente no tempo, podendo ser estacionária, reagindo de modo diferente a valores positivos ou negativos (choques) na série.

O motivo pelo qual a volatilidade seja não constante pode estar relacionado com a especulação. Quando a proporção de especuladores é alta e os sinais do mercado (informações estatísticas acerca do volume de transações, por exemplo) são interpretados de forma análoga pela maior parte dos especuladores, formam-se tendências fortes de compra ou venda que refletem no preço e, conseqüentemente, na volatilidade de dado ativo.

Crises econômicas e políticas, bem como “bolhas especulativas”⁴ estouram, ocorrem momentos de turbulência na economia e, conseqüentemente, movimentos de alta volatilidade.

A chegada de informações no mercado, também, pode ser uma explicação para movimentos de alta e baixa na volatilidade de ativos financeiros. Quando chega informação ao mercado, os agentes reavaliam as suas carteiras (ativos financeiros: ações ou opções, por exemplo) e, assim, buscam vender ou comprar ativos a fim de atingir o valor “ótimo”. Quando a chegada de informação é reduzida e pouco relevante, os mercados tenderão a exibir baixa volatilidade; caso contrário, quando a informação é intensa e relevante, poderá ocorrer períodos de forte volatilidade.

Os modelos que descrevem a variância condicional dos retornos de ativos financeiros (volatilidade), ao contrário de modelos tradicionais de séries temporais, como o ARIMA que assumem que a variância condicional é constante no tempo, permite que variância varie no tempo, modelando-a como função dependente de valores dos retornos passados. Estes modelos heterocedásticos são denominados modelos de volatilidade. O primeiro a desenvolvê-los foi Engle (1982), que, por meio dos valores dos retornos passados, fez a modelagem da volatilidade da inflação inglesa. O modelo

⁴ Relacionam-se com a valorização excessiva de um dado ativo que ultrapassa o seu valor fundamental, como exemplo, do *boom* da eletrônica dos anos 60, da *internet* no final dos anos 90 e mais recentemente no mercado imobiliário norte americano em 2008.

desenvolvido por Engle recebeu o nome de ARCH (*Autoregressive Conditional Heteroskedasticity*).

Sendo assim, a volatilidade poderia ser modelada e dividida em volatilidade incondicional (seria mesmo constante) e a volatilidade condicional (poderia oscilar ao longo do tempo, ser analisada por meio de modelos de análise de heterocedasticidade condicional).

Bollerslev (1986) apresentou uma generalização do modelo ARCH proposto por Engle, denominado modelo GARCH (*Generalized ARCH*), cuja principal diferença ao ARCH foi incluir as volatilidades anteriores na fórmula da variância condicional, sendo um modelo mais parcimonioso, uma vez que os modelos ARCH podem exigir muitos parâmetros para capturarem a forte dependência temporal da volatilidade em muitos casos.

Vários trabalhos, ao longo dos anos, deram origem a diversas variantes dos modelos ARCH e GARCH de forma a atingir uma melhor explicação do comportamento dos retornos no mercado financeiro. Porém, ambos os modelos trabalham os retornos positivos e negativos de forma similar, o que fez surgir modelos para suprir estas deficiências, dentre eles o modelo EGARCH (*exponencial GARCH*).

Como a volatilidade é considerada em um dado instante de tempo, depende dos valores passados da série, a determinação de estimadores de Máxima Verossimilhança (MV) dos parâmetros de modelos da família ARCH, requer a maximização de uma função não linear. Portanto, as estimativas só podem ser obtidas por processos numéricos. Engle (1982) sugere o uso do método de Newton como um método iterativo para o cálculo das estimativas de MV. Este procedimento relaxa as restrições impostas aos parâmetros, que asseguram estacionaridade na covariância. Por outro lado, a determinação de estimadores de MV com tais restrições envolvem algumas dificuldades para a

estimação, uma vez que não é possível encontrar uma forma fechada pelo método a máxima verossimilhança.

2.3.1 Modelos ARCH

Antes do artigo seminal de Engle (1982), as dependências temporais nos momentos superiores a um eram tratadas como simples ruído, no qual as dependências temporais do segundo momento podem explicar, razoavelmente, a evolução da volatilidade ao longo do tempo.

Conforme já dito, os retornos não têm variância constante ao longo do tempo, formando, assim, grupos com diferentes graus de volatilidade e média constante. Para modelar estes fatos, Engle (1982) propôs o modelo autorregressivo com heterocedasticidade condicional (ARCH). Sua ideia básica é que o retorno y_t (o log-retorno) é não correlacionado serialmente, porém sua volatilidade (variância condicional no tempo) depende de retornos passados por meio de uma função quadrática dos valores passados da série.

O modelo ARCH (r) é definido como:

$$y_t = \varepsilon_t \sqrt{h_t}, \quad (47)$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \dots + \alpha_p y_{t-r}^2, \quad (48)$$

em que ε_t (erro) é uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média zero e variância condicional igual a um (ruído branco), $\alpha_0 > 0$, $\alpha_i \geq 0$ com $i > 0$.

Na prática, considera-se que ε_t segue uma distribuição normal com média zero e variância um ($\varepsilon_t \sim N(0,1)$). Pode-se, também, assumir que ε_t

possui uma distribuição t de Student com ν graus de liberdade ($\varepsilon_t \sim t_\nu$) ou qualquer outra distribuição que descreva melhor as caudas pesadas das séries financeiras. Os coeficientes α_i , com $i=1, \dots, r$ devem satisfazer certas condições, dependendo do tipo de imposição que se colocar sobre o processo ε_t .

A fim de estudar algumas propriedades do processo ARCH(r), será considerado o caso mais simples do modelo, isto é, com $r = 1$. Para as inovações (erros) considerar-se-á $\varepsilon_t \sim N(0,1)$ e F_t sendo toda a informação observada até o instante t . Sendo assim, o modelo será definido como:

$$y_t = \sqrt{h_t} \varepsilon_t, \quad (49)$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2, \quad (50)$$

com $\alpha_0 > 0$, $\alpha_1 \geq 0$ e $t=1, \dots, T$.

A média e variância dos retornos y_t condicionada a F_{t-1} são dadas por:

$$E(y_t | F_{t-1}) = E(\sqrt{h_t} \varepsilon_t | F_{t-1}) = \sqrt{h_t} E(\varepsilon_t | F_{t-1}) = 0,$$

$$Var(y_t | F_{t-1}) = E(y_t^2 | F_{t-1}) = E(h_t \varepsilon_t^2 | F_{t-1}) = h_t E(\varepsilon_t^2 | F_{t-1}) = h_t.$$

A média e a variância não condicionais (ou incondicionais) são dadas por:

$$\begin{aligned}
E(y_t) &= E\left[E(y_t | F_{t-1})\right] = 0, \\
\text{Var}(y_t) &= E(y_t^2) = E\left[E(y_t^2 | F_{t-1})\right] \\
&= E(\alpha_0 + \alpha_1 y_{t-1}^2) = \alpha_0 + \alpha_1 E(y_{t-1}^2). \tag{51}
\end{aligned}$$

Se o processo for estacionário de segunda ordem, então, para todo t , tem-se que $E(y_t^2) = E(y_{t-1}^2) = \text{Var}(y_t)$. Logo, de (51) tem-se que $\text{Var}(y_t) = \alpha_0 + \alpha_1 \text{Var}(y_t)$, do que decorre que:

$$\text{Var}(y_t) = \frac{\alpha_0}{1 - \alpha_1}. \tag{52}$$

A fim de satisfazer à definição da variância e às suposições acerca dos parâmetros do modelo ARCH(r), isto é, $0 < \text{Var}(y_t) < \infty$, então, deve-se tomar $0 \leq \alpha_1 < 1$.

A covariância entre y_t e y_{t+k} é dada por:

$$\begin{aligned}
\text{Cov}(y_t, y_{t+k}) &= E(y_t y_{t+k}) = E\left[E(y_t y_{t+k} | F_{t+k-1})\right] \\
&= E\left[y_t E(y_{t+k} | F_{t+k-1})\right] = E\left[y_t E(\sqrt{h_{t+k}} \varepsilon_{t+k} | F_{t+k-1})\right] \\
&= E\left[y_t \sqrt{h_{t+k}} E(\varepsilon_{t+k} | F_{t+k-1})\right] = 0,
\end{aligned}$$

Para todo $k > 0$. Portanto, a covariância entre os retornos é nula, pois y_t está em F_{t+k-1} e $E(\varepsilon_{t+k}|F_{t+k-1})=0$. Logo, a função de autocovariância de y_t será nula, isto é, $\gamma_{y_t}(k)=0$ para $k \geq 1$.

Nos fatos estilizados viu-se que a distribuição dos retornos apresenta caudas longas, de forma que a curtose é maior que a da distribuição normal, ou seja, três. Para calcular a curtose do processo, supondo que siga o modelo (49) e (50), é preciso encontrar o momento de quarta ordem. Sendo ε_t uma sequência de variáveis aleatórias independentes e identicamente distribuídas com média nula e variância unitária, tem-se que:

$$E(y_t^4|F_{t-1}) = h_t^2 E(\varepsilon_t^4) = E(\varepsilon_t^4) (\alpha_0 + \alpha_1 y_{t-1}^2)^2. \quad (53)$$

Como a curtose da normal é três e tomando-a como referência ($E(\varepsilon_t^4) = 3$), então, a expressão (54) pode ser reescrita como:

$$E(y_t^4|F_{t-1}) = 3(\alpha_0 + \alpha_1 y_{t-1}^2)^2. \quad (54)$$

Desta forma, para encontrar o momento de quarta ordem de y_t pode-se utilizar (54) de forma que:

$$E(y_t^4) = 3E(\alpha_0 + \alpha_1 y_{t-1}^2)^2 = 3E(\alpha_0^2 + 2\alpha_0\alpha_1 y_{t-1}^2 + \alpha_1^2 y_{t-1}^4).$$

Considerando o processo estacionário de quarta ordem, o momento de ordem quatro pode ser escrito como $\mu_4 = E(y_t^4)$. Assim:

$$\begin{aligned}
 \mu_4 &= 3 \left[\alpha_0^2 + 2\alpha_0\alpha_1 E(y_{t-1}^2) + \alpha_1^2 E(y_{t-1}^4) \right] \\
 &= 3 \left[\alpha_0^2 + 2\alpha_0\alpha_1 \text{Var}(y_t) + \alpha_1^2 \mu_4 \right] \\
 &= 3 \left[\alpha_0^2 + 2\alpha_1 \left(\frac{\alpha_0^2}{1-\alpha_1} \right) + \alpha_1^2 \mu_4 \right] = 3\alpha_0^2 \left(1 + \frac{2\alpha_1}{1-\alpha_1} \right) + 3\alpha_1^2 \mu_4 \\
 \mu_4 &= \frac{3\alpha_0^2 (1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}. \tag{55}
 \end{aligned}$$

Por (55) pode-se calcular a curtose de y_t :

$$K = \frac{\mu_4}{[\text{Var}(y_t)]^2} = \frac{3\alpha_0^2 (1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} \frac{(1-\alpha_1)^2}{\alpha_0^2} = 3 \left(\frac{1-\alpha_1^2}{1-3\alpha_1^2} \right). \tag{56}$$

Vê-se por (56) que, ao admitir que y_t siga um modelo ARCH(r), as caudas serão mais pesadas que a da normal, que é uma propriedade bastante viável para o modelo, uma vez que as caudas da distribuição de retornos são pesadas e apresentam curtose maior que três.

2.3.2 Modelos GARCH

O modelo ARCH(r) pode-se apresentar inviável pelo grande número de parâmetros estimados, uma vez que a ordem deste modelo pode ser alta em

virtude da longa memória encontrada, além da existência do problema com a variância negativa. De forma a resolver este problema, Bollerslev (1986) sugere um modelo, potencialmente, mais parcimonioso, chamado modelo GARCH (“*Generalized ARCH*”) descrevendo a volatilidade com menos parâmetros se comparado ao modelo ARCH. A ideia foi introduzir na fórmula da variância condicional regressores da variância condicional passada, possibilitando ao modelo ser mais flexível, no sentido de descrever uma memória mais longa e com menos parâmetros comparado aos modelos ARCH.

Considere $\mathbf{y} = \{y_t, t=1, \dots, T\}$ como a série de retornos. O modelo GARCH (p, q) estima a volatilidade dos retornos da forma:

$$y_t = \varepsilon_t \sqrt{h_t}, \quad \varepsilon_t \sim D(0,1) \quad (57)$$

$$h_t = \omega + \sum_{i=1}^q \alpha_i y_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \quad (58)$$

sendo h_t a variância condicional (não observável) de y_t dada a informação passada $F_t = \{y_{t-1}, y_{t-2}, \dots\}$, os ε_t são variáveis aleatórias independentes e identicamente distribuídas com média zero, isto é, são os erros do modelo. $D(0,1)$ denota uma distribuição com média zero e variância um. As restrições de estacionaridade e positividade são $\alpha_0 > 0$, $\omega > 0$, $\alpha_i \geq 0$, $i=1, \dots, p$, $\beta_j \geq 0$, $j=1, \dots, q$ e $(\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j) < 1$.

O modelo GARCH (1, 1) é muito utilizado na prática. Ele será utilizado para mostrarem-se algumas propriedades dos modelos GARCH. O modelo GARCH (1,1) com $\varepsilon_t \sim N(0,1)$ pode ser expresso da forma:

$$\begin{aligned}
 y_t &= \sqrt{h_t} \varepsilon_t \\
 h_t &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1},
 \end{aligned}
 \tag{59}$$

com $\omega > 0, 0 \leq \alpha_1, \beta_1 < 1$.

A esperança e a variância de y_t dado F_t é:

$$E(y_t | F_t) = \sqrt{h_t} E(\varepsilon_t | F_t) = 0 \tag{60}$$

$$Var(y_t | F_t) = h_t E(\varepsilon_t^2 | F_t) = h_t Var(\varepsilon_t) = h_t. \tag{61}$$

A média e a variância dos retornos podem ser obtidas da forma:

$$E(y_t) = E[E(y_t | F_{t-1})] = 0, \tag{62}$$

$$Var(y_t) = E(y_t^2) = E[E(y_t^2 | F_{t-1})] = \frac{\omega}{1 - (\alpha_1 + \beta_1)}. \tag{63}$$

Para atender à definição de variância e às suposições feitas acerca do modelo, isto é, que variância não condicional dos retornos seja finita e positiva, deve-se tomar $0 \leq \alpha_1 + \beta_1 < 1$. No caso geral do modelo GARCH (p, q), a variância será dada por:

$$Var(y_t) = \frac{\omega}{1 - \left(\sum_{i=1}^p \alpha_i + \sum_{i=1}^q \beta_i \right)}. \tag{64}$$

De acordo com Morettin (2011), o coeficiente de curtose do modelo (59) é dado por:

$$K(y_t) = \frac{E(y_t^4)}{[E(y_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2}, \quad (65)$$

Dado que o denominador seja positivo, então, se y_t segue um modelo GARCH, as caudas de y_t serão mais longas do que as da normal.

A fim de ilustrar o processo GARCH, tem-se a simulação apresentada na Figura 6 de dois modelos GARCH. Em (a) modelo GARCH (1,0) que nada mais é que o modelo ARCH (1), uma vez que o parâmetro β , é nulo e considerando $\omega=0,2$ e $\alpha_1=0,7$. Em (b) o modelo GARCH (1,1) tem parâmetros $\omega=0,2$, $\alpha_1=0,6$ e $\beta_1=0,3$.

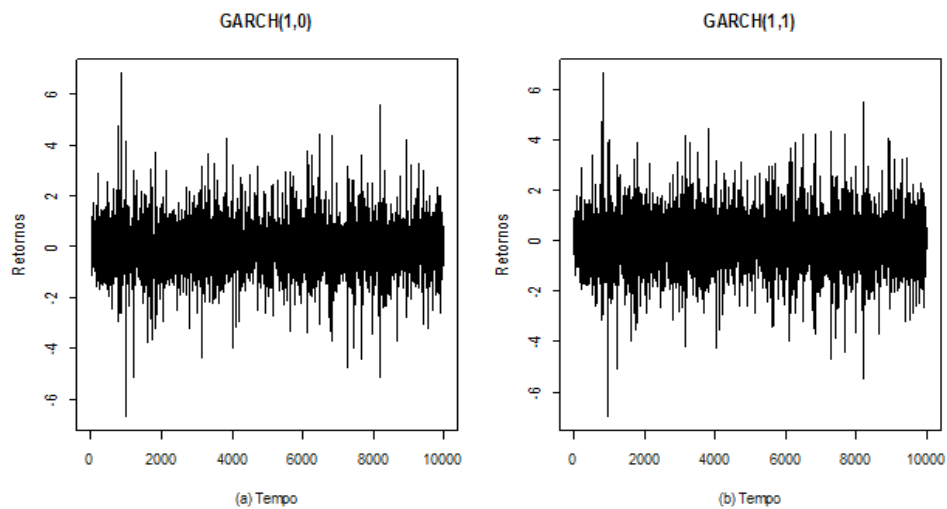


Figura 6 Simulação 10000 observações do modelo GARCH com inovações normais. Em (a) tem-se a o modelo GARCH (1,0) e em (b) um modelo GARCH (1,1)

2.3.3 Métodos de Estimação

Os modelos da classe ARCH, GARCH, bem como suas extensões, podem ser estimados, utilizando o método de máxima verossimilhança, conforme descrito em Mood et al. (1974) e Bolfarine e Sandoval (2010). Este método consiste em encontrar o valor $\hat{\theta}$ que torne máxima a função de verossimilhança.

A função de verossimilhança $L(\theta; y)$ pode ser escrita da forma $L(\theta; y) = \prod_{i=1}^n f(y_i, \theta)$, em que $f(y_i, \theta)$ é a distribuição de probabilidade de Y_i e Y_1, \dots, Y_n é uma amostra aleatória desta distribuição.

Utilizando o logaritmo na função de verossimilhança, a fim de facilitar o desenvolvimento analítico de forma a encontrar os estimadores dos parâmetros, uma vez o máximo da função de verossimilhança e máximo do seu logaritmo acontece em um mesmo ponto, então, pode-se escrever $l(\theta; y) = \log(L(\theta; y))$.

Se $\hat{\theta}$ existe, então, deve satisfazer o sistema de equações de verossimilhança, da forma:

$$\frac{\partial}{\partial \theta_i} l(\hat{\theta}; y) = 0, \text{ com } i = 1, \dots, k, \quad (66)$$

com k denotando o número de parâmetros da distribuição.

Os modelos heterocedásticos condicionais não possuem uma forma fechada para sua função densidade de probabilidade, o que requer o uso de procedimentos numéricos para a obtenção da solução do sistema de k equações de verossimilhança.

Assim, para a obtenção dos estimadores de máxima verossimilhança por um procedimento numérico, utiliza-se a função escore que é dada por:

$$U(Y; \hat{\theta}) = \frac{\partial}{\partial \theta_i} l(\hat{\theta}; y). \quad (67)$$

Se $\hat{\theta}$ é o vetor de EMV de θ , então $U(Y; \hat{\theta}) = 0$. Desta forma, por meio da expansão em série de Taylor $U(Y; \hat{\theta}) = 0$ em torno θ_0 (vetor de valores iniciais no processo iterativo), tem-se:

$$0 = U(Y; \hat{\theta}) = U(Y; \theta_0) + (\hat{\theta} - \theta_0)U'(Y; \theta_0) + \dots, \quad (68)$$

em que são desprezados os termos de ordem mais alta, obtém-se:

$$0 = U(Y; \hat{\theta}) \approx U(Y; \theta_0) + (\hat{\theta} - \theta_0)U'(Y; \theta_0), \quad (69)$$

Isto é,

$$\hat{\theta} \approx \theta_0 - U(Y; \theta_0)[U']^{-1}(Y; \theta_0) = \theta_0 + U(Y; \theta_0)I_F^{-1}(\theta_0), \quad (70)$$

sendo $I_F(\theta_0)$ a matriz de informação de Fisher.

Desta maneira, tomando um vetor de valores iniciais θ_0 e utilizando o processo iterativo:

$$\hat{\theta}^{(j+1)} = \theta^{(j)} + U(Y; \theta^{(j)}) I_F^{-1}(\theta^{(j)}), \quad (71)$$

repetidamente até que o processo se estabilize segundo algum critério de convergência, ou seja, para um dado ξ pequeno, $|\theta^{(j+1)} - \theta^{(j)}| < \xi$.

2.3.4 Estimação modelo GARCH

Para a estimação dos parâmetros do modelo GARCH (p, q), dado pelas equações (57) e (58) utilizar-se-á o método da máxima verossimilhança. A função de distribuição conjunta do modelo é dada por:

$$\begin{aligned} f(y_1, \dots, y_T) &= f(y_1) f(y_2 | y_1) \dots f(y_T | y_1, \dots, y_{T-1}) \\ &= f(y_1) \prod_{t=2}^T f(y_t | y_1, \dots, y_{t-1}). \end{aligned} \quad (72)$$

A função de log - verossimilhança do modelo é:

$$L(\Theta; y_1, \dots, y_T) = \log f(y_1, \dots, y_T) = \log f(y_1) + \sum_{t=2}^T \log f(y_t | y_1, \dots, y_{t-1})$$

sendo Θ o vetor de parâmetros do modelo GARCH (p, q), isto é,

$\Theta = [\alpha_0; \alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q]$, e a função log - verossimilhança dado y_1 será:

$$l(\Theta; \mathbf{y}) = \sum_{t=2}^T \log f(y_t | y_1, \dots, y_{t-1}). \quad (73)$$

De modo que, para uma amostra relativamente grande, o termo $f(y_t)$ poderá ser desconsiderado.

Consideramos a equação (57) do modelo GARCH (p, q). Desta tem-se que:

$$g(y_t) = \frac{y_t}{\sqrt{h_t}} \quad \text{e} \quad g'(y_t) = \frac{1}{\sqrt{h_t}}. \quad (74)$$

Sendo assim, pode-se reescrever a função de distribuição condicional da forma:

$$f(y_t | y_1, \dots, y_{T-1}) = f(g(y_t)) g'(y_t). \quad (75)$$

Usando a equação (73) na equação (74), tem-se que:

$$f(y_t | y_1, \dots, y_{T-1}) = f\left(\frac{y_t}{\sqrt{h_t}}\right) \frac{1}{\sqrt{h_t}}. \quad (76)$$

Finalmente, por (75) na equação (73), chega-se à função de log – verossimilhança que permite estimar parâmetros do modelo GARCH (1,1) por meio das inovações do modelo assumindo alguma distribuição arbitrária. Logo,

$$l(\Theta; \mathbf{y}) = \sum_{t=2}^T \log f\left(\frac{y_t}{\sqrt{h_t}}\right) + \sum_{t=2}^T \log h_t^{-1/2}. \quad (77)$$

Uma questão importante é que mesmo se $\varepsilon_t \sim N(0,1)$, a distribuição incondicional de y_t não será normal, em particular tendo caudas mais pesadas que àquela. Entretanto, há evidências na literatura de que muitas séries temporais financeiras tendem a possuir curtose observada ainda maior do que a implicada por um modelo GARCH com inovações normais. Sendo assim, alguns autores têm proposto distribuições com caudas mais pesadas do que a gaussiana para os erros ε_t . Como exemplo, pode-se citar a distribuição t de Student com V graus de liberdade em Baillie e Bollerslev (1989), Lambert e Laurent (2002) com inovações skew- t e Ardia (2008) com t de Student, por meio de uma abordagem bayesiana.

Identificar a ordem de um modelo GARCH a ser ajustado a uma série pode ser difícil. Morettin (2011) recomenda a utilização de modelos de ordem baixa (como por exemplo, GARCH (1,1), GARCH (2,1), GARCH (1,2) e GARCH (2,2)) e escolhe-se o modelo com base em critérios, como o AIC (AKAIKE, 1974) e o BIC (SCHWARZ, 1978), dentre outros. O autor afirma, ainda, que na maioria das séries financeiras um modelo GARCH (1,1) é o mais parcimonioso para descrever a volatilidade.

Neste trabalho será focado no modelo GARCH (1,1), já que modelos GARCH de outras ordens podem ser estimados de forma semelhante. Sendo assim, voltando na estimação dos modelos GARCH, considerar-se-á o GARCH (1,1), como exemplo inicial para a estimação, com inovações normais.

A função densidade de ε_t , considerando $\varepsilon_t \sim N(0,1)$ é dada por:

$$f(\varepsilon_t | y_{t-1}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \varepsilon_t^2\right). \quad (78)$$

Como $f(y_t / \sqrt{h_t}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} \frac{y_t^2}{h_t}\right)$, pois por (59) $\varepsilon_t = \sqrt{h_t} / y_t$, então:

$$\log f(y_t / \sqrt{h_t}) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \frac{y_t^2}{h_t}. \quad (79)$$

Levando em conta o resultado de (78) e aplicando este em (76), finalmente chega-se à expressão da log – verossimilhança em (80):

$$\begin{aligned} l(\Theta; \mathbf{y}) &= \sum_{t=2}^T \left(-\frac{1}{2} \log(2\pi) - \frac{1}{2} \left(\frac{y_t^2}{h_t} \right) \right) - \frac{1}{2} \sum_{t=2}^T \log h_t \\ &= -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=2}^T \log(h_t) - \frac{1}{2} \sum_{t=2}^T \frac{y_t^2}{h_t}. \end{aligned} \quad (80)$$

Logo, para encontrar os estimadores de máxima verossimilhança, por meio de (80) para os parâmetros, tem-se que maximizar a função de verossimilhança condicional. Como feito no artigo de Engle (1982), Bollerslev (1986), a fim de facilitar os resultados assintóticos, utiliza a log – verossimilhança média com erros seguindo uma distribuição normal padrão para estimar (80), $h_t = \hat{\sigma}^2$, sendo $\hat{\sigma}^2 = \sum_{t=1}^T y_t^2 / T$. Neste artigo, o autor sugere o uso do algoritmo BHHH, desenvolvido por Berndt et al. (1974). Para maiores detalhes ver o Apêndice B.

Bollerslev (1987) sugere o uso de distribuição com caudas mais pesadas para capturar o excesso de curtose. Considerando uma variável aleatória X com distribuição t de Student com ν graus de liberdade e, ainda, que

$\varepsilon_t = X / \sqrt{\nu(\nu-2)}$, a função de verossimilhança do GARCH (1,1) com inovação t de Student com ν será da forma:

$$L(\Theta; \mathbf{y}) = \prod_{t=2}^T \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)h_t}} \left(1 + \frac{y_t^2}{(\nu-2)h_t}\right)^{-(\nu+2)/2}, \quad (81)$$

em que h_t é obtida de maneira recursiva.

2.3.5 Estimação dos parâmetros do modelo GARCH com inovações skew-t

O conhecimento de assimetria e curtose é importante em inúmeras aplicações no mercado financeiro. Sendo assim, uma distribuição que possa modelar estes dois momentos parece ser fundamental. Lambert e Laurent (2001) consideraram a densidade da skew-t proposta por Fernández e Steel (1998) ao modelo GARCH. Lembrando que uma das condições para a inovação de modelos GARCH é que a média seja nula e a variância unitária para o erro. Desta forma, utilizando (46), em que, neste caso, tem-se $\varepsilon_t \sim ST(0,1)$, a log-verossimilhança do modelo GARCH (1,1) com erros skew-t é dado por:

$$l(y_t | \Theta) = T \left\{ \log \Gamma\left(\frac{\nu+1}{2}\right) - \log\left(\frac{\nu}{2}\right) - \frac{1}{2} \log[\pi(\nu-2)] + \log\left(\frac{2}{\gamma + \gamma^{-1}}\right) + \log(s) \right\} - \frac{1}{2} \sum_{t=1}^T \left[\log(h_t) + (1+\nu) \log \left[1 + \frac{(sz_t + m)^2}{\nu-2} \gamma^{-2I_t} \right] \right], \quad (82)$$

em que $\Gamma(\nu)$ é a função gama, ν é o número de graus de liberdade, $\gamma > 0$ é o parâmetro de assimetria, $m = \frac{\Gamma((\nu+1)/2)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\nu/2)}(\gamma - \gamma^{-1})$, $s = \sqrt{(\gamma^2 + \gamma^{-2} - 1) - m^2}$ e $I_t = 1$ se $z_t \geq -m/s$ ou $I_t = -1$ se $z_t < -m/s$.

Na Figura 7 tem-se em (a) a simulação do modelo GARCH (1,1) com inovações skew-t com parâmetro de assimetria $\gamma = 0,6$ e em (b) o mesmo modelo, porém com $\gamma = 1,5$. Em ambas as séries simuladas foram considerados os parâmetros do GARCH (1,1), assumindo os valores $\omega = 0,2$; $\alpha_1 = 0,3$; $\beta_1 = 0,6$ e $\nu = 4$. Ambos os modelos simulados possuem como parâmetros iniciais $\omega = 0,2$, $\alpha_1 = 0,3$ e $\beta_1 = 0,6$.

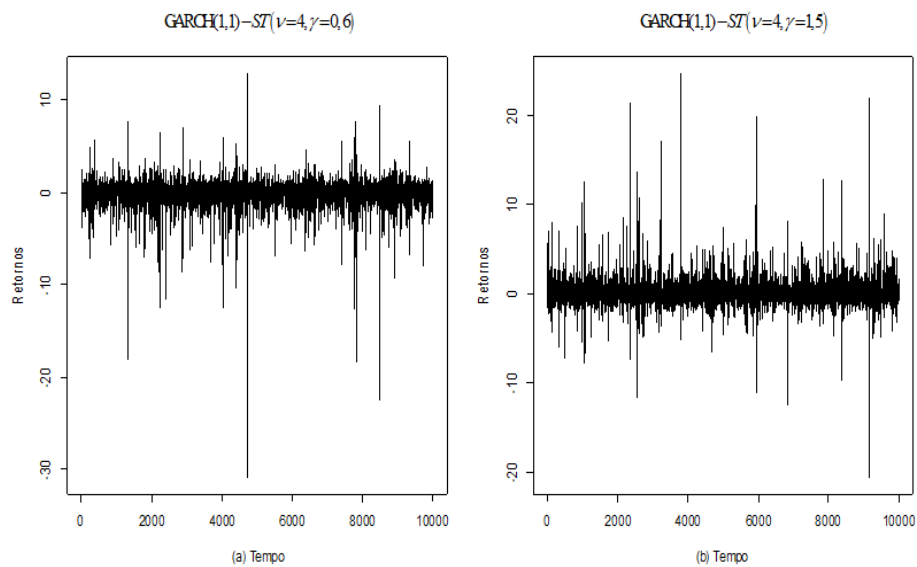


Figura 7 Simulação de 10000 observações do modelo GARCH (1,1) com inovações skew-t. Em (a) tem-se o modelo GARCH (1,1) com erros $ST(\nu = 4, \gamma = 0,6)$ e em (b) $ST(\nu = 4, \gamma = 1,5)$

2.4 Modelos Combinados (ARMA-GARCH)

Segundo Morettin (2011), pode-se considerar, inicialmente, o ajuste de um modelo ARMA a uma série de retornos para eliminar a correlação serial entre as observações e, em seguida, verificar se os resíduos do modelo apresentam heterocedasticidade condicional e ajustar um modelo ARCH (ou alguma de suas extensões) a estes resíduos. Esses modelos são conhecidos de uma forma geral como modelos ARMA-GARCH e são definidos a seguir.

Um modelo ARMA (m, n)-GARCH (p, q) de uma série temporal univariada y_t , de acordo com Aiube (2013) é da forma:

$$y_t = E(y_t | F_{t-1}) + \varepsilon_t \quad (83)$$

em que $E(\cdot | \cdot)$ é a esperança condicional, $y_t | F_{t-1}$ é o conjunto de todas as informações até o tempo $t-1$ e ε_t são os resíduos ou inovações da série temporal. Em um processo ARMA-GARCH, a esperança condicional é modelada por meio de um processo ARMA, e as inovações são modeladas de acordo com um processo GARCH.

A equação do processo ARMA (m, n)-GARCH (p, q) é dada por:

$$\begin{aligned} y_t &= \mu + \sum_{i=1}^m \phi_i y_{t-i} + \sum_{j=1}^n \theta_j \varepsilon_{t-j} + \varepsilon_t \\ \varepsilon_t &= z_t \sqrt{h_t}, \text{ em que } z_t \stackrel{i.i.d}{\sim} D(0,1), \\ h_t &= \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}, \end{aligned} \quad (84)$$

Sendo μ a média do processo, ϕ_1, \dots, ϕ_m são os coeficientes autorregressivos, $\theta_1, \dots, \theta_n$ os coeficientes de média móvel, $\alpha_1, \dots, \alpha_p \geq 0$, $\beta_1, \dots, \beta_q \geq 0$, $\omega > 0$ e $D(0,1)$ a distribuição de probabilidade das inovações com média nula e variância unitária.

Segundo Aiube (2013), a identificação do modelo (85) passa pela análise da função de autocorrelação (FAC) dos resíduos e dos resíduos ao quadrado do modelo ARMA puro. Posteriormente, o ajuste de um modelo GARCH para a variância deverá mostrar os resíduos ao quadrado descorrelacionados. A variável ε_t , assume, agora, o papel de retornos (y_t) .

Na Figura 8, em (a) há a simulação de um modelo AR (1) – GARCH (1,1) com erros skew-t cujo parâmetro de assimetria é 0,6 e quatro graus de liberdade. Em (b) o modelo MA (1) – GARCH (1,1) com inovação $ST(\nu=4, \gamma=0,6)$. Ambos os modelos têm erro $ST(\nu=4, \gamma=0,6)$ e parâmetros $\omega=0,1$; $\alpha_1=0,3$ e $\beta_1=0,6$. Já na Figura 9 tem-se a simulação de um modelo ARMA (1,1) – GARCH (1,1) com erros skew-t cujo parâmetro de assimetria é 0,6.

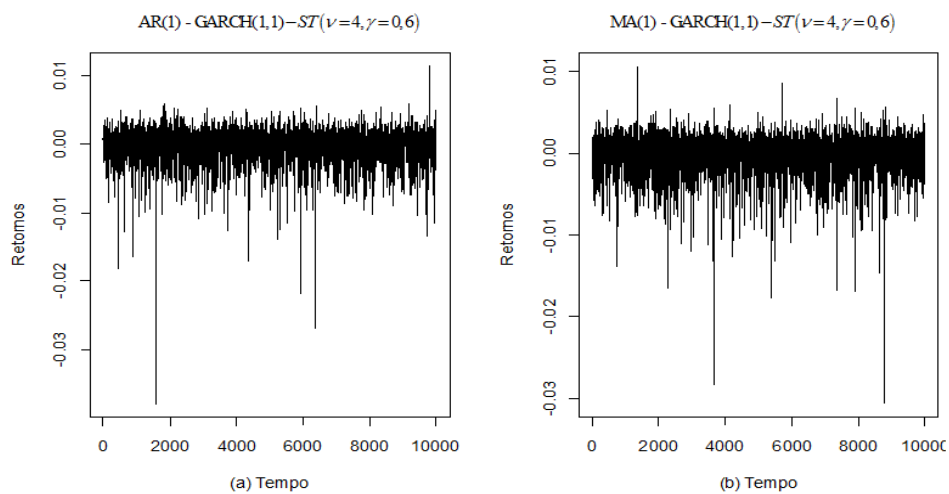


Figura 8 Simulação de 10000 observações do modelo AR (1) - GARCH (1,1), com parâmetro autoregressivo $\phi_1 = 0,5$ em (a) e (b) MA (1) - GARCH (1,1) e parâmetro média móvel $\theta_1 = 0,5$

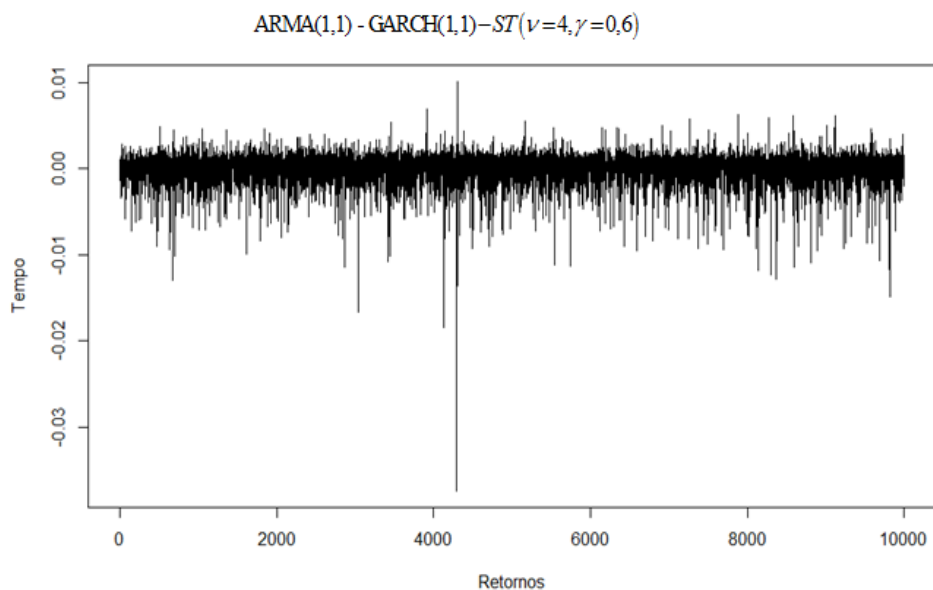


Figura 9 Simulação de 10000 observações do modelo ARMA - GARCH (1,1), com parâmetro autoregressivo $\phi_1 = 0,5$ e média móvel $\theta_1 = 0,5$ com erro $ST(v=4, \gamma=0,6)$ e parâmetros $\omega=0,1$; $\alpha_1=0,3$ e $\beta_1=0,6$

2.5 Testes para Detecção de Modelos GARCH

Nesta seção são mostrados alguns testes para identificação de modelos GARCH. Serão abordados apenas os testes mais comuns, porém há outras opções em Bollerslev, Engle e Nelson (1994).

2.5.1 Teste FAC e FACP

Os modelos GARCH (p, q) podem ser representados como processos ARMA. Para compreender esta representação, considere o GARCH (1,1): $h_t = \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1}$. Somando a ambos os termos da igualdade deste modelo a variável y_t^2 , tem-se:

$$y_t^2 + h_t = \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} + y_t^2 \quad (85)$$

em que, ao se isolar y_t^2 no lado esquerdo da equação (85) e simplificando-se a mesma até se obter o ARMA, da forma:

$$y_t^2 = \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} + \underbrace{y_t^2 - h_t}_{\nu_t} \quad (86)$$

no qual $\nu_t (y_t^2 - h_t)$ é uma diferença martingale⁵. Somando e subtraindo o termo $\beta_1 y_{t-1}^2$ após a igualdade em (86) tem-se:

⁵ $E(\nu_t) = E(y_t^2 - h_t) = E(h_t \varepsilon_t^2 - h_t) = E(h_t)(E(\varepsilon_t^2) - 1) = 0$,
 $E(\nu_t | F_{t-1}) = E(y_t^2 | F_{t-1}) - E(h_t | F_{t-1}) = h_t - h_t = 0$.

$$\begin{aligned}
y_t^2 &= \omega + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} + \beta_1 y_{t-1}^2 - \beta_1 y_{t-1}^2 + v_t \\
y_t^2 &= \omega + (\alpha_1 + \beta_1) y_{t-1}^2 - \beta_1 \underbrace{(y_{t-1}^2 - h_{t-1})}_{v_{t-1}} + v_t \\
y_t^2 &= \omega + (\alpha_1 + \beta_1) y_{t-1}^2 - \beta_1 v_{t-1} + v_t. \tag{87}
\end{aligned}$$

Como $E(v_t) = 0$ e $Cov(v_t, v_{t-k}) = 0$ conclui-se que $y_t^2 \sim \text{ARMA}(1,1)$. No caso geral, pode-se mostrar que:

$$y_t \sim \text{GARCH}(p,q) \Rightarrow y_t^2 \sim \text{ARMA}(\max[p,q], p). \tag{88}$$

De acordo com Nicolau (2012), em geral, é problemático identificar o GARCH, com base nas FAC e FACP de y_t^2 . Por dois motivos: primeiro o GARCH implica uma estrutura ARMA para y_t^2 e, como se observa, no ARMA, nenhuma das funções de autocorrelação (FAC ou FACP) é nula com base em certa ordem em diante (e é esta característica que facilita a identificação das ordens do AR ou do MA, porém não do ARMA); segundo, não existe uma correspondência perfeita entre as estruturas ARMA e GARCH (por exemplo, um ARMA(2,2) para y_t^2 pode ser um GARCH(2,1) ou um GARCH(2,2) para y_t).

Desta forma, como o modelo GARCH (p, q) assemelha-se ao modelo ARMA (max(p,q),q), as funções de autocorrelação, FAC e de autocorrelação parcial, FACP, podem sugerir se a série é heterocedástica, da mesma forma que dão indícios das ordens p e q de um modelo ARMA. Deve-se, então, proceder obtendo primeiro os quadrados dos erros estimados pela regressão do modelo estabelecido, como se a variância condicional fosse constante:

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2, \quad (89)$$

em que T é o número de resíduos.

Em seguida, calcula-se a FAC amostral para o quadrado dos resíduos, sem se esquecer de representá-los graficamente. A função é dada por:

$$\hat{\rho}_j = \frac{\sum_{t=j+1}^T (\hat{\varepsilon}_t^2 - \bar{\sigma}^2)(\hat{\varepsilon}_{t-j}^2 - \bar{\sigma}^2)}{\sum_{t=j+1}^T (\hat{\varepsilon}_t^2 - \bar{\sigma}^2)(\hat{\varepsilon}_{t-j}^2 - \bar{\sigma}^2)}, j = 1, 2, \dots \quad (90)$$

Pode aproximar o desvio padrão de $\hat{\rho}_j$ em grandes amostras por $T^{-0,5}$. Assim, pode-se dizer que um ARCH está presente se os valores de $\hat{\rho}_j$ forem estatisticamente diferentes de zero. A função de autocorrelação parcial é obtida representando-se em um gráfico o coeficiente estimado, $\hat{\phi}_{s,s}$ contra a defasagem s , considerando a equação:

$$\hat{\varepsilon}_t^2 = \phi_s + \phi_{s,1} \hat{\varepsilon}_{t-1}^2 + \dots + \phi_{s,s} \hat{\varepsilon}_{t-s}^2 + u_t, s = 1, 2, \dots, q. \quad (91)$$

Bueno (2011) argumenta que diferentemente dos modelos ARMA, a FAC fornece a ordem máxima da autorregressão do GARCH, representado pelo termo $\hat{\varepsilon}_{t-i}^2$ e a FACP oferece a ordem p das “médias móveis” do GARCH do termo $\hat{\sigma}_{t-j}^2$. Se os modelos forem assimétricos na variância, os testes ajudam a identificar a existência de heterocedasticidade condicional, sem, contudo, servir para definir as ordens p e q .

2.5.2 Teste de Box-Pierce-Ljung

Box e Pierce (1970) propuseram um teste para as autocorrelações dos resíduos estimados, que, apesar de não detectar quebras específicas no comportamento de ruído branco, pode indicar se estes valores são muito altos. Uma modificação deste teste foi proposta por Ljung e Box (1978). Assim, se o modelo for adequado, a estatística:

$$Q = n(n+2) \sum_{i=1}^K \frac{\hat{r}_i^2}{(n-i)}, \quad (92)$$

terá uma distribuição χ_r^2 com $r = K - p - q$ graus de liberdade. A hipótese de ruído branco para os resíduos é rejeitada para valores altos de Q . Geralmente, costumam-se utilizar as 15 ou 20 primeiras \hat{r}_i .

2.5.3 Teste Multiplicador de Lagrange

O teste multiplicador de Lagrange, LM é muito utilizado, veja Engle (1982), Bollerslev, Engle e Nelson (1994) e Johnston e Dinardo (1997) e depois estendido por Lee (1991) para GARCH, para outros detalhes. Define-se o teste como $H_0 : \alpha_i = 0$, para todo $i = 1, \dots, r$, na regressão:

$$y_t^2 = \omega + \alpha_1 y_{t-1}^2 + \dots + \alpha_r y_{t-r}^2 + u_t, \quad (93)$$

em que $t = r + 1, \dots, T$. A estatística do teste é $S = TR^2$, quem possui distribuição assintótica χ_r^2 sob H_0 . Aqui, R^2 é o coeficiente de determinação

(medida de ajustamento de um modelo estatístico linear generalizado, como por exemplo, a Regressão Linear).

Um teste assintoticamente equivalente, que pode ter propriedades melhores para amostras pequenas, consiste em utilizar a estatística:

$$F = \frac{(SQR_0 - SQR_1) / r}{SQR_1 / (T - 2r - 1)} \sim F(r, T - 2r - 1), \quad (94)$$

em que $SQR_0 = \sum_{t=r+1}^T (y_t^2 - \bar{y})^2$ e $SQR_1 = \sum_{t=r+1}^T \hat{u}_t^2$, com \bar{y} denotando a média amostral dos y_t^2 e \hat{u}_t os resíduos de mínimos quadrados da regressão (93). Caso o valor de F seja significativo, diz-se que há heterocedasticidade condicional na série.

2.6 Previsão e adequação do modelo

Para classes de modelos ARCH/GARCH, com inovação normal, t de Student ou skew- t , os resíduos:

$$\tilde{y}_t = \frac{y_t}{\sqrt{h_t}} \quad (95)$$

sendo \tilde{y}_t (padronizados) variáveis aleatórias independente e identicamente distribuídas normal padrão, t de Student ou skew- t . Assim, uma maneira de verificar se o modelo é adequado é calcular a estatística Q de Ljung-Box dada por (92), para a sequência \tilde{y}_t . Além disso, podem-se encontrar os coeficientes

de assimetria e curtose estimados e fazer um gráfico QxQ plot para avaliar a suposição de normalidade (t_v , ou ST).

2.7 Previsões de modelos GARCH

De acordo com Nicolau (2012), a previsão no contexto do modelo ARCH/GARCH envolve, habitualmente, a previsão de y_t e h_t . Todavia, em várias aplicações, como por exemplo, estimação de risco de mercado, valorização de opções, dentre outros, a previsão da volatilidade é mais importante do que a previsão de y_t .

Segundo Bueno (2011), a maioria das previsões segue os mesmos passos do modelo ARIMA. Prevê-se um passo à frente e, com o resultado obtido, faz-se o passo seguinte e, assim, sucessivamente.

Heynen e Kat (1994) fizeram um estudo de previsão usando índices de ações e câmbio. Seus resultados podem ser assim sintetizados: em geral, a volatilidade é mais previsível no longo do que no curto prazo. Além disso, o desenvolvimento do modelo de previsão depende, sensivelmente, do ativo com que se está trabalhando. Ainda, segundo os autores, índices de ações são mais bem previstos, utilizando-se volatilidade estocástica, enquanto para o câmbio, o modelo GARCH (1,1) apresentou melhores resultados.

As previsões de volatilidade, utilizando um modelo GARCH (1,1), podem ser calculadas considerando uma origem t , da forma:

$$\hat{h}_t(1) = \omega + \alpha_1 y_t^2 + \beta_1 h_t, \quad (96)$$

e para $l > 1$,

$$\begin{aligned}\hat{h}_t(l) &= \omega + \alpha_1 \hat{y}_t^2(l-1) + \beta_1 \hat{h}_t(l-1), \\ &= \omega + \alpha_1 \hat{h}_t(l-1) \hat{\varepsilon}_t^2(l-1) + \beta_1 \hat{h}_t(l-1),\end{aligned}\quad (97)$$

uma vez que $y_t = \sqrt{\hat{h}_t} \varepsilon_t$. Fazendo a substituição de $\varepsilon_t^2(l-1)$ por $E(\varepsilon_{t+l-1}^2 - 1) = 1$, tem-se, então:

$$\hat{h}_t(l) = \omega + (\alpha_1 + \beta_1) \hat{h}_t(l-1), \quad l > 1. \quad (98)$$

2.8 Critérios de seleção de modelos

Na seleção de modelos propostos, é preciso decidir entre em um e outro modelo baseado em algum critério. Uma vez verificada adequabilidade de dado modelo em estudo, deve-se seguir para o princípio da parcimônia, isto é, buscar por modelos que melhor expliquem um fenômeno estudado, porém com o menor número de parâmetros possível. Entretanto, isto nem sempre é fácil. Sendo assim, o uso de métodos baseados em uma função penalizadora é necessário. A ideia é escolher o modelo cujo número de parâmetros minimize a quantidade:

$$P(k) = \log(\hat{\sigma}^2) + \frac{kC(n)}{n}, \quad (99)$$

em que $\hat{\sigma}^2$ é uma estimativa da variância residual obtida por meio do ajuste do modelo às n observações da série temporal, k é o número de parâmetros a serem estimados e $C(n)$ é uma função do tamanho da série temporal.

Os critérios aqui apresentados são os mais utilizados na literatura, sendo eles o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC). Estes utilizam o estimador de máxima verossimilhança para a variância, mais uma função penalizadora que difere para cada critério.

O AIC ou Akaike Information Criterion (AKAIKE, 1974) é um índice de qualidade do ajustamento de modelos estatísticos. Quanto menor seu valor, melhor a adequabilidade do modelo em explicar a realidade. Em sua forma geral, tem-se:

$$\text{AIC} = \log(\hat{\sigma}^2) + \frac{2k}{n}. \quad (100)$$

Hurvich e Tsay (1989) propuseram uma correção para o AIC, dada por:

$$\text{AIC}_c = \text{AIC} + \frac{2(k+1)(k+2)}{n-k+2}. \quad (101)$$

O BIC (Bayesian Information Criterion), também conhecido como critério de Schwarz, assim como AIC, é índice da qualidade do ajuste de um modelo. Foi desenvolvido por Schwarz (1978), tendo este nome em função do argumento bayesiano adotado pelo autor, no qual é definido em termos da probabilidade a posterior. O BIC é dado por:

$$\text{BIC} = \log(\hat{\sigma}^2) + k \frac{\log(n)}{n}. \quad (102)$$

Em comparação como AIC, o BIC “penaliza” mais a adoção de um parâmetro a mais no modelo. Logo, a tendência é considerá-lo quanto à escolha

de valores menores para o número de parâmetros (PERRON; VOGELSANG, 1992).

2.9 Processos IGARCH e persistência na volatilidade

Segundo Morettin (2011), em muitas situações práticas, pode-se obter, por exemplo, no GARCH (1,1), $\alpha_1 + \beta_1$ próximo de um. Se a soma destes parâmetros for um, ter-se-á o modelo IGARCH (“Integrated GARCH”), sendo escrito da forma

$$\begin{aligned} y_t &= \sqrt{h_t} \varepsilon_t, \\ h_t &= \omega + \beta_1 h_{t-1} + (1 - \beta_1) y_{t-1}^2, \end{aligned} \quad (103)$$

pois $\alpha_1 = 1 - \beta_1$. No caso geral para IGARCH (p, q) tem-se:

$$\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i = 1. \quad (104)$$

Uma questão importante é verificar se os choques (que podem ser compreendidos como uma modificação súbita no curso normal da economia, por exemplo, desvalorização do real) têm efeitos transitórios ou permanentes na estrutura da variância condicional da série de retornos em estudo. Na literatura, variam não só as definições de persistência como também as ferramentas estatísticas usadas para a avaliação desta. Adotar-se-á, neste trabalho, a abordagem de Nicolau (2012).

De acordo com Nicolau (2012), como medida de persistência, pode-se adotar a *half-life* que é o valor de j tal que:

$$E(y_{t+j} - \mu | F_t) \leq \frac{1}{2}(y_t - \mu), \quad (\text{supondo } y_t > \mu). \quad (105)$$

sendo μ a média, F_t as informações no instante t e y_t a série de retorno.

Supondo que em t exista um desvio de y em face à sua média de longo prazo igual a $y_t - \mu$. Logo, uma questão importante é quantos períodos são necessários para que metade desse desvio seja eliminada (em média)? Ou, em quanto tempo o processo elimina metade do desvio $y_t - \mu$? A resposta seria j . Para exemplificar, considere o processo AR (1), $y_t = \phi y_{t-1} + \varepsilon_t$ (neste caso, $\mu = 0$). Assim, por (105) ⁶

$$E(y_{t+j} | F_t) = \phi^j y_t. \quad (106)$$

Desta forma, a *half-life* é o valor de j tal que $\phi^j y_t = \frac{1}{2} y_t$, isto é,

$$\phi^j y_t = \frac{1}{2} y_t \Rightarrow j = \frac{\log(1/2)}{\log \phi}.$$

Se há dados diários ($t = 1$ – um dia) e, por exemplo, $\phi = 0,9$, então, $j = 6,5$, que é o número de dias necessários para que, em média, metade do desvio de y_t perante sua média de longo prazo seja eliminada. Valores altos de j

⁶ $y_t = \phi y_{t-1} + \varepsilon_t = \phi(\phi y_{t-2} + \varepsilon_{t-2}) + \varepsilon_t = \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t$
 $y_t = \dots = y_{t-k} \phi^k + \varepsilon_{t-k+1} \phi^{k-1} + \dots + \varepsilon_{t-1} \phi + \varepsilon_t$, de forma que
 $E(y_t | F_{t-1}) = y_{t-k} \phi^k$, pois $E(\varepsilon_t) = 0$.

indicam maior persistência, isto é, o processo demora mais tempo para eliminar os desvios diante à média de longo prazo e, neste caso, pode-se dizer que o processo exibe reversão para a média de forma mais lenta.

No caso dos modelos GARCH, surgiu a ideia da persistência em variância (volatilidade). Choques na variância são, rapidamente, eliminados e h_t tende rapidamente para $E(h_t)$ ou, pelo contrário, os choques têm um efeito duradouro na variância? Sendo assim, pode-se definir uma *half-life* para a variância que é o valor de j tal que:

$$E(h_{t+j} - \sigma^2 | F_t) \leq \frac{1}{2}(h_t - \sigma^2), \quad (\text{supondo } h_t > \sigma^2). \quad (107)$$

sendo $\sigma^2 = E(h_t)$. Para exemplificar, considere o GARCH (1,1) cuja variância é $\sigma^2 = \omega / (1 - \alpha_1 - \beta_1)$ e

$$E(h_{t+j} - \sigma^2 | F_t) = (\alpha_1 + \beta_1)^j (h_t - \sigma^2) \quad (108)$$

em que $E(h_{t+j} | F_t)$ ⁷ é a previsão da volatilidade futura h_{t+j} , para $j > 0$.

Desta forma, a *half-life* é o valor j tal que:

$$(\alpha_1 + \beta_1)^j (h_t - \sigma^2) = \frac{1}{2}(h_t - \sigma^2) \Rightarrow j = \frac{\log(1/2)}{\log(\alpha_1 + \beta_1)}.$$

⁷ $E(h_{t+j} | F_t) = \omega + (\alpha_1 + \beta_1)E(h_{t+j-1})$

$E(h_{t+j} | F_t) = \omega \sum_{i=1}^{j-1} (\alpha_1 + \beta_1)^i + (\alpha_1 + \beta_1)^{j-1} (\alpha_1 y_t^2 + \beta_1 h_t)$ para $j > 2$.

$E(h_{t+j} - \sigma^2 | F_t) = (\alpha_1 + \beta_1)^{j-1} (E(h_{t+1}) - \sigma^2) = (\alpha_1 + \beta_1)^j (h_t - \sigma^2)$.

Se a soma $\alpha_1 + \beta_1$ tender para um, então j tenderá para mais infinito e o processo será persistente na variância (ou a volatilidade é persistente), o que corresponde ao modelo IGARCH.

De acordo com Nicolau (2012), em muitas aplicações empíricas observa-se que a estimativa para $\alpha_1 + \beta_1$ se encontra muito perto de um, sugerindo que a volatilidade pode ser persistente (forte dependência temporal).

2.10 Modelos de Memória Longa

Os modelos de memória longa (ML), em séries temporais, já eram empiricamente reconhecidos nas áreas de hidrologia e climatologia na década de 50. Estas séries apresentam persistência nas autocorrelações amostrais, isto é, existe uma dependência significativa entre as observações separadas por um longo período de tempo.

A presença de ML foi primeiramente observada por Hurst (1951) e Hurst (1957), Mandelbrot e Wallis (1968) e Mandelbrot (1972), dentre outros autores. Entretanto, sua aplicação em séries econômica – financeiras tenha acontecido a partir de 1980.

Segundo Morettin (2011), os economistas notaram que há evidências de que em processos de ML descrevem-se, de forma satisfatória, dados econômicos e financeiros, tais como taxas de juros e inflação.

2.10.1 Modelos ARFIMA

Enquanto um processo ARMA tem “memória curta”, ou seja, sua função de autocorrelação ρ_k decresce exponencialmente para zero, de forma que:

$$|\rho_k| \leq Cr^{-k}, k=1,2,\dots, \quad (109)$$

sendo $C > 0$ e $0 < r < 1$. A expressão (110) garante que a função de autocorrelação decai para zero de forma exponencial.

Um processo de memória longa é um processo estacionário cuja função de autocorrelação decresce hiperbolicamente para zero, isto é,

$$\rho_k \sim Ck^{-\alpha}, k \rightarrow \infty, \quad (110)$$

em que $C > 0$ e $\alpha \in (0,1)$.

Os modelos ARFIMA surgem como uma generalização dos modelos ARMA. Este modelo foi introduzido por Granger e Joyeux (1980) e Hosking (1981) o qual, também, é conhecido como modelo ARMA fracionário.

Um processo $\{X_t\}$ autorregressivo fracionário integrado de média móvel, ou então ARFIMA (p, d, q) com $d \in (-1/2, 1/2)$ se $\{X_t\}$ for estacionário e satisfizer a equação:

$$\phi(B)(1-B)^d X_t = \theta(B)a_t, \quad \text{para todo } t \in \mathbb{Z}, \quad (111)$$

em que $a_t \sim RB(0, \sigma_a^2)$, $\phi(B)$ e $\theta(B)$ são polinômios em B de graus p e q , respectivamente.

Segundo Morettin (2011), a razão da escolha dessa família de processos (ARFIMA), para fins de modelagem das séries de comportamento de memória longa, é que o efeito do parâmetro d em observações distantes decai hiperbolicamente, conforme a distância aumenta, enquanto os efeitos dos

parâmetros ϕ e θ decaem exponencialmente. Sendo assim, d deve ser escolhido com o objetivo de explicar a estrutura de correlação de ordens altas da série, enquanto ϕ e θ explicam a estrutura de correlação de ordens baixas.

O parâmetro d da expressão (111) é o grau de diferenciação fracionária e o operador $(1-B)^d$ é definido em termos de sua expansão, por meio de séries de Maclaurin, de acordo com Lopes e Mendes (2006) da forma:

$$(1-B)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} B^k, \quad (112)$$

sendo $\Gamma(\cdot)$ a função gama.

Hosking (1981) mostrou que o processo ARFIMA (p, d, q) dada por (111) será:

- a) estacionário se $d < 1/2$ e todas as raízes de $\phi(B)=0$ estiverem fora do círculo unitário;
- b) invertível se $d > 1/2$ e todas as raízes de $\theta(B)=0$ estiverem fora do círculo unitário.

Ainda com relação à d , Soldá (2008) argumenta que, para $0 < d < 1/2$, o processo (111) possui memória longa, agora para $-1/2 < d < 0$, ter-se-á memória curta e podendo ser identificado como antipersistente. Já para $|d| > 1/2$ o processo será considerado não estacionário. Na Figura 10 apresentam-se algumas simulações, segundo o processo ARFIMA, considerando d assumindo valor 0,2 e 0,4.

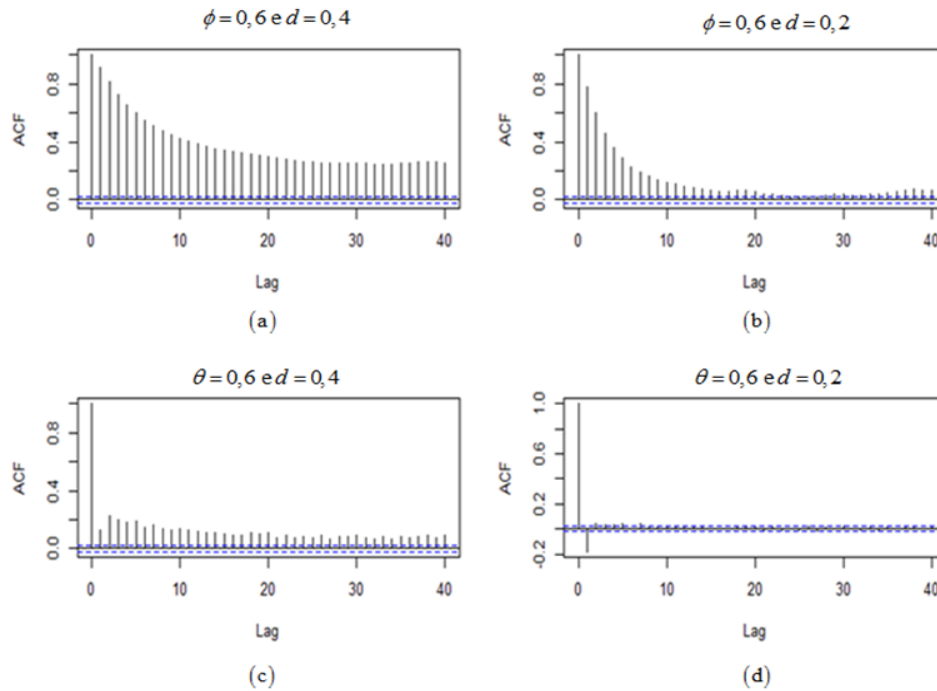


Figura 10 Séries temporais simuladas segundo o processo ARFIMA

Em (a) tem-se um ARFIMA (1,0.4,0), (b) ARFIMA (1,0.2,0), (c) ARFIMA (0,0.4,1) e (d) ARFIMA (0,0.2,1).

Para saber se determinada série temporal possui comportamento de memória longa, pode-se utilizar o teste R/S ou, então, estimar o parâmetro de longa dependência, este pode ser consultado em Geweke e Porter-Hudak (1983).

A estatística R/S foi introduzida por Hurst (1951), o qual tinha o interesse de testar a existência de memória longa numa série temporal. Dadas as observações X_1, \dots, X_T , a estatística R/S é dada por:

$$Q_T = \frac{1}{S_T} \left[\max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) \right], \quad (113)$$

em que \bar{X} é a média amostral e S_T^2 a variância amostral.

De acordo com Lo (1991), a estatística (113) não é robusta à dependência de curta memória, propondo, assim, substituir Q_T por:

$$\tilde{Q}_T = \frac{1}{\hat{\sigma}_T(q)} \left[\max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) \right], \quad (114)$$

sendo $\hat{\sigma}_T(q)$ a raiz quadrada do estimador da variância de longo prazo de Newey-West, com largura de faixa q , dado da forma:

$$\hat{\sigma}_T^2(q) = S_T^2 \left(1 + \frac{2}{T} \sum_{j=1}^q \varphi_{qj} y_j \right), \quad (115)$$

no qual $\varphi_{qj} = 1 - j/(q+1)$, $q < T$ e y_j são as autocorrelações amostrais de X_t . Newey e West (1987) propõem a escolha de $q = \left[4(T/100)^{2/9} \right]$.

2.10.2 Modelos GARCH com Memória Longa

Há vários estudos atualmente evidenciando a presença de ML nas autocorrelações dos quadrados dos retornos ou retornos absolutos de várias séries de preços em finanças e economia.

Os modelos GARCH podem ser estendidos para capturar o comportamento de memória longa e a alta persistência na volatilidade em séries de retornos.

2.10.3 Modelos FIGARCH e FIEGARCH

Na seção 2.5.1 viu-se que os modelos GARCH (p, q) podem ser escritos como um processo ARMA (m, p) em termos de y_t^2 , isto é,

$$\left[1 - \alpha(B) - \beta(B)\right]y_t^2 = \omega + \left[1 - \beta(B)\right]v_t, \quad (116)$$

em que $m = \max\{p, q\}$ e $v_t = y_t^2 - h_t$ (diferença martigale). Desta forma, um processo GARCH (p, q) integrado pode ser escrito como:

$$\left[1 - \alpha(B) - \beta(B)\right](1 - B)y_t^2 = \omega + \left[1 - \beta(B)\right]v_t. \quad (117)$$

O modelo FIGARCH pode ser obtido, por meio da substituição do primeiro operador de diferença $(1 - B)$ em (118), com o operador diferencial fracionário $(1 - B)^d$, em que $0 < d < 1$. Então, a classe de modelos FIGARCH pode ser obtida como:

$$\left[1 - \alpha(B) - \beta(B)\right](1 - B)^d y_t^2 = \omega + \left[1 - \beta(B)\right]v_t. \quad (118)$$

Para ilustrar, consideremos o modelo FIGARCH (1,d,1) que é representado da forma:

$$h_t = \omega + \beta h_{t-1} + \left(1 - \beta - (1 - \phi)(1 - B)^d\right)y_t^2. \quad (119)$$

A estimação dos parâmetros do modelo FIGARCH pode ser feita por meio do método da máxima verossimilhança. Para o caso em que $\varepsilon_t \sim N(0,1)$, a função de log-verossimilhança do FIGARCH (p, d, q) por ser escrita como:

$$\log L(\Theta, y_1, \dots, y_T) \approx -0,5 \log(2\pi) - 0,5 \sum_{t=1}^T \left[\log(h_t) + y_t^2 / h_t \right], \quad (120)$$

em que $\Theta' = (\omega, d, \beta_1, \dots, \beta_p, \phi_1, \dots, \phi_q)$.

O modelo FIEGARCH (p, d, q) foi introduzido por Bollerslev e Mikkelsen (1996), além da volatilidade variando com o tempo e os conglomerados de volatilidade (efeitos oriundos dos processos ARCH e GARCH), levam em conta a longa dependência da volatilidade e a assimetria (parte vinda do processo EGARCH). Este modelo pode ser escrito da forma:

$$\ln(h_t) = \omega + \frac{\alpha(B)}{\beta(B)(1-B)^d} g(Z_t - 1), \quad \text{para todo } t \in \mathbb{Z}, \quad (121)$$

em que $\omega \in \mathbb{R}$, $\{Z_t\}_{t \in \mathbb{Z}}$ são variáveis aleatórias independentes e identicamente distribuídas com média zero e variância um, $\{Z_t\}_{t \in \mathbb{Z}}$, $g(\cdot)$ é uma função que denota a curva de impacto de informação⁸. O operador $(1-B)^d$ é definido pela expressão (112). A estimação do modelo FIEGARCH pode ser, também, feita pelo método da máxima verossimilhança.

⁸ A função $g(\varepsilon_t) = \begin{cases} (\theta + \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|), & \text{se } \varepsilon_t \geq 0 \\ (\theta - \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|), & \text{se } \varepsilon_t < 0 \end{cases}$, em que γ representa o parâmetro de assimetria do modelo e θ é um parâmetro real.

3 MATERIAL E MÉTODOS

3.1 Dados

As séries utilizadas neste trabalho constituem-se dados diários dos índices Bovespa (IBOVESPA) e do SSE. Todos os dados podem ser obtidos no *site* <<http://finance.yahoo>>.

O índice Bovespa (IBOVESPA) é o principal indicador do mercado de capitais brasileiro, formado pelas ações com maior volume negociadas na BMF&FBOVESPA nos últimos meses.

O valor atual do índice Bovespa representa o valor de mercado, em moeda correta, de uma carteira teórica de ações, formada em 02 de janeiro de 1968, considerando uma aplicação hipotética, cujo objetivo é indicar o desempenho médio das cotações dos ativos de maior negociabilidade e representatividade do mercado de ações brasileiro.

O SSE Composite é o mais importante índice de ações da Bolsa de Valores de Xangai (Shanghai Stock Exchange), sendo um indicador de como vai a “saúde” do mercado acionário chinês. Este índice é projetado, para mostrar o desempenho geral do mercado de ações a qualquer momento, com um valor base de 100 a ser emitido em 19 dezembro de 1990.

Para a série Ibovespa, o período considerado vai de 27/04/1993 a 21/07/2014. Já para a série SSE o período é de 19/12/1990 a 27/06/2014.

3.2 Métodos

A metodologia utilizada neste trabalho fundamenta-se na análise de modelos univariados heterocedásticos condicionais por captarem de forma mais

consistente os fatos estilizados já mencionados de séries temporais financeiras, com e sem erros skew-t.

Para o estudo de análise das séries dos índices escolhidos, são utilizados os log – retornos diários, ou simplesmente retornos obtidos, conforme a expressão (19), isto é, $y_t = \ln(P_t / P_{t-1}) = \ln(P_t) - \ln(P_{t-1})$, em que P denota a série de fechamento das cotações dos índices e y_t a série de retorno.

O primeiro passo na análise dos dados constituir-se-á de uma estatística descritiva das séries de retornos e aplicação dos testes de normalidade e estacionaridade. A aplicação destes testes é de suma importância para atendimento dos pressupostos dos modelos de séries temporais que exigem a estacionaridade dos dados.

O teste de normalidade utilizado será o teste de Jarque e Bera (1987). Este teste utiliza a informação do coeficiente de assimetria e curtose amostrais, utilizando a estatística:

$$S = \left(\frac{T}{6}\right)\hat{A}^2 + \left(\frac{T}{24}\right)(\hat{K} - 3)^2, \quad (122)$$

que, sob H_0 : a série é normal, tem distribuição qui-quadrado com dois graus de liberdade, isto é, $S \sim \chi^2(2)$.

São considerados modelos combinados ARMA-GARCH com inovações t de Student e skew-t. A adequabilidade dos modelos ajustados é feita mediante o teste de Ljung-Box as séries de resíduos e também o teste LM para ver se há ou não heterocedasticidade nestas. São também considerados os gráficos FAC dos resíduos padronizados e resíduos ao quadrado, para ver se ainda há a presença de autocorrelação.

Como critério de comparação entre os modelos ajustados serão considerados os critérios de AIC e BIC, em que o maior valor em módulo indica o modelo mais adequado.

Quase todas as análises feitas neste trabalho foram realizadas utilizando-se o pacote estatístico R, que é um sistema para computação estatística e gráfica. Este é um software livre distribuído sob a Licença Pública Geral (<http://www.r-project.org>) e pode ser livremente e distribuído entre os usuários sem qualquer custo. As análises dos modelos FIGARCH e FIEGARCH foram feitas por meio dos softwares Matrixer⁹ e Oxmetrics¹⁰.

⁹ <<http://matrixer.narod.ru>>

¹⁰ <<http://www.oxmetrics.net>>.

4 RESULTADOS E DISCUSÃO

4.1 Comportamento das séries de retornos dos índices

Nesta seção são apresentadas as estatísticas descritivas das séries dos índices dos retornos do IBOVESPA e SSE mostradas nas Tabelas 1 e 2.

Tabela 1 Estatística descritiva IBOVESPA

Estatística	
Retorno médio (%)	0,1476
Curtose	10,0218
Assimetria	0,4945
Mínimo (%)	- 17,2082
Máximo (%)	28,8325

Tabela 2 Estatística descritiva SSE

Estatística	
Retorno médio (%)	0,0500
Curtose	157,9755
Assimetria	5,6332
Mínimo (%)	- 17,9000
Máximo (%)	71,9100

Por meio das Tabelas 1 e 2 percebe-se que ambas as séries de retornos possuem curtose maior que três, isto é, denotando que a série de retornos

possuem caudas mais pesadas do a distribuição normal. A Tabela 1 mostra que o coeficiente de assimetria do índice Bovespa possui assimetria à direita e a Tabela 2 indica assimetria à direita da série de retornos do índice SSE.

A Figura 11 mostra os gráficos da série IBOVESPA. Em (a) tem-se a série temporal de preços do IBOVESPA. Em (b) a série de log-retornos ou simplesmente série de retornos. Por meio deste, é possível verificar períodos de maior ou menor volatilidade no presente índice. Em (c) tem-se o histograma da série de retornos do IBOVESPA, no qual é possível verificar que esta série possui uma alta leptocurtose, isto é, uma curtose maior que três, além de assimetria à direita, que está de acordo com os fatos estilizados acerca da série de retornos do mercado financeiro. Em (d) tem-se o QxQ plot da série de retornos do IBOVESPA, no qual é possível avaliar a falta de normalidade dos dados, que também foi verificada com o teste de normalidade Jarque e Bera, cujo p-valor foi inferior $2,2 \cdot 10^{-16}$.

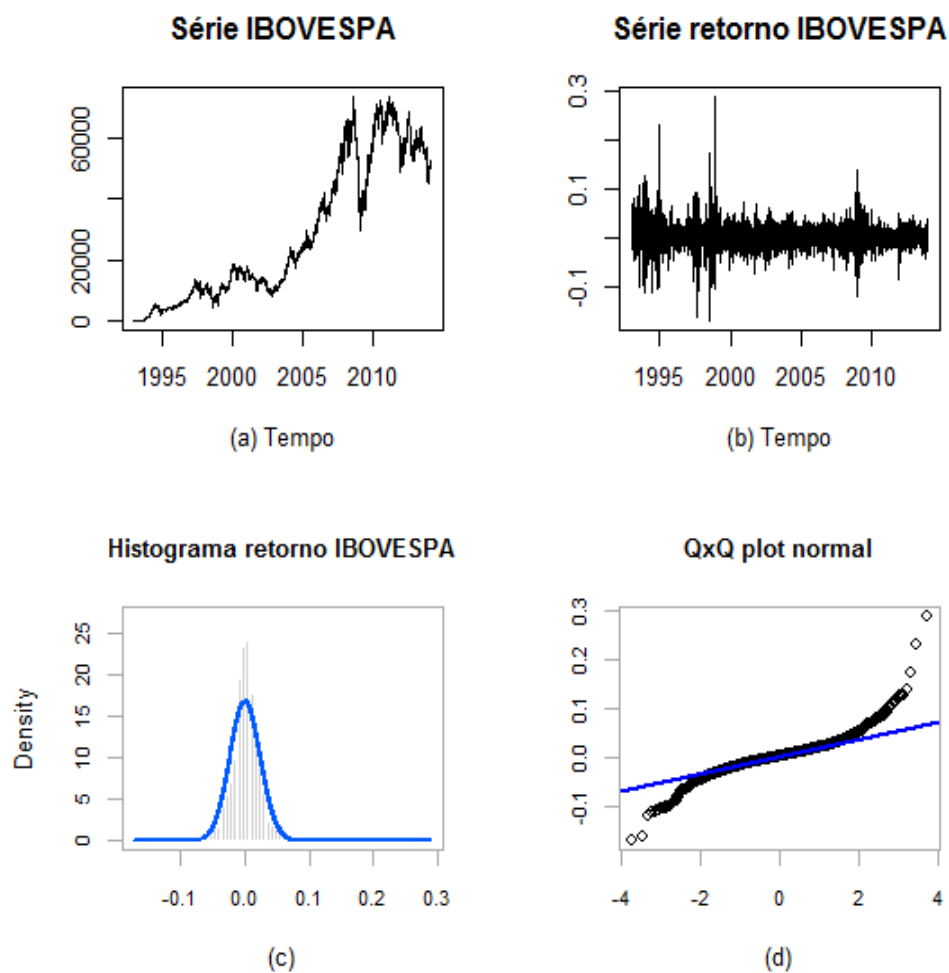


Figura 11 (a) Gráfico da série índice IBOVESPA (Brasil), (b) Série dos retornos IBOVESPA, (c) Histograma com densidade ajustada, (d) Gráfico Q x Q plot normal

Na Figura 12 tem-se em (a) a série temporal de preços do índice SSE. Em (b) a série de retornos, (c) o histograma da série de retornos e em (d) o gráfico QxQ plot da série de retornos do índice SSE. O teste de Jarque e Bera também apresentou valor inferior a $2,22e-16$, confirmando a falta de normalidade dos dados.

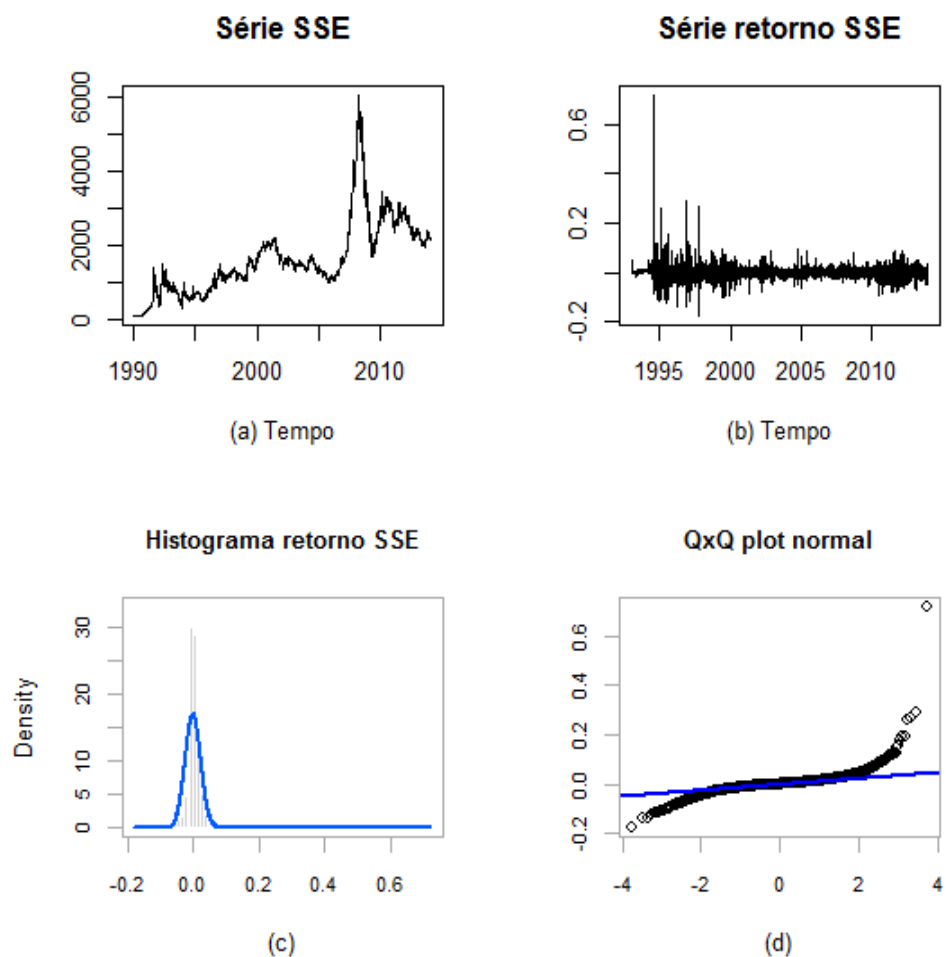


Figura 12 (a) Gráfico da série índice SSE Composite (China), (b) Série dos retornos SSE Composite, (c) Histograma com densidade ajustada, (d) Gráfico Q x Q plot normal

Consideremos, agora, a distribuição dos retornos do IBOVESPA, a função de autocorrelação e a função de autocorrelação parcial representados na Figura 13.

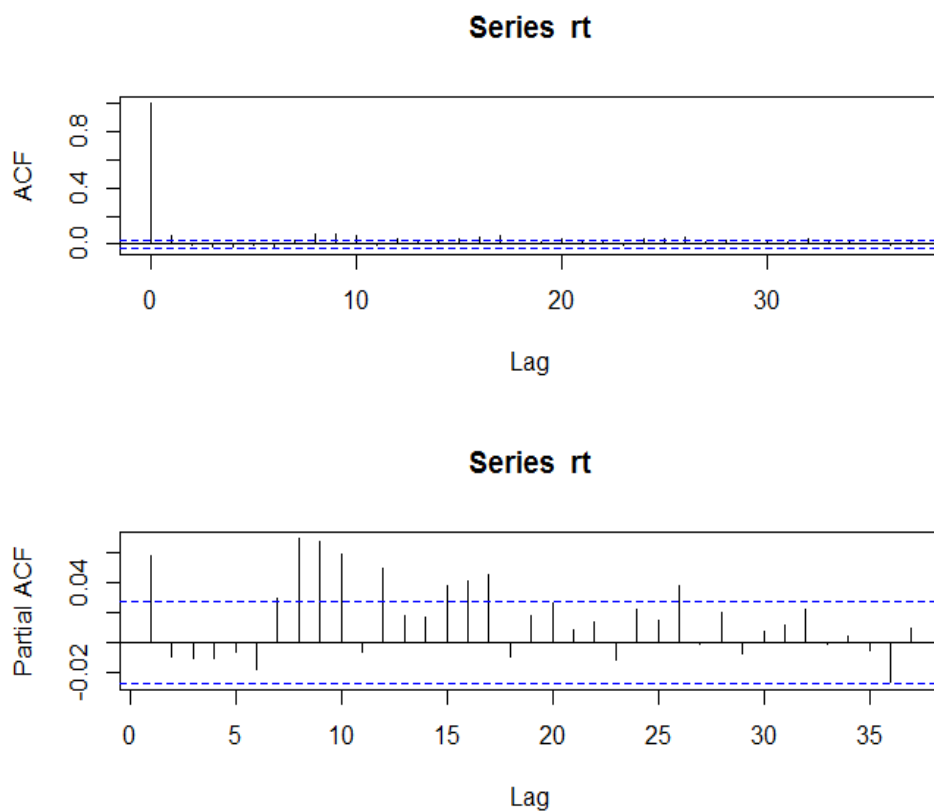


Figura 13 FAC e FACP dos retornos diários do IBOVESPA

Embora as autocorrelações sejam pequenas, podem-se encontrar alguns valores significativos para a autocorrelação e autocorrelação parcial. Em virtude desta estrutura de dependência linear entre os retornos, primeiramente ajustemos um modelo ARMA (p, q) para, posteriormente, modelar a volatilidade condicional da série de retornos.

Tabela 3 Modelos ARMA ajustados para a série de retornos IBOVESPA

Modelos	AIC
ARMA (0,0)	- 24408,73
ARMA (1,0)	- 24424,17
ARMA (2,0)	- 24422,68
ARMA (3,0)	- 24421,30
ARMA (0,1)	- 24424,44
ARMA (0,2)	- 24422,58
ARMA (0,3)	- 24421,19
ARMA (1,1)	- 24422,53
ARMA (1,2)	- 24420,60
ARMA (2,1)	- 24420,65
ARMA (2,2)	- 24419,03

De acordo com a Tabela 3 o modelo ARMA ajustado para a série de retornos é o ARMA (0,1) pelo critério de seleção de modelos AIC, pois é o que apresentou o maior valor em módulo do que os demais modelos ajustados. A Figura 14 apresenta os resíduos padronizados e também a função de autocorrelação e gráfico com os p -valores, no qual se vê a necessidade de ajustar um modelo ARMA com ordem maior do que os apresentados na Tabela 3.

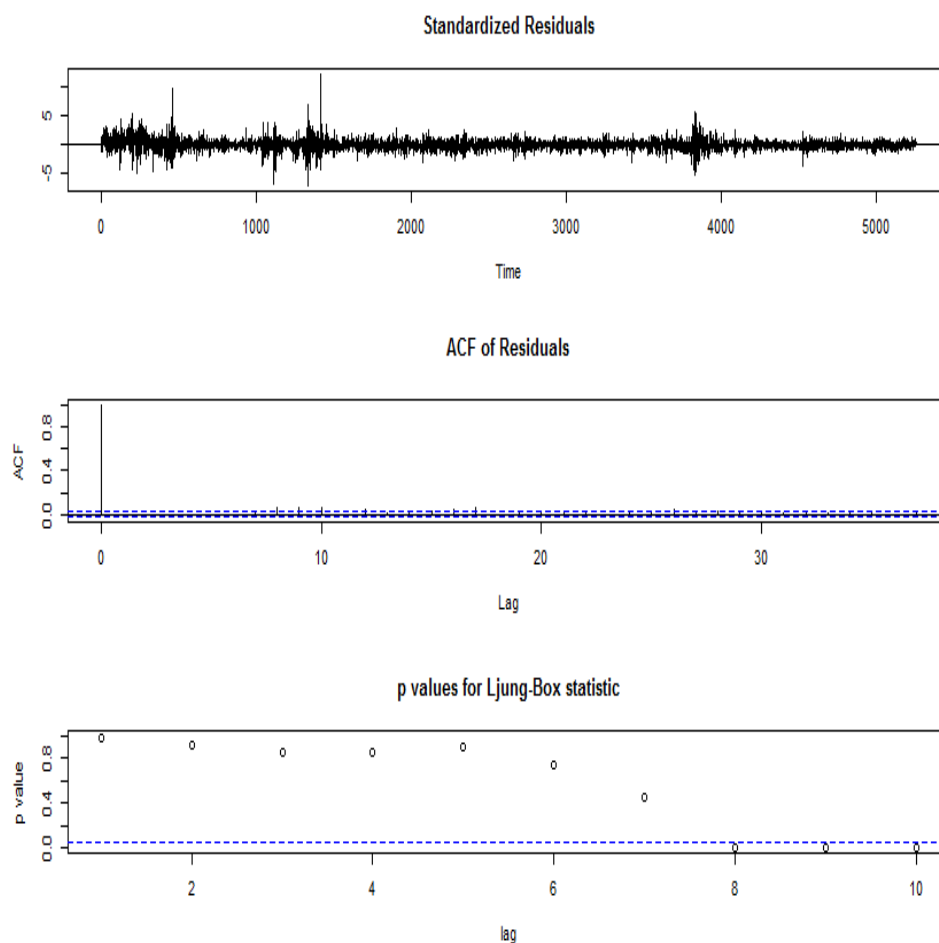


Figura 14 Resíduos padronizados, FAC dos resíduos do ajuste do modelo ARMA (0,1) para a série de retornos IBOVESPA

Consideremos, agora, a análise da função de correlação cruzada entre os resíduos e a série de retornos. Esta está representada na Figura 15. Por meio desta figura, percebe-se que tanto as funções de autocorrelação como a função de correlação cruzada indicam que se deve aumentar a ordem ARMA do modelo. Desta forma, tem-se que tentar um modelo ARMA com uma ordem maior do que as apresentadas na Tabela 3.

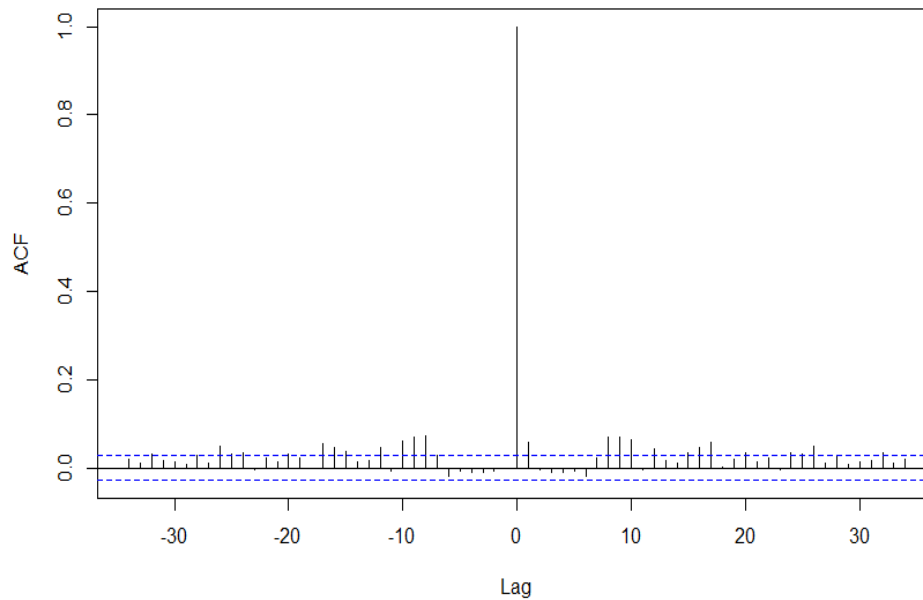


Figura 15 Função de correlação cruzada entre os resíduos e a série de retornos

Assim, consideremos, por exemplo, o modelo ARMA (10,0), cujos gráficos de resíduos encontram-se representados na Figura 16. Por meio deste, percebe-se que há não correlação entre na série de resíduos. Logo, este modelo será utilizado para a modelagem inicial da série de retornos do IBOVESPA, para posteriormente fazer a modelagem da volatilidade por meio da classe de modelos GARCH.

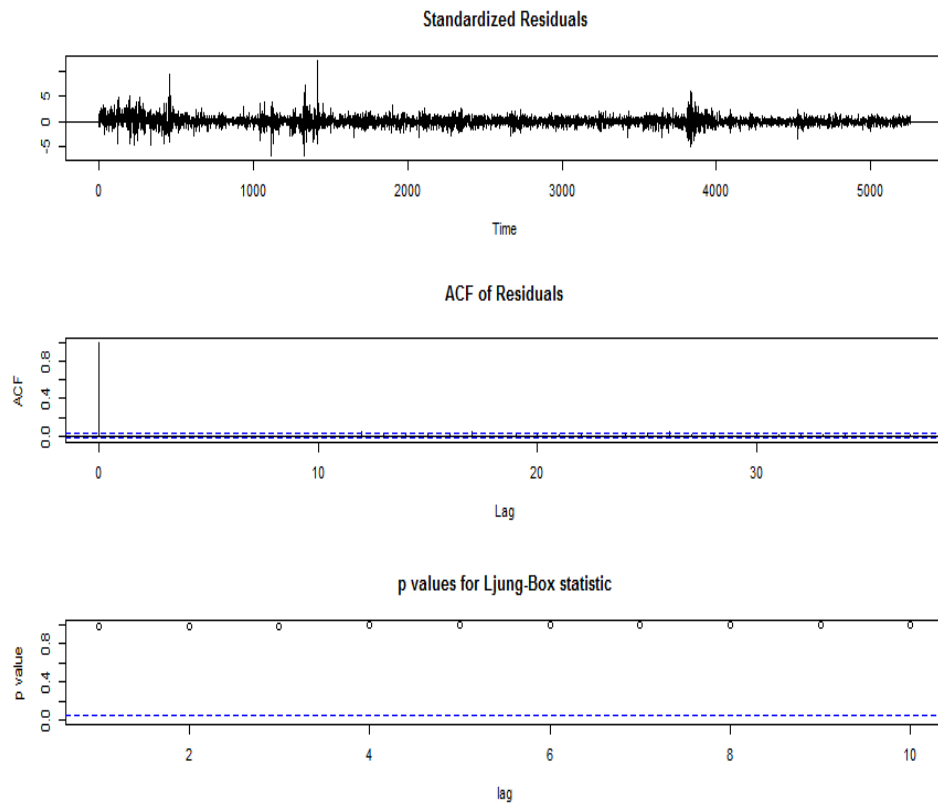


Figura 16 Resíduos padronizados FAC e FACP dos resíduos do ajuste do modelo ARMA (10,0) para a série de retornos IBOVESPA

Na Figura 17 apresentam-se as FAC e FACP dos resíduos ao quadrado. Nesta figura é importante ver que os resíduos ao quadrado possuem uma forte estrutura de autocorrelação.

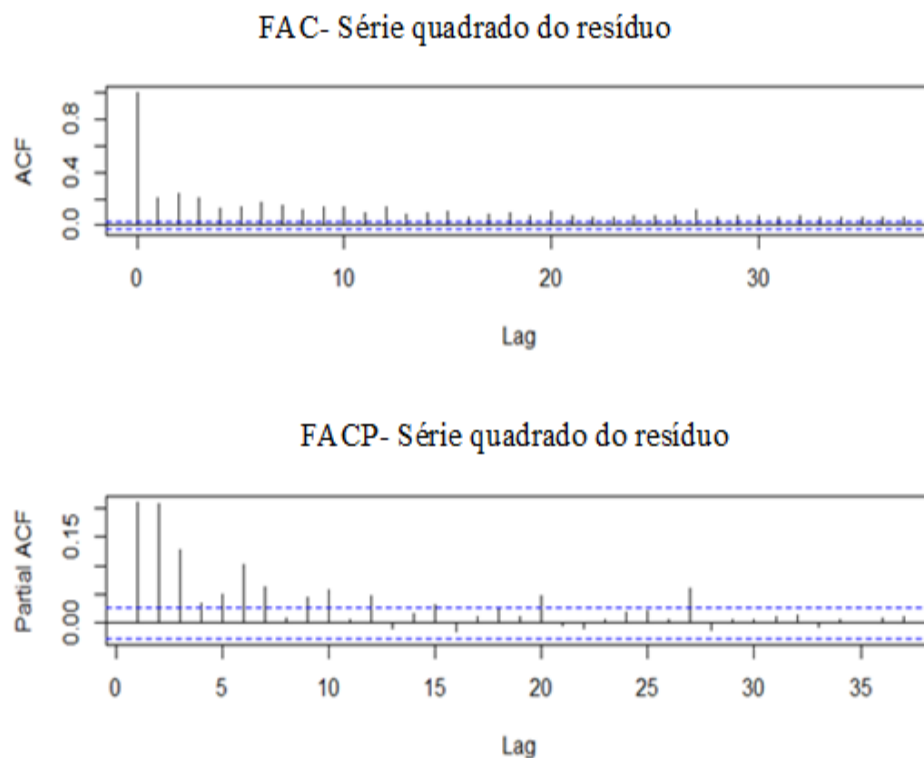


Figura 17 FAC e FACP dos resíduos ao quadrado do ajuste do modelo ARMA (10,0) para a série de retornos IBOVESPA

Ajustando o modelo AR (10)-GARCH (1,1), com erros t de Student e eliminando os coeficientes não significativos, o modelo ajustado foi:

$$y_t = 0,0014 + 0,0316y_{t-1} + 0,0276r_{t-8} + 0,0400r_{t-10} + a_t,$$

$$a_t = \sqrt{h_t} \varepsilon_t,$$

$$h_t = 0,0000062 + 0,0979a_t^2 + 0,8974h_{t-1}. \quad (123)$$

O ajuste do modelo (123) apresentou os valores da estatística de Ljung-Box para os resíduos padronizados, \tilde{a}_t , que são iguais a $Q(10) = 7,54$ com $p = 0,67$ e $Q(20) = 20,94$ com $p = 0,13$. O valor da estatística de Ljung-Box para os resíduos ao quadrado foram $Q(10)$ e $Q(20)$ foram aproximadamente iguais a um. O teste LM mostrou que não há heterocedasticidade na série de resíduos do modelo ajustado. A Figura 18 tem-se o gráfico com os resíduos padronizados e os resíduos ao quadrado do ajuste do modelo (123).

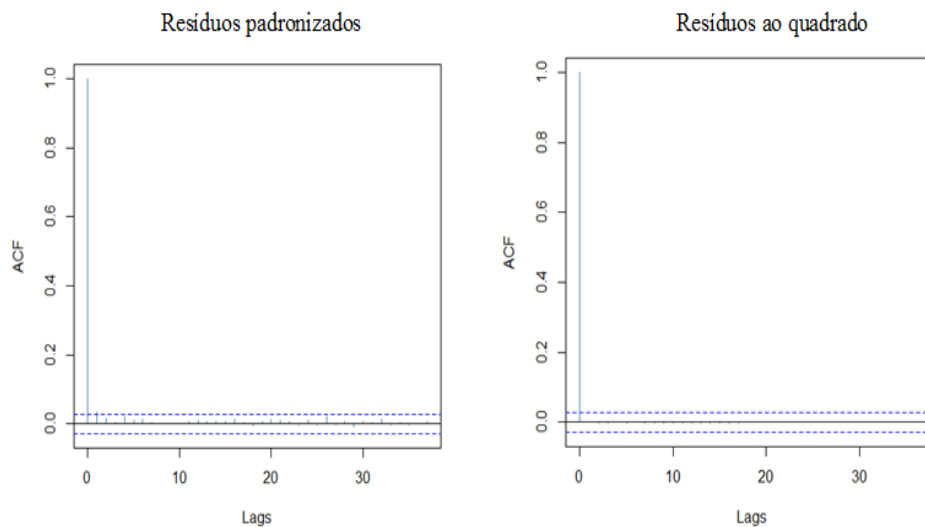


Figura 18 FAC dos resíduos padronizados e ao quadrado do ajuste do modelo AR (10)-GARCH (1,1)

Agora, considerando o ajuste do mesmo modelo AR (10)-GARCH (1,1), porém com inovações skew-t e eliminando os coeficientes não significativos, o modelo ajustado foi:

$$\begin{aligned}
 y_t &= 0,0009 + 0,0276r_{t-9} + 0,0400r_{t-10} + a_t, \\
 a_t &= \sqrt{h_t} \varepsilon_t, \\
 h_t &= 0,000057 + 0,08961a_t^2 + 0,8997h_{t-1}.
 \end{aligned} \tag{124}$$

Ambas as distribuições t de Student e skew- t para o modelo AR (10) – GARCH(1,1) foram adequadas para a modelagem da série de retornos do IBOVESPA. Na Tabela 4 tem-se o valor de AIC para ambos os modelos propostos, com inovações t de Student e skew- t .

Tabela 4 Critério de seleção de modelos – Retornos IBOVESPA

Modelo AR (10) – GARCH (1,1)	AIC	BIC
Distribuição t de Student	- 5,073581	- 5,054845
Distribuição skew- t	- 5,076053	-5,056068

O interessante, neste caso, é perceber que, para o modelo AR (10) – GARCH (1,1), com inovações t de Student e skew- t , apesar de não apresentar resultados muito distintos de AIC e BIC, ainda assim, mostra que a distribuição skew- t modela o conjunto da série de retornos de forma superior do que a distribuição t de Student.

Para a série de retornos SSE, o procedimento é o mesmo feito para a série de retornos IBOVESPA. Desta forma, um modelo indicado poderia ser o ARMA (5,3), que forma com este modelo a correlação serial presente na série que seria extinta. Considerando, então, o ajuste do ARMA (5,3) – GARCH (1,1) com erros t de Student tem-se os coeficientes do modelo mostrados na Tabela 5.

Tabela 5 Parâmetros do Modelo ARMA (5,3) – GARCH (1,1) – t de Student

Parâmetros	Modelo ARMA (5,3) – GARCH (1,1)
Constante	0,000251
AR (1)	- 0,635300
AR (2)	0,355300
AR (3)	0,795000
AR (4)	0,055480
AR (5)	0,046550
MA (1)	0,066940
MA (2)	- 0,301800
MA (3)	- 0,662400
ω	0, 000007
α_1	0,196700
β_1	0,826900
ν	3,390000

A Tabela 6 apresenta o mesmo para a série de retornos SSE, isto é, o ARMA (5,3) – GARCH (1,1), porém com inovações skew-t. O parâmetro de assimetria foi bem próximo de zero, isto é, 0,9778. Um ponto importante é que com as inovações skew-t no modelo, a constante não foi significativa. Na Tabela 7 apresentam-se os valores de AIC e BIC para o modelo ARMA (5,3)-GARCH(1,1) com inovações t de Student e skew-t.

Tabela 6 Parâmetros do Modelo ARMA (5,3) – GARCH (1,1) – skew-t

Parâmetros	Modelo ARMA (5,3) – GARCH (1,1)
AR (1)	- 0,636100
AR (2)	0,351300
AR (3)	0,786000
AR (4)	0,0564800
AR (5)	0,0480900
MA (1)	0,0669300
MA (2)	- 0,300400
MA (3)	- 0,662900
ω	0, 000007
α_1	0,198300
β_1	0,817500
ν	3,390000
γ	0,978300

Os gráficos das funções de autocorrelações para os resíduos padronizados e os resíduos ao quadrado para o ajuste do modelo ARMA (5,3)-GARCH (1,1) para a série de retornos SSE com inovações t de Student e skew- t são praticamente idênticos. Desta forma, a fim de mostrar que o ajuste do modelo foi adequado, considerar-se-á a Figura 19. Nesta, a FAC para os resíduos padronizados e resíduos ao quadrado não apresentam autocorrelação na variância, este fato também foi confirmado pelo teste de Ljung-Box para os resíduos e resíduos ao quadrado ao nível de 1 % de probabilidade. O teste LM mostrou que não há efeitos de heterocedasticidade nas séries de resíduos do modelo ajustado.

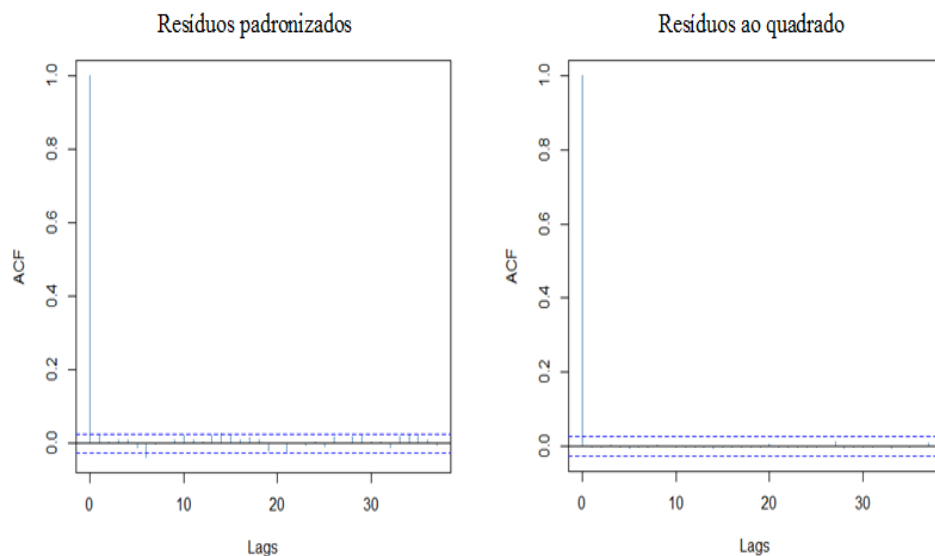


Figura 19 FAC dos resíduos padronizados e ao quadrado do ajuste do modelo ARMA (5,3) – GARCH (1,1)

Na Tabela 6 tem-se a comparação entre os ajustes do modelo ARMA (5,3) – GARCH (1,1) com inovações t de Student e skew- t .

Tabela 7 Critério de seleção de modelos – Retornos SSE

Modelo ARMA (5,3) – GARCH (1,1)	AIC	BIC
Distribuição t de Student	-5.538362	-5.523900
Distribuição skew- t	-5.538351	-5.522776

Apesar das diferenças apresentadas na Tabela 7 serem pequenas, o modelo skew- t apresentou melhor resultado no BIC do que o modelo com inovações t de Student. O interessante é que, apesar da série de retornos terem uma assimetria bem alta (5,63), o ajuste t de Student mostrou-se suficiente.

4.2 Análise da persistência da volatilidade

Relembrando que persistência de um modelo GARCH tem a ver com o quão rápido grande volatilidades decaem depois de um choque. Para o GARCH (1,1), a estatística chave é modelar a soma dos dois parâmetros principais (α_1 e β_1).

Na medida de *half-life* (em dias), para a série de retornos do IBOVESPA, apresentou-se valor de 58 dias para o modelo ajustado em (123) e 64 dias para (124). O interessante é notar que, conforme mudou a distribuição do erro, o valor de *half-life*, também, alterou-se. Para a série de retornos do SSE tem-se uma alta persistência na volatilidade, uma vez que a soma dos parâmetros α_1 e β_1 foi praticamente igual a um.

Uma das dificuldades na análise dos processos GARCH em geral, e dos IGARCH em particular, é o que certos processos podem ser persistentes num certo modo de convergência e não ser num outro modo de convergência. Assim, pode-se provar que os IGARCH, embora persistentes de acordo com a medida $E(h_{t+j} - \sigma^2 | F_t)$ não são persistentes na forma de convergência designada por “quase certamente” ou “com probabilidade 1”. Desta forma, em muitas aplicações empíricas, observa-se que a estimativa de $\alpha_1 + \beta_1$ se encontra muito próximo de um, sugerindo que a volatilidade pode ser persistente (forte dependência temporal). Um modelo estacionário, mas em volatilidade e exibe memória longa é FIGARCH (GARCH fracionário).

4.3 Análise de Memória Longa

Para a análise ML das séries de retornos utilizada neste estudo, procedeu-se a estatística R/S e por meio do procedimento GPH estimou-se o valor de \hat{d} . A Tabela 8 apresenta estes resultados.

Tabela 8 Estatística R/S e valor estimado \hat{d} das séries de retornos analisadas

Série	Estatística R/S	Valor de \hat{d}
IBOVESPA	2,7386	0,3422521
SSE	1,2095	0,06118343

O valor de \hat{d} para as séries de retornos sugere que ambas as séries de retornos possuem comportamento de memória longa. Este valor foi calculado por meio do procedimento GPH proposto por Geweke e Porter-Hudak (1983). A estatística R/S possui valores críticos iguais a 1,62 e 1,747 para os níveis de 0,05 e 0,01 de significância, respectivamente. Maiores detalhes podem ser obtidos Lo (1991), Soldá (2008), Morettin (2011). Por meio do teste R/S, percebe-se que as séries IBOVESPA e NASDAQ apresentam comportamento de ML ao nível de 1% de probabilidade. Porém, já pelo valor estimado \hat{d} , pode-se dizer que todas as séries possuem comportamento de ML, pois $0 < \hat{d} < 1/2$. Outro fato interessante foi que a série que apresentou a maior persistência na volatilidade foi a que apresentou menor valor na estatística R/S.

Finalmente, para as séries de retornos analisadas foi considerado o ajuste do modelo FIGARCH e FIEGARCH para modelar o comportamento de longo prazo da volatilidade. Fez-se o ajuste deste modelo com inovações t de Student e skew- t para as séries de retornos do IBOVESPA e SSE.

Na Tabela 9 têm-se os valores de AIC e BIC para os modelos ajustados da série de retornos IBOVESPA. Percebe-se o modelo FIGARCH (1,d,1) com inovações skew- t apresentou maior valor em módulo tanto para AIC quanto para BIC.

Tabela 9 Ajuste modelo FIGARCH – Série retornos IBOVESPA

Modelo FIGARCH (1,d,1)	AIC	BIC
Distribuição t de Student	- 5,070709	- 5,061965
Distribuição skew-t	- 5,072416	- 5,062423

Já na Tabela 10 têm-se também os valores de AIC e BIC, porém para o ajuste do modelo FIEGARCH (1,d,1). Neste caso também, o modelo com inovações skew- t apresentou melhor resultado por meio destes critérios.

Tabela 10 Ajuste modelo FIEGARCH – Série retornos IBOVESPA

Modelo FIEGARCH (1,d,1)	AIC	BIC
Distribuição t de Student	- 5,070580	- 5,060587
Distribuição skew-t	- 5,072392	-5,0611509

Nas Tabelas 11 e 12 têm-se os valores de AIC e BIC para os modelos ajustados da série de retornos SSE. Em ambos os ajustes, isto é, para os modelos FIGARCH (1,d,1) e FIEGARCH (1,d,1) as inovações skew- t apresentaram resultados superiores, se comparadas as inovações t de Student.

Tabela 11 Ajuste modelo FIGARCH – Série retornos SSE

Modelo FIGARCH (1,d,1)	AIC	BIC
Distribuição t de Student	- 5,077202	- 5,064710
Distribuição skew-t	- 5,080210	- 5,066468

Tabela 12 Critério de seleção de modelos – Série retornos SSE

Modelo FIEGARCH (1,d,1)	AIC	BIC
Distribuição t de Student	- 5,070580	- 5,060587
Distribuição skew-t	- 5,072392	-5,0611509

5 CONCLUSÕES

Neste trabalho buscou-se explorar os aspectos teóricos envolvidos na distribuição skew-t sob o método proposto por Fernandez e Steel (1998) e a sua aplicação em modelos heterocedásticos. Uma das contribuições deste trabalho foi fazer uma discussão de como é introduzida esta distribuição, principalmente, no modelo GARCH.

Os modelos combinados ARMA-GARCH mostraram-se satisfatórios, considerando inovações skew-t, que conseguiu modelar de forma adequada as séries de retornos analisadas (IBOVESPA e SSE), porém, por meio do critério de seleção de modelos AIC e BIC, não houve diferenças significativas se comparada à distribuição skew-t com a distribuição t de Student.

Uma das técnicas utilizadas neste trabalho foi a utilização da técnica Bootstrap para a construção de intervalos de confiança para previsões futuras de modelos heterocedásticos, sendo muito útil para avaliar o comportamento de previsões futuras acerca da volatilidade (*clusters*).

Nos resultados acerca da persistência da volatilidade, conhecer esta variável é fundamental para os agentes do mercado, haja vista que isto pode ser um indicador de memória longa e, também, de formação de conglomerados de volatilidade.

Finalmente, para as séries trabalhadas foi feito uma análise de memória longa nestas. A importância de se conhecer o comportamento de ML está atrelada a tomada de decisão e planejamento de estratégias dos agentes do mercado financeiro.

Um dos objetivos deste trabalho é verificar se os modelos combinados ARMA-GARCH com inovações skew-t são adequados também para modelar séries de retornos com comportamento de memória longa. De acordo com os

critérios AIC e BIC estes modelos apresentaram resultados superiores aos modelos FIGARCH e FIEGARCH.

Tanto para os modelos FIGARCH e FIEGARCH com erros skew-t aplicados às séries de retornos IBOVESPA e SSE apresentaram maior adequabilidade do que as inovações t de Student, segundo os critérios de AIC e BIC. Este fato é importante, pois as inovações skew-t além de conseguir captar o excesso de curtose das séries, também são mais robustos na presença da assimetria.

Pretende-se, em trabalhos futuros, a ampliação deste estudo levando em conta uma análise de caráter multivariado, considerando modelos de volatilidade multivariados com inovações skew-t, como por exemplo, VEC-GARCH, BEKK, GOGARCH. Também se sugere o estudo da persistência da volatilidade sob a abordagem Bayesiana para modelos GARCH, FIGARCH e FIEGARCH.

REFERÊNCIAS

AIUBE, F. A. L. **Modelos quantitativos em finanças: com enfoque em commodities**. Porto Alegre: Bookman, 2013. 438 p.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Boston, v. 19, n. 6, p. 716-723, Dec. 1974.

ARDIA, D. **Financial risk management with bayesian estimation of GARCH models: theory and applications**. Berlin: Springer-Verlag, 2009. 203 p. (Lecture Notes in Economics and Mathematical Systems, 612).

AURÉLIO, C. M. A. **Assimetria na volatilidade dos mercados de ações**. 2012. 81 p. Dissertação (Mestrado em Controle de Gestão e dos Negócios) - Instituto Superior de Contabilidade e Administração de Lisboa, Lisboa, 2012.

AZZALINI, A. Further results on a class of distributions which includes the normal ones. **Statistica**, Bologna, v. 46, p. 199-208, 1986.

AZZALINI, A.; CAPITANIO, A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. **Journal of the Royal Statistical Society: Series B**, London, v. 61, n. 3, p. 579-602, 2003.

BAILLIE, R. T.; BOLLERSLEV, T. The message in daily exchange rates: a conditional variance tale. **Journal of Business and Economic Statistics**, New Jersey, v. 7, n. 3, p. 297-305, July 1989.

BAILLIE, R. T.; BOLLERSLEV, T.; MIKKELSEN, H. O. Fractionally integrated generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, Amsterdam, v. 74, n. 1, p. 3-30, Sept. 1996.

BERA, A. K.; HIGGINS, M. L. A class of nonlinear ARCH models. **International Economic Review**, Philadelphia, v. 33, n. 1, p. 137-158, Feb. 1992.

BERNDT, E. K. et al. Estimation and inference in nonlinear structural models. **Annals of Economic and Social Measurement**, New York, v. 3, p. 653-665, 1974.

BLACK, F. Studies of stock price volatility changes. **Proceedings of the Business and Economics Section of the American Statistical Association**, Chicago, v. 1, p. 177-181, 1976.

BLACK, F.; SCHOLES, M. The pricing of options and corporate liabilities. **Journal of Political Economy**, Chicago, v. 81, n. 3, p. 637-654, 1973.

BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. 2. ed. Rio de Janeiro: SBM, 2010. 125 p.

BOLLERSLEV, T. A conditionally heterokedastic time series model for speculative process and rates of return. **Review of Economics and Statistics**, Cambridge, v. 69, p. 542-547, 1987.

BOLLERSLEV, T. Generalized autoregressive conditional heteroscedasticity. **Journal of Econometrics**, Amsterdam, v. 31, p. 307-27, 1986.

BOLLERSLEV, T.; CHOU, R. Y.; KRONER, K. F. ARCH modeling in finance. **Journal of Econometrics**, Amsterdam, v. 52, n. 3, p. 5-59, June 1992.

BOLLERSLEV, T.; ENGLE, R. F.; NELSON, D. B. Arch models. In: ENGLE, R. F.; MCFADDEN, D. L. (Ed.). **Handbook of econometrics**. New York: North Holland, 1994. v. 4, p. 2959-3038.

BOLLERSLEV, T.; WOOLDRIDGE, J. M. Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. **Econometric Reviews**, Cambridge, v. 11, n. 2, p. 143-172, Jan. 1992.

BOLSA DE VALORES, MERCADORIAS E FUTUROS DE SÃO PAULO. Disponível em: <<http://www.bmfbovespa.com.br>>. Acesso em: 16 jun. 2014.

BOX, G. E. P.; PIERCE, D. A. Distribution of the residual autocorrelations in autoregressive integrated moving average time series models. **Journal of the American Statistical Association**, New York, v. 65, p. 1509-1526, 1970.

BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: Wiley, 1992. 588 p.

BRANCO, M. D. E.; DEY, D. K. A general class of multivariate skew elliptical distributions. **Journal of Multivariate Analysis**, New York, v. 79, n. 1, p. 99-113, Oct. 2001.

BUENO, R. de L. S. **Econometria de séries temporais**. 2. ed. São Paulo: Cengage Learning, 2011. 341 p.

CAMPBELL, J. Y.; LO, A. W.; MACKINLAY, A. C. **The econometrics of financial markets**. Princeton: Princeton University, 1997. 661 p.

CARNERO, M. A.; PEÑA, D.; RUIZ, E. **Outliers and conditional autoregressive heteroscedasticity in time series**. Madrid: Universidad Carlos III, 2001. 7 p. (Statistics and Econometrics Series, 4).

CHRISTIE, A. A. The stochastic behavior of common stock variances: value, leverage and interest rate effects. **Journal of Financial Economics**, Lausanne, v. 10, p. 407-432, 1982.

DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. **Journal of the American Statistical Association**, New York, v. 74, p. 427-431, 1979.

DING, Z.; GRANGER, C. W. J.; ENGLE, R. F. A long memory property of stock market returns and a new model. **Journal of Empirical Finance**, New York, v. 1, n. 1, p. 1-131, June 1993.

ENGLE, R. F. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. **Econometria**, Rio de Janeiro, v. 50, n. 4, p. 978-1007, 1982.

FERNANDEZ, C.; STEEL, M. On bayesian modeling of fat tails and skewness. **Journal of the American Statistical Association**, New York, v. 93, n. 441, p. 359-371, 1998.

GELMAN, A.; RUBIN, D. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-511, 1992.

GEWEKE, J.; PORTER-HUDAK, S. The estimation and application of long memory time series models. **Journal of Time Series Analysis**, Clevedon, v. 4, p. 221-237, 1983.

GLOSTEN, L. R.; JAGANNATHAN, R.; RUNKLE, D. On the relation between the expected value and the volatility of the nominal excess return on stocks. **Journal of Finance**, New York, v. 48, n. 5, p. 1779-1801, Dec. 1993.

GODOI, L. G. **A distribuição t-assimétrica univariada: propriedades e inferência.** 2007. 106 p. Dissertação (Mestrado em Estatística) - Universidade de São Paulo, São Paulo, 2007.

GODOI, L. G.; BRANCO, M. D. E. **A distribuição t-assimétrica univariada: propriedade e inferência.** São Paulo: USP, 2008. 8 p.

GOLDSTEIN, A. **Constructive real analysis.** New York: Harper e Row, 1967. 178 p.

GRANGER, C. W. G. Long memory relationships and the aggregation of dynamic models. **Journal of Econometrics**, Amsterdam, v. 14, p. 227-238, 1980.

HEYNEN, R. C.; KAT, H. M. Partial barrier options. **Journal of Financial Engineering**, Singapore, v. 3, n. 3, p. 253-274, Sept./Dec. 1994.

HOSKING, J. R. M. Fractional differencing. **Biometrika**, London, v. 68, p. 165-176, 1981.

HURST, H. E. A suggested statistical model of time series that occur in nature. **Nature**, London, v. 180, p. 494, 1957.

HURST, H. E. Long term storage capacity of reservoirs. **Transactions of the American Society of Civil Engineers**, New York, v. 116, p. 770-799, 1951.

HURVICH, C. M.; TSAI, C. L. Regression and time series model selection in small samples. **Biometrika**, London, v. 76, p. 297-307, 1989.

JARQUE, C.; BERA, A. A. A test for normality of observations and regression residuals. **International Statistical Review**, Edinburgh, v. 55, p. 163-172, 1987.

JOHNSTON, J.; DINARDO, J. **Econometric methods.** New York: McGraw Hill, 1997. 480 p.

JONES, M. C. Student's simplest distribution. **The Statistician**, Washington, v. 51, n. 1, p. 41-49, Mar. 2002.

JONES, M. C.; FADDY, M. J. A skew extension of the t-distribution, with application. **Journal of Royal Statistical Society, Series B**, London, v. 65, n. 1, p. 159-174, Feb. 2003.

KENDALL, M. G. The analysis of economic time series: part 1, prices. **Journal of the Royal Statistical Society, A**, London, v. 116, p. 11-25, 1953.

LAMBERT, P.; LAUREN, S. **Modelling financial time series using GARCH-type models and a skewed student density**. Liège: Université de Liège, 2001.

LAMBERT, P.; LAUREN, S. **Modelling skewness dynamics in series of financial data**. Louvain-la-Neuve: Institut de Statistique, 2000. 32 p.

LEE, J. H. H. A Lagrange multiplier test for GARCH models. **Economics Letters**, Amsterdam, v. 37, n. 3, p. 265-271, Nov. 1991.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, London, v. 65, p. 297-303, 1978.

LO, A. W. Long term memory in stochastic market prices. **Econometrica**, Chicago, v. 59, n. 5, p. 1279-1313, Sept. 1991.

LOPES, S. R. C.; MENDES, B. V. M. Bandwidth selection in classical and robust estimation of long memory. **International Journal of Statistics and Systems**, New Delhi, v. 1, n. 2, p. 177-200, 2006.

MAHEU, J. Can GARCH models capture long-range dependence? **Studies in Nonlinear Dynamics e Econometrics**, Berlin, v. 9, n. 4, p. 1-41, 2005.

MANDELBROT, B. B.; WALLIS, J. Noah, Joseph and operational hydrology. **Water Resources Research**, Washington, v. 4, p. 909-918, 1968.

MARKOWITZ, H. Portfolio selection. **Journal of Finance**, New York, v. 7, n. 1, p. 77-91, Mar. 1952.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. Singapore: McGraw-Hill, 1974. 564 p.

MORETTIN, P. A. **Econometria financeira: um curso em séries temporais financeiras**. 2. ed. São Paulo: E. Blücher, 2011. 383 p.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2. ed. São Paulo: E. Blücher, 2006. 538 p.

NELSON, D. B. Conditional heteroskedasticity in asset returns: a new approach. **Econometrica**, Chicago, v. 59, n. 2, p. 347-370, Mar. 1991.

NEWKEY, W. K.; WEST, K. D. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. **Econometrica**, Chicago, v. 55, p. 703-708, 1987.

NICOLAU, J. **Modelação de séries temporais financeiras**. Coimbra: Fundação Económicas; Almedina, 2012. 504 p.

PENTULA, S. Modelling the persistence of conditional variances: a comment. **Econometric Review**, Atlanta, v. 5, p. 71-74, 1986.

PERRON, P.; VOGELSANG, T. J. Nonstationarity and level shifts with an application to purchasing power parity. **Journal of Business e Economic Statistics**, Oxford, v. 10, n. 3, p. 301-320, July 1992.

POON, S.; GRANGER, C. W. J. Forecasting volatility in financial markets: a review. **Journal of Economic Literature**, Nashville, v. 41, n. 2, p. 478-539, June 2003.

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 10 dez. 2013.

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, Hayward, v. 6, n. 2, p. 461-464, Mar. 1978.

SCHWERT, W. Stock volatility and the Crash of '87. **Review of Financial Studies**, Oxford, v. 3, n. 1, p. 77-102, Oct. 1990.

SENTANA, E. Quadratic ARCH models. **Review of Economic Studies**, Bristol, v. 62, n. 4, p. 639-661, Oct. 1995.

SOLDÁ, G. Y. **Modelos de memória longa, GARCH e GARCH com memória longa para séries financeiras**. 2008. 187 p. Dissertação (Mestrado em Estatística) - Universidade Federal de São Paulo, São Paulo, 2008.

SWARAY, R. B. **Volatility of primary commodity prices: some evidence from agricultural exports in Sub-Saharan Africa**. New York: University of York, 2002. 26 p. (Discussion Papers in Economics, 2002/06).

TAYLOR, S. J. **Modeling financial time series**. Chichester: J. Wiley, 1986. 268 p.

TSAY, R. S. **Analysis of financial time series**. New York: Wiley, 2002. 605 p.

WEHRENS, R.; PUTTER, H.; BUYDENS, L. The bootstrap: a tutorial. **Chemometrics and Intelligent Laboratory Systems**, Berlin, v. 54, n. 1, p. 35-52, Dec. 2000.

ZAKOIN, J. M. Threshold heteroscedasticity models. **Journal of Economic Dynamics and Control**, New York, v. 18, n. 5, p. 931-955, Sept. 1994.

APÊNCICES

APÊNDICE A - Estimação de Modelos GARCH

O modelo de regressão GARCH (p, q) pode ser obtido considerando os ε_t 's como sendo inovações de um modelo de regressão linear da forma

$$\varepsilon_t = y_t - x_t' b, \quad (1)$$

em que y_t é a variável dependente, x_t um vetor de variáveis de variáveis explicativas e b um vetor de parâmetros desconhecidos. Sendo assim, por meio de (1), diferenciando com respeito a média dos parâmetros obtém-se:

$$\frac{\partial^2 l(y_t | b)}{\partial b} = \varepsilon_t x_t h_t^{-1} + \frac{1}{2} h_t \frac{\partial h_t}{\partial b} \left(\frac{\varepsilon_t^2}{h_t} - 1 \right), \quad (2)$$

$$\begin{aligned} \frac{\partial^2 l(y_t | b)}{\partial b \partial b'} &= -h_t^{-1} x_t x_t' - \frac{1}{2} h_t^{-2} \frac{\partial h_t}{\partial b} \frac{\partial h_t}{\partial b'} \left(\frac{\varepsilon_t^2}{h_t} \right) \\ &= -2h_t^{-2} \varepsilon_t x_t \frac{\partial h_t}{\partial b} + \left(\frac{\varepsilon_t^2}{h_t} - 1 \right) \frac{\partial}{\partial b'} \left[\frac{1}{2} h_t^{-1} \frac{\partial h_t}{\partial b} \right], \end{aligned} \quad (3)$$

sendo $\frac{\partial h_t}{\partial b} = -2 \sum_{i=1}^p \alpha_i x_{t-i} \varepsilon_{t-i} + \sum_{j=1}^q \beta_j \frac{\partial h_{t-j}}{\partial b}$.

Finalmente, os elementos de fora da diagonal da matriz de informação de Fisher podem ser considerados iguais a zero. Sendo assim, para obter os estimadores de máxima verossimilhança é necessário a utilização de um método iterativo.

APÊNDICE B - Estimação por meio do algoritmo BHHH

O algoritmo BHHH (Berndt, Hall, Hall and Hausman, 1974) é um método iterativo para encontrar as estimativas de máxima verossimilhança desses parâmetros. O algoritmo BHHH pode ser descrito por

$$\theta^{(i+1)} = \theta^{(i)} + \delta^{(i)} \left(\sum_{t=1}^T \frac{\partial^2 l_t^i(\theta)}{\partial \theta \partial \theta'} \right)^{-1} \left(\sum_{t=1}^T \frac{\partial l_t^i(\theta)}{\partial \theta} \right) \quad (1)$$

sendo $l_t(\theta)$ a t - ésima log-verossimilhança condicional, θ o vetor paramétrico e $\delta^{(i)}$ uma variável que determina o tamanho do salto na i - ésima iteração, sendo uma variável escolhida de forma a maximizar a função de verossimilhança.

Berndt et al. (1974) introduziu o critério descrito a seguir, a fim de escolher a variável $\delta^{(i)}$. Sendo assim, considere $V(\theta)$ uma função contínua e diferenciável, no qual se tem o objetivo de maximizar esta para encontrar o valor de θ .

Considere o gradiente de $V(\theta)$ com relação a θ , isto é, $g = [\partial V(\theta) / \partial \theta]$ e a matriz Hessiana Q obtido da inversa da matriz de derivadas segundas de $V(\theta)$. Agora, consideremos a função $\lambda(\theta, \delta)$ da forma

$$\lambda(\theta, \delta) = \frac{V(\theta + \delta d) - V(\theta)}{\delta d' g} \quad (2)$$

com $d = Q.g$.

Se $\lambda(\theta, 1) \geq \psi$, faça $\delta = 1$. Do contrário, escolha um valor de δ que satisfaça a condição

$$\psi \leq \lambda(\theta, \delta) \leq 1 - \psi, \quad (3)$$

em que ψ é uma constante no intervalo $(0; 0,5)$.

Sendo $V(\theta)$ uma função contínua e diferenciável, então existe um valor de δ que satisfaz o critério (3). Para demonstrar esta afirmação, calculemos o seguinte limite

$$\begin{aligned} \lim_{\delta \rightarrow 0} \psi(\theta, \delta) &= \lim_{\delta \rightarrow 0} \frac{V(\theta + \delta d) - V(\theta)}{\delta d'g} \\ &= \frac{1}{d'g} \lim_{\delta \rightarrow 0} \frac{V(\theta + \delta d) - V(\theta)}{\delta} = \frac{1}{d'g} d' \frac{\partial V(\theta)}{\partial \theta} = \frac{1}{d'g} d'g = 1. \end{aligned}$$

Desta forma, pelo fato da função $V(\theta)$ ser contínua, $\lambda(\theta, 1) < \psi$, $\lim_{\delta \rightarrow 0} \lambda(\theta, \delta) = 1$ e pelo teorema do valor intermediário, tem-se que existe um δ que satisfaça o critério (3).

Satisfeitas estas condições, a convergência é verificada no processo de maximização de $V(\theta)$. Para maiores detalhes da prova de convergência para este caso, pode-se consultar Goldstein (1967).

Considerando que $V(\theta)$ possui derivada segunda em um conjunto compacto com contorno definido e a sequência $\theta_1, \theta_2, \dots$, em que

$$\theta^{(i+1)} = \theta^{(i)} + \delta^{(i)} d^{(i)}, \quad (4)$$

com $d^{(i)} = Q^{(i)} g^{(i)}$ e $Q^{(i)}$ satisfazendo $(d'g)/(d'd) > \psi$, sendo ψ um valor positivo menor que um e $\delta^{(i)}$ satisfazendo a condição (3), então $\lim_{i \rightarrow \infty} g^i = 0$.

O processo de estimação para outros modelos família GARCH segue o mesmo procedimento apresentando apenas algumas pequenas modificações,

levando em consideração que são simplificações ou complementações do modelo GARCH.

APÊNCICE C - Códigos

ALGUNS CÓDIGOS UTILIZADOS NESTE TRABALHO

```

## Estatísticas Descritivas
y<-read.table("ibovespa.txt",h=T)
attach(y)
names(y)
plot(ts(Adj)) ## Gráfico da Série Temporal
rt<-diff(log(Adj)) ### ou então da forma
summary(rt)
n <- length(Adj)
rt <- log((Adj[2:n])/(Adj[1:(n - 1)]))
plot(ts(rt)) ## Gráfico da Série de Retornos
library(fBasics)
basicStats(rt)
par(mfrow=c(2,2))
gnp1 <- ts(Adj,start = c(1993,4),end =c(2014,6), frequency = 240)
plot(gnp1,xlab="(a) Tempo",ylab="",main="Série IBOVESPA")
gnp2 <- ts(rt,start = c(1993,4),end =c(2014,6), frequency = 240)
plot(gnp2,xlab="(b) Tempo",ylab="",main="Série retorno IBOVESPA")
library(PerformanceAnalytics)
## Histograma e QxQ plot
chart.Histogram(rt,xlab="(c)",ylab="",main="Histograma retorno
IBOVESPA",breaks=90,methods = c("add.normal","add.centered"))
chart.QQPlot(rt, xlab="(d)",ylab="",main="QxQ plot normal",distribution =
"norm")

```

```
## Função de autocorrelação e autocorrelação parcial
par(mfrow = c(2, 2))
acf(rt,xlab="(a) Fac dos retornos do IBOVESPA",main="")
acf(rt^2,xlab="(b) Fac dos retornos ao quadrado do IBOVESPA",main="")
pacf(rt,xlab="(c) Facp dos retornos do IBOVESPA",main="")
pacf(rt^2,xlab="(d) Facp dos retornos ao quadrado do IBOVESPA",main="")
library(tseries)
adf.test(rt1) ## Teste ADF
pp.test(rt1) ## Teste PP
ArchTest(rt1, lags=2, demean = FALSE)
ArchTest(rt1, lags=10, demean = FALSE)
ArchTest(rt1, lags=50, demean = FALSE)
m1<-garchFit(data =rt1, formula = ~arma(0,0) + garch(1,0),cond.dist = "sstd")
summary(m1)

## Estimação ARMA-GARCH
m1<-garchFit(data =rt1, formula = ~arma(10,0) + garch(1,1),cond.dist = "sstd")
summary(m1)
```

```

## Memória Longa [R] timeseries: R/S (rescaled range) analysis
rs.test <- function(x, q, alpha)
{
  xbar <- mean(x)
  N <- length(x)
  r <- max(cumsum(x-xbar)) - min(cumsum(x-xbar))
  kovarianzen <- NULL
  for (i in 1:q)
  {
    kovarianzen <- c(kovarianzen,
sum((x[1:(N-i)]-xbar)*(x[(1+i):N]-xbar)))
  }
  if (q > 0)
  s <- sum((x-xbar)^2)/N + sum((1-(1:q)/(q+1))*kovarianzen)*2/N
  else
  s <- sum((x-xbar)^2)/N
  rs <- r/(sqrt(s)*sqrt(N))
  method <- "Teste R/S Memória Longa"
  names(rs) <- "Estatística R/S"
  names(q) <- "Bandwidth q"
  structure(list(statistic = rs, parameter = q, method = method,
data.name=deparse(substitute(x))), class="htest")
}

```