

A FEA e a USP respeitam os direitos autorais deste trabalho. Nós acreditamos que a melhor proteção contra o uso ilegítimo deste texto é a publicação online. Além de preservar o conteúdo motiva-nos oferecer à sociedade o conhecimento produzido no âmbito da universidade pública e dar publicidade ao esforço do pesquisador. Entretanto, caso não seja do interesse do autor manter o documento online, pedimos compreensão em relação à iniciativa e o contato pelo e-mail bibfea@usp.br para que possamos tomar as providências cabíveis (remoção da tese ou dissertação da BDTD).

UNIVERSIDADE DE SÃO PAULO

FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

MESTRADO PROFISSIONALIZANTE

"MODELAGEM MATEMÁTICA EM FINANÇAS"

ESTIMAÇÃO DE VOLATILIDADE REALIZADA DE TELEMAR PN COM
USO DE DADOS DE ALTA FREQUÊNCIA

Marcello D. S. Paixão

Orientador: Prof. Dr. Pedro A. Morettin

São Paulo, 10 de Agosto de 2005

T519.536 P1490

T95649



20000636870



FEAUSP

Powered by **MidPreStar** - www.logprocess.com.br

ESTIMAÇÃO DE VOLATILIDADE REALIZADA DE TELEMAR PN COM USO DE DADOS DE ALTA FREQUÊNCIA

Marcello D. S. Paixão

Orientador: Prof. Dr. Pedro A. Morettin

DEDALUS - Acervo - FEA



20600035878

Dissertação apresentada à Faculdade de
Economia, Administração e Contabilidade e ao
Instituto de Matemática e Estatística da
Universidade de São Paulo para obtenção do
Título de Mestre.

São Paulo, 10 de Agosto de 2005

USP - FEA - SBD
DATA DA DEFESA <u>13, 10, 05</u>

2

95649

FICHA CATALOGRÁFICA

Paixão, Marcello D. S.

Estimação de volatilidade realizada de Telemar PN com uso de dados
de alta frequência / Marcello D. S. Paixão. -- São Paulo, 2005.

108 p.

AGRADECIMENTOS

À Ana Lúcia e aos meus familiares por me apoiarem durante o mestrado e ao longo de minha vida;

Ao meu orientador, Prof. Dr. Pedro A. Morettin, pela valiosa orientação em todos os assuntos pertinentes à dissertação;

Ao Alberto Foltran Berti, Marcelo Sardelich e Rodrigo Chicarolli pelos conselhos relacionados ao S-Plus e tratamento de dados.

RESUMO

Essa dissertação tem como objetivo a estimação de volatilidade realizada de Telemar PN para 1 dia e 21 dias úteis, com uso de dados diários e dados de alta frequência, utilizando-se modelagem de séries temporais e modelagem de regressão. A volatilidade implícita foi usada como variável explicativa no modelo de regressão. A conclusão do estudo indica razoável eficácia dos modelos de séries temporais utilizando dados intradiários para previsão de volatilidade realizada de 1 dia e poucos dias úteis quando comparada aquela previsão utilizando dados diários. A modelagem perde eficácia quando o período para o qual se deseja projetar volatilidade cresce, ou seja, a medida em que se aumenta o número de passos à frente.

ABSTRACT

The goal of this dissertation is to project realized volatility of Telemar PN, the most liquid Brazilian stock during the period of study, for 1-day and 21-day periods, using daily and high-frequency data sets. The projections were based on time series modeling and regression modeling. It was also used implied volatility, calculated from options traded on Telemar PN, as an explanatory variable. The conclusion of the study was that, while the use of time series models with high frequency data yields good results for short projection periods, it does not generate good results for longer projection periods. The model loses predictability power as the projection horizon increases.

Índice

1	INTRODUÇÃO	9
1.1	MOTIVAÇÃO E OBJETIVO	10
2	CONCEITOS BÁSICOS DE VOLATILIDADE E REVISÃO BIBLIOGRÁFICA	11
2.1	CONCEITOS BÁSICOS DE VOLATILIDADE.....	11
2.2	REVISÃO BIBLIOGRÁFICA	17
3	DESCRIÇÃO DOS DADOS, TRATAMENTO E AMOSTRAGEM	22
3.1	<i>DESCRIÇÃO DOS DADOS</i>	22
3.1.1	<i>Tratamento dos Dados</i>	23
3.2	<i>CÁLCULO DE VOLATILIDADE</i>	26
3.2.1	<i>Cálculo da Série de Volatilidade Implícita de TNLP4 de Vinte e Um Dias Úteis de Vencimento</i> 27	
4	FATOS ESTILIZADOS	28
4.1	SUMÁRIO ESTATÍSTICO, ANÁLISE DE ESTACIONARIEDADE DOS RETORNOS, TESTE DE NORMALIDADE	29
4.2	PERSISTÊNCIA DE VOLATILIDADE OU MEMÓRIA LONGA.....	38
4.3	ANÁLISE DO TERMO DE AUTOCORRELAÇÃO NO CÁLCULO DE VOLATILIDADE REALIZADA PARA DADOS DE ALTA FREQUÊNCIA	40
4.4	ANÁLISE DE SAZONALIDADE INTRADIÁRIA	45
4.5	EFEITO DE ALAVANCAGEM.....	51
5	AJUSTE DE MODELOS DE VOLATILIDADE	52
5.1	AJUSTE DE MODELOS DE SÉRIES TEMPORAIS DA FAMÍLIA GARCH PARA DADOS DIÁRIOS	52
5.2	AJUSTE DE MODELO DA FAMÍLIA GARCH PARA SÉRIES INTRADIÁRIAS.....	56
5.3	AJUSTE DE MODELO DE REGRESSÃO PARA VOLATILIDADE REALIZADA UTILIZANDO-SE	

VOLATILIDADE IMPLÍCITA COMO VARIÁVEL EXPLICATIVA	96
6 ESTIMAÇÃO DE VOLATILIDADE REALIZADA E AVALIAÇÃO DE ACURÁCIA DO MODELO	98
6.1 ESTIMAÇÃO DE VOLATILIDADE REALIZADA DE UM DIA ÚTIL E DE VINTE E UM DIAS ÚTEIS UTILIZANDO-SE DADOS DIÁRIOS	99
6.2 ESTIMAÇÃO DE VOLATILIDADE REALIZADA DE UM DIA ÚTIL E DE VINTE E UM DIAS ÚTEIS UTILIZANDO-SE DADOS INTRADIÁRIOS DE FREQUÊNCIA DE DUAS HORAS.....	100
6.3 ESTIMAÇÃO DE VOLATILIDADE REALIZADA DE VINTE E UM DIAS ÚTEIS UTILIZANDO-SE MODELO DE REGRESSÃO COM VOLATILIDADE IMPLÍCITA COMO VARIÁVEL EXPLICATIVA.....	104
7 CONCLUSÕES	105
8 BIBLIOGRAFIA.....	106

1 INTRODUÇÃO

Em finanças o conceito de risco é associado ao desvio padrão (σ) ou à variância (σ^2). Esse conceito está relacionado com o segundo momento de distribuições. No caso contínuo, σ é um parâmetro de escala que amplifica ou reduz o tamanho das flutuações geradas por um processo Wiener padrão. Cada processo terá comportamento bastante distinto e flutuações diferentes dependendo de suas características. Desta forma, o uso do desvio padrão (σ) ou de variância (σ^2) como medida de risco só tem sentido quando associado a uma determinada distribuição, quando serão obtidas as funções densidade de probabilidade e densidade de probabilidade acumulada

O uso de dados de alta frequência no cálculo de volatilidade foi difundido possivelmente após o trabalho de Merton (1980), quando indicou que o uso de dados de alta frequência não adicionaria valor no cálculo de médias, mas sim no cálculo da variância. A intuição por trás dessa afirmação está no fato de que, para um processo aleatório, o dado inicial e o dado final já apresentam informação necessária para uma estatística completa sobre a média, enquanto que para a variância é importante obter dados referentes a todos os incrementos que constam da série de dados compreendida entre o dado inicial e o dado final. Esse fato tem como implicação que o uso de dados de alta frequência possui, na maioria das situações, informação valiosa para o cálculo da variância ou desvio padrão, adicionalmente aquela informação obtida utilizando-se dados de menor frequência como dados diários.

O uso de dados de alta frequência pode contribuir para o aumento da precisão do cálculo da variância ou desvio padrão. Ao mesmo tempo o aumento da frequência de dados gera fenômenos

novos que devem ser estudados e considerados na estimação de volatilidade. Fenômenos como sazonalidade e auto-correlação podem gerar resultados distorcidos em lugar de aumentar a precisão quando não são devidamente considerados. O procedimento nesses casos é fazer um tratamento na série intra-diária e, em seguida, identificar o modelo de volatilidade mais adequado que incorpore a informação valiosa que consta desses dados de alta frequência minimizando os efeitos indesejáveis.

1.1 MOTIVAÇÃO E OBJETIVO

O objetivo dessa dissertação é adicionar ao trabalho já desenvolvido por autores que aplicaram técnicas estatísticas de alta frequência para o cálculo de volatilidade de ativos brasileiros as seguintes perspectivas:

- Usar Telemar PN (tnlp4) como ativo por ser a ação mais líquida do mercado brasileiro;
- Avaliar diferentes frequências de dados intra-diários com relação aos fatores estilizados, incluindo a análise de características ainda pouco estudadas como o chamado fator de "bias" gerado pelo termo de auto-correlação;
- Ajustar modelos GARCH para dados diários, como também para dados de alta frequência, para modelar a volatilidade realizada de um dia útil e de vinte e um dias úteis;
- Ajustar modelo de regressão para dados de alta frequência envolvendo volatilidade realizada e volatilidade implícita de opções de compra de TNLP4 para prever volatilidade realizada de um dia útil e de vinte e um dias úteis.

2 CONCEITOS BÁSICOS DE VOLATILIDADE E REVISÃO

BIBLIOGRÁFICA

Este capítulo é dedicado a explicar alguns fundamentos básicos necessários para o entendimento dos modelos econométricos de projeção de volatilidade e aspectos relativos à modelagem de volatilidade com uso de dados intra-diários.

Também é feita uma breve revisão bibliográfica onde são expostos alguns conceitos importantes na modelagem de volatilidade, sobretudo aqueles relacionados ao uso de dados intra-diários.

2.1 CONCEITOS BÁSICOS DE VOLATILIDADE

Utiliza-se como medida de volatilidade o desvio padrão ou variância σ^2 , calculados a partir de N observações de retorno r_1, \dots, r_n : \bar{r}

$$\sigma^2 = \frac{1}{N-1} \sum_1^N (r_i - \bar{r})^2, \text{ onde } \bar{r} \text{ é o retorno médio.}$$

Para o caso contínuo assume-se que os retornos são gerados a partir da distribuição contínua de preços p_t :

$$dp_t = \sigma_t dW_{p,t},$$

Em que os retornos continuamente compostos são dados por: $r_t = \ln(p_t) - \ln(p_{t-1})$.

Caso os retornos sejam não auto-correlacionados e o processo contínuo, tem-se a partir do conceito de variação quadrática (Karatzas e Shreve, 1988) que:

$$p \lim_{m \rightarrow \infty} \left(\int_0^1 \sigma_{t+\tau}^2 d\tau - \sum_{j=1}^m r_{m,t+j/m}^2 \right) = 0,$$

Ou seja, a volatilidade pode ser calculada a partir do somatório dos retornos ao quadrado desde que esse cálculo seja feito em frequências altas.

Volatilidade Realizada. Davidian e Carroll (1987) mostram que o uso do somatório dos retornos ao quadrado para modelar a variância é apropriado somente para distribuições normais. O termo volatilidade realizada foi usado a partir dos artigos de Fung e Hsieh (1991) e Eersen e Bollerslev (1998) e significa essa soma dos quadrados dos retornos para intervalos pequenos (alta frequência). Eersen, Bollerslev, Diebold e Labys (2001) constataram que a volatilidade realizada calculada a partir de dados de alta frequência é aproximadamente lognormal. Esse cálculo é aplicável desde que a frequência não seja muito alta de forma a evitar efeitos da microestrutura dos processos e auto-correlação.

A Volatilidade Implícita. Sabe-se que para um dado ativo há apenas uma medida de volatilidade. Entretanto é fato que opções com características contratuais semelhantes, de um mesmo vencimento, mas de preços de exercício distintos, possuem volatilidades implícitas diferentes. O conceito de volatilidade implícita é bastante usado em finanças e trata-se da volatilidade calculada a partir do modelo de opções de Black-Scholes. Por ser calculada a partir do preço da opção e do ativo subjacente de forma sincronizada, a volatilidade implícita está sempre associada a uma

determinada opção, que tem como características importantes o prazo de vencimento e a medida de quanto “perto do dinheiro” está a opção, ou seja, uma combinação do chamado valor intrínseco (diferencial entre o preço de exercício da opção e o preço do ativo subjacente) e o valor temporal da opção.

Sabe-se que a volatilidade implícita não é constante para preços de exercícios distintos e tão pouco é constante para prazos de vencimentos distintos, ou seja, ela varia em duas dimensões. Essa relação não linear entre a volatilidade implícita e o preço de exercício é chamada de sorriso ou “smile”.

A teoria de apreamento de opções apresentada por Black-Scholes requer que o preço do ativo subjacente siga uma distribuição lognormal ou que o logaritmo do retorno do ativo siga uma distribuição normal. Sabe-se que isso não é verdade e que de fato os ativos possuem distribuições leptocurticas.

Para preços de exercício altos, uma opção de compra estaria bastante fora do dinheiro e a probabilidade de exercício seria baixa; entretanto, uma distribuição leptocurtica à direita apresenta maior probabilidade de exercício do que uma distribuição normal. Essa maior probabilidade de exercício leva a um maior valor para o preço da opção de compra e a uma maior volatilidade implícita para um preço de exercício maior. Quando o preço de exercício é baixo, uma opção dentro do dinheiro possui um valor intrínseco alto e um valor de temporal baixo (reflete a incerteza relativa ao tempo que falta para a opção vencer e a oscilação possível do preço do ativo em relação ao preço de exercício durante esse período). Também por conta da cauda grossa à esquerda, a probabilidade da opção terminar dentro do dinheiro é maior do que aquela de uma distribuição normal resultando em um preço maior e uma volatilidade implícita maior para essa opção. Esse fenômeno implica em se ter uma maior volatilidade implícita para opções de preços de exercícios

bem altos e bem baixos quando comparadas à volatilidade implícita das opções com preço de exercício próximo ao preço do ativo ("at the money"). Essa análise aplica-se a ativos que possuem distribuições com caudas grossas e aproximadamente simétricas como em geral ocorre com taxa de juros e moedas, sobretudo aquelas moedas com longo histórico no regime flutuante e sem restrições de conversibilidade.

No caso de ações verifica-se o fenômeno da assimetria em que há apenas a metade do sorriso (alta volatilidade implícita para preços de exercício baixos e a volatilidade implícita decai para preços de exercício altos) em que há uma tendência de maiores oscilações de preços quando o mercado cai, ou seja, a distribuição tem uma assimetria para a esquerda, apresentando uma cauda grossa à esquerda e uma cauda mais fina que a distribuição normal à direita. No caso da taxa de câmbio real/usd, ao menos até os dias atuais, esse fenômeno se dá com uma cauda mais grossa à direita (sentido de valorização do dólar contra o real ou de aumento da taxa de câmbio), ou seja, baixa volatilidade implícita para preço de exercício baixo e alta volatilidade implícita para preço de exercício alto.

Considerando-se que a volatilidade é uma variável aleatória, há também a influência da relação entre o processo do ativo subjacente e o processo da volatilidade, ou seja, o efeito da correlação entre o processo da volatilidade e o processo do preço do ativo. Isso significa que, quando essa correlação é negativa, retornos negativos geram aumento de volatilidade contribuindo para caudas grossas à esquerda enquanto retornos positivos contribuem para diminuir a volatilidade gerando caudas mais finas à direita. O inverso ocorreria caso a correlação seja positiva.

Há investidores que negociam opções em diversos mercados como ações, câmbio, juros e commodities com o objetivo de auferir ganhos a partir de uma visão de que a volatilidade realizada do ativo será menor/menor do que aquela sugerida pela opção ou em outros casos de que o

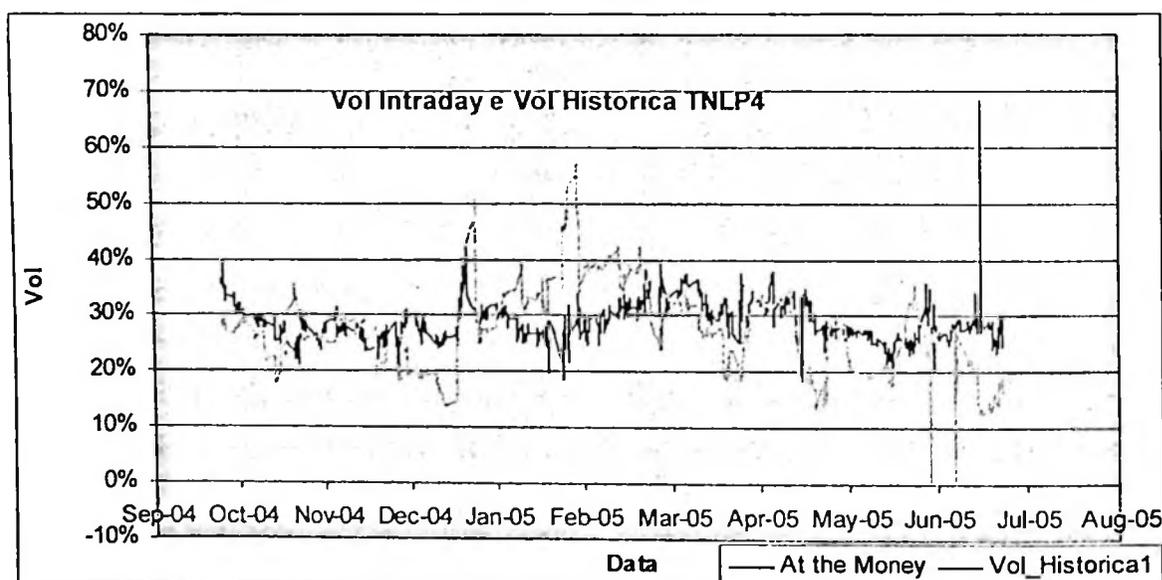
mercado irá apreciar/depreciar a opção comprada/vendida pelo investidor a partir do aumento/queda da volatilidade implícita. Essas operações são em geral executadas conjuntamente com operações de cobertura ou "hedge" em que o investidor não fica exposto à variação do preço do ativo subjacente, mas sim as derivadas de ordem superior (gamma) e/ou a efeitos gerados pela variação de volatilidade (vega), tempo (rho), etc.

Em geral a volatilidade realizada intra-diária oscila bem mais do que a volatilidade implícita intra-diária na medida em que os agentes levam algum tempo para processar as informações e definir seus preços de compra/venda de volatilidade. O apreçamento pelos investidores das opções a partir da definição da volatilidade implícita em seus modelos leva em consideração a volatilidade histórica (realizada) calculada para um prazo semelhante aquele do vencimento da opção, bem como a expectativa do investidor em relação à volatilidade futura.

A volatilidade implícita calculada a partir de opções negociadas no mercado pode ajudar a prever a volatilidade realizada. Por outro lado, há uma série de limitações no uso da volatilidade implícita como variável explicativa de um modelo de previsão de volatilidade realizada na medida em que há distorções geradas pela assimetria (volatilidade implícita distinta para preços de exercícios distintos) e pelo tempo de vencimento da opção sendo analisada (volatilidade implícita distinta para vencimentos distintos). Há também um bias positivo que aumenta o valor da volatilidade implícita em relação à volatilidade realizada de mesmo vencimento, sobretudo nos momentos de maior oscilação do mercado, por conta da tendência natural dos bancos e corretoras ("dealers") em usar um preço maior para aumentar seu lucro; esse fenômeno é tanto maior quanto menor for a liquidez das opções e tende a ser maior em mercados de balcão quando comparado aos mercados em bolsa de valores/derivativos.

Sabe-se que o diferencial entre a média da volatilidade realizada e a média da volatilidade implícita para um determinado prazo tende a ser pequeno, ou seja, o primeiro momento das distribuições dessas duas medidas de volatilidade tende a convergir para valores próximos. Além disso, o valor observado para essas duas medidas de volatilidade tende a estar mais próximo quando a opção ganha liquidez e aproxima-se do seu vencimento. Cabe entretanto ressaltar que esse diferencial ao longo do prazo que falta para vencer a opção pode variar bastante. Em geral, para dados de alta frequência, o desvio padrão da série de volatilidade realizada é bem superior aquele da série de volatilidade implícita, sobretudo para séries com poucos dados (alguns meses).

O diferencial entre volatilidade implícita de opções de compra de TNLP4 “at-the-money” e volatilidade histórica de TNLP4 pode chegar a +/- 20% em dias de maior oscilação intra-diária. Esse fato pode ser visualizado no Gráfico abaixo em que é indicada a oscilação da volatilidade histórica realizada de TNLP4 vs volatilidade implícita de opção “at-the-money” de TNLP4.



2.2 REVISÃO BIBLIOGRÁFICA

O estudo de volatilidade utilizando dados de alta frequência difundiu-se a partir da segunda metade da década de 90, sobretudo após os trabalhos de alguns autores como Eersen, Bollerslev e Lange (1999), Eersen e Bollerslev (1997, 1998, 1998-b, 1999), Eersen (2000), Taylor e Xu (1997), Olsen & Associates (diversos). Esses trabalhos indicaram que a precisão de modelos de projeção de volatilidade era tanto melhor quanto maior a frequência dos dados utilizada relativamente ao horizonte (período) para o qual se deseja projetar a volatilidade. Isso quer dizer que se perde eficácia quando o período para o qual se deseja projetar volatilidade cresce em relação à frequência dos dados. Uma das questões mais analisadas nesses trabalhos foi a presença de características intrínsecas aos dados de alta frequência que impactam de forma decisiva o comportamento desses dados e suas volatilidades. Conseqüentemente o estudo de dados de alta frequência oferece importantes subsídios para estimar a volatilidade que antes não eram considerados quando se utilizavam apenas dados diários.

Há características peculiares às séries financeiras em geral e outras características peculiares às séries financeiras de alta frequência. Essas características são conhecidas como fatos estilizados e serão analisadas em uma seção dedicada a esse assunto. Os modelos de previsão são tão mais adequados quanto maior o número de fatos estilizados que conseguem reproduzir.

Há uma série de modelos usados na previsão de volatilidade. Os principais grupos de modelos de previsão de volatilidade são:

1. Modelos que utilizam séries temporais

- Modelos que utilizam cálculos baseados no desvio padrão histórico. Exemplos desses

modelos são os modelos de média histórica, de média móvel exponencial (EWMA), etc.

- Modelos baseados em variância condicional, ou seja, na previsão para o instante t a partir do que se sabe no instante $t-1$. Essas são as classes de modelos ARCH, GARCH e suas variantes. Utilizam-se da função de verossimilhança para estimar volatilidade.
 - Modelos baseados em variância condicional que incorporam persistência da volatilidade. Sabe-se que um choque na volatilidade tem impacto prolongado, gerando efeitos de persistência chamados de memória longa. Há modelos baseados em variância condicional que incorporam esse fato estilizado como modelos FIGARCH, FIEGARCH.
2. Modelos de previsão de volatilidade a partir da volatilidade implícita obtida através de opções negociadas no mercado.
- Modelos que utilizam volatilidade calculada a partir da equação de Black-Scholes.
 - Modelos que tratam a volatilidade como variável estocástica, ou seja, consideram o processo do ativo e o processo da volatilidade.

Autores como Shu (2000) e Poteshman (2000) analisaram métodos de projeção de volatilidade realizada utilizando dados de alta frequência para o cálculo da volatilidade realizada e da volatilidade histórica, modelos que utilizam volatilidade implícita calculada por Black-Scholes e, alternativamente, modelos que utilizam volatilidade estocástica como o modelo de Heston. Concluíram que a utilização de dados de alta frequência melhora a precisão dos modelos enquanto que a utilização de um modelo com volatilidade estocástica em substituição ao modelo de Black-Scholes não melhora

significativamente a performance do modelo.

Os modelos de previsão de volatilidade que utilizam a volatilidade implícita usam em geral volatilidades calculadas a partir de opções com preço de exercício próximo ao preço atual do ativo (“at-the-money”) de forma a minimizar as distorções relacionadas com o nível de volatilidade implícita para diferentes preços de exercício. Uma série de autores como Feinstein (1989), Ederington e Guan (2000) concluíram que os modelos que compõem opções um pouco dentro do dinheiro com opções um pouco fora do dinheiro, combinando também opções de compra e opções de venda, para um prazo fixo, geram os melhores resultados obtendo-se volatilidades implícitas “at-the-money”, para o prazo em questão, sinteticamente. Alternativamente, o uso de opções “at-the-money” também gera bons resultados.

3. Modelos de previsão de volatilidade combinando-se previsões que utilizam séries temporais (item 1) com volatilidade implícita de opções (item 2).

Martens e Zein (2003) e Day e Lewis (1992) analisam modelos da família GARCH em que a volatilidade implícita é uma variável explicativa e concluem que melhora a precisão do modelo quando comparado a modelos da família GARCH sem essa variável exógena:

$$r_t = \mu + \varepsilon_t,$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma Vol_{implícita}.$$

4. Modelos de regressões simples de previsão da volatilidade realizada a partir da volatilidade

histórica e da volatilidade implícita.

Segundo Poteshman (2000) esses modelos envolvem regressões que têm a seguinte forma:

$$Vol_{realiz}(t) = \alpha + \beta Vol_{implícita}(t) + \varepsilon(t),$$

ou

$$Vol_{realiz}(t) = \alpha + \beta Vol_{implícita}(t) + \gamma Vol_{Historica}(t) + \varepsilon(t),$$

Onde $Vol_{realiz}(t)$ é a medida de volatilidade realizada projetada para o prazo (t), $Vol_{Historica}(t)$ é a medida de volatilidade, calculada de forma semelhante à volatilidade realizada, em um período histórico de prazo t. $Vol_{implícita}(t)$ é a volatilidade calculada a partir dos preços das opções de compra ATM de prazo de vencimento t, utilizando-se a metodologia de Black-Scholes. Essas três medidas de volatilidade estão na mesma escala.

5. Modelos de Volatilidade Estocástica

Esses modelos supõem que a volatilidade é um processo estocástico. São bem mais complexos, utilizam técnicas como quase máxima verossimilhança (QMLE), momentos generalizados (GMM), integração numérica ou outros métodos para estimar volatilidades. Há diversos autores como Harvey, Ruiz e Shepard (1994), Singleton (2001), Durbin e Koopman (2000) que analisaram esses modelos.

Na presente dissertação iremos nos limitar a análise de modelos GARCH de séries temporais baseadas em dados diários e de alta frequência, bem como à análise de modelos de regressão que relacionam a previsão de volatilidade realizada com informação obtida a partir de série de volatilidade implícita de opção de prazo semelhante.

3 DESCRIÇÃO DOS DADOS, TRATAMENTO E AMOSTRAGEM

Este capítulo é dedicado a descrever os dados brutos utilizados, bem como o tratamento para prepará-los para o ajuste e projeção de volatilidade. Utilizaram-se dados diários e intra-diários de TNLP4, bem como dados intra-diários de volatilidade implícita, calculados a partir de opções de compra de TNLP4. Utilizou-se o período de 02 de Janeiro de 2002 a 31 de Março de 2005 (798 dias úteis) para o estudo de fatos estilizados, tanto para os dados diários quanto para os dados intra-diários. Os modelos foram ajustados para a série de dados de 02 de Janeiro de 2002 a 30 de Dezembro de 2004 e as projeções foram realizadas para o período de 02 de Janeiro de 2005 a 30 de Junho de 2006.

3.1 *DESCRIÇÃO DOS DADOS*

Foram utilizados os seguintes dados brutos:

- Dados brutos de preço de Telemar PN (TNLP 4) com freqüência intra-diária: essa base de dados é composta por todas as cotações registradas durante o horário de pregão, ou seja, das 10:00 hs às 17:00 hs horário normal e das 11:00 hs às 18:00 hs durante horário de verão, no período de 02 de Janeiro de 2002 a 30 de Junho de 2005. Essa série de dados inicial é irregularmente espaçada pois contém todos os negócios ocorridos com Telemar PN.
- Dados brutos de Telemar PN (TNLP4) com freqüência diária, do período de 02 de Janeiro de 2002 a 30 de Junho de 2005, excluindo-se alguns

dias seguindo o mesmo critério usado para a série de dados intra-diários (ver abaixo), perfazendo um total de 860 dados diários. Capturaram-se preços de abertura, fechamento, máximo e mínimo para o período indicado.

- Dados brutos de opções de compra em Telemar PN (TNLP4), com preço de exercício "at-the-money", para os dois vencimentos com liquidez, com frequência de 5 minutos, de 19 de Outubro de 2004 a 30 de Junho de 2005, a partir do qual foram calculadas as volatilidades implícitas a cada 5 minutos para opções "at the money" com vencimento em 21 dias úteis.

3.1.1 Tratamento dos Dados

A partir da série inicial intra-diária de preços de TNLP4 irregularmente espaçada, passou-se à análise do melhor método para construir séries regularmente espaçadas com as frequências desejadas. Segundo Dacorogna, Olsen et al¹, pode-se aplicar interpolação linear ou usar o valor da cotação mais recente para preencher o espaço entre duas transações de forma a construir uma série homogênea. Segundo esses autores, a diferença entre os métodos é pequena desde que não se trate de preencher um espaço muito grande sem dados.

Segundo Corsi, et al. (2001), a utilização de interpolação linear para converter a série inicial, irregularmente espaçada e não homogênea, em uma série homogênea na frequência desejada tem mais desvantagens do que o método que utiliza a cotação mais recente. Os autores afirmam que quando duas cotações estão inicialmente separadas por um intervalo de tempo maior do que a janela de frequência que se deseja utilizar, o método de interpolação linear assume implicitamente

¹ Dacorogna M, Olsen R. et Al, An Introduction to High-Frequency Finance, Cap. 3

um nível de volatilidade mínima nesse intervalo e introduz uma correlação artificial entre retornos subsequentes da nova série regularmente espaçada. Essa correlação artificial é tanto maior quanto maior for a ocorrência de intervalos em que duas cotações estão inicialmente separadas por um intervalo de tempo maior do que a janela de frequência que se deseja utilizar.

Decidiu-se construir as séries de frequências distintas com dados homoganeamente espaçados a partir da série de dados irregularmente espaçados inicial, utilizando-se os dados mais próximos da janela de frequência desejada.

Os dados de alta frequência apresentavam ainda os seguintes problemas:

- As cotações não se encontravam regularmente distribuídas em cada dia devido aos horários distintos de pregão ao longo do ano;
- Houve dias em que a bolsa operou em expediente reduzido: vésperas de ano novo e natal, 4ª feira de cinzas e dias em que houve paralisação por problemas técnicos.
- A base também apresentou alguma inconsistência em alguns dias de cada ano.

Foram feitos os seguintes ajustes nos dados:

1. Exclusão de todas as observações de dias de expediente reduzido ou que apresentaram dados incompletos ou incorreção dos dados. Os seguintes dias foram excluídos:

- 13 de Fevereiro de 2002
- 27 de Março de 2002
- 17 de Junho de 2002
- 26 de Junho de 2002

- 05 de Março de 2003
- 12 de Março de 2003
- 25 de Fevereiro de 2004
- 30 de Agosto de 2004
- 25 de Janeiro de 2005
- 31 de Janeiro de 2005
- 09 de Fevereiro de 2005

2. Construíram-se bases de dados de retornos intra-diários, calculados a partir da diferença dos logaritmos das cotações registradas seqüencialmente em intervalos da freqüência em questão. Não foram calculados retornos a partir da diferença entre a primeira cotação do dia e a última cotação do dia anterior de modo a não incorporar a oscilação de um dia para outro ("overnight"). Os retornos foram calculados para diversas freqüências, todas construídas a partir da série inicial, transformando-as em séries homogêneas, ou seja, com o mesmo número de dados para cada dia.
3. Foram selecionadas as seguintes freqüências para estudo dos fatos estilizados da série compreendida entre 02 de Janeiro de 2002 e 31 de Março de 2005 (798 dias úteis):
 - 5 minutos. Essa série totalizou 64.638 observações de retorno de 5 minutos com 81 observações/dia.
 - 15 minutos. Essa série totalizou 21.546 observações de retorno de 15 minutos com 27 observações/dia.

- 30 minutos. Essa série totalizou 10.374 observações de retorno de 30 minutos com 13 observações/dia.
- 1 hora. Essa série totalizou 4.788 observações de retorno de 1 hora com 6 observações/dia.
- 2 horas. Essa série totalizou 2.394 observações de retorno de 2 horas com 3 observações/dia.
- 3 horas. Essa série totalizou 1.596 observações de retorno de 3 horas com 2 observações/dia.

3.2 CÁLCULO DE VOLATILIDADE

Utilizaram-se as seguintes definições de volatilidade:

- Volatilidade Realizada, calculada a partir de dados de alta frequência, apenas com o uso de retorno intra-diário, ou seja, sem o uso do retorno overnight:

$$\sqrt{\sum_1^N r_{t,d}^2}$$

onde $r_{t,d}$ são os retornos calculados a partir de séries de preços intra-diários de TNLP4.

- Volatilidade Implícita calculada a partir das opções de compra de TNLP4, como indicado a seguir.

3.2.1 Cálculo da Série de Volatilidade Implícita de TNLP4 de 21 Dias Úteis de Vencimento

A partir de cotações intra-diárias de frequência de 5 minutos de preços de TNLP4 e do preço de opções de compra de TNLP4 de preço de exercício "at-the-money" (ATM) e, utilizando-se os dois vencimentos líquidos disponíveis, calculou-se para cada um desses instantes a volatilidade implícita da opção de exercício ATM para o prazo de vencimento de 21 dias úteis.

Foi determinado o par de vencimentos mais líquidos e mais próximos de 21 dias úteis, ou seja $T_i < 21 < T_{i+1}$, que abrange o período de 21 dias úteis, de modo a calcular a volatilidade implícita para 21 dias úteis a partir de interpolação linear:

$$\sqrt{\frac{(21 - T_i)\sigma_{i+1}^2 + (T_{i+1} - 21)\sigma_i^2}{T_{i+1} - T_i}}$$

4 FATOS ESTILIZADOS

Esse capítulo analisa os fatos estilizados das séries de retornos diários e intra-diários de TNLP4. Trata-se de identificar características estatísticas em séries de retornos de ativos financeiros. Analisam-se alguns desses fatos na série de retornos logarítmicos de TNLP4 para frequência diária e para algumas frequências intradiárias, no período de 02 de Janeiro de 2002 a 31 de Março de 2005.

Os principais fatos estilizados de séries econométricas são:

- Estacionariedade dos retornos e média próxima de zero.
- As distribuições são bem distintas da normal para frequências mais altas.
 - Cauda pesada e curtose. A curtose aumenta à medida que a frequência aumenta
 - Assimetria. Para frequências menores que 1 dia a assimetria é maior
- Termo de Auto-correlação. Incidência de forte auto-correlação para frequências mais altas.
- Persistência de volatilidade ou memória longa.
- Efeito Alavancagem. O impacto dos retornos na volatilidade é assimétrico, com um efeito mais acentuado para retornos negativos do que para retornos positivos de igual magnitude. Esse fato estilizado pode ser detectado visualmente através do Gráfico

comparativo entre as séries de preços e de quadrados dos retornos (volatilidade realizada). Busca-se avaliar esse efeito também através da análise de modelos GARCH que possuem alavancagem.

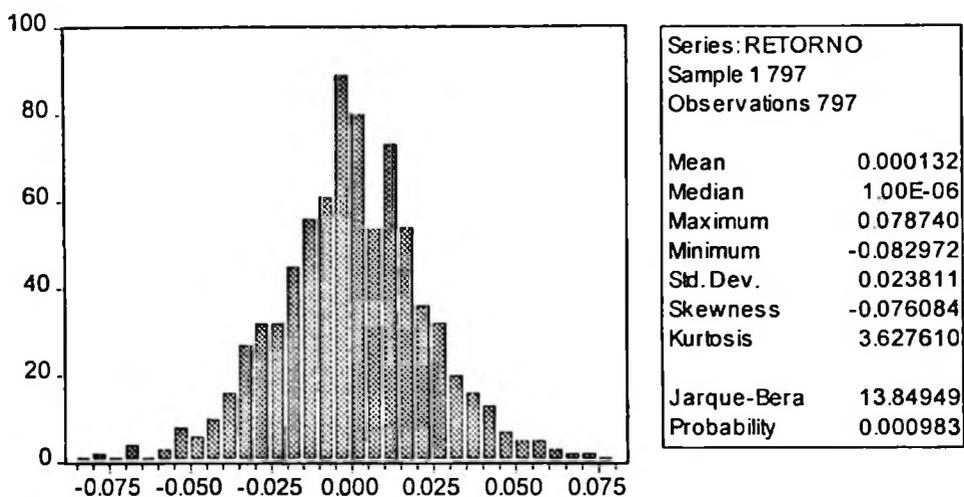
- Sazonalidade Intra-diária. A volatilidade dos retornos oscilações dependendo do horário.

4.1 SUMÁRIO ESTATÍSTICO, ANÁLISE DE ESTACIONARIEDADE DOS RETORNOS, TESTE DE NORMALIDADE

Sumário Estatístico

Série Diária. A análise do Gráfico 4.1.1 relativo ao sumário estatístico dos retornos diários indica que a distribuição possui características semelhantes à distribuição normal como, por exemplo, sua Curtose próxima de 3 e Assimetria próxima a zero.

Gráfico 4.1.1: Série de Retornos Diários



Séries Intra-diárias para Diversas Frequências

As distribuições de retorno intra-diárias apresentam Curtoses muito altas e características bem distintas da distribuição normal. A análise dos sumários estatísticos para diversas frequências (Gráficos 4.1.2 a 4.1.7), indica que, quanto maior a frequência, maior a Curtose (maior que 3) e mais negativa (diferente de zero) é a Assimetria.

Gráfico 4.1.2: Série de Retornos de Frequência de 5 minutos

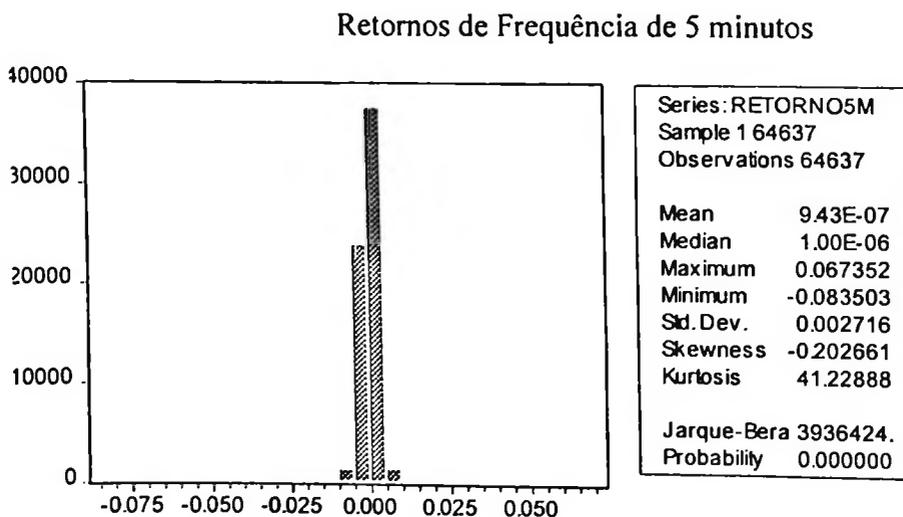


Gráfico 4.1.3: Série de Retornos de Frequência de 15 minutos

Retornos de Frequência de 15 minutos

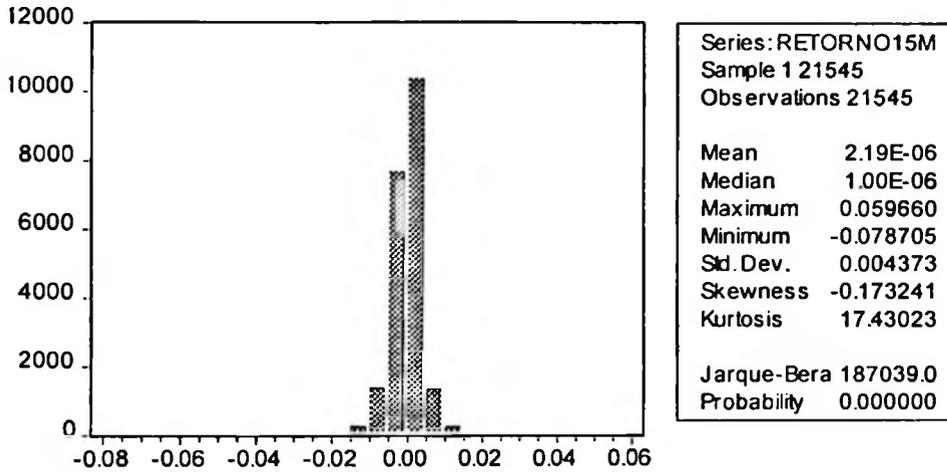


Gráfico 4.1.4: Série de Retornos de Frequência de 30 minutos

Retornos de Frequência de 30 minutos

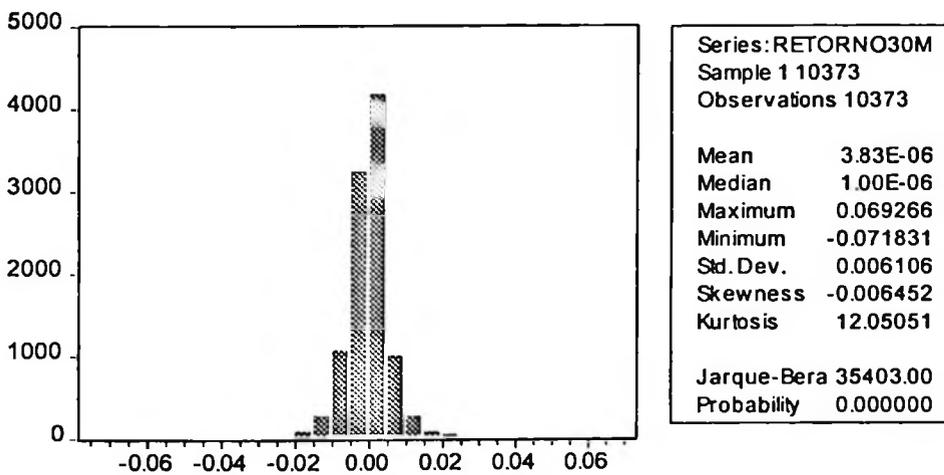


Gráfico 4.1.5: Série de Retornos de Frequência de 1 hora

Retornos de Frequência de 1 hora

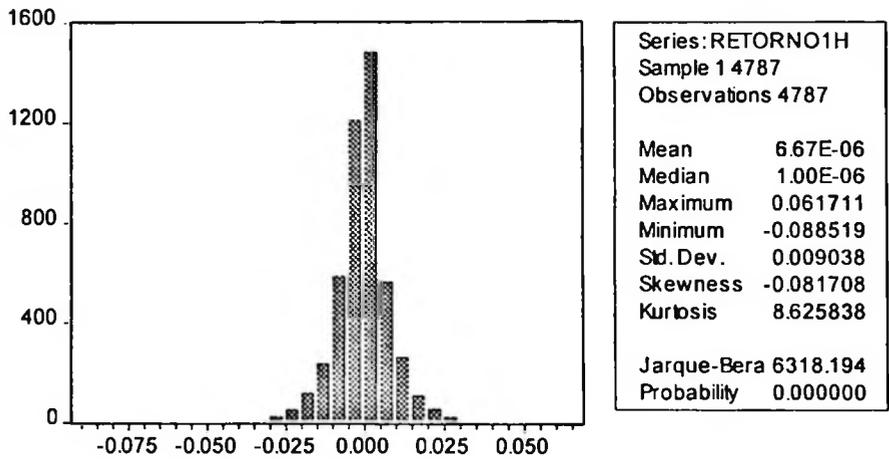


Gráfico 4.1.6: Série de Retornos de Frequência de 2 horas

Retornos de Frequência de 2 horas

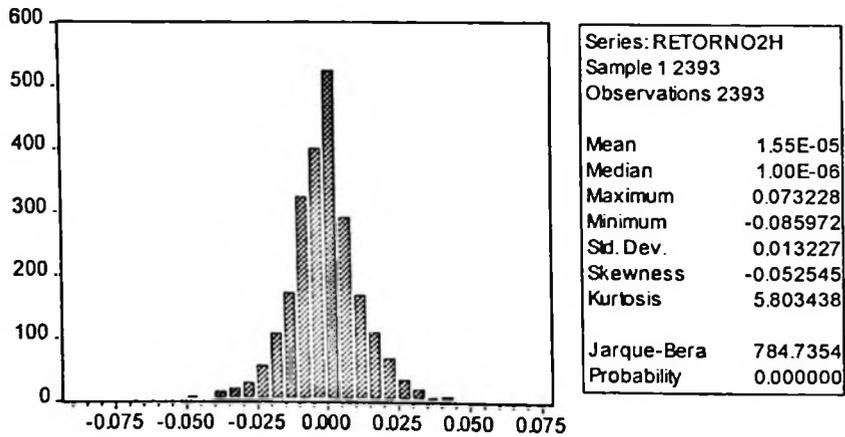
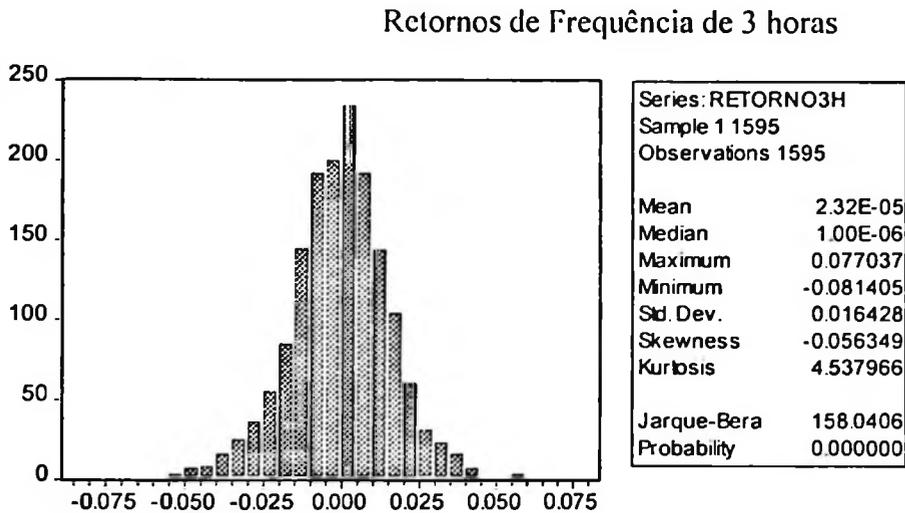


Gráfico 4.1.7: Série de Retornos de Frequência de 3 horas



Testes de Normalidade e Identificação de Auto-correlação

Série Diária. Pelo teste de Jarque-Bera, a hipótese nula de que a distribuição dos retornos diários é normal é rejeitada. O teste de Jarque-Bera (ver tabela 4.1.1) rejeitou a hipótese de normalidade e indicou a possibilidade da série ser ruído branco.

Tabela 4.1.1: Teste de normalidade para dados diários

```
> normalTest(Retorno,method="jb")
Test for Normality: Jarque-Bera
Null Hypothesis: data is normally distributed
Test Stat 14.8434
p.value 0.0006
Dist. under Null: chi-square with 2 degrees of freedom
```

Os testes de Q-estatística (Q) e Q-estatística (Q*) modificada indicam que o processo de ruído branco não pode ser descartado. Foi calculada a estatística Q (Box Pierce) e Q* (Ljung-Box/Box Pierce modificada) para verificar se os retornos diários de TNLP4 comportam-se como ruído branco. Os resultados para a série diária indicam que a hipótese nula de ruído branco não pode ser rejeitada.

Análise de Auto-correlação

O teste de auto-correlação (Ljung-Box), ver tabela 4.1.2, indicou não haver auto-correlação para os dados diários.

Tabela 4.1.2: Teste de auto-correlação

```
> autocorTest(Retorno,lag.n=4,method="lb")
Test for Autocorrelation: Ljung-Box
Null Hypothesis: no autocorrelation
Test Stat 4.6936 ; p.value 0.3202
>autocorTest(Retorno,lag.n=4,method="bp")
Test for Autocorrelation: Box-Pierce
Null Hypothesis: no autocorrelation
Test Stat 4.6632 ; p.value 0.3236
```

A análise da FAC indica tratar-se de ruído branco. A análise do Gráfico 4.1.8 (função de auto-correlação – ACF) indicou não haver auto-correlação. O Gráfico de auto-correlação parcial (Gráfico 4.1.9 – PACF) indicou que a PACF está dentro dos intervalos de significância para todos os lags.

Gráfico 4.1.8: Função de Auto-correlação para dados diários

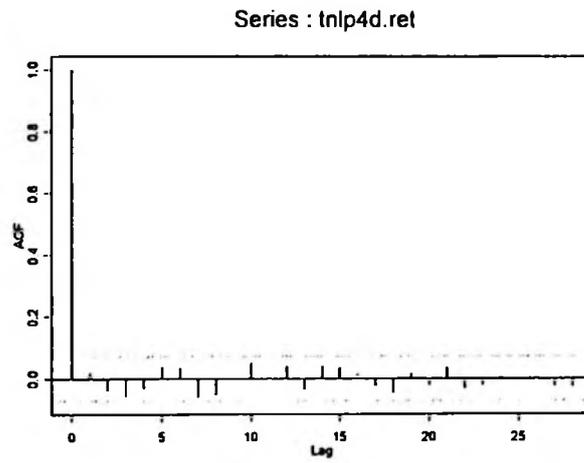
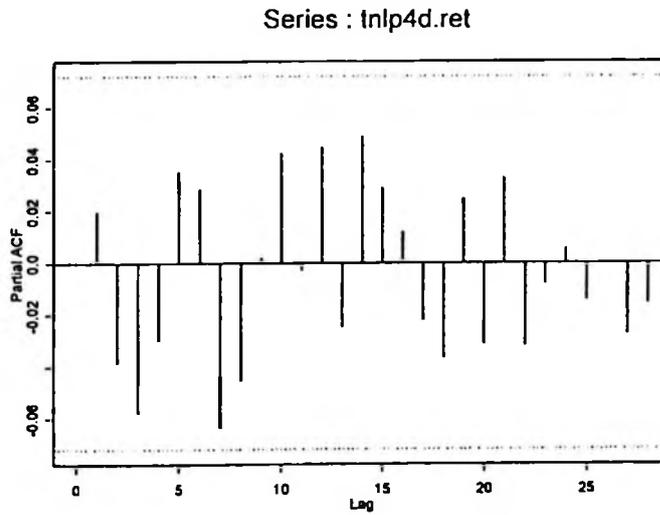


Gráfico 4.1.9: Função de Auto-correlação para dados diários



Séries Intra-diárias. O teste de Jarque-Bera indica que a hipótese nula de que a distribuição dos retornos intra-diários é normal também é rejeitada, como indicado na tabela 4.1.3 que contém as estatísticas Jarque-Bera e os p-valores para cada frequência.

Tabela 4.1.3: Teste Jarque-Bera para séries intra-diárias

Null Hypothesis: data is normally distributed

	Jarque – Bera
Estatística (série diária)	13,8495
P-valor (série diária)	0,0010
Estatística (série 5 min)	3936424
P-valor (série 5min)	0
Estatística (série 15 min)	187039
P-valor (série 15min)	0
Estatística (série 30 min)	35403
P-valor (série 30 min)	0
Estatística (série 1 h)	6318
P-valor (série 1 h)	0
Estatística (série 2 h)	784
P-valor (série 2 h)	0
Estatística (série 3 h)	158
P-valor (série 3 h)	0

Séries Intra-diárias. A análise da estatística Q e da estatística Q- modificada (Q*) para as séries intra-diárias indica que essas séries apresentam auto-correlação e não possuem comportamento de ruído branco. Ver resultados na tabela 4.1.4 abaixo com seus p-valores respectivos.

Tabela 4.1.4: Estatísticas Q e Q* para Séries Intra-diárias e P-valores

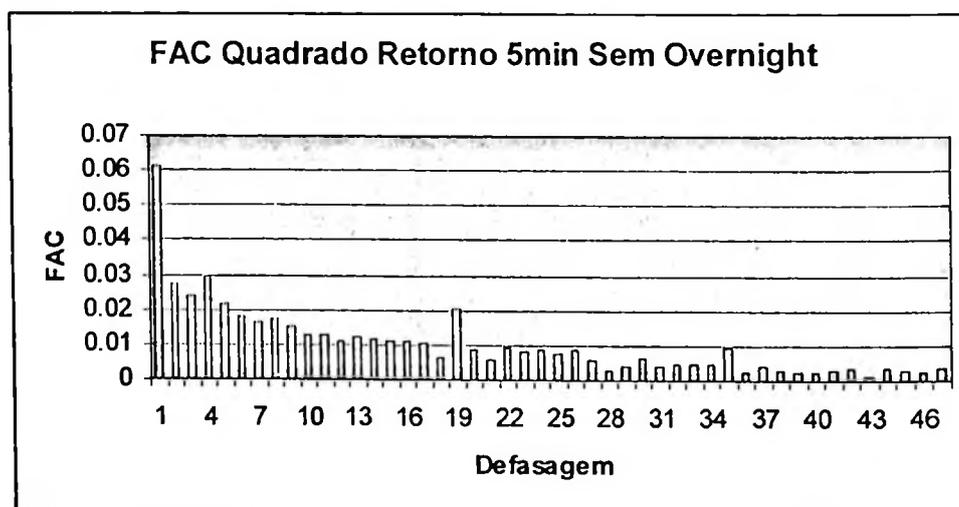
	Box-Pierce Modificado	Box-Pierce Modificado (lag 10 para diário e 100 para intra-diário)	Box-Pierce	Box-Pierce (lag 10 para diário e 100 para intra-diário)
Estatística (série diária)	28,570	15,601	28,090	15,450
P-valor (série diária)	0,480	0,111	0,512	0,116
Estatística (série 5 min)	NA	NA	623,040	713,860
P-valor (série 5 min)	NA	NA	0,000	0,000
Estatística (série 15 min)	113,990	221,300	113,890	220,840
P-valor (série 15 min)	0,00	0,000	0,000	0,000
Estatística (série 30 min)	122,070	201,62	121,860	200,860
P-valor (série 30 min)	0,00	0,000	0,000	0,000
Estatística (série 1h)	88,620	172,750	88,320	171,320
P-valor (série 1h)	0,00	0,000	0,000	0,000
Estatística (série 2h)	75,160	136,680	74,630	134,310
P-valor (série 2h)	0,00	0,000	0,000	0,012
Estatística (série 3h)	58,560	148,330	57,930	143,700
P-valor (série 3h)	0,0028	0,000	0,003	0,0028

4.2 PERSISTÊNCIA DE VOLATILIDADE OU MEMÓRIA LONGA

As séries de retornos e de volatilidade realizada (somatório de retornos ao quadrado) intra-diárias apresentam um comportamento distinto das séries diárias. As séries de volatilidade realizada para dados intra-diários possuem o que se chama de memória longa, ou seja, eventos que geram oscilações com efeitos que tendem a persistir por algum tempo. Há uma série de testes para identificar a existência de longa memória.

A persistência de volatilidade ou memória longa está muitas vezes ligada ao que se chama de aglomerados de volatilidade, ou seja, à alternância de períodos com concentração de alta e de baixa volatilidade. A análise do Gráfico da função de FAC das séries de retornos ao quadrado para a frequência de 5 minutos (Gráfico 4.2.1) indica a incidência de aglomerados de volatilidade.

Gráfico 4.2.1: FAC da série de retornos ao quadrado para frequência de 5 minutos



A estatística proposta por Geweke e Porter-Hudak (GPH) calculada a partir dos retornos absolutos intra-diários, foi calculada para os dados intra-diários (ver tabela 4.2.1) e indicou haver memória longa com nível de significância de 1% para as freqüências entre 5 minutos e 2 horas e com nível de significância de 5% para a freqüência de 3 horas. O parâmetro de longa memória (d) indicou valores na faixa de 95% de confiança para memória longa, ou seja, d no intervalo [0.337,0.578], para as freqüências de 5 minutos a 2 horas. Em resumo, quanto maior a freqüência, o fenômeno de persistência de volatilidade ou memória longa parece ser mais forte.

Tabela 4.2.1: Estatística GPH para série de retornos absolutos intra-diários

	5min	15 min	30 min	1 hora	2 horas	3 horas
GPH Test	10,634**	9,485**	7,743**	6,334**	3,540**	2,475*
Parâmetro de Longa Memória (d)	0,446	0,535	0,534	0,513	0,373	0,295

** nível de significância de 1% * nível de significância de 5%

4.3 ANÁLISE DO TERMO DE AUTO-CORRELAÇÃO NO CÁLCULO DE VOLATILIDADE REALIZADA PARA DADOS DE ALTA FREQUÊNCIA

Vários autores como Taylor e Xu (1997), Dacorogna et al. (1997) e Eersen et al.(1999) consideraram o cálculo de volatilidade a partir de dados de alta frequência. A volatilidade realizada é calculada a partir da soma cumulativa dos quadrados dos retornos intra-diários para uma determinada frequência. Esses autores demonstraram que esse somatório dos retornos ao quadrado a partir de dados de alta frequência pode ser usado para estimar a variância de uma variável aleatória i.i.d. Caso esses retornos tenham média zero e sejam descorrelacionados, o somatório dos retornos de alta frequência ao quadrado é um estimador não-viesado da variância diária σ^2 . Ainda segundo esses autores, somando-se um número suficiente de quadrados de retornos de alta frequência, é possível obter uma medida sem erros da volatilidade diária.

Esses autores indicam que qualquer definição de volatilidade realizada implica em utilizar dois parâmetros de tempo: T (o tamanho da janela de tempo móvel sobre a qual a volatilidade é calculada) e δt (o tempo relacionado com a frequência), tal que $T \gg \delta t$.

A volatilidade calculada sobre um intervalo T é definida como:

$$\sigma[T, \delta t](t) = \left[\frac{1}{n} \sum r^2(\delta t)(t) \right]^{1/2},$$

onde o retorno r tem valor esperado próximo de zero e é definido como:

$$r[\delta t](t) = \frac{Xrts(t) - Xrts(t - \delta t)}{\sqrt{\delta t / T}},$$

onde :

- X_{rts} é o log do preço da série homogênea obtida para uma determinada frequência a partir da série irregular inicial;
- $r[\delta t]$ é o retorno anualizado, observado em intervalos de tempo δt relativos à frequência escolhida ;
- $\sigma[T, \delta t]$: a volatilidade anualizada calculada a partir de $r[\delta t]$ (t) no período T;
- n: número de observações de retorno no intervalo T. No caso de intervalos igualmente espaçados $n = T / \delta t$.

Para uma distribuição aleatória Gaussiana, a magnitude do erro estatístico da estimativa de volatilidade pode ser calculada. Para esse modelo o somatório dos retornos ao quadrado tem uma distribuição qui-quadrado com grau de confiança igual ao número de termos na soma $\sigma^2[T, \delta t] \sim \chi^2$

$(T/\delta t) = \chi^2 (n)$. A medida de erro para uma distribuição χ^2 é dada por $\frac{\sigma^2 \sqrt{2n-1}}{n}$. Para uma

distribuição aleatória gaussiana, os valores usuais para o erro de medidas diárias de volatilidade com retornos calculados para distintos intervalos de tempo são:

- Erro (σ^2 [1d, 1d]) = 100% σ^2
- Erro (σ^2 [1d, 1h]) = 28,5% σ^2
- Erro (σ^2 [1d, 1min]) = 3,7% σ^2

Esses resultados indicam claramente a vantagem de calcular os retornos sobre intervalos cada vez menores (frequências mais altas) no cálculo da volatilidade realizada.

Autores como Corsi, Zumbach, Muller e Dacorogna (2001), indicam entretanto que, para frequências mais altas que poucas horas, essa definição de volatilidade realizada como estimador

não-viesado da variância diária é afetada por erros sistemáticos, indicando que essa definição não é um estimador não viesado da variância diária. Esses autores mostraram que o valor esperado da volatilidade calculada a partir de retornos em intervalos pequenos (alta frequência) não é igual à volatilidade obtida a partir de dados de retornos diários:

$$E[r^2[\delta t]] \neq E[r^2[1d]] \text{ e } E[\sigma^2[T, \delta t]] \neq E[\sigma^2[T, 1d]],$$

ou seja, a volatilidade realizada estimada a partir de dados de alta frequência é fortemente viesada quando comparada com a média dos retornos diários ao quadrado. Sendo assim, caso esse viés não seja considerado no estimador da volatilidade para dados de alta frequência, a estimativa de volatilidade diária a partir de tais dados irá conter severas distorções. O viés para dados de ações é negativo. Esse viés que distorce o estimador tende a ser relevante para frequências menores que poucas horas.

Dessa forma, se por um lado o uso de dados de alta frequência é recomendável para reduzir o erro do estimador, a microestrutura de mercado introduz um forte viés quanto maior for a frequência da série de dados, distorcendo assim os resultados e impactando negativamente a desempenho do estimador.

Esse viés que produz resultados anômalos e indesejáveis é explicado pela existência de autocorrelação $\rho(k) \neq 0$ entre os retornos para determinadas frequências, ou seja, $E[\sigma^2[T, \delta t]] \neq E[\sigma^2[T, 1d]]$ se $\rho(k) \neq 0$ para alguns lags k . Isso implica em que os retornos não são i.i.d. para essas frequências altas. A partir do retorno de frequências altas δt , o retorno em escalas de tempo maiores $\delta t' = m \delta t$ pode ser computado e as variâncias podem ser também calculadas com retornos na escala δt e $\delta t'$. Determina-se então uma expressão da variância em um período T a partir do seguinte cálculo em m intervalos de frequência δt :

$$\sigma^2[T, m\delta t] = \sigma^2[T, \delta t] * \left[1 + 2 \sum \left(1 - \frac{k}{m}\right) * \rho(k) \right],$$

onde $\rho(k)$ é a função de auto-correlação do retorno $r(\delta t)$ no momento $k(\delta t)$.

Nessa fórmula, quando :

- $\rho(k) = 0$, não há auto-correlação;
- $\rho(k) > 0$ há auto-correlação positiva e a variância final será menor do que aquela medida por $E[\sigma^2[T, 1d]]$ (viés negativo), ou seja $E[\sigma^2[T, \delta t]] < E[\sigma^2[T, 1d]]$;
- $\rho(k) < 0$ há auto-correlação negativa e a variância final será maior do que aquela medida por $E[\sigma^2[T, 1d]]$ (viés positivo), ou seja $E[\sigma^2[T, \delta t]] > E[\sigma^2[T, 1d]]$;

Espera-se uma auto-correlação positiva para ações e uma variância final menor do que aquela de um estimador não-viesado.

Os autores propõem um modelo para o processo do retorno dos ativos em alta frequência contendo um termo auto-regressivo para o retorno médio, obtendo uma estrutura de auto-covariância para o modelo relacionando o termo relativo ao preço real com uma componente de ruído relacionada com o viés.

Os autores também indicaram que o cálculo de um filtro de correção para a série de preços de modo a eliminar esse fator de viés implicaria, para o caso de ações, em usar um estimador de máxima verossimilhança com um modelo ARMA(p,q) o que seria demasiadamente trabalhoso. A alternativa então seria calcular o viés para diferentes frequências e optar pelas maiores frequências possíveis dado que esses fatores de viés têm baixa influência.

Análise do Termo de Auto-correlação Para Freqüências Intra-diárias

Fez-se o seguinte estudo de volatilidade realizada intra-diária para cada uma das freqüências de 5min, 15 min, 30min, 1 hora, 2 horas e 3 horas:

- Calculou-se a volatilidade realizada, anualizada, para períodos sucessivos de 21 dias úteis e anualizou-se o resultado. Calculou-se a média e o desvio padrão.
- Calculou-se a volatilidade realizada, anualizada, com fator de correção para o termo de auto-correlação para períodos sucessivos de 21 dias úteis. Calculou-se a média e o desvio padrão.
- Calculou-se o fator de auto-correlação para períodos sucessivos de 21 dias úteis. Calculou-se a média e o desvio padrão.

Os termos de auto-correlação para a freqüência de 5 minutos foram maiores sendo que o diferencial entre a volatilidade de 21 dias calculada com o termo de auto-correlação e a volatilidade de 21 dias calculada sem o termo de auto-correlação oscilou entre -15 pontos de volatilidade e + 11 pontos de volatilidade. Quanto menor a freqüência, menor o termo de auto-correlação. O diferencial entre a volatilidade de 21 dias calculada com o termo de auto-correlação e a volatilidade de 21 dias calculada sem o termo de auto-correlação oscilou bem menos para freqüências menores, como 2 horas.

A média das volatilidades de 21 dias para períodos longos como esse em questão tende a apresentar valores semelhantes tanto para a volatilidade com a correção do termo de auto-correlação como para a volatilidade sem essa correção (ver tabela 4.3.1 abaixo) .

Tabela 4.3.1: Média da Volatilidade Realizada com Termo de Auto-correlação vs Média da Volatilidade Realizada

	5min	15 min	30 min	1 hora	2 horas	3 horas
Média da Volatilidade Realizada com Correção do fator de Auto-correlação	37,720%	37,690%	37,651%	37,500%	37,671%	37,320%
Média da Volatilidade Realizada	37,942%	37,710%	37,520%	35,680%	35,282%	35,260%

4.4 ANÁLISE DE SAZONALIDADE INTRA-DIÁRIA

Há duas formas de padrão sazonal intra-diário comuns em dados de alta frequência: (i) Em forma de U, no qual os intervalos iniciais e finais do dia apresentam maior oscilação; (ii) O chamado modelo de J refletido, ou seja, em que a volatilidade maior ocorre na abertura, decaindo durante o meio dia e voltando a subir de forma menos intensa no fechamento. Tanto o volume de negociações, como a volatilidade do preço e o spread entre as cotações de compra/venda podem apresentar esse comportamento, ou seja, atingem seu ponto máximo na abertura, caem de patamar durante o meio dia e voltam a subir no fechamento para patamares menores do que aqueles de abertura. O padrão de J refletido para a oscilação do preço de TNLP4 foi identificado nas séries intra-diárias. A incorporação da sazonalidade na especificação do modelo pode ajudar a eliminar

distorções atribuíveis à microestrutura dos dados intra-diários e ajudar a obter modelos de previsão melhores.

Cálculo do Índice de Sazonalidade para Retornos Intra-diários

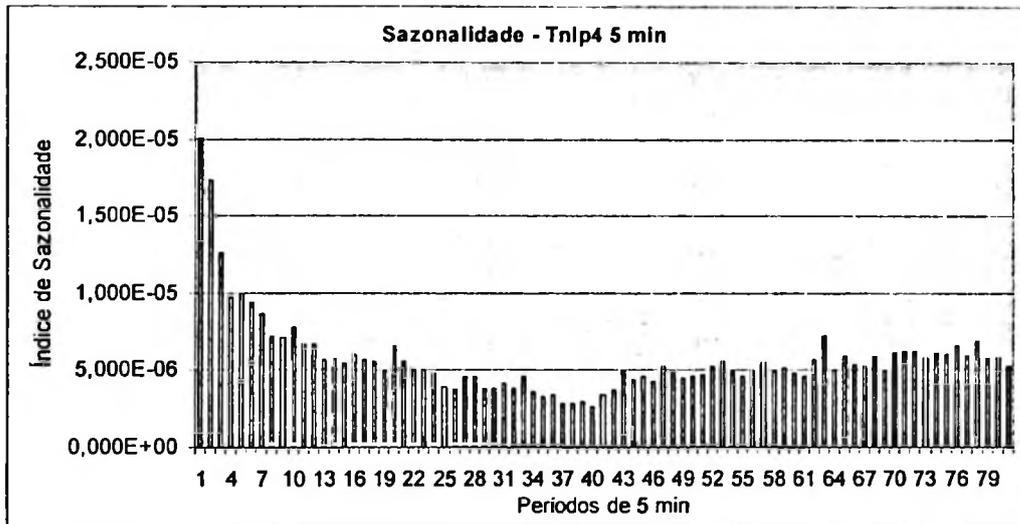
Adotou-se a técnica determinística de filtragem de sazonalidade intra-diária usada por Eersen e Bollerslev (1997, 1998, 1999) e Giot (2000).

Para a frequência de 5 minutos, calculou-se o índice de sazonalidade para cada um dos 81 intervalos de 5 minutos dos 798 dias da amostra. Desta forma, sendo I_j o índice sazonal relativo ao j-ésimo intervalo de 5 minutos, estimou-se 81 índices relativos a cada um dos 81 intervalos de 5 minutos da amostra através da média dos quadrados dos retornos verificados em cada intervalo ao longo dos 798 dias da amostra, ou seja,

$$I_j = \frac{1}{798} \sum_{k=1}^{798} \sum R^2_{k,j}, \quad j = 1, \dots, 81.$$

Encontrou-se um padrão de sazonalidade intra-diária em formato de J refletido como se esperava, sendo que a volatilidade maior ocorre no início do dia. Veja Gráfico 4.4.1

Gráfico 4.4.1: Índice de Sazonalidade de Frequência de 5 minutos

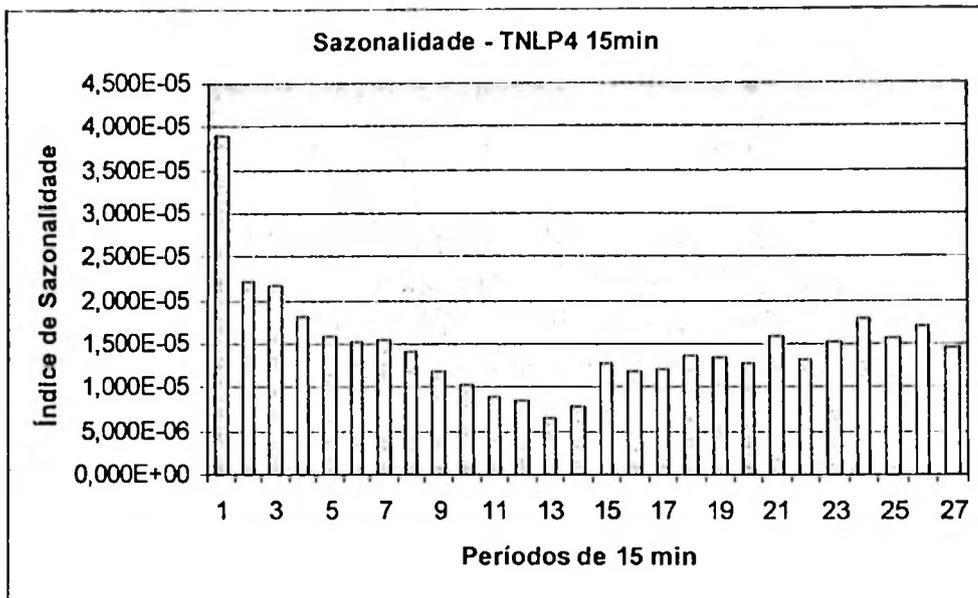


Fez-se cálculo semelhante para as freqüências de 15 minutos com 27 observações/dia:

$$I_j = \frac{1}{798} \sum_{k=1}^{798} \sum R^2_{k,j} \quad j = 1, \dots, 27$$

Encontrou-se para a freqüência de 15 minutos um padrão de sazonalidade intra-diária semelhante aquele da freqüência de 5 minutos (Gráfico 4.4.2):

Gráfico 4.4.2: Índice de Sazonalidade de Freqüência de 15 minutos

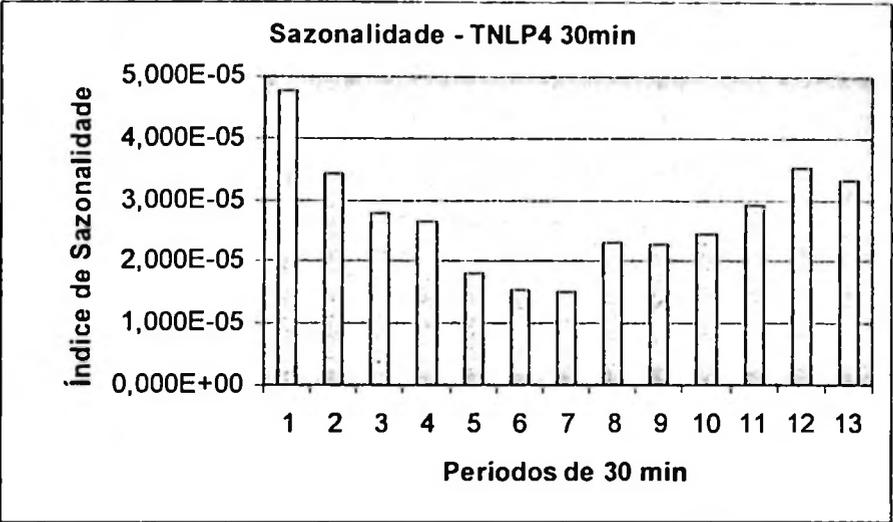


Sazonalidade para Frequência de 30 minutos: De forma equivalente, encontrou-se padrão sazonal intra-diário para as frequências de 30 minutos.

30 minutos (13 observações/dia):

$$I_j = \frac{1}{798} \sum_{k=1}^{798} \sum R^2_{k,j} \quad j = 1, \dots, 13$$

Gráfico 4.4.3: Índice de Sazonalidade de Frequência de 30 minutos

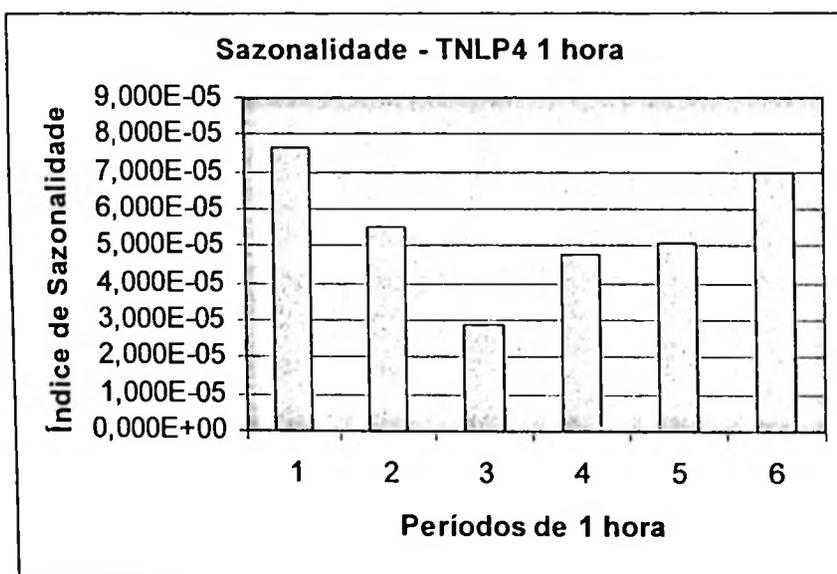


Sazonalidade para Frequência de 1 hora: De forma equivalente, encontrou-se também padrão sazonal intra-diário para as frequências de 1 hora.

1 hora (6 observações/dia):

$$I_j = \frac{1}{798} \sum_{k=1}^{798} \sum R^2_{k,j} \quad j = 1, \dots, 6$$

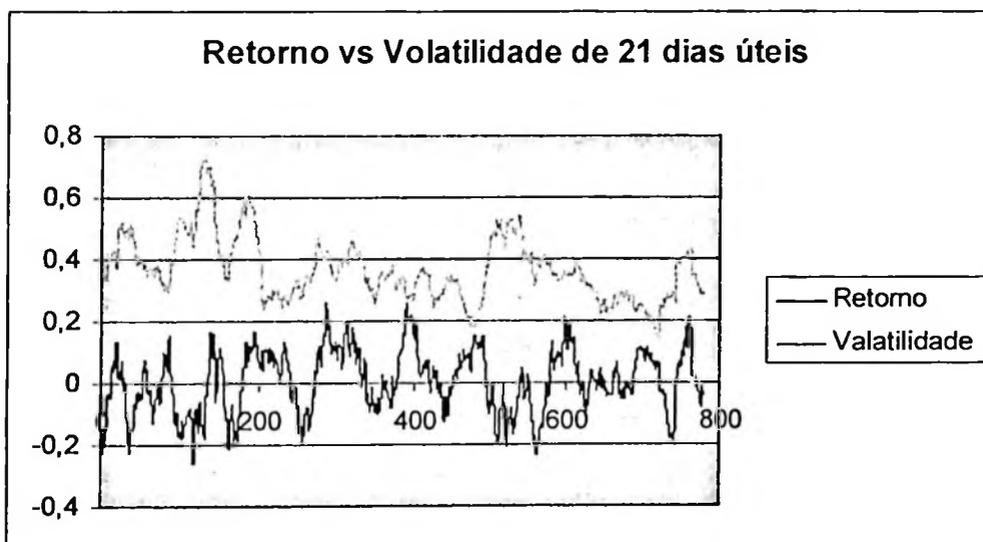
Gráfico 4.4.4: Índice de Sazonalidade de Frequência de 1 hora



4.5 EFEITO DE ALAVANCAGEM

O impacto dos retornos na volatilidade é assimétrico, com um efeito mais acentuado para retornos negativos quando comparados com retornos positivos. A comparação das séries de retornos e quadrados de retornos dá uma idéia sobre esse fato estilizado, além da análise de desempenho de modelos que incluem essa assimetria como modelos EGARCH. Ver Gráfico abaixo (4.4.5) em que o retorno está representado na linha superior e a volatilidade na linha inferior.

Gráfico 4.4.5: Gráfico de Retorno vs Volatilidade de 21 dias úteis



5 AJUSTE DE MODELOS DE VOLATILIDADE

Nesse capítulo ajustam-se modelos de projeção de volatilidade realizada, utilizando variações da família de modelos GARCH, para os dados diários e para três frequências de dados intra-diários de TNLP4, no período de 02 de Janeiro de 2002 a 30 de Dezembro de 2004. Há também o ajuste de modelo de regressão a partir de volatilidade implícita de opções de TNLP4.

5.1 AJUSTE DE MODELOS DE SÉRIES TEMPORAIS DA FAMÍLIA GARCH PARA DADOS DIÁRIOS

Como vimos na seção referente a fatos estilizados, identificou-se que a série de dados diários de TNLP4 para esse período possui comportamento semelhante aquele de um processo de ruído branco. Vimos também que a função de auto-correlação (ACF) indicou não haver auto-correlação. O Gráfico de auto-correlação parcial indicou que a PACF está dentro dos intervalos para todos os lags.

Ajuste do Modelo para Série Diária

As alternativas de modelos GARCH listados abaixo apresentaram resultados comparáveis quando se comparam os indicadores AIC, BIC e os Gráficos QQ-Plot e de função de auto-correlação e auto-correlação parcial dos quadrados dos resíduos (ver tabela 5.1.1). Escolheu-se o modelo GARCH(1,1) sem a modelagem da média por ser o modelo mais simples (ver resultado do ajuste na tabela 5.1.2).

Tabela 5.1.1: Indicadores AIC e BIC para modelagem do retorno de frequência diária

Modelo	AIC	BIC	Coef.Alavancagem / t-stat.
GARCH(1,1)	-3463	-3445	
GARCH(2,1)	-3462	-3439	
GARCH(1,2)	-3459	-3436	
EGARCH(1,1)	-3466	-3443	Negativo / não significativo
PGARCH	-3469	-3446	Negativo / não significativo
TGARCH	-3469	-3446	Negativo / não significativo
GARCH com Distribuição Student	-3463	-3440	
GARCH com Distribuição "GED"	-3466	-3443	

Tabela 5.1.2: Resultados para o ajuste do Modelo GARCH(1,1) sem a modelagem de média

```

tnlp4d=importData("tnlp4d1204.txt",type="ASCII",colNames=c("Date","Preço","Retorno",
"R2"))
attach(tnlp4d)
> tnlp4d.mod1=garch(Retorno ~1,~ garch(1, 1))
> summary(tnlp4d.mod1)
Call:
garch(formula.mean = Retorno ~ 1, formula.var = ~ garch(1, 1))
Mean Equation: Retorno ~ 1
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: gaussian
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	8.647e-004	7.995e-004	1.082	0.1399104
A	6.294e-006	5.066e-006	1.242	0.1072491
ARCH(1)	5.457e-002	1.562e-002	3.493	0.0002528
GARCH(1)	9.343e-001	1.950e-002	47.918	0.0000000

```

-----
AIC(4) = -3463.676
BIC(4) = -3445.249
Normality Test:
-----
Jarque-Bera P-value Shapiro-Wilk P-value
[1,] 2.054 0.3582 0.986 0.5033
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 11.44 0.4912 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 14.6 0.2641 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag
12 C
[1,] -0.7955 -0.1837 1.62 -0.3199 -0.04164 2.841 -0.9106 -0.31 0.3041 -0.7121 -0.635
0.05068 -0.518
TR^2 P-value F-stat P-value
[1,] 14.05 0.2972 1.303 0.3271

```

Gráfico 5.1.1: Função de auto-correlação dos resíduos ao quadrado do modelo GARCH(1,1) para dados diários

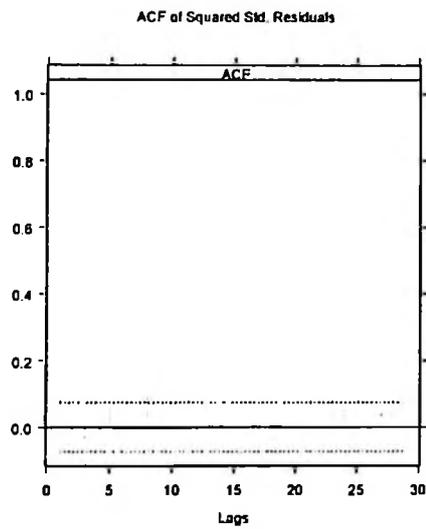
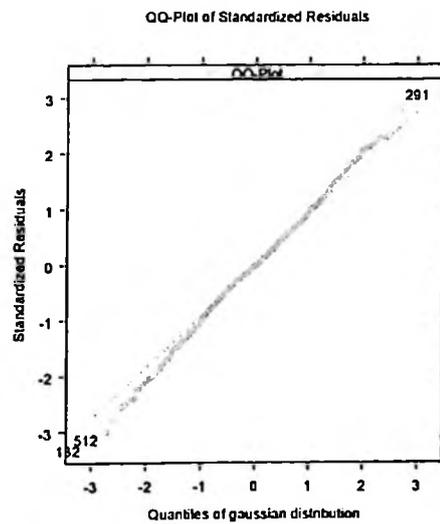


Gráfico 5.1.2: QQ-Plot do modelo GARCH(1,1) para dados diários



5.2 AJUSTE DE MODELO DA FAMÍLIA GARCH PARA SÉRIES INTRA-DIÁRIAS

Essa seção trata do ajuste de modelos da família GARCH para as séries de retornos dessazonalizados intra-diários de TNL4 para o período de 02 de Janeiro de 2002 a 30 de Dezembro de 2004.

Escolheu-se o método de filtragem de sazonalidade intra-diária como descrito no capítulo 4 e calculou-se a série de retornos dessazonalizados a partir da fórmula:

$$y_t = \frac{R_t}{\sqrt{\phi_t}},$$

onde ϕ_t é o índice de sazonalidade para o instante t .

A partir da série de retornos dessazonalizados passou-se à análise das séries de frequências de 5 minutos, 15 minutos e 2 horas de forma a ajustar um modelo da família GARCH para essas frequências.

Ajuste do Modelo para Série de 5 minutos

Analisando-se os indicadores AIC, BIC e QQ-Plot Resíduos e ACF dos resíduos ao quadrado, os modelos apresentaram resultados semelhantes sendo que damos maior destaque aqui para os modelos GARCH(1,1) + ARMA(2,0) pela simplicidade, GARCH(1,1) + ARMA(2,0) com distribuição de Student e com distribuição GED por terem melhor desempenho e o modelo de memória longa FIGARCH(1,1) + ARMA(2,0) por ter desempenho comparável aos demais.

Tabela 5.2.1 : Indicadores AIC e BIC para modelagem do retorno de frequência de 5 min

Modelo	AIC	BIC
GARCH(1,1)	161.252	161.306
EGARCH(1,1)	161.763	161.826
PGARCH	161.192	161.255
TGARCH	161.213	161.276
GARCH com Dist. Student	157.342	157.405
GARCH com Dist. "GED"	155.620	155.683
Modelo FIGARCH (longa memória)	160.526	160.589
Modelo FIEGARCH (memória longa)	160.778	160.850

Ajuste da média: A análise de auto-correlação indicou um modelo ARMA(2,0) para modelar a média de forma a eliminar a auto-correlação até o lag 2. Ver tabela 5.2.2 e gráficos 5.2.1 e 5.2.2 da função de auto-correlação e auto-correlação parcial.

Tabela 5.2.2: Análise de Auto-correlação e Auto-correlação Parcial

```
> module(finmetrics)
S+FinMetrics Version 1.0.2 for Microsoft Windows : 2003
>
tnlp45des=importData("tnlp45des_0412.txt",type="ASCII",colNames=c("Date","Preço",
"Retorno","R2","Fat","RetDes","R2Des","LR2"))
> attach(tnlp45des)
> tnlp45des.ret=RetDes

> acf(tnlp45des.ret)
Call: acf(x = tnlp45des.ret)

Autocorrelation matrix:
lag tnlp45des.ret
1 0 1.0000
2 1 -0.1060
3 2 -0.0094
4 3 0.0048
5 4 -0.0002
6 5 0.0135
7 6 0.0049
8 7 0.0039

> acf(tnlp45des.ret,type="partial")
Call: acf(x = tnlp45des.ret, type = "partial")

Partial Correlation matrix:
lag tnlp45des.ret
1 1 -0.1060
2 2 -0.0208
3 3 0.0016
4 4 0.0003
5 5 0.0138
6 6 0.0079
7 7 0.0058
```

Gráfico 5.2.1: Função de Auto-correlação

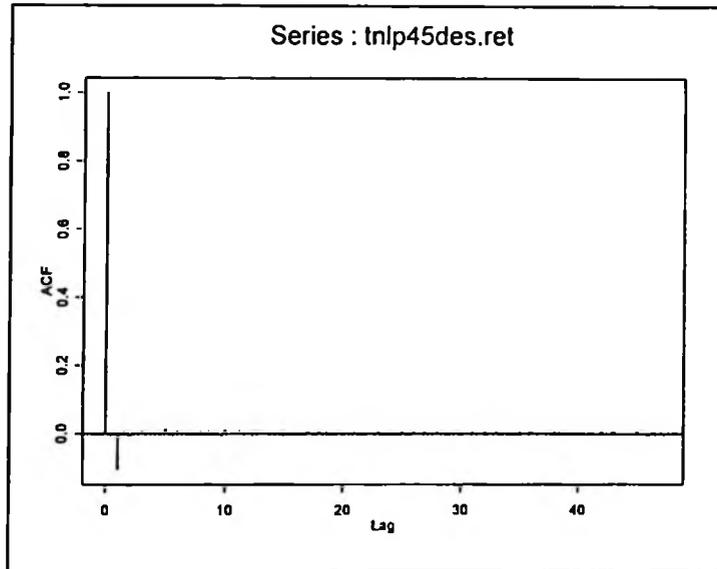
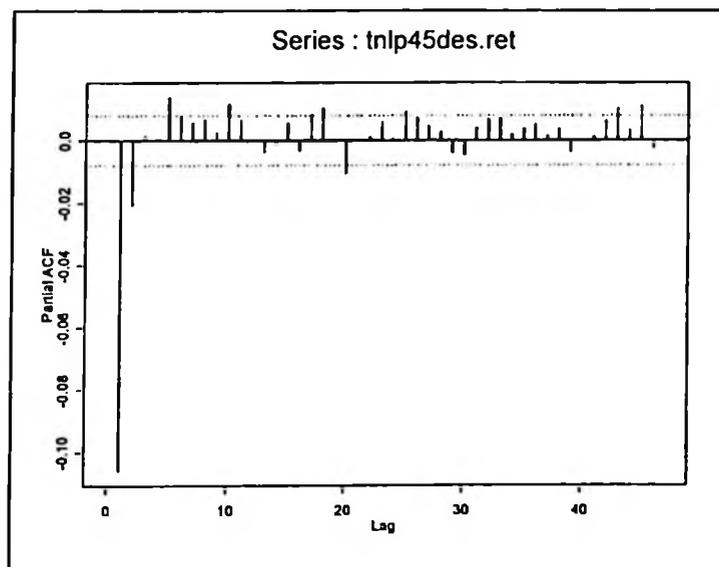


Gráfico 5.2.2: Função de Auto-correlação Parcial



O ajuste do modelo ARMA(2,0) para a média apresentou bons resultados (ver tabela 5.2.3).

Tabela 5.2.3: Ajuste de Modelo ARMA(2,0)

Dependent Variable: RETDES				
Method: Least Squares				
Date: 06/23/05 Time: 20:56				
Sample(adjusted): 3 59940				
Included observations: 59938 after adjusting endpoints				
Convergence achieved after 2 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.108133	0.004084	-26.47862	0.0000
AR(2)	-0.020807	0.004084	-5.095074	0.0000
R-squared	0.011627	Mean dependent var		-0.004751
Adjusted R-squared	0.011610	S.D. dependent var		1.000015
S.E. of regression	0.994193	Akaike info criterion		2.826262
Sum squared resid	59241.88	Schwarz criterion		2.826562
Log likelihood	-84698.24	Durbin-Watson stat		1.999904
Inverted AR Roots	-.05 -.13i	-.05+.13i		

Tabela 5.2.4: Ajuste de modelo ARMA (2,0) + GARCH(1,1) para a volatilidade

```

Modelo GARCH (1,1)
> tnlp45des.mod1=garch(formula.mean = RetDes ~ ar(2), formula.var = ~ garch(1, 1))
summary(tnlp45des.mod1)

> summary(tnlp45des.mod1)
Call:
garch(formula.mean = RetDes ~ ar(2), formula.var = ~ garch(1, 1))
Mean Equation: RetDes ~ ar(2)
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: gaussian
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	0.002121	0.0035868	0.5912	2.772e-001
AR(1)	-0.112240	0.0040773	-27.5278	0.000e+000
AR(2)	-0.021120	0.0042216	-5.0028	2.832e-007
A	0.016585	0.0004784	34.6668	0.000e+000
ARCH(1)	0.053497	0.0008128	65.8217	0.000e+000
GARCH(1)	0.929924	0.0012016	773.8749	0.000e+000

```

-----
AIC(6) = 161252.7
BIC(6) = 161306.7

Normality Test:
-----
Jarque-Bera P-value
[1,] 75036 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 33.42 0.0008333 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 59.48 2.812e-008 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12 C
[1,] 1.123 0.3249 -1.583 0.1736 -1.269 -1.145 -1.378 -1.041 -2.42 -1.298 -1.154 -0.01026 6.114

TR^2 P-value F-stat P-value
[1,] 57.68 5.966e-008 5.248 0.001871

```

Gráfico 5.2.3: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

GARCH(1,1) + ARMA(2,0)

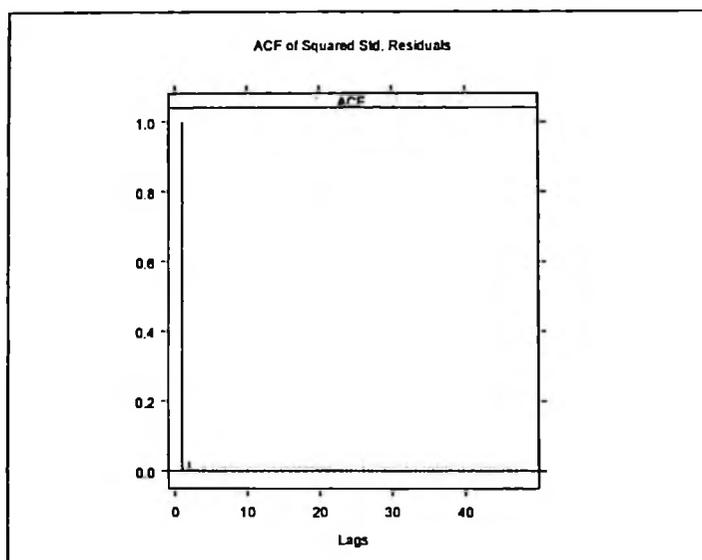


Gráfico 5.2.4: QQ-Plot para Modelo Ajustado GARCH(1,1) + ARMA(2,0)

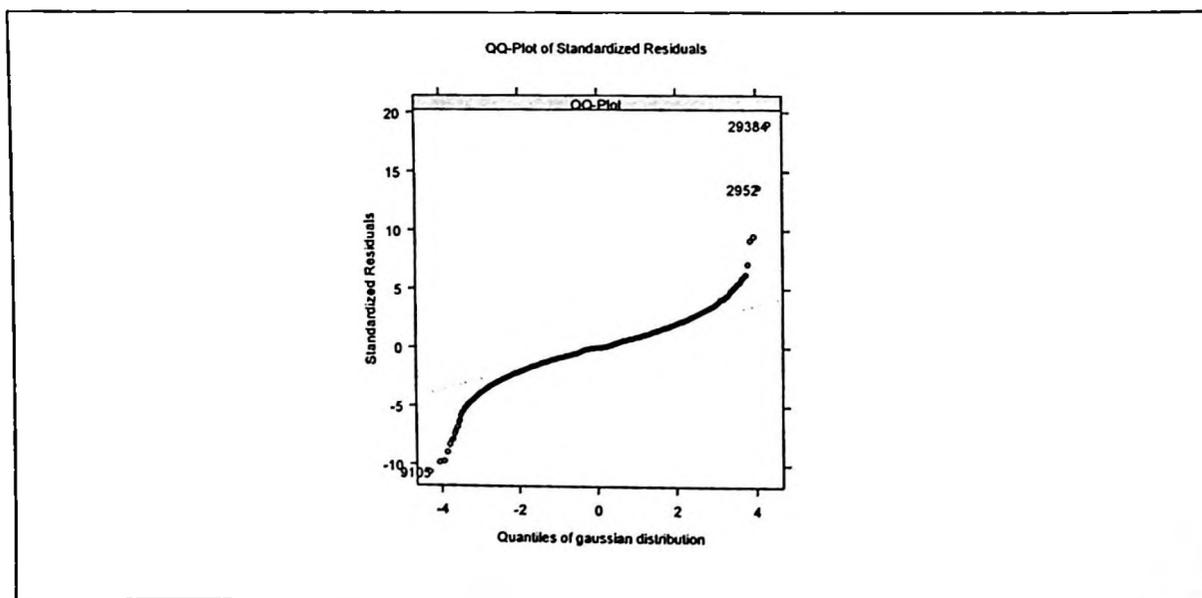


Tabela 5.2.5: Ajuste de modelo ARMA (2,0) + GARCH(1,1) com Dist. de Student para a volatilidade

```

> tnlp45des.mod11=garch(RetDes~ar(2),~garch(1,1), cond.dist="t", trace=F)
> summary(tnlp45des.mod11)

> tnlp45des.mod11=garch(RetDes~ar(2),~garch(1,1), cond.dist="t", trace=F)
> summary(tnlp45des.mod11)

Call:
garch(formula.mean = RetDes ~ ar(2), formula.var = ~ garch(1, 1), trace = F, cond.dist = "t")
Mean Equation: RetDes ~ ar(2)
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: t
with estimated parameter 5.931004 and standard error 0.1370427
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	0.001362	0.0033056	0.4119	0.3402
AR(1)	-0.125354	0.0039065	-32.0883	0.0000
AR(2)	-0.032901	0.0039817	-8.2630	0.0000
A	0.008812	0.0007078	12.4503	0.0000
ARCH(1)	0.048053	0.0017631	27.2550	0.0000
GARCH(1)	0.944373	0.0019707	479.1949	0.0000

```

-----
AIC(7) = 157342.7
BIC(7) = 157405.7
Normality Test:
-----
Jarque-Bera P-value
[1,] 90337 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 65.64 2.072e-009 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 63.21 5.83e-009 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12 C
[1,] 1.461 0.4341 -1.407 0.2377 -1.142 -1.103 -1.396 -1.104 -2.382 -1.348 -1.366 -0.01018 6.3

TR^2 P-value F-stat P-value
[1,] 61.05 1.451e-008 5.556 0.001435

```

Gráfico 5.2.5: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

GARCH(1,1) + ARMA(2,0) com Dist. Student

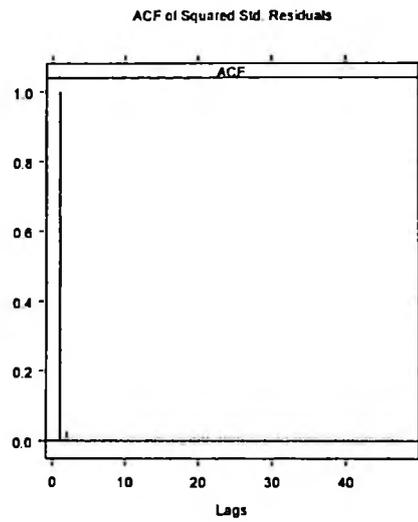


Gráfico 5.2.6: QQ-Plot para Modelo Ajustado GARCH(1,1) + ARMA(2,0) com Dist. Student

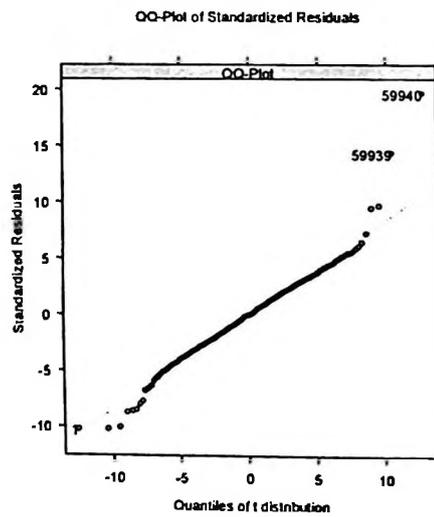


Tabela 5.2.6: Ajuste de modelo ARMA (2,0) + GARCH(1,1) com Distribuição GED para a volatilidade

```

tnlp45des.mod13=garch(RetDes~ar(2),~garch(1,1), cond.dist="ged", trace=F)
> summary(tnlp45des.mod13)
Call:
garch(formula.mean = RetDes ~ ar(2), formula.var = ~ garch(1, 1), trace = F, cond.dist = "ged")
Mean Equation: RetDes ~ ar(2)
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: ged
with estimated parameter 1.000242 and standard error 0.006350288
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	-1.271e-004	0.002759	-0.046060	0.4816
AR(1)	-9.211e-005	0.003317	-0.027772	0.4889
AR(2)	5.401e-006	0.003276	0.001649	0.4993
A	1.503e-002	0.001236	12.156458	0.0000
ARCH(1)	7.606e-002	0.003094	24.582213	0.0000
GARCH(1)	9.179e-001	0.003188	287.918035	0.0000

```

-----
AIC(7) = 155620.6
BIC(7) = 155683.6
Normality Test:
-----
Jarque-Bera P-value
[1,] 126471 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 512.6 0 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 53 4.115e-007 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12 C
[1,] -0.1174 -0.7312 -2.165 -0.9607 -1.891 -1.635 -1.851 -1.638 -2.642 -1.575 -1.574 -0.007414
4.515

TR^2 P-value F-stat P-value
[1,] 51.87 6.531e-007 4.72 0.003045

```

Gráfico 5.2.7: Função de Autocorrelação dos Resíduos ao Quadrado para Modelo Ajustado

GARCH(1,1) + ARMA(2,0) com Dist. GED

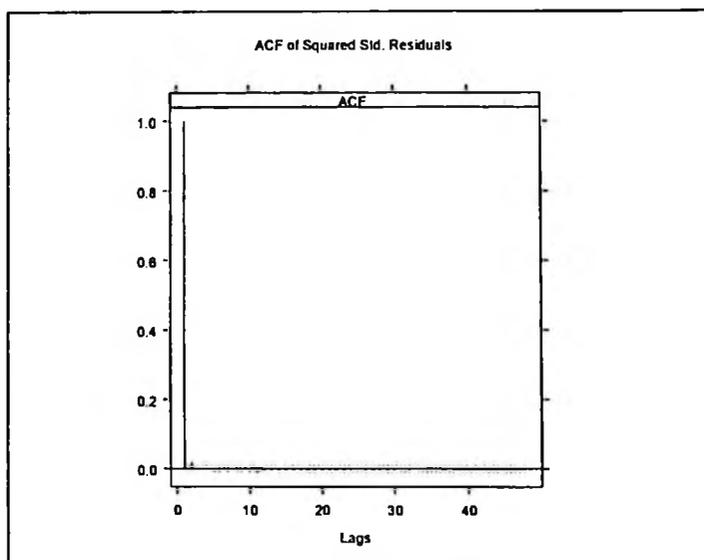


Gráfico 5.2.8: QQ-Plot para Modelo Ajustado GARCH(1,1) + ARMA(2,0) com Dist. GED

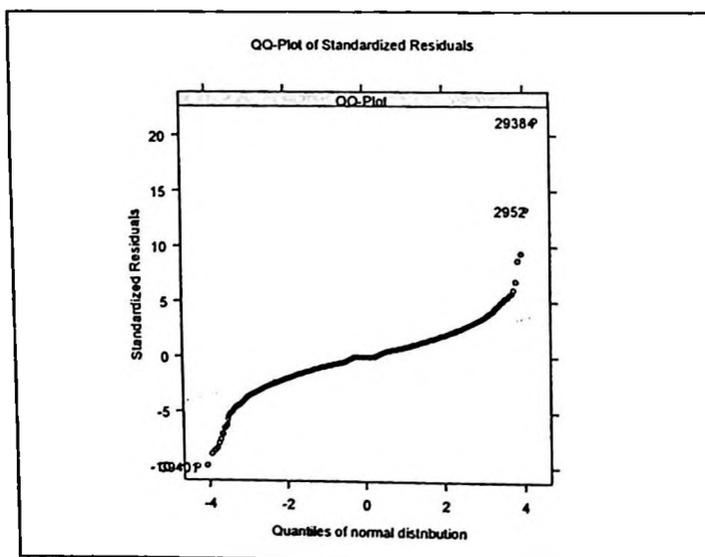


Tabela 5.2.7: Ajuste de modelo ARMA (2,0) + FIGARCH(1,1) para a volatilidade

```

> tnlp4des.figarch=fgarch(RetDes~ar(2),~figarch(1,1))
> summary(tnlp4des.figarch)
Call:
fgarch(formula.mean = RetDes ~ ar(2), formula.var = ~ figarch(1, 1))

Mean Equation: RetDes ~ ar(2)
Conditional Variance Equation: ~ figarch(1, 1)
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	0.002365	0.003512	0.6734	2.504e-001
AR(1)	-0.111925	0.004475	-25.0126	0.000e+000
AR(2)	-0.023394	0.004195	-5.5763	1.234e-008
A	0.068691	0.002435	28.2068	0.000e+000
GARCH(1)	0.562193	0.011581	48.5454	0.000e+000
ARCH(1)	0.398986	0.011753	33.9477	0.000e+000
fraction	0.264259	0.004468	59.1455	0.000e+000

```

-----
AIC(7) = 160526.9
BIC(7) = 160589.9
Normality Test:
-----
Jarque-Bera P-value
[1,] 54609 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 36.56 0.0002629 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 15.88 0.1969 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12
C
[1,] 0.6573 -0.08529 -2.148 -0.06957 -1.378 -1.162 -1.275 -0.5843 -2.164 -0.6632 -0.06524 -
0.01084 0.718
TR^2 P-value F-stat P-value
[1,] 15.9 0.1957 1.446 0.252

```

Gráfico 5.2.9: Função de Autocorrelação dos Resíduos ao Quadrado para Modelo Ajustado

FIGARCH(1,1) + ARMA(2,0)

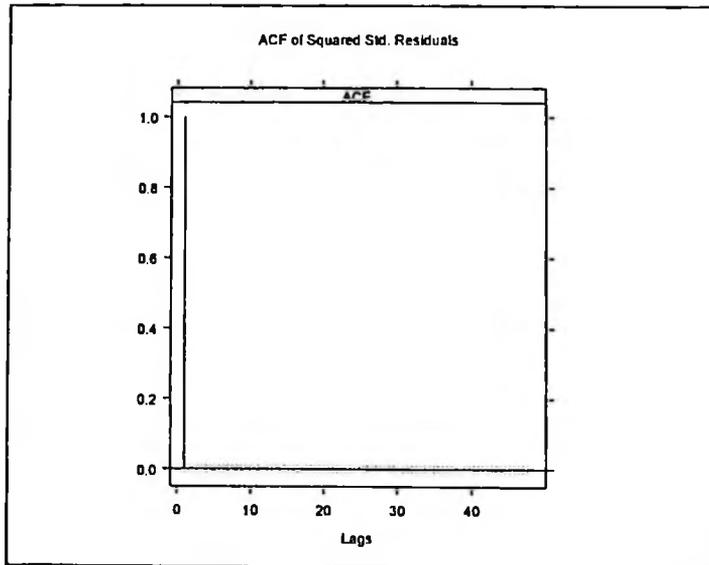
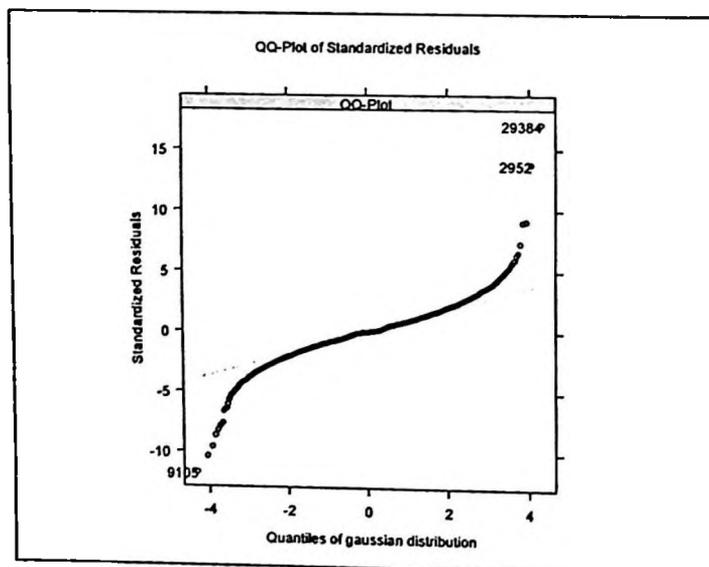


Gráfico 5.2.10: QQ-Plot para Modelo Ajustado FIGARCH(1,1) + ARMA(2,0)



Como esperado, a série de retornos de 5 minutos apresentou parâmetros de memória fracionários.

Tabela 5.2.8: Parâmetros de Memória Longa para Série de Retornos de 5 minutos

```
Estimação de Parâmetro de Memória Longa: Valores para d entre 0,16 e 0,22
> d.ros(abs(RetDes))
Warning messages:
 168 observations with NA/NaN/Inf in x, y, or wt removed. in: lsfit(log10(kVec),
log10(RoS))
[1] 0.1608362
> d.pgram(abs(RetDes))
[1] 0.2201608
```

Ajuste do Modelo para Série de 15 minutos

Analisando-se os indicadores AIC, BIC e QQ-Plot Resíduos e ACF dos resíduos ao quadrado, os modelos apresentaram resultados semelhantes sendo que damos maior destaque aqui para os modelos GARCH(1,1) + ARMA(3,0) pela simplicidade, GARCH(1,1) + ARMA(3,0) com distribuição de Student e com distribuição GED por terem melhor desempenho e o modelo de memória longa FIGARCH(1,1) + ARMA(3,0). As variantes GARCH(1,2) e GARCH(2,1) apresentaram resultados comparáveis ao GARCH(1,1). Ver tabela 5.2.9.

Tabela 5.2.9 : Indicadores AIC e BIC para modelagem do retorno de frequência de 15 min

Modelo	AIC	BIC
GARCH(1,1)	54.190	54.245
EGARCH(1,1)	54.247	54.310
PGARCH	54.183	54.246
TGARCH	54.183	54.246
GARCH(1,1) com Dist. Student	53.070	53.126
GARCH(1,2) com Dist. Student	53.055	53.126
GARCH(2,1) com Dist. Student	53.024	53.095
GARCH com Dist. *GED*	53.079	53.142
Modelo FIGARCH (longa memória)	54.069	54.132
Modelo FIEGARCH (memória longa)	54.111	54.174

Ajuste de Modelo ARMA (3,0) para a média. O gráfico da função de auto-correlação (gráfico 5.2.11) indicou haver auto-correlação significativa até o lag 3 e bem baixa para lags maiores como 7, etc. O gráfico de auto-correlação parcial indicou que a PACF está fora dos intervalos para vários lags.

Tabela 5.2.10: Função de Auto-correlação para Retornos de 15 min

```
> tnlp415des.ret=RetDes
> acf(tnlp415des.ret,type="partial")
Call: acf(x = tnlp415des.ret, type = "partial")

Partial Correlation matrix:
lag tnlp415des.ret
1 1 -0.0439
2 2 0.0157
3 3 0.0204
4 4 0.0102
5 5 -0.0004
6 6 0.0054
7 7 -0.0018
```

Gráfico 5.2.11: Função de Auto-correlação para Retornos de 15 min

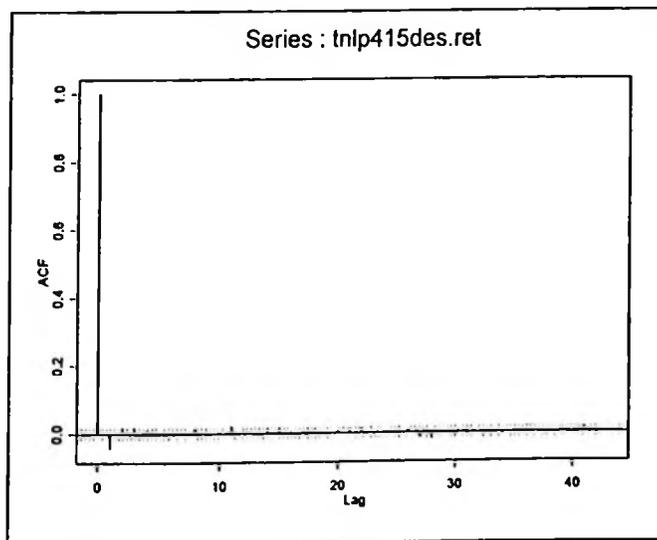
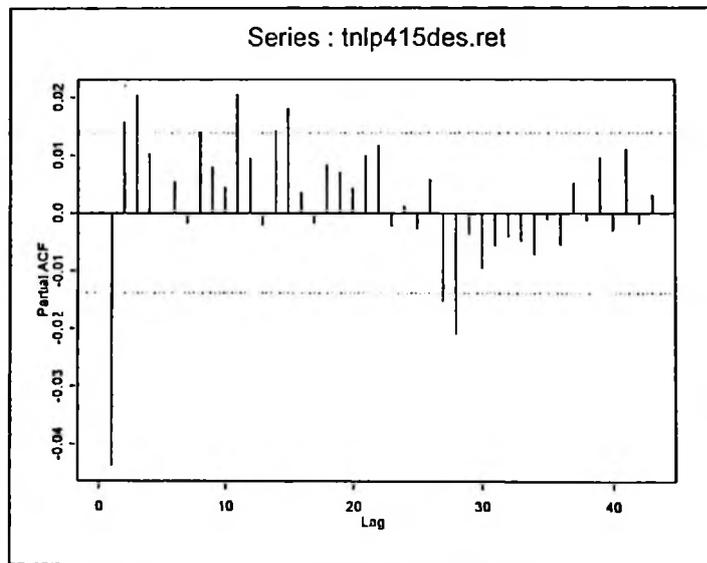


Gráfico 5.2.12: Função de Auto-correlação Parcial para Retornos de 15 min



Ajuste do Modelo ARMA(3,0) para a Média. O modelo ARMA(3,0) foi ajustado para retornos de 15 minutos de frequência com bom resultado. Ver tabela 5.2.11.

Tabela 5.2.11: Modelo ARMA(3,0) para Retornos de 15 min

Dependent Variable: RETDES
 Method: Least Squares
 Date: 07/02/05 Time: 19:21
 Sample(adjusted): 4 19980
 Included observations: 19977 after adjusting endpoints
 Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.043591	0.007074	-6.162005	0.0000
AR(2)	0.016682	0.007079	2.356471	0.0185
AR(3)	0.020501	0.007073	2.898506	0.0038
R-squared	0.002538	Mean dependent var	-0.007997	
Adjusted R-squared	0.002438	S.D. dependent var	1.005626	
S.E. of regression	1.004399	Akaike info criterion	2.846807	
Sum squared resid	20150.13	Schwarz criterion	2.847993	
Log likelihood	-28432.33	Durbin-Watson stat	1.999952	
Inverted AR Roots	.28	-.16+.22i	-.16 -.22i	

Date: 07/02/05 Time: 19:21
 Sample: 4 19980
 Included observations: 19977
 Q-statistic
 probabilities
 adjusted for 3
 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
*	*	1 0.153	0.153	468.46	
*	*	2 0.126	0.105	785.43	
*	*	3 0.104	0.073	1001.9	
*	*	4 0.097	0.063	1188.9	0.000
*	*	5 0.096	0.059	1371.3	0.000
*	*	6 0.080	0.040	1498.6	0.000
*	*	7 0.085	0.046	1644.2	0.000
*	*	8 0.095	0.055	1824.9	0.000
*	*	9 0.086	0.041	1974.2	0.000
*	*	10 0.082	0.036	2108.7	0.000
*	*	11 0.086	0.040	2257.7	0.000
*	*	12 0.104	0.057	2474.0	0.000
*	*	13 0.129	0.078	2806.7	0.000
*	*	14 0.088	0.026	2962.7	0.000
*	*	15 0.097	0.038	3150.9	0.000

Tabela 5.2.12: Ajuste do Modelo ARMA(3,0) + GARCH(1,1) para Volatilidade de dados de 15 min

```

lnlp415des.mod1=garch(formula.mean = RetDes ~ ar(3), formula.var = ~ garch(1, 1))
Call:
garch(formula.mean = RetDes ~ ar(3), formula.var = ~ garch(1, 1))
Mean Equation: RetDes ~ ar(3)
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: gaussian
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	0.001336	0.006356	0.2102	4.168e-001
AR(1)	-0.043967	0.007092	-6.1996	2.886e-010
AR(2)	0.018605	0.007239	2.5703	5.084e-003
AR(3)	0.017569	0.007201	2.4399	7.350e-003
A	0.012182	0.000927	13.1410	0.000e+000
ARCH(1)	0.046598	0.001721	27.0710	0.000e+000
GARCH(1)	0.941294	0.002244	419.4949	0.000e+000

```

-----
AIC(7) = 54190.53
BIC(7) = 54245.85
Normality Test:
-----
Jarque-Bera P-value
[1,] 5410 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 25.3 0.01347 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 48.5 2.556e-006 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag
12 C
[1,] 1.654 -0.3742 -0.1555 -1.035 -0.8281 -1.591 -1.244 -1.833 -1.894 -1.444 0.6556 -
0.04034 4.934
TR^2 P-value F-stat P-value
[1,] 45.28 9.229e-006 4.126 0.005561

```

Gráfico 5.2.13: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

GARCH(1,1) + ARMA(3,0)

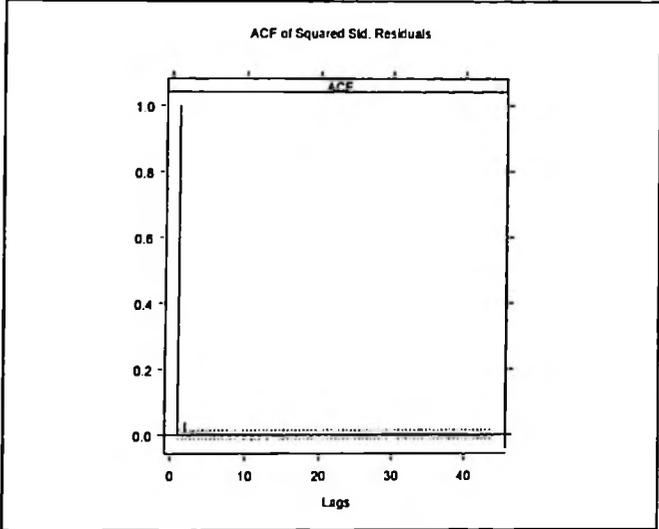


Gráfico 5.2.14: QQ-Plot para Modelo Ajustado GARCH(1,1) + ARMA(3,0)

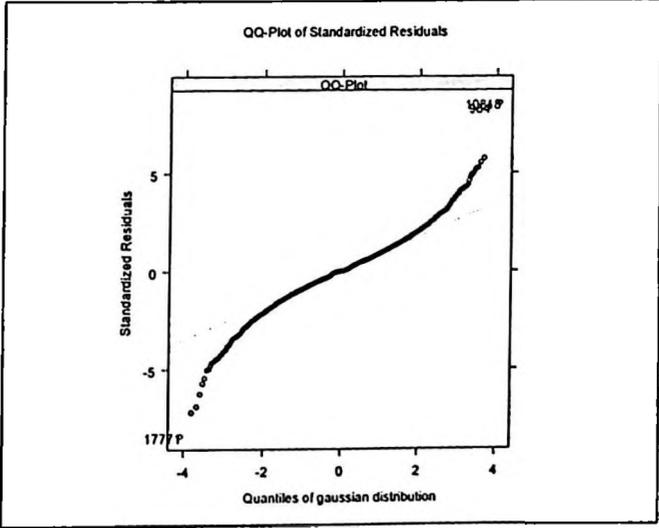


Tabela 5.2.13: Ajuste do modelo ARMA(3,0) + GARCH(1,1) com distribuição de Student

```

> tnlp415des.mod11=garch(RetDes~ar(3),~garch(1,1), cond.dist="t", trace=F)
> summary(tnlp415des.mod11)
Call:
garch(formula.mean = RetDes ~ ar(2), formula.var = ~ garch(1, 1), trace = F, cond.dist = "t")
Mean Equation: RetDes ~ ar(2)
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: t
with estimated parameter 6.027746 and standard error 0.2499735
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	0.0052426	0.005824	0.90014	1.840e-001
AR(1)	-0.0649568	0.006866	-9.46104	0.000e+000
AR(2)	0.0003789	0.006901	0.05491	4.781e-001
A	0.0103710	0.001456	7.12111	5.536e-013
ARCH(1)	0.0499557	0.003250	15.37069	0.000e+000
GARCH(1)	0.9405838	0.003786	248.43259	0.000e+000

```

-----
AIC(7) = 53070.99
BIC(7) = 53126.3
Normality Test:
-----
Jarque-Bera P-value
[1,] 5719 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 62.36 8.371e-009 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 48.67 2.389e-006 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12 C
[1,] 1.512 -0.6438 -0.4148 -1.191 -0.9606 -1.792 -1.512 -2.008 -2.116 -1.647 0.4305 -0.03905
4.502

TR^2 P-value F-stat P-value
[1,] 45.31 9.126e-006 4.128 0.005546

```

Gráfico 5.2.15: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

GARCH(1,1) + ARMA(3,0) com Distribuição de Student

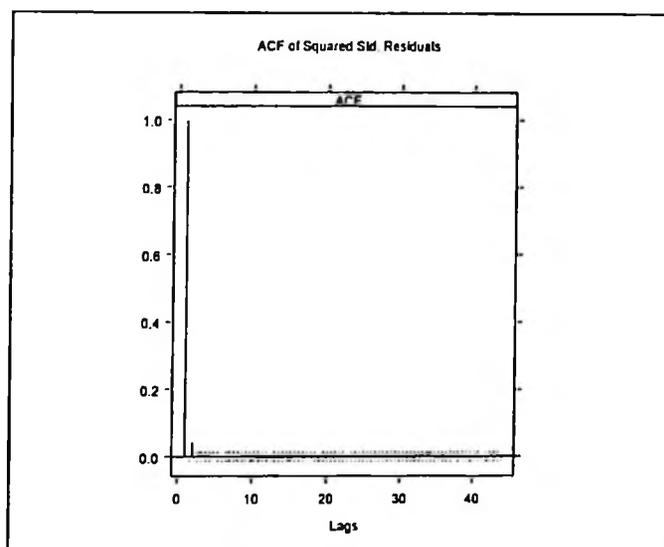


Gráfico 5.2.16: QQ-Plot dos Resíduos para Modelo Ajustado GARCH(1,1) + ARMA(3,0) com

Distribuição de Student

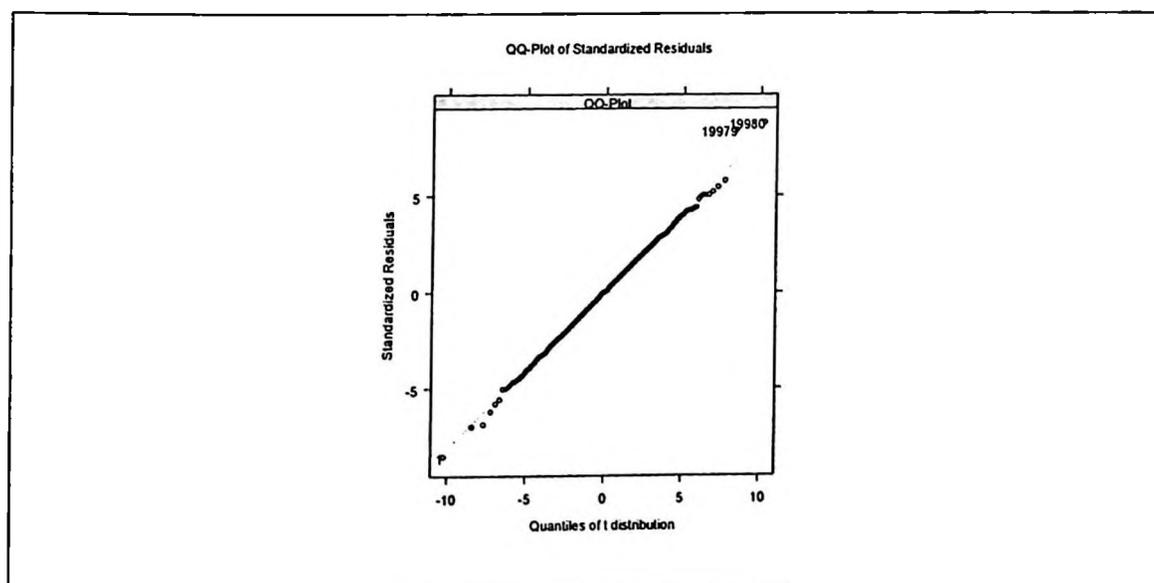


Tabela 5.2.14: Ajuste do modelo ARMA(3,0) + GARCH(1,1) com distribuição GED

```

> tnlp415des.mod12=garch(RetDes~ar(3),~garch(1,1), cond.dist="ged", trace=F)
> summary(tnlp415des.mod12)
Call:
garch(formula.mean = RetDes ~ ar(3), formula.var = ~ garch(1, 1), trace = F, cond.dist = "ged")
Mean Equation: RetDes ~ ar(3)
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: ged
with estimated parameter 1.275697 and standard error 0.01417313
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	0.002761	0.005570	0.4956	3.101e-001
AR(1)	-0.040279	0.006516	-6.1818	3.229e-010
AR(2)	-0.003933	0.006590	-0.5968	2.753e-001
AR(3)	0.004397	0.006599	0.6663	2.526e-001
A	0.008202	0.001274	6.4390	6.150e-011
ARCH(1)	0.042415	0.002815	15.0675	0.000e+000
GARCH(1)	0.949822	0.003373	281.6187	0.000e+000

```

-----
AIC(8) = 53079.03
BIC(8) = 53142.25
Normality Test:
-----
Jarque-Bera P-value
[1,] 5627 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 47.96 3.177e-006 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 54.68 2.069e-007 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12 C
[1,] 2.109 -0.1112 0.1222 -0.748 -0.6124 -1.479 -1.142 -1.792 -1.889 -1.364 0.7272 -0.04071
5.45
TR^2 P-value F-stat P-value
[1,] 50.91 9.67e-007 4.64 0.003291

```

Gráfico 5.2.17: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

GARCH(1,1) + ARMA(3,0) com Distribuição de GED

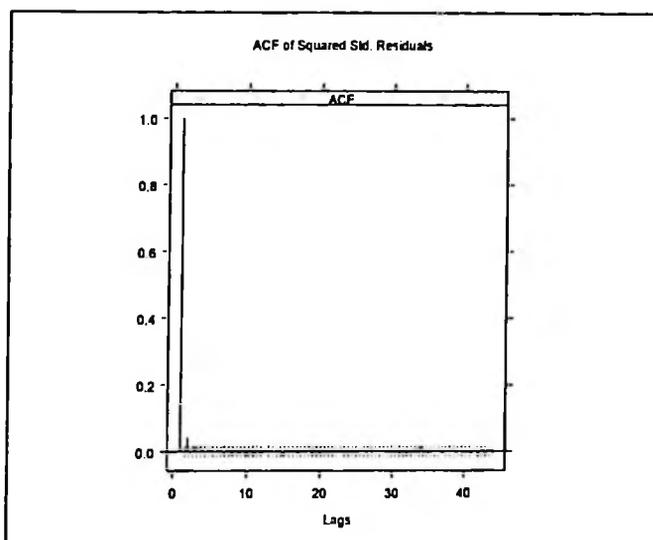


Gráfico 5.2.18: QQ-Plot dos Resíduos para Modelo Ajustado GARCH(1,1) + ARMA(3,0) com

Distribuição GED

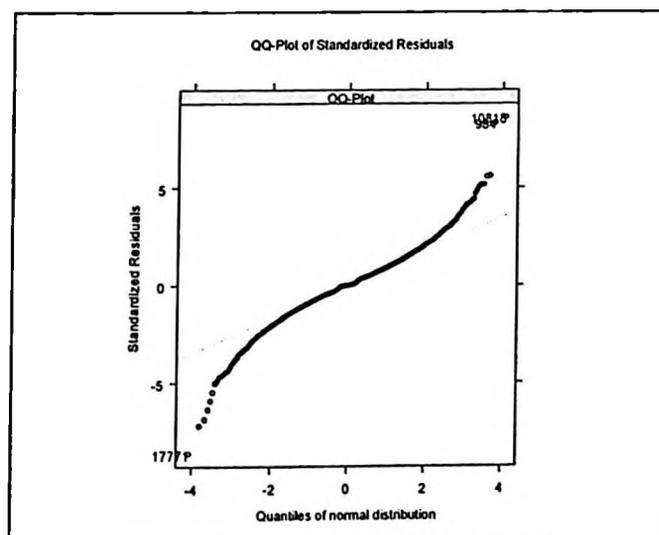


Tabela 5.2.15: Ajuste do modelo ARMA(3,0) + FIGARCH(1,1)

```

> tnlp415des.figarch=fgarch(RetDes~ar(3),~figarch(1,1))
> summary(tnlp415des.figarch)
Call:
fgarch(formula.mean = RetDes ~ ar(3), formula.var = ~ figarch(1, 1))
Mean Equation: RetDes ~ ar(3)
Conditional Variance Equation: ~ figarch(1, 1)
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	0.001555	0.006337	0.2454	4.031e-001
AR(1)	-0.047457	0.007989	-5.9405	1.444e-009
AR(2)	0.018632	0.007065	2.6373	4.182e-003
AR(3)	0.016676	0.006970	2.3927	8.368e-003
A	0.047113	0.004272	11.0293	0.000e+000
GARCH(1)	0.672143	0.016210	41.4640	0.000e+000
ARCH(1)	0.490372	0.015479	31.6795	0.000e+000
fraction	0.293319	0.012878	22.7772	0.000e+000

```

-----
AIC(8) = 54069.75
BIC(8) = 54132.97
Normality Test:
-----
Jarque-Bera P-value
[1,] 5125 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 25.9 0.0111 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 18.72 0.09542 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12 C
[1,] 2.07 0.1226 -0.02993 -0.9664 -0.823 -1.259 -1.176 -1.741 -1.746 -1.202 0.9953 -0.04081 -
1.027
TR^2 P-value F-stat P-value
[1,] 18.62 0.09806 1.695 0.1614

```

Gráfico 5.2.18: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

FIGARCH(1,1) + ARMA(3,0)

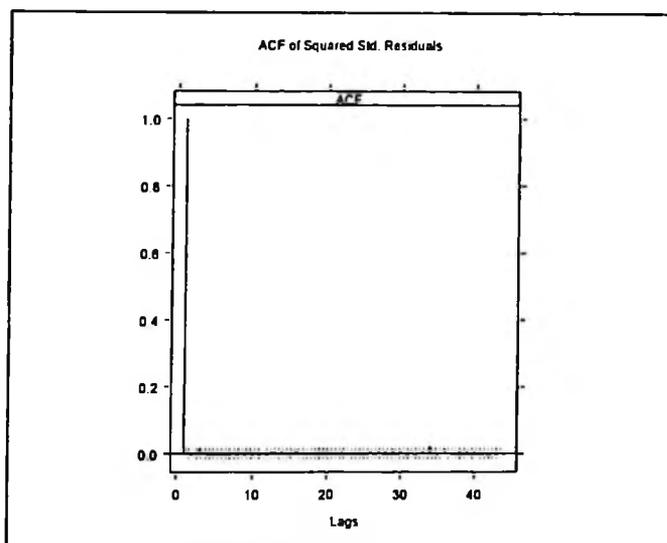


Gráfico 5.2.19: QQ-Plot dos Resíduos para Modelo Ajustado FIGARCH(1,1) + ARMA(3,0)

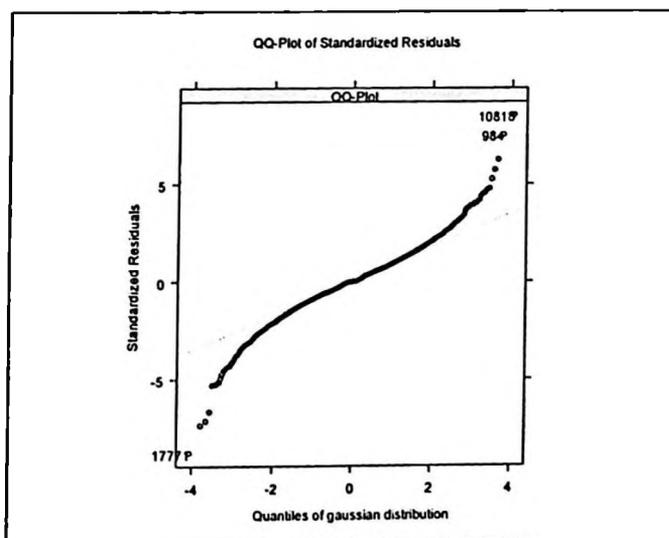


Tabela 5.2.16: Parâmetros de Memória Longa para Série de Retornos de 15 minutos

Estimação de Parâmetro de Memória Longa: Valores para d entre 0,15 e 0,31

```
> d.ros(abs(RetDes))
log10(RoS)
[1] 0.1524472
> d.pgram(abs(RetDes))
[1] 0.3164816
```

Ajuste do Modelo para Série de 2 horas

Analisando-se os indicadores AIC, BIC e QQ-Plot Resíduos e ACF dos resíduos ao quadrado, os modelos apresentaram resultados semelhantes sendo que damos maior destaque aqui para os modelos GARCH(1,1) + ARMA(3,0) pela simplicidade, GARCH(1,1) + ARMA(3,0) com distribuição de Student e com distribuição GED por terem melhor desempenho e o modelo de memória longa FIGARCH(1,1) + ARMA(3,0). As variantes GARCH(1,2) e GARCH(2,1) apresentaram resultados comparáveis ao GARCH(1,1). Ver tabela 5.2.17.

Tabela 5.2.17 : Indicadores AIC e BIC para modelagem do retorno de frequência de 2 horas

Modelo	AIC	BIC
GARCH(1,1)	6123	6163
EGARCH(1,1)	6116	6162
PGARCH	6121	6166
TGARCH	6121	6166
GARCH(1,1) com Dist. Student	6022	6068
GARCH(1,2) com Dist. Student	6023	6074
GARCH(2,1) com Dist. Student	6022	6073
GARCH com Dist. "GED"	6020	6065
Modelo FIGARCH (longa memória)	6125	6171
Modelo FIEGARCH (memória longa)	6121	6166

Ajuste de modelo para a média para frequência de 2 horas. O gráfico da função de auto-correlação (gráfico 5.2.20) indicou haver auto-correlação significativa até o lag 3 e bem baixa para lags maiores como 9, etc. O gráfico de auto-correlação parcial indicou que a PACF está fora dos intervalos para vários lags.

Tabela 5.2.18: Função de Auto-correlação para Retornos de 2 horas

```
> tnlp42hdes.ret=RetDes
>acf(tnlp42hdes.ret,type="partial")
Partial Correlation matrix:
lag tnlp42hdes.ret
1 1    0.0700
2 2    0.0176
3 3   -0.0600
4 4   -0.0071
5 5    0.0160
6 6   -0.0185
7 7   -0.0079
8 8   -0.0043
9 9   -0.0173
```

Gráfico 5.2.20: Função de Auto-correlação de série de retornos de 2 horas

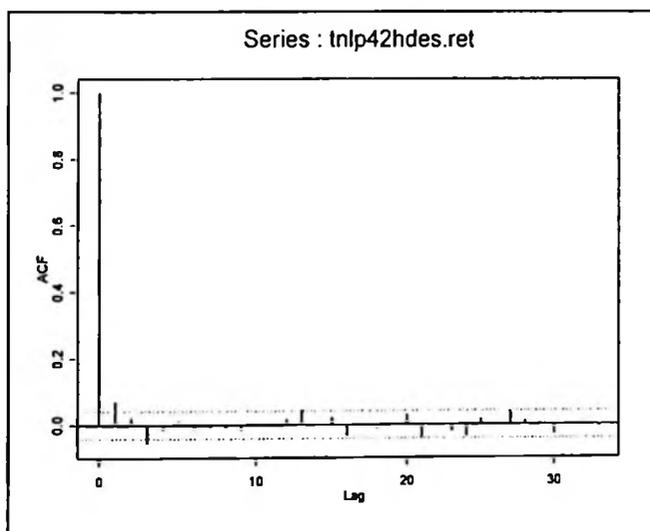
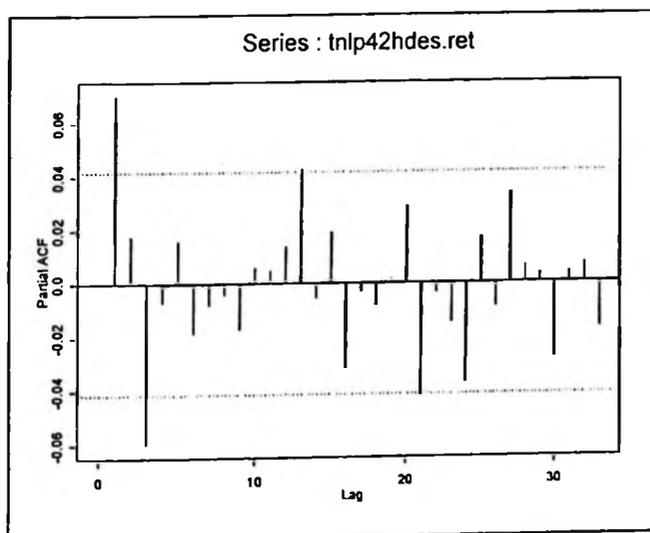


Gráfico 5.2.21: Função de Auto-correlação Parcial de série de retornos de 2 horas



Ajuste do Modelo ARMA(3,0) para a Média. O modelo ARMA(3,0) foi ajustado para retornos de 2 horas de frequência com bom resultado. Ver tabela 5.2.19.

Tabela 5.2.19: Modelo ARMA(3,0) para Retornos de 2 horas

Dependent Variable: RETDES						
Method: Least Squares						
Date: 07/03/05 Time: 20:57						
Sample(adjusted): 4 2220						
Included observations: 2217 after adjusting endpoints						
Convergence achieved after 2 iterations						
Variable	Coefficient	Std. Error	t-Statistic	Prob.		
AR(1)	0.073040	0.021164	3.451186	0.0006		
AR(3)	-0.057068	0.021162	-2.696749	0.0071		
R-squared	0.006889	Mean dependent var	-0.038766			
Adjusted R-squared	0.006441	S.D. dependent var	0.995827			
S.E. of regression	0.992615	Akaike info criterion	2.823954			
Sum squared resid	2182.406	Schwarz criterion	2.829100			
Log likelihood	-3128.353	Durbin-Watson stat	2.003960			
Inverted AR Roots	.22+.33i	.22 -.33i	-.36			
Date: 07/03/05 Time: 20:57						
Sample: 4 2220						
Included observations: 2217						
Q-statistic						
probabilities						
adjusted for 2						
ARMA term(s)						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
*	*	1	0.109	0.109	26.325	
*	*	2	0.128	0.117	62.682	
*	*	3	0.083	0.059	77.973	0.000
*	*	4	0.108	0.082	103.75	0.000
		5	0.039	0.006	107.10	0.000
		6	0.062	0.033	115.52	0.000
*	*	7	0.075	0.052	128.16	0.000
		8	0.049	0.018	133.49	0.000
*	*	9	0.087	0.062	150.23	0.000
		10	0.060	0.028	158.39	0.000
*	*	11	0.107	0.073	183.80	0.000
*	*	12	0.082	0.045	198.77	0.000
		13	0.049	0.000	204.05	0.000
		14	0.055	0.018	210.92	0.000
		15	0.058	0.020	218.42	0.000

Tabela 5.2.20: Ajuste do modelo ARMA(3,0) + GARCH(1,1)

```

> tnlp42hdes.mod1=garch(formula.mean = RetDes ~ ar(3), formula.var = ~ garch(1, 1))
> summary(tnlp42hdes.mod1)
Call:
garch(formula.mean = RetDes ~ ar(3), formula.var = ~ garch(1, 1))
Mean Equation: RetDes ~ ar(3)
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: gaussian
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	-0.02356	0.019555	-1.205	1.142e-001
AR(1)	0.06198	0.021205	2.923	1.752e-003
AR(2)	0.02656	0.022143	1.200	1.152e-001
AR(3)	-0.06717	0.022095	-3.040	1.196e-003
A	0.02052	0.006153	3.334	4.346e-004
ARCH(1)	0.04484	0.006630	6.764	8.577e-012
GARCH(1)	0.93398	0.010579	88.289	0.000e+000

```

-----
AIC(7) = 6123.154
BIC(7) = 6163.091
Normality Test:
-----
Jarque-Bera P-value
[1,] 229.6 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 3.855 0.9859 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 9.744 0.6384 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12
C
[1,] 0.06738 -0.02677 1.472 -2.208 0.02386 -0.06086 -1.105 0.1269 -0.2707 0.8788 0.8466
0.07063 0.7459
TR^2 P-value F-stat P-value
[1,] 10.02 0.6143 0.9149 0.6375

```

Gráfico 5.2.22: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

$$\text{GARCH}(1,1) + \text{ARMA}(3,0)$$

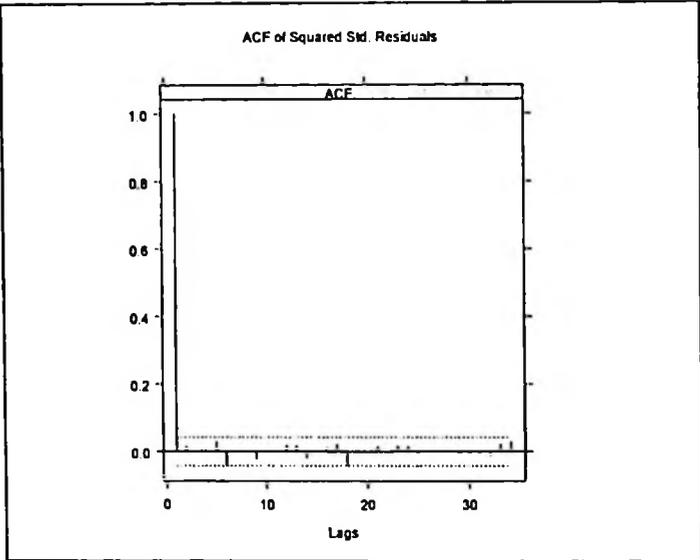


Gráfico 5.2.23: QQ-Plot dos Resíduos para Modelo Ajustado GARCH(1,1) + ARMA(3,0)

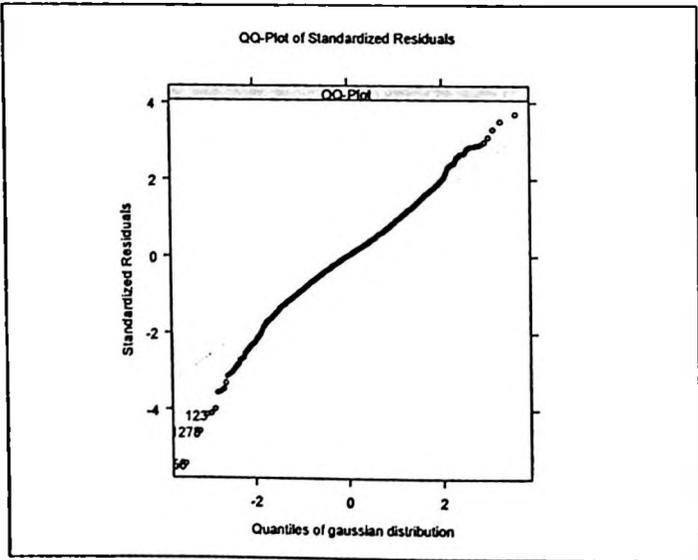


Tabela 5.2.21: Ajuste do modelo ARMA(3,0) + GARCH(1,1) com Student

```

> tnlp42hdes.mod11=garch(RetDes~ar(3),~garch(1,1), cond.dist="t", trace=F)
> summary(tnlp42hdes.mod11)
Call:
garch(formula.mean = RetDes ~ ar(3), formula.var = ~ garch(1, 1), trace = F, cond.dist = "t")
Mean Equation: RetDes ~ ar(3)
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: t
with estimated parameter 5.737651 and standard error 0.7988007
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	-0.01327	0.01835	-0.7233	2.348e-001
AR(1)	0.04857	0.02073	2.3427	9.618e-003
AR(2)	0.01611	0.02102	0.7663	2.218e-001
AR(3)	-0.06539	0.02090	-3.1288	8.891e-004
A	0.02929	0.01107	2.6470	4.090e-003
ARCH(1)	0.06399	0.01372	4.6640	1.643e-006
GARCH(1)	0.90904	0.01944	46.7568	0.000e+000

```

-----
AIC(8) = 6022.537
BIC(8) = 6068.179
Normality Test:
-----
Jarque-Bera P-value
[1,] 259.1 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 5.877 0.9222 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 10.56 0.5666 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12 C
[1,] -0.5018 -0.603 0.9452 -2.437 -0.324 -0.5421 -1.413 -0.22 -0.4522 0.5268 0.8135 0.07458
0.1368
TR^2 P-value F-stat P-value
[1,] 10.78 0.5474 0.9853 0.57

```

Gráfico 5.2.24: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

GARCH(1,1) + ARMA(3,0) com Dist. Student

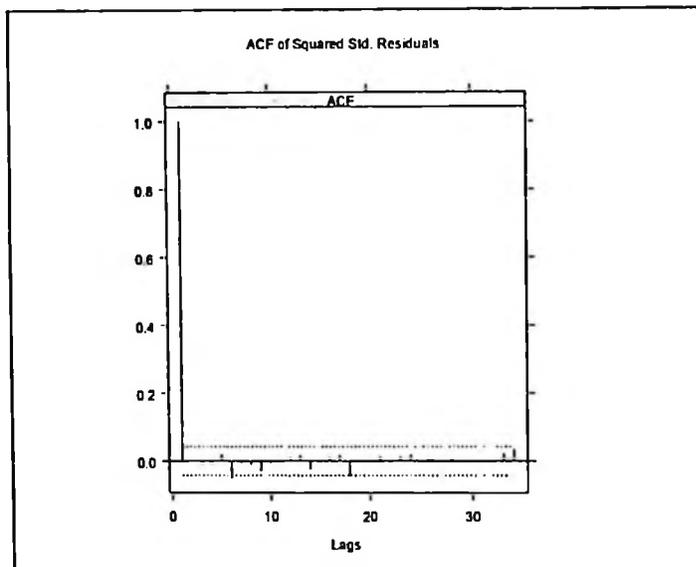


Gráfico 5.2.25: QQ-Plot dos Resíduos para GARCH(1,1) + ARMA(3,0) com Dist. Student

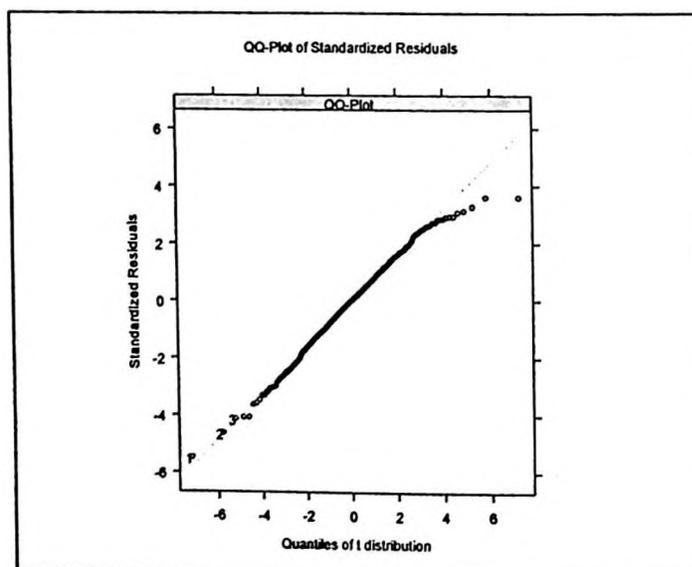


Tabela 5.2.22: Ajuste do modelo ARMA(3,0) + GARCH(1,1) com GED

```

> tnlp42hdes.mod12=garch(RetDes~ar(3),~garch(1,1), cond.dist="ged", trace=F)
> summary(tnlp42hdes.mod12)
Call:
garch(formula.mean = RetDes ~ ar(3), formula.var = ~ garch(1, 1), trace = F, cond.dist = "ged")
Mean Equation: RetDes ~ ar(3)
Conditional Variance Equation: ~ garch(1, 1)
Conditional Distribution: ged
with estimated parameter 1.321384 and standard error 0.05305524
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	-0.01179	0.01769	-0.6667	2.525e-001
AR(1)	0.04538	0.02002	2.2670	1.174e-002
AR(2)	0.01539	0.02020	0.7620	2.231e-001
AR(3)	-0.07017	0.02024	-3.4674	2.678e-004
A	0.02562	0.01034	2.4786	6.633e-003
ARCH(1)	0.05496	0.01196	4.5933	2.304e-006
GARCH(1)	0.91861	0.01845	49.8009	0.000e+000

```

-----
AIC(8) = 6020.207
BIC(8) = 6065.849
Normality Test:
-----
Jarque-Bera P-value
[1,] 247.9 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 6.241 0.9034 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,]
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12
C
[1,] -0.2517 -0.3844 1.14 -2.334 -0.2008 -0.3515 -1.289 -0.08077 -0.3893 0.6627 0.801 0.07419
0.3597
TR^2 P-value F-stat P-value
[1,] 9.913 0.6236 0.9053 0.647

```

Gráfico 5.2.26: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

GARCH(1,1) + ARMA(3,0) com Dist. GED

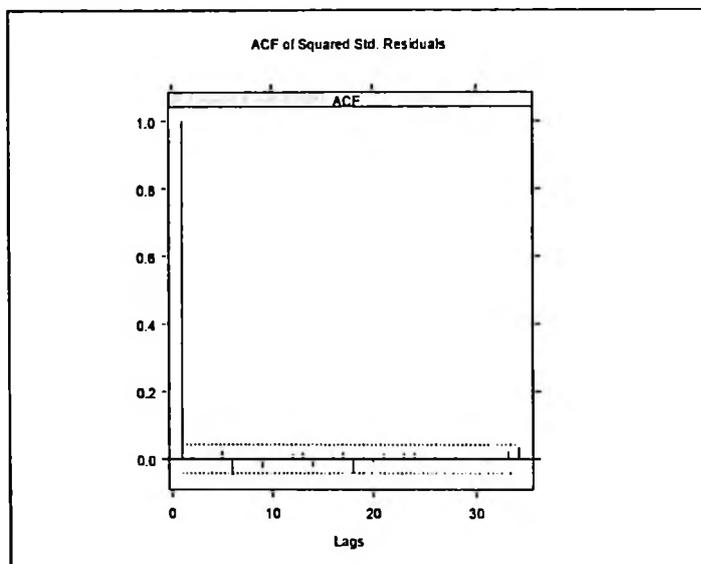


Gráfico 5.2.27: QQ-Plot dos Resíduos para GARCH(1,1) + ARMA(3,0) com Dist. GED

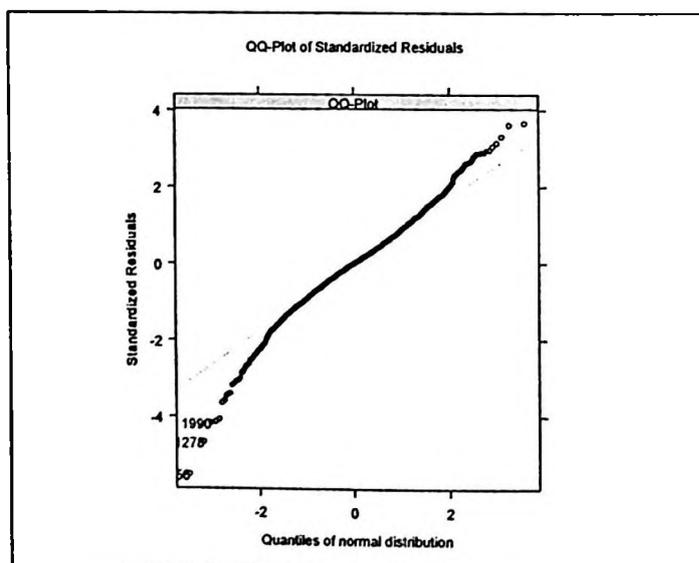


Tabela 5.2.23: Ajuste do modelo ARMA(3,0) + FIGARCH(1,1)

```

tnlp42hdes.figarch=figarch(RetDes~ar(3),~figarch(1,1))
> summary(tnlp42hdes.figarch)
Call:
figarch(formula.mean = RetDes ~ ar(3), formula.var = ~ figarch(1, 1))
Mean Equation: RetDes ~ ar(3)
Conditional Variance Equation: ~ figarch(1, 1)
-----
Estimated Coefficients:
-----

```

	Value	Std.Error	t value	Pr(> t)
C	-0.02509	0.01959	-1.281	1.002e-001
AR(1)	0.06887	0.02282	3.018	1.287e-003
AR(2)	0.02612	0.02176	1.200	1.151e-001
AR(3)	-0.06694	0.02174	-3.079	1.051e-003
A	0.03891	0.01573	2.473	6.738e-003
GARCH(1)	0.69139	0.05301	13.043	0.000e+000
ARCH(1)	0.46898	0.05395	8.692	0.000e+000
fraction	0.32781	0.05608	5.845	2.903e-009

```

-----
AIC(8) = 6125.549
BIC(8) = 6171.191
Normality Test:
-----
Jarque-Bera P-value
[1,] 243.7 0
Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 3.658 0.9888 12
Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
[1,] 10.62 0.5616 12
Lagrange multiplier test:
-----
Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9 Lag 10 Lag 11 Lag 12 C
[1,] 0.2918 0.09541 1.634 -2.116 0.04347 -0.2171 -1.116 0.1116 -0.3647 0.7428 1.001 0.06978 -
0.7217
TR^2 P-value F-stat P-value
[1,] 10.66 0.558 0.9741 0.5805

```

Gráfico 5.2.28: Função de Auto-correlação dos Resíduos ao Quadrado para Modelo Ajustado

FIGARCH(1,1) + ARMA(3,0)

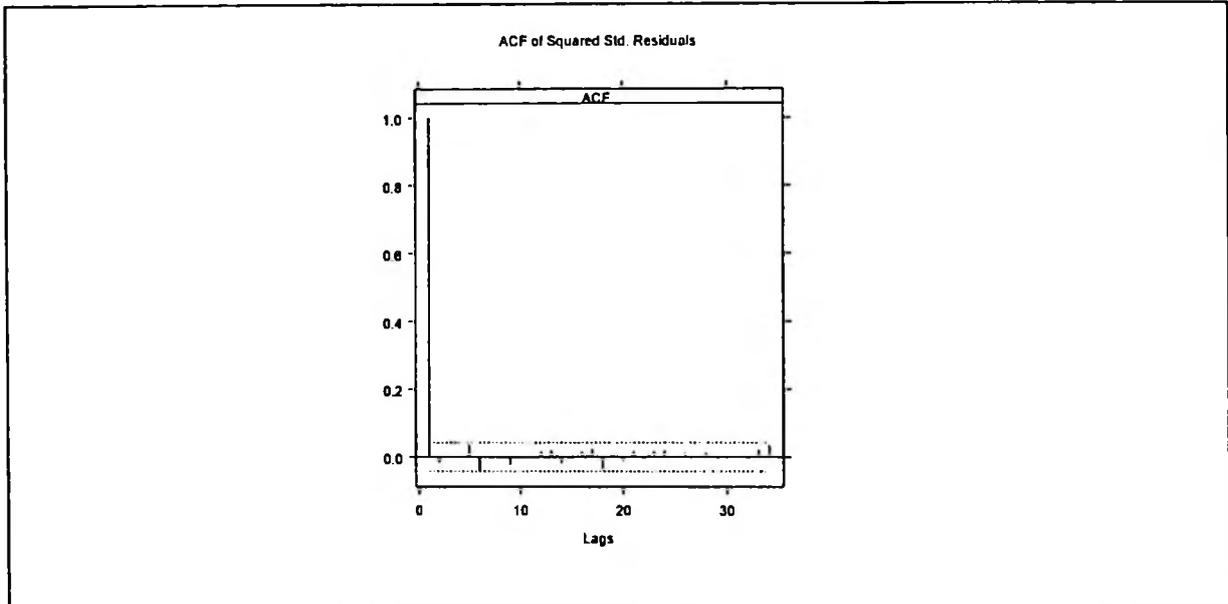


Gráfico 5.2.29: QQ-Plot dos Resíduos para FIGARCH(1,1) + ARMA(3,0)

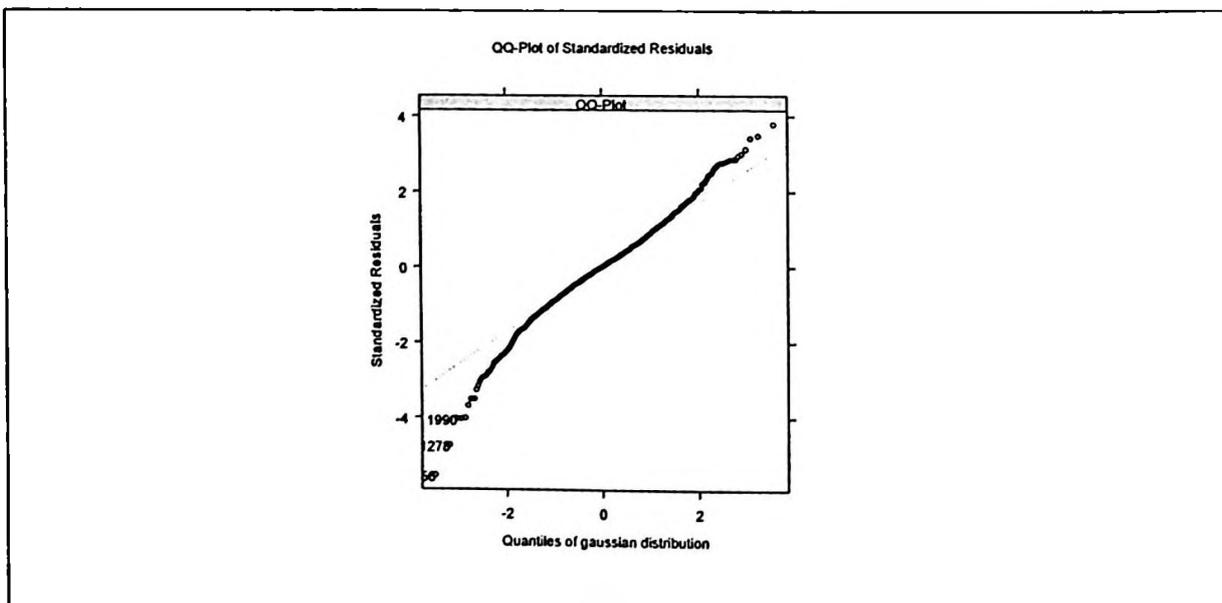


Tabela 5.2.24: Parâmetros de Memória Longa para Série de Retornos de 2 horas

Estimação de Parâmetro de Memória Longa: Valores para d entre 0,14 e 0,35

```
> d.ros(abs(RetDes))  
[1] 0.1430558  
> d.pgram(abs(RetDes))  
[1] 0.3576798
```

5.3 AJUSTE DE MODELO DE REGRESSÃO PARA VOLATILIDADE REALIZADA UTILIZANDO-SE VOLATILIDADE IMPLÍCITA COMO VARIÁVEL EXPLICATIVA

O objetivo dessa seção é ajustar um modelo de regressão para prever volatilidade de 21 dias úteis utilizando a volatilidade realizada histórica e a volatilidade implícita calculada a partir de série de preços de opção de compra "at-the-money" de TNLP4 de 21 dias de prazo de vencimento.

O modelo usado foi:

$$Vol_{realiz}(t) = \alpha + \beta Vol_{implicita}(t) + \varepsilon(t) ,$$

Onde:

$Vol_{realiz}(t)$ é a medida de volatilidade realizada no prazo (t) que é de 21 dias úteis.

$Vol_{implicita}(t)$ é a volatilidade calculada a partir dos preços das opções de compra ATM de prazo de vencimento $t=21$ dias úteis, utilizando-se a metodologia de Black-Scholes.

Utilizaram-se as seguintes séries de 5 minutos de frequência para ajustar o modelo para o período de 19 de Outubro de 2004 a 31 de Março de 2005:

- Série de volatilidade realizada histórica de 21 dias úteis, calculada a partir de dados intra-diários de 5 minutos de TNLP4, corrigida pelo fator de auto-correlação;
- Série intra-diária de 5 minutos de volatilidade implícita de opções de compra de TNLP4 "at the money" de 21 dias úteis de vencimento.

Resultado da Regressão

O modelo de regressão tendo volatilidade implícita de 21 dias como variável explicativa apresentou resultado apenas razoável e semelhante aos resultados apresentados pelos modelos estudados por Jihghong Shu e Jin E. Zhang. Ver tabela 5.3.1.

Tabela 5.3.1: Ajuste do Modelo de Regressão Volatilidade Histórica vs Volatilidade Implícita

```
tnlp4v=importData("vol_20d_0503.txt",type="ASCII",colNames=c("Date","volhist","volimpl"))
> attach(tnlp4v)

> tnlp4.fit1=OLS(volhist~volimpl)
> summary(tnlp4.fit1)
Call:
OLS(formula = volhist ~ volimpl)
Residuals:
  Min   1Q Median   3Q   Max
-0.8104 -0.0470 -0.0239  0.0465  0.2242

Coefficients:
              Value      Std. Error    t value    Pr(>|t|)
(Intercept) -0.0326      0.0070     -4.6346    0.0000
volimpl      1.2191      0.0240     50.7676    0.0000

Regression Diagnostics:
  R-Squared      0.2276
Adjusted R-Squared 0.2275
Durbin-Watson Stat 0.1284

Residual Diagnostics:
  Stat P-Value
Jarque-Bera 49967.07      0.00
Ljung-Box 271908.46      0.00

Residual standard error: 0.06823 on 8746 degrees of freedom
F-statistic: 2577 on 1 and 8746 degrees of freedom, the p-value is 0
```

6 ESTIMAÇÃO DE VOLATILIDADE REALIZADA E AVALIAÇÃO DO MODELO

Esse capítulo trata da estimação de volatilidade realizada de 1 dia útil e de 21 dias úteis para o período de 03 de Janeiro de 2005 a 30 de Junho de 2005, a partir de modelos GARCH ajustados na seção 5 para o período de 02 de Janeiro de 2002 a 30 de Dezembro de 2004. Optou-se pela frequência de 2 horas para as projeções com uso de dados de alta frequência na medida em que, para a projeção com a janela de 21 dias, a janela ensejaria muitos passos à frente para dados de maior frequência, o que impactaria significativamente a precisão do modelo.

Além disso, é também avaliado o modelo de regressão ajustado no capítulo 5 para previsão de volatilidade de 21 dias úteis, utilizando a volatilidade implícita de opção de 21 dias úteis de vencimento como variável explicativa. A período de previsão foi de Abril de 2005 a Junho de 2005.

Utilizaram-se as seguintes medidas para avaliação dos modelos :

- Raiz Quadrada do Erro Quadrado Médio (RMSE)

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (h_i^2 - \sigma_i^2)^2}$$

- Erro Absoluto Médio (MAE)

$$MAE = \frac{1}{T} \sum_{i=1}^T |h_i^2 - \sigma_i^2|$$

6.1 ESTIMAÇÃO DE VOLATILIDADE REALIZADA DE 1 DIA ÚTIL E DE 21 DIAS ÚTEIS UTILIZANDO-SE DADOS DIÁRIOS

A partir do modelo selecionado na seção 5 para dados diários, o GARCH(1,1) sem modelagem de média, projetou-se a volatilidade de 1 dia útil e de 21 dias úteis para o período de 03 de Janeiro de 2005 a 30 de Junho de 2005, perfazendo 120 dias úteis de projeção.

Projeção de Volatilidade de 1 dia útil. A raiz quadrada do erro quadrático médio (RMSE) das projeções de volatilidade de 1 dia útil em relação às volatilidades realizadas no período foi de 15,50% sendo que o erro absoluto médio (MAE) foi de 13,73%.

Projeção de Volatilidade de 21 dias úteis. A raiz quadrada do erro quadrático médio (RMSE) das projeções de volatilidade de 21 dias em relação às volatilidades realizadas no período foi de 7,69% sendo que o erro absoluto médio (MAE) foi de 6,56%.

6.2 ESTIMAÇÃO DE VOLATILIDADE REALIZADA DE 1 DIA ÚTIL E DE 21 DIAS ÚTEIS UTILIZANDO-SE DADOS INTRA-DIÁRIOS DE FREQUÊNCIA DE 2 HORAS

A partir dos modelos selecionados na seção 5 para a frequência de 2 horas: (i) GARCH(1,1); (ii) GARCH(1,1) com distribuição de Student; (iii) GARCH(1,1) com distribuição GED; (iv) FIGARCH(1,1) e modelagem de média ARMA(3,0), são obtidas as projeções de volatilidade de 1 dia útil e de 21 dias úteis para o período de 03 de Janeiro de 2005 a 30 de Junho de 2005.

Essa seção trata da previsão de volatilidade realizada, na frequência de 2 horas, k passos à frente, sendo k igual a 3 (3 observações por dia para a frequência de 2 horas) de modo a cobrir o período de um dia.

Ajustou-se o modelo para a janela inicial de 2.220 dados de retornos dessazonalizados (740 dias, referentes ao período de 02 de Janeiro de 2002 a 30 de Dezembro de 2004, vezes 3 observações por dia) e foi realizada a previsão $k=3$ passos à frente (um dia). Para cada novo dia da projeção (3 passos) reajustou-se o modelo mantendo a janela inicial de 2.220 dados, ou seja, excluindo o primeiro dia da série e incorporando-se o próximo dia do período de projeção. Dessa forma, calcularam-se as projeções de volatilidade realizada de 1 dia útil e de 21 dias úteis para todo o período de projeção (Janeiro/05 a Junho/05), mantendo-se o mesmo tamanho de janela e incorporando-se dados mais recentes.

Ao final de cada dia, as previsões foram agregadas para obter a volatilidade diária.

$$h_t^2(k) = \sum_{n=1}^k h_{t,n}^2,$$

onde: $h_t(k)$ é a estimativa da volatilidade realizada k períodos à frente, dado a informação até o instante t e $h_{t,n}(k)$ é a estimativa da volatilidade realizada para o horizonte n dado a informação até o instante t .

A previsão para a volatilidade k passos à frente, a partir da informação no instante t , é obtida através do seguinte somatório que utiliza pesos da representação de um processo ARMA (p,q) e da projeção da volatilidade obtida através do modelo GARCH usado:

$$h_t^2(k) = \sum_{p=1}^k \left(\sum_{j=0}^{k-p} \psi_j \right)^2 h_{t,p}^2,$$

em que ψ_j são pesos da representação de um processo ARMA. No presente modelo utilizaram-se pesos para o processo ARMA(3,0).

O termo $h_{t,p}^2$ já é a projeção da volatilidade sazonalizada para cada dia à frente, após re-introduzir a sazonalidade através da fórmula:

$$h_t = h_{t,d} * \phi_d,$$

onde ϕ_d é o fator sazonal para cada um dos intervalos de 2 horas. Após calcular-se a projeção de volatilidade sazonalizada para cada dia à frente (projeção de 1 dia útil), somou-se a volatilidade para janelas consecutivas de 21 dias úteis de projeção de modo a obter projeções de volatilidade realizada para 21 dias úteis.

Projeção de Volatilidade de 1 dia útil para Série de 15 minutos

Projeção de Volatilidade de 1 dia útil. A projeção de volatilidade de 1 dia útil utilizando os modelos da família GARCH para dados de 2 horas apresentou resultados melhores do que aqueles do modelo com dados diários. O modelo FIGARCH(1,1) + AR(3) apresentou o melhor comportamento, medido pelos critérios da raiz quadrada do erro quadrático médio (RMSE) e do erro absoluto médio (MAE) das projeções de volatilidade de 1 dia útil em relação às volatilidades realizadas.

	RMSE	MAE
GARCH(1,1) + AR(3)	14,85%	10,55%
GARCH(1,1) dist. St. + AR(3)	14,80%	10,52%
GARCH(1,1) dist GED + AR(3)	14,83%	10,53%
FIGARCH(1,1) + AR(3)	13,71%	9,58%

Projeção de Volatilidade de 21 dias úteis para Série de 15 minutos

Projeção de Volatilidade de 21 dias úteis. A projeção de volatilidade de 21 dias úteis utilizando os modelos da família GARCH para dados de 2 horas apresentou resultados piores do que aqueles do modelo com dados diários. O modelo FIGARCH(1,1) + AR(3) apresentou o melhor desempenho, inclusive com resultados comparáveis aqueles de modelos utilizando dados diários; esse resultado foi obtido a partir da utilização dos critérios da raiz quadrada do erro quadrático médio (RMSE) e do erro absoluto médio (MAE) das projeções de volatilidade de 21 dias úteis em relação às volatilidades realizadas. Esse resultado evidencia o fato de que o modelo com dados intra-diários perde eficiência para janelas de projeção maiores.

	RMSE	MAE
GARCH(1,1) + AR(3)	12,17%	10,86%
GARCH(1,1) dist. St. + AR(3)	12,80%	11,30%
GARCH(1,1) dist GED + AR(3)	13,01%	11,50%
FIGARCH(1,1) + AR(3)	9,74%	8,68%

6.3 ESTIMAÇÃO DE VOLATILIDADE REALIZADA DE 21 DIAS ÚTEIS UTILIZANDO-SE MODELO DE REGRESSÃO COM VOLATILIDADE IMPLÍCITA COMO VARIÁVEL EXPLICATIVA

A projeção de Volatilidade de 21 dias úteis para o período de Abril de 2005 a Junho de 2005, utilizando o modelo de regressão tendo a volatilidade implícita como variável explicativa, apresentou bom desempenho quando avaliados os critérios da raiz quadrada do erro quadrático médio (RMSE) e do erro absoluto médio (MAE). Os indicadores de desempenho são comparáveis aqueles obtidos com o modelo de dados diários e com o modelo FIGARCH(1,1) para dados intra-diários.

	RMSE	MAE
$Vol_{realiz}(t) = -0.032 + 1,219 * Vol_{implicita}(t) + \varepsilon(t)$	8,29%	7,39%

7 CONCLUSÃO

A análise de fatos estilizados de séries de dados de alta frequência indicou que se faz necessário o tratamento de questões como auto-correlação e sazonalidade antes de ajustar modelos para séries intra-diárias. Foi realizado o tratamento de sazonalidade antes do ajuste de modelos de séries temporais para a série de retornos de alta frequência bem como o tratamento de auto-correlação para construção da série de dados de alta frequência de volatilidade implícita.

Os modelos utilizando séries temporais com dados de alta frequência apresentaram bom ajuste sendo que o desempenho das projeções de volatilidade realizada de TNLP4 foi melhor de que o desempenho de modelos utilizando dados diários para a projeção de volatilidade de 1 dia útil.

A projeção de volatilidade realizada de TNLP4 para 21 dias úteis utilizando dados de alta frequência apresentou pior desempenho do que a projeção de volatilidade de 1 dia útil. O modelo FIGARCH(1,1) com dados intra-diários apresentou desempenho comparável aquele utilizando dados diários. Esse resultado demonstra que as projeções utilizando dados intra-diários devem estar, em geral, limitadas a poucos passos à frente.

A desempenho do modelo de regressão utilizando dados intra-diários de volatilidade implícita de 21 dias úteis para prever volatilidade realizada de 21 dias úteis apresentou resultado comparável aquele obtido em outros trabalhos. A desempenho desse modelo pode ter sido prejudicado pela pequena amostra de dados usada para ajustar o modelo.

8 BIBLIOGRAFIA

Andersen T. e Bollerslev T. (1997 e 1998). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*.

Andersen T. , Bollerslev T., Diebold F., Labys P. (2001). Modeling and Forecasting Realized Volatility.

Bandi F. e Russell J. (2003). Microstructure noise, realized volatility, and optimal sampling.

Barndorff-Nielsen O., Shephard N. (2000) Econometric analysis of realized volatility and its use in estimating stochastic volatility models .

Berti, A. (2005). Estimação da Volatilidade Diária com Dados de Alta Frequência: Aplicação ao Cálculo do Valor em Risco do Ibovespa. Dissertação apresentada ao IME. Orientador: Prof. Dr. Pedro A. Morettin.

Corsi, F. , Zumbach G., Müller U., Dacorogna M. (2001). Consistent High-Precision Volatility from High Frequency Data.

Dacorogna M., Gençay R., Müller U., Olsen R. e Pictet O. An Introduction to High-Frequency Finance (AP Academic Press).

Davidian M. e Carroll R.J. (1987). Variance function estimation, *Journal of American Statistical Association*, 82, 1079-1091.

Day T.E. e Lewis CM. (1993). Forecasting futures market volatility, *Journal of Derivatives*, 1, 33-50.

Durbin J. e Koopman S.J. (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives, *Journal of Royal Statistical Society Series*, 62, 1, 3-56.

Ederington L.H. e Guan W. (2000a). Forecasting volatility, working paper, University of Oklahoma.

Ederington L.H. e Guan W. (2000b). Measuring implied volatility. Is an average better? working paper, University of Oklahoma.

Feinstein Steven P. (1989b). Forecasting stock market volatility using options on index futures, *Economic Review*, Federal Reserve Bank of Atlanta, 74, 3, 12-30.

Fung W.K.H. e Hsieh D. A. (1991). Empirical Analysis of Implied Volatility: Stocks, Bonds and Currencies, Working paper, Department of Finance, Fuqua School of Business.

Giot, P. (2000). Intraday value-at-risk. Department of Quantitative Economics, Maastricht University (Holland) and Center for Operations Research and Econometrics (CORE), UCL (Belgium)

Hansen, P.H. and Lunde, A. (2004). A forecast comparison of volatility models: does anything beat a GARCH(1,1)? , *Journal of Applied Econometrics*.

Harvey C.R., Ruiz E. e Shepard N. (1994). Multivariate stochastic variance models, *Review of Economic Studies*, 61, 247-264.

Hol, E. and S. J. Koopman (2002). Stock index volatility forecasting with high frequency data, discussion paper no. 2002-068/4, Tinbergen Institute.

Martens M. e Zein J. (2003). Predicting financial volatility: high-frequency time-series forecasts vis-à-vis implied volatility, working paper, Erasmus University.

Moreira J. M. (2002), O Uso de Dados de Alta Frequência na Estimação de Volatilidade e Valor em Risco para o Ibovespa. Orientador: Prof. Eduardo Facó Lengruber. UFRJ.

Morettin, P.A. (2004). Econometria Financeira: Um Curso de Séries Temporais Financeiras. IME-USP.

Oomen R. (2001). Using High Frequency Stock Market Index Data to Calculate, Model & Forecast Realized Volatility.

Pong, S, Shackleton, B. M., Taylor, S.J. And Xu, X. (2004). Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models, Journal of Banking & Finance, 28, 2541-2563.

Poon S. e Granger C. (2002). Forecasting Volatility in Financial Markets: A Review.

Poteshman, Allen (2000). Forecasting Future Volatility from Option Prices.

Shu, J., The Relation between Implied and Realized Volatility of S&P500 Index, School of International Trade and Economics, University of International Business & Economics, Beijing, China.

Taylor e Xu (1997). The incremental volatility information in one million foreign exchange quotations. Journal of Empirical Finance 4, 317-340

Tsay R., Analysis of Financial Time Series. John Wiley & Sons, Inc.

Zivot, E. and J. Wang. Modeling Financial Time Series with S-Plus. New York: Springer-Verlag.

**BIBLIOTECA FEA/USP
DOAÇÃO**

Procedência PÓS-GRADUAÇÃO

Data 13 / 04 / 09