

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO  
DEPARTAMENTO DE ECONOMIA

MONOGRAFIA DE FINAL DE CURSO

**A VOLATILIDADE REALIZADA ROBUSTA À RUIDOS DE  
MICROESTRUTURA: UMA APLICAÇÃO AO ÍNDICE S&P 500**

Roberto Pougy Ferreira da Cunha

Matrícula: 05-1-1990

Orientador: Marcelo Cunha Medeiros

Novembro de 2007

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO  
DEPARTAMENTO DE ECONOMIA

MONOGRAFIA DE FINAL DE CURSO

**A VOLATILIDADE REALIZADA ROBUSTA À RUÍDOS DE MICROESTRUTURA:  
UMA APLICAÇÃO AO ÍNDICE S&P 500**

“Declaro que o presente trabalho é de minha autoria e que não recorro para realizá-lo a nenhuma forma de ajuda externa, exceto quando autorizado pelo professor tutor”.

---

Roberto Pougy Ferreira da Cunha

Matrícula: 05-1-1990

Orientador: Marcelo Cunha Medeiros

Novembro de 2007

“As opiniões expressas neste trabalho são de responsabilidade única e exclusiva do autor”

## **AGRADECIMENTOS**

Dedico este trabalho a Sergio Henrique Ferreira da Cunha, Emilio Lèbre La Rovere e Mario Veiga Ferraz Pereira, cujas vidas profissionais desde sempre têm servido de inspiração diária em minha incipiente carreira profissional.

Gostaria de agradecer a Paulo Vitor Costa de Carvalho, Felipe Figueiredo Pires Del Negro, Paulo Mattosinho Neto e Quemil Manasfi; amigos de infância do Colégio de São Bento cujo convívio tanto motiva à superação dos obstáculos do dia a dia.

Gostaria de agradecer ao professor Marcelo Cunha Medeiros, orientador deste trabalho, aos demais professores do Departamento de Economia da PUC-Rio e aos professores do Departamento de Economia da Universidade da Califórnia, San Diego, em especial à Marina Figueira de Mello, Márcio Scalécio, Márcio Gomes Pinto Garcia, João Barbosa de Oliveira, Sérgio Volchan, Magdalena Sokalska, David Miller, James Andreoni e Marc Andreas Muendler, por terem sido os pivôs da minha paixão pela economia.

Por fim, não posso deixar de agradecer à minha família, em especial a meus pais Sergio e Alice e à minha irmã Bruna, dos quais sem o apoio e compreensão nestes últimos meses este trabalho não teria sido concluído.

## ÍNDICE

I.	INTRODUÇÃO .....	7
II.	ESTIMAÇÃO DA VOLATILIDADE REALIZADA SOB EFEITOS DE MICROESTRUTURA .....	9
	A. O conceito de variância realizada .....	9
	B. Um modelo em tempo contínuo .....	9
	C. Ruídos de microestrutura.....	11
	D. Contornando os ruídos de microestrutura 1 – O estimador ZMA (2005) .....	13
	E. Contornando os ruídos de microestrutura 2 – O estimador BHLS (2006).....	14
III.	DISTRIBUIÇÃO DOS RETORNOS PADRONIZADOS DO ÍNDICE S&P 500 .....	16
	A. Dados .....	16
	B. A distribuição dos retornos padronizados .....	18
IV.	MODELAGEM E PREVISÃO COM ESTIMADORES ZMA e BHLS.....	23
	A. Resultados dentro da amostra .....	23
	B. Resultados fora da amostra.....	27
V.	CONCLUSÃO .....	32
VI.	REFERÊNCIAS BIBLIOGRÁFICAS .....	33

## ÍNDICE DE TABELAS

Tabela 1: Estatísticas descritivas dos retornos diários do índice S&P 500.

Tabela 2: Estatísticas descritivas dos retornos padronizados.

Tabela 3: Testes de "boa aproximação" para os retornos padronizados.

Tabela 4: Estatísticas descritivas das séries de volatilidades realizadas.

Tabela 5: Testes de "goodness of fit" para a variação da volatilidade.

Tabela 6: Estimativas para o estimador de diferenciação fracional GPH.

Tabela 7: Funções de perda - dentro da amostra.

Tabela 8: Funções de perda modificadas – dentro da amostra.

Tabela 9: Testes de cobertura - dentro da amostra

Tabela 10: Funções de perda - fora da amostra

Tabela 11: Testes de cobertura - Fora da amostra

## I. INTRODUÇÃO

O termo volatilidade, no mundo das finanças, refere-se à medida estatística da dispersão dos retornos de um ativo ou índice financeiro. É uma medida do *risco* associado a cada ativo financeiro. É de grande interesse para economistas, operadores e investidores em geral o desenvolvimento de medidas apuradas e confiáveis da volatilidade como um instrumento no apoio à decisão.

Historicamente, a principal ferramenta desenvolvida para análise e previsão da volatilidade, cuja utilização é disseminada nos mercados financeiros mundiais, é a classe de modelos “Auto-Regressivos de Heteroscedasticidade Condicional” (ARHC), com respeito ao modelo ARHC desenvolvido por Engle (1982) e aprimorado para o modelo ARHCG por Bollerslev (1986). Esta classe de modelos assume o excesso do retorno à média, em cada período de tempo, como normalmente distribuído, com média zero e uma variância heteroscedástica, sendo esta descrita por um processo auto-regressivo de defasagem distribuída - ADD(p,q).

O ponto fraco da metodologia ARCH é no entanto, ser paramétrica. Isto implica na impossibilidade de se *observar* a volatilidade diretamente, podendo esta apenas ser estimada. Por tratar-se de um modelo que estima a volatilidade de um período com base na observada em períodos anteriores, ficamos obrigados à imposição de uma condição inicial arbitrária.

Neste sentido, o conceito de volatilidade realizada é a grande novidade na pesquisa econométrica em volatilidade de séries temporais financeiras. Seu grande avanço em relação aos modelos do tipo ARHC foi justamente tornar possível a observação da volatilidade, sem a dependência de um parâmetro. Tal avanço foi possibilitado pela disponibilidade, há aproximados 12 anos, de séries temporais intradiárias de ativos financeiros.

Ao dividirmos um pregão em, por exemplo, 84 intervalos de 5 minutos, e assumirmos que em cada intervalo os retornos são bem descritos por variáveis aleatórias independentes, de mesma distribuição e média igual a zero, temos que o retorno para o dia pode ser calculado como o somatório dos retornos dos intra-períodos. A volatilidade realizada é justamente a soma, para cada intra-período, dos retornos ao quadrado. Logo temos um estimador da volatilidade tão melhor quanto maior for a frequência de observações durante o pregão. No limite, gostaríamos de construir nosso estimador a partir de observações contínuas, ou no jargão, *tick by tick*.

Há, contudo, um problema na contramão deste novo conceito. Os chamados “efeitos de microestrutura” - características dos sistemas de negociação, como por exemplo *spreads* entre preços demandados e ofertados – introduzem maior viés na estimação da volatilidade quanto maior for a frequência de observações. Assim sendo, a variação entre dois preços consecutivos deixa de refletir exclusivamente uma real variação no valor do ativo.

O objetivo deste trabalho será aplicar estimadores robustos a ruídos de microestrutura recém desenvolvidos por Zhang, Mykland e Aït-Sahalia (2005), a partir daqui referido por ZMA (2005), e por Barndorff-Nielsen, Hansen, Lunde e Shephard (2006), referido por BHLS (2006), à séries temporais de aproximadamente 10 anos de observações do índice de ações norte-americano S&P 500.

Com isso investigaremos se a distribuição dos retornos padronizados pela volatilidade realizada tem distribuição normal ou log-normal, o que não acontecia nos modelos ARHCG, ARHCG-E e GJR. Em seguida proporemos um modelo ARFIMA para análise e previsão fora da amostra da série temporal de volatilidade realizada, podendo assim comparar seu desempenho de previsão com o alcançado pelos modelos tradicionais.

A seção II caracteriza o conceito de volatilidade realizada através de suas propriedades teóricas, diferenças para a modelagem ARHCG, consistência e eficiência na estimação da volatilidade. Em seguida detalha o impacto gerado pela presença de efeitos de microestrutura, apresentando em seguida as soluções propostas por ZMA (2005) e por BHLS (2006) para contorná-los.

A seção III apresenta uma análise empírica sobre a distribuição de probabilidade dos retornos logaritmizados do índice S&P 500. São confrontadas as distribuições dos retornos padronizados construídas a partir dos estimadores de volatilidade realizada ZMA (2005) e BHLS (2006) com os construídos a partir dos modelos ARHCG(1,1), ARHCG-E, GJR e MA(1)-ARHCG(1,1). Para tal, fazemos uso dos testes estatísticos apropriados - a saber, os de Jarque-Bera, Kolmogorov-Smirnov e Lilliefors - para a hipótese nula de que a distribuição é normal.

A seção IV realiza exercícios de pseudo previsão fora da amostra a partir das séries de volatilidades anteriormente analisadas. A capacidade de previsão dos modelos são comparadas entre si sob o critério de minimização da raiz do erro de previsão quadrático médio (REPQM) e do erro absoluto médio (EAM).

A seção V sintetiza as conclusões obtidas nas seções anteriores.

## II. ESTIMAÇÃO DA VOLATILIDADE REALIZADA SOB EFEITOS DE MICROESTRUTURA

Esta seção apresenta o modelo de tempo discreto no qual se baseia o estimador de volatilidade realizada. A sub-seção A introduz o conceito. A sub-seção B apresenta a síntese do modelo feita em Carvalho, Freire, Medeiros e Souza (2007). A sub-seção C introduz a modelagem dos ruídos de micro-estrutura. A sub-seção D apresenta os estimadores consistentes sob ruídos de micro-estrutura ZMA (2005). A sub-seção E apresenta os estimadores consistentes sob ruídos de micro-estrutura BHLS (2006).

### A. O conceito de variância realizada

A estimação da volatilidade de um ativo financeiro durante um período de tempo pode ser tão mais precisa quanto maior for a frequência de observação dos retornos, adotando as hipóteses de que retornos intra-períodos são não correlacionados e de média nula. Definimos um período  $t$  como um dia de negociação do ativo com retorno diário representado por  $r_t$ , este equivalente à soma de  $n_t$  retornos intradiários  $r_{t,i}$ :

$$r_t = \sum_{i=0}^{n_t} r_{t,i}$$

Definimos a variância realizada (*realized variance*) para o dia  $t$  como

$$RV_t^{(all)} = \sum_{i=0}^{n_t} r_{t,i}^2$$

e à sua raiz quadrada nos referimos como volatilidade realizada (1). O estimador  $RV_t^{(all)}$  equivale a  $n_t$  vezes a variância amostral de  $n_t$  retornos intra-diários, sendo portanto um estimador natural da variância do dia  $t$ .

### B. Um modelo em tempo contínuo

Considere que o processo de difusão representado pela equação diferencial estocástica

$$dp(t + \tau) = \mu(t + \tau) + \sigma(t + \tau)dW(t + \tau), \quad 0 \leq \tau \leq 1, t = 1, 2, 3 \dots,$$

onde  $p(t + \tau)$  representa o logaritmo do preço no momento  $t + \tau$ ,  $\mu(t + \tau)$  o componente de arrastamento,  $\sigma(t + \tau)$  a volatilidade instantânea ou desvio padrão e  $dW(t + \tau)$ , um movimento browniano padrão, representa o comportamento do logaritmo dos preços ao longo de um dia  $t$ . O componente de arrastamento  $\mu(t + \tau)$  é usualmente assumido constante.

Andersen, Bollerslev, Diebold e Labys (2003), e Barndorff-Nielsen e Shephard (2002) mostraram que os retornos diários compostos, definidos como  $r_t = p(t) - p(t - 1)$ , são normalmente distribuídos, condicionalmente ao conjunto de informação  $\mathcal{F}\{\mu(t + \tau - 1), \sigma(t + \tau - 1), 0 \leq \tau \leq 1\}$ , tal que

$$r_t | \mathcal{F}_t \sim N \left( \int_0^1 \mu(t + \tau) d\tau, \int_0^1 \sigma^2(t + \tau) d\tau \right)$$

Sob estas condições uma medida de interesse é a volatilidade *ex-post*, chamada *variância integrada* do dia  $t$  e definida como

$$IV_t = \int_0^1 \sigma^2(t + \tau) d\tau$$

Em aplicações práticas os preços são observados em intervalos discretos e irregularmente espaçados, havendo diversas maneiras de serem amostrados. Suponha que em um dado dia  $t$  particionemos o intervalo  $[0,1]$  em subintervalos e definamos a grade de observações

$$\mathcal{G} = \{\tau_0, \dots, \tau_n\}, 0 = \tau_0 \leq \tau_1 \leq \dots \leq \tau_{n-1} \leq \tau_n = 1$$

O tamanho do  $i$ -ésimo subintervalo é dado por  $\delta_i = \tau_i - \tau_{i-1}$ . Assumiremos que o tamanho de cada subintervalo convergirá para zero conforme o número de observações  $n$  aumentarem. A variância integrada para cada intra-período é definida como

$$IV_{i,t} = \int_{\tau_{i-1}}^{\tau_i} \sigma^2(t + \tau) d\tau$$

O mais popular sistema de amostragem utiliza um sistema de calendário como referência, fazendo os intervalos equidistantes no tempo-calendário, isto é  $\delta_i = 1/n$ . Este sistema será adotado no restante deste trabalho.

Definindo  $p_{i,t}$ ,  $t = 1, \dots, n$ , como o  $i$ -ésimo preço observado durante o dia  $t$ , tal que  $r_t = p_{t,i} - p_{t,i-1}$  seja o  $i$ -ésimo retorno intra-diário do dia  $t$ , podemos construir a variância e a volatilidade realizada como

$$r_t = \sum_{i=0}^{n_t} r_{t,i} \quad (1)$$

Sob algumas condições de regularidade adicionais, incluindo a hipótese de retornos intra-diários não-correlacionados, Andersen, Bollerslev, Diebold e Labys (2003) mostram que a variância realizada, ao usar toda a informação disponível, como definido em (1), é um estimador consistente da variância integrada, tal que  $RV_t \xrightarrow{p} IV_t$ . Barndorff-Nielsen e Shephard (2002) e Bandi e Russell (2005) derivaram a distribuição assintótica da variância realizada como

$$\sqrt{n} \frac{1}{\sqrt{2IQ_t}} (RV_t - IV_t) \xrightarrow{d} N(0,1)$$

onde a quarticidade integrada  $IQ_t$  é definida como  $IQ_t = \int_0^1 \sigma^4(t + \tau) d\tau$ .

A quarticidade integrada é, sob condições de regularidade similares, consistentemente estimada pela quarticidade realizada, definida por

$$RQ_t = \frac{1}{3} \sum_{i=1}^n r_{t,i}^4$$

No entanto tal análise não considera a presença de ruídos de micro-estrutura. Na prática isto significa um fator externo ao processo de formação de preços que introduz uma correlação entre retornos sub-seqüentes, sendo esta tão maior quanto mais freqüentes forem nossas observações de preços. O próximo tópico sumariza a incorporação destes efeitos em nossa modelagem.

### C. Ruídos de microestrutura

O comércio de ativos em mercados financeiros está organizado de maneiras diferentes em diversos países do mundo. As particularidades da microestrutura de cada mercado irá determinar a disponibilidade dos dados disponíveis.

Ativos podem ser negociados em contratos de troca (*exchange trade*) e contratos de balcão (*over the counter ou OTC*). O sistema de negociações pode ser tanto dirigido por ordens (*order-driven structure*) quanto dirigido por cotação (*quote-driven structure*), diferindo no método que estabelece o preço entre compradores e vendedores.

Cotações são emitidas por *market makers*, um agente preparado para ser tanto vendedor quanto comprador das operações, provendo liquidez ao mercado. Podemos ter um especialista monopolista ou muitos especialistas competindo entre si. Alternativamente, ordens podem ser casadas automaticamente por sistemas eletrônicos, removendo assim intermediários do processo de negociação. Mecanismos de negociação mudam ao longo do tempo e nos últimos anos houve um grande crescimento dos sistemas de negociação eletrônica com estruturas dirigidas por ordens.

No entanto todas estas características variam de mercado a mercado, de país a país e determinam as características das séries temporais de preços disponíveis para estudo e análise. Logicamente as variações nos preços representadas em tais séries não obrigatoriamente refletem variações reais *puras* no valor do ativo ou do conjunto de ativos, mas incluem um *viés de microestrutura*.

Para efeito de modelagem, consideraremos que os preços são observados tais que

$$p_{t,i} = p_{t,i}^* + \mathcal{E}_{t,i}$$

onde  $p_{t,i}^*$  é o preço verdadeiro ou eficiente e  $\mathcal{E}_{t,i}$  representa um ruído de microestrutura. Segue que

$$r_{t,i} = r_{t,i}^* + \varepsilon_{t,i} - \varepsilon_{t,i-1} = r_{t,i}^* + \nu_{t,i}$$

onde,  $r_{t,i}^*$  é o retorno verdadeiro ou eficiente. Resta claro a presença de autocorreção em  $r_{t,i}$ , de modo que  $RV_t$  torna-se um estimador viesado da volatilidade realizada, como podemos ver abaixo:

$$RV_t^{(all)} = \sum_{i=1}^{n_t} (r_{t,i}^*)^2 + \sum_{i=1}^{n_t} r_{t,i}^* \nu_{t,i} + \sum_{i=1}^{n_t} \nu_{t,i}^2$$

$$E(RV_t^{(all)} | r^*) = RV_t^{*(all)} + 2n_t E(\varepsilon_{t,i}^2)$$

Como é comum ao processo de modelagem econômica, a solução para o problema está na adoção de um conjunto apropriado de hipóteses simplificadoras que equilibrem o trade-off entre afastamento da realidade e manuseabilidade do problema.

Alguns conjuntos diferentes de hipóteses foram propostos. Com base em diferentes hipóteses, diferentes autores construíram estimadores não-viesados e consistentes da volatilidade realizada. As próximas sub-seções detalham dois destes estimadores.

#### D. Contornando os ruídos de microestrutura 1 – O estimador ZMA (2005)

Considere o seguinte conjunto de hipóteses com respeito a estrutura dos ruídos de microestrutura  $\mathcal{E}_{t,i}$ .

- a. O ruído de microestrutura  $\mathcal{E}_{t,i}$  é uma variável aleatória de média zero, independente e identicamente distribuída para todo  $t$  e para todo  $i$ .
- b. O processo de ruído é independente do processo de preços.
- c. A variância de  $v_{t,i}$  é  $O(1)$ .

ZMA (2005) mostra que sob estas hipóteses a volatilidade realizada  $RV_t^{(all)}$  converge em distribuição, quando multiplicada por um escalar apropriado, para a soma da variância integrada (nosso alvo), um choque com distribuição normal de média zero e um viés introduzido pelo ruído de micro estrutura equivalente à  $2n_t E(\mathcal{E}_{t,i}^2)$ . Logo quanto maior  $n_t$ , maior será o viés do nosso estimador  $RV_t^{(all)}$ .

Por tal motivo foi argumentado que não seria ótimo estimar a volatilidade realizada a partir da observação da série completa de preços (*tick by tick*), mas gerar sub-amostras espaçadas no tempo, como por exemplo de 5 em 5 minutos ou de 10 em 10 minutos, de modo a reduzir  $n_t$  e assim reduzir o viés  $2n_t E(\mathcal{E}_{t,i}^2)$ . No entanto a variância do choque aleatório normal aumenta quanto diminuimos  $n_t$ , obrigando-nos a enfrentar um trade-off entre dimensão do viés e variância.

Bandi e Russell (2005, 2006) e ZMA (2005) propuseram um método para seleção de uma frequência ótima de observação utilizando o critério de minimização do erro quadrático médio (EQM):

$$n_t^* = \left\{ \frac{IQ_t}{4[E(\mathcal{E}_{t,i}^2)]^2} \right\}^{\frac{1}{3}}$$

Em seguida ZMA (2005) foi além e optou por expandir o estimador  $RV_t^{(all)}$  a partir da idéia de que um dia de negócios com início às 9:30 da manhã pode ser medido usando os intervalos de 10 minutos começados em qualquer momento do dia, por exemplo de 9:30-9:40, 9:31-9:41, 9:32-9:42k, etc. Então foi proposto em seguida que o estimador  $RV_t^{(all)}$  fosse atualizado para

$$RV_t^{(ZMA)} = \frac{1}{K} \sum_{i=1}^K RV_t^{(k)} - \frac{\bar{n}_t}{n_t} RV_t^{(all)}$$

onde

$$\bar{n}_t = \frac{1}{K} \sum_{i=1}^K n_t^{(k)} = \frac{n_t^{(k)} - K + 1}{K}$$

Este estimador considera a variância realizada como o excesso da média das observações sob diferentes grades  $k$ , subtraídas de um componente múltiplo do estimador original  $RV_t^{(all)}$ . ZMA(2005) mostra que este estimador converge em distribuição para a variância integrada, somada a um choque normalmente distribuído, mas sem a presença do viés de microestrutura anterior. O viés foi eliminado quando subtraímos a média das grades do estimador original.

AMZ (2006) propõe um pequeno ajuste de amostra para o estimador acima, de modo que a fórmula final será

$$RV_t^{(ZMA,adj)} = \left(1 - \frac{\bar{n}_t}{n_t}\right)^{-1} RV_t^{(ZMA)} = \left(1 - \frac{\bar{n}_t}{n_t}\right)^{-1} \left(\frac{1}{K} \sum_{i=1}^K RV_t^{(k)} - \frac{\bar{n}_t}{n_t} RV_t^{(all)}\right)$$

## E. Contornando os ruídos de microestrutura 2 – O estimador BHLS (2006)

Uma outra alternativa para solucionar o problema do viés de microestrutura é o estimador BHLS (2006), construído a partir de funções kernel. Uma função kernel em estatística é uma função peso usada em técnicas não-paramétricas de estimação da distribuição de estimadores.

BHLS (2006) propõe que o estimador original  $RV_t^{(all)}$  tenha seu viés corrigido somando-se um termo que pondera o produto dos retornos de diferentes períodos por uma função kernel apropriada

$$RV_t^{(BHLS)} = RV_t^{(all)} + \sum_{h=1}^H k\left(\frac{h-1}{H}\right) (\widehat{Y}_h - \widehat{Y}_{-h})$$

onde  $k(x)$  para  $x \in [0,1]$  é uma função peso não estocástica com  $k(0) = 0$  e  $k(1) = 0$  e

$$\widehat{Y}_h = \frac{n_t}{n_t - h} \sum_{j=1}^{n_t-h} r_{t,j} r_{t+h}$$

Para especificar a função  $k$  da equação do estimador foram testados três kernels diferentes:

- a. Barlett:  $k(x) = 1 - x$
- b. 2ª ordem:  $k(x) = 1 - 2x + x^2$
- c. Epanechnikov:  $k(x) = 1 - x^2$

O estimador com a função Barlett apresentou a mesma distribuição assintótica do estimador ZMA (2005) da seção anterior, sendo mais eficiente que o estimador com a função Epanechnikov, mas no entanto menos eficiente que o estimador com a função kernel de 2ª ordem.

Os autores em seu trabalho testaram outras oito funções kernel com a propriedade de que  $k'(0) = k'(1) = 0$ , tendo encontrado que a função kernel de Tukey-Hanning -  $k(x) = \frac{1 - \cos \pi(1-x)^2}{2}$  - aparenta ser a melhor opção em termos de eficiência.

BHLS (2006) também infere que suas conclusões são robustas para séries de ruídos dependentes ao longo do tempo e para ruídos endógenos.

### III. DISTRIBUIÇÃO DOS RETORNOS PADRONIZADOS DO ÍNDICE S&P 500

Investigamos nesta seção a distribuição dos retornos diários do índice de ações norte-americano S&P 500, padronizados segundo diferentes modelos para sua variância. Nosso objetivo é verificar se, após a correção dos ruídos de microestrutura, as regularidades empíricas constatadas por Andersen, Bollerslev, Diebold e Labys (2003) continuam presentes nas séries padronizadas.

#### A. Dados

Para análise da performance relativa dos estimadores de volatilidade realizada, analisaremos as propriedades das séries de retornos do índice de ações norte-americano S&P 500 com e sem overnight.

Duas séries deste mesmo índice são utilizadas neste trabalho. A série R, de retornos sem overnight, considera o retorno diário como a variação logarítmica entre o preço do primeiro negócio e o preço do último negócio realizado em cada dia observado. A série RET, de retornos com overnight, considera a variação logarítmica entre o último negócio de um dado dia e o último negócio do dia anterior. Os períodos de observação das duas séries são idênticos, ambos compostos por 2798 observações diárias entre 2 de janeiro de 1996 e 29 de março de 2007.

A diferença relevante entre as duas séries está no impacto do acúmulo de informações durante o período noturno. Neste período o mercado norte-americano está fechado, enquanto outros mercados mundiais, em especial o mercado financeiro asiático, estão em operação. Todas as notícias relevantes deste período influenciam uma única variação do índice, a saber a variação entre o preço de fechamento de um dia e o preço de abertura do dia seguinte. Já variações do índice durante o dia refletem apenas informações surgidas no breve período de tempo que as separam. Considera-se que a retorno overnight pode introduzir ruídos na série de retornos, e por tal analisaremos as duas séries em paralelo.

Foram obtidas as séries diárias de variâncias realizadas estimadas segundo os modelos propostos em ZMA (2005) e BHLS (2006), a partir daqui mencionadas por ZMA e BHLS. Foram geradas, para comparação, as séries de variâncias para o mesmo período a partir dos modelos ARHCG(1,1), ARHCG-Exponencial, GJR ou TARCH e MA(1)-ARHCG(1,1).

Foi arbitrariamente definido como período amostral as primeiras 2433 observações (~87% do total), equivalente ao período com início em 2/1/1996 e término em 13/10/2005. A tabela 1 mostra as estatísticas descritivas da série de retornos R e RET, tanto para o período amostral quanto para a base de dados completa. As figuras 1 e 2 mostram respectivamente os retornos diários e seus histogramas para as séries R e RET.

Base Dados (2798 obs)	Média	D.P.	Assimetria	Curtose	Jarque-Bera	P-Valor
R (sem overnight)	0.0182	1.0992	-0.1435	6.3817	1342.8042	0.0000
RET (com overnight)	0.0296	1.1225	-0.1781	6.3915	1355.7912	0.0000

Amostra (2433 obs)	Média	D.P.	Assimetria	Curtose	Jarque-Bera	P-Valor
R (sem overnight)	0.0132	1.1515	-0.1259	6.0083	923.8282	0.0000
RET (com overnight)	0.0263	1.1765	-0.1627	6.0054	926.3741	0.0000

Tabela 1- Estatísticas descritivas dos retornos diários do índice S&P 500. Como esperado, a série RET possui desvio-padrão maior por refletir o impacto do acúmulo de informações *overnight* no preço de abertura.

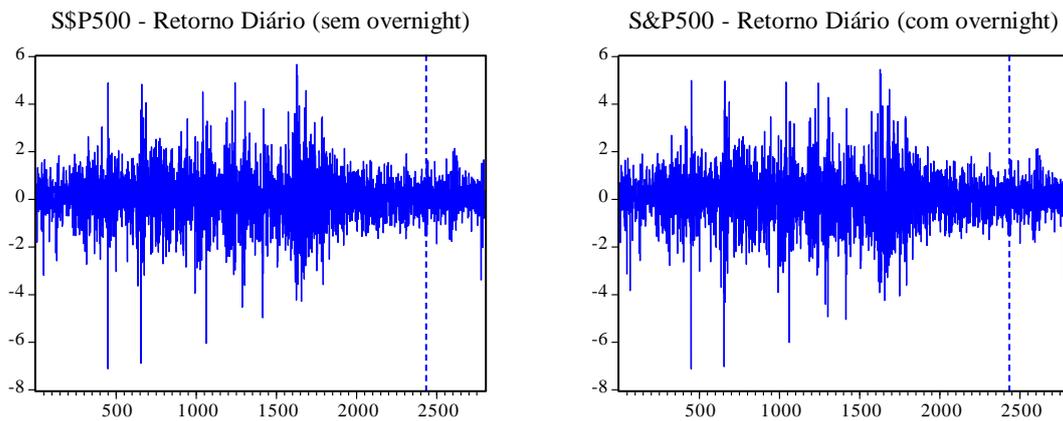


Figura 1 - Retornos diários do índice S&P 500 com e sem overnight. Os dados à direita da linha pontilhada compõem as observações excluídas da amostra para exercícios de pseudo-previsão fora-da-amostra.

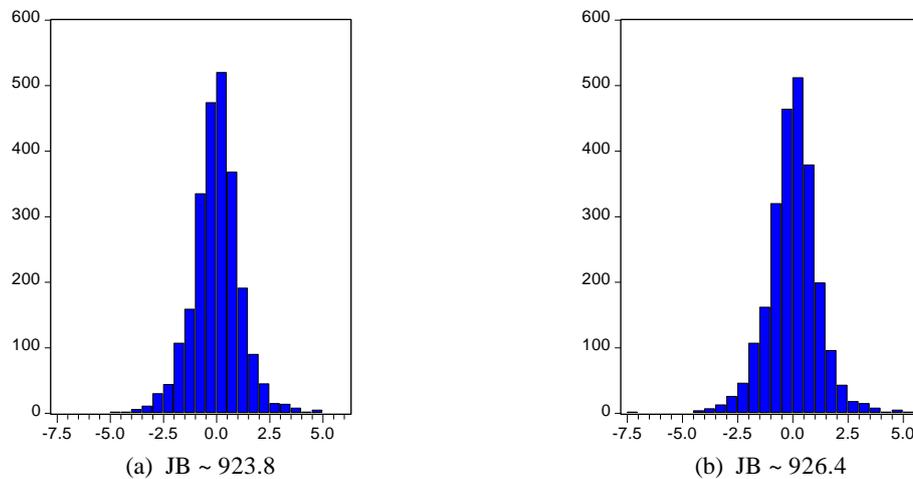


Figura 2 – Histogramas das séries (a) R (sem overnight) e (b) RET (com overnight)

## B. A distribuição dos retornos padronizados

Andersen, Bollerslev, Diebold e Labys (2003) determinam três regularidades empíricas, no jargão conhecidos como “*stylized facts*”, que compõem o arcabouço para o uso da variância realizada como estimador da volatilidade de ativos financeiros.

São elas (1) que os retornos, quando padronizados pelas volatilidades realizadas correspondentes, apresentam distribuições de probabilidade aproximadamente normais; (2) que as volatilidades realizadas apresentam distribuições de probabilidade assimétricas para direita, tendo distribuição aproximadamente log-normal e (3) que a dinâmica de longo-prazo das volatilidades realizadas podem ser aproximadas por um processo de memória longa fracionalmente integrado.

Carvalho, Freire, Medeiros e Souza (2007), verificam as duas primeiras regularidades empíricas em ativos brasileiros, utilizando no entanto estimadores para volatilidade realizada não robustos à ruídos de microestrutura. Em nossa análise do índice financeiro S&P 500, verificamos as três regularidades, embora a segunda apenas parcialmente.

Apresentamos na tabela 2 as estatísticas descritivas dos retornos padronizados.

<b>Sem overnight (R)</b>	Média	Desvio-Padrão	Assimetria	Curtose	Jarque-Bera	P-Valor
BHLS(2006)	0.0577	1.0173	0.0435	2.5011	26.00	0.0000
ZMA(2005)	0.0610	1.0433	0.0475	2.5505	21.39	0.0000
ARHCG(1,1)	0.0025	0.9997	-0.4258	4.6986	366.02	0.0000
ARHCG-E(1,1)	0.0009	1.0004	-0.3887	4.2530	220.42	0.0000
GJR(1,1,1)	-0.0011	1.0004	-0.4288	4.4373	283.98	0.0000
Ma(1)-Archg(1,1)	0.0025	0.9997	-0.4256	4.6973	365.49	0.0000

<b>Com overnight (RET)</b>	Média	Desvio-Padrão	Assimetria	Curtose	Jarque-Bera	P-Valor
BHLS(2006)	0.0743	1.0419	0.0245	2.6217	14.75	0.0006
ZMA(2005)	0.0784	1.0691	0.0344	2.6788	10.94	0.0042
ARHCG(1,1)	0.0141	1.0000	-0.4314	4.6870	363.95	0.0000
ARHCG-E(1,1)	0.0142	1.0008	-0.3990	4.3071	237.75	0.0000
GJR(1,1,1)	0.0116	1.0003	-0.4321	4.4547	290.24	0.0000
Ma(1)-Archg(1,1)	0.0141	1.0000	-0.4317	4.6894	364.90	0.0000

**Tabela 2 - Estatísticas descritivas dos retornos, com e sem overnight, padronizados pelas variâncias realizadas BHLS e ZMA e pelos modelos Garch(1,1), Egarch(1,1), GJR(1,1,1) e MA(1)-Garch(1,1)**

Como podemos ver, os retornos padronizados pelos estimadores BHLS e ZMA não podem ser considerados como normalmente distribuídos pelo teste de Jarque-Bera (valor crítico igual a 9,21 para nível de confiança de 1%). No entanto é evidente a

relativa proximidade destes ao valor crítico, quando comparados com a mesmas estatísticas referentes aos modelos da classe ARHC. Tal fato nos induz a buscar testes estatísticos alternativos, como o de Kolmogorov-Smirnov e o de Lilliefors.

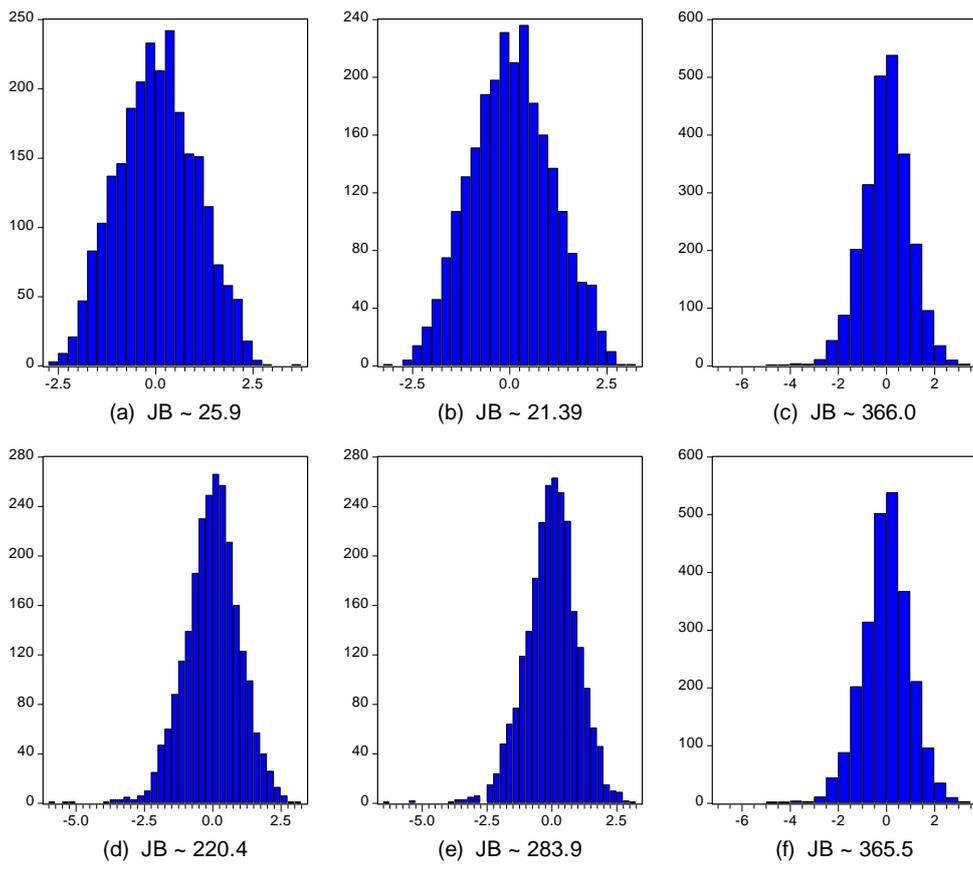
O teste de Kolmogorov-Smirnov confronta os dados contra a hipótese nula de distribuição normal com média igual a zero e variância unitária. Já o teste de Lilliefors, mais abrangente, verifica se a distribuição é normal com média e variância indeterminadas. A tabela 3 apresenta os resultados.

Como podemos ver, nenhum retorno padronizado apresenta distribuição normal *padronizada*, dada a rejeição ao nível de confiança de 5% de todas as séries analisadas pelo teste de Kolmogorov-Smirnov. Já para o teste de Lilliefors, mais abrangente e com o mesmo nível de significância, verificamos que os retornos padronizados por estimadores de volatilidade realizada, tanto BHLS quanto ZMA, apresentam a distribuição aproximadamente normal, confirmando assim a primeira regularidade empírica destacada por Andersen, Bollerslev, Diebold e Labys (2003).

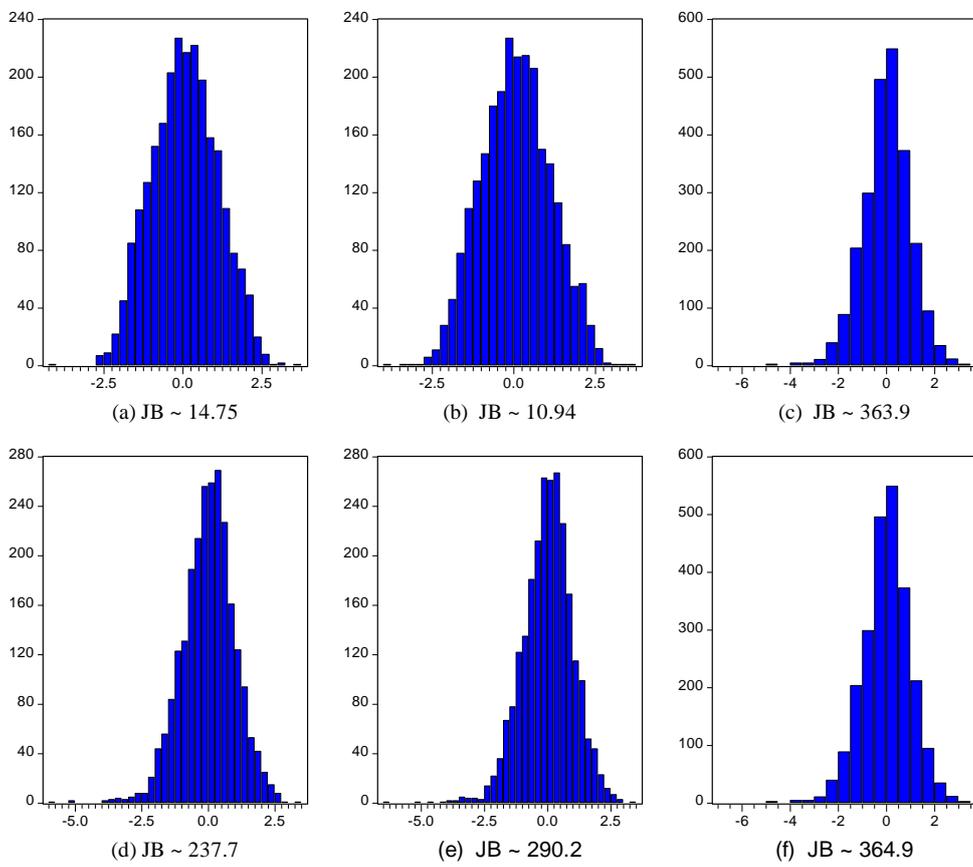
Retornos Padronizados	Kolmogorov-Smirnov		Lilliefors	
	Estatística	P-valor	Estatística	P-valor
R/BHLS	1	0.0045	0	0.0520
R/ZMA	1	0.0007	0	> 0.1
R/ARHCG(1,1)	1	0.0061	1	0.0000
R/ARCHG-E(1,1)	1	0.0435	1	0.0002
R/GJR(1,1,1)	1	0.0296	1	0.0000
R/MA(1)-G(1,1)	1	0.0062	1	0.0000
RET/BHLS	1	0.0003	0	0.0503
RET/ZMA	1	0.0001	0	> 0.1
RET/ARHCG(1,1)	1	0.0006	1	0.0000
RET/ARCHG-E(1,1)	1	0.0016	1	0.0000
RET/GJR(1,1,1)	1	0.0020	1	0.0000
RET/MA(1)-G(1,1)	1	0.0006	1	0.0000

**Tabela 3 - Testes de "boa aproximação" para os retornos padronizados. "R" é a série de retornos diários sem overnight. "RET" é a série de retornos diários com overnight.**

As figuras 3 e 4 a seguir apresentam os histogramas dos retornos padronizados para série R (sem overnight) e para série RET (com overnight), respectivamente, onde pode ser vista a maior aproximação da distribuição normal pelos retornos padronizados pelas volatilidades realizadas.



**Fig. 3 – Série R, padronizada por: (a) BHLS, (b) ZMA, (c) Archg, (d) Archg-e, (e) GJR, (f) Ma1-G11**



**Fig. 4 – Série RET, padronizada por: (a) BHLS, (b) ZMA, (c) Archg, (d) Archg-e, (e) GJR, (f) Ma1-g11**

Com relação à segunda regularidade empírica, passamos à análise das estatísticas descritivas das séries de volatilidade. Podemos ver na tabela 4 abaixo, em seu segmento superior, que as séries de volatilidades realizadas apresentam-se claramente assimétricas e leptocúrticas, conforme era esperado. Pelo teste de Jarque-Bera não restariam dúvidas sobre a não-normalidade destas séries.

No entanto, ao contrário da regularidade empírica apontada por Andersen, Bollerslev, Diebold e Labys (2003), as séries logarítmicas das volatilidades continuam não-gaussianas. Podemos perceber que os **logaritmos** das volatilidades estão *relativamente mais próximos* da distribuição normal do que a volatilidade em si, mas ainda muito distantes do valor crítico do teste de Jarque-Bera. Este resultado se confirma nos testes de Kolmogorov-Smirnov e Lilliefors, como pode ser visto na tabela 5. A figura 5 mostra os histogramas dos **logaritmos** das séries BHLS e ZMA.

<b>Volatilidade</b>	Média	Desvio-Padrão	Assimetria	Curtose	Jarque-Bera	P-Valor
BHLS(2006)	1.2245	1.5479	5.3249	46.4154	202,579.04	0.0000
ZMA(2005)	1.1299	1.2582	4.2969	30.9852	86,880.89	0.0000

<b>Log(Volatilidade)</b>						
	Média	Desvio-Padrão	Assimetria	Curtose	Jarque-Bera	P-Valor
BHLS(2006)	-0.1794	0.8102	0.5429	3.4006	135.79	0.0000
ZMA(2005)	-0.2156	0.7689	0.5336	3.2413	121.34	0.0000

Tabela 4 - Estatísticas descritivas das séries de volatilidades realizadas BLHS(2006) e ZMA(2005) e das volatilidades condicionais da classe ARHC estimadas a partir da série R (sem variações overnight)

<b>Log (volatilidade)</b>	Kolmogorov-Smirnov		Lilliefors	
	Estatística	P-valor	Estatística	P-valor
BHLS	1	0.0000	1	0.0000
ZMA	1	0.0000	1	0.0000

Tabela 5 - Testes de "goodness of fit" para a variação da volatilidade

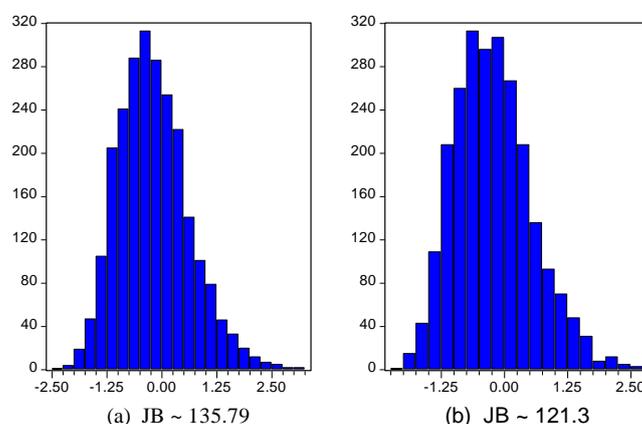


Figura 5 – Histogramas dos logaritmos das volatilidades realizadas (a) BHLS e (b) ZMA

A terceira regularidade empírica diz respeito à dinâmica de longo prazo das séries de volatilidade realizada. Uma série temporal possui memória longa quando há autocorrelação significativa em uma quantidade de defasagens superior ao que seria esperado de um processo ARMA.

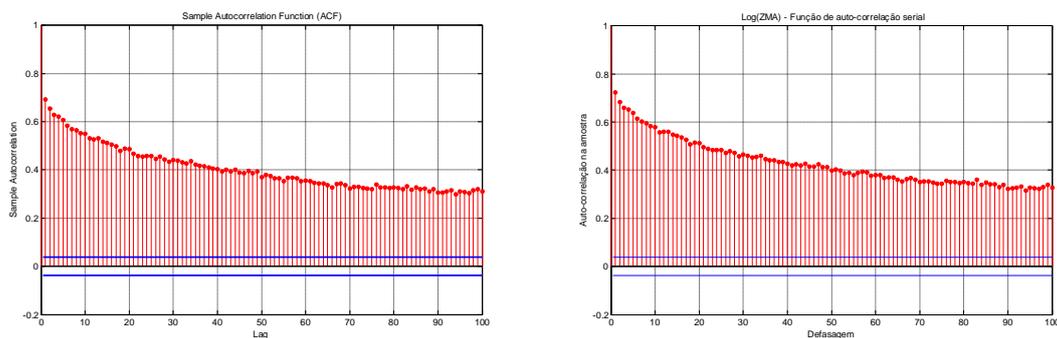
Um processo estocástico  $Y_t$  é dito integrado de ordem  $d$  e representados por  $I(d)$  quando  $Y_t$  precisa ser diferenciado  $d$  vezes para eliminar sua tendência estocástica. Analogamente,  $Y_t$  é  $I(d)$  se:

$$(1 - L)^d Y_t = u_t$$

onde  $L$  é o operador de defasagem. Séries integradas  $I(0)$  ou  $I(1)$  não são compatíveis com níveis elevados de persistência na função de autocorrelação. Um processo de memória longa *fracionalmente integrado* é definido para  $-0,5 < d < 0,5$  e  $u_t$  estacionário, tendo espectro positivo e limitado em todas as frequências. Quando  $d$  é positivo temos um decaimento hiperbólico e positivo das autocorrelações parciais, sendo esta a terceira regularidade empírica sugerida por Andersen, Bollerslev, Diebold e Labys (2003).

Analisando a figura 6, que contém as funções de autocorrelação serial para os logaritmos das séries BHLS e ZMA, fica evidente a **persistência** em mais de 100 defasagens, quando 20 já seriam suficientes para ratificar nossa conclusão.

Em paralelo, calculamos o estimador GPH, proposto em Geweke e Porter-Hudak (1983) para  $d$ , mostrado em seguida na tabela 6. Podemos ver que tanto para o estimador BHLS quanto para o estimador ZMA o estimador é positivo e inferior a 0,5, confirmando a validade terceira regularidade empírica nestas séries.



**Figura 6 - Função de autocorrelação serial para 100 defasagens do Log(BHLS) e do Log(ZMA), respectivamente. Evidências de memória longa na série.**

Log(Volatilidade)	$d^{GPH}$
ZMA	0.4335
BHLS	0.3863

**Tabela 6 - Estimativas para o estimador de diferenciação fracional de Geweke Porter-Hudak (GPH)**

#### IV. MODELAGEM E PREVISÃO COM ESTIMADORES ZMA e BHLS

Nesta seção realizamos exercícios de **pseudo previsão** fora da amostra para a volatilidade do índice S&P 500. Comparamos a performance, tanto dentro (sub-seção A) quanto fora da amostra (sub-seção B), dos estimadores de volatilidade realizada ZMA e BHLS, comparando-os com os mesmos modelos analisados na seção III sob o critério de minimização da raiz do erro de previsão quadrático médio (REPQM) e do erro absoluto médio (EAM).

##### A. Resultados dentro da amostra

A análise dentro da amostra tem por objetivo avaliar qual dos métodos para estimação da volatilidade mais se aproxima, em média, da volatilidade teórica da série. Adotamos como critério de comparação duas funções de perdas estatísticas (*statistical loss functions*) de uso disseminado em comparações da capacidade descritiva de modelos econométricos: a raiz do erro de previsão quadrático médio (REPQM) e do erro absoluto médio (EAM).

O nosso objetivo é ajustar estas funções ao tamanho de nossa amostra de modo que elas penalizem os erros dos modelos de estimação. Como normalmente utilizado em comparações entre modelos da classe ARHC, os erros são medidos com relação ao quadrado do retorno observado. Este é considerado um bom estimador da volatilidade teórica, sob a hipótese de que os retornos têm média zero, pois deste modo:

$$\text{Var}(r_t) = E(r_t^2) + E(r_t)^2 = E(r_t^2)$$

As funções de perdas adotadas são então descritas pelas seguintes equações:

$$\text{REPQM} = \sqrt{\frac{1}{T} \sum_{t=1}^T (r_{t+1}^2 - h_{t+1})^2}$$
$$\text{EAM} = \frac{1}{T} \sum_{t=1}^T |r_{t+1}^2 - h_{t+1}|$$

onde  $h_t$  é a variância estimada pelo modelo em questão e  $r_t$  é o retorno observado no período  $t$ . Utilizaremos para efeito de consistência tanto os retornos da série R (sem overnight) quanto os retornos da série RET (com overnight).

O resultados podem ser visualizados na tabela 7 abaixo. Podemos ver que a volatilidade realizada gera, tanto para a série de retornos R quanto para série de retornos RET, menores valores das funções estatísticas de perdas. Entre os diferentes estimadores de volatilidade realizada analisados, há um empate: a série BHLS sai ganhadora sob o critério de minimização do REPQM, enquanto a série ZMA sai ganhadora sob o critério de minimização do EAM.

R	REPQM	EAM	REPQM %BHLS	EAM %ZMA
<b>BHLS</b>	<b>2.4656</b>	1.0914	<b>100.00%</b>	100.49%
<b>ZMA</b>	2.5283	<b>1.0860</b>	102.55%	<b>100.00%</b>
ARHCG(1,1)	2.8237	1.1747	114.53%	108.16%
ARHCG-E	2.7324	1.1463	110.82%	105.55%
GJR	2.7493	1.1564	111.51%	106.48%
MA(1)-ARCHG(1,1)	2.8236	1.1746	114.52%	108.16%

RET	REPQM	EAM	REPQM %BHLS	EAM %ZMA
<b>BHLS</b>	<b>2.5981</b>	1.1119	<b>100.00%</b>	100.47%
<b>ZMA</b>	2.6662	<b>1.1067</b>	102.62%	<b>100.00%</b>
ARHCG(1,1)	2.9482	1.2027	113.48%	108.68%
ARHCG-E	2.8629	1.1740	110.19%	106.08%
GJR	2.8804	1.1853	110.87%	107.10%
MA(1)-ARCHG(1,1)	2.9485	1.2028	113.48%	108.69%

Tabela 7 – Funções de perda dentro da amostra. A parte superior utiliza o quadrado do retorno sem overnight (R) como proxy para volatilidade. Já a parte inferior utiliza o quadrado do retorno com overnight (RET).

Muitas vezes não se dispõe de dados intra-diários para o ativo financeiro que se quer prever. Sabemos que o conceito de volatilidade realizada consiste na soma de diversos retornos intradiários ao quadrado, e portanto, é uma *proxy* da volatilidade verdadeira superior ao quadrado do retorno diário, por absorver um volume consideravelmente maior de informações.

Com isso podemos melhor avaliar a performance dos modelos da classe ARHC utilizando a volatilidade realizada como proxy nas funções de perdas estatísticas em questão. Em síntese, ajustaremos as funções REPQM e EAM para penalizar desvios destes modelos *em relação a volatilidade realizada*, ao invés do quadrado do retorno diário. Para efeitos de consistência realizamos o exercício variando tanto as séries de retorno (R e RET) quanto as séries de volatilidade realizada (BHLS e ZMA).

Os estimadores modificados são descritos pelas seguintes equações:

$$REPQM' = \sqrt{\frac{1}{T} \sum_{t=1}^T (\sigma_{RV,t+1}^2 - h_{t+1})^2}$$

$$EAM' = \frac{1}{T} \sum_{t=1}^T |\sigma_{RV,t+1}^2 - h_{t+1}|$$

onde  $\sigma_{RV,t+1}^2$  é a variância realizada, estimada tanto pelo estimador BHLS quanto pelo estimador ZMA, e  $h_t$  é a variância estimada pelos modelos da classe ARHC. A tabela 8 apresenta os resultados.

Dentre os quatro métodos avaliados, podemos ver que o modelo ARHCG-Exponencial é o que em média erra menos na descrição da volatilidade realizada, resultado este válido tanto para o estimador BHLS quanto ZMA. Em síntese, nossa análise empírica mostra que na ausência de dados intradiários, o modelo ARHCG-E tem desempenho superior aos demais modelos da classe ARHC na descrição da variância dos retornos.

Séries R e BHLS			Séries R e ZMA		
	REPQM	EAM		REPQM	EAM
GARCH11	1.3046	0.8229	GARCH11	1.0653	0.7866
<b>EGARCH</b>	<b>1.1475</b>	<b>0.7686</b>	<b>EGARCH</b>	<b>0.8836</b>	<b>0.7249</b>
GJR	1.1847	0.7849	GJR	0.9818	0.7524
MA1G11	1.3044	0.8229	MA1G11	1.0651	0.7866

Série RET e BHLS			Séries RET e ZMA		
	REPQM	EAM		REPQM	EAM
GARCH11	1.3323	0.8438	GARCH11	1.1015	0.8104
<b>EGARCH</b>	<b>1.1618</b>	<b>0.7849</b>	<b>EGARCH</b>	<b>0.9079</b>	<b>0.7452</b>
GJR	1.2282	0.8089	GJR	1.0332	0.7782
MA1G11	1.3327	0.8439	MA1G11	1.1020	0.8105

Tabela 8 – Funções de perda modificadas – dentro da amostra. Para séries de volatilidade estimadas a partir dos retornos R e RET, considerando como volatilidade teórica tanto a série BLHS quanto a série ZMA.

Partimos agora para a estimação de intervalos de confiança para os retornos diários a partir das séries de variâncias realizadas e condicionais geradas pelos seis métodos analisados, com níveis de significância de 5% da tabela da distribuição normal padronizada, seguindo o modelo abaixo:

$$I_t = (\mu - 1,96\sqrt{h_t}; \mu + 1,96\sqrt{h_t})$$

onde  $I_t$  é o intervalo de confiança,  $\mu$  é a média e  $h_t$  é a variância condicional.

Os resultados podem ser vistos na tabela 9 e nas figuras 7 e 8 abaixo. A estatística referente à cobertura reporta o percentual de pontos do retorno observado que ficaram no interior do intervalo de confiança estimado.

Embora todos os métodos tenham gerados percentuais de cobertura próximos, é interessante notar que o método ZMA aparece como o pior método analisado. Tal conclusão está em sintonia com o resultado obtido em Carvalho (2003). Neste estudo o autor atribui o resultado a uma incapacidade do estimador de volatilidade realizada de estimar propriamente a variância em momentos de estresse, onde a volatilidade aumenta rapidamente. Choques na volatilidade seriam, ainda segundo o autor, melhor modelados seguindo a metodologia usual da classe ARHC.

Na contramão desta conclusão está o método BHLS, que é o melhor na descrição dos retornos sem overnight, sendo no entanto o segundo pior na descrição dos retornos com overnight. Este curioso resultado sugere que a eliminação dos ruídos de microestrutura neste método torna o estimador mais eficaz na previsão de momentos de volatilidade intensa, tornando-o melhor inclusive que os modelos da classe ARHC. No entanto, este estimador ainda é calculado sobre a totalidade de variações dos retornos intradiários, nada contendo em sua estimação de informações relativas ao período overnight, assim explicando o mau desempenho para a série RET.

<b>R</b>			
Ranking	Variância	Cobertura	
1	BHLS	2317	95.23%
2	GJR	2315	95.15%
3	ARHCG-E	2314	95.11%
4	ARHCG(1,1)	2308	94.86%
5	MA(1)-ARHCG(1,1)	2308	94.86%
6	Variância Incondicional	2298	94.45%
7	ZMA	2285	93.92%

<b>RET</b>			
Ranking	Variância	Cobertura	
1	ARHCG-E	2317	95.23%
2	GJR	2317	95.23%
3	Variância Incondicional	2306	94.78%
4	ARHCG(1,1)	2305	94.74%
5	MA(1)-ARHCG(1,1)	2305	94.74%
6	BHLS	2302	94.62%
7	ZMA	2268	93.22%

Tabela 9 - Testes de cobertura - dentro da amostra

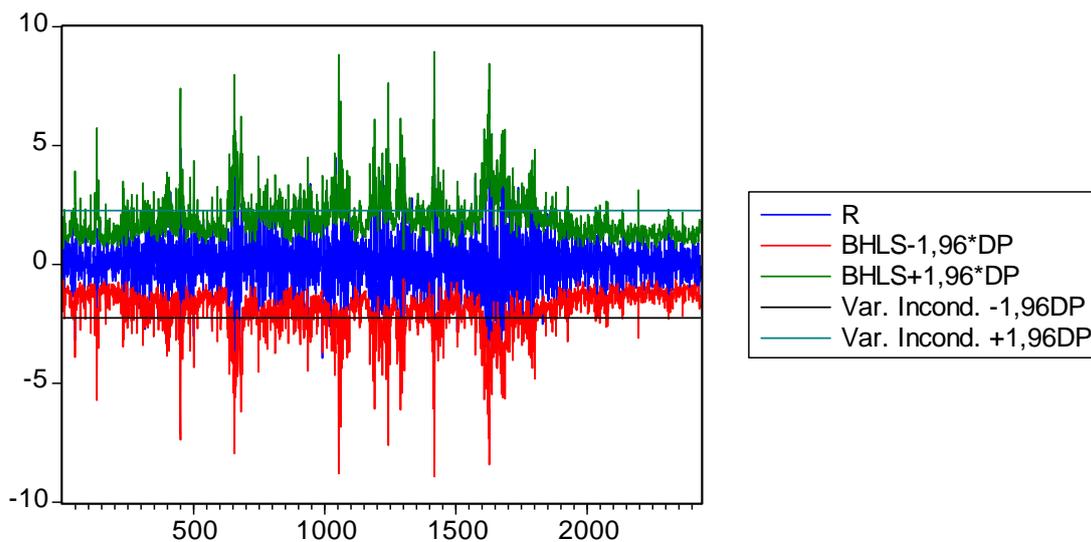


Figura 7 - Intervalos de confiança para série R (sem overnight) - dentro da amostra

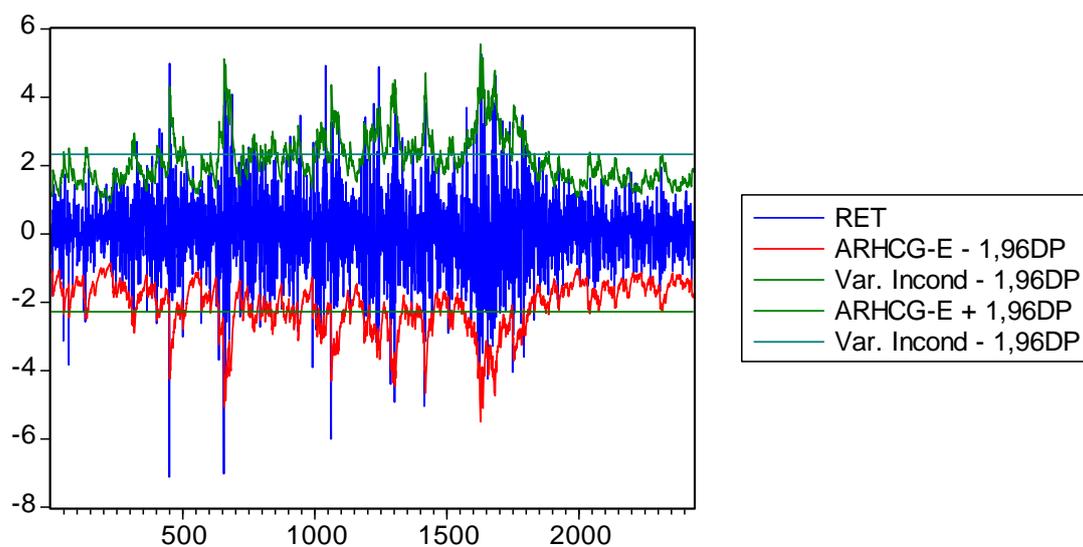


Figura 8 - Intervalos de confiança para série RET (com overnight) - dentro da amostra

## B. Resultados fora da amostra

Deste o início da seção III foram reservadas as últimas 365 observações contidas nas séries para exercícios de pseudo-previsão fora da amostra.

Este exercício consiste em simular qual teria sido a decisão ótima de um investidor financeiro no dia 13 de Outubro de 2005, que, dispondo das séries de

retornos desde 1996, pretende escolher dentre os diferentes intervalos de confiança para os retornos dos 15 meses seguintes.

As previsões um passo a frente para o período fora da amostra são facilmente geradas para os modelos da classe ARHC. Apenas realizamos a estimação dos coeficientes dos modelos a partir dos dados contidos na amostra e, em seguida, mantendo-se fixos os parâmetros calculados, geramos uma nova variância condicional a cada novo conjunto de informação.

Já a previsão dos modelos de volatilidade realizada não é tão simples. A princípio elas são geradas a partir dos retornos intra-diários dos dias futuros, não havendo um modelo específico que relacione as observações entre dias diferentes. Intuitivamente deveríamos regredir um modelo ARMA para o período observado, tentando buscar padrões recorrentes na série histórica de volatilidade realizadas, para então gerar previsões para o período fora da amostra.

No entanto, devido a excessiva persistência verificada na seção III, **concluimos** que a volatilidade realizada é um processo de memória longa fracionalmente integrada. Com isso o uso de modelos ARMA torna-se inapropriado. Mesmo o uso de modelos da classe ARIMA produziria estimações viesadas da volatilidade futura devido à excessiva **persistência** das autocorrelações.

O modelo apropriado para filtrar este viés é o modelo ARFIMA, proposto inicialmente por Hosking (1981) e por Granger e Joyeux (1980)<sup>1</sup>. Em síntese, este modelo introduz um filtro no modelo ARIMA, que transforma o processo estocástico em questão, no caso as séries de volatilidades realizadas, em uma constante somada de um **ruído** branco.

Uma análise semelhante à realizada dentro da amostra pode ser vista nas tabelas 10 e 11. Novamente são utilizadas as duas funções de perda REPQM e EAM propostas na sub-seção A, assim como a mesma metodologia é adotada para a construção dos intervalos de confiança. Fica claro, pela tabela 11 que o período fora da amostra foi consideravelmente menos volátil que o período dentro da amostra. Isto pode ser percebido pela posição da variância incondicional como melhor estimador de variância para intervalos de confiança.

Observando as figuras 9 e 10 abaixo, podemos entender o aparente bom desempenho da volatilidade realizada no exercício de pseudo-previsão fora da amostra.

---

<sup>1</sup> Para mais detalhes, ver Taylor (2005), página 46.

Claramente a série ZMA é insensível a variações bruscas da série de retornos. Por acaso, o período em questão apresentou baixa volatilidade, de modo que o intervalo de confiança **construído** pela volatilidade realizada acabou por absorver 97,26% dos retornos do período. Mas podemos ter certeza de que as chances de ocorrência deste evento não eram os 95% sugeridos pelo intervalo de confiança.

Com isso confirmamos a **hipótese** de Carvalho (2003) de que o estimador de volatilidade realizada falha em perceber choques de volatilidade. Na ausência destes choques, o estimador de ZMA passou a ser o melhor dentre os demais para retornos sem overnight (série R) e o segundo melhor para a série de retornos com overnight. O estimador BHLS continua obtendo um péssimo desempenho na presença do retorno overnight, conforme já havíamos verificado na análise dentro da amostra. O modelo GJR, de características semelhantes ao modelo ARCHG-Exponencial, obteve o melhor desempenho para a série RET.

R	REPQM	EAM	REPQM %BHLS	EAM %ZMA
BHLS	0.9287	0.6145	100.00%	95.52%
ZMA	0.9408	0.6433	101.30%	100.00%
ARHCG(1,1)	0.9084	0.5326	97.81%	82.79%
<b>ARCHG-E</b>	<b>0.8929</b>	<b>0.5093</b>	<b>96.14%</b>	<b>79.16%</b>
GJR	0.9100	0.5355	97.98%	83.24%
MA(1)-ARHCG(1,1)	0.9084	0.5326	97.81%	82.79%

RET	REPQM	EAM	REPQM %BHLS	EAM %ZMA
BHLS	0.9731	0.6229	100.00%	115.53%
<b>ZMA</b>	<b>0.9445</b>	<b>0.5392</b>	<b>97.06%</b>	<b>100.00%</b>
ARHCG(1,1)	0.9562	0.5488	98.26%	101.78%
ARCHG-E	0.9606	0.5624	98.71%	104.30%
GJR	0.9563	0.5488	98.27%	101.79%
MA(1)-ARHCG(1,1)	0.9841	0.6515	101.13%	120.82%

Tabela 10 - Funções de perda - fora da amostra

**R**

Ranking	Variância	Cobertura
1	Incondicional	364 99.73%
2	ZMA	355 97.26%
3	GJR	355 97.26%
4	BHLS	352 96.44%
5	ARCHG-E	352 96.44%
6	MA(1)-ARCGH(1,1)	352 96.44%
7	ARHCG(1,1)	351 96.16%

**RET**

Ranking	Variância	Cobertura
1	Incondicional	364 99.73%
2	GJR	357 97.81%
3	ZMA	355 97.26%
4	ARCHG-E	352 96.44%
5	ARHCG(1,1)	351 96.16%
6	MA(1)-ARCGH(1,1)	351 96.16%
7	BHLS	350 95.89%

Tabela 11 - Testes de cobertura - Fora da amostra

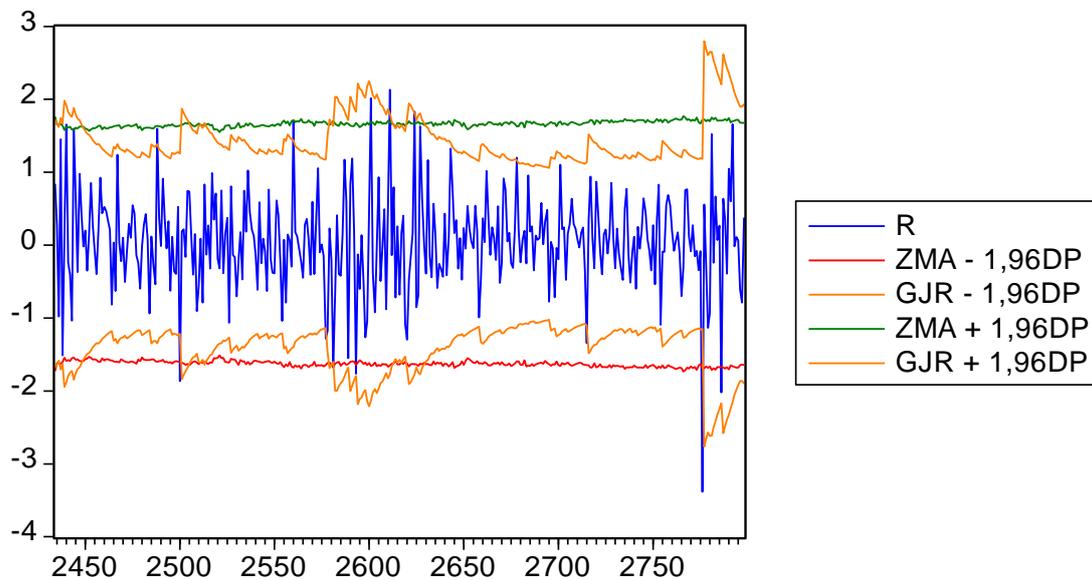
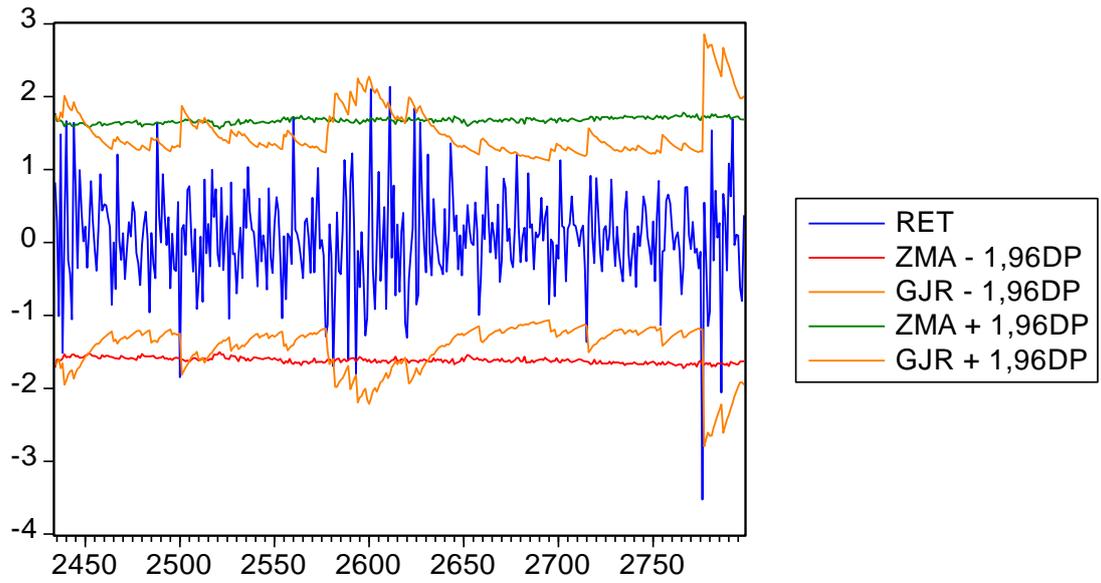


Figura 9 - Intervalos de confiança para série R - Fora da amostra



**Figura 10 - Intervalos de confiança para série RET - Fora da amostra**

## V. CONCLUSÃO

A principal conclusão que podemos retirar deste trabalho é que a correção dos ruídos de microestrutura nos estimadores de volatilidade realizada pouco nos afastou dos resultados anteriormente obtidos em trabalhos empíricos com estimadores de volatilidade realizada. A julgar pelas hipóteses simplificadoras adotadas por ZMA (2005) e BHLS (2006) para o problema do viés de microestrutura, seus estimadores não provocam grandes mudanças em resultados empíricos já conhecidos.

Verificamos que as ditas regularidades empíricas constatadas por Andersen, Bollerslev, Diebold e Labys (2003) continuam presentes nas séries de retornos padronizadas por estimadores de volatilidade realizada robustos à microestrutura. Verificamos no entanto que, após a correção, tais regularidades são menos significantes, o que sugere que os ruídos anteriormente presentes atuavam para intensificá-las.

Também pudemos perceber que as séries de volatilidade realizada apresentam memórias muito longas, em ordens superiores a 400 defasagens. Este problema já foi identificado anteriormente, onde foi levantada a hipótese de se dever à estrutura do próprio estimador e não à uma característica intrínseca às séries de volatilidades dos ativos financeiros. Isto porque faz pouco sentido imaginar que a volatilidade de 6 meses atrás está determinando *fortemente* a volatilidade de hoje.

Confirmamos na seção IV a incapacidade preditiva dos estimadores de volatilidade realizada, resultado já conhecido para estimadores não robustos aos ruídos de microestrutura. Mesmo robustos aos ruídos, os modelos ARFIMA estimados para a previsão das séries de volatilidades realizadas mostraram-se incapazes de prever movimentos bruscos na série de retornos. Isto é uma conclusão crucial, pois esta é a variável chave na adoção de modelos para uso no dia-a-dia dos mercados financeiros.

A solução para o problema pode estar no aprimoramento da modelagem para a previsão das séries de volatilidade realizada. Uma idéia para o futuro seria incluir, como regressor na equação de variância condicional de um modelo ARHCG, a série de variâncias realizadas. Deste modo estariam inseridas nesta bem-sucedida classe de modelos as informações contidas neste estimador.

## VI. REFERÊNCIAS BIBLIOGRÁFICAS

**Aït-Sahalia, Y., P.A. Mykland e L. Zhang.** “Ultra high frequency volatility estimation with dependent microstructure noise”, Working Paper, w11380, NBER, 2006.

**Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys (2003)**, Modeling and forecasting realized volatility, *Econometrica*, 71, 529 – 626.

**Bandi, F.M. and J.R. Russell (2005)**, Microstructure noise, realized volatility, and optimal sampling, unpublished paper, Graduate School of Business, University of Chicago.

**Bandi, F.M., and J.R. Russell (2006)**, Separating market microstructure noise from volatility, *Journal of Financial Economics*, 79, 655 – 692.

**Barndorff-Nielsen, O.E., P.H. Hansen, A. Lunde e N. Shephard**, “Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise”, manuscrito não publicado, Stanford University, 2006.

**Barndorff-Nielsen, O.E. and N. Shephard (2002)**, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society B*, 64, 253 – 280.

**Bollerslev, T. (1986)**, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 21, 307 – 328.

**Carvalho, Marcelo R.C.**, “A volatilidade realizada como uma metodologia para modelagem e previsão da variância dos retornos de ativos financeiros”. PUC-Rio, 2003.

**Carvalho, M.; Freire, M.A.; Medeiros, M., Souza, L.** “Modeling e forecasting the volatility of Brazilian asset returns”, *Revista Brasileira de Finanças*, 2006.

**Engle, R.F.** “Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation”, *Econometrica*, 50, 987 – 1007, 1982.

**Geweke, J. and S. Porter-Hudak (1983)**, "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221-238.

**Granger, C.W.G, Joyex, R. (1980)**, “An introduction to long-memory time series models.” *Journal of Time Series Analysis* 1, 15-29

**Hosking, J. (1981)**, “Fractional Differencing.” *Biometrika*, Vol. 68, No. 1, 165-176

**McAleer, M. and Medeiros, M.C.** “Realized Volatility: A Review”. *Econometric Reviews*, 2007. Por aparecer.

**Taylor, S.** “Asset price dynamics, volatility and prediction”. New Jersey: Princeton University Press, 2005.

**Zhang, L., P.A. Mykland and Y. Aït-Sahalia (2005).** A tale of two time scales: Determining integrated volatility with noisy high frequency data, Journal of the American Statistical Association, 100, 1394 – 1411.