

# **MODELANDO A VOLATILIDADE DOS RETORNOS DE PETROBRÁS USANDO DADOS DE ALTA FREQUÊNCIA**

**Leonardo Cappa**

**Rosas Capital Management**

**Rua dos Pinheiros, 498, cj. 111**

**05422-000 São Paulo, S.P. Brasil**

**Tel: 11- 3014-6419**

**Fax: 11-3014-6415**

**e-mail: [leonardo.cappa@rosasoffice.com](mailto:leonardo.cappa@rosasoffice.com)**

**Pedro L. Valls Pereira<sup>1</sup>**

**CEQEF e EESP – FGV**

**Rua Itapeva 474 sala 1202**

**01332-000 São Paulo, S.P. Brasil**

**Tel: 11-3799-3726**

**Fax: 11-37993350**

**e-mail: [pedro.valls@fgv.br](mailto:pedro.valls@fgv.br)**

## **RESUMO**

O objetivo do presente trabalho é analisar as características empíricas de uma série de retornos de dados em alta frequência para um dos ativos mais negociados na Bolsa de Valores de São Paulo. Estamos interessados em modelar a volatilidade condicional destes retornos, testando em particular a presença de memória longa, entre outros fenômenos que caracterizam este tipo de dados. Nossa investigação revela que além da memória longa, existe forte sazonalidade intradiária, mas não encontramos evidências de um fato estilizado de retornos de ações, o efeito alavancagem. Utilizamos modelos capazes de captar a memória longa na variância condicional dos retornos dessazonalizados, com resultados superiores a modelos tradicionais de memória curta, com implicações importantes para precificação de opções e de risco de mercado.

**Palavras-chave:** memória longa; sazonalidade intradiária; FIGARCH; dados de alta frequência.

---

<sup>1</sup> O segundo autor foi financiado parcialmente pelo CNPq Grant n. 480831/2007-6 e pela FAPESP.

## ABSTRACT

The aim of this paper is to assess the empirical characteristics of a high-frequency return series of one of the main assets traded at the São Paulo Stock Exchange. We are interested in modeling the conditional volatility of these return series, particularly testing for long-memory. Our findings reveal that besides long memory, there is strong intraday periodicity, but we found no evidence of leverage effect. We use models that are able to account for the long memory in the conditional variance of the seasonally adjusted returns, yielding superior results when compared to traditional short-memory volatility models, with important implications to option pricing and risk management

**Keywords:** long-memory; intraday periodicity; FIGARCH; high frequency data.

## INTRODUÇÃO

Neste artigo testamos através da existência de memória longa alguns dos fatos estilizados presentes nos retornos de ativos financeiros, usando dados de alta frequência. Trabalhos utilizando base de dados de alta frequência ainda são raros para ativos brasileiros, devido à inexistência de tais bases, (até anos recentes). Este tipo de dado geralmente apresenta memória longa na volatilidade, ou seja, alta persistência de choques nas previsões da variância condicional. Testamos ainda a ocorrência do chamado efeito de alavancagem, mas não encontramos evidências do mesmo.

Estudos da volatilidade intradiária são relativamente recentes na literatura. Uma descrição pioneira de seu comportamento e padrões utilizando dados de ações negociadas na Bolsa de Valores de New York estão no artigo de Wood, McInish e Ord (1985). Paralelamente, a partir do surgimento do modelo ARCH, proposto em Engle (1982), diversas extensões e desenvolvimentos adequando-se cada vez mais a realidade dos dados de retornos foram publicadas, mas somente em meados da década seguinte foram propostos modelos capazes de captar a memória longa, respectivamente o FIGARCH e o FIEGARCH nos trabalhos de Baillie, Bollerslev e Mikkelsen (1996) e Bollerslev e Mikkelsen (1996). Estes trabalhos no entanto empregaram tais modelos para descrição da volatilidade de séries diárias. Outro desenvolvimento de grande relevância para o tema foi a constatação, já em 1994 e especialmente em 1997, de que séries de retornos intradiários possuem volatilidade com características distintas para diferentes períodos do dia, apresentando portanto periodicidade (sazonalidade) intradiária que deve ser tratada para a correta modelagem da variância condicional.

A partir destas observações, modelos determinísticos e estocásticos para volatilidade, e métodos de extração do componente periódico mais complexos apareceram na literatura em abundância por pelo menos uma década. Neste trabalho aplicamos alguns destes modelos, utilizando uma série de retornos de alta frequência de ações preferenciais da Petrobrás, o ativo mais líquido negociado na Bolsa de Valores de São Paulo para modelar o comportamento da volatilidade condicional. Nossos resultados mostram que apesar de serem os primeiros modelos apropriados para modelagem de memória longa na variância condicional, o FIGARCH e o FIEGARCH são bastante robustos, revelando-se tão adequados quanto modelos mais recentes, e bastante superiores a modelos tradicionais de memória curta.

A estrutura deste artigo é: na próxima seção colocamos considerações mais gerais sobre modelagem financeira, sendo o principal ponto de interesse tornar claro porque optamos

por utilizar os dados de alta frequência no presente estudo, e mostrar que resultados são frequentemente encontrados neste caso. O aparato econométrico utilizado para modelar e analisar a volatilidade de uma série temporal de retornos é explicado na seção 2. A seção 3 fornece a aplicação empírica da modelagem apresentada anteriormente, discutindo os resultados a luz dos fatos estilizados delineados na segunda seção. A seção 4 conclui o trabalho, apresentando algumas críticas e sugestões para futuros estudos.

## **1. Modelagem financeira**

### **1.1 Retornos**

Algumas considerações sobre a modelagem de retornos de ativos financeiros iniciam esta seção. A principal motivação para modelarem-se retornos e não preços dos ativos na maioria dos estudos financeiros é que retornos possuem propriedades estatísticas mais interessantes, destacando-se que retornos são em geral, estacionários e ergódicos<sup>2</sup>. Outra razão, apontada em Campbell, Lo e MacKinlay (1997), é que retornos compreendem para o investidor médio uma medida completa e livre de escala da oportunidade de investimento.

Em geral, retornos de ativos financeiros possuem média amostral bastante próxima de zero, e usaremos o termo retorno e inovação intercambiavelmente<sup>3</sup>, notando que a rigor, a inovação é definida por  $\varepsilon_t = r_t - \bar{r}$ .

### **1.2 Fatos Estilizados sobre Retornos**

Muitos fatos estilizados sobre retornos estão documentados em estudos empíricos ou teóricos, dentre eles destacamos Bollerslev, Chou e Kroner (1992), e Taylor (2005). Uma pesquisa abrangente está em Cont (2001), que expõe entre outros fenômenos:

F.E. a - **Ausência de autocorrelação:** retornos de ativos em geral não apresentam dependência linear, exceto em altas frequências - Cont (2001) sugere intervalos de até cerca de 20 minutos - aonde efeitos de microestruturas tendem a causar uma modesta autocorrelação negativa em defasagens pequenas, normalmente um período<sup>4</sup>.

<sup>2</sup> Para uma definição de estacionaridade e ergodicidade, ver Hamilton (1994).

<sup>3</sup> Também os termos log-retorno e retorno terão o mesmo significado ao longo da discussão.

<sup>4</sup> Em especial devido ao bid-ask bounce, descrito primeiramente em Roll (1984). Transações tendem a ocorrer próximas ao preço de compra ou ao preço de venda, e então o preço efetivo transacionado oscila entre esses dois limites, gerando a autocorrelação negativa. Uma investigação pioneira sobre as microestruturas de mercado

F.E. b - **Agrupamento de volatilidade:** grandes variações de preço mais provavelmente são seguidas por outras grandes variações, ou seja, existe dependência temporal não linear<sup>5</sup> nos incrementos do preço, e podemos dizer que a volatilidade dos retornos é serialmente correlacionada<sup>6</sup>. Este fato está ligado ao fato dos retornos serem leptocúrticos (distribuição não condicional com curtose superior a da Gaussiana, ver F.E. c a seguir).

F.E. c - **Caudas pesadas:** a distribuição não condicional dos retornos parece possuir uma distribuição de caudas pesadas (leptocúrtica), isto é, com maior frequência de ocorrência de valores extremos do que no caso da distribuição Normal, ainda que a forma exata das caudas seja difícil de determinar. Estes fatos foram descritos inicialmente por Fama (1965) e Mandelbrot (1963), que também foi pioneiro ao descrever F.E. b acima.

F.E. d - **Assimetria:** a distribuição não condicional é negativamente assimétrica, sugerindo que valores extremos negativos são mais frequentes que os positivos<sup>7</sup>.

F.E. e - **Efeito de alavancagem:** conforme notado primeiramente em Black (1976), a maioria das medidas de volatilidade de um ativo são negativamente correlacionadas com os preços passados do mesmo. Desta forma, em geral a volatilidade é maior em momentos de baixa de seus papéis, possivelmente devido ao aumento subsequente da alavancagem da companhia (endividamento em relação ao valor do ativo total) nestes períodos.

F.E. f - **Memória longa e persistência da volatilidade:** além do agrupamento de volatilidade notada acima, um resultado comum está relacionado à alta persistência da volatilidade condicional, especialmente para amostras grandes, como em períodos de tempo longos com

dando suporte teórico a existência do bid-ask bounce, como devido ao custo de transação de um especialista de mercado e aversão ao risco está em Glosten e Milgrom (1985).

<sup>5</sup> De fato, em uma linha de pesquisa mais recente há resultados indicando que não somente a volatilidade dos retornos captada pela variância condicional quanto a assimetria e curtose condicionais também apresentam dependência temporal. Coroneo e Veredas (2006) e Brooks et al. (2005) por exemplo, apontam nessa direção, e Coroneo e Veredas (2006) encontram ainda sazonalidade intradiária para estes momentos de ordem superior, especialmente na curtose, utilizando retornos de ações da bolsa de valores da Espanha. Por ser uma medida de associação entre variáveis aleatórias, ligada a dispersão dos retornos, a curtose condicional intradiária possui relevância maior para modelagem de risco.

<sup>6</sup> Dentre as proxies para volatilidade, o retorno absoluto (módulo dos retornos) parece ser a medida mais correlacionada serialmente. De fato, Ding, Granger e Engle (1993) comparam a função de autocorrelação amostral para várias potências distintas a partir da fórmula  $C_\alpha(\tau) = \text{corr}(|r(t + \tau, \Delta t)|^\alpha, |r(t, \Delta t)|^\alpha)$  e notam que para uma determinada defasagem  $\tau$ , a correlação é maior quando  $\alpha = 1$ , indicando que o retorno absoluto é mais previsível que potências superiores dos retornos.

<sup>7</sup> Segundo Jondeau, Poon e Rockinger (2007) entre outros, a assimetria e o excesso de curtose persistem mesmo após ajustado um modelo que capte a heteroscedasticidade condicional, significando que a distribuição condicional é também não gaussiana. Entretanto, a série dos resíduos padronizados apresenta em geral menor grau desses fenômenos. Ademais, conforme a frequência de amostragem diminui, a distribuição dos retornos se aproxima de uma Normal, fenômeno por vezes chamado de normalidade agregada. Dacorogna et al. (2001), utilizando dados da taxa de câmbio do marco alemão/dólar americano, notam que o peso das caudas aumenta na mesma direção da frequência.

observações diárias<sup>8</sup> ou dados de alta frequência. Estas séries possuem um comportamento que exibe dependência entre observações bastante espaçadas no tempo denominado de memória longa. Definições formais de memória longa fogem do escopo deste trabalho e podem ser encontradas em Palma (2007). Para nossos objetivos, seguindo Granger e Ding (1996) uma série com memória longa é definida como uma em que a função de autocorrelação amostral declina lentamente para zero - a uma taxa hiperbólica - em oposição a um declínio mais rápido, a uma taxa exponencial, nas séries de memória curta.

### ***1.3 Retornos em alta frequência***

#### **1.3.1 Uso de dados em alta frequência e microestruturas**

Estudos empíricos sobre o mercado financeiro brasileiro utilizando dados de alta frequência para modelagem de risco e microestruturas de mercado vêm se tornando mais frequentes. Exemplos recentes estão em Moreira e Lemgruber (2004) ou Laurini, Furlani e Portugal (2008), que investigam seu uso na estimação do valor em risco (VaR) para o índice IBOVESPA e nas microestruturas existentes nos determinantes do spread entre os preços de compra e venda da taxa de câmbio real/dólar americano, respectivamente.

Há evidências de que a incorporação de informação contida em dados de alta frequência melhora as projeções da volatilidade dos retornos diários, ao permitir captar com maior precisão o processo real seguido pela volatilidade dos ativos financeiros. O motivo pelo qual aplicações empíricas de dados em alta frequência são raras ainda na nossa literatura e em tantos outros países é principalmente o custo de obter, selecionar, armazenar e recuperar esses dados. Recentemente, com o desenvolvimento tecnológico computacional, a disponibilidade de dados em alta frequência vem lentamente tornando-se um problema secundário. Em nossa análise, não iremos focar na extensa literatura sobre microestruturas de mercado, mas apresentaremos alguns fenômenos, que chamaremos de fatos estilizados de retornos intradiários, seguindo Taylor (2005) que são de interesse primordial, e em geral cabe a teoria de microestruturas explicarem sua origem.

---

<sup>8</sup> Ver, por exemplo Baillie, Bollerslev e Mikkelsen (1996), que modelam a volatilidade condicional de uma série diária de retornos da taxa de câmbio dólar-marco alemão no período entre março de 1979 e dezembro de 1992, e Bollerslev e Mikkelsen (1996), que utilizam dados diários do índice composto S&P 500 entre 1953 e 1990. Ding, Granger e Engle (1993) usam uma base ainda maior para o mesmo índice, no período entre 1928 e 1991. Baillie e Morana (2007) também utilizam uma série de retornos do S&P 500, entre 1928 e 2007. Todos estes estudos encontram evidência de memória longa nas séries analisadas.

### 1.3.2 Periodicidade intradiária

Wood, McInish e Ord (1985) e Harris (1986) estão entre os primeiros a investigar as propriedades intradiárias dos retornos do mercado acionário norte-americano, documentando o formato em “U” assumido pela volatilidade ao longo do dia, ou seja, a volatilidade assume um pico logo nos primeiros momentos após a abertura do mercado, caindo gradativamente até o horário do almoço para voltar a subir lentamente conforme se aproxima do fechamento. Estudos posteriores documentaram o mesmo padrão de regularidade para outros mercados internacionais com abertura e fechamento bem definidos e padrões sazonais mais complexos no mercado de câmbio<sup>9</sup>, geralmente utilizando como proxy para a volatilidade o retorno absoluto ou elevado ao quadrado. Estes padrões apresentam ainda idiosincrasias para cada dia da semana. De fato, não somente a volatilidade como volume, *spread* e frequência entre *trades* exibem este formato. Goodhart e O'Hara (1997) oferecem uma revisão da literatura sobre o uso de dados em alta frequência em finanças e notam que alguns estudos revelam mais precisamente um formato em “J” reverso, mas de qualquer forma, a evidência é de que há fortes padrões diurnos (ou sazonalidade intradiária) neste tipo de dado, ainda que haja controvérsias entre as causas teóricas destes fenômenos<sup>10</sup>. Taylor (2005) apresenta uma estimação dos efeitos periódicos para o índice de futuros FTSE 100 para cada dia da semana, utilizando retornos de cinco minutos entre 1993 e 1998, obtendo novamente o padrão de um “J” invertido<sup>11</sup>.

Andersen e Bollerslev (1997) e Guillaume, Dacorogna e Pictet (1994) chamam a atenção para que não corrigir os dados por esta sazonalidade intradiária pode acarretar sérios problemas na estimação de modelos para a volatilidade condicional, introduzindo distorções e erros de especificação. Dacorogna, et al. (2001) revela que este componente periódico leva a subestimação das autocorrelações amostrais, gerando a impressão de que não há persistência além de periodicidades específicas. Andersen e Bollerslev (1997) e Dacorogna, et al. (2001)

---

<sup>9</sup> Em especial para o marco alemão/dólar americano e o yen/dólar americano. Ver por exemplo, Andersen e Bollerslev (1997) e Andersen, Bollerslev e Das (2001). O primeiro, para o caso do marco alemão, sendo que os autores notam que a sazonalidade neste caso pode ser vista aproximadamente como a soma de formatos em “U” ao longo do funcionamento do mercado nos principais centros financeiros mundiais.

<sup>10</sup> Segundo Andersen, Bollerslev e Das, (2001) as principais vertentes centram na prevalescência da informação pública versus privada, além do risco e custos enfrentados pelo market maker de manter inventário. Para definições destes fenômenos ver O'Hara (1995).

<sup>11</sup> O autor nota ainda um pico isolado às 13:30 EST (*Eastern Standard Time*) por conta do horário em que muitos dados macroeconômicos são divulgados nos Estados Unidos, adicionando volatilidade momentaneamente. Ademais, estes picos são cerca de duas vezes maiores nas sextas-feiras, sugerindo maior relevância dos anúncios neste dia.

mostram que a extração da volatilidade sazonal é possível, apresentando métodos determinísticos para tal<sup>12</sup>.

## ***2. Modelos de Séries Temporais***

Nesta seção apresentamos alguns dos principais modelos determinísticos utilizados para descrever o comportamento da variância condicional de retornos de ativos especulativos, focando nos modelos cujos resultados receberão ênfase em nossa análise da volatilidade das ações da Petrobrás na seção seguinte, ou seja, os modelos de memória longa.

### ***2.1 Modelos para a variância condicional***

Uma infinidade de modelos para volatilidade está proposta na literatura baseados na seguinte equação para retornos:

$$r_t = \sigma_t z_t \quad (15)$$

Segundo Dacorogna, et al. (2001), distinguimos três tipos de modelagem da variável volatilidade ( $\sigma_t$ ) de uso difundido e debatido: modelos do tipo ARCH, modelos de volatilidade estocástica e modelos baseados na volatilidade realizada.

A idéia básica de modelagem da volatilidade por uma função determinística está representada no modelo ARCH, proposto em Engle (1982). Heteroscedasticidade é o termo estatístico referente à variância não constante. Modelos do tipo ARCH – e suas inúmeras extensões encontradas na literatura – possuem em comum a definição da variância condicional  $\sigma_t^2$  como uma função dos retornos passados. Esta função pode ser simples ou complexa, mas sempre há uma representação equivalente em que  $\sigma_t^2$  ou alguma transformação, em geral  $\ln(\sigma_t^2)$ , está definida em relação aos retornos passados somente.

Em modelos de volatilidade estocástica (VE), a variável  $\sigma_t$  não depende dos retornos passados, não sendo observável ou computável a partir destes. De fato, ela depende de seus

---

<sup>12</sup> Andersen e Bollerslev (1997) utilizam o intervalo intradiário e o fator de volatilidade diário calculado com base na volatilidade intradiária, empregando um tipo de regressão não linear, a Forma Flexível de Fourier. Dacorogna, et al. (2001) utilizam uma abordagem não paramétrica baseada em ondaletas. Ao introduzirem termos harmônicos e polinomiais nas estimativas da sazonalidade, ambas formas são mais suaves que as adotadas em nossa aplicação, baseadas em metodologia empregada em Taylor e Xu (1997); ver seção 4.3. Beltratti e Morana (1999) estimam sazonalidade estocástica para a volatilidade intradiária.



próprios valores defasados, sendo como consequência mais complexa a estimação dos parâmetros desse tipo de modelo. Como nos modelos do tipo ARCH, diversas variantes foram propostas e seu uso é promissor, pois em geral estudos comparando modelos de VE com modelos do tipo ARCH favorecem o primeiro, e as dificuldades técnicas são cada vez menos relevantes dados os avanços computacionais, levando alguns autores a clamarem por sua disseminação, como Gouriéroux e Jasiak (2001). Não obstante, as evidências apontam que em geral os resultados não divergem em elevado grau, e a flexibilidade de modelos ARCH continua oferecendo aos teóricos base para proporem modelos que se ajustem cada vez melhor aos dados, captando diversos dos fatos estilizados mencionados. Novamente, outra referência para o assunto está em Taylor (2005).

Andersen, Bollerslev e Diebold, e Labys (2000) propõem definir  $\sigma_t$  como a volatilidade realizada computada no instante  $t-1$  ao invés de modelar seu comportamento. Esta é calculada através de retornos em alta frequência, porém é empregada no modelo principal em que o intervalo de tempo é maior, geralmente um dia, sendo que a principal vantagem do modelo é não necessitar de suposições de um modelo como nos outros casos, que poderiam estar equivocadas.

Todos os tipos de modelos possuem vantagens e desvantagens, e buscamos aqui apenas demonstrar que a abordagem que adotamos não é a única, ainda que seja atualmente a mais difundida. Neste trabalho seguiremos a linha determinística dos modelos do tipo ARCH, pois devido facilidade de se estimar este tipo de processo, há pacotes econométricos estimando modelos relativamente sofisticados em questão de segundos, ao passo que nos idos da década de 80, modelos mais rudimentares requeriam um período de dias para serem estimados.

### 2.1.1 Modelos de Memória Curta

No modelo ARCH original para a variância condicional, assume-se que os retornos não são correlacionados ( $r_t = \varepsilon_t$  é um ruído branco), e a dependência das inovações é descrita pela seguinte equação:

$$\begin{aligned}\varepsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2\end{aligned}\tag{16}$$

para um ARCH( $q$ ), em que  $\{z_t\}$ , também chamado de erro padronizado, é um processo independente de média zero e variância unitária, ou seja,  $z_t \sim i.i.d. D(0,1)$ , e  $D(\cdot)$  é uma função densidade de probabilidade constante, sobre a qual devemos fazer alguma suposição ao ajustar qualquer um dos modelos expostos. Em nossa aplicação abaixo, usaremos a distribuição Normal, T-student e T-student assimétrica. Notamos que a ordem do processo é dada por  $q$ , e na prática um modelo ARCH apresenta ordem elevada, o que requer a estimação de um número grande de parâmetros para capturar a persistência na volatilidade (F.E. d) e obter um bom ajuste aos dados.

No modelo GARCH de Bollerslev (1986), adiciona-se o componente da variância passada, resultando num modelo que se revelou ser mais parcimonioso em muitas aplicações empíricas com uma parametrização GARCH(1,1). A forma geral que um processo GARCH( $p,q$ ) segue é:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (17)$$

O modelo GARCH captura o chamado agrupamento de volatilidade (F.E. b) e portanto, a dependência de curto prazo encontrada em muitos ativos especulativos; este modelo não capta, no entanto, a ocorrência do efeito alavancagem (F.E. e). Entre os modelos mais utilizados para avaliar a presença deste efeito, temos o GJR, de Glosten, Jagannathan e Runkle (1993) e APARCH, ou ARCH de potência assimétrica. No primeiro modelo da classe ARCH capaz de levar em conta este efeito está em Nelson (1991), que propôs o EGARCH (*exponential* GARCH), que modela o logaritmo da variância condicional, parametrizado como um modelo ARMA, da seguinte forma:

$$\ln(\sigma_t^2) = \omega + [1 - \beta(L)]^{-1} [1 + \alpha(L)] g(z_{t-1}) \quad (18)$$

onde

$$g(z_t) = \theta_1 z_t + \theta_2 [z_t - E(|z_t|)] \quad (19)$$

em que  $g(z_t)$  é uma função da magnitude (captada pelo parâmetro  $\theta_2$ ) e do sinal ( $\theta_1$ ) do erro padronizado. Assim, para  $\theta_1 < 0$ , a variância condicional futura aumentará proporcionalmente mais como resultado de um choque negativo do que positivo, para o mesmo tamanho absoluto do choque. O modelo GJR é escrito em sua forma generalizada como:

$$\sigma_t^2 = \omega + \sum_{i=1}^q (\alpha_i \varepsilon_{t-i}^2 + \gamma_i S_{t-i}^- \varepsilon_{t-i}^2) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2$$

em que  $S_t^-$  é uma variável *dummy* que assume o valor 1 quando  $\varepsilon_t$  é negativo e 0 quando  $\varepsilon_t$  é positivo. Neste modelo, se a estimativa de  $\gamma_i$  for estatisticamente igual a zero, temos que a resposta na volatilidade de choques passados negativos ou positivos é simétrica, não havendo evidências do efeito alavancagem.

Já o modelo APARCH aparece a partir da afirmação em Ding, Granger e Engle (1993) de que não há uma razão óbvia pela qual deve-se assumir que a variância condicional seja uma função linear dos quadrados dos retornos passados ou alternativamente, que o desvio padrão condicional seja função linear dos retornos absolutos, oferecendo então uma forma funcional mais geral capaz de estimar a potência do processo gerador de dados<sup>13</sup>, e que engloba diversos modelos da classe ARCH, como o GJR ou GARCH. A formulação do modelo impõe uma transformação da potência do processo do desvio padrão condicional e dos resíduos absolutos, dada pela seguinte equação:

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta \quad (20)$$

Desta forma, mais uma vez podemos testar se existe assimetria na resposta da volatilidade após choques positivos ou negativos nos retornos pelo parâmetro  $\gamma_i$ .

### 2.1.2 Modelos de Memória Longa

A volatilidade varia em geral de maneira lenta no tempo (F.E. f) e efeitos de um choque na volatilidade em um período podem demorar tempo considerável para se dissipar, sendo comum encontrar autocorrelações amostrais positivas nos retornos ao quadrado para defasagens acima de dois mil períodos em estudos com amostras como as supracitadas<sup>14</sup>.

<sup>13</sup> O parâmetro  $\delta$  permite testar se o modelo real é bem representado por um modelo baseado nos quadrados dos retornos ou outra potência. Os autores estimam o valor de 1,43 para este parâmetro para os dados do estudo (ver nota de rodapé 8), rejeitando as hipóteses nulas que  $\delta$  seja igual a um ou 2 (como nos processos estudados aqui) em favor do APARCH.

<sup>14</sup> Andersen e Bollerslev (1997) notam que as autocorrelações amostrais são em geral negativamente viesadas e se tornam menos precisas quando a defasagem aumenta. Seguindo seus passos, limitamos nossa análise na seção

Modelos da classe ARCH capturam agrupamentos de volatilidade, permitindo reproduzir corretamente o comportamento de choques no curto prazo, uma vez que choques na variância condicional se dissipam a uma taxa exponencial. No modelo GARCH integrado, descrito abaixo, estes choques não se dissipam. Os primeiros modelos determinísticos que permitem captar a dependência de longo prazo, ou memória longa na volatilidade, caracterizada por choques que se dissipam lentamente, são os modelos GARCH integrados fracionalmente.

A equação (17) de um processo GARCH( $p, q$ ) acima pode ser também expressa como um modelo ARMA ( $m, p$ ) para  $[\varepsilon_t]$ , resultando em:

$$[1 - \alpha(L) - \beta(L)]\varepsilon_t^2 = \omega + [1 - \beta(L)]v_t, \quad (21)$$

em que  $m \equiv \max\{p, q\}$ , e  $v_t \equiv \varepsilon_t^2 - \sigma_t^2$  possui média zero e é não correlacionado, ou seja, pode ser interpretado como o processo das inovações da variância condicional. Em muitas aplicações do modelo GARCH, os parâmetros estimados no polinômio auto-regressivo contém uma raiz unitária, indicando a alta persistência da volatilidade, e motivados por esta regularidade empírica, Engle e Bollerslev (1986) propõem um modelo integrado na variância, em que este polinômio é fatorado como  $[1 - \alpha(L) - \beta(L)] = (1 - L)\phi(L)$ , com  $\phi(z) = 0$  possuindo todas as raízes fora do círculo unitário. Temos então o modelo IGARCH( $p, q$ ), escrito sucintamente como:

$$\phi(L)(1 - L)\varepsilon_t^2 = \omega + [1 - \beta(L)]v_t \quad (22)$$

Ainda que nem sempre se rejeite empiricamente que os parâmetros (excluindo a constante) estimados em um GARCH( $p, q$ ) somem a unidade, indicando um processo integrado para a variância, podemos pensar que na maioria dos casos a volatilidade segue um processo que reverte a média, ao passo que no caso de um processo integrado, as previsões de longo prazo continuam sensíveis aos choques para todos os horizontes, ou seja, a persistência de choques na volatilidade é infinita. Para viabilizar um meio termo entre processos de memória curta, estacionários, em que choques se dissipam rapidamente (taxa de decaimento exponencial), e processos integrados, Baillie, Bollerslev e Mikkelsen (1996) introduzem um modelo para variância condicional que permite uma ordem fracional “ $d$ ” de integração, o FIGARCH ( $p, d, q$ ), de *fractionally integrated* GARCH. O equação do processo FIGARCH é

---

4 a períodos de defasagem que podem parecer altos, mas correspondem a uma fração modesta da amostra, bastante abaixo ainda da defasagem máxima de  $t/4$  recomendada em Enders (2003).

obtida utilizando o operador de diferenças fracional no lugar do operador de primeira diferença do modelo para um IGARCH, como a seguir:

$$\phi(L)(1-L)^d \varepsilon_t^2 = \omega + [1 - \beta(L)] v_t \quad (23)$$

Para definição do operador de diferenças fracional, referimos o leitor a Bollerslev e Mikkelsen (1996).

De maneira análoga ao problema encontrado na prática quando estimamos um modelo GARCH, a maior raiz do polinômio estimado  $\beta(x)=1$  em um processo EGARCH é frequentemente bastante próxima da unidade. Bollerslev e Mikkelsen (1996) sugerem fatorar o polinômio autoregressivo como  $[1 - \beta(L)] = \phi(L)(1-L)^d$ , novamente com a restrição que  $\phi(z)=0$  possui todas as raízes fora do círculo unitário obtendo, estendendo o modelo para o logaritmo da variância como na equação a seguir:

$$\ln(\sigma_t^2) = \omega + \phi(L)^{-1} (1-L)^{-d} [1 + \alpha(L)] g(z_{t-1}) \quad (24)$$

Como em Baillie, Bollerslev e Mikkelsen (1996), podemos escrever o FIGARCH da equação (23) como:

$$\sigma_t^2 = \omega [1 - \beta(L)]^{-1} + \left\{ 1 - [1 - \beta(L)]^{-1} \phi(L)(1-L)^d \right\} \varepsilon_t^2 \quad (25)$$

ou seja,  $\sigma_t^2 = \omega^* + \lambda(L) \varepsilon_t^2$ . Se substituirmos  $\lambda(L)$ , por:

$$\left\{ 1 - [1 - \beta(L)]^{-1} \phi(L) [1 + \psi((1-L)^d - 1)] \right\} \quad (26)$$

temos o HYGARCH de Davidson (2004)<sup>15</sup>.

Finalmente, temos o modelo FIAPARCH de Tse (1998), que por sua vez é uma extensão do modelo APARCH utilizando o aparato da integração fracionária. Uma maneira sucinta de escrever o processo da variância condicional no modelo FIAPARCH( $p, d, q$ ) é:

$$\sigma_t^2 = \omega + \left\{ 1 - [1 - \beta(L)]^{-1} \phi(L)(1-L)^d \right\} (|\varepsilon_t| - \varepsilon_t)^\delta \quad (27)$$

---

<sup>15</sup> Este modelo fornece uma generalização do FIGARCH, e podemos notar que quando  $\psi=1$  ou seja,  $\log(\psi)=0$  este modelo equivale ao FIGARCH e quando  $\psi=0$ , equivale ao GARCH.

### **3. Aplicação**

Nesta seção empreendemos uma aplicação empírica dos modelos GARCH, EGARCH, GJR, IGARCH, FIGARCH, FIEGARCH, FIAPARCH e HYGARCH comparando os modelos em sua capacidade de descrição do processo real seguido pela volatilidade, utilizando dados em alta frequência de um ativo negociado na Bolsa de Valores de São Paulo.

#### **3.1 Dados**

A base de dados que será analisada neste trabalho consiste na série de preços de negócios efetuados das ações a vista da Petrobrás S.A. do tipo preferencial, negociadas na Bolsa de Valores de São Paulo (BOVESPA). O sistema de captura de informação em tempo real é provido pela empresa de informações financeiras Bloomberg. Os dados são referentes ao período de 3 de janeiro de 2005 a 13 de abril de 2005, compreendendo um total de aproximadamente 120 mil observações ao longo de 69 dias de pregão (negociação). Além dos preços, a base inclui informações de data, horário (registrado com precisão de um segundo) e volume de cada negócio. A escolha do ativo Petrobrás se deve primordialmente a sua liquidez (ação com maior volume negociado na BOVESPA durante o período), sendo que há em média, um negócio a cada 15 segundos com um volume médio de cerca de 700 ações ou R\$72.500. Houve pagamento de proventos (dividendos e complemento de juros sobre capital próprio) no dia 15 de fevereiro, o que impacta o preço de abertura das negociações, porém isto não afeta a série de retornos.

A série analisada primordialmente, construída a partir desta série original, consiste nos log-retornos calculados com base nos últimos preços registrados em cada intervalo consecutivo de cinco minutos<sup>16</sup>, desde o horário de abertura das operações da bolsa, encerrando no período pós fechamento das negociações normais (after-market)<sup>17</sup>. Em geral são registrados preços em 85 intervalos diferentes consecutivos por dia, incluindo o after-market, havendo 4 dias com 86 intervalos e dois dias com horário de negociação reduzido

---

<sup>16</sup> Em geral, utilizam-se retornos calculados em intervalos de cinco, 10 ou 15 minutos. Andersen (2000), nota que a frequência de cinco minutos é aproximadamente a mais alta para a qual as propriedades da série não são distorcidas por microestruturas de mercado. Optou-se aqui por trabalhar com retornos de 5 minutos pois não houve evidência de autocorrelação além da primeira ordem, a ainda pela comparabilidade com outros estudos realizados nesta frequência, sendo que em um texto recente, Taylor (2005) nota que frequências de amostragem de cinco minutos são provavelmente a escolha mais popular nestes estudos. O autor nota ainda que apesar dos ruídos provocados por microestruturas, frequências maiores carregam maior conteúdo informacional, citando estudos de volatilidade que defendem uso de retornos na faixa de 5 a 30 minutos.

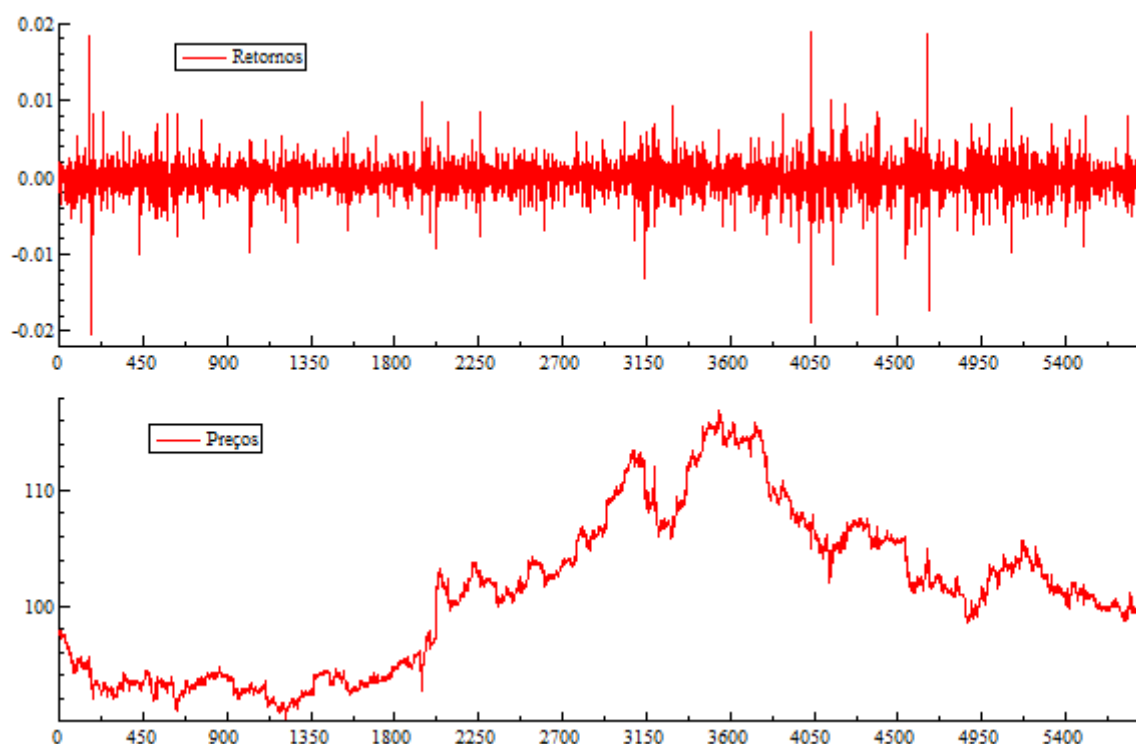
<sup>17</sup> Devido a ajustes oriundos do horário de verão, o horário de abertura e fechamento da bolsa no período da amostra fica compreendido entre 10h e 17h, até 28 de janeiro, e das 11h as 18h, de 2 de fevereiro até o final da amostra.

(pregões de 31 de janeiro e 9 de fevereiro, com 69 e 63 intervalos respectivamente). Apesar da liquidez do ativo analisado, existe ao longo da amostra um total de 21 intervalos sem observações de negócios na base original durante o horário de funcionamento normal. Neste caso, e também para compor os retornos de eventuais lacunas entre o último preço do intervalo final do horário regular no dia e o último preço registrado no after-market, (em geral registrados no segundo intervalo de cinco minutos após o encerramento do horário normal), foi utilizada a média aritmética simples (não ponderada por volume). Procedimento semelhante foi utilizado em Andersen e Bollerslev, (1997) em seu estudo sobre a volatilidade com dados de contratos futuros do índice S&P500, gerando retornos idênticos entre os intervalos em questão, o que suaviza a série ao redor dos intervalos vazios, mitigando o efeito de eventuais alterações bruscas após o retorno da atividade<sup>18</sup>. Finalmente, não foram utilizados os retornos computados entre a última informação de preço de uma data qualquer e a primeira do dia subsequente, seguindo Andersen e Bollerslev (1997), uma vez em que este intervalo incorpora ajustes ao conjunto de informações recebido entre o encerramento de um pregão e o início do próximo, constituindo um retorno *overnight* cuja variabilidade média é bastante superior a de qualquer outro intervalo<sup>19</sup>. Desta forma, obtemos uma série completa com um total de 5.831 log-retornos na frequência de cinco minutos. Uma segunda série de retornos, que chamaremos de homogênea, foi construída a partir desta série original, excluindo-se os últimos retornos dos dias com 86 retornos, bem como os dois dias com horário de mercado reduzido, totalizando 5.695 retornos intradiários. Abaixo, vemos o gráfico da série completa de preços e retornos:

---

<sup>18</sup> Ainda que este não pareça ser o caso, já que os preços subsequentes a um intervalo vazio são em geral bastante próximos ou iguais aos últimos preços registrados, mesmo nos casos em que há dois intervalos seguidos sem registro de *trade*. Provavelmente, isto ocorre pois enquanto na base em Andersen e Bollerslev (1997) os intervalos vazios decorrem da suspensão das negociações (em caso de índice de mercado caindo acima de um determinado limite, com aumento da volatilidade), em nosso trabalho eles são provenientes de momentos de pouca liquidez.

<sup>19</sup> Andersen e Bollerslev (1997) notam que seria possível ajustar estes retornos específicos utilizando uma variável dummy para este intervalo, porém o procedimento seria *ad hoc* por natureza, e ademais, investigações informais revelaram pouca sensibilidade ao tratamento empregado para este intervalo.



**Figura 1 - Retornos e Preços da ação da Petrobrás**

### **3.2 Análise descritiva da série de retornos**

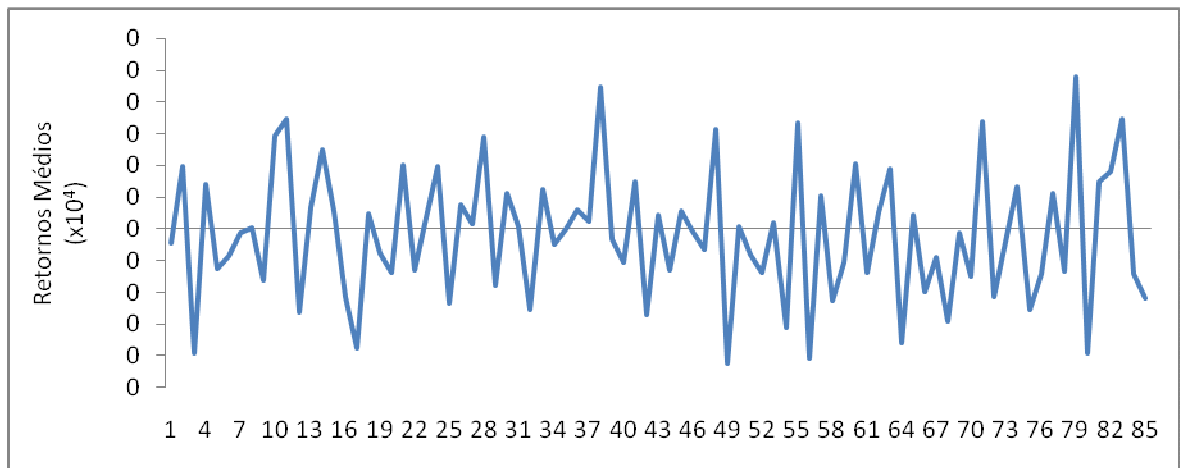
Cabe apresentar uma breve análise descritiva da série de dados. A média dos log-retornos calculados a partir da série homogênea com 5.695 retornos é  $-1,8 \cdot 10^{-5}$ , estatisticamente indistinguível de zero a níveis de confiança usuais, dado o desvio padrão de  $1,8 \cdot 10^{-3}$ . A amplitude atinge 3,916%, com o maior retorno de 5 minutos sendo 1,89% e o menor, -2,02%. A assimetria é -0,28, e o excesso de curtose, 12,65. O teste de Jarque & Bera possui estatística 38.964, sendo rejeitada a hipótese de normalidade a qualquer nível de significância prático<sup>20</sup>.

Os retornos médios durante os 85 intervalos ao longo do dia são em geral muito próximos de zero, não havendo horários específicos em que pareça haver evidência de retornos anormais, como pode ser visto a seguir<sup>21</sup>:

<sup>20</sup> Estas estatísticas são condizentes com outros estudos utilizando retornos de 5 minutos, como Andersen e Bollerslev (1997) e Chan, Chan e Karolyi (1991), ainda que estes se baseiem em dados de um índice de ações (S&P500) e contratos futuros deste índice. Os desvios padrões encontrados nestes estudos são consideravelmente menores (entre 0,044% e 0,104%) , devido a se tratar de um índice composto por papéis de muitas companhias. Para o mercado brasileiro, o autor desconhece estudos com estatísticas descritivas baseadas em retornos de cinco minutos para qualquer ativo.

<sup>21</sup> Andersen e Bollerslev (1997) encontram retornos médios positivos e estatisticamente diferentes de zero a 5% em especial logo após a abertura dos mercados para o índice S&P500 (mas não para a taxa de câmbio marco





**Figura 2 - Retornos médios por intervalo intradiário de 5 minutos**

### 3.3 Fatores sazonais

Conforme vimos na seção 1.3.2, a filtragem do componente sazonal intradiário da volatilidade se faz necessária antes que se possa efetuar corretamente a modelagem estatística proposta nesta seção. Utilizando a série homogênea de retornos de 5 minutos<sup>22</sup>, podemos calcular o componente sazonal para cada intervalo a partir da média dos retornos ao quadrado de cada intervalo ao longo da amostra, uma vez que retornos ao quadrado oferecem estimativas simples de variabilidade. Giot (2000) utiliza intervalos de 30 minutos obtendo um padrão bastante cru que é então suavizado via interpolação por cubic spline. Moreira e Lemgruber (2004), utilizando retornos de 15 minutos, procedem um alisamento inicial em seus fatores sazonais, calculando como fator para o  $j$ -ésimo intervalo, a média dos quadrados dos retornos nos intervalos  $j$ ,  $j+1$  e  $j-1$ , obtendo uma janela móvel horizontal de 45 minutos, utilizando ainda procedimentos adicionais para suavizar os filtros. Estes autores calculam ainda também sazonais para cada dia da semana, de maneira a captar diferenças na periodicidade intrasemanal da volatilidade<sup>23</sup>. Devido a sua simplicidade, adotaremos para o

---

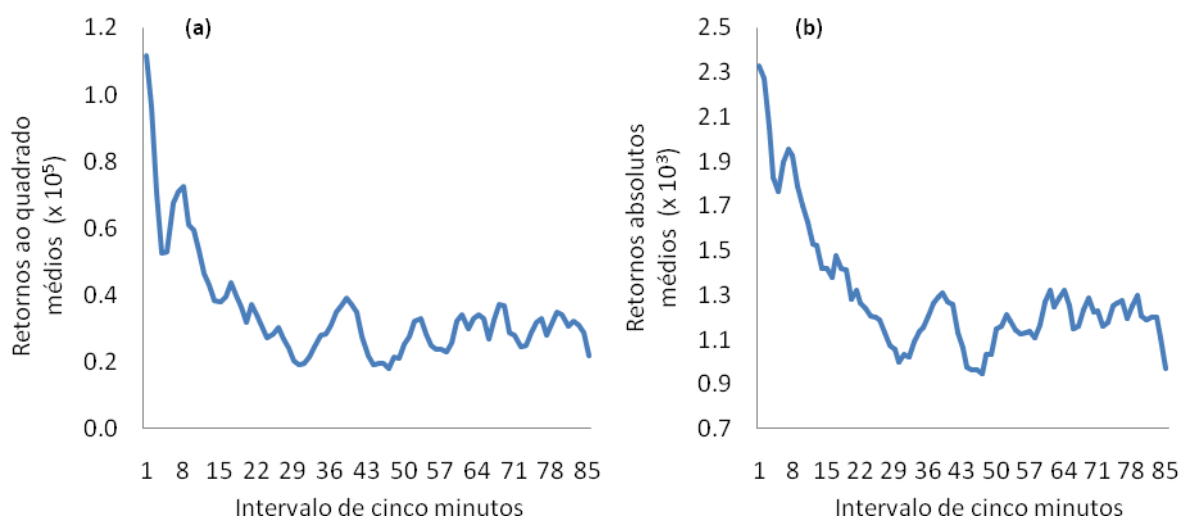
alemão/dólar americano), a partir das violações da banda de confiança de 5% para a hipótese nula de uma série i.i.d.

<sup>22</sup> A série de retornos utilizada aqui não é a completa, pois devido aos diferentes horários de abertura e encerramento do pregão, bem como à presença de dias com pregão reduzido ao longo do período de nossa amostra, foi julgado um procedimento mais adequado utilizarmos para o cálculo dos fatores a série de retornos homogeneizada para 85 intervalos por dia. Com isso deixamos de incluir 136 retornos em nossa análise, porém as séries possuem características bastante semelhantes. De fato, foram testados os mesmos modelos para a série completa dessazonalizada, e os resultados (não reportados) foram bastante semelhantes.

<sup>23</sup> Giot (2000) e Moreira e Lemgruber (2004) usam séries com 13 e 14 semanas na construção dos fatores sazonais, respectivamente, e confirmam que os padrões são consideravelmente diferentes entre os dias da

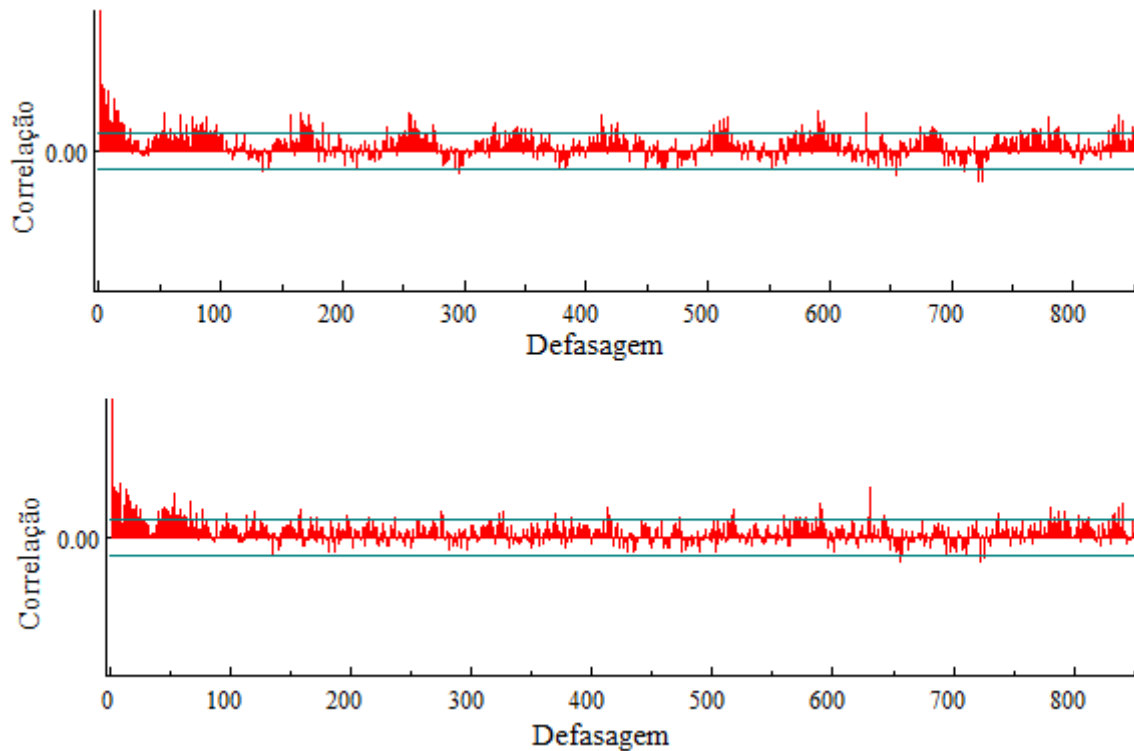
cálculo dos fatores sazonais a linha determinística proposta em Taylor e Xu (1997) e seguida em Moreira e Lemgruber (2004).

Procedemos da seguinte forma para o cálculo dos fatores sazonais intradiários. Computamos inicialmente a média aritmética simples dos quadrados dos retornos em cada intervalo. Para cada um dos 85 intervalos de 5 minutos, um suavizamento por médias móveis considerando o intervalo anterior e o subsequente fornece o componente sazonal intradiário. A série de retornos dessazonalizada utilizada em nossa análise a seguir é obtida dividindo os log-retornos originais pela raiz quadrada dos fatores sazonais respectivos aos mesmos intervalos. Alternativamente, foram calculados fatores sazonais para cada intervalo em cada dia da semana, e detalhes sobre sua construção e os resultados dos modelos estimados com retornos corrigidos por estes fatores são reportados no apêndice . Abaixo, mostramos o gráfico (a) dos retornos ao quadrado médios suavizados em cada um dos 85 intervalos intradiários ao longo dos 67 dias da amostra, em que podemos observar o formato em “J” invertido, conforme esperado; estes são nossos fatores de filtragem do componente periódico intradiário. Apenas para efeito de comparação foram calculados os retornos absolutos médios, vistos abaixo no gráfico (b):



**Figura 3 - Retornos intradiários médios suavizados: (a) ao quadrado, (b) absolutos**

Abaixo vemos o correlograma amostral para a série de retornos antes e depois da filtragem do componente sazonal, onde nota-se que a periodicidade intradiária foi praticamente eliminada:



**Figura 4 - FAC dos retornos absolutos: (a) antes da dessazonalização, (b) depois da filtragem**

### 3.4 Resultados

Todos os resultados reportados na tabela abaixo foram estimados usando o módulo G@RCH 5.0 do pacote econométrico OxMetrics versão 5, sendo apresentados os resultados para a série homogeneizada de 67 dias com 85 intervalos cada, após filtragem pelos fatores sazonais calculados de acordo com a metodologia na sub-seção anterior. Foram testadas algumas parametrizações para a média condicional dos retornos, e concluímos que um modelo AR(1) sem constante é suficiente para ajustar sua estrutura de dependência. De acordo com o modelo proposto para a variância condicional, o coeficiente de autocorrelação de primeira ordem estimado varia aproximadamente de -0,1 a -0,08, sendo em todos os casos altamente significativo, e seu valor negativo corrobora o resultado esperado devido ao efeito de

microestrutura do *bid-ask bounce*. Já as autocorrelações de defasagem maior que a unidade são em geral não significativas a um nível de significância de 5%<sup>24</sup>. Na tabela abaixo comparamos os modelos ajustados com a suposição de que o resíduo padronizado segue uma distribuição Normal :

**Tabela 1 - Estatísticas e medidas de qualidade de ajuste para os modelos estimados (distribuição Normal)**

	GARCH(1,1)	EGARCH(1,1)	GJR(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)	FIEGARCH(1,d,1)	FIAPARCH(1,d,1)	HYGARCH(1,d,1)
LL	-7714.240	-7681.267	-7713.332	-7755.547	-7678.121	-7675.107	<b>-7674.942</b>	-7677.944
AIC	2.710532	2.699655	2.710564	2.724687	2.698199	2.697843	<b>2.697785</b>	2.698488
BIC	2.715201	2.706658	2.716400	2.728189	<b>2.704035</b>	2.706013	2.705955	2.705491
Assimetria	-0.14816	-0.14805	-0.15164	-0.15734	-0.12833	-0.12608	-0.13253	-0.12965
Ex. Curtose	2.1784	2.0877	2.1841	2.3959	2.0677	2.0754	2.0649	2.0618
Q(10)	0.0000170	0.0451195	0.0000103	0.0000001	0.4278356	0.1753631	0.3638225	0.4398221
Q(50)	0.0031314	0.1120537	0.0028391	0.0000319	0.4517712	0.2492684	0.4201687	0.4572387
Q(425)	0.0533411	0.0936276	0.0499174	0.0189230	0.3996160	0.1645655	0.3244611	0.3962082
Q(850)	0.0684681	0.1485989	0.0615476	0.0223342	0.4589758	0.1857197	0.3846768	0.4600056

LL denota o log da verossimilhança; AIC e BIC são os critérios de informação de Akaike e Schwarz, respectivamente. Q(l) representa o p-valor da estatística de Box-Pierce para o quadrado dos resíduos padronizados, em que aceitamos a hipótese nula de não correlação serial das "l" primeiras defasagens a um nível de significância qualquer quando o p-valor é maior que este nível. Valores em destaque denotam o melhor ajuste.

A seguir vemos a mesma tabela assumindo que  $[z_t]$  segue uma distribuição T-student, notando que não houve convergência nos modelos EGARCH e FIEGARCH:

**Tabela 2 - Estatísticas e medidas de qualidade de ajuste para os modelos estimados (distribuição T-student)**

	GARCH(1,1)	EGARCH(1,1)	GJR(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)	FIEGARCH(1,d,1)	FIAPARCH(1,d,1)	HYGARCH(1,d,1)
LL	-7496.236	-	-7496.004	-7510.665	-7469.547	-	<b>-7467.918</b>	-7469.542
AIC	2.634323	-	2.634593	2.639040	<b>2.625302</b>	-	2.625432	2.625651
BIC	2.640160	-	2.641597	2.643708	<b>2.632305</b>	-	2.634770	2.633822
Assimetria	-0.14637	-	-0.14910	-0.16719	-0.12560	-	-0.13038	-0.12528
Ex. Curtose	2.1803	-	2.1834	2.3283	2.0793	-	2.0735	2.0809
Q(10)	0.0000186	-	0.0000126	0.0000000	0.4370029	-	0.4042945	0.4364116
Q(50)	0.0032725	-	0.0029568	0.0000357	0.4451208	-	0.4305609	0.4446133
Q(425)	0.0565772	-	0.0524903	0.0173193	0.3658179	-	0.2967428	0.3656655
Q(850)	0.0854167	-	0.0771946	0.0188715	0.4402658	-	0.3871416	0.4385237

Ver nota de rodapé da tabela anterior.

E abaixo os resultados análogos, supondo distribuição T-student assimétrica:

<sup>24</sup> As estatísticas do teste de Box-Pierce para não auto-correlação serial nos resíduos padronizados ao quadrado são reportadas até a defasagem máxima de ordem 850, correspondendo a 10 dias.

**Tabela 3 - Estatísticas e medidas de qualidade de ajuste para os modelos estimados (distribuição T-student assimétrica)**

	GARCH(1,1)	EGARCH(1,1)	GJR(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)	FIEGARCH(1,d,1)	FIAPARCH(1,d,1)	HYGARCH(1,d,1)
LL	-7494.890	-7469.357	-7494.654	-7509.594	-7468.435	<b>-7466.470</b>	-7466.715	-7468.434
AIC	2.634202	2.625938	2.634470	2.639014	<b>2.625263</b>	2.625275	2.625361	2.625613
BIC	2.641205	2.635275	2.642641	2.644851	<b>2.633433</b>	2.635780	2.635866	2.634951
Assimetria	-0.14677	-0.14462	-0.14949	-0.16715	-0.12579	-0.12598	-0.13072	-0.12564
Ex. Curtose	2.1805	2.1134	2.1836	2.3287	2.0792	2.1490	2.0739	2.0800
Q(10)	0.0000172	0.0478653	0.0000118	0.0000000	0.4375617	0.3352701	0.4012823	0.4372569
Q(50)	0.0031720	0.1002428	0.0028695	0.0000354	0.4458030	0.3423728	0.4298237	0.4455509
Q(425)	0.0556541	0.0737152	0.0515805	0.0172524	0.3665179	0.1218719	0.2948459	0.3664662
Q(850)	0.0835489	0.1397771	0.0756110	0.0189044	0.4408292	0.2667778	0.3849019	0.4400450

Da mesma forma, apresentamos abaixo os resultados das estimativas para os parâmetros de cada modelo, conforme especificações na seção 2.3.2 com a suposição de normalidade condicional dos resíduos padronizados:

**Tabela 4 - Estimativas dos parâmetros (distribuição Normal)**

	GARCH(1,1)	EGARCH(1,1)	GJR(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)	FIEGARCH(1,d,1)	FIAPARCH(1,d,1)	HYGARCH(1,d,1)
$\varphi$	-0.094635 (-0.015769)	-0.098641 (-0.01538)	-0.096095 (-0.015698)	-0.100585 (-0.014072)	-0.094679 (-0.015403)	-0.098314 (-0.015318)	-0.096712 (-0.015649)	-0.094382 (-0.015497)
$\omega$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\omega$	0.07147* (-0.036848)	0.13741* (-0.093674)	0.06729 (-0.033742)	0.017337* (-0.0089052)	0.047142 (-0.018774)	0.433568* (-0.23396)	0.067315 (-0.02192)	0.056339* (-0.029186)
	0.0525	0.1425	0.0462	0.0516	0.0121	0.0639	0.0021	0.0536
$\alpha$	0.103837 (-0.029731)	-0.623483 (-0.083126)	0.091698 (-0.032452)	0.106467 (-0.033826)	0.681969 (-0.094513)	-0.808126 (-0.10474)	0.589947 (-0.098563)	0.646132 (-0.11815)
	0.0000	0.0000	0.0047	0.0017	0.0000	0.0000	0.0000	0.0000
$\beta$	0.823684 (-0.065296)	0.969027 (-0.011607)	0.831039 (-0.060814)	0.893533	0.745548 (-0.081835)	0.913651 (-0.041027)	0.722841 (-0.067478)	0.731657 (-0.081964)
	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000
$\theta_1$		-0.026114* (-0.019378)				-0.017362* (-0.019912)		
		0.1778				0.3833		
$\theta_2$		0.335054 (-0.041486)				0.344081 (-0.041213)		
		0.0000				0.0000		
$d$					0.247523 (-0.049803)	0.336223 (-0.10911)	0.327658 (-0.074488)	0.292182 (-0.10869)
					0.0000	0.0021	0.0000	0.0072
$\gamma$			0.017586* (-0.018501)				0.040782* (-0.056302)	
			0.3419				0.4689	
$\delta$							1593079 (-0.217)	
							0.0000	
$\ln(\Psi)$								-0.066888* (-0.1288)
								0.6036

Para cada parâmetro, a tabela reporta a estimativa de Quasi Máxima Verossimilhança, seguida pelo erro padrão robusto (entre parênteses) e o p-valor. O sinal "\*" denota parâmetro não significativo a 5% de significância.

Abaixo, os resultados considerando a distribuição T-student:

**Tabela 5 - Estimativas dos parâmetros (distribuição T-student)**

	GARCH(1,1)	EGARCH(1,1)	GJR(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)	FIEGARCH(1,d,1)	FIAPARCH(1,d,1)	HYGARCH(1,d,1)
G.L.	4.519525 (0.27407) 0.0000	- - 0.0000	4.525785 (0.27518) 0.0000	3.781841 (0.23322) 0.0000	4.604138 (0.28393) 0.0000	- - 0.0000	4.650254 (0.28383) 0.0000	4.600036 (0.28257) 0.0000
$\varphi$	-0.081732 (0.014648) 0.0000	- - 0.0000	-0.081984 (0.014585) 0.0000	-0.084981 (0.013781) 0.0000	-0.079290 (0.013995) 0.0000	- - 0.0000	-0.079180 (0.014161) 0.0000	-0.079341 (0.014009) 0.0000
$\omega$	0.077637* (0.056758) 0.1714	- - 0.0000	0.074610* (0.052587) 0.1560	0.027002* (0.014430) 0.0614	0.040385 (0.013105) 0.0021	- - 0.0000	0.060177 (0.019662) 0.0022	0.038685* (0.020193) 0.0554
$\alpha$	0.114548 (0.047559) 0.0160	- - 0.0000	0.105075 (0.047842) 0.0281	0.105879 (0.036286) 0.0035	0.696655 (0.073590) 0.0000	- - 0.0000	0.610105 (0.094019) 0.0000	0.702986 (0.092277) 0.0000
$\beta$	0.815911 (0.098969) 0.0000	- - 0.0000	0.821475 (0.092690) 0.0000	0.894121	0.783733 (0.057512) 0.0000	- - 0.0000	0.761329 (0.058457) 0.0000	0.786501 (0.061229) 0.0000
$d$					0.301261 (0.054000) 0.0000	- - 0.0000	0.373574 (0.076185) 0.0000	0.293710 (0.091488) 0.0013
$\gamma$			0.012689* (0.017917) 0.4788				0.015340* (0.048579) 0.7522	
$\delta$							1.603345 (0.23890) 0.0000	
$\ln(\psi)$								0.011262* (0.11048) 0.9188

Para cada parâmetro, a tabela reporta a estimativa de Quasi Máxima Verossimilhança, seguida pelo erro padrão robusto (entre parênteses) e o p-valor. O sinal "\*" denota parâmetro não significativo a 5% de significância. G.L. denota o número de graus de liberdade estimado para a distribuição condicional dos resíduos padronizados.

E finalmente, a tabela a seguir reporta as estimativas sob a suposição de que  $z_t$  segue uma distribuição T-student assimétrica:

**Tabela 6 - Estimativas dos parâmetros (distribuição T-student assimétrica)**

	GARCH(1,1)	EGARCH(1,1)	GJR(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)	FIEGARCH(1,d,1)	FIAPARCH(1,d,1)	HYGARCH(1,d,1)
$\ln(\text{assim.})$	-0.024004* (0.012908) 0.0630	-0.023597* (0.012967) 0.0688	-0.024052* (0.012909) 0.0625	-0.020879* (0.012668) 0.0994	-0.021916* (0.012960) 0.0909	-0.021148* (0.011929) 0.0763	-0.022810* (0.012985) 0.0790	-0.021891* (0.012975) 0.0916
G.L.	4.537494 (0.27610) 0.0000	4.558215 (0.28063) 0.0000	4.543842 (0.27721) 0.0000	3.788811 (0.23451) 0.0000	4.619193 (0.28635) 0.0000	4.600386 (0.28357) 0.0000	4.666891 (0.28627) 0.0000	4.617262 (0.28484) 0.0000
$\varphi$	-0.081211 (0.014631) 0.0000	-0.077324 (0.013896) 0.0000	-0.081370 (0.014570) 0.0000	-0.084460 (0.013795) 0.0000	-0.078857 (0.013995) 0.0000	-0.077241 (0.013885) 0.0000	-0.078702 (0.014158) 0.0000	-0.078881 (0.014011) 0.0000
$\omega$	0.077061* (0.054998) 0.1612	-0.192673* (0.11335) 0.0892	0.074170* (0.051052) 0.1463	0.026946* (0.014276) 0.0591	0.040569 (0.013215) 0.0022	-0.338250* (0.28043) 0.2278	0.060766 (0.019615) 0.0020	0.039763* (0.020519) 0.0527
$\alpha$	0.113772 (0.046186) 0.0138	-0.665554 (0.066616) 0.0000	0.104379 (0.046417) 0.0246	0.105960 (0.035958) 0.0032	0.696431 (0.073700) 0.0000	-0.421556* (0.36369) 0.2465	0.607058 (0.093776) 0.0000	0.699437 (0.092884) 0.0000
$\beta$	0.816837 (0.096134) 0.0000	0.974214 (0.0092816) 0.0000	0.822185 (0.090172) 0.0000	0.894040	0.782674 (0.057900) 0.0000	0.369084* (0.27914) 0.1861	0.759836 (0.058441) 0.0000	0.783982 (0.061826) 0.0000
$\theta_1$		-0.014964* (0.019172) 0.4351				-0.007981* (0.018675) 0.6691		
$\theta_2$		0.374416 (0.037198) 0.0000				0.387936 (0.037799) 0.0000		
$d$					0.299450 (0.053950) 0.0000	0.524425 (0.082463) 0.0000	0.374227 0.076490 0.0000	0.295867 (0.091558) 0.0012
$\gamma$			0.012722* (0.017745) 0.4735				0.016161* (0.048549) 0.7392	
$\delta$							1.593360 (0.23727) 0.0000	
$\ln(\Psi)$								0.005301* (0.10883) 0.9612

Para cada parâmetro, a tabela reporta a estimativa de Quasi Máxima Verossimilhança, seguida pelo erro padrão robusto (entre parênteses) e o p-valor. O sinal "\*" denota parâmetro não significativo a 5% de significância. G.L. denota o número de graus de liberdade estimado para a distribuição condicional dos resíduos padronizados.

Os resultados que ressaltamos são: 1) o parâmetro  $d$  é altamente significativo em todos os modelos integrados fracionalmente ajustados, indicando que o processo gerador da variância condicional possui de fato memória longa<sup>25</sup>; 2) enquanto que nos modelos de memória curta podemos ver pelas estatísticas de Box-Pierce que resta estrutura nos resíduos padronizados ao quadrado (ou seja, mesmo após ajustado este tipo de modelo para a variância condicional, ainda há correlação serial nos quadrados dos retornos<sup>26</sup>), nos modelos de

<sup>25</sup> Evidência adicional é fornecida pelo teste GPH aplicado na série dos quadrados dos retornos dessazonalizados antes da modelagem, que estima o parâmetro  $d$  como 0,1436 com desvio padrão de 0,01267 e p-valor de 0,0000. Para detalhes sobre este teste, ver Geweke e Porter-Hudak (1983).

<sup>26</sup> Podemos ver ainda nos resultados de todos os modelos ajustados para os retornos filtrados que, em acordo com a nota de rodapé 7, as curtoses e assimetrias amostrais dos resíduos padronizados ainda não são compatíveis com a distribuição Normal, ainda que menores do que as mesmas estatísticas não condicionais da série de retornos dessazonalizados, que possui assimetria de -0,19896 e excesso de curtose de 2,85067.

memória longa não rejeitamos que as autocorrelações sejam em conjunto nulas até uma defasagem elevada, a qualquer grau de significância usual; 3) a qualidade superior dos modelos que permitem em sua formulação captar a presença de memória longa no processo, o que pode ser visto pela melhora no valor da log-verossimilhança, que recompensa modelos que ajustem melhor os dados ao processo gerador destes, e ainda nos critérios de informação<sup>27</sup>; 4) quando permitimos que os resíduos padronizados possuam distribuição condicional T-student, que permite captar o elevado peso nas caudas observado nesse tipo de dado, obtemos melhoras significativas na verossimilhança dos modelos ajustados – porém uma distribuição T-student assimétrica produz melhorias apenas marginais, sendo que para todos os modelos estimados neste caso não rejeitamos que o parâmetro de assimetria seja não significativo a 5%, e finalmente<sup>28</sup>; 5) não significância dos parâmetros  $\theta_1$  (EGARCH e FIEGARCH) e  $\gamma$  (GJR e FIAPARCH), capazes de captar o efeito de alavancagem (F.E. e).

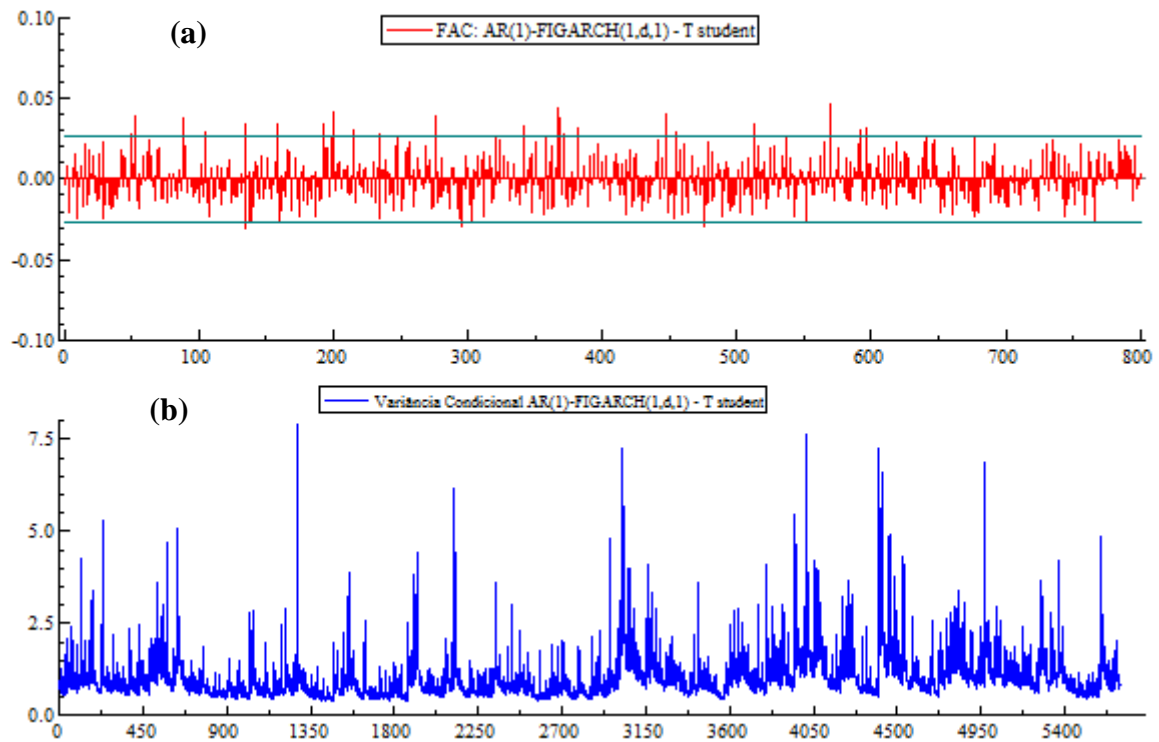
Mostramos abaixo o gráfico do processo para a variância condicional modelada por um AR(1)-FIGARCH(1,d,1) com distribuição condicional T-student dos resíduos padronizados:

---

<sup>27</sup> Aqui notamos ainda que não possa rejeitar o FIGARCH como modelo verdadeiro em favor do modelo HYGARCH, uma vez que a estimativa de  $\log(\psi)$  não é estatisticamente diferente de zero.

<sup>28</sup> Uma vez que a distribuição T assimétrica equivale a T quando a assimetria é igual a um, o que ocorre no caso do logaritmo do parâmetro de assimetria ser zero. Veja Bauwens e Laurent (2005) para mais detalhes sobre o uso de distribuição não gaussianas para a estrutura dos modelos ARCH.





**Figura 5 - Resultados do Modelo AR(1)-FIGARCH(1,d,1) T student: (a) FAC dos Resíduos Padronizados ao Quadrado; (b) Variância Condicional**

### 3.5 Implicações dos Resultados

Os resultados obtidos indicam que os modelos que captam a memória longa são mais adequados para descrever o comportamento verdadeiro do processo seguido pela volatilidade, quando comparados aos modelos de memória curta mais largamente empregados. Este resultado era esperado devido ao fato estilizado F.E. f, que caracteriza a ocorrência freqüente na prática da persistência da volatilidade, que muda ao longo do tempo de maneira relativamente lenta. Ainda, o fato estilizado do efeito alavancagem (F.E. e) não foi suportado pelos dados do estudo, sugerindo que em intervalos curtos de tempo, não haja correlação bem definida entre a volatilidade e movimentos de alta ou queda dos ativos da empresa.

Ainda que os modelos da classe ARCH ou volatilidade estocástica tenham amplo sucesso em descrever o processo seguido pela volatilidade, esta na prática não é observável e projeções feitas diretamente com estes modelos devem ser utilizadas com precaução, especialmente quando o horizonte de previsão desejado passa de alguns dias. Não obstante,

medidas de volatilidade histórica podem ser construídas com sucesso e são de fato empregadas em gerenciamento de risco de mercado. Previsões de volatilidade podem ser obtidas estimando modelos ARCH através da informação presente nos retornos passados e volatilidades implícitas nos preços de mercado das opções compra ou venda de uma ação ou taxa de câmbio, e neste caso, a constatação obtida neste estudo de que o processo da volatilidade apresenta memória longa representa um ponto de interesse prático.

Ao menos em tese, modelos superiores em descrever o comportamento da volatilidade de um ativo, como o caso do FIGARCH e FIEGARCH estão relacionados a melhores processos na gestão do risco deste ativo. Para detalhes sobre a construção das medidas de volatilidade implícita e previsão da mesma usando modelos ARCH de memória longa, ver Taylor (2005). Finalmente, a evidência de memória longa na volatilidade motiva a utilização destes efeitos em métodos de precificação de opções. Taylor (2005) fornece uma revisão do arcabouço metodológico apresentado para esta incorporação. De fato, ainda que sugerindo investigações mais detalhadas sobre os resultados, Bollerslev e Mikkelsen (1996) estudam o fenômeno no escopo de precificação de opções, concluindo que decisões de alocação de carteiras de investimento podem se tornar extremamente sensíveis ao horizonte de investimento na presença de memória longa, e estas dependências de longo alcance devem ser levadas em conta na definição de estratégias ótimas de proteção.

#### **4. CONCLUSÃO**

Neste estudo, mostramos que existe evidência de memória longa nos retornos de alta frequência de um ativo de grande liquidez negociados na principal bolsa de valores nacional, e que os modelos FIGARCH, FIEGARCH, FIAPARCH e HYGARCH são capazes de captar este fenômeno, descrevendo o processo seguido pelo quadrado dos retornos, nossa proxy para volatilidade, de maneira mais adequada do que modelos tradicionais de memória curta. Estes resultados, esperados teoricamente e confirmados pelos dados podem ser utilizados na prática com sucesso em aplicações como gerenciamento de risco de investimentos especulativos. No entanto, não foram encontradas evidências da presença do chamado efeito de alavancagem.

Não obstante os resultados significativos e animadores obtidos em nossa aplicação, cabe fazer algumas considerações. Ainda que haja justificativas para preferir modelar a volatilidade de retornos de 5 minutos, outros estudos apontam que frequências maiores são preferíveis, como Giot (2000), que defende que uma amostragem a cada 15 minutos constitui

intervalos ótimos para o cômputo dos retornos. Em outro ponto, uma amostra maior contribuiria de maneira a validar intervalos mais longos mantendo um número total de retornos elevado, o que é desejável, por exemplo, para corroborar ou não o resultado das estimações globais via subamostras, e daria robustez aos fatores de dessazonalização calculados. A maior crítica ao método utilizado talvez recaia justamente sobre esses fatores, uma vez que os resultados são sensíveis ao filtro utilizado para extrair o componente intradiário da volatilidade, e é possível que estimativas mais precisas sejam obtidas com filtros que ex-post provem-se mais adequados.

Ademais, como mencionamos, existem diversas abordagens para estimação ou cálculo de volatilidade, e adotamos apenas uma dessas linhas. Mesmo dentro dessa metodologia, diversos outros modelos com distribuições condicionais dos erros padronizados mais apropriadas a dados reais de retornos ou parametrizações mais recentes, que oferecem vantagens (ao menos em teoria) sobre as empregadas aqui estão disponíveis na literatura. A título de exemplo, mencionamos o modelo AFIGARCH, proposto em Baillie and Morana (2007). O primeiro autor, devemos notar, é um dos responsáveis pela criação dos modelos originais de memória longa na variância condicional, mais de uma década previamente. Neste novo modelo, é feita uma extensão que permite uma adaptação de um processo FIGARCH de maneira a captar além da memória longa, quebras estruturais no processo de volatilidade, com resultados superiores ao do modelo FIGARCH utilizando simulações de Monte Carlo.

## REFERÊNCIAS

- Andersen, T. G. “Some Reflections on Analysis of High-Frequency Data.” *Journal of Business & Economic Statistics* 18, 2000: 146-153.
- Andersen, T. G., e T. Bollerslev. “Intraday Periodicity and Volatility Persistence in Financial Markets.” *Journal of Empirical Finance* 4, 1997: 115-158.
- Andersen, T. G., T. Bollerslev, e A. Das. “Variance-Ratio Statistics and High-Frequency Data: Testing for Changes in Intraday Volatility Patterns.” *The Journal of Finance* 56, 2001: 305-327.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, e P. Labys. “Exchange rate returns standardized by realised volatility are (nearly) Gaussian.” *NBER Working Paper*, 7488, 2000.

Baillie, R. T., e C. Morana. "Modeling Long Memory and Structural Breaks in Conditional Variances: an Adaptive FIGARCH Approach." *Working Papers 593, Queen Mary University of London, Department of Economics*, 2007.

Baillie, R. T., T. Bollerslev, e H. O. Mikkelsen. "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 74, 1996: 3-30.

Bauwens, L., e S. Laurent. "A new class of multivariate skew densities, with application to GARCH models." *Journal of Business and Economic Statistics* 23, 2005: 346-354.

Beltratti, A., e C. Morana. "Computing Value at Risk with High Frequency Data." *Journal of Empirical Finance* 53, 1999: 431-455.

Black, F. "Studies of stock price volatility changes." *Proceedings from the American Statistical Association, Business and Economic Statistics Section*, 1976: 177-181.

Bollerslev, T. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31, 1986: 307-327.

Bollerslev, T., e H. O. Mikkelsen. "Modeling and Pricing Long-Memory in Stock Market Volatility." *Journal of Econometrics* 73, 1996: 151-184.

Bollerslev, T., R. Chou, e K. Kroner. "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence." *Journal of Econometrics* 52, 1992: 5-59.

Brooks, C., S. P. Burke, e G. Persaud. "Autoregressive Conditional Kurtosis." *Journal of Financial Econometrics* 3, 2005: 399-421.

Bussab, W., e P. A. Morettin. *Estatística Básica, 5a Edição*. São Paulo: Editora Saraiva, 2002.

Campbell, J. Y., A. W. Lo, e A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton: Princeton University Press, 1997.

Chan, K., K. C. Chan, e G. A. Karolyi. "Intraday Volatility in the Stock Index and Stock Index Futures Market." *Review of Financial Studies* 4, 1991: 657-684.

Cont, R. "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues." *Quantitative Finance* 1, 2001: 223-236.

Coroneo, L., e D. Veredas. "Intradaily seasonality of returns distribution: A Quantile Regression approach and Intradaily VaR." *Revised Version of CORE Discussion Paper No. 2006/77*, 2006: 1-24.

Dacorogna, M. M., R. Gencay, U. Müller, R. B. Olsen, e V. P. Pictet. *An Introduction do High Frequency Finance*. London: Academic Press, 2001.

Davidson, J. "Moment and memory properties of linear conditional heteroscedasticity models, and a new model." *Journal of Business and Economic Statistics* 22, 2004: 16-29.

Ding, Z., C. W. J. Granger, e R. F. Engle. "A Long Memory Property of Stock Returns and a New Model." *Journal of Empirical Finance* 1, 1993: 83-106.

Enders, W. *Applied Econometric Time Series, Second Edition*. New York: Wiley, 2004.

Engle, R. F. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica* 50, 1982: 987-1008.

Engle, R.F., e T. Bollerslev. "Modelling the persistence of conditional variances." *Econometric Reviews* 5, 1986: 1-50.

Fama, E. F. "The Behaviour of Stock Market Prices." *Journal of Business* 38, 1965: 34-105.

Geweke, J., e S. Porter-Hudak. "The estimation and application of long memory time series models." *Journal of Time Series Analysis* 4, 1983: 221-238.

Giot, P. "Time Transformations, Intraday Data and Volatility Models." *Journal of Computational Finance* 4, 2000: 31-62.

Glosten, L. R., R. Jagannathan, e D. E. Runkle. "On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks." *Journal of Finance* 48, 1993: 1779-1801.

Glosten, L.R., e P.R. Milgrom. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders." *Journal of Financial Economics*, 1985: 71-100.

Goodhart, C. A. E., e M. O'Hara. "High Frequency Data in Financial Markets: Issues and Applications." *Journal of Empirical Finance* 4, 1997: 73-114.

- Gouriéroux, C., e J. Jasiak. *Financial Econometrics: Problems, Models, and Methods*. Princeton: Princeton University Press, 2001.
- Granger, C.W.J., e Z. Ding. “Modeling Volatility Persistence of Speculative Returns.” *Journal of Econometrics*, 73, 1996: 185-215.
- Guillaume, D. M., M. M. Dacorogna, e O. V. Pictet. “On the Intradaily Performance of GARCH Processes.” *Proceedings of the First International Conference on High Frequency Data in Finance, Zürich, Switzerland*, 1994.
- Hamilton, J. D. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994.
- Harris, L. “A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns.” *Journal of Financial Economics* 16, 1986: 99-117.
- Jondeau, E., S. H. Poon, e M. Rockinger. *Financial Modeling Under Non-Gaussian Distributions*. London: Springer, 2007.
- Laurini, M. P., L. G. C. Furlani, e M. S. Portugal. “Microestrutura Empírica de Mercado: Uma Análise Para a Taxa de Câmbio BRL/US\$ Usando Dados de Alta Frequência.” *IBMEC Working Paper* 55, 2008: 1-20.
- Mandelbrot, B. “The Variation of Some Speculative Prices.” *Journal of Business* 36, 1963: 394-419.
- Moreira, J. M. S., e E. J. Lemgruber. “O uso de dados de alta frequência na estimação da volatilidade e do valor em risco para o IBOVESPA.” *Revista Brasileira de Economia*, Jan./Mar. de 2004, vol. 58, no 1: 100-120.
- Morettin, P. A. *Econometria Financeira: Um Curso em Séries Temporais Financeiras*. São Paulo: Associação Brasileira de Estatística, 2006.
- Morettin, P. A., e C. M. C. Toloi. *Análise de Séries Temporais, Segunda Edição*. São Paulo: Editora E. Blücher-Associação Brasileira de Estatística, 2006.
- Nelson, D. B. “Conditional Heteroskedasticity in Asset Returns: a New Approach.” *Econometrica* 59, 1991: 349-370.
- O'Hara, M. *Market Microstructure Theory*. Cambridge, MA: Blackwell Publishers, 1995.

Palma, W. *Long-Memory Time Series*. Hoboken, NJ: John Wiley & Sons, 2007.

Roll, R. "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market." *The Journal of Finance* 39, 1984: 1127-1139.

Taylor, S. J. *Asset Price Dynamics, Volatility and Prediction*. Princeton, NJ: Princeton University Press, 2005.

Taylor, S. J., e X. Xu. "The Incremental Volatility Information in One Million Foreign Exchange Quotations." *Journal of Empirical Finance* 4, 1997: 317-340.

Tsay, R. S. *Analysis of Financial Time Series, Second Edition*. New York: Wiley, 2005.

Tse, Y. K. "The conditional heteroscedasticity of the yen-dollar exchange rate." *Journal of Applied Econometrics* 13, 1998: 49-55.

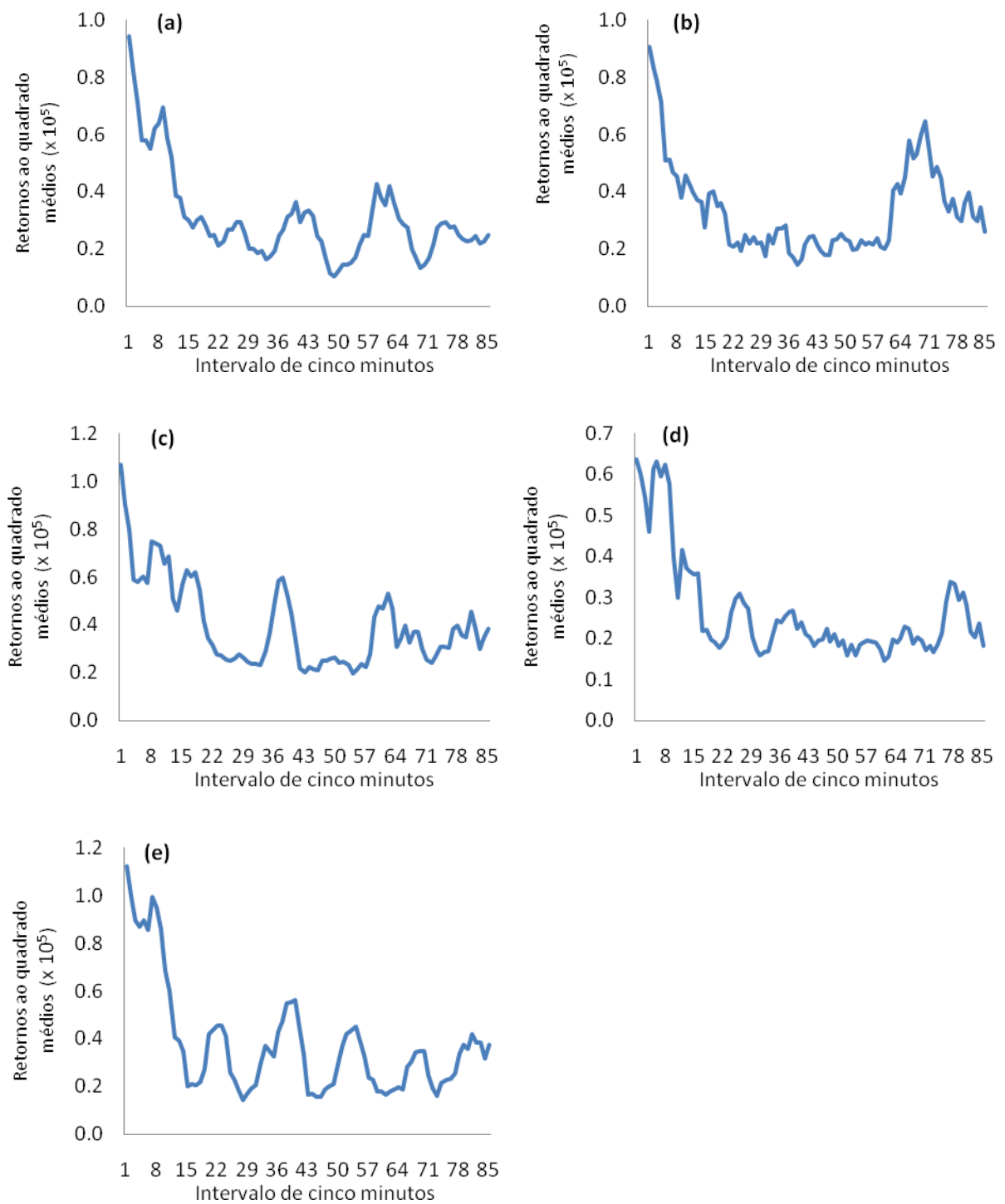
Wood, R. A., T. H. McInish, e J. K. Ord. "An Investigation of Transaction Data for NYSE Stocks." *Journal of Finance* 25, 1985: 723-739.

## APÊNDICE

### A.1 Construção dos fatores sazonais específicos para cada dia da semana

Para o cálculo dos fatores intradiários específicos de cada dia da semana, identificaram-se os 5 maiores retornos absolutos da série (3 deles em segundas-feiras, um em uma terça-feira e o último numa quinta-feira), que foram alterados para refletir o retorno médio entre os preços dos intervalos anterior e subsequente. Procedeu-se desta forma, pois estes retornos anormais (*outliers*) produziriam um viés considerável no filtro devido à reduzida amostra de intervalos para cada dia da semana, que se estenderia aos retornos dessazonalizados em todos os mesmos dias da semana naqueles intervalos. Não obstante, estes retornos suspeitos foram mantidos nas séries de retornos, pois aí constituem apenas 5 dentre mais de 5 mil retornos da amostra; nota-se ainda que conforme visto na descrição dos dados, a amplitude da amostra completa é menor que a encontrada em Andersen e Bollerslev (1997), e a amplitude, sem estes considerados *outliers*, passa a cerca de 2% (de -1,12% a 1,01%). Novamente, calculamos o fator de dessazonalização a partir dos retornos quadrados médios de cada intervalo. Para promover um alisamento adicional destes fatores, uma vez que nossa série consiste em uma amostra relativamente curta para se calcular fatores específicos de cada dia da semana (ver, no entanto, nota de rodapé 14), a janela móvel utilizada neste caso foi de 30 minutos, ou seja, o filtro empregado é a média aritmética do fator calculado para o próprio intervalo e dos dois intervalos anteriores e subsequentes. Os padrões da sazonalidade intradiária assumem a forma observada nas figuras a seguir:





**Figura 6 - Retornos intradiários ao quadrado médios suavizados: (a) 2<sup>a</sup> -feira, (b) 3<sup>a</sup> -feira, (c) 4<sup>a</sup> -feira, (d) 5<sup>a</sup> -feira, (e) 6<sup>a</sup> -feira**

## A.2 Resultados

Abaixo seguem os resultados quando corrigimos a periodicidade intradiária em cada dia da semana pelos fatores específicos para estes. Novamente, a série de retornos homogênea com 67 dias de 85 intervalos foi dividida pela raiz quadrada dos filtros de cada respectivo intervalo e dia da semana, e reportamos os resultados sob a suposição de normalidade condicional:

**Tabela 7 - Estatísticas e medidas de qualidade de ajuste para os modelos estimados com fatores intradiários de dessazonalização calculados para cada dia da semana**

	GARCH(1,1)	EGARCH(1,1)	GJR(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)	FIEGARCH(1,d,1)	FIAPARCH(1,d,1)	HYGARCH(1,d,1)
LL	-7870.188	-7846.378	-7868.126	-7903.381	-7849.944	<b>-7843.182</b>	-7844.305	-7849.915
AIC	2.765299	2.757639	2.764926	2.776604	2.758540	<b>2.756868</b>	2.757263	2.758881
BIC	2.769967	2.764643	2.770762	2.780106	<b>2.764376</b>	2.765039	2.765433	2.765885
Assimetria	-0.12398	-0.11457	-0.12703	-0.11972	-0.10143	-0.10097	-0.10922	-0.10191
Ex. Curtose	1.7804	1.7119	1.7818	2.0206	1.7355	1.7171	1.7143	1.7330
Q(10)	0.0040815	0.3723608	0.0036923	0.0011910	0.6334283	0.4397885	0.5329184	0.6388822
Q(50)	0.0475394	0.1831050	0.0468249	0.0010520	0.4417443	0.2644772	0.3587702	0.4449913
Q(425)	0.0556900	0.1369932	0.0490690	0.0139229	0.5495012	0.2723821	0.3829021	0.5470796
Q(850)	0.0433607	0.1100293	0.0422140	0.0152786	0.4334060	0.1844163	0.2804333	0.4313009

LL denota o log da verossimilhança; AIC e BIC são os critérios de informação de Akaike e Schwarz, respectivamente. Q(l) representa o p-valor da estatística de Box-Pierce para o quadrado dos resíduos padronizados, em que aceitamos a hipótese nula de não correlação serial das "l" primeiras defasagens a um nível de significância qualquer quando o p-valor é maior que este nível. Valores em destaque denotam o melhor ajuste.

**Tabela 8 - Estimativas para parâmetros dos modelos com fatores intradiários de dessazonalização calculados para cada dia da semana**

	GARCH(1,1)	EGARCH(1,1)	GJR(1,1)	IGARCH(1,1)	FIGARCH(1,d,1)	FIEGARCH(1,d,1)	FIAPARCH(1,d,1)	HYGARCH(1,d,1)
$\varphi$	-0.095584 (0.014768) 0.0000	-0.095809 (0.016535) 0.0000	-0.097047 (0.014735) 0.0000	-0.098569 (0.013798) 0.0000	-0.093369 (0.014678) 0.0000	-0.095917 (0.021700) 0.0000	-0.096107 (0.014959) 0.0000	-0.093282 (0.014703) 0.0000
$\omega$	0.044743 (0.017433) 0.0103	0.172994 (0.087274) 0.0475	0.043462 (0.016095) 0.0069	0.008238* (0.0066818) 0.2176	0.069701 (0.025094) 0.0055	0.347111 (0.16256) 0.0328	0.088131 (0.022304) 0.0001	0.073751 (0.035873) 0.0398
$\alpha$	0.071037 (0.015631) 0.0000	-0.570119 (0.10249) 0.0000	0.059586 (0.016127) 0.0002	0.065082 (0.031659) 0.0399	0.507359 (0.085605) 0.0000	-0.715857 (0.10728) 0.0000	0.448523 (0.064790) 0.0000	0.497412 (0.10316) 0.0000
$\beta$	0.884916 (0.030922) 0.0000	0.971908 (0.010112) 0.0000	0.887440 (0.028918) 0.0000	0.934918	0.633685 (0.083143) 0.0000	0.919520 (0.036447) 0.0000	0.684518 (0.061608) 0.0000	0.636726 (0.079109) 0.0000
$\theta_1$	-0.029300* (0.015766) 0.0632					-0.023704* (0.015807) 0.1338		
$\theta_2$	0.262144 (0.042300) 0.0000					0.260817 (0.041542) 0.0000		
$d$					0.256035 (0.046081) 0.0000	0.278637 (0.11041) 0.0116	0.376246 (0.075375) 0.0000	0.279051 (0.13763) 0.0426
$\gamma$	0.019702* (0.012445) 0.1135						0.089306* (0.070197) 0.2033	
$\delta$							1.347831 (0.26948) 0.0000	
$\ln(\psi)$								-0.033614* (0.17790) 0.8501

Para cada parâmetro, a tabela reporta a estimativa de Quasi Máxima Verossimilhança, seguida pelo erro padrão robusto (entre parênteses) e o p-valor. O sinal "\*" denota parâmetro não significativo a 5% de significância.