



Revista Brasileira de Finanças

ISSN: 1679-0731

rbfin@fgv.br

Sociedade Brasileira de Finanças

Brasil

Ceretta, Paulo Sérgio; Galvão de Barba, Fernanda; Mendes Vieira, Kelmara; Casarin, Fernando
Previsão da Volatilidade Intradiaria: Análise das Distribuições Alternativas
Revista Brasileira de Finanças, vol. 9, núm. 2, 2011, pp. 209-226

Sociedade Brasileira de Finanças

Rio de Janeiro, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=305824887003>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

Previsão da Volatilidade Intradiária: Análise das Distribuições Alternativas

(Intraday Volatility Forecasting: Analysis of Alternative Distributions)

Paulo Sérgio Ceretta*

Fernanda Galvão de Barba**

Kelmara Mendes Vieira***

Fernando Casarin****

Resumo

A previsão da volatilidade tem sido de grande interesse tanto no âmbito acadêmico quanto profissional em todo o mundo. Não há, entretanto, um consenso sobre o melhor modelo a ser aplicado no processo de estimar a volatilidade. Novas modelagens têm incluído medidas de assimetria, mudanças de regime e diferentes distribuições. Destes fatos estilizados, poucos estudos têm sido feitos considerando as diferentes distribuições. Este trabalho visa investigar como a especificação da distribuição influencia a performance da previsão da volatilidade em dados intradiários do Ibovespa, usando o modelo APARCH. As previsões foram realizadas supondo seis distribuições distintas: normal, normal assimétrica, *t-student*, *t-student* assimétrica, generalizada e generalizada assimétrica. Os resultados obtidos evidenciam que o modelo considerando a distribuição *t-student* assimétrica foi o que melhor se ajustou aos dados dentro da amostra, porém, na previsão fora da amostra, o modelo com distribuição normal apresentou melhor desempenho.

Palavras-chave: volatilidade; modelos de previsão; diferentes distribuições.

Códigos JEL: C01; C53; G17.

Abstract

Volatility forecasting has been of great interest both in academic and professional fields all over the world. However, there is no agreement about the best model to estimate volatility. New models include measures of skewness, changes of regimes and different distributions; few studies, though, have considered different distributions. This paper aims to investigate how the specification of a distribution influences the performance of volatility forecasting on Ibovespa intraday data, using the APARCH model. The forecasts were carried out assuming six distinct distributions: normal, skewed normal, *t-student*, skewed *t-student*, generalized and skewed generalized. The results evidence that the model considering the skewed *t-student* distribution offered the best fit to the data inside the sample, on the other hand,

Submetido em fevereiro de 2010. Aceito em dezembro de 2010.

*Universidade Federal de Santa Maria, Santa Maria, RS, Brazil. E-mail: ceretta@smail.ufsm.br

**Universidade Federal de Santa Maria, Santa Maria, RS, Brazil. E-mail: fernandadebarba@yahoo.com.br

***Universidade Federal de Santa Maria, Santa Maria, RS, Brazil. E-mail: kelmara@smail.ufsm.br

****Universidade Federal de Santa Maria, Santa Maria, RS, Brazil. E-mail: ferinvest@gmail.com

the model assuming a normal distribution provided a better out-of-the-sample performance forecast.

Keywords: volatility; forecasting models; different distributions.

1. Introdução

A volatilidade dos mercados financeiros tem sido um assunto de grande interesse no meio acadêmico em face à incessante busca de antecipar seu comportamento. A sua previsão mostra-se importante na elaboração de estratégias de investimento, análise de risco e apreçamento de ativos. Não há, entretanto, consenso sobre o melhor modelo a ser aplicado para estimar a volatilidade (Otuki *et al.*, 2008).

As modelagens utilizadas em dados financeiros diferem em parte das modelagens tradicionais de séries temporais. Esses dados apresentam peculiaridades que requerem adaptações e estruturações específicas para seu comportamento. Os principais fatos estilizados relativos aos retornos financeiros são: a) os retornos são em geral não-auto-correlacionados; b) os quadrados dos retornos são auto-correlacionados, apresentando uma correlação de pequenas defasagens (1 ou 2 defasagens) e depois uma queda lenta das demais; c) séries de retornos apresentam agrupamentos de volatilidades ao longo do tempo; d) a distribuição (incondicional) dos retornos apresenta caudas mais pesadas do que uma distribuição normal; além disso, a distribuição, embora aproximadamente simétrica, é em geral leptocúrtica; e) algumas séries de retornos são não-lineares, no sentido de que são impactadas de modo diferente por choques negativos e positivos, ou choques grandes e pequenos, (Morettin, 2008) fato este muito encontrado na volatilidade.

Os modelos tradicionais de séries temporais geralmente assumem uma distribuição normal. Entretanto, a literatura financeira frequentemente enfatiza que as distribuições não são normais apresentando distribuições leptocúrticas e de caudas grossas. Recentes trabalhos sobre a previsão da volatilidade têm testado a performance de modelos ARCH/GARCH assimétricos obtendo bons resultados (Brooks *et al.*, 2000, Awartani & Corradi, 2005, Bali, 2007, Otuki *et al.*, 2008). Poucos desses trabalhos, entretanto, buscaram verificar o efeito do tipo de distribuições dos erros nos modelos usados nesta previsão.

Este artigo visa investigar como a especificação da distribuição influencia a performance da previsão da volatilidade em dados intradiários do Ibovespa, usando o modelo APARCH. A previsão é realizada supondo seis distribuições distintas: normal, normal assimétrica, *t-student*, *t-student* assimétrica, generalizada e generalizada assimétrica.

O artigo está estruturado da seguinte forma: após esta breve introdução, a sessão seguinte aborda as modelagens utilizadas para estimar a volatilidade, a sessão 3 conceitua e caracteriza as distribuições objeto de estudo, a sessão 4 trata dos trabalhos empíricos realizados anteriormente sobre o assunto. Na sessão 5 são apresentados os aspectos metodológicos e os dados. Os resultados são apresentados e analisados na sessão 6 e a sessão 7 sumariza o estudo e relata as conclusões.

2. Modelagens da Volatilidade

Desde 1952, quando Markowitz utilizou a volatilidade dos retornos das ações como medida de risco, formas de modelá-la têm sido buscadas por estudiosos de Finanças. Os modelos das famílias ARCH (*Autoregressive Conditional Heteroskedastic*), inicialmente propostos por Engle (1982), apresentam um grupo de características que os torna atrativos para aplicações econômicas. Dentre estas características, o modelo ARCH considera que a variância condicional pode mudar ao longo do tempo e é prevista pelos erros ao quadrado passados, capturando um importante fato estilizado: os agrupamentos de volatilidade.

Em 1986, Bollerslev apresentou uma extensão do modelo ARCH, o GARCH (Generalized ARCH), que permite uma estrutura de defasagem mais flexível. A modelagem ARCH geralmente necessita de uma defasagem relativamente longa na equação da variância condicional e, para evitar problemas com parâmetros negativos de variância, uma estrutura de defasagem fixa é tipicamente imposta. Estes problemas levam ao interesse prático na extensão dos modelos ARCH para permitir tanto uma memória mais longa, quanto uma estrutura de defasagem mais flexível. A modelagem GARCH oferece um ajuste levemente melhor do que o modelo ARCH, assim como uma estrutura de defasagem mais razoável.

Em 1987, Bollerslev propôs uma extensão do modelo ARCH que permite uma distribuição condicional *t-student* dos erros. Neste trabalho, o autor chegou à conclusão de que a distribuição *t-student* falha em considerar a dependência temporal e os modelos ARCH e GARCH com distribuição condicional dos erros não parecem capturar completamente a leptocurtose.

Atualmente, variações dos modelos ARCH/GARCH, que presumem as especificidades dos dados financeiros, têm sido testadas. Estas variações levam em conta a assimetria, diferentes distribuições, mudança de regimes, entre outros. Dentro destas especificidades, a assimetria tem sido a mais enfatizada. Os modelos ARCH/GARCH assimétricos, como o EGARCH de Nelson (1991), o TARCH, desenvolvido por Zakoian (1994), e o GJR-GARCH de Glosten *et al.* (1993) levam em consideração que choques positivos e negativos impactam de modo distinto na volatilidade. As diferentes distribuições do erro têm sido pouco abordadas nos estudos empíricos.

Estudos sobre a volatilidade de ativos financeiros em mercados de diversos países têm indicado que as variações dos modelos ARCH/GARCH apresentam bom desempenho na previsibilidade do mercado. Os estudos voltam-se principalmente ao comportamento de mercados maduros. Nos últimos anos, entretanto, a proeminência econômica dos países emergentes (principalmente as quatro maiores economias desse grupo: Brasil, Rússia, Índia e China) tem gradativamente atraído a atenção de estudiosos. O comportamento da volatilidade nos mercados emergentes tem revelado um comportamento distinto dos mercados maduros, principalmente por apresentarem risco mais elevado.

3. Distribuições dos Erros

Balakrishnan & Nevzorov (2003) destacam que a função densidade da probabilidade (fdp) normal é a distribuição mais comumente utilizada como parâmetro de comparação em procedimentos estatísticos (análise de regressão, séries temporais, experimentações, etc). A forma padronizada dessa distribuição é dada por (1).

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] \quad (1)$$

A função densidade é chamada de padronizada quando x for uma variável aleatória normal com média zero e desvio padrão unitário, $N(\mu = 0, \sigma = 1)$, simétrica e coeficiente de curtose igual a 3. A forma de função densidade empírica da normal padronizada é ilustrada na parte superior da Figura 1.

Na parte superior da Figura 1 são apresentadas três funções de densidade empírica normais. A linha cheia (—) representa a normal simétrica e mesocúrtica, a linha tracejada (....) representa a normal assimétrica positiva mesocúrtica e a linha pontilhada (···) representa a normal assimétrica negativa mesocúrtica.

Diretamente associadas a procedimentos de séries temporais, algumas distribuições alternativas têm sido implementadas em alguns softwares (Stata, PcGive, Eviews, Splus, R, etc). A idéia básica das distribuições alternativas é permitir que o termo de erro possa ter um componente distinto do previsto por uma distribuição normal.

Autores como Hsieh (1989) e Baillie & Bollerslev (1989) têm enfatizado que a distribuição *t-student* pode capturar melhor características dos log-retornos em séries temporais associados à curtose. Krishnamoorthy (2006) examina em detalhes a distribuição *t-student*. A função densidade da probabilidade da função *t-student* padronizada é definida por (2).

$$f(x|v) = \frac{1}{\sqrt{v-2}B(\frac{1}{2}, \frac{v}{2})} \frac{1}{(1 + \frac{x^2}{v-2})^{\frac{v+1}{2}}} \quad (2)$$

onde v = graus de liberdade e $B(ab) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ sendo Γ a função gama. Krishnamoorthy (2006) demonstrou que para grandes valores de v a distribuição *t*-

student tende a uma normal. O gráfico na parte central da Figura 1 ilustra a função densidade de uma distribuição *t-student* em três situações diferentes. A linha cheia no gráfico central da Figura 1 ilustra a função densidade *t-student* com $v = 30$. É possível verificar visualmente grande semelhança da função densidade *t-student* ($v = 30$) com uma densidade normal (gráfico superior da Figura 1). À medida que v diminui a distribuição *t-student* torna-se mais leptocúrtica.

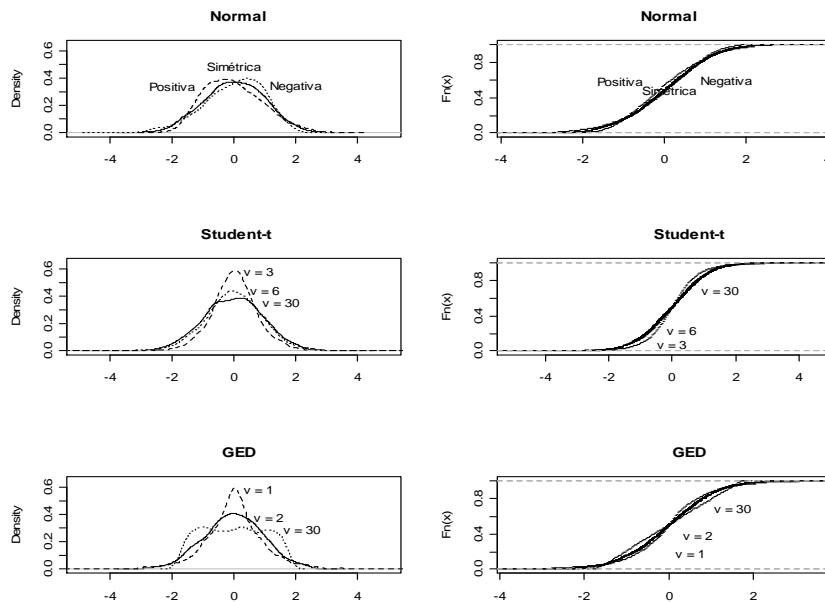


Figura 1
Função densidade da probabilidade e função densidade cumulativa empírica

A distribuição generalizada do erro (GED) é outra possibilidade para um melhor ajuste dos procedimentos estatísticos aos dados. Conforme destaca Theodosiou (2000), essa distribuição foi introduzida por Subbotin em 1923 e inclui como casos especiais as distribuições de Laplace, normal e uniforme, sendo inicialmente utilizada por Box & Tiao (1973), e posteriormente por Nelson (1991). A escolha da densidade GED é dada pela inabilidade dos processos GARCH Gaussianos em levar em conta a leptocurtose da maioria das séries de retorno, uma questão que é ainda mais relevante quando são usados dados de mercados emergentes (Lee *et al.*, 2001). A função densidade da probabilidade GED é dada por (3) onde v são os graus de liberdade e Γ é a função gama. Essa função densidade é ilustrada no gráfico inferior da Figura 1.

$$f(x|v) = \sqrt{\left(\frac{2^{-2/v}\Gamma(\frac{1}{v})}{\Gamma(\frac{3}{v})}\right)} \quad (3)$$

Na parte inferior da Figura 1 tem-se o plot de três possibilidades da fdp de uma distribuição GED. Assumindo que $v = 1$, a distribuição GED tem a propriedade de capturar a leptocurtose reduzindo-se a uma função Laplace. Quando $v = 2$, a GED assemelha-se a uma distribuição Normal. À medida que v aumenta, a GED tende a uma distribuição Uniforme. Para uma descrição mais detalhada das distribuições Laplace, Normal, *t-student* e Uniforme, ver Krishnamoorthy (2006).

Uma visualização alternativa sobre a diferença das distribuições Normal (simétrica e assimétrica), *t-student* e GED para diferentes graus de liberdade é apresentada no lado esquerdo da Figura 1. A função densidade cumulativa (fdc) descreve completamente a distribuição da probabilidade de uma variável aleatória de valor X . Para cada número real x a fdc é dada por $F(x) = P(X \leq x)$. É convenção usar um F maiúsculo para fdc, em contraste com o f minúsculo usado para a função densidade da probabilidade (fdp). A mudança na assimetria (no gráfico da densidade Normal) e as variações nos graus de liberdade (nos gráficos *t-student* e GED) determinam o grau de inclinação da curva ascendente das funções densidade cumulativas. É possível verificar também que os gráficos das densidades *t-student* e GED abrangem valores mais negativos que a densidade Normal.

4. Estudos Empíricos Anteriores

Um número limitado de estudos busca verificar empiricamente se pressuposições de diferentes distribuições (com e sem assimetria) melhoram a previsão da volatilidade fora da amostra. Os trabalhos seminais de Mandelbrot (1963) e Fama (1965) evidenciaram as características dos retornos financeiros exibirem excesso de curtose positiva e caudas grossas. A partir disto, distribuições distintas da Normal têm sido propostas para levar em consideração o excesso de curtose. Nelson (1991), Taylor (1994), Lopez (2001), Lee *et al.* (2001), Marcucci (2005), por exemplo, propuseram o uso da distribuição generalizada do erro. Do mesmo modo, Politis (2004) propôs a distribuição de caudas grossas para aliviar o problema.

Em 2007, Bali realizou um estudo propondo modelos paramétricos generalizados para a taxa de juros de curto prazo nos Estados Unidos. O autor utilizou diversos modelos da família GARCH pressupondo as distribuições normal, normal assimétrica, generalizada, generalizada assimétrica, *t-student* e *t-student* assimétrica para tentar prever a taxa de juros. Os resultados obtidos nesse estudo induziram à conclusão de que os modelos temporais discretos que incorporaram o nível e os efeitos GARCH acomodando as caudas grossas têm melhor performance na previsão da volatilidade condicional da taxa de juros.

Mais recentemente, buscando analisar como a especificação da distribuição influencia a performance da previsão da volatilidade, Liu *et al.* (2009) utilizaram dois modelos GARCH (GARCH-N, que assume a distribuição normal, e GARCH-SGED, com distribuição generalizada assimétrica) em dados diários das bolsas de Shanghai e Shenzhen durante o período de 4 de janeiro de 2000 a 29 de dezembro de 2006. Os resultados empíricos indicam que o modelo GARCH-SGED é superior ao modelo GARCH-N para prever a volatilidade dos mercados financeiros da China. Este resultado aponta que tanto a assimetria quanto a característica de caudas grossas da distribuição condicional dos retornos são significativas, especialmente para mercados financeiros emergentes.

No Brasil, os estudos de volatilidade direcionaram-se à comparação entre modelos para verificar o de melhor ajuste ou à análise do papel da assimetria na modelagem das séries financeiras. Carvalho *et al.* (2006) investigaram se os fatos estilizados verificados no mercado americano também ocorriam no mercado brasileiro, usando modelos GARCH e EWMA (*exponentially weighted moving average*). Para esse estudo, os autores analisaram as cinco ações com maior liquidez na Bolsa de Valores de São Paulo com dados intradiários. Os resultados evidenciaram que os log-retornos de dados intradiários apresentam uma distribuição muito próxima da Gaussiana. Em segundo lugar, não foi encontrada evidência de memória longa no log da variância realizada (o contrário do mercado americano) e, por final, em média, o modelo EWMA apresentou previsões menos precisas que as efetuadas pelo modelo linear.

Otuki *et al.* (2008) buscaram evidência do efeito assimétrico na volatilidade das séries de retornos dos índices de ações na Argentina, Brasil e México durante o período de janeiro de 2000 a dezembro de 2005. Neste estudo, foram utilizados os modelos GARCH, EGARCH e TARCH, sendo que os melhores ajustes foram o modelo EGARCH (1, 1) para o índice Imex e o modelo TARCH(1, 1) para os índices Ibovespa e Merval. Os resultados mostraram que os eventos negativos exercem maior influência do que os positivos. O mesmo resultado foi confirmado pelo estudo de Jubert *et al.* (2008).

5. Metodologia e Dados

A base de dados utilizada foi composta por observações intradiárias de 15 em 15 minutos do Ibovespa durante o período de 15 de junho de 2009 a 27 de outubro de 2009. Cada dia é composto por 28 observações intradiárias no horário de 10:00 às 16:45. Quando não constava observação em um determinado intervalo, foi utilizada a observação do minuto mais próximo. A amostra obtida no período de 94 dias totalizou 2632 observações. Deste total, foram utilizadas as 2492 observações iniciais para ajuste dos modelos dentro da amostra e as 140 observações posteriores serviram para comparar a previsão dos diferentes modelos fora da amostra.

Da base de dados inicial foram calculados os log-retornos pela expressão $r_{t,j} = \ln P_{t,j} - \ln P_{t,j-1}$, onde $r_{t,j}$ é o log-retorno do j -ésimo intervalo de quinze minutos do dia t e $P_{t,j}$ é o preço de fechamento do índice no j -ésimo intervalo de quinze minutos do dia t , onde $t = 1, 2, 3, \dots, 94$ e $j = 1, 2, 3, \dots, 28$. O primeiro retorno de cada dia foi calculado considerando o minuto $j - 1$ como o último minuto do dia anterior ($t - 1, 28$). A título de exemplo, considere o cálculo do primeiro retorno do dia $t = 24$, a expressão seria $r_{24,1} = \ln P_{24,1} - \ln P_{23,28}$. Os finais de semana e feriados foram desconsiderados para o cálculo dos retornos.

Os estudos empíricos a respeito da previsão da volatilidade utilizam-se principalmente de dados diários, semanais ou mensais. No Brasil, a utilização de dados de alta frequência (intradiários) ainda é relativamente baixa em função da dificuldade de obtenção dos mesmos. Andersen & Bollerslev (1998, 1999) salientam que a incorporação da informação contida em dados de alta freqüência melhora significativamente as projeções da volatilidade dos retornos diários, tanto em termos teóricos como práticos. Giot (2000) considera ótimo o intervalo de 15 minutos, enquanto Andersen (2000) enfatiza que a frequência de 5 minutos seria a mais alta para a qual as propriedades da série não seriam distorcidas por aspectos da microestrutura de mercado.

Andersen & Bollerslev (1997) enfatizam que os padrões sistemáticos observados em séries financeiras de alta freqüência, nos mercados de ações e de câmbio, são ignorados em boa parte da literatura empírica sobre microestrutura de mercado. Isso explicaria a aparente inadequação dos modelos tradicionais de volatilidade quando aplicados a séries de alta freqüência. A estimativa e filtragem do componente periódico intradiário da volatilidade permite eliminar a maior parte das distorções atribuídas à sazonalidade. Seguindo a orientação de Moreira & Lemgruber (2004), foi realizada a dessazonalização dos dados por dia da semana.

Para a série dessazonalizada foi verificada a estacionaridade a partir dos testes ADF (Dickey-Fuller Aumentado), PP (Phillips-Perron) e KPSS (Kwiatkowski-Phillips-Schmidt-Shin). O teste de Ljung-Box foi utilizado para checar a existência de autocorrelação na série de ajuste do modelo.

Em função da volatilidade não ser uma variável diretamente observável, existem três enfoques principais para o seu cálculo: (i) equacionando um preço de mercado observado com o preço modelado de uma opção – *volatilidade implícita*; (ii) *volatilidade estatística*, que é modelada diretamente a partir da série de retornos usando modelos da família ARCH e (iii) *volatilidade histórica*, obtida por meio de uma média de uma função dos últimos k retornos (Morettin, 2008). Neste estudo foi utilizada a volatilidade estatística modelada a partir da série de retornos do Ibovespa.

Foram realizadas estimativas do retorno ao quadrado (*proxy* da volatilidade) utilizando o modelo APARCH para cada uma das distribuições (normal, normal assimétrica, *t-student*, *t-student* assimétrica, generalizada, generalizada assimétrica). A ordem do modelo foi definida a partir dos gráficos ACF (função de autocorrelação) e PACF (função de autocorrelação parcial). O modelo APARCH(p, q), proposto por Ding *et al.* (1993), pode ser definido da seguinte forma:

$$y_t = x'_{1,t}\mu + \epsilon_t \quad (4)$$

$$\epsilon_t = \sigma_t z_t \quad (5)$$

$$\sigma_t^\delta = x'_{2,t}\omega + \sum_{i=1}^q \alpha_i k(\epsilon_{t-i})^\delta + \sum_{j=1}^p \beta_j \alpha_{t-j}^\delta \quad (6)$$

$$k(\epsilon_{t-i}) = |\epsilon_{t-i}| \gamma_i \epsilon_{t-i} \quad (7)$$

onde y_t é o retorno, $x_{1,t}$ e $x_{2,t}$ são dois vetores de, respectivamente, η_1 e η_2 variáveis fracamente exógenas (incluindo o intercepto), μ , ω , α_i 's, γ_i 's, β_j 's e σ são parâmetros (ou vetores de parâmetros) a serem estimados. $\sigma(\sigma > 0)$ assume o papel de uma transformação Box-Cox do desvio padrão condicional γ_t , enquanto os γ_i 's refletem o assim chamado efeito alavancagem. Um valor positivo (negativo) de γ_i 's significa que choques negativos (positivos) passados têm um impacto mais forte na volatilidade condicional corrente do que choques positivos passados.

Charles (2010) afirma que o modelo APARCH foi considerado particularmente relevante em muitas aplicações recentes. Segundo Laurent (2004) esse modelo é um dos mais promissores da família ARCH. Esta afirmação fundamenta-se na noção de que esta modelagem abrange ao menos sete modelos ARCH como casos especiais:

- ARCH, quando $\sigma = 2$, $\gamma_i = 0(i = 1, \dots, p)$ e $\beta_j = 0(j = 1, \dots, q)$
- GARCH, quando $\sigma = 2$ e $\gamma_i = 0(i = 1, \dots, p)$
- GARCH de Taylor (1986) e Schwert (1990), quando $\sigma = 1$ e $\gamma_i = 0(i = 1, \dots, p)$
- GJR-GARCH, introduzido por Glosten *et al.* (1993), quando $\sigma = 2$
- TARCH de Zakoian (1994), quando $\sigma = 1$
- ARCH não-linear de Higgins & Bera (1992), quando $\gamma_i = 0(i = 1, \dots, p)$ e $\beta_j = 0(j = 1, \dots, q)$
- Log-ARCH de Geweke (1986) e Pantula (1986), quando $\sigma \rightarrow 0$.

O ajuste do modelo foi confirmado através da análise dos resíduos e dos critérios informacionais. Os resíduos foram testados para a normalidade e para a presença de dependência serial. Os critérios informacionais utilizados foram os de Akaike (AIC), *Log Likelihood* e *Log Likelihood* normalizado.

A partir das estimativas foram realizadas previsões um passo a frente para 140 períodos fora da amostra. Em cada passo da previsão, o modelo foi reestimado incorporando a observação anterior. As previsões geradas fora da amostra foram comparadas com o retorno ao quadrado (*proxy* da volatilidade) e o melhor modelo para previsão foi escolhido através dos indicadores raiz do erro absoluto médio percentual, raiz do erro quadrático médio percentual e desigualdade de U-Theil. Em todos os três testes, o melhor modelo é o que apresenta o valor mais próximo de zero. Quanto mais próximo de zero, menor a margem de erro da previsão. Para a desigualdade de U-Theil, quanto mais próximo de um, significa que a série estimada pelo modelo tem pouco ajustamento à série original.

6. Análise dos Resultados

A Tabela 1 resume as características estatísticas básicas da série de retornos do Ibovespa dividida em dois grupos: dentro e fora da amostra. Os valores dentro da amostra apresentam valores mínimos e máximos de -0.010565 e 0.010740, assimetria negativa e excesso de curtose positiva. Os valores de mínimo e máximo dos dados fora da amostra foram menores (-0.009392 e 0.007819), uma assimetria negativa mais elevada e leve excesso de curtose positiva.

Tabela 1
Estatísticas descritivas

Estatísticas	Retorno intradiário	
	Amostra	Fora-da-amostra
N de Observações	2492	140
Mínimo	-0.010565	-0.009392
Máximo	0.010740	0.007819
Média Aritmética	0.000103	-0.000324
Desvio Padrão	0.002278	0.003064
Assimetria	-0.150292	-0.472786
Excesso de Curtose	1.844.186	1.113.128

Inicialmente foi realizada a dessazonalização dos dados por dia da semana (Moreira & Lemgruber, 2004). Os dados apresentaram alta sazonalidade no início do dia, diferenciando o grau de decaimento em cada dia da semana. Em seguida foram realizados os testes ADF, PP e KPSS para verificar a estacionariedade da série de retornos dessazonalizada, confirmando que a série é $I(0)$.

Os gráficos ACF e PACF determinam, respectivamente, as ordens p e q do modelo autoregressivo a ser aplicado na série. Processos mistos tipicamente apresentam decaimento exponencial tanto no gráfico ACF quanto no PACF. Quando o decaimento exponencial dos gráficos é amortecido, o modelo mais apropriado é o de ordem (1, 1). Estes gráficos da série de retornos estudada não demonstraram

dependência serial (não apresentaram decaimento exponencial), portanto, não foi necessário realizar a filtragem inicial dos dados (Figura 2). Neste contexto foram utilizados os retornos ao quadrado como *proxy* de volatilidade. O teste de Ljung-Box com 30 defasagens (*p*-valor de 0,9387) confirma que não há autocorrelação na série de retornos dentro da amostra.

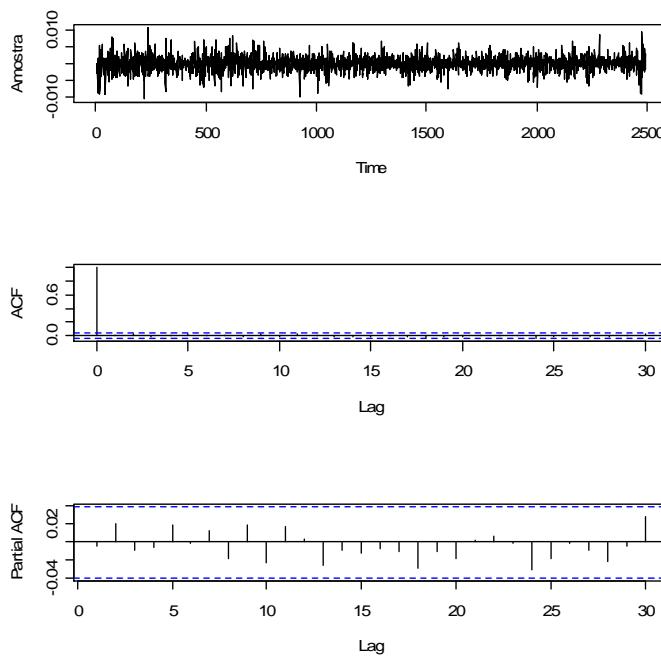


Figura 2
Retornos, função autocorrelação e autocorrelação parcial dos dados dentro da amostra do Ibovespa

A Figura 3 apresenta o gráfico contra o tempo dos retornos ao quadrado que é uma proxy para a variância condicional, assim como o ACF e o PACF dos mesmos. Os gráficos ACF e PACF identificam dependência serial, sendo adequada a aplicação de uma modelagem autoregressiva heterocedástica condicional de ordem (1, 1). O diagnóstico visual foi complementado com o teste Ljung-Box. O $p - \text{valor} < 2.2e - 16$ comprova a existência de autocorrelação na série.

Para realizar o ajuste do modelo, inicialmente, foram realizadas seis estimativas utilizando o modelo APARCH (1, 1). Cada uma das estimativas considerou uma das diferentes distribuições: normal, normal assimétrica, *t-student*, *t-student* assimétrica, generalizada e generalizada assimétrica. A comparação entre o ajuste das estimativas com diferentes distribuições foi realizada utilizando os critérios *Log Likelihood*, *Log Likelihood* Normalizada e AIC (Tabela 2).

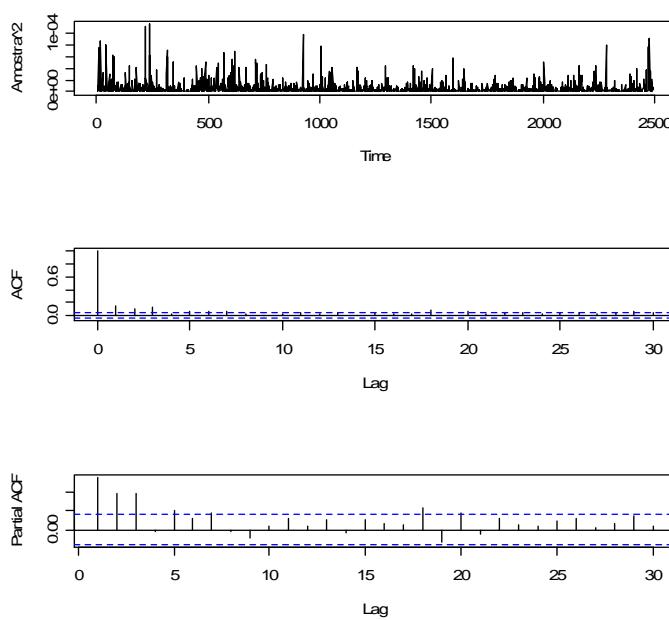


Figura 3
Função autocorrelação e autocorrelação parcial da volatilidade

Tabela 2
Critérios de seleção entre as distribuições dos modelos APARCH (1,1)

Distribuição*	Log likelihood	Log likelihood normalizada	AIC
Normal	11697.24	4.693.915	-9.383.014
Normal assimétrica	11698.33	4.694.353	-9.383.088
<i>t-student</i>	11754.97	4.717.082	-9.428.547
<i>t-student</i> assimétrica	11756.67	4.717.763	-9.429.106
GED	11741.21	4.711.563	-9.417.508
GED assimétrica	11743.49	4.712.475	-9.418.529

AIC: Akaike Informational Criteria.

*Software R.

A Tabela 2 apresenta os valores obtidos nos critérios informacionais dos modelos gerados para cada uma das distribuições. A distribuição *t-student* assimétrica obteve o melhor grau de ajuste de acordo com todos os critérios utilizados (*Log Likelihood*, *Log Likelihood* Normalizado e AIC). A distribuição normal foi a que apresentou o pior ajuste, considerando todos os critérios adotados. Os coeficientes obtidos pelo modelo APARCH (1,1) com distribuição *t-student* assimétrica são apresentados na Tabela 3 e foram obtidos através do *software R*.

Tabela 3Resultados do modelo APARCH (1,1) com distribuição *t-student* assimétrica – dentro da amostra

Coeficiente	Estimativa	Erro Padrão	T	p-valor(> t)
Omega (Ω)	0.0000	0.0000	24.820	0.0131
Alpha (α)	0.1079	0.0286	37.770	0.0002
Gamma (γ)	0.2044	0.0901	22.690	0.0233
Beta (β)	0.8078	0.0585	138.030	< 2e-16
Delta (δ)	15.610	0.3667	42.560	0.0000
Skew	0.9510	0.0280	339.460	< 2e-16
Shape	65.740	0.8128	80.880	0.0000

Na Tabela 3 estão relacionados os coeficientes estimados utilizando a distribuição *t-student* assimétrica, que através dos critérios informacionais é a que melhor se ajusta à série. Tanto a assimetria quanto o tipo de distribuição são significativos, com confiança de 99%. O valor positivo de 0,2044 obtido para o coeficiente gama (γ) significa que choques negativos passados têm um impacto mais forte na volatilidade condicional corrente do que choques positivos passados.

Tabela 4Diagnóstico do modelo APARCH (1, 1) com distribuição *t-student* assimétrica

Teste			Estatística	p-valor
Jarque-Bera	R	Chi2	377.416	0
Shapiro-Wilk	R	W	0.9838802	3,31E-10
Ljung-Box	R	Q(10 defasagens)	4.409.607	0.9269832
Ljung-Box	R	Q(15 defasagens)	6.054.085	0.9788249
Ljung-Box	R	Q(20 defasagens)	8.274.382	0.9898912
Ljung-Box	R $\hat{2}$	Q(10 defasagens)	7.640.767	0.6638799
Ljung-Box	R $\hat{2}$	Q(15 defasagens)	1.002.508	0.8181596
Ljung-Box	R $\hat{2}$	Q(20 defasagens)	1.329.622	0.864316
LM-Arch	R	TR $\hat{2}$	1.346.520	0.3361531

R: Resíduos Padronizados.

R $\hat{2}$: Resíduos Padronizados ao Quadrado.

A análise dos resíduos do modelo APARCH com distribuição *t-student* assimétrica, apresentada na Tabela 4, confirma a não normalidade dos resíduos (testes de Jarque-Bera e Shapiro-Wilk). Porém os testes de Ljung-Box (com 10, 15 e 20 defasagens) e LM ARCH confirmam que os resíduos são “ruído branco” e não apresentam dependência serial, evidenciando um ajuste apropriado.

Para verificar o desempenho da previsão fora da amostra para cada um dos modelos, procedeu-se a previsão com todos os modelos, um passo a frente para 140 períodos, perfazendo uma semana completa (cinco dias). Estas previsões foram comparadas com a série de volatilidade dessazonalizada para o período analisado na Tabela 5 utilizando como critério de comparação o erro médio quadrado, o erro absoluto médio e a desigualdade de U-Theil (Theil, 1966).

Tabela 5

Desempenho dos modelos fora-da-amostra

Distribuição	Erro Médio Quadrado	Erro Absoluto Médio	Desigualdade U-Theil
Normal	0.0018377	0.0014068	0.47013
Normal assimétrica	0.0018444	0.0014091	0.45308
<i>t</i> -student	0.0018399	0.0014092	0.46607
<i>t</i> -student assimétrica	0.0018394	0.0014094	0.46534
GED	0.0018389	0.0014082	0.46879
GED assimétrica	0.0018395	0.0014095	0.46686

A Tabela 5 apresenta o desempenho das previsões através dos critérios mencionados. A desigualdade de U-Theil mostra que a distribuição normal assimétrica possui o melhor grau de poder preditivo. Por outro lado, o erro médio quadrado e o erro absoluto médio apontam para a distribuição normal como a que melhor prevê o comportamento da volatilidade da série. A contradição entre o melhor desempenho dos modelos dentro e fora da amostra destaca que o melhor ajuste num período longo não se confirma como melhor ajuste num período curto.

7. Sumário e Conclusões

O presente estudo teve como objetivo estimar e comparar a performance do modelo APARCH aplicado a dados intradiários do índice da Bolsa de Valores de São Paulo presumindo diferentes distribuições (normal, normal assimétrica, generalizada, generalizada assimétrica, *t*-student e *t*-student assimétrica). Os retornos do índice não apresentaram dependência serial, portanto foram utilizados os retornos ao quadrado como medida de volatilidade.

Os resultados apresentaram o modelo APARCH (1,1) com distribuição *t*-student assimétrica como melhor modelo ajustado dentro da amostra, conforme todos os critérios informacionais utilizados. A previsão fora da amostra, entretanto, apontou o modelo APARCH com distribuição normal como o que melhor prevê a volatilidade do índice (conforme dois dos três indicadores utilizados). Este resultado foi diverso do obtido por Liu *et al.* (2009), que verificou que a distribuição generalizada assimétrica melhor se adequava aos índices das bolsas de Shanghai e Shenzhen. É importante salientar que os valores obtidos nos critérios de comparação das previsões parecem ser muito próximos uns dos outros, levando a crer que a diferença entre as distribuições é bastante sutil.

Uma possível explicação para a diferença entre o melhor modelo ajustado (com distribuição *t*-student assimétrica) e o que obteve melhor desempenho fora da amostra (distribuição normal) pode ser o fato de que o ajuste dentro da amostra leva em consideração o perfil de longo prazo dos dados da série, enquanto que a previsão do comportamento fora da amostra considerou um perfil de curto prazo. A peculiaridade dos dados intradiários pode ser a razão dessa divergência. Verificando-se também as estatísticas descritivas dentro e fora da amostra, percebe-se que as características dos dois grupos são ligeiramente distintas. Os testes de normalidade evidenciaram que os dados fora da amostra apresentaram distribuição

normal, diferentemente da série utilizada para ajuste do modelo, que possui curtose elevada.

Outra possível explicação poderia ser que a modelagem feita a partir de uma série longa pode incorporar no modelo efeitos atípicos (evidenciados pelos *outliers*), viciando a previsão. Neste caso, a previsão é realizada supondo tais efeitos, que geralmente são esporádicos, enviesando os valores previstos. Uma possível solução seria realizar o ajuste do modelo utilizando uma série menor, com menos efeitos esporádicos, onde a previsão poderia ter um comportamento mais aproximado ao habitual para a série, minimizando o efeito de eventuais variações acentuadas. Outra contribuição deste estudo é o fato de que em muitos casos, o modelo que melhor se ajusta acaba por não ser o que fornecerá a melhor previsão. Essa evidência ressalta a importância da comparação entre os modelos estudados também fora da amostra (além dos critérios informacionais utilizados no ajuste) para encontrar o modelo que melhor prevê o comportamento futuro da série.

A partir dessa divergência entre ajuste e previsão, pode-se supor que o contexto macroeconômico poderia influenciar tanto o ajuste quanto a previsão de uma série financeira. É possível que características distintas de um período para outro possa resultar que o modelo ajustado não seja o mais adequado ao período que se deseja prever. Por outro lado, é possível que um modelo apenas não seja suficiente para capturar todas as características de uma série financeira. A previsão, portanto, poderia ser feita através de uma combinação de previsões gerando uma estimativa mais provável dos valores posteriores. Dando suporte a essa possibilidade, um estudo dos autores Cavalieri & Ribeiro (2009) demonstrou que a combinação de previsões alcança valores de previsão mais acurados do que os modelos individuais. Sugere-se também que seja feito o mesmo estudo em dados com outras freqüências para verificar se os resultados são mantidos. E por fim, pode-se realizar um estudo com a mesma base de dados, porém considerando um período menor para o ajuste do modelo para verificar os efeitos tanto na adequação do ajuste quanto na previsão.

Referências

- Andersen, Torben G. 2000. Some Reflections on Analysis of High-Frequency Data. *Journal of Business & Economic Statistics*, **18**, 146–153.
- Andersen, Torben G., & Bollerslev, Tim. 1997. Intraday Periodicity and Volatility Persistence in Financial Markets. *Journal of Empirical Finance*, **4**, 115–158.
- Andersen, Torben G., & Bollerslev, Tim. 1998. Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies. *The Journal of Finance*, **53**, 219–265.
- Andersen, Torben G., & Bollerslev, Tim. 1999. Forecasting Financial Market Volatility: Sample Frequency Vis-À-Vis Forecast Horizon. *Journal of Empirical Finance*, **6**, 457–477.

- Awartani, Basel, & Corradi, Valentina. 2005. Predicting the Volatility of the S&P-500 Stock Index Via GARCH Models: The Role of Asymmetries. *International Journal of Forecasting*, **21**, 167–183.
- Baillie, Richard T., & Bollerslev, Tim. 1989. The Message in Daily Exchange Rates: A Conditional Variance Tale. *Journal of Business and Economic Statistics*, **7**, 297–305.
- Balakrishnan, Narayanaswamy, & Nevzorov, Valery B. 2003. *A Primer on Statistical Distributions*. New Jersey: Wiley.
- Bali, Turan. 2007. Modeling the Dynamics of Interest Rate Volatility with Skewed Fat-Tailed Distributions. *Annals of Operational Research*, **151**, 151–178.
- Box, George E., & Tiao, George E. 1973. *Bayesian Inference in Statistical Analysis*. Massachusetts: Addison-Wesley, Reading.
- Brooks, Robert, Faff, Robert, McKenzie, Michael, & Mitchell, Heather. 2000. A Multi-Country Study of Power ARCH Models and National Stock Market Returns. *Journal of International Money and Finance*, **19**, 377–397.
- Carvalho, Marcelo, Freire, Marco Aurélio, Medeiros, Marcelo, & Souza, Leonardo. 2006. Modeling and Forecasting the Volatility of Brazilian Asset Returns: A Realized Variance Approach. *Revista Brasileira de Finanças*, **4**, 321–343.
- Cavaleri, Rosângela, & Ribeiro, Eduardo. 2009. *Combinação de Previsões de Volatilidade: Um Estudo Exploratório*. IX Encontro Brasileiro de Finanças, São Leopoldo, RS.
- Charles, Amelie. 2010. The Day-of-the-Week Effects on the Volatility: The Role of the Asymmetry. *European Journal of Operational Research*, **202**, 143–152.
- Ding, Zhuanxin, Granger, Clive W. J., & Engle, Robert F. 1993. A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, **1**, 83–106.
- Engle, Robert F. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of Variance of UK Inflation. *Econometrica*, **50**, 987–1008.
- Fama, Eugene. 1965. The Behavior of Stock Market Prices. *Journal of Business*, **38**, 34–105.
- Geweke, John. 1986. Modeling the Persistence of Conditional Variances: A Comment. *Econometric Review*, **5**, 57–61.

- Giot, Pierre. 2000. *Intraday Value-at-Risk. Department of Quantitative Economics, Maastricht University (Holland) and Center for Operations Research and Econometrics (CORE), UCL (Belgium)*. Disponível em <http://www.gloriamundi.org/var/wps.html>. Acesso em: 28/01/2010.
- Glosten, Lawrence, Jagannathan, Ravi, & Runkle, David. 1993. On the Relation Between Expected Value and the Volatility of the Nominal Excess Returns on Stocks. *Journal of Finance*, **48**, 1779–1801.
- Higgins, Matthew, & Bera, Anil. 1992. A Class of Nonlinear ARCH Models. *International Economic Review*, **33**, 137–158.
- Hsieh, David. 1989. Modeling Heteroskedasticity in Daily Foreign Exchange Rates. *Journal of Business and Economics Statistics*, **7**, 307–317.
- Jubert, Roberto, Paixão, Márcia, Monte, Paulo, & Lima, Willemberg. 2008. Um Estudo Do Padrão de Volatilidade Dos Principais Índices Financeiros Do Ibovespa: Uma Aplicação de Modelos ARCH. *Revista UnB Contábil*, **11**, 221–239.
- Krishnamoorthy, K. 2006. *Handbook of Statistical Distributions with Applications*. Boca Raton, NW: Chapman & Hall/CRC.
- Laurent, Sébastien. 2004. Analytical Derivatives of the APARCH Model. *Computational Economics*, **24**, 51–57.
- Lee, Cheng F., Chen, Gong-Meng, & Rui, Oliver M. 2001. Stock Returns and Volatility on China Stock Markets. *Journal of Financial Research*, **24**, 523–543.
- Liu, Hung-Chun, Lee, Yen-Hsien, & Lee, Ming-Chih. 2009. Forecasting China Stock Markets Volatility Via GARCH Models under Skewed-GED Distribution. *Journal of Money, Investment and Banking*, **7**.
- Lopez, Jose. 2001. Evaluating the Predictive Accuracy of Variance Models. *Journal of Forecasting*, **20**, 87–109.
- Mandelbrot, Benoit. 1963. The Variation of Certain Speculative Prices. *Journal of Business*, **36**, 394–419.
- Marcucci, Juri. 2005. Forecasting Stock Market Volatility with Regime Switching GARCH Models. *Studies on Nonlinear Dynamics & Econometrics*, **9**, 1–53.
- Moreira, João Maurício, & Lemgruber, Eduardo. 2004. O Uso de Dados de Alta Freqüência Na Estimação Da Volatilidade e Do Valor Em Risco Para O Ibovespa. *Revista Brasileira de Economia*, **58**.
- Morettin, Pedro Alberto. 2008. *Econometria Financeira*. São Paulo: Blucher.

- Nelson, Daniel. 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, **59**, 347–370.
- Otuki, Thiago, Radavelli, Carlos Henrique, Seabra, Fernando, & Costa, Newton, Jr. 2008. Assimetria Na Volatilidade Dos Retornos Revisitada: Ibovespa, Merval e Inmex. *Revista de Gestão da USP*, **15**, 71–84.
- Pantula, Sastry. 1986. Modeling the Persistence of Conditional Variances: A Comment. *Econometric Review*, **5**, 71–74.
- Politis, Dimitris N. 2004. A Heavy-Tailed Distribution for ARCH Residuals with Application to Volatility Prediction. *Annals of Economics and Finance*, **5**, 283–298.
- Schwert, William. 1990. Stock Volatility and the Crash of '87. *Review of Financial Studies*, **3**, 77–102.
- Taylor, Stephen. 1986. *Modeling Financial Time Series*. New York: Wiley.
- Taylor, Stephen J. 1994. Modeling Stochastic Volatility: A Review and Comparative Study. *Mathematical Finance*, **4**, 183–204.
- Theil, Henri. 1966. *Applied Economic Forecasting*. Amsterdam: North-Holland.
- Theodossiou, Panayiotis. 2000. *Skewed Generalized Error Distribution of Financial Assets and Option Pricing*. Working Paper, School of Business, Rutgers University. Disponível em: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=219679. Acesso em: 28/01/2010.
- Zakoian, Jean-Michel. 1994. Threshold Heteroskedasticity Models. *Journal of Economic Dynamics and Control*, **15**, 931–955.