

Volatilidade e Previsão de Retorno com Modelos de Alta Frequência e GARCH: Evidências para o Mercado Brasileiro*

Volatility and Return Forecasting with High-Frequency and GARCH Models: Evidence for the Brazilian Market

Flávio de Freitas Val

Doutorando, Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro
E-mail: flavio.val@uol.com.br

Antonio Carlos Figueiredo Pinto

Professor Doutor, Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro
E-mail: figueiredo@iag.puc-rio.br

Marcelo Cabus Klotzle

Professor Doutor, Departamento de Administração, Pontifícia Universidade Católica do Rio de Janeiro
E-mail: klotzle@iag.puc-rio.br

Recebido em 21.4.2013 - Aceito em 29.4.2013 – 2ª versão aceita em 7.3.2014

RESUMO

Com base em estudos desenvolvidos em anos recentes sobre o uso de dados de alta frequência para a estimação da volatilidade, este artigo implementa o modelo Autorregressivo Heterogêneo (HAR) desenvolvido por Andersen, Bollerslev, e Diebold (2007) e Corsi (2009), e o modelo Componente (2-Comp) desenvolvido por Maheu e McCurdy (2007) e os compara com a família de modelos Autorregressivos com Heteroscedasticidade Generalizados (GARCH) para estimar a volatilidade e os retornos. Durante o período analisado, os modelos que usam dados *intraday* obtiveram melhores previsões de retornos dos ativos avaliados, tanto dentro como fora da amostra, confirmando assim que esses modelos possuem informações importantes para uma série de agentes econômicos.

Palavras-chave: Volatilidade realizada. Estimativa de volatilidade. Retorno *intraday*. HAR.

ABSTRACT

Based on studies developed over recent years about the use of high-frequency data for estimating volatility, this article implements the Heterogeneous Autoregressive (HAR) model developed by Andersen, Bollerslev, and Diebold (2007) and Corsi (2009), and the Component (2-Comp) model developed by Maheu and McCurdy (2007) and compare them with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family models in order to estimate volatility and returns. During the period analyzed, the models using intraday data obtained better returns forecasts of the assets assessed, both in and out-of-sample, thus confirming these models possess important information for a variety of economic agents.

Keywords: Realized volatility. Volatility estimation. Intraday return. HAR.

* Os autores são gratos ao Prof. Dr. Claudio Henrique da Silveira Barbedo do IBMEC-Rio e Banco Central do Brasil por seus comentários valiosos.

1 INTRODUÇÃO

Dados de alta frequência resultam de observações disponibilizadas ao longo de curtos períodos de tempo. Para séries históricas financeiras, isso pode ser descrito como observações disponibilizadas com frequência diária ou com intervalos ainda menores, quando existem vários bancos de dados fornecendo informações a cada negociação de ativos financeiros.

A disponibilidade de bancos de dados de negociadores e os avanços nos cálculos cada vez mais têm disponibilizado esses dados aos pesquisadores e negociadores e gerou um enorme crescimento nas pesquisas empíricas na área de finanças.

Este desenvolvimento abriu o caminho para uma ampla gama de aplicações empíricas, particularmente em mercados financeiros líquidos, com grandes volumes e alta frequência de negócios e baixos custos de transação. Entre essas aplicações, pesquisas aplicadas à estimativa, previsão e comparação da volatilidade dos retornos sobre ativos financeiros com diferentes frequências se destacam.

Além disso, dados de alta frequência também estão sendo amplamente utilizados para estudar questões relacionadas à microestrutura do mercado, tais como: o comportamento dos participantes em um mercado específico, a dinâmica de preços e como afetam as transações e ofertas de compra e venda de um ativo em particular, concorrência entre mercados relacionados e modelagem em tempo real da dinâmica do mercado.

Este artigo contribui à literatura de estudos da eficácia das estimativas de retorno produzidas por modelos de volatilidade de dados de alta frequência. Foram propostos dois modelos de retornos bivariados e volatilidade realiza-

da, e sua contribuição à melhoria das previsões de retorno. Deve-se observar que as evidências empíricas sugerem que as previsões de variação quadrática, baseadas em dados de alta frequência, são melhores que os modelos padrão de estimativa de volatilidade. Portanto, os resultados apresentados no presente estudo oferecem uma contribuição importante para melhores estimativas de volatilidade e precificação de ativos financeiros. Em termos práticos, podem-se usar os modelos implementados para validar e refinar os modelos de preço e retorno *intraday*. Portanto, podem ser úteis em estratégias de investimento *intraday*, em estratégias com posições longas-curtas e na gestão de risco, por exemplo, para calcular diferentes volatilidades condicionais no intuito de comparar e melhorar as metodologias de Valor em Risco.

O artigo foi organizado da seguinte maneira. A próxima seção apresenta um resumo da literatura relevante. Na seção 3, são descritos os dados que foram usados para construir retornos diários da estimativa da volatilidade realizada (VR) diária. Nessa estimativa, os ajustes de VR se destacam para remover os efeitos da microestrutura do mercado. A seção 4 descreve a metodologia e estima o retorno e os modelos de VR com base em dados *intraday*, e os modelos de referência com base nos retornos diários. Na seção 5, os resultados empíricos são apresentados e validados mediante o uso da variância condicional *intraday* para estimar o Modelo de Precificação de Ativos Capitais. E, finalmente, a seção 6 destaca as conclusões deste estudo.

2 RESUMO DA LITERATURA RELEVANTE

A análise de dados de alta frequência impõe novos desafios aos pesquisadores, já que esses dados têm características únicas, não presentes em bancos de dados com frequências inferiores.

Desde a apresentação por Hsieh (1991) de uma das primeiras estimativas de variância nos retornos diários baseados nos retornos *intraday* do índice de ações S&P500, alcançaram-se progressos em várias áreas diferentes. Entre outros artigos seminais que discutem as propriedades e características únicas das distribuições dos dados *intraday*, podem-se mencionar: Zhou (1996), que usou dados de frequência hiper-alta relevantes aos mercados de câmbio para explicar a autocorrelação negativa da primeira ordem de retornos e para estimar a volatilidade para dados de alta-frequência; Goodhart e O'Hara (1997) que destacam os efeitos da estrutura de mercado na interpretação e análise dos dados, os efeitos sazonais *intraday* e os efeitos da volatilidade temporal; e Andersen e Bollerslev (1997, 1998a) que analisaram o comportamento da volatilidade *intraday*, os choques de volatilidade devido a pronunciamentos macroeconômicos e a persistência de longo prazo na série temporal da volatilidade realizada, também no mercado de câmbio.

Outros trabalhos importantes, tais como os de Andersen e Bollerslev (1998b), Andersen, Bollerslev, Diebold, e Ebens (2001a), Andersen, Bollerslev, Diebold, e Labys (2001b), Barndorff-Nielsen e Shephard (2002) e Meddahi

(2002), estabeleceram as propriedades teóricas e empíricas da estimativa da variação quadrática para uma ampla classe de processos estocásticos na área de finanças, permitindo assim as pesquisas empíricas com uma nova classe de estimadores, incluindo a volatilidade realizada.

Andersen e Benzoni (2008) relatam as aplicações empíricas derivadas das medidas construídas com base em dados de alta frequência, destacando pelo menos quatro grandes áreas de pesquisa: (i) previsão de volatilidade, com ênfase em pesquisas com foco na melhoria do desempenho dessa previsão, na literatura relevante relacionada à detecção de saltos e na investigação de problemas relacionados à microestrutura no desempenho da previsão; (ii) implicações na distribuição de retornos sob condições de não-arbitragem; (iii) medidas multivariadas da variação quadrática; e (iv) volatilidade realizada, modelos de especificação e estimativa.

Dentro dessas subáreas de pesquisa, este artigo enfoca a melhoria do desempenho da previsão de volatilidade, com atenção especial às propriedades das séries temporais e à melhoria dos procedimentos de estimativa, a saber, mediante o uso da volatilidade realizada.

A seguir apresentamos alguns dos artigos principais nesta subárea de pesquisa.

Andersen et al. (2001a) estimam a volatilidade diária realizada de uma série de ações listadas no índice *Dow Jo-*

nes *Industrial Average* (DJIA). Os autores alcançam resultados que afirmam que a distribuição incondicional da variância e covariância realizadas é altamente assimétrica para direita, enquanto o desvio padrão logarítmico realizado e as correlações são aproximadamente Gaussianas. O mesmo vale para a distribuição dos retornos escalados pelos desvios padrão realizados.

Andersen, Bollerslev, Diebold, e Labys (2003) oferecem uma estrutura geral para o uso de dados *intraday* de alta frequência para medir, modelar e prever as distribuições das volatilidades e retornos com frequência diária ou em períodos inferiores.

Ghysels, Santa-Clara, e Valkanov (2005) introduzem um novo estimador que prevê a variância mensal, utilizando os retornos quadrados diários passados e o denominam amostragem de dados mistos (MIDAS).

Andersen et al. (2007) afirmam que mais e mais fontes na literatura confirmam os ganhos na previsão da volatilidade dos ativos financeiros, utilizando-se medidas baseadas em dados de alta frequência. Implementam uma nova medida de volatilidade (medida de variação *bipower*) e testes não paramétricos correspondentes para saltos. A análise empírica das taxas de câmbio, índice de retornos de ações e taxas de títulos sugere que o componente de volatilidade devido a saltos seja muito importante e menos persistente que o componente contínuo, e que a separação entre movimentos de salto e movimentos suaves (contínuos) resulte em uma melhoria significativa na previsão de volatilidade fora da amostra. Além disso, muitos saltos significativos são associados com novos anúncios de eventos macroeconômicos.

Maheu e McCurdy (2007) propõem um modelo flexível e parcimonioso da dinâmica combinada entre o retorno e o risco do mercado para prever o prêmio de capital de mercado variável no tempo. Esse modelo de volatilidade permite diferentes taxas de decaimento para seus componentes, gerando previsões de retornos médios e permitindo metas de variância.

Corsi (2009) propõe um modelo aditivo de volatilida-

de de componentes definidos em diferentes horizontes de tempo. Esse modelo possui componentes autorregressivos na volatilidade realizada e é denominado o Modelo Autorregressivo Heterogêneo de Volatilidade Realizada – HAR-LOG(VR). Fácil de ser implementado, os resultados simulados mostram que este modelo consegue reproduzir as principais características dos retornos sobre ativos financeiros (memória longa, caudas gordas e autossimilaridade). Além disso, os resultados empíricos mostram excelente desempenho de previsão.

Poucos artigos estudaram os benefícios da incorporação da VR na distribuição dos retornos. Esses incluem Andersen et al. (2003) e Giot e Laurent (2004), que consideram o valor da VR para estimativas e para calcular o Valor em Risco, comparando o desempenho de um modelo ARCH, que usa os retornos diários, com o desempenho de um modelo baseado na volatilidade realizada diária – que usa retornos *intraday* – em carteiras do índice de ações e taxas de câmbio. Essas abordagens separam a dinâmica dos retornos e volatilidades e assumem que a VR seja uma medida suficiente para representar a variância condicional dos retornos. Ghysels et al. (2005) observam que as medidas de volatilidade de alta frequência identificam o *tradeoff* entre risco e retorno com frequências menores.

Entre os estudos aplicados ao mercado brasileiro, Moreira e Lemgruber (2004) avaliam o uso de dados de alta frequência na previsão de volatilidade e no VaR, utilizando os modelos GARCH com horizontes diários e *intraday*. Entre os resultados, destacaram que os dados *intraday* podem melhorar significativamente o VaR de um dia. O artigo mais notável no mercado brasileiro foi desenvolvido por Wink Junior e Valls Pereira (2012) que, de forma pioneira, escolhem o intervalo de tempo *intraday* ótimo, tratam da questão do ruído gerado por uma microestrutura de mercado e implementam dois modelos recentes, que usam dados de alta frequência para estimar e prever a volatilidade de cinco ações representativas do Índice Bovespa.

3 DADOS E ESTIMATIVA DA VOLATILIDADE REALIZADA

Neste estudo, foram utilizados os preços negociados para as ações PETR4 e VALE5, as duas ações com maior liquidez no mercado brasileiro de ações. Os preços das negociações dessas ações foram obtidos diretamente da BM&FBovespa.

Os dados na amostra se referem ao período entre 1º de dezembro de 2009 e 23 de março de 2012 para ambas as ações.

Após remover os erros dos dados das negociações, foi construída uma grade de cinco minutos dentro do período das negociações do leilão eletrônico, encontrando-se o preço negociado igual a ou depois mais próximo de cada intervalo na grade. Dessa grade, foram construídos retornos continuamente compostos de cinco minutos (retornos log). Esses retornos foram multiplicados por 100 e anotados como $r_{t,i} = 1, \dots, I$, em que I é a quantidade de retornos *intraday* no dia t . Para essa grade de cinco minutos, a média é de $I=83$ para cada dia de negociação. Essa rotina gerou, respectivamente, 47.334 e 47.322 retornos de cinco minutos para as ações PETR4 e VALE5 ao longo dos 573 dias em

que as ações foram negociadas.

O aumento na variação quadrática é uma medida natural da variância *ex-post* dentro de um período específico. A variância realizada, também conhecida como volatilidade realizada, é um dos mais populares estimadores da variância quadrática, calculada como a soma dos retornos quadrados ao longo de um período de tempo específico.

Assim, diante dos retornos *intraday* $r_{t,i} = 1, \dots, I$, um estimador diário não ajustado da VR é

$$VR_{t,u} = \sum_{i=1}^I r_{t,i}^2$$

3.1

Porém, na presença da dinâmica da microestrutura do mercado, a VR pode ser distorcida e é um estimador inconsistente da variação quadrática (Bandi & Russell, 2004). Portanto, a volatilidade realizada diária foi ajustada com o método da média móvel, usado por Andersen et al. (2001a) e posteriormente generalizado por Hansen, Large, e Lunde (2008) e também implementado no mercado brasileiro de

ações por Wink Junior e Valls Pereira (2012).

Assim, se o retorno *intraday* de um ativo seguir um processo de média móvel na ordem de (MA(q)), dado por $r_{t,m} = \varepsilon_t - \theta_1 \varepsilon_{t-1,m} - \dots - \theta_q \varepsilon_{t-q,m}$, Hansen et al. (2008) mostram que, diante de algumas hipóteses, o estimador que corrige o viés não ajustado na VR, baseado no processo MA(q), é dado por:

$$VR_{t,MAq} = \frac{(1 - \hat{\theta}_1 - \dots - \hat{\theta}_q)^2}{1 + \hat{\theta}_1^2 + \dots + \hat{\theta}_q^2} VR_{t,u} \tag{3.2}$$

Assim, para evitar uma lacuna entre as medidas de volatilidade *interday* e *intraday*, os retornos diários r_t usados nos modelos da família GARCH foram calculados pela diferença logarítmica entre o último preço do dia e o último preço do dia anterior, ambos capturados na grade de cinco minutos. Esses retornos também foram multiplicados por 100.

A Figura 1 mostra que as volatilidades realizadas das ações analisadas têm autocorrelação serial significativa.

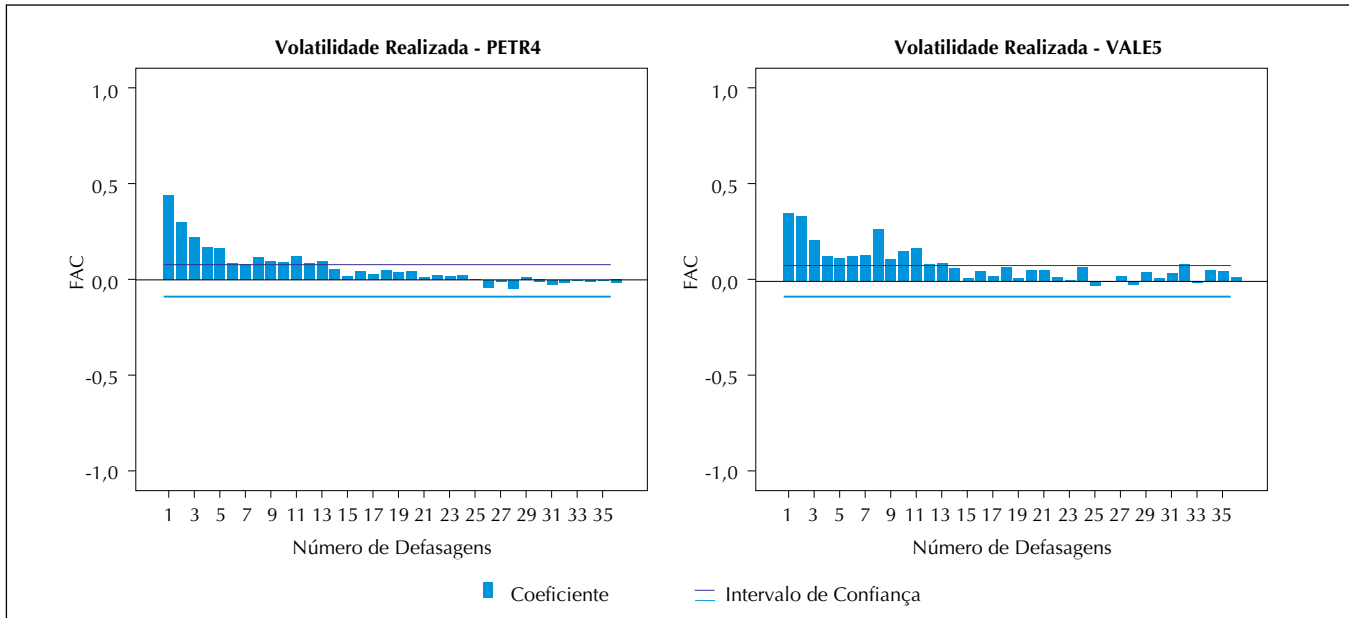


Figura 1 Autocorrelação das Volatilidades Realizadas para PETR4 e VALE5

A Tabela 1 mostra as estatísticas descritivas para os retornos diários e para a VR diária estimada, utilizando-se a grade de cinco minutos. Há um certo viés na VR não ajustada. Após a análise dos correlogramas da VR diária e o critério adotado por Maheu e McCurdy (2011) para remover

esse viés, um processo MA com q=8 parece necessário para os retornos da PETR4, enquanto q=11 é apropriado para os retornos da VALE5. A partir daqui, $RV_t = RV_{t,MAq}$ será usado com q=8 e q=11 para as estimativas, respectivamente, da PETR4 e VALE5.

Tabela 1 Estatísticas resumidas: retornos diários e volatilidades realizadas

		Média	Variância	Assimetria	Curtose	Mínimo	Máximo
PETR4	r_t	-0,090	3,121	-0,404	1,359	-7,596	5,328
	RV_u	3,155	11,124	4,580	28,561	0,417	32,743
	RV_{ma1}	0,629	0,443	4,580	28,561	0,083	6,532
	RV_{ma2}	0,023	0,001	4,580	28,561	0,003	0,237
	RV_{ma3}	0,083	0,008	4,580	28,561	0,011	0,863
	RV_{ma4}	0,083	0,008	4,580	28,561	0,011	0,863
	RV_{ma4}	0,330	0,121	4,580	28,561	0,044	3,420
	RV_{ma5}	1,031	1,188	4,580	28,561	0,136	10,699
	RV_{ma6}	1,746	3,405	4,580	28,561	0,231	18,115
	RV_{ma7}	2,223	5,521	4,580	28,561	0,294	23,067
	RV_{ma8}	3,083	10,624	4,580	28,561	0,408	31,999
	RV_{ma9}	3,645	14,844	4,580	28,561	0,482	37,823
	RV_{ma10}	0,083	0,008	4,580	28,561	0,011	0,863
RV_{ma10}	3,951	17,443	4,580	28,561	0,522	41,001	

continua

continuação

		Média	Variância	Assimetria	Curtose	Mínimo	Máximo
VALE5	r_t	-0,011	3,015	-0,364	3,009	-9,958	5,815
	RV_u	2,866	14,783	6,344	56,496	0,303	46,845
	RV_{ma1}	0,967	1,684	6,344	56,496	0,102	15,812
	RV_{ma2}	0,166	0,050	6,344	56,496	0,018	2,717
	RV_{ma3}	0,000	0,000	6,344	56,496	0,000	0,001
	RV_{ma4}	0,204	0,075	6,344	56,496	0,022	3,328
	RV_{ma5}	0,499	0,448	6,344	56,496	0,053	8,155
	RV_{ma6}	0,646	0,750	6,344	56,496	0,068	10,555
	RV_{ma7}	0,699	0,879	6,344	56,496	0,074	11,421
	RV_{ma8}	1,568	4,427	6,344	56,496	0,166	25,635
	RV_{ma9}	1,646	4,878	6,344	56,496	0,174	26,908
	RV_{ma10}	1,830	6,029	6,344	56,496	0,193	29,917
	RV_{ma11}	3,016	16,374	6,344	56,496	0,319	49,302
RV_{ma12}	3,468	21,650	6,344	56,496	0,367	56,691	

4 METODOLOGIA

Neste artigo, modelos bivariados foram propostos com base em duas formas alternativas de relação entre a VR e a variância condicional dos retornos e os modelos da família GARCH servem como arcabouço para a análise de desempenho dos modelos *intraday*.

De acordo com Maheu e McCurdy (2011), propuseram-se duas formas funcionais para os modelos bivariados dos retornos e da VR. O primeiro modelo usa a função autorregressiva heterogênea (HAR) do logaritmo defasado (VR) (Corsi, 2009; Andersen, Bollerslev, & Diebold, 2007). O segundo modelo permite que os componentes do logaritmo(VR) tenham diferentes taxas de decaimento (Maheu e McCurdy, 2007).

Uma maneira de se vincular a VR à variância nos retornos também foi considerada, impondo a restrição de que a variância condicional nos retornos diários seja igual à expectativa condicional da VR diária.

Da mesma forma que com os modelos EGARCH e TGARCH, modelos bivariados permitem o assim chamado efeito de alavancagem, ou assimetrias, das inovações negativas contra as inovações positivas nos retornos.

Uma maneira de se confirmar que a informação *intraday* contribui à melhoria nas estimativas das distribuições dos retornos é comparar as estimativas dos modelos bivariados do retorno e logaritmo(VR) especificado nas estimativas dos modelos da família GARCH:

$$r_t = \rho_1 r_{t-1} + \rho_2 r_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t, \\ e_t = \sigma_t u_t, u_t \sim NID(0,1) \quad 4.1$$

$$\text{GARCH: } \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha e_{t-1}^2 \quad 4.2$$

$$\text{EGARCH: } \log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \gamma \frac{e_{t-1}}{\sigma_{t-1}} + \alpha \left| \frac{e_{t-1}}{\sigma_{t-1}} \right| \quad 4.3$$

$$\text{TGARCH: } \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha e_{t-1}^2 + \gamma e_{t-1}^2 I_{t-1} \text{ em que } I_t = 0 \\ \text{se } e_t < 0 \text{ e } 0 \text{ caso contrário} \quad 4.4$$

O principal método de comparação implementado usa a raiz quadrada do erro quadrático médio e o teste Modificado de Diebold e Mariano (1995), baseado na obra de Harvey, Leybourne, e Newbold (1997). Intuitivamente, modelos mais bem estimados terão menos erros de previsão e, em comparação com os outros modelos de desempenho inferior, apresentarão diferenças estatísticas nos seus erros. Portanto, para avaliar os modelos implementados neste artigo, enfocamos a acurácia relativa desses modelos na estimativa dos retornos dentro e fora da amostra.

Um aspecto importante da abordagem utilizada é a possibilidade de comparar diretamente as especificações de volatilidade tradicionais, tais como os modelos da família GARCH, com os modelos dos retornos bivariados e VR, porque os modelos implementados têm um critério em comum – a previsão dos retornos. A média e o teste estatístico desses erros de previsão nos permitem investigar a contribuição relativa da VR nas previsões.

4.1 Modelos de Retornos Bivariados e Variância Realizada.

Nesta subseção, duas especificações combinadas do retorno diário e da VR foram implementadas. Esses modelos bivariados se distinguem pelas suas condições alternativas para a dinâmica da VR. Em cada caso, as restrições entre as equações vinculam a variância dos retornos à especificação da VR.

O primeiro corolário de Andersen et al. (2003) mostra que, sob condições empíricas realistas, a expectativa condicional da variação quadrática (VQ_t) é igual à variância condicional dos retornos, i.e. $E_{t-1}(VQ_t) = Var_{t-1}(r_t) \equiv \sigma_t^2$. Se a VR for um estimador não enviesado da variação quadrática, segue que a variância condicional dos retornos pode ser vinculada a VR como $\sigma_t^2 = E_{t-1}(VR_t)$, em que a combinação das informações se define como $\sigma_{t-1} \equiv \{r_{t-1}, VR_{t-1}, r_{t-2}, VR_{t-2}, \dots, r_1, VR_1\}$. Supondo-se

que a VR tenha distribuição log-normal, a restrição assume a seguinte forma:

$$\sigma_t^2 = E_{t-1}(VR_t) = \exp(E_{t-1} \log(VR_t) + \frac{1}{2} Var_{t-1}(\log(VR_t))) \quad 4.5$$

4.1.1 Especificação autorregressiva heterogênea: Modelo HAR.

O primeiro modelo implementado possui uma especificação bivariada dos retornos diários e da VR em que os retornos condicionais são dirigidos por inovações normais e a dinâmica do $\log(VR_t)$ é capturada em uma função autorregressiva heterogênea (HAR) do $\log(VR_t)$ defasado. Corsi (2009) e Andersen et al. (2007) usam as funções HAR para captar a dependência da memória de longo prazo com parcimônia. Motivados por esses estudos, definimos:

$$\log(VR_{t-h,h}) \equiv \frac{1}{h} \sum_{i=0}^{h-1} \log(VR_{t-h+i}) \quad 4.6$$

$$\log(VR_{t-1,1}) \equiv \log(VR_{t-1})$$

Por exemplo, $\log(VR_{t-22,22})$ é estimado mediante o cálculo da média do $\log(VR)$ para os últimos 22 dias, i.e. de t-22 até t-1, $\log(VR_{t-5,5})$ considera a média dos últimos cinco dias.

Isso toma a especificação dos retornos diários e da VR com a dinâmica do $\log(VR_t)$ sendo modelada como uma função HAR assimétrica do $\log(VR)$ passado¹. Este sistema bivariado é resumido da seguinte forma:

$$r_t = \rho_1 r_{t-1} + \rho_2 r_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t, \quad 4.7$$

$$e_t = \sigma_t u_p, u_t \sim NID(0,1)$$

$$\log(VR_t) = \omega + \theta_1 \log(VR_{t-1}) + \theta_2 \log(VR_{t-5}) + \theta_3 \log(VR_{t-22}) + \gamma e_{t-1} + \eta v_t, v_t \sim NID(0,1) \quad 4.8$$

Essa especificação bivariada dos retornos diários e da VR impõe a equação de restrição que relaciona a variância condicional nos retornos diários com a expectativa condicional da VR diária, conforme mostrado em (4.5).

Como o banco de dados analisado neste artigo refere-se a retornos de ações, é importante permitir efeitos assimétricos na volatilidade. Para facilitar as comparações com o modelo de referência EGARCH, a parametrização na equação (4.8) inclui o termo assimétrico γe_{t-1} , associado com as inovações nos retornos e_{t-1} . O coeficiente de impacto para as inovações negativas nos retornos será γ . Tipicamente, $\tilde{\gamma} < 0$, o que significa que as inovações negativas nos retornos implicam maior variância condicional para o período seguinte.

4.1.2 Especificação Component-Log(RV): Modelo 2-Comp.

Essa especificação bivariada dos retornos diários e VR possui retornos condicionais direcionados por inovações normais, mas a dinâmica do $\log(VR)$ é capturada por dois componentes (2-Comp) com taxas de decaimento diferentes, conforme demonstrada em Maheu e McCurdy (2007). Particularmente, esse sistema bivariado pode ser representado na seguinte equação:

$$r_t = \rho_1 r_{t-1} + \rho_2 r_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t, \quad 4.9$$

$$e_t = \sigma_t u_p, u_t \sim NID(0,1)$$

$$\log(VR_t) = \omega + \sum_{i=1}^2 \theta_i s_{i,t} + \gamma e_{t-1} + \eta v_t, v_t \sim NID(0,1) \quad 4.10$$

$$s_{i,t} = (1 - \alpha_i) \log(VR_{t-1}) + \alpha_i s_{i,t-1}, 0 < \alpha_{i,t} < 1, i=1,2 \quad 4.11$$

Novamente uma restrição foi imposta pela equação (4.5), que relacionou a variância condicional dos retornos diários com a expectativa condicional da VR diária. Para essa especificação, a dinâmica do $\log(VR)$ diário foi parametrizada pelas equações (4.10) e (4.11), substituindo a função HAR na equação (4.8).

A média da série temporal de retornos foi estimada por modelos ARMA (p, q), utilizando a função do *software* R *auto.arima*. Para todas as outras estimativas foi utilizado o *software* *eviews* 7.1. No intuito de combinar a parcimônia e solidez dessas estimativas, estabelecemos a defasagem máxima (p+q) de 4 e seleção automática da defasagem com o Critério de Informação de Schwarz (BIC). Assim, para a média da série temos um AR (1) modelando as séries diárias e semanais das ações VALE5 e PETR4 e um ARMA (2,2) para as séries semanais e mensais de ambos os ativos.

Os sistemas bivariados foram estimados em dois passos. Inicialmente, as equações das médias foram estimadas e, depois, as inovações nos retornos foram modeladas com o uso de diferentes modelos de volatilidade. As previsões da volatilidade foram feitas com uma sequência de previsões um passo à frente, utilizando os valores atuais para a variável dependente defasada e as previsões de retorno, considerando $r_{i,t} = f_{i,t} + e_{i,t}$, $e_{i,t} = \sigma_{i,t} u_p$, $u_t \sim NID(0,1)$ em que $f_{i,t}$ é a equação estimada para o processo médio do ativo i no tempo t , $e_{i,t}$ é a inovação do retorno para o ativo i no tempo t e $\alpha_{i,t}$ é o desvio padrão condicional estimado para o ativo i no tempo t .

Para fins de um exercício prático da aplicabilidade dos modelos apresentados aqui, usamos a variância condicional estimada com o modelo mensal 2-Comp, que retornou as previsões de erro mais baixas entre os modelos analisados, na estimativa com o modelo *Capital Asset Pricing Model* (CAPM). O CAPM, desenvolvido por Sharpe (1964), Treynor (1961), Lintner (1965) e Black, Jensen, e Scholes (1972), tem se tornado nas últimas décadas o modelo mais amplamente divulgado para a determinação dos preços de ativos (Barros, Famá, & Silveira, 2002). Esse modelo determina que os ativos são precificados de forma compatível com um *trade-off* entre o risco não-diversificável e as expectativas de retorno.

O CAPM pode ser apresentado formalmente como $E(r_i) = r_f + \beta_i (E(r_m) - r_f)$, em que $E(r_i)$ é o retorno esperado do ativo i ao longo de um único período, r_f é o risco da taxa de juros livre de risco ao longo do período, $E(r_m)$ o retorno esperado no mercado ao longo do período, e $\beta = \frac{Cov(r_m, r_i)}{\sigma^2(r_m)}$ identifica a exposição do ativo i ao mercado.

Para estimar β , usamos: (i) a covariância estatística $Cov(r_m, r_i)$ da PETR4 e VALE5 na BOVA11 em uma janela mó-

¹ A série temporal RV para os ativos usados aqui é estacionária de acordo com o teste de raízes unitárias, que rejeita a hipótese nula de não estacionariedade.

vel de 22 dias e (ii) a variância condicional $\sigma^2(r_m)$ estimada pelo modelo 2-Comp na mesma janela temporal. O Certificado Interbancário de Depósitos (CDI) foi selecionado para representar o risco da taxa de juros livre de risco, de acordo com Barros, Famá, e Silveira (2002).

Para representar o portfólio do mercado, escolhemos o fundo de índice BOVA11, considerando que seu retorno esperado para o período subsequente é uma função do re-

torno do período anterior. As qualidades do BOVA11 incluem: (i) é negociado efetivamente em um mercado ativo, permitindo a extração da volatilidade realizada e seu uso nos modelos de volatilidade *intraday*, (ii) possui uma correlação média de mais de 99% com o Ibovespa² no período analisado, (iii) é um ativo com liquidez crescente, com giro diário médio de R\$44.1mi em 2011, e de longe o fundo de índice mais negociado no mercado brasileiro.

5 RESULTADOS EMPÍRICOS

Nesta seção, apresentaremos os resultados estimados dos modelos para as janelas de 1, 5 e 22 dias.

A Tabela 2 apresenta as estimativas dos modelos da família GARCH. O Critério de Informação de Schwarz (BIC) indica que modelos assimétricos são tão bem ajustados aos dados quanto o modelo GARCH estimado, confirmado em

quase todas as previsões dentro da amostra desses modelos. Com base nos modelos EGARCH e TGARCH ajustados, exceto a estimativa mensal da PETR4, todos os coeficientes do efeito de alavancagem são significativos no nível de 10%, confirmando o impacto assimétrico entre os retornos positivos e negativos dos ativos.

Tabela 2 Estimativas com os modelos GARCH, EGARCH e TGARCH

$$r_t = \rho_1 r_{t-1} + \rho_2 r_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t, e_t = \sigma_t u_t, u_t \sim NID(0,1)$$

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha e_{t-1}^2$$

Horizonte	Ação/ Parâmetro	ρ_1	ρ_2	θ_1	θ_2	R ² Ajust.	ω	β	α	γ	BIC
1 dia	PETR4						0,159***	0,883*	0,07**		3,774
	VALE5	0,104**				0,011	0,090	0,060	0,030		3,668
5 dias	PETR4	0,815*				0,661	0,072	0,057	0,034		4,325
	VALE5	1,299*	-0,664*	-0,664*	0,779*	0,707	0,211***	0,856*	0,092**		4,222
22 dias	PETR4		0,908*	0,978*	-0,041	0,889	0,285	0,892*	0,053**		4,483
	VALE5	0,228**	0,638*	0,907*	0,276*	0,905	0,215	0,062	0,029		4,325
		0,119	0,114	0,118	0,043		0,032	0,021	0,018		

$$r_t = \rho_1 r_{t-1} + \rho_2 r_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t, e_t = \sigma_t u_t, u_t \sim NID(0,1)$$

$$\log(\sigma_t^2) = \omega + \beta \log(\sigma_{t-1}^2) + \gamma \frac{e_{t-1}}{\sigma_{t-1}} + \alpha \left| \frac{e_{t-1}}{\sigma_{t-1}} \right|$$

Horizonte	Ação/ Parâmetro	ρ_1	ρ_2	θ_1	θ_2	R ² Ajust.	ω	β	α	γ	BIC
1 dia	PETR4						-0,051	0,9405*	0,115**	-0,138*	3,778
	VALE5	0,104**				0,011	0,044	0,033	58,000	0,041	3,637
5 dias	PETR4	0,815*				0,661	-0,005	0,904*	0,101**	-0,183*	4,325
	VALE5	1,299*	-0,664*	-0,664*	0,779*	0,707	0,032	0,022	0,044	0,037	4,221
22 dias	PETR4		0,908*	0,978*	-0,041	0,889	0,044	0,022	0,059	0,034	4,499
	VALE5	0,228**	0,638*	0,907*	0,276*	0,905	0,613	0,496***	0,245**	-0,078	4,329
		0,119	0,114	0,118	0,043		0,429	0,287	0,111	0,073	
							-0,069*	1,003*	0,078*	-0,045**	
							0,021	0,005	0,028	0,020	

continua

² O Índice Bovespa (Ibovespa) é o indicador mais importante dos preços médios das ações negociadas na Bolsa de Valores de São Paulo e inclui as ações com os maiores volumes de negócios nos meses recentes.

continuação

$$r_t = \rho_1 r_{t-1} + \rho_2 r_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t, e_t = \sigma_t u_t, u_t \sim NID(0,1)$$

$$\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha e_{t-1}^2 + \gamma e_{t-2}^2 + I_{t-1}$$

Horizonte	Ação/ Parâmetro	ρ_1	ρ_2	θ_1	θ_2	R ² Ajust.	ω	β	α	γ	BIC
1 dia	PETR4						0,155**	0,858*	-0,003	0,152**	3,769
							0,081	0,051	0,027	0,060	
	VALE5	0,104**				0,011	0,142*	0,879*	-0,041**	0,216*	3,629
		0,042					0,036	0,030	0,017	0,046	
5 dias	PETR4	0,815*				0,661	0,125	0,902*	-0,015	0,164**	4,316
		0,025					-0,015	0,048	0,032	0,054	
	VALE5	1,299*	-0,664*	-0,664*	0,779*	0,707	0,196**	0,888*	0,002	0,123**	4,213
		0,042	0,040	0,034	0,029		0,079	0,041	0,031	0,051	
22 dias	PETR4		0,908*	0,978*	-0,041	0,889	0,215	0,913*	0,031	0,026	4,497
			0,017	0,043	0,044		0,157	0,048	0,031	0,039	
	VALE5	0,228**	0,638*	0,907*	0,276*	0,905	-0,027***	0,979*	-0,013	0,073*	4,328
		0,119	0,114	0,118	0,043		0,016	0,012	0,016	0,022	

Nota: *, ** e *** indicam a rejeição da hipótese nula com níveis de significância em 1%, 5% e 10%, respectivamente.

Ambos os modelos assimétricos oferecem efeitos de alavancagem semelhantes para os horizontes temporais estimados. Considerando um choque estandardizado de dois desvios padrão, o efeito de alavancagem para EGARCH(1,1,1) pode ser estimado (Tsay, 2010) como $\frac{\sigma_t^2(u_{t-1} = -2)}{\sigma_t^2(u_{t-1} = +2)} = \frac{\exp[-(\alpha-\gamma)(u_{t-1})]}{\exp[(\alpha+\gamma)(u_{t-1})]}$, e para TGARCH(1,1,1) como $\frac{\sigma_t^2(u_{t-1} = -2)}{\sigma_t^2(u_{t-1} = +2)} = \frac{[(\alpha+\gamma)(u_{t-1})^2 + \beta]}{[\alpha(u_{t-1})^2 + \beta]}$.

A seguinte Tabela 3 compara esses modelos assimétricos, assumindo que $e_{t-1} = \pm 2\sigma_{t-1}$ de tal forma que $u_{t-1} = \pm 2$. O impacto de um choque negativo de dois desvios padrão

é mais potente para os retornos diários. Assim, o efeito de alavancagem se reduz com o aumento dos horizontes temporais.

Por exemplo, para a VALE5 com horizonte de um dia, o impacto de um choque negativo de dois desvios padrão é aproximadamente 121% e 108% superior, respectivamente, para os modelos EGARCH e TGARCH, quando comparado ao impacto de um choque positivo da mesma magnitude. Para um horizonte de 22 dias, esses impactos são reduzidos para 31% e 20%.

Tabela 3 Efeito de Alavancagem (considerando 2 desvios padrão)

Modelo	Ação	1 dia	5 dias	22 dias
EGARCH	PETR4	1,74	1,60	1,37
	VALE5	2,21	1,51	1,31
TGARCH	PETR4	1,72	1,78	1,10
	VALE5	2,08	1,44	1,20

As Tabelas 4 e 5, a seguir, mostram as estimativas dos modelos HAR-Log(RV) e 2-Comp, respectivamente, para os horizontes temporais de 1, 5 e 22 dias, respectivamente.

Os resultados mostrados na Tabela 4 estão de acordo com Andersen et al. (2007) e Wink Junior e Valls Pereira (2012). As estimativas estatisticamente significativas dos coeficientes dos componentes de volatilidade diária (θ_1), semanal (θ_2) e mensal (θ_3) confirmam a presença de alta persistência na volatilidade³. O peso relativo do componente de volatilidade diária diminui das regressões diárias

para o componente semanal e mensal, enquanto o componente mensal tende a ter relativamente maior importância nas regressões ao longo de períodos maiores⁴.

Além disso, quando se compara o R² ajustado do HAR, que inclui o componente γ assimétrico com a mesma estatística HAR padrão (última coluna da Tabela 3), observa-se pouca melhoria na estimativa dos modelos, indicando que os coeficientes do HAR já estavam capturando alguma dinâmica assimétrica dos retornos sobre os ativos.

³ Somente três dos 18 coeficientes de volatilidade não são significantes ao nível de 10%. Essas exceções se referem aos componentes de volatilidade estimados em equações com horizontes de um e cinco dias.

⁴ Conforme observado por Andersen et al. (2007), apesar de a estrutura do modelo HAR não conter formalmente uma memória, a combinação entre poucos componentes de volatilidade consegue reproduzir uma queda suave notável na autocorrelação dessa volatilidade, sendo quase indistinguível da decaimento hiperbólico (memória longa).

Tabela 4 Estimativas com o modelo HAR-log(RV)

$$r_t = \rho_1 r_{t-1} + \rho_2 r_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t, e_t = \sigma_t u_t, u_t \sim NID(0,1)$$

$$\log(VR_t) = \omega + \theta_1 \log(VR_{t-1}) + \theta_2 \log(VR_{t-2}) + \theta_3 \log(VR_{t-22}) + \gamma e_{t-1} + \eta v_t, v_t \sim NID(0,1)$$

Horizonte	Ação/ Parâmetro	ρ_1	ρ_2	θ_1	θ_2	R ² Ajust.	ω	θ_1	θ_2	θ_3	γ	η	R ² _j Ajust.	R ² Ajust.
1 dia	PETR4	-	-	-	-	-	0,156**	0,290*	0,345*	0,093	-0,038**	-0,021	0,3354	0,3310
		-	-	-	-	-	0,064	0,060	0,091	0,090	0,017	0,026		
	VALE5	0,104**	-	-	-	0,011	0,057	0,312*	0,220**	0,267*	-0,058*	-0,018	0,3891	0,3753
		0,042	-	-	-		0,055	0,060	0,089	0,088	0,019	0,027		
5 dias	PETR4	0,815*				0,661	0,053*	0,113*	0,855*	-0,016	-0,011*	-0,001	0,9049	0,9040
		0,025					0,019	0,018	0,028	0,028	0,004	0,008		
	VALE5	1,299*	-0,664*	-0,664*	0,779*	0,707	0,037**	0,138*	0,807*	0,022	-0,002	-0,007	0,8991	0,8990
		0,042	0,040	0,034	0,029		0,019	0,020	0,030	0,030	0,005	0,009		
22 dias	PETR4		0,908*	0,978*	-0,041	0,889	0,004	0,018*	0,039*	0,944*	-0,001	0,001	0,9859	0,9859
			0,017	0,043	0,044		0,006	0,006	0,009	0,008	0,001	0,002		
	VALE5	0,228**	0,638*	0,907*	0,276*	0,905	0,005	0,014*	0,036*	0,949*	-0,004*	-0,002	0,9891	0,9888
		0,119	0,114	0,118	0,043		0,005	0,005	0,008	0,008	0,001	0,002		

Nota: *, ** e *** indicam a rejeição da hipótese nula com níveis de significância em 1%, 5% e 10%, respectivamente.

Porém, os resultados apresentados na Tabela 5 mostram que o modelo 2-Comp estimado conseguiu captar eficazmente as diferentes dinâmicas da volatilizada, distinguindo claramente entre os coeficientes persistentes α_1 e α_2 de acordo com cada horizonte temporal (veja a Figura 2). Além disso, o coeficiente α inferior em cada equação mostra um efeito menos persistente, sendo mais

influenciado pelas observações de VR mais recentes.

Assim, da mesma forma que no modelo HAR utilizado, os componentes γ assimétricos do modelo 2-Comp são negativos, mas também relativamente muito pequenos, indicando que esses modelos sem γ também conseguem captar parcialmente o equilíbrio assimétrico dos retornos.

Tabela 5 Estimativas com o modelo 2-Comp

$$r_t = \rho_1 r_{t-1} + \rho_2 r_{t-2} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + e_t, e_t = \sigma_t u_t, u_t \sim NID(0,1)$$

$$\log(VR_t) = \omega + \sum_{i=1}^2 \theta_i s_{it} + \gamma e_{t-1} + \eta v_t, v_t \sim NID(0,1)$$

$$s_{it} = (1 - \alpha_i) \log(VR_{t-1}) + \alpha_i s_{i,t-1}, 0 < \alpha_i < 1, i=1,2$$

Horizonte	Ação/ Parâmetro	ρ_1	ρ_2	θ_1	θ_2	R ² Ajust.	ω	θ_1	θ_2	α_1	α_2	γ	η
1 dia	PETR4						0,071*	0,544*	0,242*	0,784	0,001	-0,017**	-0,007
							0,023	0,093	0,064			0,007	0,011
	VALE5	0,104**				0,011	0,034	0,516*	0,345*	0,408	0,904	-0,027*	-0,006
		0,042					0,022	0,076	0,107			0,008	0,011
5 dias	PETR4	0,815*				0,661	0,027*	1,505*	-0,579*	0,003	0,354	-0,005*	-0,002
		0,025					0,007	0,122	0,125			0,002	0,003
	VALE5	1,299*	-0,664*	-0,664*	0,779*	0,707	0,024*	-0,507*	1,432*	0,349	0,018	-0,003	-0,004
		0,042	0,040	0,034	0,029		0,007	0,138	0,135			0,002	0,004
22 dias	PETR4		0,908*	0,978*	-0,041	0,889	0,009*	-0,222*	1,199*	0,770	0,001	-0,001	-0,001
			0,017	0,043	0,044		0,002	0,026	0,025			0,001	0,001
	VALE5	0,228**	0,638*	0,907*	0,276*	0,905	0,006*	-0,307*	1,291*	0,628	0,001	-0,002*	-0,001
		0,119	0,114	0,118	0,043		0,002	0,034	0,033			0,000	0,001

Nota: *, ** e *** indicam a rejeição da hipótese nula com níveis de significância em 1%, 5% e 10%, respectivamente.

A Figura 2 mostra os gráficos da série histórica $s_{1,t}$ e $s_{2,t}$,⁵ gerados pela diminuição nos fatores α_1 e α_2 (gráfico 2.1) e Log(VR) realizado e estimado (gráfico 2.2) para VALE5 no horizonte temporal de um dia.

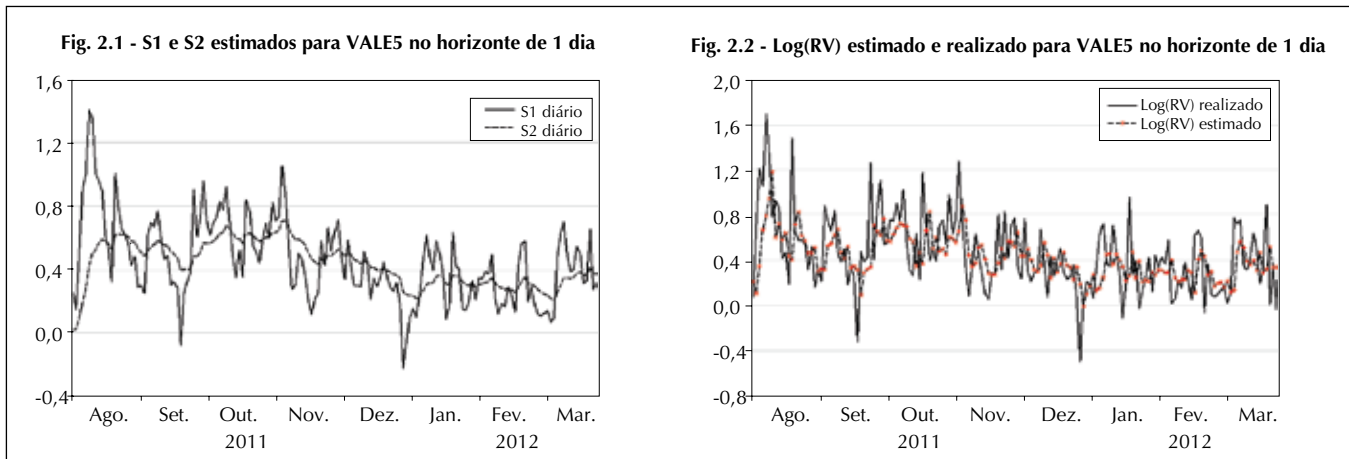


Figura 2 Fatores de decaimento e log(RV) realizado e estimado para VALE5

Para avaliar a acurácia dos modelos usados, foram avaliados para a previsão dos horizontes subsequentes de 1, 5 e 22 dias, dentro e fora da amostra, utilizando a medida da raiz quadrada do erro quadrático médio (RMSQ). O teste Modificado de Diebold Mariano foi utilizado⁵ para estimar as diferenças estatísticas entre os modelos.

Para o período dentro da amostra, os dados coletados entre 07/01/2010 e 29/7/2011 foram considerados (388 observações) e, para o período fora da amostra, os dados en-

tre 01/08/2011 e 21/03/2012 (160 observações).

A Tabela 6 apresenta a RMSQ das previsões nos três horizontes temporais definidos. O modelo 2-Comp proporcionou melhores previsões para todos os horizontes temporais, e HAR a segunda melhor previsão para os horizontes de 5 e 22 dias. Mas a qualidade estatística desses resultados é melhor que os modelos da família GARCH? Na Tabela 7, tentou-se responder essa e outras perguntas.

Tabela 6 Raiz quadrada do erro quadrático médio

Na amostra						
Horizonte	Modelo	GARCH	EGARCH	TGARCH	HAR-log(VR)	2-Comp
1 dia	PETR4	2,15	2,16	2,16	2,19	1,91
5 dias		2,88	2,89	2,87	2,56	2,34
22 dias		3,13	3,13	3,13	2,72	2,49
1 dia	VALE5	2,14	2,12	2,14	2,17	1,89
5 dias		2,76	2,74	2,76	2,49	2,26
22 dias		3,04	3,04	3,03	2,69	2,44
Fora da amostra						
Horizonte	Modelo	GARCH	EGARCH	HAR-VR	HAR-log(VR)	2-Comp
1 dia	PETR4	3,07	2,97	3,15	3,02	2,59
5 dias		4,55	4,31	4,55	3,85	3,44
22 dias		4,19	4,00	4,20	3,80	3,42
1 dia	VALE5	3,14	3,06	3,37	3,19	2,63
5 dias		3,85	3,80	3,98	3,58	3,09
22 dias		4,40	4,32	4,47	3,91	3,49

A Tabela 7 mostra os valores de p do teste estatístico Modificado de Diebold Mariano.

Com relação às previsões de retorno tanto nos pe-

ríodos dentro como fora da amostra, os valores de p mostram que: (i) o modelo 2-Comp proporciona as melhores previsões nos três horizontes temporais; (ii)

⁵ Este teste estatístico envolve testar a hipótese nula de igualdade entre o erro médio quadrático de duas previsões, utilizando-se os valores críticos de uma distribuição de Student-t com (n-1) graus de liberdade.

o modelo HAR tem a segunda melhor previsão para os horizontes de 5 e 22 dias.

Diante das previsões de retorno no período da amostra: (i) não há diferença significativa entre os modelos GARCH, EGARCH e HAR no horizonte de um dia; (ii) não há diferença significativa entre os modelos da família GARCH nos horizontes de 5 e 22 dias.

Além disso, as previsões de retorno somente no período fora da amostra mostram: (i) no horizonte de um dia: TGARCH não manteve o desempenho do EGAR-

CH *vis a vis* o similar nas previsões amostrais de ambos os modelos; GARCH e EGARCH produzem melhores previsões que o modelo HAR para VALE5; não há diferenças significativas de previsão dos modelos HAR, GARCH e EGARCH para PETR4; (ii) no horizonte de cinco dias, o modelo EGARCH gera melhores previsões que o TGARCH; (iii) no horizonte de 22 dias, o modelo EGARCH gera melhores previsões que os modelos GARCH e TGARCH.

Tabela 7 Teste Modificado Diebold Mariano (valor de P)

7.1 Na amostra							7.2 Fora da amostra							
Horizonte de 1 dia							Horizonte de 1 dia							
		GARCH	EGARCH	TGARCH	HAR-VR (log)	2-Comp			GARCH	EGARCH	TGARCH	HAR-VR (log)	2-Comp	
GARCH	PETR4	100,0%	94,4%	96,9%	67,2%	3,8%	GARCH	PETR4	100,0%	8,7%	3,3%	11,6%	0,0%	
EGARCH			100,0%	99,4%	81,0,0%	0,4%	EGARCH			100,0%	0,0%	9,9%	0,0%	
TGARCH				100,0%	77,0,0%	0,8%	TGARCH					100,0%	0,0%	0,0%
HAR-VR(log)					100,0%	100,0%	HAR-VR						100,0%	0,0%
2-Comp						100,0%	2-Comp							100,0%
GARCH	VALE5	100,0%	82,1%	96,3%	74,6%	0,2%	GARCH	VALE5	100,0%	7,5%	0,0%	2,5%	0,0%	
EGARCH			100,0%	91,6%	52,8%	0,4%	EGARCH			100,0%	0,0%	0,0%	0,0%	0,0%
TGARCH				100,0%	74,0,0%	0,0%	TGARCH					100,0%	0,0%	0,0%
HAR-VR(log)					100,0%	0,0%	HAR-VR						100,0%	0,0%
2-Comp						100,0%	2-Comp							100,0%
Horizonte de 5 dias							Horizonte de 5 dias							
GARCH	PETR4	100,0%	85,0%	95,4%	0,0%	0,0%	GARCH	PETR4	100,0%	0,0%	57,4%	0,0%	0,0%	
EGARCH			100,0%	93,2%	0,0%	0,0%	EGARCH			100,0%	0,0%	0,0%	0,0%	0,0%
TGARCH				100,0%	0,0%	0,0%	TGARCH					100,0%	0,0%	0,0%
HAR-VR(log)					100,0%	0,0%	HAR-VR(log)						100,0%	0,0%
2-Comp						100,0%	2-Comp							100,0%
GARCH	VALE5	100,0%	80,3%	93,2%	0,0%	0,0%	GARCH	VALE5	100,0%	32,6%	0,0%	0,0%	0,0%	
EGARCH			100,0%	93,0%	0,0%	0,0%	EGARCH			100,0%	0,0%	0,0%	0,0%	0,0%
TGARCH				100,0%	0,0%	0,0%	TGARCH					100,0%	0,0%	0,0%
HAR-VR(log)					100,0%	0,0%	HAR-VR(log)						100,0%	0,0%
2-Comp						100,0%	2-Comp							100,0%
Horizonte de 22 dias							Horizonte de 22 dias							
GARCH	PETR4	100,0%	95,1%	99,4%	0,0%	0,0%	GARCH	PETR4	100,0%	0,0%	91,7%	0,0%	0,0%	
EGARCH			100,0%	97,3%	0,0%	0,0%	EGARCH			100,0%	0,0%	0,0%	0,0%	0,0%
TGARCH				100,0%	0,0%	0,0%	TGARCH					100,0%	0,0%	0,0%
HAR-VR(log)					100,0%	0,0%	HAR-VR(log)						100,0%	0,0%
2-Comp						100,0%	2-Comp							100,0%
GARCH	VALE5	100,0%	93,5%	89,8%	0,0%	0,0%	GARCH	VALE5	100,0%	0,9%	1,5%	0,0%	0,0%	
EGARCH			100,0%	97,2%	0,0%	0,0%	EGARCH			100,0%	0,0%	0,0%	0,0%	0,0%
TGARCH				100,0%	0,0%	0,0%	TGARCH					100,0%	0,0%	0,0%
HAR-VR(log)					100,0%	0,0%	HAR-VR(log)						100,0%	0,0%
2-Comp						100,0%	2-Comp							100,0%

A Figura 3 representa os gráficos das duas melhores previsões fora da amostra para as ações PETR4 e VALE5 nos três horizontes. Pode-se observar que os modelos de alta frequência implementados apresen-

tam previsões com grandes semelhanças mútuas. Além disso, essas previsões parecem aderir fortemente aos retornos realizados em todos os horizontes temporais analisados.

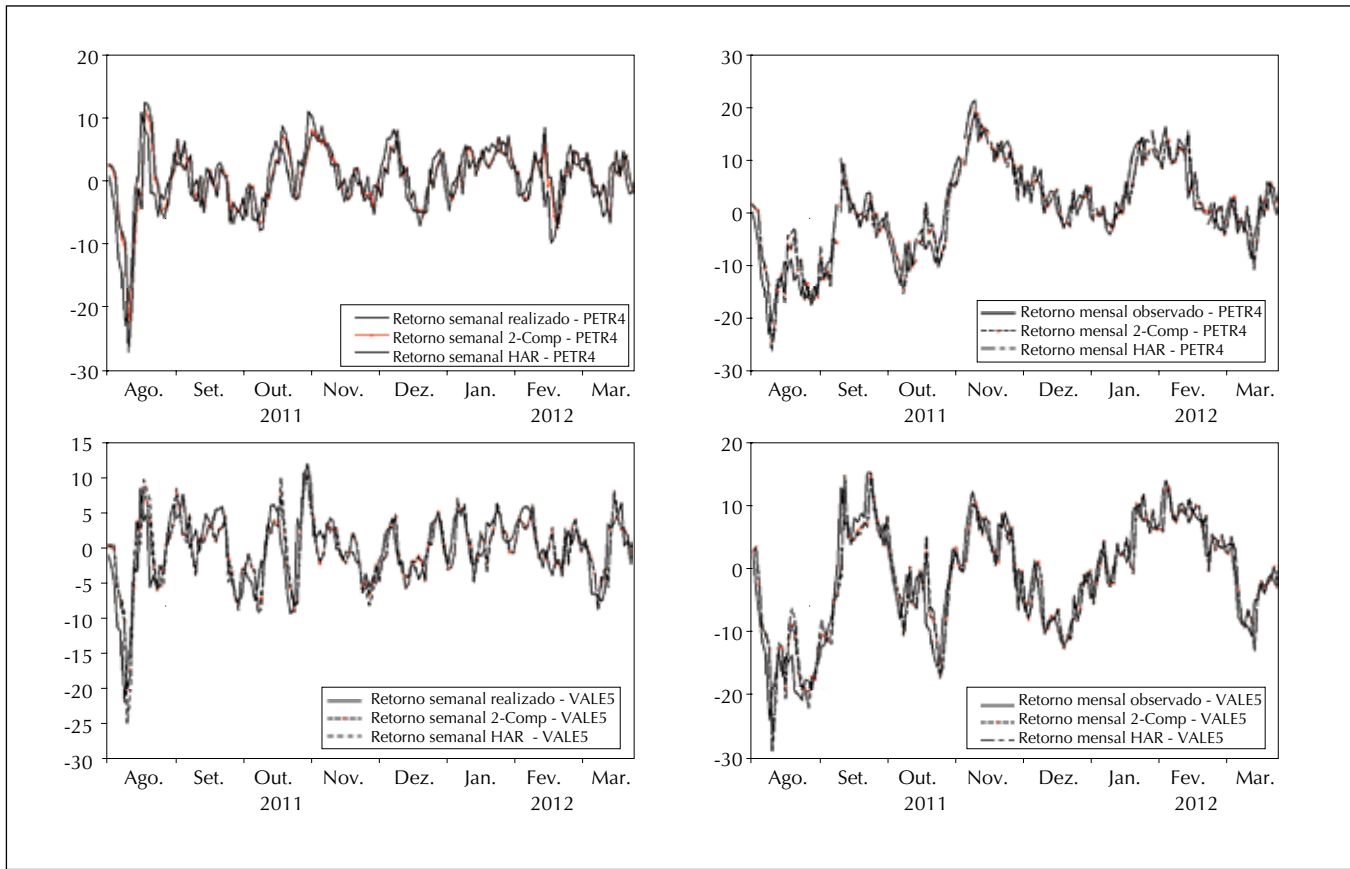


Figura 3 Retornos realizados e previsão fora da amostra pelos modelos HAR-Log(RV) e 2-Comp

Considerando os resultados do CAPM, a Figura 4 mostra o comportamento da série de retornos esperados e as séries de retornos da PETR4 e VALE5 para o horizonte de 22 dias. Há alto nível de

adesão dos retornos estimados aos retornos realizados, com correlação de 65% e 90%, respectivamente, para PETR4 e VALE5.

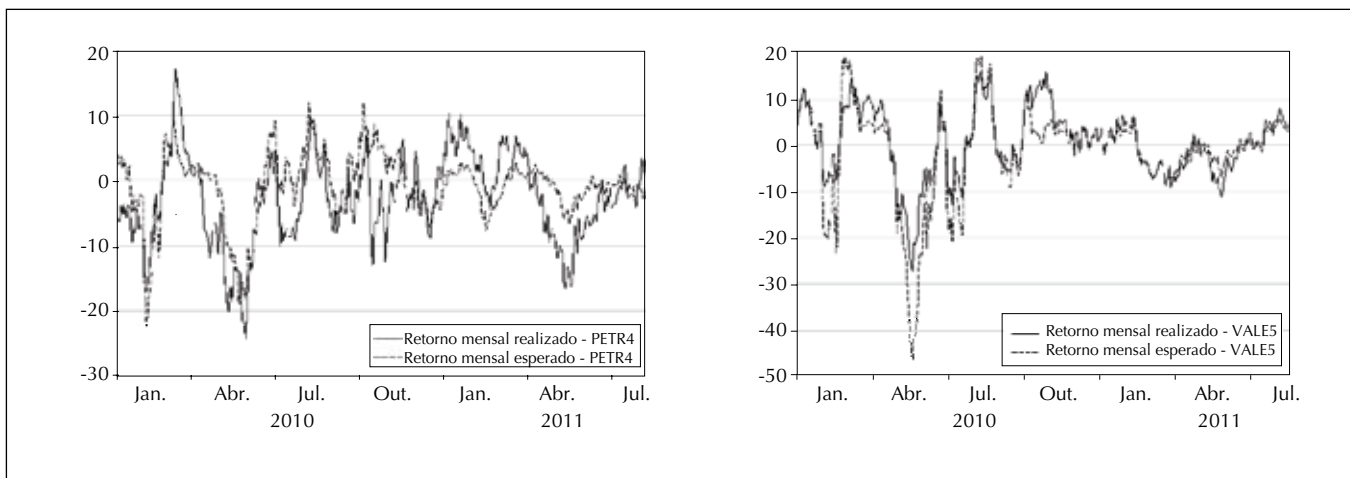


Figura 4 Retornos mensais esperados e observados para PETR4 e VALE5

A seguinte tabela resume as estatísticas descritivas para os retornos mensais esperados e observados da PETR4 e VALE5. Entre os resultados, podemos mencionar que: (i) a distribuição dos retornos esperados de ambas as ações é

assimétrica para a esquerda e mostra mais caudas gordas (mais leptocúrticas) que os retornos realizados e (iii) as distribuições dos retornos realizados e estimados não são normais no nível de significância de 5%.

Tabela 8 Estatística descritiva para retornos mensais esperados e observados

Ação	Retornos	Média	Mediana	Máximo	Mínimo	Desvio padrão	Assimetria	Curtose	Jarque-Bera
PETR4	Observado	(2,70)	(2,48)	17,15	(24,16)	6,70	(0,26)	3,29	5,93
	Esperado	(0,85)	(0,27)	11,99	(22,31)	5,32	(1,11)	5,38	170,93
VALE5	Observado	0,50	0,42	15,95	(27,30)	7,60	(0,46)	3,59	19,42
	Esperado	(1,19)	0,64	19,20	(46,36)	10,12	(1,48)	7,05	407,92

6 CONCLUSÕES

Neste artigo foram propostos modelos alternativos que usaram os retornos diários e a VR, relacionando a VR com a variância dos retornos. Além disso, buscou explorar os possíveis benefícios de se usar dados *intraday* para alcançar melhores estimativas de volatilidade e previsões de retornos.

As aplicações empíricas implementadas nos retornos da PETR4 e VALE5 desvelam a importância das informações nos retornos *intraday* e o uso do log(VR). Os resultados encontrados confirmam que: (i) modelos bivariados que usam dados de alta frequência proporcionam melhorias significativas nas previsões em comparação com os modelos padrão, baseados em dados diários, confirmando os resultados encontrados por Maheu e McCurdy (2011) na bolsa norte-americana; (ii) os dois modelos bivariados de alta frequência, de forma parcimoniosa e singular, tiveram êxito na modelagem da volatilidade, conforme apresentado por Wink Junior e Valls Pereira (2012), demonstrando desempenho excelente na previsão dos retornos e confirmando os resultados em Corsi (2009).

Esses resultados podem ser úteis em estratégias de in-

vestimentos *intraday*, em estratégias com posições longas-curtas e na gestão de riscos. As volatilidades condicionais HAR e 2-Comp podem ser usadas, por exemplo, para comparar e refinar o desempenho de diferentes metodologias de Valor em Risco.

Ao final deste artigo, também buscamos responder a pergunta: os modelos de preço de alta frequência oferecem melhores previsões de retorno que os modelos aceitos com o uso dos preços de fechamento?

Pode-se confirmar que sim. Os modelos com o uso de dados de alta frequência implementados na presente pesquisa parecem contribuir para melhores previsões de volatilidade e retorno. Esses resultados foram alcançados nos períodos dentro e fora da amostra, mediante o uso da raiz quadrada do erro quadrático médio e o teste Modificado de Diebold Mariano das previsões nos horizontes de 1, 5 e 22 dias. Apesar disso, a estimativa desses modelos para outros ativos financeiros e séries históricas maiores poderia confirmar e validar os resultados alcançados no presente estudo.

Referências

- Andersen, T. G., & Benzoni, L. (2008). *Realized volatility*. Federal Reserve Bank of Chicago, Working Paper 2008-14.
- Andersen, T. G., & Bollerslev, T. (1997). Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. *Journal of Finance*, 52 (3), 975-1005.
- Andersen, T. G., & Bollerslev, T. (1998a). Deutsche mark-dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance*, 53 (1), 219-265.
- Andersen, T. G., & Bollerslev, T. (1998b). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39 (4), 885-905.
- Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics*, 89 (4), 701-720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001a). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61 (1), 43-76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001b). The distribution of exchange rate volatility. *Journal of the American Statistical Association*, 96 (453), 42-55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71 (2), 529-626.
- Bandi, F. M., & Russell, J. R. (2004). *Microstructure noise, realized volatility, and optimal sampling*. Manuscript, University of Chicago.
- Barndorff-Nielsen, O., Hansen, P., Lunde, A., & Shephard, N. (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76 (6), 1481-1536.
- Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, 64 (2), 253-280.
- Barros, L. A., Famá, R., & Silveira, B. P. (2002). Conceito de taxa livre de risco e sua aplicação no Capital Asset Pricing Model. Um estudo exploratório para o mercado brasileiro. *Anais do Encontro Brasileiro de Finanças da Sociedade Brasileira de Finanças*, Rio de Janeiro, RJ, Brasil, 2.
- Black, F., Jensen M. C., & Scholes M. (1972). The capital asset pricing model: some empirical tests. In M. Jensen (Ed.). *Studies in the theory of capital markets*. (pp. 79-121). New York: Praeger Publishers.
- Corsi, F. (2009). A simple approximate long memory model of realized volatility. *Journal of Financial Econometrics*, 7 (2), 174-196.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13 (3), 252-263.
- Ghysels, E., Santa-Clara, P., & Valkanov, R. (2005). There is a risk-return tradeoff after all. *Journal of Financial Economics*, 76 (3), 509-548.
- Giot, P., & Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, 11 (3), 379-398.
- Goodhart, C. A. E., & O'Hara, M. (1997). High frequency data in financial markets: issues and applications. *Journal of Empirical Finance*, 4 (2), 73-114.
- Hansen, P. R., Large, J. H., & Lunde, A. (2008). Moving average-based estimators of integrated variance. *Econometric Reviews*, 27 (1-3), 79-111.
- Harvey, D., Leybourne S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13 (2), 281-291.
- Hsieh, D. A. (1991). Chaos and nonlinear dynamics: application to financial markets. *Journal of Finance*, 46 (5), 1839-1877.
- Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47 (1), 13-37.
- Maheu, J. M., & McCurdy, T. H. (2007). Components of market risk and return. *Journal of Financial Econometrics*, 5 (4), 560-590.
- Maheu, J. M., & McCurdy, T. H. (2011). Do high-frequency measures of volatility improve forecasts of returns distributions? *Journal of Econometrics*, 160 (1), 69-76.
- Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17 (5), 479-508.
- Moreira, J. M. S., & Lemgruber, E. F. (2004). O uso de dados de alta frequência na estimação da volatilidade e do valor em risco para o IBOVESPA. *Rev. Bras. Econ.*, Rio de Janeiro, 58 (1), 99-120.
- Sharpe, W. F. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, 19 (3), 425-442.
- Treynor, J. L. (1961). Market value, time, and risk. Unpublished manuscript.
- Tsay, R. S. (2010). *Analysis of financial time series*. (3rd ed.). New Jersey: Wiley.
- Wink Junior, M. V., & Valls Pereira, P. L. (2012). *Modelagem e previsão de volatilidade realizada: evidências para o Brasil*. Texto para Discussão 313, 3, FGV-EESP, CEQEF.
- Zhou, B. (1996). High-frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics*, 14 (1), 45-52.