

# Modelagem e Previsão de Volatilidade Realizada: Evidências para o Brasil

Marcos Vinício Wink Junior\*

Pedro Luiz Valls Pereira†

JEL Classification: C22,C53,C58

Área 7. Microeconomia, Métodos Quantitativos e Finanças.

18 de julho de 2011

## Resumo

Usando dados intradiários dos ativos mais negociados do BOVESPA, este trabalho considerou dois modelos recentemente desenvolvidos na literatura de estimação e previsão de volatilidade realizada. São eles; *Heterogeneous Autorregressive Model of Realized Volatility* (HAR-RV), desenvolvido por Corsi (2009) e o *Mixed Data Sampling* (MIDAS-RV), desenvolvido por Ghysels et al. (2004). Através de medidas de comparação de previsão dentro e fora da amostra, constatou-se resultados superiores do modelo MIDAS-RV apenas para previsões dentro da amostra. Para previsões fora da amostra, no entanto, não houve diferença estatisticamente significativa entre os modelos. Também encontram-se evidências que a utilização da volatilidade realizada induz distribuições dos retornos padronizados mais próximas da normal.

**Palavras-chave:** Volatilidade Realizada. HAR. MIDAS. Dados financeiros intradiários

## Abstract

Using intraday data for the most actively traded stocks of BOVESPA, this work has considered two recently developed models in the literature of the estimation and forecasting of realized volatility; The Heterogeneous Autorregressive Model of Realized Volatility (HAR-RV), developed by Corsi (2009) and the Mixed Data Sampling (MIDAS-RV), developed by Ghysels et al. (2004). Through statistical comparison of forecasts in-sample and out-of-sample, it was found that superior results of the MIDAS-RV model occurred only for the in-sample forecasting. However, for out-of-sample forecasts no statistically different results were found between the models. Also, there are evidences that the use of realized volatility induces normality in standardized returns.

**Keywords:** Realized Volatility. HAR. MIDAS. High frequency financial data.

---

\*Fundação de Economia e Estatística (FEE-RS) e Escola de Economia de São Paulo (FGV)

†Professor da Escola de Economia de São Paulo (FGV)

# Introdução

A tentativa de modelar volatilidade de retorno de ativos é uma literatura que cresceu muito nos últimos anos. Ultimamente, os modelos que estimam volatilidade realizada com dados intradiários são os mais freqüentes na literatura. Dentro desses modelos, surgiram hipóteses a respeito do comportamento dos agentes e a tentativa de verificar fatos estilizados a respeito de séries financeiras. Acredita-se que os retornos padronizados pela volatilidade realizada de ativos financeiros têm distribuições próximas da Normal, o que é muito relevante para modelagem de risco e utilização de modelos de Valor em Risco (V@R). Este trabalho se propõe a comparar modelos de estimação de volatilidade realizada de cinco ativos presentes na Bolsa de Valores de São Paulo (BOVESPA). Os métodos que serão considerados nesse trabalho são *Heterogeneous Autorregressive Model of Realized Volatility* (HAR-RV), desenvolvido por Corsi (2009) e o *Mixed Data Sampling* (MIDAS-RV), desenvolvido por Ghysels et al. (2004).

Para fazermos as estatísticas de comparação da capacidade de ajuste e de previsão dos modelos, utilizaremos medidas de comparação dentro e fora da amostra. Para previsões dentro da amostra, comparamos os  $R^2$  ajustados e os erros quadráticos médios (EQM) de dentro da amostra. Para comparações fora da amostra, utilizaremos os EQM fora da amostra e o teste de Diebold-Mariano modificado. Este trabalho, portanto, se justifica na medida em que contribui com a literatura de previsão de volatilidade realizada com dados de alta freqüência para países emergentes, assim como o estudo de Chung et al. (2008) para o mercado de Taiwan e Carvalho et al. (2006) para o Brasil.

Conforme afirma Carvalho et al. (2006), a previsão de retornos acionários é bastante difícil de ser feita. Modelagem de volatilidade de retornos, no entanto, é mais fácil de ser feita e conseqüentemente gera previsões melhores. Dessa forma, a literatura de modelagem de volatilidade cresceu muito nos últimos anos.

A tentativa de modelar volatilidade começou principalmente através da estimação de modelos da classe ARCH-GARCH e através de modelos de volatilidade estocástica. No entanto, como afirma Corsi (2009), esses modelos sofrem de uma “dupla fraqueza”. Primeiro, eles não conseguem representar características empíricas específicas de dados financeiros e, segundo, também tendem a ter estimações não triviais, principalmente quando tratamos de modelos de volatilidade estocástica.

A partir disso, o próximo passo da literatura foi tentar calcular volatilidade através do quadrado do retorno diário do ativo. Dessa forma, a estimação da volatilidade se daria a partir de uma variável observada que seria a proxy para a verdadeira volatilidade, que é uma variável latente. Andersen e Bollerslev (1998), no entanto, verificaram que a estimação da volatilidade através de dados diários gerava problema em função do grande ruído associado a essas séries. Dessa forma, os autores sugeriram que dados intradiários trariam mais precisão na estimação da volatilidade. A estimação da volatilidade através de dados intradiários ficou conhecida na literatura como cálculo de volatilidade realizada. Os métodos de estimação de volatilidade realizada que serão considerados nesse trabalho são *Heterogeneous Autorregressive Model of Realized Volatility* (HAR-RV), desenvolvido por Corsi (2009) e o *Mixed Data Sampling* (MIDAS), desenvolvido por Ghysels et al. (2004).

O sucesso do método HAR-RV de estimação de volatilidade realizada baseia-se em modelar comportamentos de volatilidade de memória longa de uma maneira parcimoniosa e simples. Este modelo foi inspirado pela Hipótese de Mercados Heterogêneos e também

pela assimetria da propagação da volatilidade. Corsi (2009) afirma que em mercados heterogêneos existem diferentes tipos de agentes e que estes se contentam com preços diferentes e, dessa forma, decidem realizar transações em momentos diferentes. Para o desenvolvimento do modelo de estimação de volatilidade realizada, Corsi (2009) considera apenas como fonte de heterogeneidade os diferentes horizontes de tempo das transações.

A literatura de séries temporais desenvolve normalmente modelos envolvendo amostras com frequências iguais. O modelo MIDAS, desenvolvido por Ghysels et al. (2004), no entanto, considera diferentes níveis de frequências amostrais. O modelo MIDAS, como afirma Ghysels et al. (2004) tem como virtude uma estimação parcimoniosa com um reduzido número de parâmetros a serem estimados. Os regressores MIDAS também têm, conforme demonstrado em Ghysels et al. (2004), propriedades desejáveis como eficiência e ausência de viés quando a amostra tende a ser mais freqüente.

A utilização de dados de alta frequência também tem uma desvantagem. Andersen et al. (2007) afirmam que a extração do estimador da volatilidade realizada diária é viesado em virtude do alto ruído associado a microestrutura de mercado verificado em dados intradiários. Esse ruído, como afirmam os autores, está associado principalmente ao fato de que os preços observados não são contínuos e, mais, na verdade os preços de mercado são cotados em um *grid* de preços discretos. Dessa forma, como será discutido posteriormente, o preço intradiário observado não é um preço único de mercado em um preciso instante do tempo, mas sim um preço com um ruído associado a microestrutura de mercado. Andersen et al. (2007), ainda, discutem algumas prováveis fontes de microestrutura de mercado. Segundo os autores, as fontes mais freqüentes são: a diferença entre preços de compradores e vendedores (*bid-ask spread*), diferentes precificações entre diferentes participantes dos mercados devido a crenças, informações ou posição de compra ou venda.

Já houve no Brasil trabalhos que tentaram modelar a volatilidade dos retornos de ativos do Bovespa. Destacam-se Carvalho et al. (2006) e de Sá Mota e Fernandes (2004). Estes trabalhos, de maneira geral, compararam previsões de modelos de volatilidade realizada com modelos da classe GARCH. Nenhum desses trabalho, no entanto, utilizou alguma forma de controlar os efeitos da microestrutura de mercado. Para controlar esse possível viés, este trabalho utiliza o processo de filtragem dos dados como sugerido por Hansen et al. (2008).

Além desta introdução, o trabalho apresenta no capítulo 2 o referencial teórico sobre variância realizada, no capítulo 3 a descrição dos modelos utilizados, no capítulo 4 as estatísticas descritivas dos dados utilizados, no capítulo 5 os resultados das estimações e, por fim, a conclusão do trabalho.

## 1 Variância Realizada

Segundo McAleer e Medeiros (2008) e Andersen et al. (2003), suponha que o processo do logaritmo do preço  $p_t$  de um determinado ativo em tempo contínuo segue a seguinte difusão:

$$dp_t = \sigma_t dW_t, t = 1, 2, \dots, \quad (1.1)$$

onde  $p_t$  é o logaritmo do preço no tempo  $t$ ,  $\sigma_t$  é a volatilidade instantânea, estritamente estacionária e  $W_t$  é um movimento Browniano padrão. Assumimos que o processo descrito em (1.1) não contém um componente de arrastamento (*drift*).

Os retornos amostrais com  $M$  observações por período podem ser calculados através de:

$$r(m), t = p_t - p_{t-1/M} = \int_0^{1/M} \sigma_{t-1/(M+\tau)} dW_{t-1/(M+\tau)}, t = \frac{1}{M}, \frac{2}{M}, \dots \quad (1.2)$$

Utilizando a definição de esperança dos retornos iguais a zero para qualquer horizonte de tempo e padronizando o intervalo de tempo por  $M$  observações intradiárias e, ainda, assumindo  $W_t$  e  $\sigma_t$  independentes e condicionando a esperança matemática na trajetória amostral da volatilidade  $\{\sigma_{t+\tau}\}_{\tau=0}^h$  temos que a variância do retorno para  $h$  períodos pode ser escrita como:

$$\sigma_{t,h}^2 = \int_0^h \sigma_{t+\tau}^2 d\tau \quad (1.3)$$

A equação (1.3) descreve a chamada variância integrada (VI), que é uma medida de volatilidade *ex post*. Em outras palavras, a volatilidade para  $h$  períodos é a integral das volatilidades intradiárias passadas. A VI, no entanto, não é observada e por isso se torna o objetivo de interesse e precisa ser estimada. Define-se o retorno intradiário no período  $m$  e dia  $t$  por:

$$r_{t,m} = p_{t,m} - p_{t,m-1}, m = 1, \dots, M, t = 1 \dots n \quad (1.4)$$

A variância realizada (VR) diária é definida como:

$$VR_t = \sum_{m=1}^n r_{t,m}^2 \quad (1.5)$$

Andersen et al. (2003) demonstraram sob algumas condições de ausência de autocorrelação dos retornos, que a variância realizada definida na equação (1.5) é um estimador consistente da variância integrada, já que  $VR_t \xrightarrow{p} VI_t$ . Volatilidade realizada (RV) é a raiz quadrada da variância realizada. Barndorff-Nielsen et al. (2002) derivaram a distribuição assintótica do estimador da variância integrada como:

$$\sqrt{n} \frac{1}{\sqrt{2IQ_t}} (VR_t - VI_t) \xrightarrow{d} N(0, 1), \quad (1.6)$$

onde  $IQ_t$  é a *integrated quarticity* e é definida como:

$$IQ_t = \int_0^1 \sigma^4(t + \tau) d\tau \quad (1.7)$$

Sob a hipótese de ausência de correlação dos retornos intradiários,  $IQ_t$  pode ser consistentemente estimada pela *realized quarticity*, definida como:

$$RQ_t = \frac{1}{3} \sum_{i=1}^n r_{i,t}^4 \quad (1.8)$$

## 2 HAR-RV e MIDAS

Os modelos considerados aqui serão o HAR-RV e MIDAS. O modelo HAR-RV proposto por Corsi (2009), propõe um método de estimar volatilidade considerando diferentes tamanhos de intervalos. Defini-se  $\tilde{\sigma}_t^{(\cdot)}$  como a volatilidade parcial gerada por certos componentes de mercado. O modelo é proposto através da adição de volatilidades parciais em cascatas, onde elas seguem uma espécie de processo autorregressivo. Simplificando, considere 3 diferentes componentes de volatilidade geradas por diferentes horizontes de tempo, mais especificamente nesse exemplo:  $\tilde{\sigma}_t^{(d)}$ , referente a um dia,  $\tilde{\sigma}_t^{(w)}$ , referente a uma semana e, por fim,  $\tilde{\sigma}_t^{(m)}$ , referente a um mês. Suponha que o retorno diário seja dado por:

$$r_t = \sigma_t^{(d)} \varepsilon_t \quad (2.1)$$

com  $\varepsilon_t \sim \text{NID}(0,1)$  e  $\sigma_t^{(d)}$  sendo a volatilidade integrada diária.

O processo de volatilidade parcial  $\tilde{\sigma}_t^{(\cdot)}$  em cada escala de tempo é função da volatilidade parcial passada, da mesma escala de tempo, e da esperança de valores do próximo período de mais longo prazo. Como a maior escala de tempo é a mensal, o modelo segue:

$$\tilde{\sigma}_{t+1m}^{(m)} = c^{(m)} + \phi_m RV_t^m + \tilde{w}_{t+1m}^m \quad (2.2)$$

$$\tilde{\sigma}_{t+1w}^{(w)} = c^{(w)} + \phi_w RV_t^w + \gamma E_t[\tilde{\sigma}_{t+1m}^{(m)}] + \tilde{w}_{t+1w}^w \quad (2.3)$$

$$\tilde{\sigma}_{t+1d}^{(d)} = c^{(d)} + \phi_d RV_t^d + \gamma E_t[\tilde{\sigma}_{t+1w}^{(w)}] + \tilde{w}_{t+1d}^d, \quad (2.4)$$

onde  $RV_t^m, RV_t^w$  e  $RV_t^d$  são as volatilidades realizadas mensais, semanais e diárias observadas ex post.  $\tilde{w}_{t+1m}^m, \tilde{w}_{t+1w}^w$  e  $\tilde{w}_{t+1d}^d$ , são inovações na volatilidade que são variáveis contemporaneamente e serialmente não correlacionados.

Através de substituições recursivas de volatilidades parciais, o modelo pode ser escrito como:

$$\sigma_{t+1d}^d = c^{(m)} + \beta^d RV_t^d + \beta^w RV_t^w + \beta^m RV_t^m + \tilde{w}_{t+1d}^d \quad (2.5)$$

Podemos escrever a volatilidade parcial ex post como:

$$\sigma_{t+1d}^d = RV_{t+1}^d + w_{t+1d}^d, \quad (2.6)$$

onde  $w_{t+1d}^d$  representa o erro de estimação da volatilidade diária latente.

Dessa forma, substituindo a equação (2.6) na (2.5), encontramos

$$RV_{t+1d}^d = c + \beta^d RV_t^d + \beta^w RV_t^w + \beta^m RV_t^m + w_{t+1d}^d, \quad (2.7)$$

onde  $w_{t+1d}^d = \tilde{w}_{t+1d}^d - w_{t+1d}^d$

A equação (2.7), portanto, descreve a volatilidade realizada através de um processo autorregressivo simples. Como utilizado em Corsi (2009),  $RV_t^m$ , corresponde a média da volatilidade do mês. O mesmo vale para a semana em  $RV_t^w$ .

O modelo HAR-RV de volatilidade de um dia pode ser estendido facilmente, como mostrado em Andersen et al. (2007) e em Forsberg e Ghysels (2006), para  $h$  horizontes de tempo, indexados em  $t$ ,  $RV_{t,t+h}$ . Definimos a volatilidade realizada para multi-períodos:

$$RV_{t,t+h} = h^{-1}(RV_{t,t+1} + RV_{t+1,t+2} + \dots + RV_{t+h-1,t+h}) \quad (2.8)$$

onde  $RV_{t,t+h}$  refere-se ao incremento na  $RV$  de  $t$ , para  $t+h$  períodos, com  $h = 1, 5, 10, 15$  e  $20$ , indicando, respectivamente, um dia, uma semana, duas semanas, três semanas e um mês. Dessa forma, o modelo HAR-RV para multi-períodos é dado por:

$$RV_{t,t+h} = c + \beta^d RV_t^d + \beta^w RV_t^w + \beta^m RV_t^m + w_{t,t+h} \quad (2.9)$$

Corsi (2009) ainda sugere que o modelo seja estimado na sua forma logarítmica, devido a distribuição lognormal da volatilidade realizada. Dessa forma, o modelo estimado nesse trabalho é descrito pela equação (2.10).

$$\ln RV_{t,t+h} = c + \beta^d \ln RV_t^d + \beta^w \ln RV_t^w + \beta^m \ln RV_t^m + w_{t,t+h} \quad (2.10)$$

Os modelos MIDAS com estruturas lag polinomial, introduzido por Ghysels et al. (2007) envolvem regressores com diferentes frequências, assim ele não é um modelo autor-regressivo. No modelo MIDAS, portanto, a variável dependente  $Y_t$  tem uma frequência fixa (anual, trimestral, mensal ou diária) chamada de intervalo de referência. Seja  $X_t^{(m)}$  uma amostra de  $m$  intervalos de tempo, temos, por exemplo, com dados anuais, que  $X_t^{(4)}$  diz respeito à amostra com dados trimestrais. Nesse sentido, uma regressão MIDAS é:

$$Y_t = \alpha + B(L^{1/m})X_t^{(m)} + \varepsilon_t^{(m)}, \quad (2.11)$$

onde  $B(L)$  é um operador de defasagens finito, ou infinito, geralmente parametrizado como um pequeno conjunto de hiperparâmetros<sup>1</sup>.  $B(L^{1/m}) = \sum_{j=0}^{j^{max}} B(j)L^{j/m}$  é um polinômio de grau  $j^{max}$  (possivelmente infinito) no operador  $L^{1/m}$  e  $L^{j/m}X_t^{(m)} = X_{t-j/m}^{(m)}$ . O operador  $L^{1/m}$ , dessa forma, produz valores de  $X_t^{(m)}$  defasados  $j/m$  períodos.

Seguindo Forsberg e Ghysels (2006) e Chung et al. (2008), escrevemos uma especificação particular do modelo MIDAS-RV, baseado na função Beta, para  $h$  horizontes de tempo.

$$RV_{t,t+h} = \alpha + \phi \sum_{k=0}^{k=50} b(k, \theta_1, \theta_2) RV_{t-k-1,t-k} + \varepsilon_{t,t+h} \quad (2.12)$$

$$b(k, \theta_1, \theta_2) = \frac{f(k/50; \theta_1, \theta_2)}{\sum_{k=1}^{k=50} f(k/50; \theta_1, \theta_2)} \quad (2.13)$$

$$f(x; \theta_1, \theta_2) = \frac{x^{\theta_1-1}(1-x)^{\theta_2-1}}{\beta(\theta_1, \theta_2)}, \quad (2.14)$$

onde  $\beta(\theta_1, \theta_2)$  é a função beta, isto é,  $\beta(\theta_1, \theta_2) = \Gamma(\theta_1)\Gamma(\theta_2)/\Gamma(\theta_1 + \theta_2)$  e  $f(x; \theta_1, \theta_2)$  é a função de densidade de probabilidade da distribuição beta. A função beta que define os 50<sup>2</sup> pesos que determinam a memória do processo de volatilidade realizada. Os parâmetros  $\theta_1$  e  $\theta_2$  que determinam a forma da função de pesos. Quanto maior for  $\theta_2$ , mais rapidamente a função declina. O parâmetro  $\theta_1$  define a trajetória inicial da função, assim, se  $\theta_1 > 1$ ,

<sup>1</sup>Definição que surge do modelo estilizado de defasagens distribuídos

<sup>2</sup>A escolha de  $k = 50$  é sugerida por Ghysels (2006). Segundo os autores, os resultados produzidos por  $k > 50$  são bastante semelhantes.

a função tem um salto antes de começar a decair<sup>3</sup>. Como processos com memória longa, em geral, dependem mais das observações quanto mais recentes elas forem, geralmente o que ocorre é  $\theta_1 \simeq 1$  e  $\theta_2 > 1$ .

### 3 Dados

O trabalho utiliza dados dos 5 ativos mais líquidos, no período, presentes na bolsa de valores de São Paulo (BOVESPA). São eles: Bradesco (BBDC4), Petrobrás (PETR4), Vale do Rio Doce (VALE5), Telemar (TNLP4) e Usiminas (USIM5). A amostra abrange dados intradiários em 3 janelas diferentes: intervalos de 5, 15 e 30 minutos. O período analisado é de 01/11/2007 até 30/04/2010.

Como discutido em McAleer e Medeiros (2008), existe um debate na literatura a respeito da escolha da frequência dos dados intradiários. Quando maior a frequência (M grande), maior a precisão, no entanto, são maiores as chances de haver ruído associado a microestrutura, como por exemplo, ausência de negociação. Andersen et al. (2001) propõe intervalos de 5 minutos, Oomen (2002) argumenta que a frequência ótima é dada em intervalos de 25 minutos. Já Giot e Laurent (2004) encontram como frequência ótima, dados de 15 em 15 minutos. Diante deste debate da literatura, optou-se neste trabalho por utilizar dados em intervalos de 5 minutos, 15 minutos e 30 minutos. A tabela 1 mostra o número de observações para cada ativo em cada janela intradiária. Há diferentes números de observações porque obtemos o preço do ativo negociado imediatamente após a janela. Em outras palavras, estamos obtendo o preço da primeira negociação após a janela determinada e, dessa forma, ativos mais líquidos terão mais observações já que tenderão a ter negociação logo após o fechamento da janela.

Tabela 1: Número de Observações

Ativo	Janela de 5 min	Janela de 15 min	Janela de 30 min
<b>BBDC4</b>	41.900	15.663	8.330
<b>PETR4</b>	45.858	16.142	8.456
<b>VALE5</b>	45.532	16.138	8.458
<b>TNLP4</b>	32.800	14.407	7.947
<b>USIM5</b>	40.413	15.475	8.280

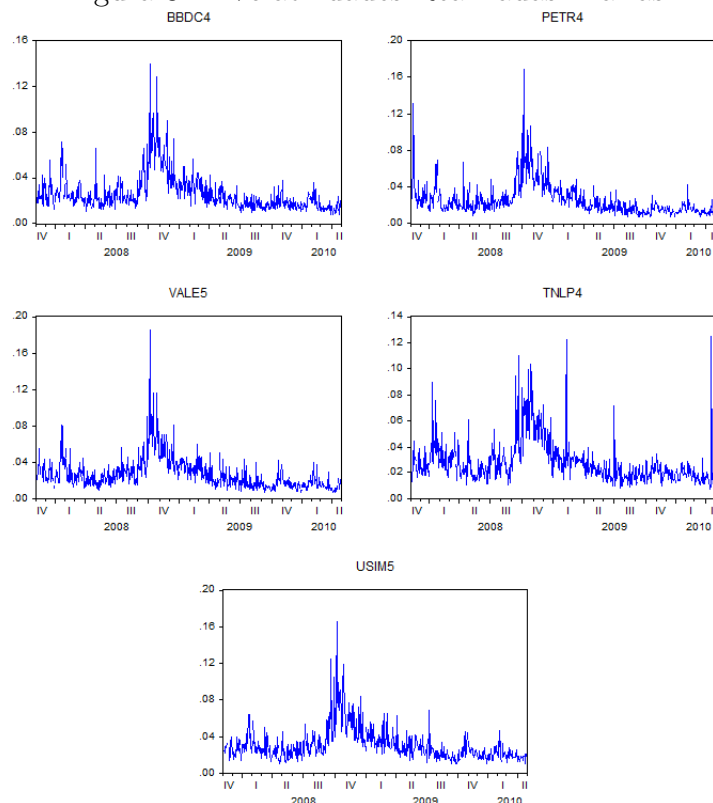
Fonte: BM&F BOVESPA

A figura 3.1 mostra o gráfico da volatilidade realizada diária, utilizando janelas de 5 minutos, para cada ativo. Percebe-se um aumento da volatilidade no segundo semestre de 2008. Tal fato ocorreu devido a crise do *subprime*. Observa-se também saltos de volatilidade em todos os ativos<sup>4</sup>

<sup>3</sup>Para mais detalhes ver Ghysels et al. (2007)

<sup>4</sup>Andersen et al. (2007) afirmam que a maioria dos saltos de volatilidade ocorrem devido a anúncios macroeconômicos.

Figura 3.1: Volatilidades Realizadas Diárias



Fonte: BM&F/Bovespa

A tabela 2 apresenta as estatísticas descritivas dos retornos diários dos 5 ativos selecionados. Como usual na literatura, os retornos apresentam excesso de curtose, implicando que os retornos não tenham distribuições normais, como pode ser verificado no teste de Jarque-Bera. Outro fato relevante é que todos os ativos apresentaram médias de retornos negativas. Tal fato ocorreu, principalmente, porque a amostra utilizada engloba o período caracterizado pela crise do *subprime*.

Tabela 2: Estatísticas Descritivas dos Retornos Diários

Ativo	Média	Desvio Padrão	Assimetria	Curtose	Jarque-Bera
<b>BBDC4</b>	-0,00015	0,029	0,678	8,484	0,00
<b>PETR4</b>	-0,00014	0,031	-0,034	6,322	0,00
<b>VALE5</b>	-0,00023	0,032	-0,153	6,284	0,00
<b>TNLP4</b>	-0,00061	0,031	0,470	7,541	0,00
<b>USIM5</b>	-0,00012	0,035	0,008	6,271	0,00

Fonte: Elaboração Própria

Nota: O valor do Jarque-Bera corresponde ao p-valor do teste.

Para escolher o estimador da volatilidade realizada mais preciso, dentro das diferentes frequências, construímos o tamanho médio do intervalo de confiança de 95% do estimador, calculado a partir da sua variância descrita na equação (1.6). Os resultados são apresentados na tabela 3. Os resultados sugerem que o menor tamanho de intervalo de



confiança é gerado pela frequência de 5 minutos para todos os ativos. Dessa forma, daqui para frente utilizaremos no trabalho dados intradiários com frequência de 5 minutos.

Tabela 3: Tamanho Médio do Intervalo de Confiança da Volatilidade Realizada Diária

Ativo	Janela de 5 min	Janela de 15 min	Janela de 30 min
<b>BBDC4</b>	0,00016	0,00045	0,00044
<b>PETR4</b>	0,00016	0,00030	0,00047
<b>VALE5</b>	0,00019	0,00032	0,00049
<b>TNLP4</b>	0,00021	0,00034	0,00046
<b>USIM5</b>	0,00023	0,00039	0,00059

Fonte: Elaboração Própria

A tabela 4 mostra as estatísticas descritivas dos retornos padronizados. As volatilidades diárias foram estimadas por 5 métodos diferentes: GARCH, EGARCH, EWMA (utilizando o parâmetro de decaimento estimado), GJR e volatilidade realizada. A ordem dos modelos e as distribuições condicionais dos erros foram escolhidas utilizando o critério de informação de Schwarz.

Os resultados descritos na tabela 4 sugerem que a estimação da volatilidade realizada induz distribuições dos retornos padronizados mais próximas da distribuição normal em relação a estimação da volatilidade diária por outros métodos. Tal resultado é muito útil na utilização de gerenciamento de risco através de modelos de Valor em Risco (V@R). Outro resultado importante presente na tabela 5 diz respeito a superioridade dos modelos assimétricos de volatilidade (EGARCH e GJR) sobre modelos GARCH para gerar retornos padronizados com distribuições normais<sup>5</sup>.

<sup>5</sup>Tal fato já havia sido encontrado para retornos de ativos brasileiros em Carvalho et al. (2006).

Tabela 4: Estatísticas Descritivas dos Retornos Padronizados Diários

Ativo	Média	Desvio Padrão	Assimetria	Curtose	Jarque-Bera
<b>GARCH</b>					
BBDC4	0,008	1,003	0,216	3,824	0,000
PETR4	-0,007	0,998	-0,140	3,351	0,077
VALE5	0,006	1,000	-0,224	4,380	0,000
TNLP4	-0,013	1,010	-0,412	7,781	0,000
USIM5	0,010	1,002	0,078	3,408	0,088
<b>EGARCH</b>					
BBDC4	-0,002	1,000	0,096	3,352	0,130
PETR4	-0,005	1,000	-0,130	3,236	0,207
VALE5	0,06	1,000	-0,224	4,380	0,000
TNLP4	-0,020	1,019	-0,602	1,001	0,000
USIM5	0,010	1,003	0,084	4,380	0,093
<b>EWMA</b>					
BBDC4	0,010	1,038	0,205	3,792	0,000
PETR4	-0,008	1,036	-0,160	3,527	0,008
VALE5	0,007	1,047	-0,338	4,944	0,000
TNLP4	-0,016	1,069	-0,806	1,145	0,000
USIM5	0,016	1,047	0,120	3,508	0,018
<b>GJR</b>					
BBDC4	0,000	1,001	0,103	3,387	0,087
PETR4	-0,007	1,000	-0,123	3,301	0,146
VALE5	0,002	1,000	-0,211	4,067	0,000
TNLP4	-0,019	0,909	-0,407	7,478	0,000
USIM5	0,011	1,001	0,056	3,230	0,437
<b>VOLATILIDADE REALIZADA</b>					
BBDC4	0,024	0,949	0,140	2,601	0,049
PETR4	0,050	1,047	-0,015	2,729	0,388
VALE5	0,055	01,040	0,016	2,588	0,114
TNLP4	0,021	0,964	0,189	2,742	0,070
USIM5	0,031	1,028	0,039	2,500	0,038

Fonte: Elaboração Própria

Nota: O valor do Jarque-Bera corresponde ao p-valor do teste.

### 3.1 Filtro

Como já discutido, a variância realizada é um estimador consistente da variância integrada se os retornos forem não autocorrelacionados. Andersen et al. (2001), no entanto, mostram como a microestrutura de mercado gera autocorrelação dos retornos. Para mostrar isso

suponhamos que o logarítmo do preço observado de um ativo  $p_{t,i}$  seja dado por:

$$p_{t,i} = p_{t,i}^* + \epsilon_{t,i}, \quad (3.1)$$

onde  $p_{t,i}^*$  é o preço latente eficiente e  $\epsilon_{t,i}$  é o ruído associado a microestrutura de mercado. O retorno do ativo  $r_{t,i}$  será dado por:

$$r_{t,i} = p_{t,i} - p_{t-1,i} = r_{t,i}^* + \epsilon_{t,i} - \epsilon_{t-1,i}, \quad (3.2)$$

onde  $r_{t,i}^* = p_{t,i}^* - p_{t-1,i}^*$  é o retorno eficiente.

Dessa forma, fica claro como o ruído de microestrutura gera autocorrelação dos retornos e, conseqüentemente, gera uma estimação viesada da variância integrada. Para corrigir o viés causado pela microestrutura de mercado, utilizamos a volatilidade realizada diária filtrada pelo método de médias móveis (filtro MA). O filtro MA foi utilizado por Ebens (1999) e Andersen et al. (2001). Os autores sugeriram a utilização de um processo médias móveis de primeira ordem para modelar os retornos intradiários dos ativos e a filtragem decorreria da utilização dos resíduos dessa estimação. Hansen et al. (2008), no entanto, mostraram que o filtro MA(1) seria válido apenas sob a hipótese de que o ruído associado a microestrutura de mercado seja *iid*. Hansen et al. (2008), portanto, proporam um método de filtragem cuja hipótese de *iid* do ruído não seja necessária.

Suponha que o retorno intradiário de um ativo siga um processo 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 sob algumas condições o estimador da VR baseado no processo MA(q),  $VR_{t,MA(q)}$  é:

$$VR_{t,MA(q)} = (1 - \hat{\theta}_1 - \dots - \hat{\theta}_m)^2 \sum_{m=1}^n \hat{\varepsilon}_{t,m}^2 \quad (3.3)$$

Temos que:

$$E[VR_t] = E\left[\sum_{m=1}^n r_{t,m}^2\right] = (1 + \theta_1^2 + \dots + \theta_q^2) E\left[\sum_{m=1}^n \varepsilon_{t,m}^2\right] \quad (3.4)$$

Portanto

$$\frac{E[VR_t]}{VI} \simeq \frac{1 + \theta_1^2 + \dots + \theta_q^2}{(1 - \theta_1 - \dots - \theta_q)^2} \quad (3.5)$$

A equação (3.5) sugere um estimador corrigido de viés para a variância realizada,  $\widetilde{VR}_{t,MA(q)}$ , dado por:

$$\widetilde{VR}_{t,MA(q)} = \frac{(1 - \hat{\theta}_1 - \dots - \hat{\theta}_q)^2}{1 + \hat{\theta}_1^2 + \dots + \hat{\theta}_q^2} VR_t \quad (3.6)$$

As estimações feitas nesse trabalho foram realizadas a partir da volatilidade realizada filtrada pelo método MA(q) descrito na equação (3.6).

## 4 Resultados das Estimações

Neste capítulo apresentamos os resultados dos modelos HAR-RV e MIDAS-RV para os 5 ativos e para os horizontes de tempo  $h = 1, 5, 10, 15$  e  $20$ , conforme descrito nas equações (2.10) e (2.12) respectivamente. Para ambos os modelos, a volatibilidade realizada utilizada foi filtrada, conforme (3.6), para corrigir problemas associados a microestrutura. Os coeficientes estimados dos modelos HAR-RV são apresentados na tabela 6, assim como seus respectivos valores p.

Tabela 6: Coeficientes Estimados do Modelo HAR-RV

Horizontes	$c$	P-Valor	$\beta^d$	P-Valor	$\beta^w$	P-Valor	$\beta^m$	P-Valor
<b>BBDC4</b>								
1 dia	-0,277	0,02	0,165	0,00	0,472	0,00	0,299	0,00
1 semana	-0,339	0,03	0,171	0,00	0,387	0,00	0,354	0,00
2 semanas	-0,444	0,01	0,130	0,00	0,401	0,00	0,352	0,00
3 semanas	-0,529	0,00	0,129	0,00	0,379	0,00	0,353	0,00
4 semanas	-0,627	0,00	0,119	0,01	0,395	0,01	0,322	0,02
<b>PETRA</b>								
1 dia	-0,290	0,01	0,198	0,00	0,461	0,00	0,273	0,00
1 semana	-0,363	0,02	0,174	0,00	0,413	0,00	0,320	0,00
2 semanas	-0,456	0,01	0,158	0,00	0,364	0,00	0,359	0,00
3 semanas	-0,529	0,00	0,145	0,00	0,315	0,00	0,402	0,00
4 semanas	-0,599	0,00	0,130	0,00	0,307	0,01	0,407	0,00
<b>VALE5</b>								
1 dia	-0,307	0,02	0,186	0,00	0,476	0,00	0,267	0,00
1 semana	-0,376	0,02	0,151	0,00	0,436	0,00	0,315	0,00
2 semanas	-0,469	0,01	0,130	0,00	0,430	0,00	0,316	0,00
3 semanas	-0,559	0,00	0,127	0,00	0,385	0,00	0,342	0,00
4 semanas	-0,655	0,00	0,114	0,00	0,378	0,00	0,335	0,00
<b>TNLP4</b>								
1 dia	-0,297	0,03	0,182	0,00	0,291	0,02	0,459	0,00
1 semana	-0,448	0,01	0,100	0,02	0,300	0,00	0,482	0,00
2 semanas	-0,564	0,00	0,088	0,02	0,345	0,00	0,417	0,00
3 semanas	-0,625	0,00	0,097	0,00	0,287	0,02	0,450	0,00
4 semanas	-0,709	0,00	0,083	0,01	0,309	0,02	0,418	0,00
<b>USIM5</b>								
1 dia	-0,282	0,04	0,206	0,00	0,303	0,01	0,422	0,00
1 semana	-0,350	0,05	0,131	0,00	0,323	0,00	0,451	0,00
2 semanas	-0,459	0,01	0,096	0,00	0,395	0,00	0,383	0,00
3 semanas	-0,570	0,00	0,098	0,00	0,372	0,00	0,372	0,00
4 semanas	-0,691	0,00	0,093	0,01	0,375	0,00	0,341	0,01

Fonte: Elaboração Própria

Nota: Coeficientes e seus respectivos p-valores resultantes da estimação dos modelos HAR-RV, desenvolvido por Corsi (2009), conforme descrito na equação (2.10). Os dados foram filtrados conforme (3.6). Os erros padrões foram corrigidos por Newey-West.

Os resultados descritos na tabela 6 mostram que os coeficientes estimados  $\beta^d$ ,  $\beta^w$  e  $\beta^m$

são significativos, confirmando assim a hipótese de alta persistência da volatilidade. Verifica-se também que, em média, as estimativas de  $\beta^d$  e de  $\beta^w$  decaem ao longo dos horizontes, diferentemente de  $\beta^m$  que tende a ser relativamente mais importante para a volatilidade de longo prazo. Tal fato já havia sido evidenciado por Andersen et al. (2007) e por Chung et al. (2008). Assim como neste trabalho, Andersen et al. (2007) também não encontrou um padrão comum a todos betas para os diferentes horizontes de tempo. Vale ressaltar que apenas o ativo PETR4 apresentou queda das estimativas de  $\beta^d$  e de  $\beta^w$  e elevação das estimativas de  $\beta^m$  conforme aumento do horizonte de previsão para todas as estimações.

Os coeficientes estimados dos modelos MIDAS-RV são apresentados na tabela 7, assim como seus respectivos valores p.

Tabela 7: Coeficientes Estimados do Modelo MIDAS-RV

Horizontes	$\alpha$	P-Valor	$\phi$	P-Valor	$\theta_1$	P-Valor	$\theta_2$	P-Valor
<b>BBDC4</b>								
1 dia	0,0016	0,03	0,9293	0,00	0,9988	0,00	15,9066	0,00
1 semana	0,0021	0,00	0,9013	0,00	0,9905	0,00	12,2882	0,00
2 semanas	0,0029	0,00	0,8660	0,00	0,9926	0,00	11,9034	0,00
3 semanas	0,0036	0,00	0,8314	0,00	0,9975	0,00	13,9990	0,00
4 semanas	0,0045	0,00	0,7909	0,00	0,9984	0,00	15,8491	0,00
<b>PETR4</b>								
1 dia	0,0015	0,05	0,9337	0,00	0,9807	0,00	10,7500	0,00
1 semana	0,0021	0,00	0,9031	0,00	0,9897	0,00	10,3859	0,00
2 semanas	0,0027	0,00	0,8707	0,00	0,9920	0,00	10,4316	0,00
3 semanas	0,0034	0,00	0,8401	0,00	0,9938	0,00	11,4309	0,00
4 semanas	0,0041	0,00	0,8071	0,00	0,9944	0,00	11,8838	0,00
<b>VALE5</b>								
1 dia	0,0019	0,02	0,9181	0,00	0,9918	0,00	13,0563	0,00
1 semana	0,0026	0,00	0,8847	0,00	0,9943	0,00	12,1346	0,00
2 semanas	0,0033	0,00	0,8495	0,00	0,9935	0,00	11,0012	0,00
3 semanas	0,0040	0,00	0,8146	0,00	0,9960	0,00	12,0207	0,00
4 semanas	0,0049	0,00	0,7771	0,00	0,9954	0,00	12,3643	0,00
<b>TNLP4</b>								
1 dia	0,0019	0,04	0,9184	0,00	0,9896	0,00	8,6825	0,00
1 semana	0,0025	0,00	0,8873	0,00	0,9936	0,00	8,6456	0,00
2 semanas	0,0033	0,00	0,8546	0,00	0,9961	0,00	9,7436	0,00
3 semanas	0,0038	0,00	0,8289	0,00	0,9970	0,00	10,3687	0,00
4 semanas	0,0045	0,00	0,7982	0,00	0,9992	0,00	11,5155	0,00
<b>USIM5</b>								
1 dia	0,0019	0,05	0,9302	0,00	0,9721	0,00	8,7542	0,00
1 semana	0,0026	0,00	0,9002	0,00	0,9876	0,00	9,0655	0,00
2 semanas	0,0034	0,00	0,8674	0,00	0,9972	0,00	10,3530	0,00
3 semanas	0,0044	0,00	0,8317	0,00	0,9970	0,00	11,6103	0,00
4 semanas	0,0055	0,00	0,7917	0,00	0,9974	0,00	12,3630	0,00

Fonte: Elaboração Própria

Nota: Coeficientes e seus respectivos p-valores resultantes da estimação dos modelos MIDAS-RV, desenvolvido por Ghysels et al. (2004), conforme descrito na equação (2.12). Os dados foram filtrados conforme (3.6).

Os resultados descritos na tabela 7 mostram que todos os coeficientes estimados são significativamente diferentes de zero. O principal parâmetro de interesse é o  $\phi$ , pois ele que determina se a volatilidade realizada futura é função das volatilidades realizadas passadas. Esse coeficiente se mostrou significativo para todos os ativos e em todos os horizontes de previsão. Como já mencionado, os parâmetros  $\theta_1$  e  $\theta_2$  determinam o formato da função de pesos.

Como esperado,  $\theta_1$  é próximo de 1 em todas as estimações, indicando que as funções de pesos decaem desde o primeiro lag.  $\theta_2$  indica a velocidade de decaimento da função de pesos. Como os valores estimados para  $\theta_2$  são relativamente altos, a função de pesos da volatilidade realizada decai rápido, chegando próximo de zero antes do lag 50.

## 4.1 Resultados Dentro da Amostra

Para compararmos os resultados dos modelos HAR-RV e MIDAS-RV dentro da amostra, seguimos Forsberg e Ghysels (2006) e Chung et al. (2008) e utilizamos o  $R^2$  ajustado e o erro quadrático médio (EQM) de dentro da amostra. Como o  $R^2$  ajustado é comparável apenas entre modelos com mesma variável dependente, contruímos um  $R^2$  para o HAR-RV retirando o logaritmo das variáveis. O EQM foi construído através da estimação dos modelos com todas as observações, comparando a série da  $RV$  estimada com a verificada para toda a amostra. Assim como no caso do  $R^2$ , também ajustamos o EQM para o fato de no modelo HAR-RV termos estimado as variáveis em sua forma logarítmica. Os resultados dos  $R^2$  ajustados estão descritos na tabela 8.

Tabela 8:  $R^2$  ajustado

Horizontes	BBDC4	PETRA4	VALE5	TNLP4	USIM5
<b>HAR-RV</b>					
<b>1 dia</b>	0,62	0,60	0,57	0,48	0,56
<b>1 semana</b>	0,74	0,75	0,71	0,70	0,71
<b>2 semanas</b>	0,73	0,74	0,70	0,71	0,72
<b>3 semanas</b>	0,70	0,71	0,67	0,68	0,69
<b>4 semanas</b>	0,66	0,68	0,63	0,65	0,65
<b>MIDAS-RV</b>					
<b>1 dia</b>	0,64	0,62	0,58	0,50	0,57
<b>1 semana</b>	0,78	0,77	0,75	0,72	0,73
<b>2 semanas</b>	0,76	0,76	0,74	0,74	0,75
<b>3 semanas</b>	0,74	0,74	0,71	0,73	0,72
<b>4 semanas</b>	0,69	0,70	0,67	0,70	0,68

Fonte: Elaboração Própria

Nota: Os  $R^2$  ajustados dos modelos HAR-RV foram construídos de modo que podem ser comparados aos dos modelos MIDAS-RV

Os resultados presentes na tabela 8, mostram que os modelos MIDAS-RV, para o mesmo ativo e mesmo horizonte de previsão, apresentam  $R^2$  ajustados levemente superiores aos dos modelos HAR-RV. É interessante salientar que os  $R^2$  ajustados para  $h = 1$  são sempre menores e que são crescentes até 2 semanas. Tal fato ocorre porque a variável dependente é a média da volatilidade realizada do horizonte de previsão e assim, como

afirma Forsberg e Ghysels (2006), é mais fácil prever horizontes maiores já que essas séries são relativamente mais suaves. Nesse sentido, para os dados em questão, encontrou-se um horizonte de previsão melhor com  $h = 10$ .

A tabela 9 mostra os EQM de dentro da amostra.

Tabela 9: Erros Quadráticos Médios (EQM) Dentro da Amostra

Horizontes	BBDC4	PETR4	VALE5	TNLP4	USIM5
<b>HAR-RV</b>					
<b>1 dia</b>	8,13	9,83	12,00	11,10	11,69
<b>1 semana</b>	4,16	4,63	5,65	4,27	5,49
<b>2 semanas</b>	4,12	4,43	5,51	3,69	4,91
<b>3 semanas</b>	4,40	4,69	5,77	3,86	5,19
<b>4 semanas</b>	4,84	5,08	6,16	4,11	5,70
<b>MIDAS-RV</b>					
<b>1 dia</b>	7,26	8,93	10,84	9,83	10,75
<b>1 semana</b>	3,35	3,97	4,66	3,60	4,74
<b>2 semanas</b>	3,40	3,78	4,48	3,06	4,18
<b>3 semanas</b>	3,61	3,99	4,68	3,11	4,38
<b>4 semanas</b>	4,11	4,49	5,21	3,33	4,94

Fonte: Elaboração Própria

Nota: Os EQM estão multiplicados por  $10^5$ .

Os EQM de dentro da amostra, sugerem que o modelo MIDAS-RV teve um melhor desempenho dentro da amostra para todas as estimações, na comparação com o mesmo ativo e mesmo horizonte de tempo. Pelo mesmo fator que faz com que os  $R^2$  ajustados sejam maiores em horizontes maiores, os EQM também apresentam um comportamento descendente até o horizonte de 2 semanas.

## 4.2 Resultados Fora da Amostra

Para compararmos o desempenho dos modelos para fora da amostra, seguimos o trabalho de Chung et al. (2008). Utilizamos o EQM fora da amostra e comparamos os erros de previsão fora da amostra através do teste de Diebold-Mariano modificado. Os erros de previsão fora da amostra foram calculados para 60 passos a frente a partir da estimação dos modelos HAR-RV e MIDAS-RV sem as últimas 60 observações (3 meses). Assim como no caso do EQM dentro da amostra, também transformamos o EQM fora da amostra do modelo HAR-RV retirando os logaritmos das variáveis usadas na construção da estatística. O teste de Diebold-Mariano modificado, proposto por Harvey et al. (1997), compara a diferença dos erros de previsão através de uma função específica  $g(e)$ , que neste trabalho, assim como em Harvey et al. (1997), usa-se a função quadrática. Sendo assim, suponha que tenhamos um par de erros de previsão  $(e_{1t}, e_{2t})$ ;  $t = 1, \dots, n$ . A hipótese nula do teste é:

$$H_0 : E [g(e_{1t}) - g(e_{2t})] = 0 \quad (4.1)$$

A estatística do teste é dada por:

$$S_1 = \left[ \frac{n+1 - 2\tau + n^{-1}\tau(\tau-1)}{n} \right]^{\frac{1}{2}} [\hat{V}(\bar{d})]^{-\frac{1}{2}} \bar{d}, \quad (4.2)$$

onde  $n$  é o número de observações,  $\tau$  é o número de períodos a frente que serão feitas as previsões,  $d_t = g(e_{1t}) - g(e_{2t}); t = 1, \dots, n$ ,  $\bar{d} = n^{-1} \sum_{t=1}^n d_t$  e  $\hat{V}(\bar{d})$  é a variância de  $\bar{d}$  que é dada por:

$$\hat{V}(\bar{d}) = n^{-1} \left[ \gamma_0 + 2 \sum_{k=1}^{\tau-1} \hat{\gamma}_k \right], \quad (4.3)$$

onde  $\hat{\gamma}_k$  é a  $k$ -ésima autocovariância de  $d_t$ .

Na tabela 10 estão presentes os EQM fora da amostra.

Tabela 10: Erros Quadráticos Médios (EQM) Fora da Amostra

Horizontes	BBDC4	PETR4	VALE5	TNLP4	USIM5
<b>HAR-RV</b>					
<b>1 dia</b>	1,96	2,45	3,22	18,60	1,52
<b>1 semana</b>	0,79	1,19	1,28	4,39	1,07
<b>2 semanas</b>	1,05	1,18	1,55	1,54	0,95
<b>3 semanas</b>	1,38	1,40	1,80	1,00	1,17
<b>4 semanas</b>	1,58	1,56	1,95	0,93	1,19
<b>MIDAS-RV</b>					
<b>1 dia</b>	2,10	2,42	3,23	18,97	3,23
<b>1 semana</b>	0,81	1,17	1,35	4,55	1,16
<b>2 semanas</b>	1,10	1,15	1,67	1,43	1,04
<b>3 semanas</b>	1,47	1,38	1,94	0,93	1,30
<b>4 semanas</b>	1,77	1,62	2,16	0,91	1,38

Fonte: Elaboração Própria

Nota: Os EQM estão multiplicados por  $10^5$ .

Os EQM apresentados na tabela 10 são muito próximos entre os diferentes modelos, não parecendo haver um padrão de qual modelo apresenta desempenho para fora da amostra superior. Cabe ressaltar também o comportamento próximo dos modelos com relação aos *outliers*, como no caso do EQM para fora da amostra do TNLP4 no horizonte  $h = 1$ . Outro fator interessante diz respeito ao valor do EQM fora da amostra em relação ao EQM dentro da amostra. Os EQM fora da amostra são menores independentemente do modelo estimado ou do horizonte de previsão. Isso ocorre devido ao fato de a crise do *subprime*, ocorrida em 2008, estar presente somente nos erros dentro da amostra, já que os erros fora da amostra são computados apenas nas últimas 60 observações.

A tabela 11 apresenta o valor p do teste de Diebold-Mariano modificado.



Tabela 11: Teste Diebold-Mariano para Fora da Amostra (P-VALOR)

Horizontes	BBDC4	PETR4	VALE5	TNLP4	USIM5
<b>1 dia</b>	0,990	0,997	0,999	0,970	0,994
<b>1 semana</b>	0,998	0,998	0,994	0,987	0,993
<b>2 semanas</b>	0,997	0,997	0,991	0,991	0,993
<b>3 semanas</b>	0,993	0,998	0,989	0,994	0,989
<b>4 semanas</b>	0,985	0,996	0,983	0,998	0,985

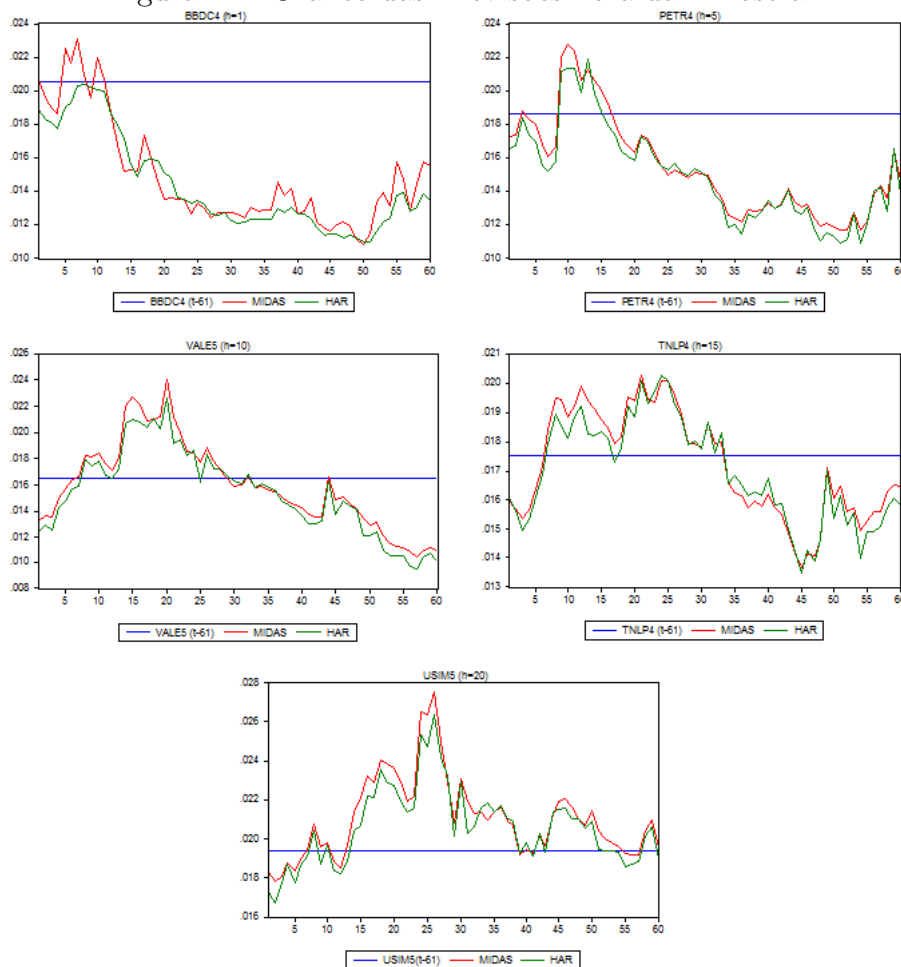
Fonte: Elaboração Própria

Nota: O teste foi realizado conforme a equação (4.2) e função perda utilizada foi a quadrática.

Os p-valores do teste de Diebold-Mariano, exibidos na tabela 11, confirmam que os erros de previsão para fora da amostra entre ambos os modelos não são estatisticamente diferentes. Dessa forma, para previsões fora da amostra, não encontram-se evidências que o desempenho do modelo MIDAS-RV para os dados utilizados seja superior ao do modelo HAR-RV, como ocorreu na análise dentro da amostra.

Como os resultados das previsões fora da amostra são muito semelhantes, construímos o gráfico, (4.1), das 60 previsões fora da amostra, para cada ativo em um horizonte de tempo. Junto, plotamos a volatilidade realizada do último dia da estimação e a traçamos constante para todas as previsões. Verifica-se que os modelos HAR-RV e MIDAS-RV, apresentam também comportamento semelhante ao longo das previsões.

Figura 4.1: Gráfico das Previsões Fora da Amostra



Nota: Gráficos com as previsões fora da amostra de cada ativo com um horizonte de tempo diferente. As previsões foram realizadas pelos modelos HAR-RV e MIDAS-RV, além disso, utilizamos também a última volatilidade realizada verificada, retirando as últimas 60 observações.

Fonte: Elaboração Própria

## Conclusão

Utilizando o referencial teórico de volatilidade realizada, proposto por Andersen et al. (2003), este trabalho teve o objetivo de comparar dois modelos de previsão de volatilidade para os 5 ativos mais líquidos do BOVESPA, no período analisado, utilizando dados intradiários com frequência de 5 minutos. Os ativos utilizados, poratanto, são: Bradesco (BBDC4), Petrobrás (PETR4), Vale do Rio Doce (VALE5), Telemar (TNLP4) e Usiminas (USIM5). Para corrigir o viés de microestrutura de mercado, utilizou-se o filtro MA(q), proposto por Hansen et al. (2008). Os modelos considerados neste trabalho foram o HAR-RV, proposto por Corsi (2009), e o MIDAS-RV, proposto por Ghysels et al. (2004).

Para comparação entre os modelos, utilizamos medidas de previsão de volatilidade realizada dentro e fora da amostra. A partir da consideração dos erros quadráticos médios (EQM) e dos  $R^2$  ajustados, encontraram-se evidências, para os ativos utilizados, de que

o modelo MIDAS-RV é superior ao modelo HAR-RV apenas dentro da amostra. Para previsões fora da amostra, segundo o teste de Diebold-Mariano modificado de comparação de EQM, não há diferença significativa entre os modelos. Sugere-se, assim, a utilização do modelo HAR-RV para previsões fora da amostra devido a sua maior facilidade de estimação.

Outro resultado importante deste trabalho diz respeito as evidências de que a utilização da volatilidade realizada induz distribuições dos retornos padronizados mais próximas da distribuição normal, como já havia sido encontrado para dados brasileiros por Carvalho et al. (2006).

Inúmeras possibilidades de estudos se apresentam para complementar o presente trabalho, principalmente no que se refere a estudos com dados brasileiros. Sugerem-se trabalhos, por exemplo, que contemplem os efeitos dos saltos de volatilidade que mostramos que ocorrem nos dados, que testem outros tipos de filtros para a correção da microestrutura de mercado ou, ainda, a aplicação de outros modelos de volatilidade realizada.

## Referências

- ANDERSEN, T., E BOLLERSLEV, T. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39(4):885–905, 1998.
- ANDERSEN, T., BOLLERSLEV, T., DIEBOLD, F., E EBENS, H. The distribution of realized stock return volatility\* 1. *Journal of Financial Economics*, 61(1):43–76, 2001.
- ANDERSEN, T., BOLLERSLEV, T., DIEBOLD, F., E LABYS, P. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.
- ANDERSEN, T., BOLLERSLEV, T., E DIEBOLD, F. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4):701–720, 2007.
- BARNDORFF-NIELSEN, O., ET AL. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280, 2002.
- CARVALHO, M., FREIRE, M., MEDEIROS, M., E SOUZA, L. Modeling and forecasting the volatility of Brazilian asset returns: a realized variance approach. *Revista Brasileira de Finanças*, 4:321–343, 2006.
- CHUNG, H., HUANG, C., E TSENG, T. Modeling and forecasting of realized volatility based on high-frequency data: evidence from Taiwan. *International Research Journal of Finance and Economics*, 22:178–191, 2008.
- CORSI, F. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 2009.
- DE SÁ MOTA, B., E FERNANDES, M. Desempenho de Estimadores de Volatilidade na Bolsa de Valores de Sao Paulo. *Revista Brasileira de Economia*, 58:429–448, 2004.

- EBENS, H. Realized stock volatility. *Department of Economics, Johns Hopkins University*, 1999.
- FORSBERG, L., E GHYSELS, E. Why do absolute returns predict volatility so well? *Journal of Financial Econometrics*, 5:31–67, 2006.
- GHYSELS, E., SANTA-CLARA, P., VALKANOV, R., E CIRANO. *The MIDAS touch: Mixed data sampling regression models*. Citeseer, 2004.
- GHYSELS, E., SINKO, A., E VALKANOV, R. MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90, 2007.
- GHYSELS, E.SANTA CLARA, P. V. R. Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131:59–95, 2006.
- GIOT, P., E LAURENT, S. Modelling daily Value-at-Risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, 11(3):379–398, 2004.
- HANSEN, P., LARGE, J., E LUNDE, A. Moving average-based estimators of integrated variance. *Econometric Reviews*, 27(1):79–111, 2008.
- HARVEY, D., LEYBOURNE, S., E NEWBOLD, P. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281 – 291, 1997.
- MCALEER, M., E MEDEIROS, M. Realized volatility: a review. *Econometric Reviews*, 27(1):10–45, 2008.
- OOMEN, R. *Modelling realized variance when returns are serially correlated*. Warwick Business School, 2002.