

**Universidade de São Paulo
Instituto de Matemática e Estatística**

Centro de Estatística Aplicada

Relatório de Análise Estatística

RAE-CEA–25P01

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

“Previsão de Volatilidade usando Dados Intradiários e Sentimento de Mercado”

Guilherme Vinicius Afonso Dias de Freitas

Leandro dos Santos Maciel

São Paulo, junho de 2025

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “Previsão de Volatilidade usando Dados Intradiários e Sentimento de Mercado”.

PESQUISADOR: Guilherme Vinicius Afonso Dias de Freitas

INSTITUIÇÃO: Faculdade de Economia, Administração, Contabilidade e Atuária da Universidade de São Paulo – FEA-USP

FINALIDADE DO PROJETO: Publicação de artigo científico

RESPONSÁVEIS PELA ANÁLISE: Guilherme Vinicius Afonso Dias de Freitas
Leandro dos Santos Maciel

REFERÊNCIA DESTE TRABALHO: FREITAS, G.V.A.D; MACIEL, L.S. **Relatório de análise estatística sobre o projeto: “Previsão de Volatilidade usando Dados Intradiários e Sentimento de Mercado”**. São Paulo, IME-USP, 2025. (RAE–CEA-25P01)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. (2011). Parametric and Nonparametric Volatility Measurement. **Handbook of the Economics of Finance**, 2, 67-122.

BAKER, S. R., BLOOM, N., DAVIS, S. J. (2016). Measuring Economic Policy Uncertainty. **Quarterly Journal of Economics**, Vol. 131, Nº 4, 1593–1636.

BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. (2004). Power and bipower variation with stochastic volatility and jumps. **Journal of Financial Econometrics**, Vol. 2, Nº 1, 1–37.

BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. **Journal of Financial Econometrics**, Vol. 4, Nº 1, 1–30.

BOLLERSLEV, T., GHYSELS, E. (1996). Periodic Autoregressive Conditional Heteroscedasticity. **Journal of Business & Economic Statistics**, Vol. 14, Nº 2, 139-51.

CORSI, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. **Journal of Financial Econometrics**, Vol. 7, Nº 2, 174–196.

CORSI, F., PIRINO, D., RENÒ, R. (2010). Threshold Bipower Variation and the Impact of Jumps on Volatility Forecasting. **Journal of Econometrics**, 159, 276–288.

CORSI, F.; RENÒ, R. (2012). Discrete-Time Volatility Forecasting with Persistent Leverage Effect and the Link with Continuous-Time Volatility Modeling. **Journal of Business & Economic Statistics**, Vol. 30, Nº 3, 368–380.

IZZELDIN, M., LIU, H., MURPHY, D. (2019). Forecasting Realized Volatility Using ARFIMA and HAR Models. **Quantitative Finance**, 19 (6), 963-976.

MINCER, J., ZARNOWITZ, V. (1969). The Evaluation of Economic Forecasts. Em: MINCER, J. (Ed.), Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance. **New York: National Bureau of Economic Research**, 3–46.

NEWBY, W. K., WEST, K. D. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. **Econometrica: Journal of the Econometric Society**, 703-708.

PATTON, A. J. (2011). Volatility Forecast Comparison Using Imperfect Volatility Proxies. **Journal of Econometrics**, Vol. 160, Nº 1, 246–256.

VAL, A. L. S., PINTO, A. C. F., KLOTZLE, M. C. (2014). Modelos de Volatilidade Realizados e Previsão de Volatilidade de Ações Brasileiras. **Revista Contabilidade & Finanças**, 25 (66), 172-185.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Python (3.11.3)

Microsoft Word para Windows (2024)

Microsoft Excel para Windows (2024)

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Unidimensional (03:010)

Estimação Paramétrica Multidimensional (04:160)

Testes de Hipóteses Paramétricas (05:010)

Análise de Regressão Clássica (07:020)

Outros (07:990)

Séries Temporais (11:010)

ÁREA DE APLICAÇÃO

Econometria (14:070)

Sumário

1. Introdução	8
2. Objetivo	8
3. Variáveis.....	9
3. Modelos HAR	10
Modelo HAR	10
Modelo HAR-CJ.....	10
Modelo HAR-TCJ	11
Modelo LHAR-TCJ.....	11
Modelo MLHAR-TCJ.....	11
Modelo ULHAR-TCJ	11
Modelo ALHAR-TCJ	12
Modelos HAR para funções de <i>RV</i>	12
4. Testes de <i>jumps</i>	12
5. Análise <i>in-sample</i>	13
6. Análise <i>out-of-sample</i>	15
7. Conclusão	17
APÊNDICE A	19
APÊNDICE B	34
APÊNDICE C	40

1. Introdução

Em finanças, a volatilidade de um ativo representa a variabilidade dos seus retornos ao longo do tempo, sendo frequentemente utilizada como uma medida de risco. Altos níveis de volatilidade implicam em maior incerteza quanto ao comportamento futuro dos preços, o que impacta diretamente o processo de tomada de decisão por parte de investidores, gestores de risco e formuladores de políticas econômicas.

No contexto atual, marcado por intensa dinâmica dos mercados financeiros, as criptomoedas se destacam como ativos de elevada volatilidade. Essa característica torna essencial o desenvolvimento de modelos robustos capazes de antecipar os padrões de variação de seus preços. Modelos baseados em dados intradiários têm se mostrado particularmente promissores nesse aspecto, conforme estudos como o de Val et al. (2014).

O presente trabalho se insere nesse contexto, visando contribuir para a literatura de previsão de volatilidade com foco em ativos digitais, por meio de modelos heterogêneos autoregressivos (HAR) e suas extensões.

2. Objetivo

O objetivo central deste estudo é desenvolver modelos estatísticos preditivos da volatilidade de ativos, com ênfase no uso de dados intradiários e variáveis exógenas que reflitam informações de mercado e sentimento econômico. Almeja-se:

- Avaliar o desempenho preditivo dos modelos HAR, HAR-CJ, HAR-TCJ, LHAR-TCJ e HAR-SJ;
- Investigar o valor informacional de variáveis como retorno negativo e indicadores de saltos;
- Adicionar medidas de retorno excessivo ao mercado;
- Incorporar medidas de sentimento, como o Índice de Incerteza de Política Econômica (GEPU) e Google Trends;
- Comparar os modelos em termos de previsões dentro e fora da amostra.

3. Variáveis

A variável dependente principal dos modelos é a *Realized Volatility (RV)*, obtida a partir da soma dos quadrados dos retornos intradiários:

$$RV_{i,t} = \sum_{i=1}^n r_{i,t}^2, \quad (1)$$

onde $r_{i,t}$ representa o retorno no intervalo i do dia t .

As Figuras B.1 a B.3 mostram a série de RV para os ativos Bitcoin (BTC), Amazon (AMZN) e Apple (AAPL), respectivamente.

As variáveis explanatórias incluem:

- Médias móveis de RV;
- “Saltos” (*jumps*): distinguem movimentos bruscos dos preços da variabilidade global;
- Retornos negativos: capturam efeitos de assimetria (*leverage effect*);
- S&P500: um índice de mercado, mantido pela empresa Dow Jones, que acompanha o desempenho de 500 das maiores empresas de capital aberto dos Estados Unidos, amplamente utilizado como *proxy* para movimentos do mercado de ações como um todo;
- GEPU - *Global Economic Policy Uncertainty Index*: índice desenvolvido em Baker et al. (2016) que mede a incerteza da política econômica a partir da frequência de termos relacionados à incerteza econômica em jornais de diversos países. Quanto maior o índice, maior a percepção de incerteza;
- Google Trends: usado como índice de atenção, fornece dados sobre o volume de buscas por palavras-chave ao longo do tempo, refletindo o interesse público ou de investidores por certos termos.

3. Modelos HAR

Os modelos do tipo heterogêneos autoregressivos (HAR) utilizam a volatilidade realizada do ativo para realizar previsões de volatilidade em janelas futuras com uma estrutura simples de regressão linear.

Estudos como o de Izzeldin et al. (2019) e Corsi (2009) mostram que o modelo HAR apresenta previsões de volatilidade robustas em diferentes condições de mercado e horizontes temporais, quando comparado com modelos tradicionais, como ARFIMA e GARCH.

Modelo HAR

Desenvolvido em Corsi (2009), o modelo HAR simples assume que a volatilidade realizada pode ser explicada por sua própria média ao longo de diferentes janelas temporais:

$$RV_t = \alpha + \beta_d RV_{t-1} + \beta_w RV_{t-5}^{(W)} + \beta_m RV_{t-22}^{(M)} + \epsilon_t, \quad (2)$$

com $RV_{t-5}^{(W)} = \frac{1}{5} \sum_{t=t-5}^{t-1} RV_t$ e $RV_{t-22}^{(M)} = \frac{1}{22} \sum_{t=t-22}^{t-1} RV_t$ médias móveis semanais e mensais de RV, respectivamente.

Se baseia na ideia de que diferentes tipos de agentes de mercado possuem comportamentos e tempos de reação distintos quando expostos a novas informações, gerando impacto na volatilidade em diversas janelas temporais.

Modelo HAR-CJ

O modelo HAR-CJ, apresentado em Corsi et al. (2010), separa a volatilidade realizada em componentes contínuos e discretos, utilizando *jumps*. Inclui saltos que afetam a dinâmica da volatilidade de forma diferente da variação contínua:

$$RV_t = \alpha + \beta_d C_{t-1} + \beta_w C_{t-5}^{(W)} + \beta_m C_{t-22}^{(M)} + \gamma J_{t-1} + \epsilon_t, \quad (3)$$

onde J_{t-1} representa a componente de *jumps* da volatilidade e $C_{t-1} = RV_{t-1} - J_{t-1}$.

Essa decomposição dos componentes gera uma melhor previsão pois permite captar de forma mais eficiente a dinâmica dos choques no mercado ao distingui-la de variações contínuas.

Modelo HAR-TCJ

O modelo HAR-TCJ (Andersen et al., 2011), possui a mesma estrutura do HAR-CJ, porém utiliza uma correção robusta no teste de saltos, gerando mais precisão e assertividade:

$$RV_t = \alpha + \beta_d TC_{t-1} + \beta_w TC_{t-5}^{(W)} + \beta_m TC_{t-22}^{(M)} + \gamma TJ_{t-1} + \epsilon_t, \quad (4)$$

onde TJ_{t-1} é a parcela que representa *jumps* da volatilidade, corrigida por um limitador, e $TC_{t-1} = RV_{t-1} - TJ_{t-1}$.

A utilização de um *threshold* (limitante) para calcular os saltos elimina parte do ruído nos dados, criando um modelo que performa melhor. Sua descrição e cálculo serão abordados de forma mais aprofundada na seção subsequente.

Modelo LHAR-TCJ

Corsi e Renò (2012) adicionam uma componente de assimetria (*leverage*) ao HAR-TCJ, seguindo a premissa de que os retornos negativos têm um impacto maior na volatilidade do que os positivos, criando o LHAR-TCJ:

$$RV_t = \alpha + \beta_d TC_{t-1} + \beta_w TC_{t-5}^{(W)} + \beta_m TC_{t-22}^{(M)} + \gamma TJ_{t-1} + \delta e^{r_{t-1}^-} + \epsilon_t, \quad (5)$$

em que $r_{t-1}^- = \min(r_{t-1}, 0)$.

Modelo MLHAR-TCJ

Agregando o retorno excedente do ativo sobre o mercado, nesse caso considerando o retorno do S&P500 como aproximação, buscamos metrificar o impacto do resultado excessivo na volatilidade, baseado na conexão intrínseca de risco e retorno. Assim, temos o modelo MLHAR-TCJ:

$$RV_t = \alpha + \beta_d TC_{t-1} + \beta_w TC_{t-5}^{(W)} + \beta_m TC_{t-22}^{(M)} + \gamma TJ_{t-1} + \delta e^{r_{t-1}^-} + \tau D_{t-1} + \epsilon_t, \quad (6)$$

com $D_{t-1} = \max(r_{t-1}^{ativo} - r_{t-1}^{S\&P500}, 0)$.

Modelo ULHAR-TCJ

Adicionando o GEPV como variável independente, criamos o modelo UHAR-TCJ, que busca cumular poder de previsão de volatilidade dos ativos utilizando o nível de incerteza política global refletida no mercado:

$$RV_t = \alpha + \beta_d TC_{t-1} + \beta_w TC_{t-5}^{(W)} + \beta_m TC_{t-22}^{(M)} + \gamma TJ_{t-1} + \delta e^{r_{t-1}^-} + \mu GEPU_{t-1} + \epsilon_t, \quad (7)$$

Modelo ALHAR-TCJ

Utilizando o Google Trends, esperamos capturar a influência do número de investidores atentos ao ativo em sua volatilidade, devido a variação do volume de transações e/ou a expectativa de acontecimentos relevantes:

$$RV_t = \alpha + \beta_d TC_{t-1} + \beta_w TC_{t-5}^{(W)} + \beta_m TC_{t-22}^{(M)} + \gamma TJ_{t-1} + \delta e^{r_{t-1}^-} + \psi G_{t-1} + \epsilon_t, \quad (8)$$

com G_t sendo o volume normalizado de pesquisas do termo referente ao ativo no Google Trends no dia t .

Modelos HAR para funções de RV

Os modelos serão estimados usando 3 funções de RV . Exemplificando com o modelo HAR-CJ, temos:

$$RV_t = \alpha + \beta_d C_{t-1} + \beta_w C_{t-5}^{(W)} + \beta_m C_{t-22}^{(M)} + \gamma J_{t-1} + \epsilon_t$$

$$\sqrt{RV_t} = \alpha + \beta_d \sqrt{C_{t-1}} + \beta_w \sqrt{C_{t-5}^{(W)}} + \beta_m \sqrt{C_{t-22}^{(M)}} + \gamma \sqrt{J_{t-1}} + \epsilon_t$$

$$\ln(RV_t) = \alpha + \beta_d \ln(C_{t-1}) + \beta_w \ln(C_{t-5}^{(W)}) + \beta_m \ln(C_{t-22}^{(M)}) + \gamma \ln(J_{t-1} + 1) + \epsilon_t$$

4. Testes de jumps

Para testar a existência de *jumps* no dia t utilizaremos um teste de hipóteses proposto por Barndorff-Nielsen e Shephard (2006), baseado na ideia de que, na ausência de saltos, a diferença entre a soma de quadrados dos retornos (RV) e a soma dos produtos dos pares de retornos imediatamente seguintes (BPV) deve ser pequena.

O teste pode ser realizado com as estatísticas de teste Z simples ou C -Tz, uma estatística robusta com limitador (*threshold*):

$$Z = \delta^{-\frac{1}{2}} \frac{(RV_t - BPV_t) \cdot RV_t^{-1}}{\sqrt{\theta \cdot \max \left\{ 1, \frac{TriPV_t}{(BPV_t)^2} \right\}}}$$

$$C-Tz = \delta^{-\frac{1}{2}} \frac{(RV_t - C-TBPV_t) \cdot RV_t^{-1}}{\sqrt{\theta \cdot \max \left\{ 1, \frac{C-TTriPV_t}{(C-TBPV_t)^2} \right\}}}$$

com $\theta = \frac{\pi^2}{4} + \pi - 5$, $\delta = \frac{T}{n}$, o tamanho dos subintervalos em que dividimos o período T (dia) e BPV_t , $TriPV_t$, $C-TBPV_t$, $C-TBPV_t$ descritos no Apêndice C. Ambas seguem uma Normal Padrão sob a hipótese nula de não existência de salto.

As Figuras B.4 a B.6 mostram que o número de dias em que saltos são detectados utilizando a estatística com *threshold* C-Tz é consistentemente maior do que com os testes feitos com a estatística Z para todos os ativos.

Finalmente, os saltos são calculados com base no teste considerado:

$$J_t = \mathbf{1}_{(z_t > \Phi_\alpha)} \max(RV_t - BPV_t, 0)$$

ou

$$TJ_t = \mathbf{1}_{(C-Tz_t > \Phi_\alpha)} \max(RV_t - TBPV_t, 0)$$

em que Φ_α é a função de distribuição acumulada (f.d.a.) da Normal Padrão, com nível de confiança α , e $TBPV_t$, é descrito no Apêndice C.

5. Análise *in-sample*

O propósito da análise *in-sample* (dentro da amostra), é observar se as variáveis explicativas possuem poder preditivo sobre a volatilidade dos ativos.

Para isso, utilizamos a série completa para estimar a regressão do modelo utilizando o método de mínimos quadrados ordinários (OLS), porém com o estimador de Newey-West (1987) para calcular a matriz de variância-covariância, dos estimadores dos coeficientes, de forma robusta na presença de heterocedasticidade e autocorrelação.

Avaliamos o desempenho preditivo dos modelos, nas transformações RV_t , $\sqrt{RV_t}$ e $\ln(RV_t)$, utilizando as medidas:

- R^2 : de Mincer-Zarnowitz (1969) para previsões de regressão;
- $HRMSE$: Raiz do Erro Quadrático Médio Heterocedástico Ajustado, proposto em Bollerslev e Ghysels (1996), dado por:

$$HRMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\frac{RV_t - \widehat{RV}_t}{RV_t} \right)^2};$$

onde \widehat{RV}_t é o valor previsto pelo modelo para a RV no tempo t .

- $QLIKE$: Função de Perda de Quase-Verossimilhança, robusta na avaliação de previsões de volatilidade, na forma definida em Patton (2011):

$$QLIKE = \frac{1}{T} \sum_{t=1}^T \left(\log RV_t - \frac{\widehat{RV}_t}{RV_t} \right),$$

As Tabelas A.1 a A.9 mostram os resultados da análise para o Bitcoin, Amazon e Apple.

Os resultados revelam padrões consistentes e algumas distinções importantes entre os ativos no desempenho dos modelos considerados.

Um resultado comum a todos os ativos foi a melhora sistemática nas métricas de previsão ao aplicarmos transformações na variável dependente. Especificamente, a transformação logarítmica da RV apresentou o melhor desempenho preditivo, seguida pela transformação por raiz quadrada, enquanto a modelagem da RV em sua forma original resultou nos ajustes menos eficazes. Por exemplo, no caso do BTC, os valores de R^2 do modelo mais simples, HAR, foram 0,279, 0,522 e 0,589, respectivamente. Tendência semelhante foi observada para AMZN e AAPL, indicando que transformações estabilizadoras de variância são eficazes para melhorar a qualidade da previsão.

Para o BTC, o modelo que apresentou o melhor desempenho geral foi o MLHAR-TCJ, que utiliza o retorno excedente sobre o mercado, com R^2 de 0,6, HRMSE de 0,893 e QLIKE de 1,867, porém, no caso da previsão de RV, o modelo com as melhores métricas foi o ALHAR-TCJ, com R^2 de 0,365, HRMSE de 2,005 e QLIKE de 2,171. Modelos mais

simples apresentaram R^2 inferiores e erros preditivos maiores, o que reforça a percepção de que a dinâmica de ativos voláteis como as criptomoedas possuem reflexo em seus retornos e devem sofrer maior impacto conforme os agentes de mercado desviam sua atenção em direção a eles.

No caso da AMZN, o modelo de melhor desempenho foi o HAR-CJ, com R^2 de 0,724, HRMSE de 0,581 e QLIKE de 0,713. A inclusão de variáveis adicionais em modelos mais complexos, como o ULHAR-TCJ ou ALHAR-TCJ, melhoram gradativamente as medidas de desempenho do modelo LHAR-TCJ, mas não o suficiente para superarem a previsibilidade do HAR-CJ. Isso indica que, para a AMZN, a estrutura básica de heterogeneidade de horizontes com componente de salto já é suficiente para capturar a dinâmica da volatilidade.

Por fim, para a AAPL, o modelo com melhor desempenho, considerando também a complexidade, foi o LHAR-TCJ, que apresentou R^2 de 0,594, HRMSE de 0,619 e QLIKE de 0,174. As tentativas de adicionar novas variáveis exógenas não resultou em ganhos expressivos: tanto o R^2 quanto os erros de previsão permaneceram sem melhoras significativas. Isso sugere que a estrutura do LHAR-TCJ é adequada para a dinâmica da AAPL, e que modelos mais complexos podem apenas aumentar o risco de sobreajuste sem ganhos substanciais em previsão ou que as variáveis exógenas estudadas não possuem poder explicativo sobre esse ativo.

Podemos observar o comportamento das previsões dos modelos que performaram melhor em cada ativo pelas Figuras B.7 a B.9.

6. Análise *out-of-sample*

Na análise *out-of-sample* (fora da amostra), cujo objetivo é validar a qualidade das previsões realizadas pelos modelos, a estimação é feita de forma iterativa:

- Partindo de uma janela inicial de tamanho $R = 0.4 * T$ (limitada a 3 anos), com T sendo o tamanho da série, a estimação do modelo é feita como na análise *in-sample*, com a variável dependente igual a média móvel de RV , de 1 dia, 1 semana, 2 semanas ou 1 mês, à frente;
- Em seguida, é feita a previsão da observação imediatamente seguinte ao final da janela, que é então armazenada em um vetor;

- A janela de estimação é movida 1 passo à frente (com tamanho R , fixo) e a previsão seguinte é feita, armazenada, e assim por diante.
- Por fim, é feita uma regressão simples entre os valores previstos e os reais, denominada Regressão OS.

As medidas utilizadas para comparação dos modelos, novamente para as funções RV_t , $\sqrt{RV_t}$ e $\ln(RV_t)$, são:

- R^2 : da Regressão OS;
- $HRMSE$, como descrita anteriormente, da Regressão OS;
- MAE : Erro Absoluto Médio, também da Regressão OS, dado por:

$$MAE = \frac{1}{T} \sum_{t=1}^T |RV_t - \widehat{RV}_t|.$$

Os resultados dessa análise (Tabelas A.10 a A.18), mostram o mesmo padrão quanto as transformações: melhores previsões são alcançadas utilizando a transformação logarítmica, seguida da raiz quadrada e por fim as previsões menos eficazes são obtidas sem transformação alguma (RV), com exceção da AAPL, onde as melhores previsões são realizadas com a transformação pela raiz quadrada.

No caso do BTC, a melhora das previsões também ocorre conforme aumentamos a janela de previsão, de 1 dia até 1 mês, algo que pode ser específico de ativos tão voláteis quanto as criptomoedas, reduzindo o ruído nos dados e amortecendo grandes variações. Novamente, o melhor modelo é o MLHAR-TCJ, porém com menos disparidades nas medidas comparativas, fazendo previsões menos precisas em janelas curtas, de 1 dia e 1 semana, do que outros modelos mais simples, como o HAR-CJ, em especial para a transformação do logaritmo de RV .

De forma similar ao BTC, as previsões da AMZN também melhoram ao longo do aumento das janelas temporais, porém, o R^2 começa a decair a partir da janela mensal, indicando que, para esse ativo, estender muito a janela de previsão pode causar uma suavização exagerada da volatilidade, reduzindo assim as informações específicas contidas em janelas menores. Em linha com a análise *in-sample*, o modelo que apresenta as melhores previsões fora da amostra é o HAR-CJ, porém, não com o destaque anterior,

tendo previsões menos precisas do que o modelo ALHAR-TCJ, em janelas de 1 dia e 1 semana na transformação quadrática.

Já para AAPL, janelas preditivas de 2 semanas ou mais apresentam medidas comparativas piores do que da janela de 1 semana, onde ocorre a maior precisão das previsões, replicando o movimento de perda de informação observado na AMZN. Divergente da análise *in-sample*, o modelo com a melhor qualidade preditiva fora da amostra é o ULHAR-TCJ, que apresenta medidas consistentemente maiores do que o modelo LHAR-TCJ, apontando que, provavelmente, ativos ligados diretamente ao poder de compra e políticas econômicas, sofrem maiores impactos dos índices de incerteza como o GEPU.

Mais uma vez, podemos observar o comportamento das previsões, em todas as janelas temporais, dos modelos que performaram melhor em cada ativo pelas Figuras B.10 a B.12.

7. Conclusão

As análises sobre o comportamento dos modelos HAR para previsão de volatilidade realizada (RV), reforçam tanto a importância de considerar a transformação da variável dependente quanto a adequação específica do modelo à dinâmica de cada ativo. Enquanto o BTC exige maior flexibilidade estrutural, ativos como AMZN e AAPL parecem responder melhor a modelos mais parcimoniosos, desde que adequadamente especificados.

Em ambos os cenários, *in-sample* e *out-of-sample*, a transformação logarítmica da RV se destacou como a mais eficaz no geral, seguida pela transformação por raiz quadrada, com a modelagem da RV em sua forma original apresentando os piores desempenhos. Essa tendência se confirmou para BTC, AMZN e AAPL, refletindo a eficácia das transformações estabilizadoras de variância na melhora das métricas preditivas.

Os desempenhos dos modelos variaram significativamente entre os ativos. Para o BTC, o modelo MLHAR-TCJ apresentou o melhor ajuste *in-sample* e também o melhor desempenho preditivo em janelas maiores fora da amostra, evidenciando a importância de variáveis de retorno de mercado em ativos de alta volatilidade. Para a AMZN, o HAR-

CJ demonstrou desempenho superior *in-sample* e manteve sua competitividade *out-of-sample*, principalmente em janelas curtas, sugerindo que a inclusão de saltos já é suficiente para capturar a dinâmica de sua volatilidade. No caso da AAPL, embora o LHAR-TCJ tenha se destacado *in-sample*, o ULHAR-TCJ superou os demais modelos fora da amostra, especialmente em janelas curtas, indicando que variáveis relacionadas à incerteza econômica (como o GEPU) podem ter maior relevância nesse ativo.

Adicionalmente, observou-se que o horizonte de previsão influencia fortemente a acurácia preditiva: no BTC e AMZN, janelas maiores tendem a suavizar o ruído e melhorar a performance, até certo ponto; por outro lado, para a AAPL, janelas maiores levam à perda de informação, com o melhor desempenho concentrado em horizontes curtos (1 semana).

Esses resultados reforçam que não há um modelo universalmente superior, sendo necessário considerar tanto a natureza do ativo quanto o horizonte temporal e a forma de transformação da variável de interesse. Além disso, modelos mais complexos nem sempre geram ganhos significativos em previsão, podendo inclusive aumentar o risco de sobreajuste, especialmente quando as variáveis exógenas adicionadas possuem baixo poder explicativo.

APÊNDICE A

Tabelas

Tabela A.1 Resultados *in-sample* para previsão de RV do Bitcoin (BTC)

	HAR	HAR-CJ	HAR-TCJ	LHAR-TCJ	MLHAR-TCJ	ULHAR-TCJ	ALHAR-TCJ
α	0,011***	0,012***	0,007*	0,971*	1,000**	0,968*	-3,657*
β_d	0,392***	0,390***	0,176	0,102	0,102	0,101	0,377
β_w	0,052	0,070	0,140	0,204	0,204	0,201	0,079
β_m	0,219**	0,217*	0,353	0,294*	0,295*	0,315*	0,186*
γ		-0,154	1,050	0,863*	0,862*	0,859*	3,898*
δ				-0,973*	-1,002**	-0,979*	-0,237*
$T / \mu / \psi$					-0,031	0,000	0,000
R^2	0,245	0,247	0,302	0,362	0,362	0,362	0,365
HRMSE	2,146	2,208	2,001	2,073	2,078	2,095	2,005
QLIKE	2,712	2,728	2,563	2,292	2,292	2,299	2,171

As estimativas com valores-p menores do que 5%, 1% e 0,1% são denotadas por *, ** e ***, respectivamente.

Tabela A.2 Resultados *in-sample* para previsão de \sqrt{RV} do Bitcoin (BTC)

	HAR	HAR-CJ	HAR-TCJ	LHAR-TCJ	MLHAR-TCJ	ULHAR-TCJ	ALHAR-TCJ
α	0,028***	0,031***	0,035***	1,790*	3,567**	1,787*	1,769*
β_d	0,488***	0,498***	0,485***	0,355***	0,326***	0,354***	0,345***
β_w	0,137*	0,132*	0,162*	0,249**	0,254***	0,249**	0,244**
β_m	0,157**	0,158**	0,143**	0,132**	0,144**	0,135**	0,112*
γ		-0,036	0,188**	0,129**	0,101*	0,129**	0,136**
δ				-1,755*	-3,520**	-1,758*	-1,749*
$\tau / \mu / \psi$					-0,284*	0,000	0,004*
R^2	0,478	0,484	0,478	0,513	0,529	0,513	0,516
HRMSE	1,223	1,227	1,202	1,186	1,226	1,187	1,147
QLIKE	2,105	2,102	2,093	2,084	2,093	2,085	2,071

As estimativas com valores-p menores do que 5%, 1% e 0,1% são denotadas por *, ** e ***, respectivamente.

Tabela A.3 Resultados *in-sample* para previsão de $\ln(RV)$ do Bitcoin (BTC)

	HAR	HAR-CJ	HAR-TCJ	LHAR-TCJ	MLHAR-TCJ	ULHAR-TCJ	ALHAR-TCJ
α	-0,565***	-0,504***	-0,587***	-0,730***	-0,944***	-0,394	-1,083***
β_d	0,491***	0,512***	0,461***	0,403***	0,391***	0,404***	0,389***
β_w	0,233***	0,209***	0,210***	0,251***	0,253***	0,250***	0,250***
β_m	0,169***	0,170***	0,153***	0,147***	0,150***	0,141***	0,132***
γ		-4,828	3,562***	2,555*	2,024	2,596*	2,719*
δ				-3,979***	-6,042***	-3,937***	-4,001***
$T / \mu / \psi$					-0,021**	-0,065	0,080*
R^2	0,568	0,578	0,589	0,597	0,600	0,597	0,598
HRMSE	0,962	0,965	0,982	0,917	0,893	0,922	0,902
QLIKE	1,889	1,882	1,877	1,870	1,867	1,870	1,868

As estimativas com valores-p menores do que 5%, 1% e 0,1% são denotadas por *, ** e ***, respectivamente.

Tabela A.4 Resultados *in-sample* para previsão de RV da Amazon (AMZN)

	HAR	HAR-CJ	HAR-TCJ	LHAR-TCJ	MLHAR-TCJ	ULHAR-TCJ	ALHAR-TCJ
α	0,003**	0,002*	0,003***	0,121*	0,291**	0,143*	0,123*
β_d	0,082*	0,437***	0,583**	0,549**	0,511*	0,561**	0,550**
β_w	0,133	0,342*	0,588*	0,588*	0,621*	0,601*	0,581*
β_m	0,595***	0,561**	0,440*	0,434*	0,422*	0,392*	0,436*
γ		0,002	0,090**	0,080*	0,091*	0,064	0,079*
δ				-0,119*	-0,288**	-0,134*	-0,121*
$T / \mu / \psi$					-0,257**	0,000***	0,000*
R^2	0,172	0,207	0,179	0,181	0,184	0,191	0,182
HRMSE	1,809	1,348	1,853	1,762	1,756	1,600	1,773
QLIKE	1,588	1,403	1,717	1,667	1,658	1,462	1,666

As estimativas com valores-p menores do que 5%, 1% e 0,1% são denotadas por *, ** e ***, respectivamente.

Tabela A.5 Resultados *in-sample* para previsão de \sqrt{RV} da Amazon (AMZN)

	HAR	HAR-CJ	HAR-TCJ	LHAR-TCJ	MLHAR-TCJ	ULHAR-TCJ	ALHAR-TCJ
α	0,011***	0,000	0,005**	0,325	0,388	0,512**	0,427*
β_d	0,256***	0,449***	0,400***	0,379***	0,377***	0,391***	0,391***
β_w	0,158***	0,298***	0,401***	0,407***	0,409***	0,413***	0,393***
β_m	0,427***	0,387***	0,293***	0,298***	0,298***	0,274***	0,283***
γ		0,027	0,245***	0,239***	0,239***	0,195***	0,206***
δ				-0,319	-0,382	-0,469*	-0,410*
$T / \mu / \psi$					-0,010	-0,003***	-0,001***
R^2	0,460	0,516	0,467	0,468	0,468	0,482	0,479
HRMSE	1,063	0,724	0,903	0,904	0,903	0,838	0,911
QLIKE	0,978	0,870	0,947	0,945	0,945	0,912	0,945

As estimativas com valores-p menores do que 5%, 1% e 0,1% são denotadas por *, ** e ***, respectivamente.

Tabela A.6 Resultados *in-sample* para previsão de $\ln(RV)$ da Amazon (AMZN)

	HAR	HAR-CJ	HAR-TCJ	LHAR-TCJ	MLHAR-TCJ	ULHAR-TCJ	ALHAR-TCJ
α	-0,525***	0,410***	0,667***	0,560***	0,592***	2,479***	0,582***
β_d	0,487***	0,499***	0,222***	0,201***	0,203***	0,220***	0,259***
β_w	0,184***	0,286***	0,535***	0,536***	0,536***	0,499***	0,477***
β_m	0,242***	0,260***	0,287***	0,292***	0,292***	0,245***	0,282***
γ		-1,679***	4,789***	4,270***	4,318***	2,749**	1,611**
δ				-4,362***	-3,933**	-6,155***	-5,761***
$T / \mu / \psi$					0,003	-0,458***	-0,049***
R^2	0,689	0,724	0,631	0,633	0,633	0,668	0,687
HRMSE	0,925	0,581	0,723	0,706	0,707	0,647	0,661
QLIKE	0,760	0,713	0,768	0,765	0,765	0,744	0,738

As estimativas com valores-p menores do que 5%, 1% e 0,1% são denotadas por *, ** e ***, respectivamente.

Tabela A.7 Resultados *in-sample* para previsão de RV da Apple (AAPL)

	HAR	HAR-CJ	HAR-TCJ	LHAR-TCJ	MLHAR-TCJ	ULHAR-TCJ	ALHAR-TCJ
α	0,002***	0,001***	0,002***	0,109*	0,205***	0,108*	0,110*
β_d	0,191***	0,366***	0,592***	0,526**	0,521**	0,527**	0,526**
β_w	0,353**	0,682***	0,859***	0,895***	0,893***	0,893***	0,895***
β_m	0,231**	0,072	-0,128	-0,148	-0,144	-0,148	-0,151
γ		0,069	0,106*	0,086	0,092*	0,086	0,086
δ				-0,108*	-0,203***	-0,107**	-0,108*
$T / \mu / \psi$					-0,170**	0,000	0,000
R^2	0,240	0,282	0,295	0,299	0,302	0,299	0,299
HRMSE	1,714	1,324	1,268	1,157	1,171	1,174	1,150
QLIKE	0,927	0,791	0,765	0,691	0,690	0,697	0,687

As estimativas com valores-p menores do que 5%, 1% e 0,1% são denotadas por *, ** e ***, respectivamente.

Tabela A.8 Resultados *in-sample* para previsão de \sqrt{RV} da Apple (AAPL)

	HAR	HAR-CJ	HAR-TCJ	LHAR-TCJ	MLHAR-TCJ	ULHAR-TCJ	ALHAR-TCJ
α	0,012***	0,007***	0,008***	0,773***	1,061***	0,763***	0,768***
β_d	0,320***	0,553***	0,646***	0,559***	0,555***	0,560***	0,559***
β_w	0,220***	0,315***	0,351***	0,394***	0,394***	0,393***	0,393***
β_m	0,257***	0,141**	0,063	0,063	0,064	0,064	0,065
γ		-0,002	0,085***	0,067**	0,063**	0,066**	0,067**
δ				-0,764***	-1,051***	-0,758***	-0,762***
$T / \mu / \psi$					-0,044**	0,000	0,000
R^2	0,405	0,471	0,482	0,489	0,490	0,489	0,489
HRMSE	1,179	0,805	0,760	0,740	0,735	0,744	0,741
QLIKE	0,446	0,346	0,333	0,322	0,322	0,324	0,323

As estimativas com valores-p menores do que 5%, 1% e 0,1% são denotadas por *, ** e ***, respectivamente.

Tabela A.9 Resultados *in-sample* para previsão de $\ln(RV)$ da Apple (AAPL)

	HAR	HAR-CJ	HAR-TCJ	LHAR-TCJ	MLHAR-TCJ	ULHAR-TCJ	ALHAR-TCJ
α	-0,911***	-0,324***	-0,364***	-0,634***	-0,783***	-0,782***	-0,85***0
β_d	0,435***	0,513***	0,485***	0,420***	0,418***	0,421***	0,422***
β_w	0,178***	0,221***	0,240***	0,268***	0,267***	0,268***	0,267***
β_m	0,243***	0,181***	0,151***	0,153***	0,152***	0,154***	0,156***
γ		-2,261*	2,627***	0,614	0,136	0,564	0,607
δ				-9,346***	-11,567***	-9,271***	-9,274***
$T / \mu / \psi$					-0,014***	0,033	0,059
R^2	0,523	0,581	0,584	0,594	0,596	0,594	0,594
HRMSE	0,924	0,650	0,627	0,619	0,621	0,620	0,621
QLIKE	0,230	0,182	0,178	0,174	0,173	0,174	0,174

As estimativas com valores-p menores do que 5%, 1% e 0,1% são denotadas por *, ** e ***, respectivamente.

Tabela A.10 Resultados *out-of-sample* para previsão de RV do Bitcoin (BTC)

	1 dia			1 semana			2 semanas			1 mês		
	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE
HAR	0,032	3,713	2,974	0,047	1,930	2,361	0,030	1,641	2,298	0,008	1,224	2,235
HAR-CJ	0,031	3,732	2,974	0,047	1,931	2,355	0,032	1,637	2,288	0,009	1,216	2,226
HAR-TCJ	0,025	3,763	3,044	0,035	1,959	2,385	0,023	1,666	2,305	0,010	1,211	2,217
LHAR-TCJ	0,039	3,523	2,966	0,037	1,949	2,375	0,027	1,647	2,290	0,013	1,204	2,209
MLHAR-TCJ	0,039	3,478	2,969	0,031	1,976	2,403	0,023	1,660	2,300	0,011	1,212	2,213
ULHAR-TCJ	0,041	3,534	2,970	0,045	2,144	2,411	0,038	1,703	2,298	0,033	1,146	2,212
ALHAR-TCJ	0,041	3,350	2,939	0,080	1,697	2,158	0,077	1,429	2,109	0,064	1,074	2,093

Tabela A.11 Resultados *out-of-sample* para previsão de \sqrt{RV} do Bitcoin (BTC)

	1 dia			1 semana			2 semanas			1 mês		
	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE
HAR	0,093	2,739	2,578	0,173	1,198	1,571	0,145	1,087	1,433	0,094	0,896	1,339
HAR-CJ	0,091	2,823	2,585	0,175	1,199	1,559	0,151	1,079	1,420	0,108	0,876	1,326
HAR-TCJ	0,056	3,205	2,729	0,162	1,280	1,554	0,162	1,099	1,389	0,146	0,844	1,275
LHAR-TCJ	0,095	2,590	2,518	0,165	1,262	1,542	0,164	1,095	1,387	0,146	0,847	1,277
MLHAR-TCJ	0,160	2,004	2,396	0,173	1,229	1,520	0,173	1,081	1,372	0,151	0,850	1,272
ULHAR-TCJ	0,098	2,582	2,517	0,169	1,333	1,567	0,167	1,142	1,411	0,166	0,826	1,301
ALHAR-TCJ	0,098	2,446	2,497	0,247	0,950	1,341	0,292	0,788	1,171	0,292	0,664	1,136

Tabela A.12 Resultados *out-of-sample* para previsão de $\ln(RV)$ do Bitcoin (BTC)

	1 dia			1 semana			2 semanas			1 mês		
	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE
HAR	0,147	1,469	2,370	0,324	0,781	1,157	0,260	0,817	1,048	0,178	0,763	0,963
HAR-CJ	0,152	1,530	2,340	0,331	0,779	1,141	0,274	0,803	1,031	0,210	0,730	0,946
HAR-TCJ	0,084	3,597	2,925	0,310	0,879	1,122	0,314	0,769	0,975	0,277	0,677	0,900
LHAR-TCJ	0,371	2,870	2,480	0,323	0,836	1,111	0,314	0,763	0,977	0,276	0,678	0,901
MLHAR-TCJ	0,554	3,054	2,388	0,362	0,757	1,078	0,318	0,763	0,974	0,275	0,686	0,897
ULHAR-TCJ	0,358	2,863	2,483	0,321	0,861	1,123	0,309	0,790	0,993	0,282	0,687	0,912
ALHAR-TCJ	0,394	2,825	2,450	0,450	0,582	0,924	0,516	0,498	0,769	0,484	0,506	0,736

Tabela A.13 Resultados *out-of-sample* para previsão de RV da Amazon (AMZN)

	1 dia			1 semana			2 semanas			1 mês		
	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE
HAR	0,191	1,922	0,712	0,405	1,566	0,653	0,476	1,450	0,625	0,485	1,278	0,579
HAR-CJ	0,232	1,475	0,648	0,503	1,284	0,587	0,598	1,239	0,563	0,591	1,193	0,534
HAR-TCJ	0,220	1,333	0,652	0,482	1,171	0,595	0,570	1,155	0,572	0,553	1,162	0,545
LHAR-TCJ	0,221	1,298	0,644	0,486	1,160	0,589	0,576	1,156	0,568	0,558	1,162	0,542
MLHAR-TCJ	0,224	1,285	0,644	0,490	1,137	0,588	0,580	1,141	0,567	0,564	1,149	0,540
ULHAR-TCJ	0,219	1,307	0,651	0,477	1,136	0,597	0,559	1,110	0,579	0,512	1,076	0,565
ALHAR-TCJ	0,224	1,468	0,643	0,497	1,321	0,582	0,593	1,304	0,561	0,583	1,279	0,534

Tabela A.14 Resultados *out-of-sample* para previsão de \sqrt{RV} da Amazon (AMZN)

	1 dia			1 semana			2 semanas			1 mês		
	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE
HAR	0,211	1,479	0,676	0,561	0,986	0,458	0,608	0,902	0,420	0,589	0,813	0,386
HAR-CJ	0,246	1,166	0,629	0,661	0,748	0,407	0,716	0,709	0,371	0,679	0,689	0,347
HAR-TCJ	0,240	1,002	0,628	0,630	0,669	0,416	0,670	0,655	0,385	0,625	0,667	0,365
LHAR-TCJ	0,241	0,981	0,621	0,632	0,671	0,413	0,676	0,655	0,383	0,629	0,665	0,364
MLHAR-TCJ	0,240	0,981	0,621	0,630	0,671	0,413	0,674	0,655	0,383	0,627	0,664	0,364
ULHAR-TCJ	0,241	1,002	0,621	0,628	0,676	0,414	0,665	0,653	0,389	0,611	0,643	0,377
ALHAR-TCJ	0,249	1,089	0,613	0,663	0,736	0,397	0,714	0,703	0,365	0,665	0,696	0,346

Tabela A.15 Resultados *out-of-sample* para previsão de $\ln(RV)$ da Amazon (AMZN)

	1 dia			1 semana			2 semanas			1 mês		
	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE
HAR	0,225	1,287	0,652	0,643	0,750	0,348	0,666	0,700	0,324	0,638	0,652	0,314
HAR-CJ	0,251	1,097	0,624	0,727	0,554	0,315	0,752	0,533	0,291	0,714	0,523	0,283
HAR-TCJ	0,222	0,921	0,652	0,621	0,560	0,360	0,626	0,547	0,338	0,599	0,542	0,324
LHAR-TCJ	0,218	0,934	0,643	0,631	0,562	0,356	0,641	0,550	0,336	0,609	0,540	0,322
MLHAR-TCJ	0,216	0,954	0,644	0,630	0,559	0,354	0,639	0,550	0,336	0,607	0,541	0,322
ULHAR-TCJ	0,247	0,990	0,628	0,691	0,598	0,334	0,709	0,587	0,314	0,669	0,582	0,310
ALHAR-TCJ	0,219	0,952	0,638	0,632	0,563	0,354	0,636	0,549	0,338	0,600	0,534	0,325

Tabela A.16 Resultados *out-of-sample* para previsão de RV da Apple (AAPL)

	1 dia			1 semana			2 semanas			1 mês		
	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE
HAR	0,254	1,928	0,476	0,402	1,421	0,421	0,398	1,288	0,395	0,300	1,090	0,380
HAR-CJ	0,303	1,607	0,438	0,485	1,290	0,391	0,488	1,200	0,376	0,372	1,041	0,370
HAR-TCJ	0,315	1,546	0,434	0,500	1,274	0,391	0,494	1,205	0,378	0,366	1,048	0,372
LHAR-TCJ	0,311	1,482	0,434	0,504	1,217	0,388	0,498	1,170	0,377	0,367	1,038	0,370
MLHAR-TCJ	0,309	1,501	0,436	0,504	1,238	0,389	0,503	1,180	0,377	0,370	1,042	0,371
ULHAR-TCJ	0,309	1,511	0,436	0,508	1,191	0,385	0,510	1,146	0,379	0,401	1,028	0,388
ALHAR-TCJ	0,310	1,479	0,435	0,502	1,212	0,387	0,493	1,187	0,379	0,363	1,052	0,374

Tabela A.17 Resultados *out-of-sample* para previsão de \sqrt{RV} da Apple (AAPL)

	1 dia			1 semana			2 semanas			1 mês		
	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE
HAR	0,284	1,632	0,450	0,495	1,008	0,313	0,475	0,913	0,286	0,375	0,798	0,272
HAR-CJ	0,319	1,360	0,422	0,561	0,898	0,290	0,535	0,826	0,268	0,420	0,739	0,260
HAR-TCJ	0,334	1,277	0,415	0,580	0,857	0,286	0,550	0,800	0,266	0,422	0,729	0,260
LHAR-TCJ	0,327	1,245	0,413	0,586	0,831	0,283	0,556	0,783	0,264	0,426	0,722	0,259
MLHAR-TCJ	0,323	1,271	0,414	0,586	0,845	0,283	0,557	0,791	0,264	0,425	0,725	0,259
ULHAR-TCJ	0,331	1,187	0,410	0,608	0,775	0,280	0,596	0,725	0,262	0,504	0,676	0,262
ALHAR-TCJ	0,327	1,235	0,412	0,587	0,815	0,281	0,556	0,774	0,264	0,429	0,720	0,260

Tabela A.18 Resultados *out-of-sample* para previsão de $\ln(RV)$ da Apple (AAPL)

	1 dia			1 semana			2 semanas			1 mês		
	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE	R ²	HRMSE	MAE
HAR	0,298	1,437	0,440	0,548	0,824	0,256	0,525	0,753	0,237	0,441	0,667	0,218
HAR-CJ	0,318	1,191	0,420	0,594	0,650	0,232	0,551	0,613	0,216	0,461	0,580	0,205
HAR-TCJ	0,296	1,344	0,425	0,467	0,832	0,244	0,371	0,856	0,235	0,348	0,691	0,212
LHAR-TCJ	0,255	1,532	0,439	0,481	0,816	0,242	0,392	0,829	0,231	0,360	0,679	0,210
MLHAR-TCJ	0,235	1,674	0,450	0,493	0,804	0,241	0,403	0,814	0,229	0,361	0,676	0,210
ULHAR-TCJ	0,271	1,460	0,432	0,572	0,699	0,230	0,520	0,676	0,214	0,487	0,586	0,201
ALHAR-TCJ	0,255	1,530	0,439	0,483	0,808	0,241	0,395	0,822	0,231	0,368	0,671	0,210

APÊNDICE B

Figuras

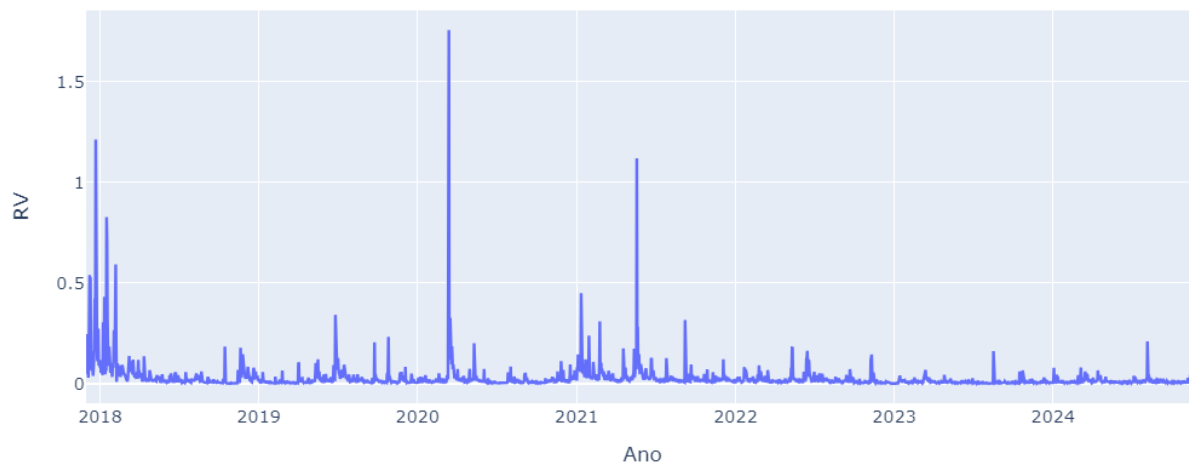


Figura B.1 Série temporal da volatilidade realizada (RV) do Bitcoin (BTC)

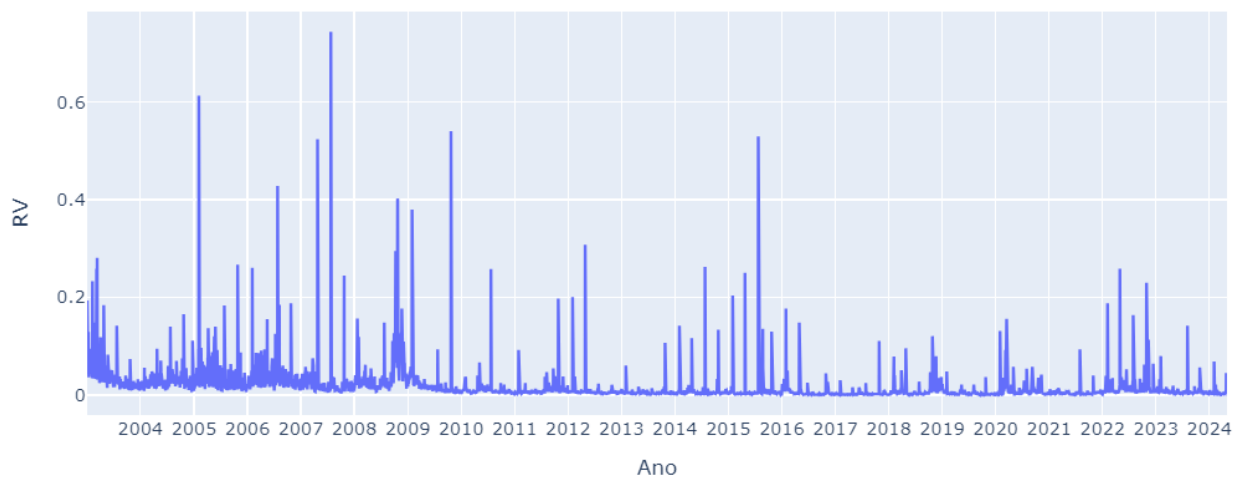


Figura B.2 Série temporal da volatilidade realizada (RV) da Amazon (AMZN)

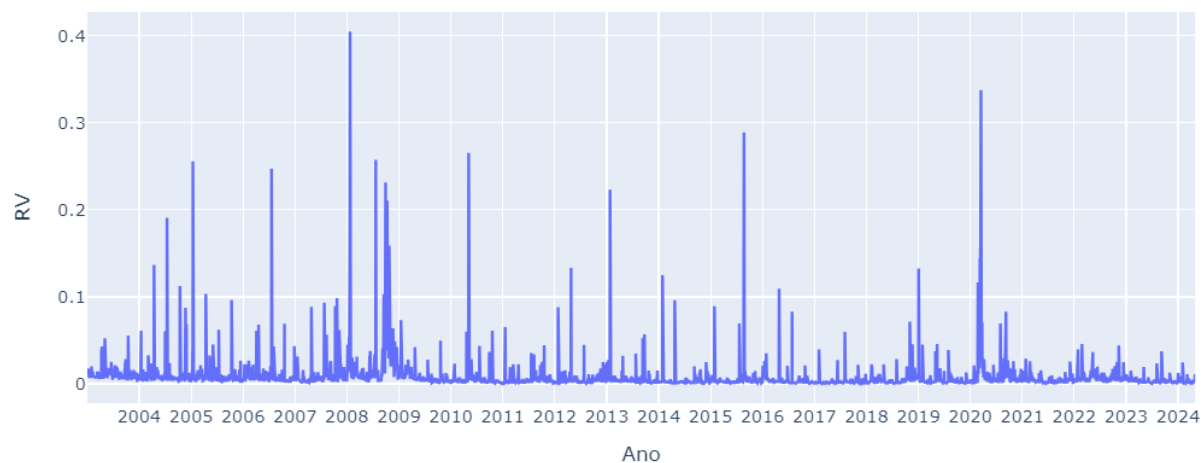


Figura B.3 Série temporal da volatilidade realizada (RV) da Apple (AAPL)

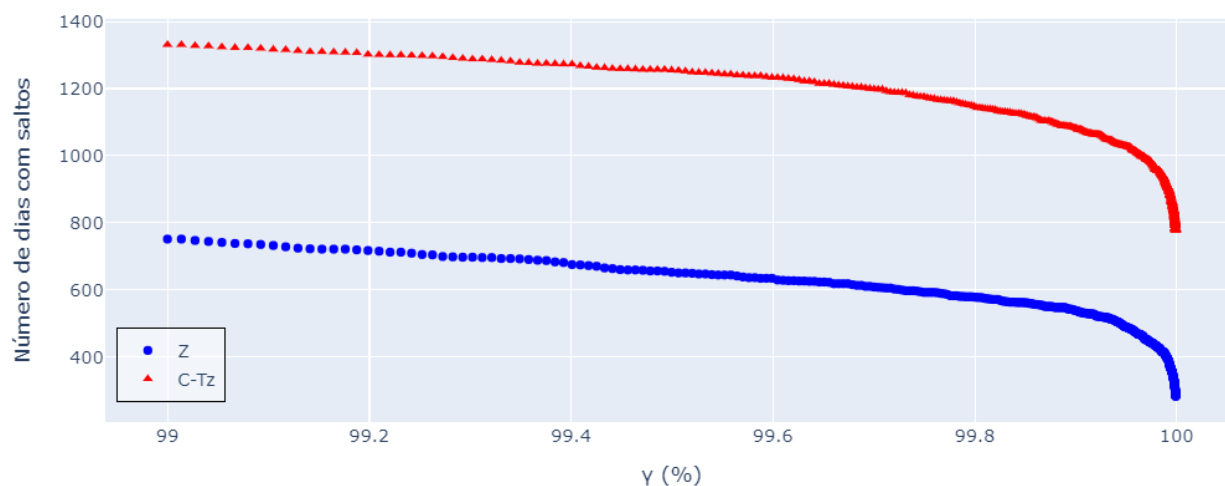


Figura B.4 Número de dias com saltos, de acordo com o nível de confiança γ e a estatística de teste, para o Bitcoin (BTC)

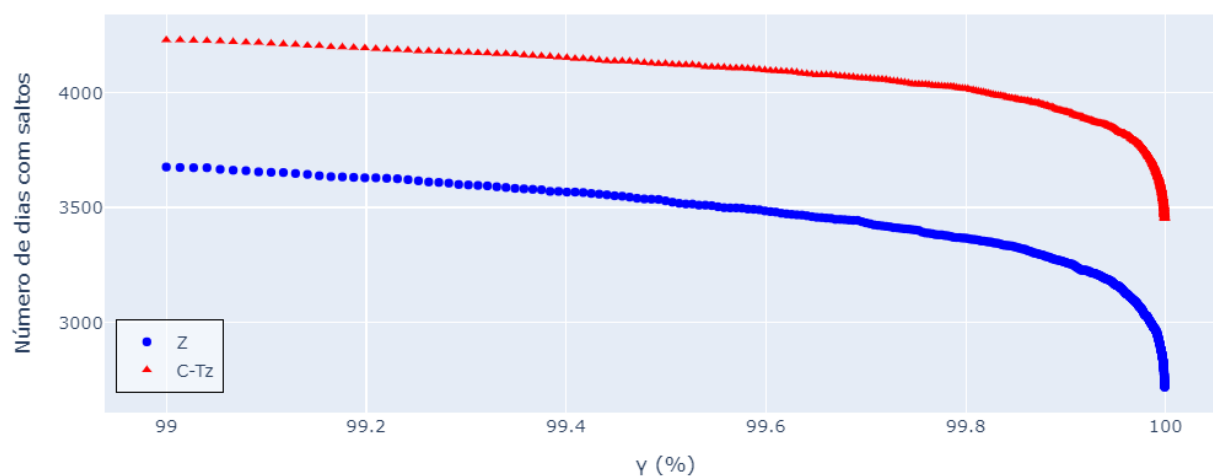


Figura B.5 Número de dias com saltos, de acordo com o nível de confiança γ e a estatística de teste, para a Amazon (AMZN)

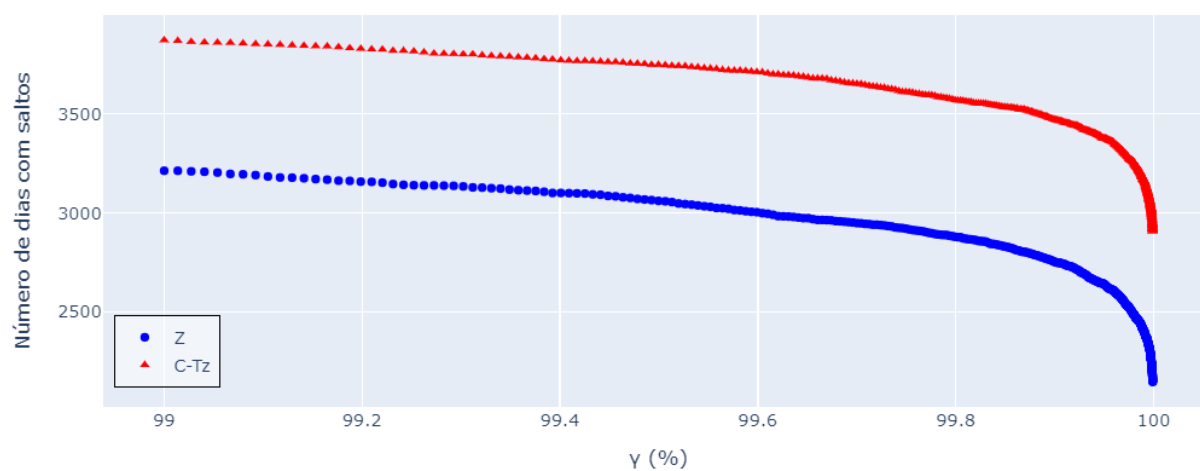


Figura B.6 Número de dias com saltos, de acordo com o nível de confiança γ e a estatística de teste, para a Apple (AAPL)

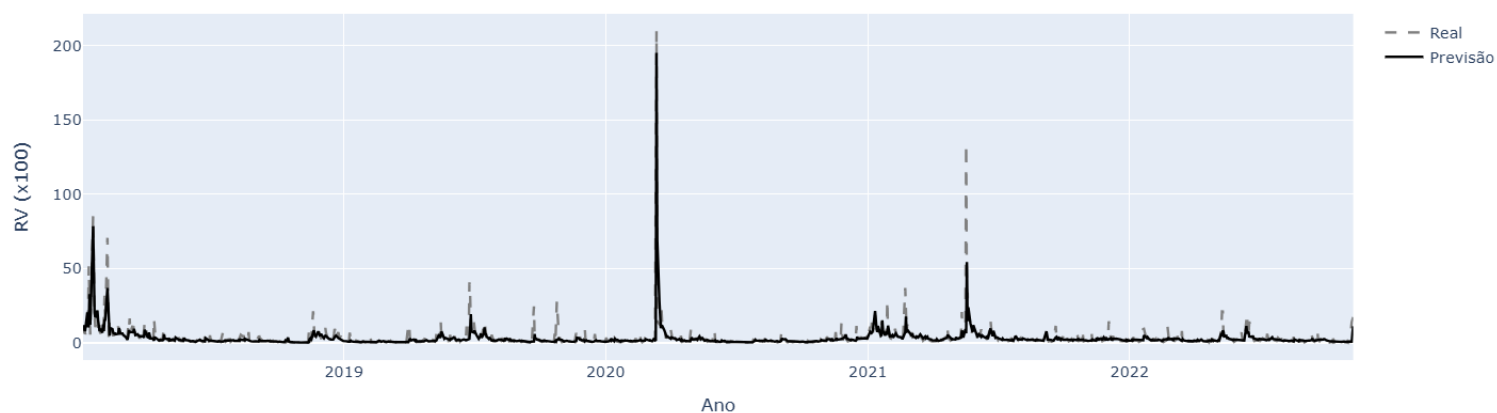


Figura B.7 Previsões *in-sample* de RV , pela transformação logarítmica, do modelo MLHAR-TCJ para o Bitcoin (BTC)

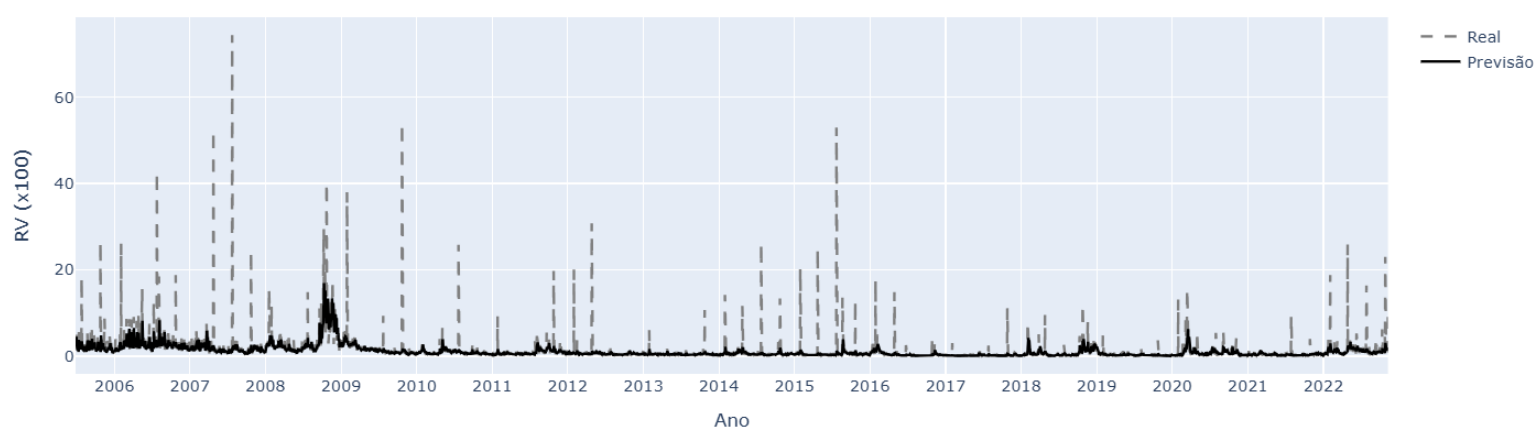


Figura B.8 Previsões *in-sample* de RV , pela transformação logarítmica, do modelo HAR-CJ para a Amazon (AMZN)

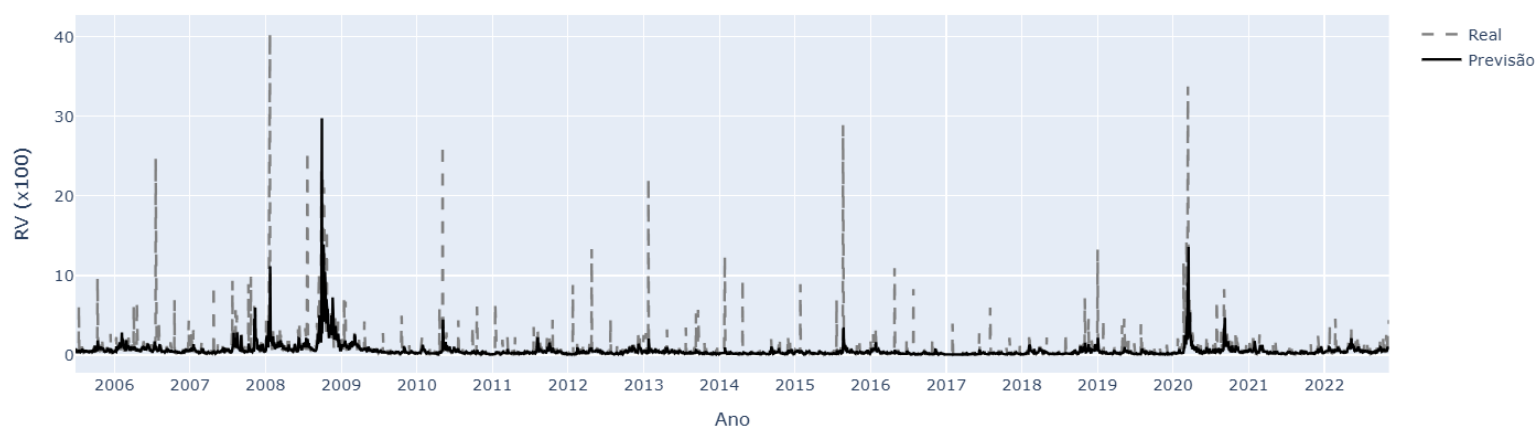


Figura B.9 Previsões *in-sample* de RV , pela transformação logarítmica, do modelo LHAR-TCJ para a Apple (AAPL)

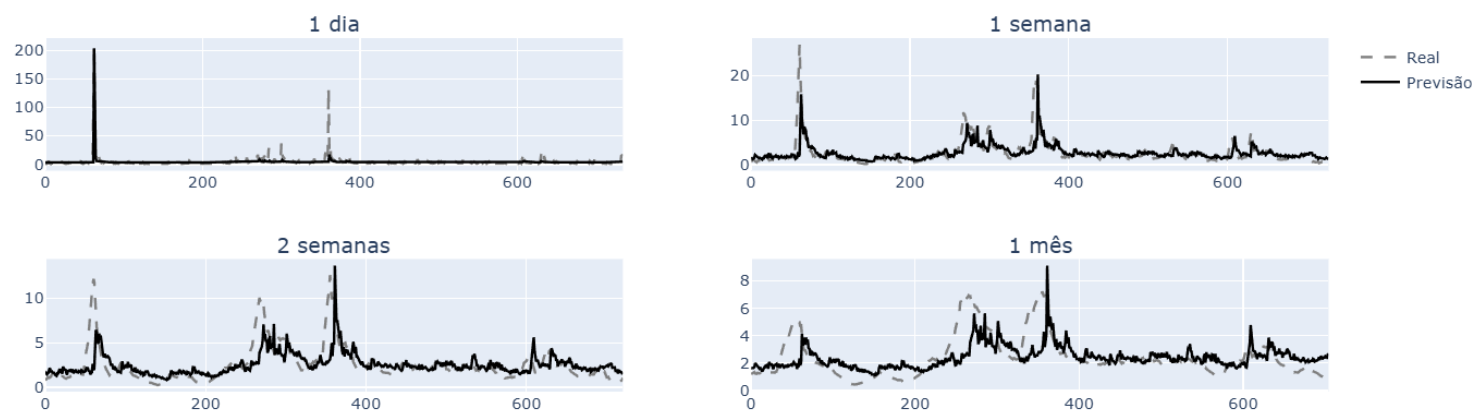


Figura B.10 Previsões *out-of-sample* de *RV*, pela transformação logarítmica, do modelo MLHAR-TCJ para o Bitcoin (BTC)

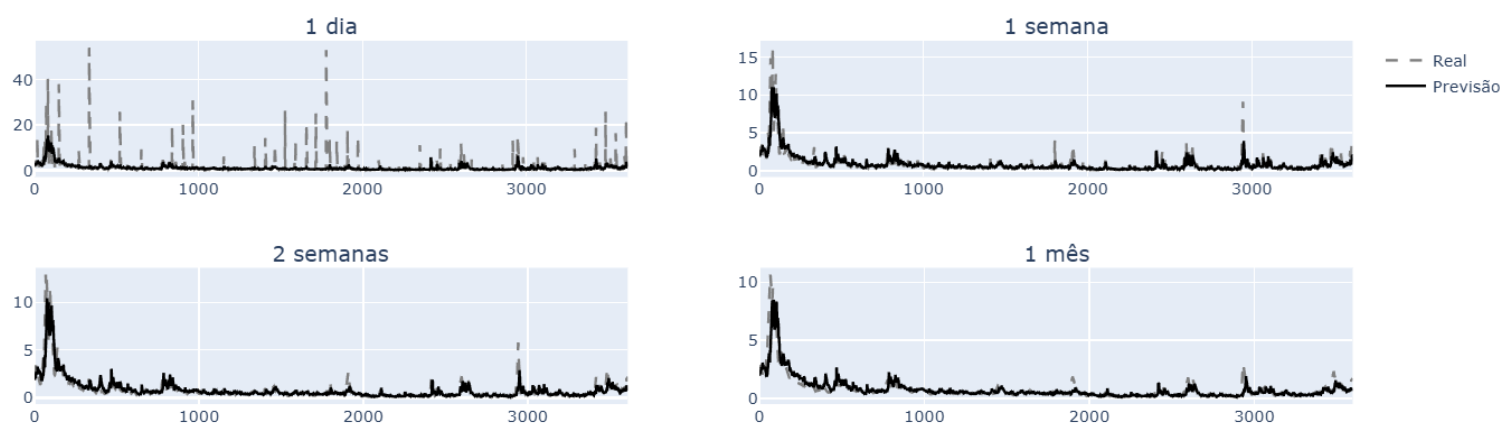


Figura B.11 Previsões *out-of-sample* de *RV*, pela transformação logarítmica, do modelo HAR-CJ para a Amazon (AMZN)

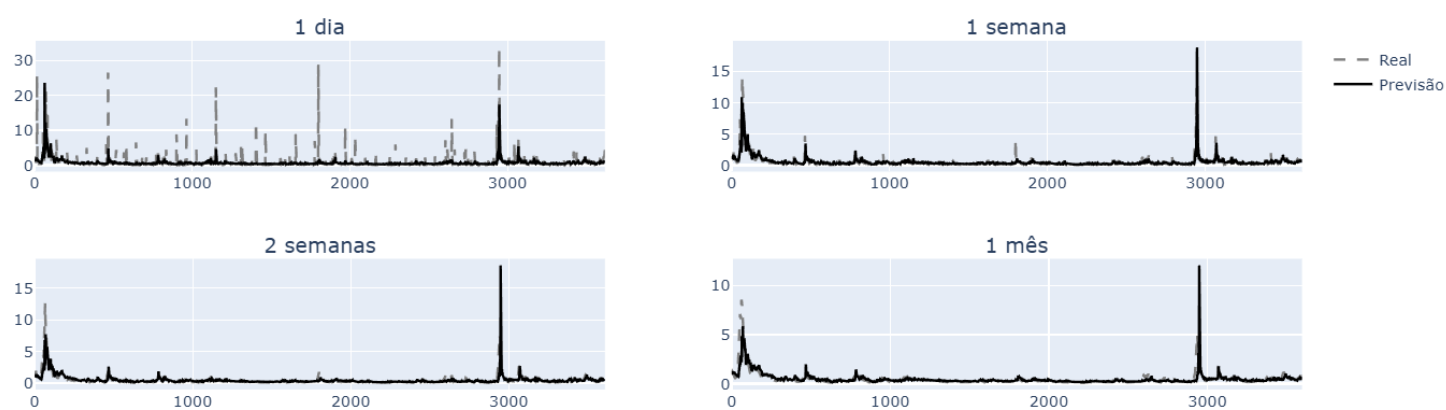


Figura B.12 Previsões *out-of-sample* de *RV*, pela transformação logarítmica, do modelo ULHAR-TCJ para a Apple (AAPL)

APÊNDICE C

Detecção de Saltos

Casos particulares de *multipower variation*, desenvolvidos em Barndorff-Nielsen e Shephard (2004):

- *BPV – Bipower variation*:

$$BPV_t = \mu_1^{-2} \sum_{i=2}^n |r_{i-1,t}| \cdot |r_{i,t}|$$

com $\mu_1 \cong 0,7979$.

- *TriPV – Tripower variation*:

$$TriPV_t = \mu_{\frac{4}{3}}^{-3} \sum_{i=3}^n |r_{i-2,t}|^{\frac{4}{3}} \cdot |r_{i-1,t}|^{\frac{4}{3}} \cdot |r_{i,t}|^{\frac{4}{3}}$$

com $\mu_{\frac{4}{3}} \cong 0,8309$.

Medidas alternativas de *RV* com *threshold*:

- *TBPV – Threshold bipower variation*:

$$TBPV_t = \mu_1^{-2} \sum_{i=2}^n |r_{i-1,t}| \mathbf{1}_{(r_{i-1,t}^2 \leq \vartheta_{i-1,t})} \cdot |r_{i,t}| \mathbf{1}_{(r_{i,t}^2 \leq \vartheta_{i,t})}$$

- *C-TBPV – Corrected threshold bipower variation*:

$$C-TBPV_t = \mu_1^{-2} \sum_{i=2}^n Z_1(r_{i,t}; \vartheta_{i,t}) \cdot Z_1(r_{i-1,t}; \vartheta_{i-1,t})$$

- *C-TTriPV – Corrected threshold tripower variation*:

$$C-TTriPV_t = \mu_{\frac{4}{3}}^{-3} \delta^{-1} \sum_{i=3}^n Z_{\frac{4}{3}}(r_{i,t}; \vartheta_{i,t}) \cdot Z_{\frac{4}{3}}(r_{i-1,t}; \vartheta_{i-1,t}) \cdot Z_{\frac{4}{3}}(r_{i-2,t}; \vartheta_{i-2,t})$$

sendo

$$Z_{\zeta}(x, y) = \begin{cases} |x|^{\zeta} & \text{se } x^2 \leq y \\ \frac{1}{2N(-c_{\vartheta})\sqrt{\pi}} \left(\frac{2}{c_{\vartheta}^2} y\right)^{\frac{\zeta}{2}} \Gamma\left(\frac{\zeta+1}{2}; \frac{c_{\vartheta}^2}{2}\right) & \text{se } x^2 > y \end{cases}$$

E ϑ_t o *threshold* estocástico dado por:

$$\vartheta_t = c_{\vartheta}^2 \cdot \hat{V}_t$$

onde c_{ϑ} ($= 3$) é uma constante sem escala usada para mudar o limitador e \hat{V}_t uma estimativa auxiliar da variância local σ_t^2 calculada utilizando um filtro não-paramétrico de tamanho $2L + 1$ ($L = 25$), adaptado para saltos iterando em Z :

$$\hat{V}_t^Z = \frac{\sum_{i=-L, i \neq -1, 0, 1}^L K\left(\frac{i}{L}\right) r_{t+i}^2 \mathbf{1}_{(r_{t+i}^2 \leq c_v^2 \cdot \hat{V}_t^{Z-1})}}{\sum_{i=-L, i \neq -1, 0, 1}^L K\left(\frac{i}{L}\right) \mathbf{1}_{(r_{t+i}^2 \leq c_v^2 \cdot \hat{V}_t^{Z-1})}}, \quad Z = 1, 2, \dots$$

com $K(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$, um kernel gaussiano, $c_v = 3$ e $\hat{V}^0 = +\infty$. Aceitamos a convergência do método quando a máxima diferença absoluta entre 2 vetores seguidos da variância local é menor do que 10^{-6} .