

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
DEPARTAMENTO DE ECONOMIA E RELAÇÕES INTERNACIONAIS**

TAINAN DE BACCO FREITAS BOFF

**COMPARAÇÃO DO DESEMPENHO DE MEDIDAS REALIZADAS PARA
PREVISÃO DE VOLATILIDADE DE AÇÕES DA B3**

Porto Alegre

2018

TAINAN DE BACCO FREITAS BOFF

**COMPARAÇÃO DO DESEMPENHO DE MEDIDAS REALIZADAS PARA
PREVISÃO DE VOLATILIDADE DE AÇÕES DA B3**

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia.

Orientador: Prof. Dr. João Fróis Caldeira

Porto Alegre

2018

CIP - Catalogação na Publicação

Boff, Tainan de Bacco Freitas
Comparação do desempenho de medidas realizadas
para previsão de volatilidade de ações da B3 / Tainan
de Bacco Freitas Boff. -- 2018.
50 f.
Orientador: João Fróis Caldeira.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Faculdade
de Ciências Econômicas, Curso de Ciências Econômicas,
Porto Alegre, BR-RS, 2018.

1. Dados de alta frequência. 2. Volatilidade
realizada. 3. Previsão de volatilidade. 4. HAR-RV.
5. Model Confidence Set. I. Caldeira, João Fróis,
orient. II. Título.

TAINAN DE BACCO FREITAS BOFF

**COMPARAÇÃO DO DESEMPENHO DE MEDIDAS REALIZADAS PARA
PREVISÃO DE VOLATILIDADE DE AÇÕES DA B3**

Trabalho de conclusão submetido ao Curso de Graduação em Ciências Econômicas da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título Bacharel em Economia.

Orientador: Prof. Dr. João Fróis Caldeira

Aprovado em: Porto Alegre, 29 de junho de 2018.

BANCA EXAMINADORA:

Prof. Dr. João Fróis Caldeira – Orientador

UFRGS

Prof. Dr. Fernando Augusto Boeira Sabino da Silva

UFRGS

Prof. Dr. Marcelo Savino Portugal

UFRGS

RESUMO

Este trabalho compara previsões de volatilidade dos preços de ações geradas com base em dados de alta frequência. O objetivo principal é fornecer orientação para a escolha da medida realizada e da frequência amostral a ser utilizada em aplicações na área de econometria financeira com dados da bolsa de valores B3. A hipótese básica é a de que a variância realizada combinada com uma frequência amostral de 5 minutos é a opção mais adequada. Inicialmente, é feito o tratamento dos dados para a obtenção de séries homogêneas de retornos financeiros, sem quaisquer distorções que não reflitam movimentos do mercado. Em seguida, são calculadas 7 classes de medidas realizadas em 8 frequências amostrais. Estas estimativas são combinadas ao modelo preditivo HAR-RV e a comparação das previsões é realizada através do procedimento *Model Confidence Set*. Os resultados sugerem que a variância realizada de 5 minutos é superada por estimadores mais sofisticados, em especial o estimador em duas escalas de tempo com correção de viés e o estimador em duas escalas de tempo robusto. Além da medida realizada, a frequência amostral também parece ter um papel relevante no desempenho preditivo: frequências mais altas geram melhores previsões de volatilidade para ativos mais líquidos.

Palavras-chave: Dados de alta frequência. Volatilidade realizada. Previsão de volatilidade. HAR-RV. *Model Confidence Set*.

ABSTRACT

This work compares volatility forecasts of stock prices based on high frequency data. The main goal is to provide guidance on the choice of a realized measure and a sampling frequency to be used in applications in the area of financial econometrics using data from the B3 stock exchange. The basic hypothesis is that the realized variance combined with a sampling frequency of 5 minutes is the ideal option. Initially, it is necessary to clean and aggregate the price data in order to obtain homogeneous series of financial returns, without any distortions that do not reflect market fluctuations. After that, 7 classes of realized measures are estimated based on 8 sampling frequencies. These estimates are combined with the predictive model HAR-RV and the comparison of the predictions is performed through the Model Confidence Set procedure. The results suggest that the 5-minute realized variance is outperformed by more sophisticated estimators, especially the two time scale variance with bias correction and the robust two time scale variance. In addition to the choice of realized measure, the sampling frequency also seems to play a relevant role in predictive performance: higher frequencies generate better predictions of volatility for more liquid assets.

Keywords: High-frequency data. Realized volatility. Volatility forecast. HAR-RV. Model Confidence Set.

SUMÁRIO

1.	INTRODUÇÃO	6
2.	REVISÃO DA LITERATURA	12
3.	METODOLOGIA.....	15
3.1.	TRATAMENTO DOS DADOS	16
3.1.1.	Limpeza	16
3.1.2.	Ajuste dos preços devido a eventos corporativos.....	17
3.1.3.	Cálculo dos retornos intradiários	18
3.2.	MEDIDAS REALIZADAS	19
3.2.1.	Variância realizada	20
3.2.2.	Estimador em duas escalas de tempo.....	20
3.2.3.	Estimador em duas escalas de tempo com correção de viés.....	21
3.2.4.	Varição realizada bipotente	22
3.2.5.	MedRV e minRV	23
3.2.6.	Estimador em duas escalas de tempo robusto	23
3.3.	PREVISÃO DA VOLATILIDADE: O MODELO HAR-RV	24
3.4.	COMPARAÇÃO DAS PREVISÕES: MODEL CONFIDENCE SET	25
3.4.1.	Proxy de volatilidade e função perda	27
4.	RESULTADOS EMPÍRICOS.....	29
5.	CONSIDERAÇÕES FINAIS	32
	REFERÊNCIAS.....	35
	APÊNDICE A – ESTATÍSTICAS DESCRITIVAS	40
	APÊNDICE B – GRÁFICOS DAS PREVISÕES.....	43
	APÊNDICE C – BOXPLOTS DAS PERDAS	47

1. INTRODUÇÃO

Ao discursar sobre o papel do setor financeiro como motor do desenvolvimento econômico, a economista Anne Krueger declarou que à medida em que as economias se desenvolvem, o sistema financeiro torna-se cada vez mais importante como facilitador do crescimento (se está desempenhando suas funções e se desenvolvendo com o resto da economia) ou como um inibidor do crescimento (se continuar subdesenvolvido) (KRUEGER, 2006). A função mais relevante do mercado financeiro é promover uma alocação eficiente dos recursos, canalizando excedentes monetários de poupadores para investidores. Naturalmente, essa transferência só ocorre porque há vantagens para todos os envolvidos: os poupadores têm a oportunidade de receber retornos atrativos, os tomadores têm acesso aos recursos necessários aos seus projetos de investimento e os intermediários são remunerados através de taxas e comissões.

Um importante segmento do mercado financeiro é o mercado de capitais, o qual funciona como um mecanismo de financiamento de médio e longo prazo. No mercado primário, as companhias captam recursos via emissão e oferta pública de ações, debêntures e outros títulos privados. Já no mercado secundário, os investidores negociam entre si os valores mobiliários, conferindo-lhes liquidez. A instituição central do mercado de capitais é a bolsa de valores. No Brasil, existe uma única bolsa de valores em operação: a B3 - Brasil, Bolsa, Balcão. A fim de ilustrar a magnitude dos recursos financeiros movimentados através dela, cabe apresentar alguns dados recentes. Em maio de 2018, a B3 foi a sexta maior bolsa global em termos de valor de mercado (R\$45 bilhões). A capitalização de mercado, ou seja, o valor total das ações listadas, foi de R\$3,2 trilhões e o valor médio negociado diariamente no mercado de ações foi de R\$14,2 bilhões. Metade (50%) das transações do segmento Bovespa, em termos de volume, foram realizadas por investidores estrangeiros, 28% por investidores institucionais, 17% por pessoas físicas e 4% por instituições financeiras (B3, 2018).

Uma característica do mercado de capitais é a sua ampla transparência. As informações são divulgadas em alta velocidade e podem ser recebidas da mesma forma por todos os participantes do mercado. O preço das ações varia constantemente refletindo os sentimentos e as expectativas dos investidores sobre o desempenho

financeiro das companhias. Aliás, uma peculiaridade deste mercado é sua sensibilidade a fatores psicológicos e emocionais, constituindo uma considerável fonte de oscilação dos preços. A hipótese do mercado eficiente, conforme definida por Fama (1970), postula que os preços das ações refletem instantaneamente as informações disponíveis¹. Desta forma, os agentes seriam incapazes de obter retornos superiores à média do mercado de forma consistente. Na prática, os *traders* (ou especuladores) empregam uma série de ferramentas para identificar oportunidades de ganhos, comprando e vendendo ações em janelas relativamente curtas de tempo. Um caso extremo é o das negociações de alta frequência, em que as transações de compra e venda são realizadas em intervalos de poucos segundos ou até microssegundos. A variação dos preços é conhecida como volatilidade. Se, para muitos participantes do mercado ela é vista com temor, por ser sinônimo de risco, para os *traders* ela é parte crucial da estratégia de ganho.

Mensurar e prever a volatilidade das ações de forma precisa tem diversas utilidades práticas para os investidores: esta informação é crucial, por exemplo, para a precificação de ativos e seleção de carteiras ótimas. A precificação de ativos está assentada no binômio risco-retorno, segundo o qual ativos financeiros mais voláteis devem receber maior prêmio pelo risco. Já a otimização de carteiras é uma forma de reduzir a exposição ao risco através da diversificação dos investimentos. De acordo com a teoria do portfólio de Markowitz (1952), dado um conjunto de alternativas, é possível encontrar uma combinação ponderada de ativos capaz de maximizar o retorno esperado dado um nível de risco ou minimizar o risco dado um nível de retorno esperado.

A academia tem muito a contribuir para o desenvolvimento do mercado de capitais através da proposição de modelos sofisticados para a previsão de volatilidade. Evidentemente, dada a profusão de modelos teóricos, surge a necessidade de testá-los empiricamente e comparar o seu desempenho. Este é o papel da econometria aplicada na área de finanças. Os pesquisadores desta área procuram compreender o funcionamento e a dinâmica do mercado (ou de algum ativo ou conjunto de ativos em

¹ A hipótese do mercado eficiente tem três versões: fraca, semiforte e forte. Em sua forma fraca, diz que os preços refletem informações históricas (preços passados). Em sua forma semiforte, considera que os preços refletem toda a informação publicamente disponível. Por fim, em sua forma forte, assume que os preços refletem toda informação existente, inclusive informações privilegiadas

especial) com o objetivo último de possibilitar a identificação de oportunidades de investimento ou meios para controlar a exposição ao risco. Para isso, extraem informações de séries temporais de preços de ativos financeiros através da utilização de ferramentas estatísticas e matemáticas.

Em estudos que envolvem a variabilidade dos preços, é preciso usar uma variável *proxy*, visto que a volatilidade em si não é observável. Tradicionalmente, utiliza-se o quadrado dos retornos² diários. Porém, em anos recentes, a crescente disponibilidade de dados financeiros em alta frequência motivou o desenvolvimento de uma área de pesquisa dedicada à construção de *proxies* alternativas, conhecidas como medidas realizadas. Através das medidas realizadas é possível estimar volatilidades a partir de preços observados em intervalos intradiários, chegando ao limite de haver um registro para cada evento (transação ou oferta de compra/venda). Em teoria, as medidas realizadas são mais precisas e contêm mais informações sobre a volatilidade quando comparadas aos (tradicionalmente empregados) retornos quadráticos diários.

O presente trabalho propõe-se a identificar quais medidas realizadas e quais frequências amostrais intradiárias geram melhores previsões de volatilidade para ações negociados na bolsa de valores B3 - Brasil, Bolsa, Balcão. O objetivo da pesquisa é fornecer orientação para a escolha da medida realizada a ser utilizada em aplicações na área de econometria financeira. Para tanto, a análise será implementada em três etapas: (a) prever a volatilidade dos preços de ações um dia à frente utilizando as medidas realizadas descritas na seção 3.2 em frequências amostrais de 1, 5, 10, 15, 30, 45, 60 e 90 minutos; (b) estimar o *Model Confidence Set* de Hansen, Lunde e Nason (2011); e (c) comparar o desempenho das medidas realizadas e frequências amostrais. Serão utilizadas séries de preços para duas ações com diferentes níveis de liquidez: VALE3 e BBDC3. A análise levará em consideração dados para o período compreendido entre 3 de novembro de 2009 e 17 de fevereiro de 2017, o que equivale a 1,793 dias.

A hipótese básica é a de que a variância realizada combinada com uma frequência amostral de 5 minutos é a opção mais adequada para fins de previsão de volatilidade.

² Os retornos financeiros podem ser calculados como a diferença entre o logaritmo dos preços nos tempos t e $t - 1$.

Esta hipótese tem respaldo teórico à medida em que a variância realizada é o equivalente amostral da variável alvo que desejamos prever: a variação quadrática de uma série temporal de retornos financeiros. Além disso, na ausência de ruídos nos dados, a variância realizada é o estimador não paramétrico de máxima verossimilhança, logo, é eficiente. A presença de ruídos nos dados tem relação direta com a frequência em que os preços são amostrados e resulta na obtenção de uma estimativa viesada. Para evitar, ou ao menos reduzir, o nível de ruído nos dados, uma alternativa é a diminuição da frequência amostral. Obviamente, isso não deve comprometer os ganhos de informação que se espera obter quando se utilizam dados intradiários. Por isso, é uma escolha comum entre pesquisadores utilizar a frequência de 5 minutos.

A quantidade de pesquisas que envolvem a utilização de dados de alta frequência é crescente, apesar de ainda esbarrar na disponibilidade deste tipo de dados. A primeira dificuldade está relacionada à sua obtenção, já que, em geral, as bases não são publicamente acessíveis. A segunda está relacionada à capacidade computacional que é exigida para a transferência, o armazenamento e a manipulação do grande volume de dados. No caso brasileiro, no que diz respeito ao acesso aos dados, a B3 disponibiliza, via FTP, cotações intradiárias para ações negociadas nos mercados à vista, a termo, de opções e futuros, índices, renda fixa corporativa, derivativos financeiros e agropecuários, câmbio e títulos públicos federais. São disponibilizados dados relativos aos dois anos mais recentes, porém é possível ter acesso a registros mais antigos sob demanda. No que diz respeito à sua manipulação, recentemente foi publicado um pacote para o *software* estatístico R (GetHFData³) que auxilia no *download* e agregação por intervalos temporais. Com a facilitação do acesso aos dados, espera-se um aumento do número de pesquisas nesta área.

A utilização de medidas realizadas envolve um grande número de decisões que, em geral, são feitas *ad hoc*: qual medida realizada utilizar entre as diversas propostas na literatura? Que frequência amostral (30 segundos, 1 minuto, 5 minutos, 1 hora, ...) deve ser escolhida para calcular estas medidas? De que forma amostrar os dados intradiários, a cada x segundos ou a cada x transações? Quais preços utilizar, os de ordens de

³ Ramos & Perlin (2016)

compra e venda ou somente aqueles relativos a transações efetivadas? Estas escolhas são as mesmas para diferentes classes de ativos? Entre outras questões.

Para análises de mercados financeiros internacionais, existem estudos que auxiliam nestas decisões. Em especial, o trabalho de Liu, Patton e Sheppard (2015) procura fornecer respostas para estas questões utilizando dados americanos e britânicos. Uma medida muito popular na literatura é a variância realizada com frequência amostral de 5 minutos. Essa medida alia a simplicidade de cálculo com uma frequência amostral que é alta, mas não a ponto de gerar vieses associados à microestrutura de mercado. O artigo supramencionado compara o desempenho desta medida com o de outras centenas de opções e conclui que, em geral, é difícil superá-la. Além disso, os autores identificam que medidas realizadas mais sofisticadas, em geral, apresentam um desempenho significativamente pior em mercados que não são o norte-americano.

No Brasil, um reduzido número de trabalhos fornece algumas evidências a respeito do desempenho de diferentes medidas realizadas e frequências amostrais. Porém, não existe nenhuma investigação sistemática que sirva como um guia para a tomada de decisão. Por este motivo, o presente trabalho visa preencher esta lacuna na literatura. Em geral, as pesquisas com dados intradiários da bolsa de valores brasileira apoiam-se sobre a hipótese de que a utilização de frequências amostrais mais altas (relativamente à frequência diária, que é amplamente utilizada em finanças) resultará em um melhor desempenho das previsões ou estratégias de investimento. Contudo, os resultados obtidos têm sido mistos: alguns estudos corroboram e outros refutam essa hipótese. Ocorre que, na falta de um estudo como o que está sendo aqui proposto, os pesquisadores fazem suas escolhas baseados na literatura para dados internacionais (em especial, dados americanos). Alguns dos trabalhos⁴ que testaram sua hipótese através da utilização de um maior número de medidas realizadas e frequências amostrais sugeriram que o bônus de desempenho associado ao uso de dados intradiários pode estar justamente relacionado a estas escolhas.

Os resultados do presente estudo apontam para a rejeição da hipótese de que a variância realizada combinada à frequência amostral de 5 minutos é a opção mais

⁴ Ver, por exemplo, Borges, Caldeira, & Ziegelmann (2015).

acertada para fins de previsão de volatilidade. Essa combinação foi eliminada do conjunto que contém o melhor estimador a um nível de confiança de 95%. Em geral, tanto a frequência amostral quanto a escolha da medida realizada parecem ser relevantes para a qualidade da previsão. Para ativos mais líquidos, as frequências mais altas apresentam melhor desempenho. Com relação aos estimadores, o estimador em duas escalas de tempo com correção de viés de Zhang, Mykland e Ait-Sahalia (2005) e o estimador em duas escalas de tempo robusto a saltos de Boudt e Zhang (2015) mostraram-se superiores, especialmente para o ativo mais líquido.

O trabalho está organizado da seguinte forma: o capítulo 2 revisa a literatura relevante e apresenta seus principais resultados. O capítulo 3 trata da metodologia. Ele inicia expondo o tratamento dos dados, que não é uma tarefa trivial quando se manipulam dados de preços em alta frequência. Em seguida, expõe as medidas realizadas, o modelo preditivo HAR-RV e o método utilizado para comparar as previsões: *Model Confidence Set*. O capítulo 4 apresenta os resultados do estudo e, finalmente, o capítulo 5 apresenta as considerações finais.

2. REVISÃO DA LITERATURA

Os estudos comparativos entre diversas medidas realizadas podem ser divididos entre os de caráter teórico e empírico. Entre as contribuições teóricas mais recentes, McAleer e Medeiros (2008) fazem uma ampla revisão da literatura sobre volatilidade realizada nos casos univariado e multivariado, analisando os impactos da presença de ruídos de microestrutura nas séries e fazendo uma exposição crítica de diferentes técnicas de estimação e previsão propostas na literatura. Bandi e Russell (2006) revisam a literatura com foco em processos de preços contaminados por ruídos de microestrutura. Barndorff-Nielsen e Shephard, (2007) tratam da estimação não paramétrica de volatilidade com foco principal em processos sem fricção, com ou sem efeitos de saltos.

Entre as contribuições empíricas, o trabalho de Liu, Patton e Sheppard (2015) merece destaque. Neste artigo, comparam-se oito classes de medidas realizadas em termos do seu desempenho na estimação da variação quadrática e doze classes de medidas realizadas em termos da acurácia na previsão, em diferentes horizontes temporais, quando combinadas ao modelo HAR-RV. Cada classe de medida realizada é estimada utilizando-se diferentes frequências amostrais e métodos de amostragem. Além disso, as aplicações envolvem 31 ativos de 5 diferentes classes. Quando a variância realizada de 5 minutos é tomada como *benchmark*, os autores encontram poucas evidências de que ela possa ser superada por outras medidas. Por outro lado, ao utilizarem métodos de inferência que não requerem o uso de um *benchmark*, eles concluem que a variância realizada de 5 minutos é superada, em média, por alternativas mais sofisticadas (quando bem escolhidas). Um ponto relevante deste estudo é que, ao utilizarem dados americanos e britânicos, os autores verificam que as medidas realizadas mais sofisticadas geralmente apresentam um desempenho significativamente pior no mercado britânico. Segundo eles, isso é potencialmente indicativo de diferentes efeitos de microestrutura, que podem ser tratados mais adequadamente com novas abordagens. Para comparar a acurácia dos estimadores de variação quadrática, os autores utilizam os métodos indicados por Patton (2011a). Este último artigo também faz um estudo empírico de classificação de estimadores da variação quadrática envolvendo uma série de preços da ação da IBM. O autor compara a variância realizada com quatro

estimadores mais sofisticados. Para o cálculo do estimador mais simples, a amostragem é baseada tanto em cotações de ofertas de compra e venda como em preços de negociação. São utilizadas diversas frequências amostrais, tanto em termos de tempo-calendário – segundos/minutos – quanto em termos de *ticks* (amostragem a cada x transações). Os resultados indicam que estimadores baseados em retornos amostrados em intervalos que vão de 15 segundos a 2 minutos são significativamente mais precisos do que a variância realizada de 5 minutos. Além disso, alguns dos estimadores de variação quadrática mais sofisticados apresentam desempenho significativamente superior ao da variância realizada. Outra contribuição interessante é a de Cardoso (2016). Este autor simula processos univariados e multivariados e estuda a interferência de diferentes fontes de perturbação no preço - ruídos de microestrutura, saltos e *outliers*, individualmente e simultaneamente - sobre cada estimador de medida realizada, utilizando diferentes frequências amostrais. Este trabalho também apresenta aplicações usando dados de taxas cambiais. No caso univariado são feitos diagnósticos de estimação. No caso multivariado são avaliados os desempenhos de carteiras igualmente ponderadas e de mínima variância. Entre as conclusões do trabalho, Cardoso identifica que, na ausência de perturbações, a variância realizada apresenta os melhores resultados, e quanto maior a frequência utilizada, melhor seu desempenho. No entanto, conforme esperado, ao adicionar perturbações ao processo, seu desempenho passa a ser o pior dentre as medidas estudadas, sendo, portanto, não robusta a nenhuma contaminação nos preços. Sob presença de ruídos de microestrutura, as medidas realizadas têm melhores resultados quando calculadas a partir de retornos tomados a cada 5 minutos, conforme indicado por Andersen e Bollerslev (1998). Ghysels e Sinko (2011) fazem previsão de volatilidade utilizando a abordagem de regressão MIDAS (*Mixed Frequency Data Sampling*) com diversas medidas realizadas e frequências amostrais. Assim como Zhang, Mykland e Ait-Sahalia (2012) e Ait-Sahalia e Mancini (2008), concluem que o estimador em duas escalas de tempo com correção de viés superam a variância realizada.

Com relação à frequência amostral, Ait-Sahalia, Mykland e Zhang (2005) concluem que, quando o ruído de microestrutura de mercado está presente, mas não é considerado, a frequência ótima de amostragem está entre 1 e 5 minutos. Porém, de

acordo com Bauwens, Hafner e Laurent (2012), modelar o ruído e usar todos os dados deve gerar melhores resultados. Griffin e Oomen (2008) observam que, quando a amostragem é realizada a cada x transações (em contraposição à amostragem feita a cada x minutos), os retornos apresentam menos correlação serial e os ruídos de microestrutura ficam mais próximos de serem independentes e identicamente distribuídos. Por isso, as medidas que incorporam correção de viés baseadas na hipótese de ruído i.i.d. funcionam melhor com este tipo de amostragem. Patton (2011a) identifica que, em geral, usar dados amostrados a cada x transações gera melhores estimadores da variação quadrática, particularmente quando as transações são espaçadas de forma muito irregular.

Na literatura empírica com dados do mercado de ações brasileiro, Borges, Caldeira e Ziegelmann (2015) comparam a contribuição de diferentes medidas realizadas para a otimização de carteiras de variância mínima. Trata-se de uma análise multivariada na qual são estimadas matrizes de covariância utilizando diferentes frequências amostrais intradiárias, além de um estimador que possibilita o uso de dados amostrados de forma não sincronizada. Os resultados do estudo indicam que a frequência de 5 minutos gera carteiras com menores custos de transação, enquanto frequências mais baixas apresentam melhores resultados em termos de risco (120 minutos) e retorno (frequência diária). Assim como em Pooter, Martens e Dijk (2008), os autores concluem que utilizar uma frequência amostral apropriada parece ser mais relevante do que usar medidas robustas aos ruídos de microestrutura de mercado. Uma pesquisa semelhante é apresentada em Caldeira, Moura, Perlin e Santos (2017). Neste artigo, a frequência amostral de 5 minutos gera as carteiras com menores risco e *turnover*. Ambos os artigos supramencionados, além de compararem medidas realizadas e frequências amostrais, investigam as potenciais vantagens de incorporar dados de alta frequência em aplicações na área de finanças. Espera-se que novos estudos nesta área de pesquisa sejam beneficiados pelo esforço aqui empreendido, uma vez que a escolha do estimador de volatilidade e da frequência amostral tem impacto sobre os resultados econômicos.

3. METODOLOGIA

A pesquisa aqui proposta é de natureza aplicada, à medida em que pretende gerar conhecimentos para aplicação prática - dirigidos à solução de um problema específico - que é a previsão da volatilidade de ativos do mercado de capitais brasileiro. Quanto aos objetivos, trata-se de pesquisa exploratória, na qual procura-se testar a hipótese de que a variância realizada de 5 minutos tem um desempenho preditivo superior ao de outras medidas realizadas através de procedimentos estatísticos. A abordagem é quantitativa, sendo a análise dos dados e as conclusões baseadas em métodos econométricos. As ferramentas estatísticas permitem utilizar uma amostra de dados para fazer inferências sobre a população, o que configura o método indutivo, ou seja, parte-se do específico para o geral. Entre os procedimentos adotados, utilizam-se pesquisa bibliográfica e documental. A pesquisa bibliográfica apoia-se em artigos científicos publicados nas principais revistas nacionais e internacionais da área de econometria financeira. A pesquisa documental baseia-se nos registros de preços de ativos e eventos corporativos da bolsa de valores.

A população de interesse são as ações negociadas no mercado à vista da B3. A amostra é do tipo intencional, e consiste em observações de preços de transação *tick by tick* das ações ordinárias da Vale S.A. (VALE3) e do Banco Bradesco S.A. (BBDC3) para o período que vai de 3 de novembro de 2009 a 17 de fevereiro de 2017. A determinação deste período não segue critérios estatísticos, mas está relacionada à disponibilidade de dados. Apesar disso, possibilita a construção de uma série temporal contendo 1.793 dias, que pode ser considerada longa o suficiente para a realização das análises propostas. As ações diferem entre si em termos de liquidez: no período considerado, VALE3 e BBDC3 apresentaram transações em 96% e 73% dos intervalos de 1 minuto, respectivamente. O percentual de retornos não nulos neste mesmo intervalo foi de 77% para VALE3 e 56% para BBDC3. A utilização de dados de alta frequência em aplicações empíricas não é trivial e requer uma série de procedimentos. A seção 3.1 descreve as etapas de limpeza e agregação dos dados, ajuste dos preços para eventos corporativos e o cálculo dos retornos.

A partir da obtenção dos retornos intradiários em diferentes frequências amostrais, é possível dar início ao cálculo das medidas realizadas, as quais serão utilizadas para prever a volatilidade dos papéis através do modelo preditivo HAR-RV. A seção 3.2 descreve a metodologia de cálculo das medidas realizadas, a seção 3.3 apresenta o modelo preditivo HAR-RV e a seção 3.4 encerra o capítulo com a exposição do método *Model Confidence Set*, que será utilizado para comparar o desempenho das medidas realizadas. Todas as análises serão conduzidas através da utilização do *software R*, com o auxílio dos pacotes:

- a) xts de Ryan, Ulrich e Bennett (2018);
- b) dplyr de Wickham, François, Henry, Müller e RStudio (2018);
- c) highfrequency de Boudt, Cornelissen e Payseur (2018);
- d) GetHFDData de Ramos e Perlin (2016);
- e) MCS de Catania e Bernardi (2017).

3.1. TRATAMENTO DOS DADOS

3.1.1. Limpeza

Engle (2000) se refere ao tipo de dados que utilizamos neste trabalho como “dados de ultra alta frequência”. Trata-se de uma série temporal de preços de ações na qual todas as transações são registradas. De acordo com Barndorff-Nielsen, Hansen, Lunde e Shephard (2009) a limpeza cuidadosa é um dos aspectos mais importantes para a estimação da volatilidade quando se utiliza este tipo de dados. Neste trabalho, a limpeza é realizada através das seguintes etapas:

- a) descartar quaisquer transações que ocorram fora do horário de negociação;
- b) descartar transações canceladas;
- c) a fim de produzir séries temporais homogêneas, agregar os preços de acordo com o método de tempo-calendário, ou seja, montar grades de 1, 5, 10, 15, 30, 45, 60 e 90 minutos e selecionar o último preço dentro de cada intervalo;

- d) descartar datas em que o horário de negociação foi reduzido devido a falhas técnicas, feriados ou outras circunstâncias;
- e) na ausência de transações em alguns intervalos, preencher com o preço imediatamente anterior.

3.1.2. Ajuste dos preços devido a eventos corporativos

Existe uma série de eventos corporativos que, apesar de não terem impacto sobre a posição financeira dos investidores, afetam o preço e a quantidade das ações sob sua custódia. Alguns exemplos são os dividendos, juros sobre capital próprio, bonificações, subscrições de capital, desdobramentos e agrupamentos. Sempre que ocorre um evento deste tipo, as cotações anteriores (ou posteriores) a ele devem ser ajustadas. Caso contrário, o cálculo dos retornos acusará grandes variações de preços que não foram causadas por forças de mercado, emitindo, assim, sinais enganosos sobre a volatilidade dos ativos.

Para ajustar a série de preços, calcula-se um fator de ajuste - cuja fórmula varia de acordo com o tipo de evento - e multiplicam-se todas as cotações, desde o início da série até o dia anterior ao evento (data ex), por este fator. Isso permite que as cotações anteriores ao fato sejam comparáveis às cotações posteriores a ele. Neste trabalho, o cálculo dos fatores é feito de acordo com a metodologia descrita no manual do *software* Economática:

- a) dividendo: $F = 1 - D/Pu$;
- b) bonificação: $F = 1/(1 + b)$;
- c) desdobramento: $F = 1/d$;
- d) redução de Capital: $F = 1/(1 - r)$;
- e) grupamento: $F = g$;
- f) subscrição: $F = (Pu + s * S)/((1 + s) * Pu)$;
- g) cisão: $F = 1 - c/100$.

Onde,

F = fator de ajuste;

D = valor do dividendo;

P_u = preço original na última data "com" (data imediatamente anterior à data ex);

b = quantidade de ações novas recebidas por cada ação possuída;

d = quantidade de ações novas que substituirão cada ação antiga;

r = quantidade de ações canceladas para cada ação possuída;

g = quantidade de ações antigas que serão substituídas por cada ação nova;

S = preço da subscrição;

s = quantidade de ações novas ofertadas por cada ação possuída;

c = porcentagem que a parte cindida representava do valor de mercado da empresa antes da cisão.

As informações sobre eventos corporativos estão disponíveis no *site* <http://www.bmfbovespa.com.br>.

3.1.3. Cálculo dos retornos intradiários

Retornos logarítmicos são definidos como a diferença entre os logaritmos naturais dos preços nos tempos t e $t - 1$. O cálculo dos retornos logarítmicos intradiários usualmente é realizado no formato *open-to-close*, isto é, são considerados apenas preços dentro de um mesmo dia. Dito de outra forma, não é calculada a diferença entre os logaritmos do preço de abertura do dia t e do preço de fechamento do dia $t - 1$ (retornos *overnight*). A obtenção das séries de retornos intradiários é pré-requisito para a estimação das medidas realizadas.

A Tabela 1 apresenta as estatísticas descritivas dos retornos para cada ação e frequência amostral.

Tabela 1 - Estatísticas descritivas dos retornos intradiários

Frequência dos retornos	Média	Máximo	Mínimo	Desvio padrão	Assimetria	Curtose
<i>BBDC3</i>						
1 minuto	9.61E-08	0.0374	-0.0355	1.13E-03	-0.0372	22.04
5 minutos	-4.82E-07	0.0387	-0.0355	2.20E-03	0.0093	9.19
10 minutos	1.27E-06	0.0387	-0.0346	2.94E-03	-0.0090	6.95
15 minutos	3.38E-06	0.0387	-0.0346	3.45E-03	-0.0004	5.18
30 minutos	2.34E-05	0.0378	-0.0362	4.59E-03	0.0428	3.77
45 minutos	5.86E-05	0.0450	-0.0344	5.14E-03	0.0991	3.62
60 minutos	8.73E-05	0.0577	-0.0359	5.96E-03	0.1452	3.47
90 minutos	1.15E-04	0.0577	-0.0364	6.66E-03	0.0531	3.02
<i>VALE3</i>						
1 minuto	-2.31E-06	0.0326	-0.0356	1.13E-03	-0.0262	16.25
5 minutos	-1.18E-05	0.0514	-0.0404	2.33E-03	0.0996	14.12
10 minutos	-2.34E-05	0.0505	-0.0425	3.23E-03	0.0196	11.94
15 minutos	-3.68E-05	0.0535	-0.0635	3.89E-03	-0.0237	12.17
30 minutos	-6.07E-05	0.0559	-0.0580	5.46E-03	0.0055	9.37
45 minutos	-5.36E-05	0.0760	-0.0653	6.18E-03	0.0386	9.47
60 minutos	-6.68E-05	0.0932	-0.0892	7.18E-03	-0.0359	11.30
90 minutos	-1.38E-04	0.0742	-0.0772	8.21E-03	0.1754	6.86

Fonte: elaboração própria.

3.2. MEDIDAS REALIZADAS

Medidas realizadas são estimadores da variabilidade dos preços ao longo de algum horizonte temporal (por exemplo, um dia ou uma semana) que são construídos utilizando dados amostrados em períodos de tempo mais curtos (por exemplo, a cada 5 minutos). Em geral, essas medidas visam a estimar a variação quadrática ou a variação integrada de um processo de preços. Denote por $\{p_t\}$ o processo de logaritmo de preços em tempo contínuo. Assumindo que os preços no dia $t + 1$ foram observados n vezes, $\left\{p_{t+1/n}, \dots, p_{t+1-\frac{1}{n}}, p_{t+1}\right\}$, a variação quadrática do processo no período $t + 1$ é definida como:

$$VQ_{t+1} = \text{plim}_{n \rightarrow \infty} \sum_{i=1}^n r_{t+i/n}^2,$$

$$r_{t+i/n} = p_{t+i/n} - p_{t+(i-1)/n}.$$

3.2.1. Variância realizada

A medida realizada mais difundida na literatura é conhecida como variância realizada. Trata-se do análogo empírico da variação quadrática. Ela é definida como a soma do quadrado dos retornos a uma frequência arbitrária. A variância realizada é calculada conforme a equação abaixo:

$$VR_{t+1}^{(k)} = \sum_{i=1}^n \left(r_{t+i/n}^{(k)} \right)^2$$

onde n é o número de observações no período $t + 1$ resultantes da amostragem realizada em uma frequência k .

Andersen e Bollerslev (1998) alertam que à medida que se aumenta a frequência amostral, os ruídos de microestrutura de mercado tornam-se mais importantes. Os autores sugerem utilizar a frequência de 5 minutos.

3.2.2. Estimador em duas escalas de tempo

Uma alternativa proposta por Zhang, Mykland e Ait-Sahalia (2005) é a utilização de duas escalas de tempo (ou reamostragem). Por exemplo, para calcular o estimador em duas escalas de tempo com a frequência de 5 minutos é possível amostrar os retornos financeiros a cada 30 segundos, obtendo-se, assim, 10 séries diferentes de retornos com intervalos de 5 minutos entre si. Calcula-se, então, a variância realizada para cada uma das 10 séries e obtém-se a média entre elas. Esse procedimento permite aproveitar a maior quantidade de informação contida em frequências mais altas e reduz a variância do estimador. Apesar disso, não elimina o viés de microestrutura contido na variância realizada.

Seja k a frequência amostral dos dados – por exemplo, $k = 5$ minutos – e seja C o número de reamostragens realizadas – por exemplo, se cada amostra iniciar com 30 segundos de distância entre si, será possível obter 10 amostras - o estimador em duas escalas de tempo é calculado através da seguinte equação:

$$VR2E_{t+1}^{(k)} = \frac{1}{C} \sum_{i=1}^C VR_{t+1,i}^{(k)}$$

3.2.3. Estimador em duas escalas de tempo com correção de viés

Para eliminar o viés de microestrutura contido na variância realizada com reamostragem, Zhang (2011) propôs incluir na equação um termo de correção do viés:

$$VR2Evies_{t+1}^{(k)} = \frac{1}{C} \sum_{i=1}^C VR_{t+1,i}^{(k)} - 2\bar{n}\hat{\sigma}_u^2$$

onde \bar{n} é o número médio de observações ao longo do dia t utilizadas em cada uma das C estimativas e $\hat{\sigma}_u^2$ é estimado da mesma forma que a variância realizada, mas utilizando todos os dados disponíveis:

$$\hat{\sigma}_u^2 = \frac{1}{2N} \sum_{t=1}^T \sum_{i=1}^n r_{t+i/n}^2$$

onde n é o número total de retornos observados durante o dia t e $N = \sum_{t=1}^T n_t$ é o número total de retornos até o fechamento do dia T . Neste trabalho, esta medida realizada é calculada para frequências amostrais entre 5 e 30 minutos, devido a restrições relacionados ao tamanho da amostra.

3.2.4. Variação realizada bipotente

Um componente significativo da volatilidade é atribuído a variações abruptas dos preços que ocorrem em um curto espaço de tempo, em episódios aleatórios, conhecidas como saltos. As principais explicações para os saltos são geralmente atribuídas à reação do mercado a notícias inesperadas e à falta de liquidez (veja, por exemplo, Boudt e Petitjean (2014)).

Assumindo que os logaritmos de preços podem ser representados por um processo de difusão com saltos, tal com definido em Merton (1976), a variação quadrática para o processo de retornos cumulativos pode ser decomposta em um componente contínuo (variância integrada) e um componente induzido pelos saltos:

$$VQ_{t+1} = \text{plim}_{n \rightarrow \infty} \sum_{i=1}^n r_{t+i/n}^2 = \int_t^{t+1} \sigma^2(s) ds + \sum_{t < s \leq t+1} \kappa^2(s)$$

A ideia subjacente à construção de medidas realizadas robustas aos saltos é estimar apenas a parte contínua da variação quadrática. Neste sentido, Barndorff-Nielsen e Shephard (2004) propuseram a medida “variação realizada bipotente”, na qual a volatilidade é estimada através da soma cumulativa do produto de retornos absolutos adjacentes:

$$VRBP_{t+1}^{(k)} = \frac{\pi}{2} \left(\frac{n}{n-1} \right) \sum_{i=1}^{n-1} \left| r_{t+i/n}^{(k)} \right| \left| r_{t+(i+1)/n}^{(k)} \right|$$

De modo simplificado, a robustez aos saltos baseia-se na ideia de que, como os retornos cobrem intervalos muito pequenos, deve haver no máximo um salto entre dois retornos adjacentes. Deste modo, o efeito do salto será atenuado pela multiplicação entre os retornos.

3.2.5. MedRV e minRV

As medidas propostas por Andersen, Dobrev e Schaumburg (2012) baseiam-se no truncamento de retornos adjacentes com a intenção de driblar o efeito dos saltos na estimação de volatilidade. O estimador minRV utiliza o quadrado do mínimo entre dois retornos consecutivos. Se um destes retornos for grande devido à presença de um salto, ele é automaticamente descartado:

$$\min RV_{t+1}^{(k)} = \frac{\pi}{\pi - 2} \left(\frac{n}{n-1} \right) \sum_{i=1}^{n-1} \min \left(\left| r_{t+i/n}^{(k)} \right|, \left| r_{t+(i+1)/n}^{(k)} \right| \right)^2$$

Assim como a variação realizada bipotente, a minRV é sensível à ocorrência de retornos muito pequenos ou iguais a zero. Além disso, retornos pequenos são menos informativos sobre a volatilidade subjacente. Para contornar essa inconveniência, os autores também propuseram o estimador medRV, que utiliza o quadrado da mediana entre três retornos adjacentes:

$$\text{med}RV_{t+1}^{(k)} = \frac{\pi}{6 - 4\sqrt{3} + \pi} \left(\frac{n}{n-2} \right) \sum_{i=2}^{n-1} \text{med} \left(\left| r_{t+(i-1)/n}^{(k)} \right|, \left| r_{t+i/n}^{(k)} \right|, \left| r_{t+(i+1)/n}^{(k)} \right| \right)^2$$

onde n é o número de observações no período t . Em ambas as equações, os fatores à esquerda do somatório asseguram uma estimativa não viesada para a variância pontual subjacente se os retornos forem independentes e identicamente distribuídos (i.i.d.) de acordo com a distribuição gaussiana.

3.2.6. Estimador em duas escalas de tempo robusto

O estimador em duas escalas de tempo robusto a saltos de Boudt e Zhang (2015) remove do estimador em duas escalas de tempo com correção de viés aqueles retornos que excedem um determinado limite (relacionado à sua distribuição, sob a hipótese de ausência de saltos). Os autores definem uma função indicadora que assume o valor 1

caso não haja suspeita da ocorrência de salto e o valor 0 caso contrário. Essa função indicadora é aplicada aos dois componentes do estimador em duas escalas de tempo com correção de viés, apresentado na seção 3.2.3, juntamente com um fator de correção do viés oriundo do truncamento. Neste trabalho, essa medida realizada é calculada para frequências amostrais entre 5 e 30 minutos, devido a restrições relacionados ao tamanho da amostra.

O apêndice A apresenta estatísticas descritivas (média, máximo, mínimo, desvio padrão, assimetria e curtose) para todas as medidas realizadas calculadas para as ações BBDC3 e VALE3 em diversas frequências amostrais durante o período entre 3 nov. 2009 e 17 fev. 2017.

3.3. PREVISÃO DA VOLATILIDADE: O MODELO HAR-RV

Um dos principais usos práticos dos estimadores de volatilidade é a construção de previsões, que poderão ser aplicadas a problemas de gestão de risco, seleção de carteiras, apreçamento de ativos, entre outros. Para gerar previsões de volatilidade, este trabalho utiliza o modelo HAR-RV de Corsi (2009).

O *Heterogeneous Autoregressive Realized Volatility* (HAR-RV) é um modelo que alia simplicidade à capacidade de reproduzir os principais fatos estilizados dos retornos quadráticos, tais como memória longa e padrões de comportamento que variam de acordo com diferentes horizontes temporais. O HAR-RV assemelha-se a um modelo autorregressivo, porém, em vez de incluir apenas defasagens da variável dependente, seus termos representam agregações temporais desta variável. Corsi (2009) utiliza uma especificação para a volatilidade realizada diária que inclui três componentes, correspondendo aos horizontes temporais diário, semanal e mensal. O componente diário é uma defasagem da variável dependente. Já os componentes semanal e mensal são a média simples da volatilidade realizada dos últimos 5 e 22 dias, respectivamente. O modelo é dado pela seguinte equação:

$$MR_{i,t+1} = \beta_{0,i} + \beta_{1,i}MR_{i,t} + \beta_{2,i} \frac{1}{5} \sum_{k=0}^4 MR_{i,t-k} + \beta_{3,i} \frac{1}{22} \sum_{k=0}^{21} MR_{i,t-k} + \varepsilon_{i,t+1}$$

onde $MR_{i,t}$ representa um estimador da variação quadrática ou da variação integrada diária (medida realizada), o qual pode assumir diferentes formas funcionais, como as apresentadas na seção 3.2., ou ser computado utilizando diferentes frequências amostrais intradiárias. A motivação por trás desta especificação provém do fato de que os participantes do mercado financeiro atuam em diversas frequências de negociação, percebendo, reagindo a, e causando os diferentes componentes da volatilidade. O modelo funciona como uma cascata de volatilidades das frequências mais baixas para as frequências mais altas e baseia-se na observação de que os negociadores de curto prazo reagem à volatilidade de longo prazo, enquanto que o oposto não é necessariamente verdadeiro.

A estimação dos parâmetros é feita por mínimos quadrados ordinários. Empregase o método de Newey-West para a correção dos erros padrão a fim de considerar a possível presença de auto correlação serial nos dados.

No presente trabalho, o modelo HAR-RV é empregado para prever a volatilidade das ações BBDC3 e VALE3 utilizando-se cada uma das medidas realizadas apresentadas na seção 3.2, bem como 8 frequências amostrais diferentes (1, 5, 10, 15, 30, 45, 60 e 90 minutos). Utiliza-se uma janela móvel de 500 dias para estimar o modelo e fazer a previsão um passo a frente (fora da amostra). Para cada método, são geradas 1.292 previsões um passo à frente. O apêndice B apresenta os gráficos das previsões agrupados por ação e por frequência amostral. Estes gráficos mostram que as previsões têm uma trajetória semelhante no tempo, especialmente para o ativo de maior liquidez (VALE3) e para frequências amostrais mais altas. A seção seguinte descreve o método utilizado para comparar estas previsões.

3.4. COMPARAÇÃO DAS PREVISÕES: MODEL CONFIDENCE SET

A fim de identificar as medidas realizadas e frequências amostrais capazes de gerar as melhores previsões de volatilidade, utilizamos a abordagem *Model Confidence Set* (MCS), proposta por Hansen, Lunde e Nason (2011). Um MCS é um conjunto de

modelos⁵ construído de tal forma a conter o melhor modelo a um dado nível de confiança. Ele é análogo a um intervalo de confiança para um parâmetro.

Considere um conjunto inicial \mathcal{M}^0 contendo um número finito de objetos que se deseja comparar. O procedimento consiste em “peneirar” \mathcal{M}^0 com o intuito de obter um conjunto reduzido, \mathcal{M}^* . Cada objeto é avaliado em termos de uma função perda. Denote por $L_{i,t}$ a perda associada ao objeto i no período $t, t = 1, \dots, n$. Denote por $d_{ij,t} \equiv L_{i,t} - L_{j,t}, \forall i, j \in \mathcal{M}^0$ o desempenho relativo entre dois objetos. Assume-se que $\mu_{ij} \equiv E(d_{ij,t})$ é finita e não depende de t para todo $i, j \in \mathcal{M}^0$. O objeto i será preferível ao j quando $\mu_{ij} < 0$ e o objeto j será preferível ao i quando $\mu_{ij} > 0$. A determinação de quais objetos farão parte de \mathcal{M}^* é realizada através de uma sequência de testes de significância, nos quais a hipótese nula e a hipótese alternativa assumem a seguinte forma:

$$H_0: \mu_{ij} = 0 \text{ para todo } i, j \in \mathcal{M}$$

$$H_A: \mu_{ij} \neq 0 \text{ para algum } i, j \in \mathcal{M}$$

onde $\mathcal{M} \subset \mathcal{M}^0$. O algoritmo funciona da seguinte forma: inicialmente, define-se $\mathcal{M} = \mathcal{M}^0$. A seguir, testa-se a hipótese nula ao nível de significância α . Em caso de rejeição, elimina-se do conjunto \mathcal{M} o objeto identificado como sendo significativamente inferior. Caso não seja possível rejeitar H_0 , o objeto permanece no conjunto \mathcal{M} . Em ambos os casos, repete-se o procedimento até que todos os pares de objetos sejam testados. Deste modo, os objetos permanecem no conjunto até que se provem significativamente inferiores a algum dos seus concorrentes. O conjunto de objetos estatisticamente superiores ao nível de significância α será o *Model Confidence Set*, denotado por $\widehat{\mathcal{M}}_{1-\alpha}^*$. Quanto mais informativa for a amostra, maior será o poder dos testes para rejeitar as hipóteses falsas, resultando em um MCS com apenas um ou poucos objetos. Esta é a situação ideal. Por outro lado, caso os dados sejam pouco informativos, é possível que nenhum objeto seja eliminado.

⁵ Hansen, Lunde e Nason (2011), utilizam o termo “modelo”, mas salientam que ele pode se referir não apenas a modelos econométricos, mas também a previsões, regras de negociação, etc. Alternativamente, em alguns trechos do artigo, os autores empregam o termo “objeto”.

No teste empírico conduzido neste trabalho, o conjunto de modelos inicial, \mathcal{M}^0 , conta com um total de 47 objetos, apresentados na primeira coluna da Tabela 4 (Apêndice A). Obtemos um *Model Confidence Set* ao nível de confiança de 95% para cada ação (BBDC3 e VALE3). Este procedimento permitirá identificar se os resultados são consistentes para ações com diferentes níveis de liquidez. A variância utilizada no teste de hipótese é estimada através de *bootstrap*, no qual utilizamos 10.000 reamostragens. A seção 3.4.1 versa sobre a escolha da função perda. Os resultados são apresentados na seção 4.

3.4.1. Proxy de volatilidade e função perda

A comparação entre previsões geradas por diferentes modelos é comumente realizada através de uma função perda. Esta função mede a distância entre o valor previsto e o de fato “observado” e pode assumir diferentes formas funcionais. O modelo que tiver o melhor desempenho preditivo irá apresentar o menor valor para a função perda.

No presente trabalho, denotamos por $\hat{\sigma}_{t+h}^2$ a previsão pontual para a “verdadeira” variância condicional no período $t + h$, σ_{t+h}^2 . Ocorre que σ_{t+h}^2 não é diretamente observável, mesmo *ex post*, e daí surge a necessidade de substituí-la por uma variável *proxy*, $\tilde{\sigma}_{t+h}$. Uma *proxy* de volatilidade é um estimador imperfeito da verdadeira variância condicional. De acordo com Patton (2011b), dependendo da função perda que se utiliza para classificar previsões, a presença de ruído na *proxy* pode acarretar distorções na classificação. O autor define uma função perda robusta como aquela capaz de gerar classificações de previsões invariantes ao nível de ruídos na *proxy* e demonstra a robustez de duas funções perda amplamente utilizadas na literatura: MSE (EQM em português) e QLIKE. Simulações realizadas em Patton e Sheppard (2009a) e resultados empíricos em Hansen e Lunde (2005), Patton e Sheppard (2009b), Patton (2011a) e Liu, Patton e Sheppard (2015) sugerem que testes baseados na MSE têm menos poder relativamente àqueles baseados na QLIKE.

Patton (2011b) mostra que os *rankings* de previsões de volatilidade elaborados usando funções perda robustas e uma *proxy* de volatilidade condicionalmente não

viesada são assintoticamente equivalentes aos *rankings* que usam a verdadeira variável alvo latente (PATTON 2011a). De acordo com este autor, a proxy da variável alvo latente deve ser não viesada, porém não é necessário que seja muito precisa: uma opção simples e amplamente disponível é o retorno quadrático diário. Para muitos ativos, é razoável assumir que essa variável seja condicionalmente não viesada, pois tanto o retorno médio quanto o impacto dos efeitos de microestrutura de mercado são geralmente negligenciáveis na frequência diária. Cabe ressaltar, entretanto, que na presença de saltos no processo gerador dos dados, os retornos quadráticos funcionarão como uma *proxy* da variação quadrática, que é a variação integrada mais a soma dos saltos quadráticos. A *proxy* utilizada neste trabalho é o retorno diário quadrático, calculado como a diferença entre o preço de fechamento e o preço de abertura do pregão.

Denote por $\hat{\sigma}_{i,t+1}^2$ a previsão pontual de volatilidade e por $\tilde{\sigma}_{i,t+1}$ a proxy de volatilidade correspondente. A função perda QLIKE associada ao objeto i para o dia t é definida como:

$$L_{i,t} = L(\tilde{\sigma}_{i,t+1}, \hat{\sigma}_{i,t+1}^2) = \log \hat{\sigma}_{i,t+1}^2 + \frac{\tilde{\sigma}_{i,t+1}}{\hat{\sigma}_{i,t+1}^2}$$

4. RESULTADOS EMPÍRICOS

A Tabela 2 ilustra o *Model Confidence Set* resultante para as previsões de volatilidade das ações do Banco Bradesco S.A. (BBDC3) ao nível de confiança de 95%. As colunas representam as medidas realizadas e as linhas representam as frequências amostrais. As células hachuradas significam que o estimador correspondente à coluna não foi calculado para a frequência indicada pela linha. As células preenchidas com a cor preta representam as combinações de estimadores e frequências amostrais que foram excluídos do conjunto. As demais células correspondem às combinações que fazem parte do *Model Confidence Set*. O número no interior das células indica a posição daquela combinação no *ranking* dos melhores modelos. Pode-se fazer algumas inferências interessantes com base nesta tabela:

- a) o tradicional método da variância realizada com frequência de 5 minutos foi excluído do conjunto;
- b) as frequências amostrais muito baixas (90 minutos) ou muito altas (1 minuto) demonstraram desempenho inferior para prever a volatilidade das ações do BBDC3, classificadas como papéis de baixa liquidez neste trabalho;
- c) a combinação que obteve o melhor desempenho foi o estimador em duas escalas de tempo com frequência amostral de 1 hora;
- d) a frequência amostral – entre 10 e 15 minutos – parece ter mais relevância sobre o desempenho das previsões do que o estimador de volatilidade.

A Tabela 3 apresenta os resultados para a ação da Vale S.A. (VALE3), um dos papéis de maior liquidez na B3. Para este ativo, tanto o estimador de volatilidade quanto a frequência amostral parecem ser relevantes: permaneceram no *Model Confidence Set* apenas o estimador em duas escalas de tempo com correção de viés e o estimador em duas escalas de tempo robusto. Para ambas as medidas, nenhuma frequência amostral foi eliminada, porém a classificação mostra que as frequências mais altas geraram as melhores previsões. Todas as demais combinações de medidas realizadas e frequências amostrais foram consideradas significativamente inferiores ao nível de significância de 5%, incluindo a variância realizada com amostra a cada 5 minutos. Corroborando este resultado, o estimador em duas escalas de tempo com correção de viés foi considerado

o melhor entre os cinco estimadores propostos em Zhang, Mykland e Ait-Sahalia (2005). Ele também apresentou o melhor desempenho nas comparações conduzidas por Ait-Sahalia e Mancini (2008), Ghysels e Sinko (2011) e Zhang, Mykland e Ait-Sahalia (2012).

Para melhor ilustrar as diferenças entre as medidas realizadas, a Figura 3 do Apêndice C apresenta as distribuições empíricas das perdas (distância das previsões em relação à *proxy* de volatilidade) para a VALE3. Os gráficos de caixa estão agrupados por frequência amostral, permitindo a comparação entre os estimadores. Para todas as frequências em que são calculados, o estimador em duas escalas de tempo com correção de viés e o estimador em duas escalas de tempo robusto a saltos são aqueles cujas perdas apresentam as menores medianas e estão menos dispersas.

Tabela 2 - Model confidence set para BBDC3

	VR	VR2E	VR2Evies	VRBP	MedRV	MinRV	VR2Erob
1 min							
5 min			29	9	13	21	24
10 min	27		22	3	2	16	18
15 min	4	31	5	8	11	6	7
30 min	17	20	14	28	26	10	15
45 min	23	19				30	
60 min	32	1				25	
90 min		12					

Fonte: elaboração própria.

Tabela 3 - Model confidence set para VALE3

	VR	VR2E	VR2Evies	VRBP	MedRV	MinRV	VR2Erob
1 min							
5 min			2				1
10 min			4				3
15 min			6				5
30 min			7				8
45 min							
60 min							
90 min							

Fonte: elaboração própria.

Em suma, os resultados apontam para a rejeição da hipótese de que a variância realizada combinada com uma frequência amostral de 5 minutos é a opção mais adequada para fins de previsão de volatilidade. Assim como em Liu, Patton e Sheppard (2015) e Patton (2011a), os resultados mostram que essa combinação de medida realizada e frequência amostral, tradicionalmente utilizada em aplicações na área de finanças, é superada por alternativas mais sofisticadas.

A frequência amostral parece ter um papel importante no desempenho preditivo e aparentemente está relacionada à liquidez do ativo: ativos mais líquidos funcionam melhor com frequências mais altas e ativos de menor liquidez requerem amostras mais espaçadas. Este resultado está de acordo com Patton (2011a), que utilizou ações da IBM e encontrou os melhores desempenhos em frequências mais altas – entre 15 segundos e 2 minutos. Por fim, duas medidas realizadas destacaram-se: o estimador em duas escalas de tempo com correção de viés de Zhang, Mykland e Ait-Sahalia (2005) e o estimador em duas escalas de tempo robusto a saltos, proposto por Boudt e Zhang (2015).

5. CONSIDERAÇÕES FINAIS

O tema central do presente trabalho foi a previsão da volatilidade de ações. Uma vez que a volatilidade é uma variável latente - não observável - é preciso utilizar uma *proxy* em seu lugar. Tradicionalmente utiliza-se o quadrado dos retornos financeiros obtidos a partir dos preços de fechamento dos pregões da bolsa. Porém, nos últimos anos tem se difundido o uso de medidas realizadas como *proxy* de volatilidade. As medidas realizadas são estimadores da variabilidade dos preços ao longo de algum horizonte temporal - geralmente 1 dia - que são construídos utilizando dados amostrados em períodos de tempo mais curtos - minutos ou segundos. A teoria sugere que estas medidas têm potencial de gerar ganhos de precisão, uma vez que utilizam uma quantidade maior de informações em sua construção. A literatura tem sido ativa na proposição de diferentes formas de cálculo destas medidas e, por isso, o presente trabalho se propôs a comparar o desempenho de diferentes medidas realizadas na geração de previsões. A pergunta que norteou o seu desenvolvimento foi: "quais medidas realizadas e quais frequências amostrais geram as melhores previsões de volatilidade para ações negociadas na B3?".

Esta pergunta é relevante por diversas razões: ela interessa aos participantes do mercado de capitais à medida em que previsões precisas da volatilidade desempenham um papel crucial na precificação de ativos financeiros, na construção de carteiras de investimento ótimas, na gestão de riscos, bem como na elaboração de estratégias de negociação. Além disso, ela interessa aos acadêmicos da área de finanças empíricas à medida em que investigações envolvendo dados de alta frequência têm se tornado cada vez mais frequentes e elas implicam na escolha da medida realizada e frequência amostral que será utilizada. Inclusive, há evidências na literatura de que o bônus de desempenho associado ao uso de dados intradiários está justamente relacionado a estas escolhas e também de que elas variam de acordo com as características de cada bolsa de valores.

A hipótese desta pesquisa era de que a variância realizada combinada à frequência amostral de 5 minutos seria a opção mais adequada para fins de previsão de volatilidade. Essa hipótese tinha respaldo teórico no fato de que a variância realizada é o

equivalente amostral da variável alvo que desejávamos prever: a variação quadrática de uma série de retornos financeiros. Já a frequência de 5 minutos seria baixa o suficiente para evitar a presença de ruídos relacionados à microestrutura de mercado, mas sem comprometer os ganhos de informação que se espera obter quando se utilizam dados de alta frequência. Ocorre que os resultados apontaram para a rejeição desta hipótese. Medidas realizadas mais sofisticadas, em especial o estimador em duas escalas de tempo com correção de viés e o estimador em duas escalas de tempo robusto geraram previsões de volatilidade significativamente superiores. Além disso, a frequência amostral parece ter um papel importante no desempenho preditivo e, aparentemente, está relacionada à liquidez do ativo: ativos mais líquidos funcionam melhor com frequências mais altas e ativos de menor liquidez requerem amostras mais espaçadas.

O objetivo geral do trabalho era o de fornecer orientação para a escolha da medida realizada a ser utilizada em aplicações na área de econometria financeira. Após gerarmos previsões de volatilidade utilizando 7 estimadores e 8 frequências amostrais diferentes e compararmos seus desempenhos através da metodologia *Model Confidence Set*, concluímos que dentre as opções analisadas, o estimador em duas escalas de tempo com correção de viés e a sua versão robusta a saltos geram as melhores previsões de volatilidade. Já a escolha da frequência amostral depende do nível de liquidez do ativo. Para o ativo mais líquido da B3 no período considerado (PETR3), as frequências mais altas foram as que geraram os melhores resultados. Para a ação BBDC3, que é menos líquida, a frequência ideal ficou entre 10 e 30 minutos.

Estudos futuros poderiam ampliar o escopo desta pesquisa em diversos sentidos:

- a) utilizando um maior número de ações para adicionar robustez aos resultados;
- b) incluindo outras classes de ativos, a fim de abranger ações negociadas nos mercados à vista, a termo, de opções e futuros, índices, renda fixa corporativa, derivativos financeiros e agropecuários, câmbio e títulos públicos federais;
- c) considerando outros métodos de agregação dos preços, além do último preço negociado, tais como preço ponderado pelo volume, preço ponderado pelo logaritmo do volume, preço ponderado pelo número de negociações (média dos preços), mediana dos preços e preços de maior volume associado;

- d) utilizando outros métodos de amostragem, por exemplo, a cada x transações;
- e) comparando outras medidas realizadas propostas na literatura;
- f) utilizando frequências amostrais alternativas;
- g) empregando outros métodos de avaliação das previsões, além do Model Confidence Set. Por exemplo, empregando critérios econômicos para a comparação do desempenho, tal como o retorno de carteiras;
- h) realizando uma análise multivariada, através da previsão da matriz de covariância dos ativos.

REFERÊNCIAS

AIT-SAHALIA, Yacine.; MYKLAND, Per. A.; ZHANG, Lan. How often to sample a continuous- time process in the presence of market microstructure noise. **Review of Financial Studies**, New York, v. 18, n. 2, p. 351–416, 10 fev. 2005. Disponível em: <<https://academic.oup.com/rfs/article-abstract/18/2/351/1599888>>. Acesso em 05 jan. 2018.

AIT-SAHALIA, Yacine.; MANCINI, Lorian. Out of sample forecasts of quadratic variation. **Journal of Econometrics**, Amsterdam, v. 147, n. 1, p. 17-33, nov. 2008. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0304407608001218>>. Acesso em 05 jan. 2018.

ANDERSEN, Torben G.; BOLLERSLEV, Tim. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. **International Economic Review**, Hoboken, v. 39, n. 4, p. 885-905, nov. 1998. Disponível em: <<http://www.jstor.org/stable/2527343>>. Acesso em 05 jan. 2018.

ANDERSEN, Torben G.; DOBREV, Dobrislav; SCHAUMBURG, Ernst. Jump-robust volatility estimation using nearest neighbor truncation. **Journal of Econometrics**, Amsterdam, v. 169, n. 1, p. 75-93, jul. 2012. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0304407612000127>>. Acesso em 05 jan. 2018.

B3. **Relação com investidores**: apresentação mensal. São Paulo, jun. 2018. Disponível em: <http://ri.bmfbovespa.com.br/ptb/2147/B3%20-%20Apresentao%20Mensal_Maio.pdf>. Acesso em: 4 jul. 2018.

BANDI, Federico M.; RUSSELL, Jeffrey R. Volatility. In: BIRGE, John R.; LINETSKY, Vadim (Eds.). **Handbooks in Operations Research and Management Science**, v. 15 – Financial Engineering, [S.I.], Elsevier, 2007, p. 183-222. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0927050707150052>>. Acesso em 05 jan. 2018.

BARNDORFF-NIELSEN, Ole E.; SHEPHARD, Neil. Power and bipower variation with stochastic volatility and jumps. **Journal of financial econometrics**, Cary, v. 2, n. 1, p. 1-37, 2004. Disponível em: <<https://academic.oup.com/jfec/article-abstract/2/1/1/960705>>. Acesso em: 01 jun. 2018.

BARNDORFF-NIELSEN, Ole E.; SHEPHARD, Neil. Variation, jumps and high frequency data in financial econometrics. In: BLUNDELL, Richard; TORSTEN, Persson; NEWAY, Whitney K. (Eds.). **Advances in economics and econometrics: theory and applications**, Ninth World Congress. Cambridge: Cambridge University Press, 2006, p. 328-372. Disponível em: <<https://scholar.harvard.edu/files/world07.pdf>>. Acesso em 05 jan. 2018.

BARNDORFF-NIELSEN, Ole E.; et al. Realised kernels in practice: trades and quotes. **The Econometrics Journal**, Oxford, v. 12, n. 3, p. C1-C32, 24 nov. 2009. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1368-423X.2008.00275.x/full>>. Acesso em 05 jan. 2018.

BAUWENS, Luc; HAFNER, Christian M.; LAURENT, Sebastien. **Handbook of volatility models and their applications**. Hoboken: John Wiley & Sons, 2012.

BORGES, Bruna K.; CALDEIRA, João Fróis; ZIEGELMANN, Flavio A. A. Selection of minimum variance portfolio using intraday data: An empirical comparison among different realized measures for BM&FBovespa data. **Brazilian Review of Econometrics**, Rio de Janeiro, v. 35, n. 1, p. 23-46, 2015. Disponível em: <<http://bibliotecadigital.fgv.br/ojs/index.php/bre/article/view/21453>>. Acesso em 05 jan. 2018.

BOUDT, Kris; CROUX, Christophe; LAURENT, Sébastien. Outlyingness weighted covariation. **Journal of Financial Econometrics**, Cary, v. 9, n. 4, p. 657-684, 2011. Disponível em: <<https://academic.oup.com/jfec/article-abstract/9/4/657/929909>>. Acesso em 14 jun. 2018.

BOUDT, Kris; PETITJEAN, Mikael. Intraday liquidity dynamics and news releases around price jumps: evidence from the DJIA stocks. **Journal of Financial Markets**, Amsterdam, v. 17, p. 121-149, 2014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1386418113000177>>. Acesso em: 1 jun. 2018.

BOUDT, Kris; ZHANG, Jin. Jump Robust Two Time Scale Covariance Estimation and Realized Volatility Budgets. *Quantitative Finance*, , v. 15, n. 6, p. 1041-1054, 2015. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/14697688.2012.741692>>. Acesso em 17 jun 2018.

BOUDT, Kris.; CORNELISSEN, Jonathan; PAYSEUR, Scott. **Highfrequency**: tools for highfrequency data analysis. Pacote R versão 0.5.3. [S.l.], 31 Jul. 2017. Disponível em: <<https://cran.r-project.org/web/packages/highfrequency/highfrequency.pdf>>. Acesso em 05 jan. 2018.

CALDEIRA, João F.; MOURA, Guilherme V.; PERLIN, Marcelo S., SANTOS, André A. P. Portfolio management using realized covariances: evidence from Brazil. **EconomiA**, São Paulo, v. 18, n. 3, p. 328-343, 2017. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1517758017300395>>. Acesso em: 26 mai. 2018.

CARDOSO, Francisco de Arruda Botelho. **Estimação da variância integrada em processos univariados e multivariados para preços através de medidas realizadas**. 2015. Dissertação (Mestrado em Estatística) – Programa de Pós-Graduação em Estatística, Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, 2015.

CATANIA, Leopoldo; BERNARDI, Mauro. **MCS**: model confidence set procedure. Pacote R versão 0.1.3. [S.l.]. 4 ago. 2017. Disponível em <<https://cran.r-project.org/package=MCS>>. Acesso em 26 mai. 2018.

CORSI, Fulvio. A simple approximate long-memory model of realized volatility. **Journal of Financial Econometrics**, Cary, v. 7, n. 2, p. 194-196, mar. 2009. Disponível em: <<https://academic.oup.com/jfec/article-abstract/7/2/174/856522>>. Acesso em 05 jan. 2018.

ECONOMÁTICA. **Apresentação das cotações em janela específica**. (s.l.), (s.d.). Disponível em <http://economica.com/support/manual/pt_javas/whnjs.htm>. Acesso em 24 de novembro de 2017,

ENGLE, Robert F. The econometrics of ultra-high-frequency data. **Econometrica**, Chicago, v. 68, n. 1, p. 1-22, jan. 2000. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1111/1468-0262.00091/full>>. Acesso em 05 jan. 2018.

GHYSELS, Eric; SINKO, Arthur. Volatility forecasting and microstructure noise. **Journal of Econometrics**, Amsterdam, v. 160, n. 1, p. 257-271, 2011. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0304407610000771>>. Acesso em 05 jan. 2018.

GRIFFIN, Jim; OOMEN, Roel C. A. Sampling returns for realized variance calculations: tick time or transaction time? **Econometric Reviews**, New York, v. 27, n. 1-3, p. 230–253, 7 Mar. 2008. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/07474930701873341>>. Acesso em 05 jan. 2018.

HANSEN, Peter R.; LUNDE, Asger. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? **Journal of Applied Econometrics**, Chichester, v. 20, n. 7, p. 873-889, mar. 2005. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/jae.800/full>>. Acesso em: 8 jun. 2018.

HANSEN, Peter R.; LUNDE, Asger.; NASON, James M. The model confidence set. **Econometrica**, Chicago, v. 79, n. 2, p. 453-497, 14 fev. 2011. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.3982/ECTA5771/full>>. Acesso em 05 jan. 2018.

LIU, Lily Y.; PATTON, Andrew J.; SHEPPARD, Kevin. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. **Journal of Econometrics**, Amsterdam, v. 187, n. 1, p. 293-311, jul. 2015. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0304407615000329>>. Acesso em 05 jan. 2018.

FAMA, Eugene F. Efficient capital markets: A review of theory and empirical work. **The Journal of Finance**, New York, v. 25, n. 2, p. 383-417, 1970.

KRUEGER, Anne O. **Financial Markets and Economic Growth**. Tóquio, set. 2006. Disponível em: <http://www.imf.org/en/News/Articles/2015/09/28/04/53/sp092806>. Acesso em: 10 jul. 2018.

MARKOWITZ, Harry. Portfolio selection. **The journal of finance**, New York, v. 7, n. 1, p. 77-91, 1952. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1952.tb01525.x>. Acesso em: 10 jul. 2018.

MCALEER, Michael; MEDEIROS, Marcelo C. Realized volatility: a review. **Econometric Reviews**, New York, v. 27, n. 1-3, p. 10-45, 7 mar. 2008. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/07474930701853509>>. Acesso em 05 jan. 2018.

MERTON, Robert C. Option pricing when underlying stock returns are discontinuous. **Journal of Financial Economics**, Lausanne, v. 3, n. 1-2, p. 125-144, 1976. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0304405X76900222>>. Acesso em 1 jun. 2018.

PATTON, Andrew J.; SHEPPARD, Kevin. Evaluating volatility and correlation forecasts. In: ANDERSEN, Torben G.; DAVIS, Richard A.; KREISS, Jens-Peter; MIKOSH, Thomas (Eds.). **Handbook of Financial Time Series**. Verlag: Springer, 2009. p. 801-838.

PATTON, Andrew J.; SHEPPARD, Kevin. Optimal combinations of realised volatility estimator. **International Journal of Forecasting**, Amsterdam, v. 25, n. 2, abr.-jun. 2009, p. 218-238. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169207009000107>>. Acesso em: 8 jun. 2018.

PATTON, Andrew J. Data-based ranking of realised volatility estimators. **Journal of Econometrics**, Amsterdam, v. 161, n.2, p. 284-303, 2011. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304407610002551>>. Acesso em: 9 jun. 2018.

PATTON, Andrew J. Volatility forecast comparison using imperfect volatility proxies. **Journal of Econometrics**, Amsterdam, v. 160, n. 1, p. 246-256, 2011. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S030440761000076X>>. Acesso em: 8 jun. 2018.

POOTER, Michiel de.; MARTENS, Martin; DIJK, Dick van. Predicting the daily covariance matrix for S&P 100 stocks using intraday data—but which frequency to use?. **Econometric Reviews**, New York, v. 27, n. 1-3, p. 199-229, 2008. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/07474930701873333>>. Acesso em 26 mai. 2018.

RAMOS, Henrique P.; PERLIN, Marcelo S. GetHFData: A R package for downloading and aggregating high frequency trading data from Bovespa. **Revista Brasileira de Finanças**, Rio de Janeiro, v. 14, n. 3, p. 443-478, jul.-set. 2016. Disponível em: <<http://www.redalyc.org/pdf/3058/305850693005.pdf>>. Acesso em 05 jan. 2018.

RYAN, Jeffrey A., ULRICH, Joshua M., BENNETT, Ross. **xts**: extensible times series. Pacote R, versão 0.10-2. 14 mar. 2018. Disponível em: <<https://cran.r-project.org/package=xts>>.

WICKHAM, Hadley; FRANÇOIS, Romain; HENRY, Lionel; MÜLLER, Kirill; RStudio. **dplyr**: a grammar of data manipulation. Pacote R, versão 0.7.5. 19 mai. 2018. Disponível em: <<https://cran.r-project.org/package=dplyr>>.

ZHANG, Lan; MYKLAND, Per.; AIT-SAHALIA, Yacine. A tale of two time scales: determining integrated volatility with noisy high-frequency data. **Journal of the American Statistical Association**, New York, v. 100, n. 472, p. 1394-1411, 2005. Disponível em: <<http://amstat.tandfonline.com/doi/abs/10.1198/016214505000000169>>. Acesso em 05 jan. 2018.

ZHANG, Lan. Estimating covariation: epps effect, microstructure noise. **Journal of Econometrics**, Amsterdam, v. 160, n. 1, p. 33-47, jan. 2011. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0304407610000540>>. Acesso em 05 jan. 2018.

APÊNDICE A – ESTATÍSTICAS DESCRITIVAS

Tabela 4 - Estatísticas descritivas dos retornos quadráticos diários e medidas realizadas da BBDC3

(continua)

Estimador	Frequência	Média	Máximo	Mínimo	Desvio padrão	Assimetria	Curtose
R ²	diária	2.81E-04	5.84E-03	0	4.61E-04	3.6991	21.6545
VR	1 min	5.40E-04	4.35E-03	0	3.52E-04	3.3574	21.6793
VR	5 min	4.05E-04	4.27E-03	0	3.01E-04	3.8879	28.1611
VR	10 min	3.60E-04	4.31E-03	0	3.03E-04	4.3641	34.9107
VR	15 min	3.25E-04	4.00E-03	0	2.77E-04	3.8498	27.8766
VR	30 min	2.78E-04	2.60E-03	0	2.57E-04	3.1156	14.6129
VR	45 min	2.40E-04	2.88E-03	0	2.47E-04	3.9388	25.8908
VR	60 min	2.19E-04	4.67E-03	0	2.59E-04	5.8841	69.4061
VR	90 min	1.88E-04	3.42E-03	0	2.34E-04	4.617	37.2259
VR2E	5 min	4.17E-04	4.53E-03	0	3.07E-04	4.2439	33.8701
VR2E	10 min	3.83E-04	4.91E-03	0	3.04E-04	4.5225	40.1682
VR2E	15 min	3.65E-04	5.18E-03	0	3.00E-04	4.7868	47.8752
VR2E	30 min	3.34E-04	4.30E-03	0	2.90E-04	4.0685	31.5263
VR2E	45 min	3.24E-04	3.24E-03	0	2.88E-04	3.552	21.4793
VR2E	60 min	3.04E-04	3.67E-03	0	2.78E-04	3.4292	21.322
VR2E	90 min	2.97E-04	4.09E-03	0	2.84E-04	3.4982	24.8239
VR2Evies	5 min	3.71E-04	4.57E-03	0	2.98E-04	4.5263	38.3871
VR2Evies	10 min	3.37E-04	4.86E-03	0	2.86E-04	4.8361	47.4898
VR2Evies	15 min	3.16E-04	5.13E-03	0	2.76E-04	5.2938	61.525
VR2Evies	30 min	2.76E-04	4.21E-03	0	2.52E-04	4.5907	43.9632
VRBP	1 min	3.64E-04	3.34E-03	0	2.99E-04	2.6927	15.215
VRBP	5 min	3.35E-04	3.34E-03	0	2.59E-04	3.2069	19.9611
VRBP	10 min	3.09E-04	4.41E-03	0	2.71E-04	4.4086	40.7321
VRBP	15 min	2.79E-04	4.25E-03	0	2.52E-04	4.6148	44.9998
VRBP	30 min	2.30E-04	2.94E-03	0	2.39E-04	4.1791	30.9237
VRBP	45 min	2.01E-04	2.37E-03	0	2.17E-04	3.8321	24.3047
VRBP	60 min	1.77E-04	4.21E-03	0	2.53E-04	6.3216	69.1859
VRBP	90 min	1.32E-04	2.20E-03	0	1.78E-04	4.2762	28.8247
MedRV	1 min	3.99E-04	3.30E-03	0	3.06E-04	2.5734	13.445
MedRV	5 min	3.45E-04	3.29E-03	0	2.58E-04	3.0565	17.8701
MedRV	10 min	3.12E-04	5.68E-03	0	2.88E-04	6.1809	81.0226
MedRV	15 min	2.80E-04	5.05E-03	0	2.65E-04	5.8163	71.1829
MedRV	30 min	2.39E-04	3.73E-03	0	2.56E-04	4.5527	38.3873
MedRV	45 min	2.14E-04	2.86E-03	0	2.36E-04	3.8342	26.1415
MedRV	60 min	2.06E-04	5.71E-03	0	3.05E-04	7.2208	92.5177
MedRV	90 min	1.69E-04	4.42E-03	0	2.58E-04	6.3984	76.8866
MinRV	1 min	3.80E-04	3.20E-03	0	3.10E-04	2.6064	13.5309
MinRV	5 min	3.50E-04	3.53E-03	0	2.76E-04	3.2243	19.8224
MinRV	10 min	3.24E-04	5.66E-03	0	3.02E-04	5.4237	64.6138

Tabela 4 - Estatísticas descritivas dos retornos quadráticos diários e medidas realizadas da BBDC3

(continuação)

MinRV	15 min	3.00E-04	5.29E-03	0	2.91E-04	5.3874	60.7625
MinRV	30 min	2.71E-04	4.26E-03	0	3.08E-04	4.8395	42.2727
MinRV	45 min	2.18E-04	2.98E-03	0	2.45E-04	3.4761	20.4969
MinRV	60 min	2.55E-04	7.01E-03	0	3.96E-04	7.0345	87.3431
MinRV	90 min	1.68E-04	4.04E-03	0	2.66E-04	5.2704	46.0225
VR2Erob	5 min	3.62E-04	3.53E-03	0	2.63E-04	3.5547	24.0968
VR2Erob	10 min	3.35E-04	5.13E-03	0	2.81E-04	5.3299	59.9205
VR2Erob	15 min	3.14E-04	5.31E-03	0	2.71E-04	5.747	73.7766
VR2Erob	30 min	2.77E-04	4.29E-03	0	2.54E-04	4.6873	46.0535

Fonte: elaboração própria.

Notas: as estatísticas descritivas referem-se ao período entre 03 nov. 2009 e 17 fev. 2017, o que equivale a 1,793 dias úteis. R^2 : retorno quadrático diário. VR: variância realizada. VR2E: estimador em duas escalas de tempo. VR2Evies: estimador em duas escalas de tempo com correção de viés. VRBP: variação realizada bipotente. VR2Erob: estimador em duas escalas de tempo robusto.

Tabela 5 - Estatísticas descritivas dos retornos quadráticos diários e medidas realizadas da VALE3

(continua)

Estimador	Frequência	Média	Máximo	Mínimo	Desvio padrão	Assimetria	Curtose
R^2	diária	4.96E-04	1.25E-02	0	1.04E-03	5.2577	38.2871
VR	1 min	5.38E-04	7.64E-03	0	5.52E-04	3.9379	27.1451
VR	5 min	4.51E-04	8.41E-03	0	5.36E-04	4.5735	38.2707
VR	10 min	4.35E-04	8.96E-03	0	5.57E-04	4.8675	43.0096
VR	15 min	4.15E-04	9.61E-03	0	5.66E-04	5.6581	57.3885
VR	30 min	3.94E-04	1.27E-02	0	6.02E-04	7.5683	112.3416
VR	45 min	3.47E-04	1.14E-02	0	5.57E-04	7.5446	103.5217
VR	60 min	3.17E-04	1.75E-02	0	6.39E-04	13.5383	309.4838
VR	90 min	2.85E-04	1.17E-02	0	5.36E-04	8.8262	136.1504
VR2E	5 min	4.63E-04	8.03E-03	0	5.37E-04	4.2433	31.9388
VR2E	10 min	4.51E-04	9.32E-03	0	5.54E-04	4.9389	46.8297
VR2E	15 min	4.44E-04	1.01E-02	0	5.66E-04	5.4204	58.3334
VR2E	30 min	4.39E-04	1.21E-02	0	6.09E-04	6.4762	85.1168
VR2E	45 min	4.45E-04	1.31E-02	0	6.48E-04	6.8706	93.583
VR2E	60 min	4.36E-04	1.29E-02	0	6.61E-04	6.7062	85.5209
VR2E	90 min	4.41E-04	1.35E-02	0	6.98E-04	6.6876	83.4447
VR2Evies	5 min	4.30E-04	7.15E-03	0	5.22E-04	4.0448	26.3603
VR2Evies	10 min	4.20E-04	8.54E-03	0	5.36E-04	4.7642	40.8013
VR2Evies	15 min	4.10E-04	9.32E-03	0	5.43E-04	5.2761	52.4286
VR2Evies	30 min	3.94E-04	1.15E-02	0	5.73E-04	6.7319	90.0704
VRBP	1 min	4.71E-04	6.17E-03	0	4.82E-04	3.637	22.3074
VRBP	5 min	4.13E-04	5.99E-03	0	4.76E-04	3.8644	23.1645
VRBP	10 min	3.87E-04	7.59E-03	0	4.91E-04	4.9309	42.7327

Tabela 5 - Estatísticas descritivas dos retornos quadráticos diários e medidas realizadas da VALE3

(continuação)

VRBP	15 min	3.54E-04	9.89E-03	0	5.00E-04	7.054	92.537
VRBP	30 min	3.04E-04	1.12E-02	0	5.02E-04	8.7309	141.8543
VRBP	45 min	2.85E-04	4.21E-03	0	4.00E-04	3.8185	19.8655
VRBP	60 min	2.28E-04	7.23E-03	0	4.16E-04	7.4695	84.856
VRBP	90 min	2.08E-04	6.41E-03	0	4.10E-04	6.8794	69.5657
MedRV	1 min	4.83E-04	6.08E-03	0	4.85E-04	3.5466	20.8765
MedRV	5 min	4.10E-04	5.85E-03	0	4.71E-04	3.8211	22.5933
MedRV	10 min	3.81E-04	9.13E-03	0	5.00E-04	6.0706	69.1604
MedRV	15 min	3.47E-04	9.05E-03	0	4.77E-04	6.3053	76.2138
MedRV	30 min	3.18E-04	1.38E-02	0	5.52E-04	10.7051	213.0765
MedRV	45 min	3.12E-04	7.35E-03	0	4.71E-04	4.9471	42.3783
MedRV	60 min	2.49E-04	5.82E-03	0	4.11E-04	5.7227	50.9171
MedRV	90 min	2.58E-04	8.04E-03	0	5.16E-04	6.9554	70.9654
MinRV	1 min	4.80E-04	6.00E-03	0	4.81E-04	3.4669	20.0942
MinRV	5 min	4.17E-04	5.54E-03	0	4.76E-04	3.6968	20.7301
MinRV	10 min	3.91E-04	8.01E-03	0	5.09E-04	5.4604	51.8164
MinRV	15 min	3.64E-04	1.12E-02	0	5.33E-04	7.7518	113.166
MinRV	30 min	3.43E-04	1.36E-02	0	5.78E-04	9.3032	167.5251
MinRV	45 min	3.09E-04	6.50E-03	0	4.71E-04	4.5332	31.8268
MinRV	60 min	3.07E-04	8.40E-03	0	5.37E-04	6.2642	61.6084
MinRV	90 min	2.69E-04	1.03E-02	0	5.97E-04	7.7787	91.9677
VR2Erob	5 min	4.16E-04	7.54E-03	0	4.93E-04	4.3734	34.6722
VR2Erob	10 min	4.11E-04	8.66E-03	0	5.14E-04	4.9673	47.4045
VR2Erob	15 min	4.03E-04	9.40E-03	0	5.24E-04	5.5252	60.0125
VR2Erob	30 min	3.94E-04	1.15E-02	0	5.74E-04	6.7749	91.0857

Fonte: elaboração própria.

Notas: as estatísticas descritivas referem-se ao período entre 03 nov. 2009 e 17 fev. 2017, o que equivale a 1,793 dias úteis. R^2 : retorno quadrático diário. VR: variância realizada. VR2E: estimador em duas escalas de tempo. VR2Evies: estimador em duas escalas de tempo com correção de viés. VRBP: variação realizada bipotente. VR2Erob: estimador em duas escalas de tempo robusto.

APÊNDICE B – GRÁFICOS DAS PREVISÕES

Figura 1 - Previsões de volatilidade para BBDC3

(continua)

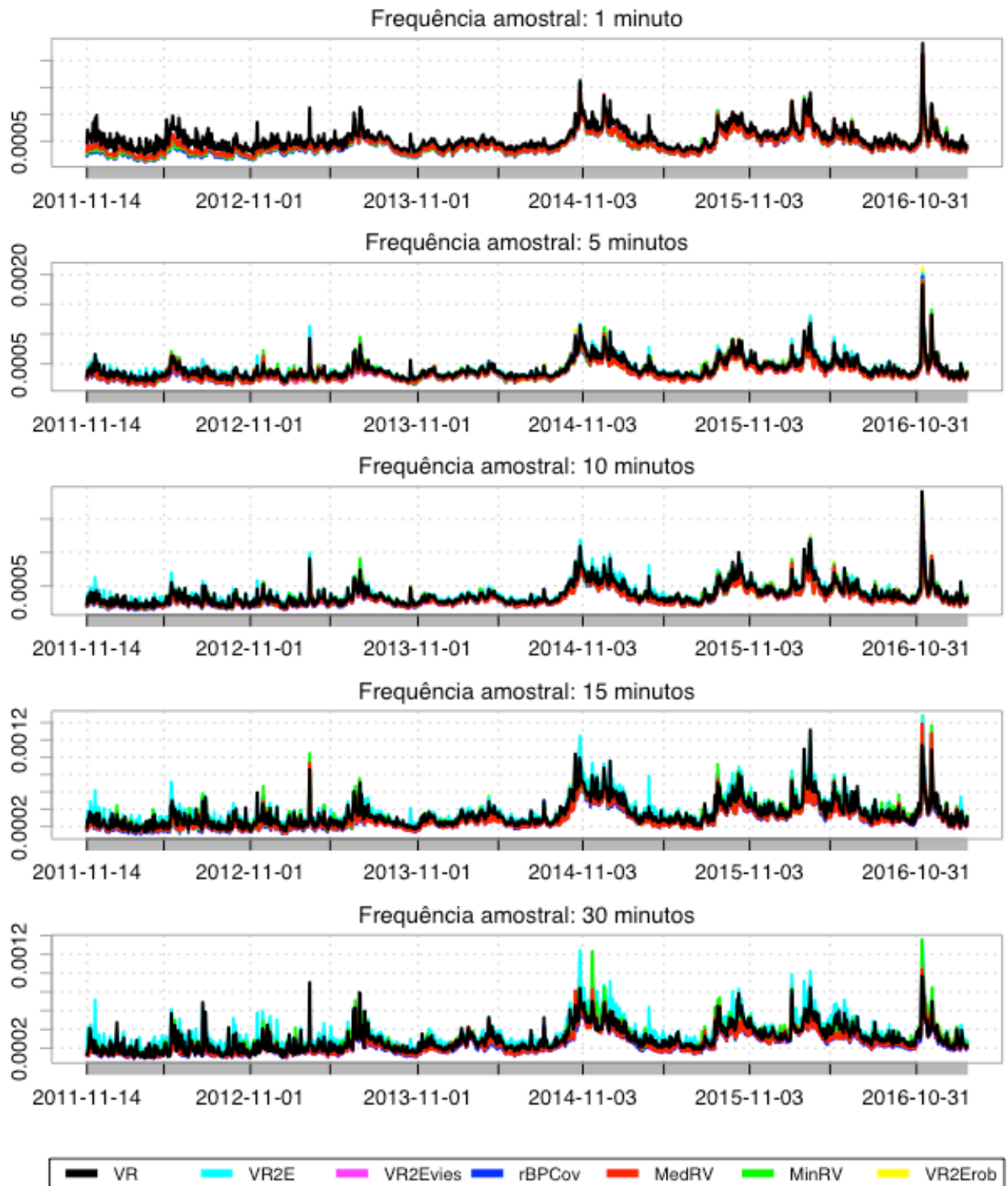
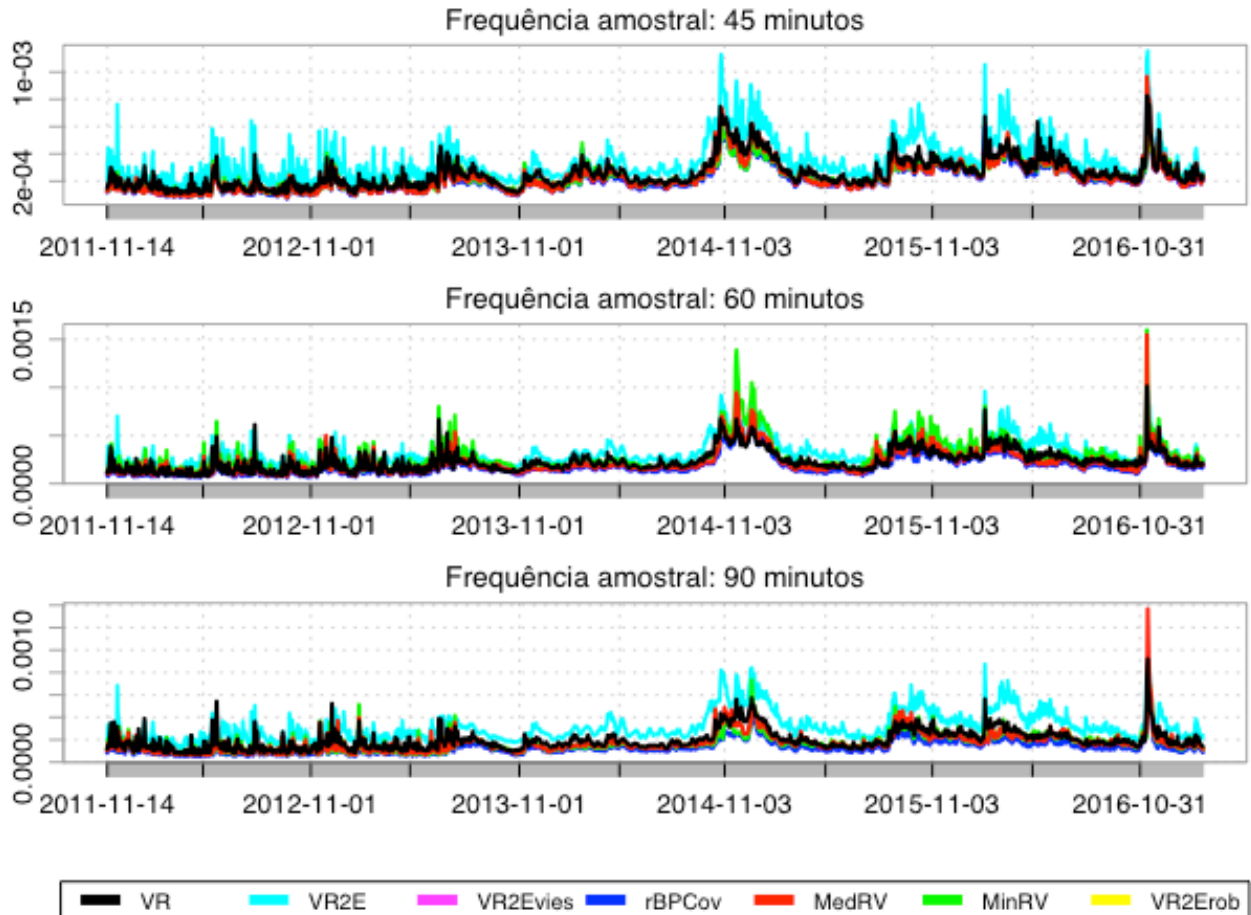


Figura 1 - Previsões de volatilidade para BBDC

(continuação)



Fonte: elaboração própria.

Notas: a figura apresenta previsões da volatilidade diária geradas a partir do modelo preditivo HAR-RV, utilizando as diversas combinações de medidas realizadas e frequências amostrais. A estimação do modelo e previsão 1 passo à frente foi realizada através de uma janela móvel de 500 dias. Os gráficos apresentam previsões para 1,292 dias. VR: variância realizada. VR2E: estimador em duas escalas de tempo. VR2Eviés: estimador em duas escalas de tempo com correção de viés. VRBP: variação realizada bipotente. VR2Erob: estimador em duas escalas de tempo robusto.

Figura 2 - Previsão de volatilidade para VALE3

(continua)

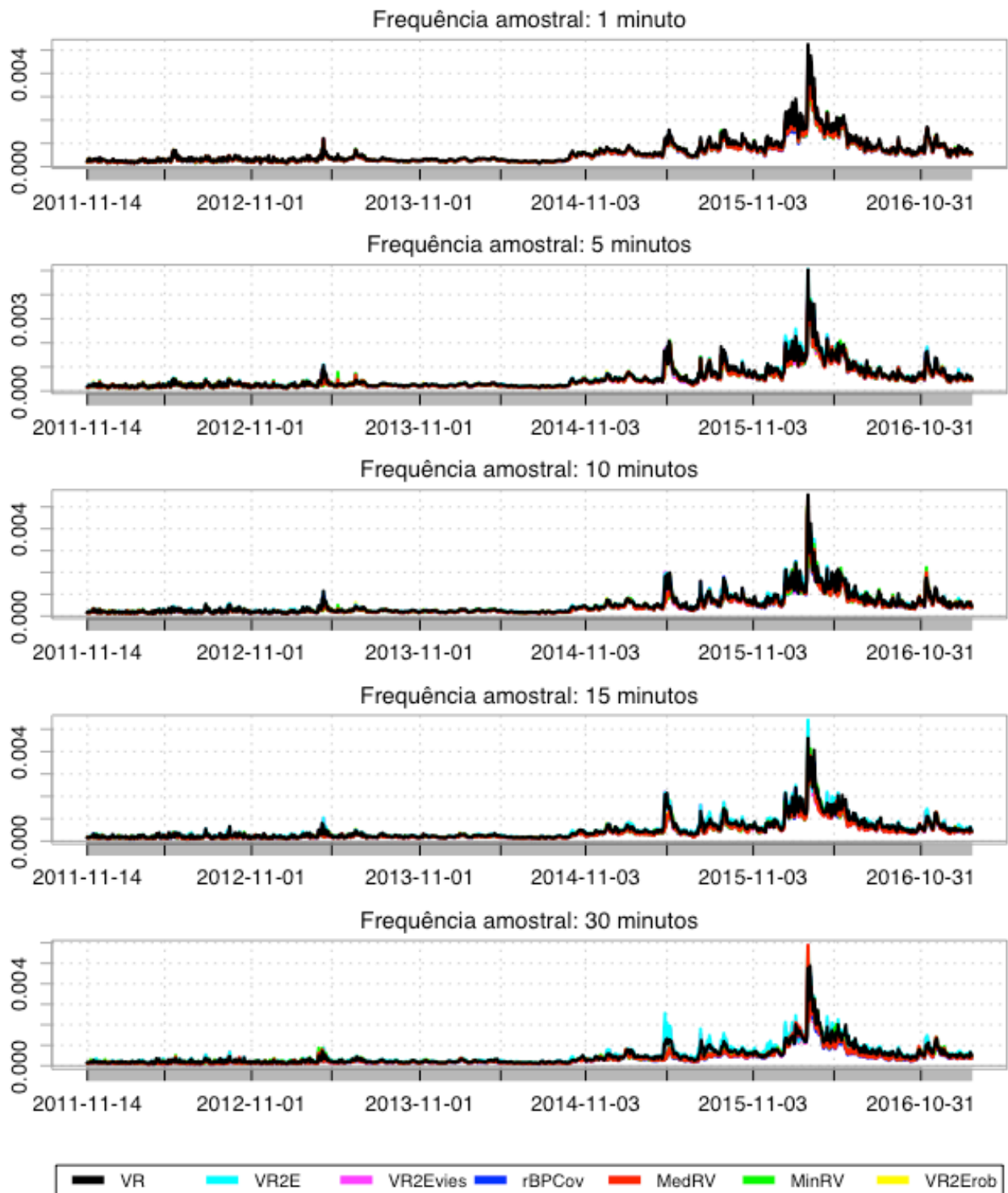
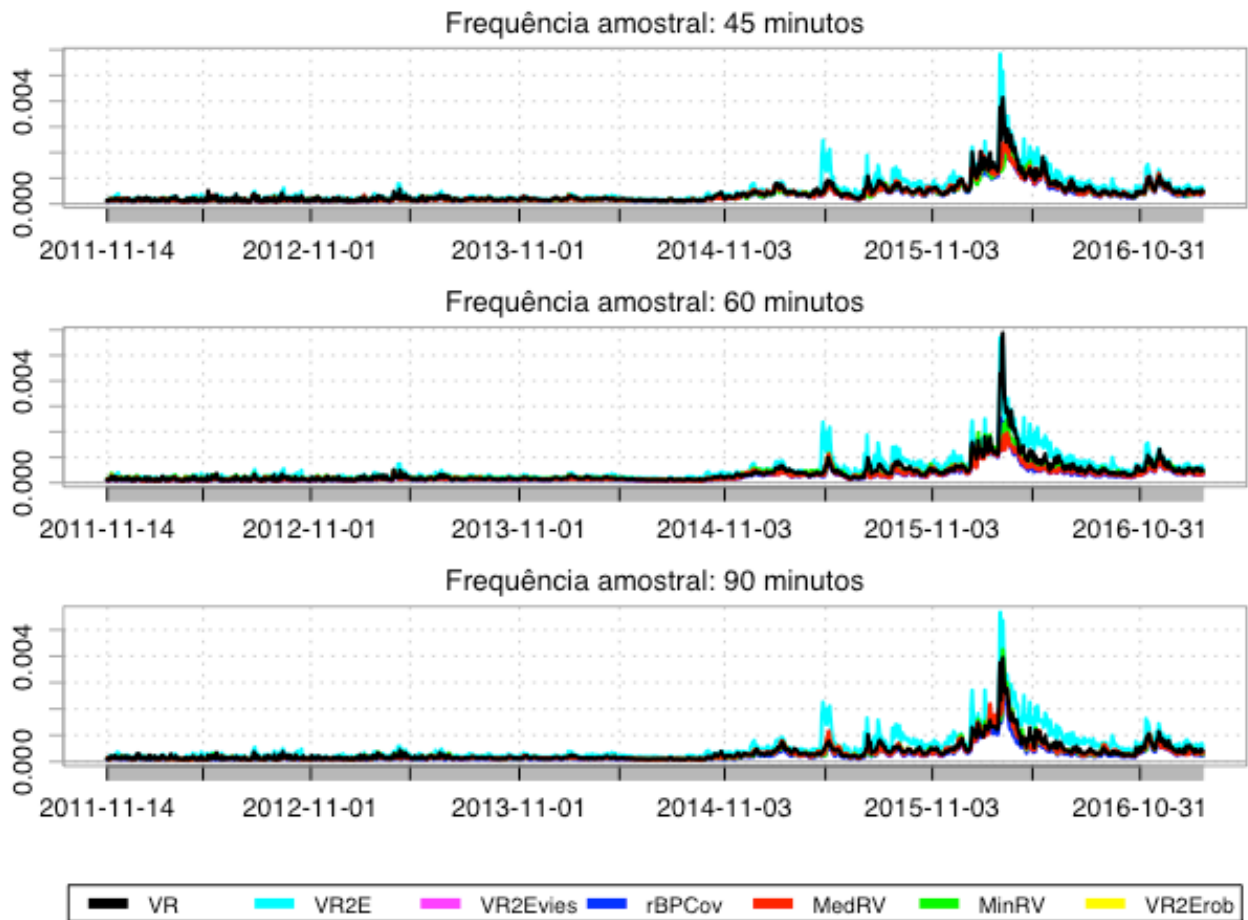


Figura 2 - Previsão de volatilidade para VALE3

(continuação)



Fonte: elaboração própria.

Notas: a figura apresenta previsões da volatilidade diária geradas a partir do modelo preditivo HAR-RV, utilizando as diversas combinações de medidas realizadas e frequências amostrais. A estimação do modelo e previsão 1 passo à frente foi realizada através de uma janela móvel de 500 dias. Os gráficos apresentam previsões para 1,292 dias. VR: variância realizada. VR2E: estimador em duas escalas de tempo. VR2Evis: estimador em duas escalas de tempo com correção de viés. VRBP: variação realizada bipotente. VR2Erob: estimador em duas escalas de tempo robusto.

APÊNDICE C – BOXPLOTS DAS PERDAS

Figura 3 - Boxplot das perdas em relação à proxy de volatilidade para VALE3 agrupado por frequência amostral

(continua)

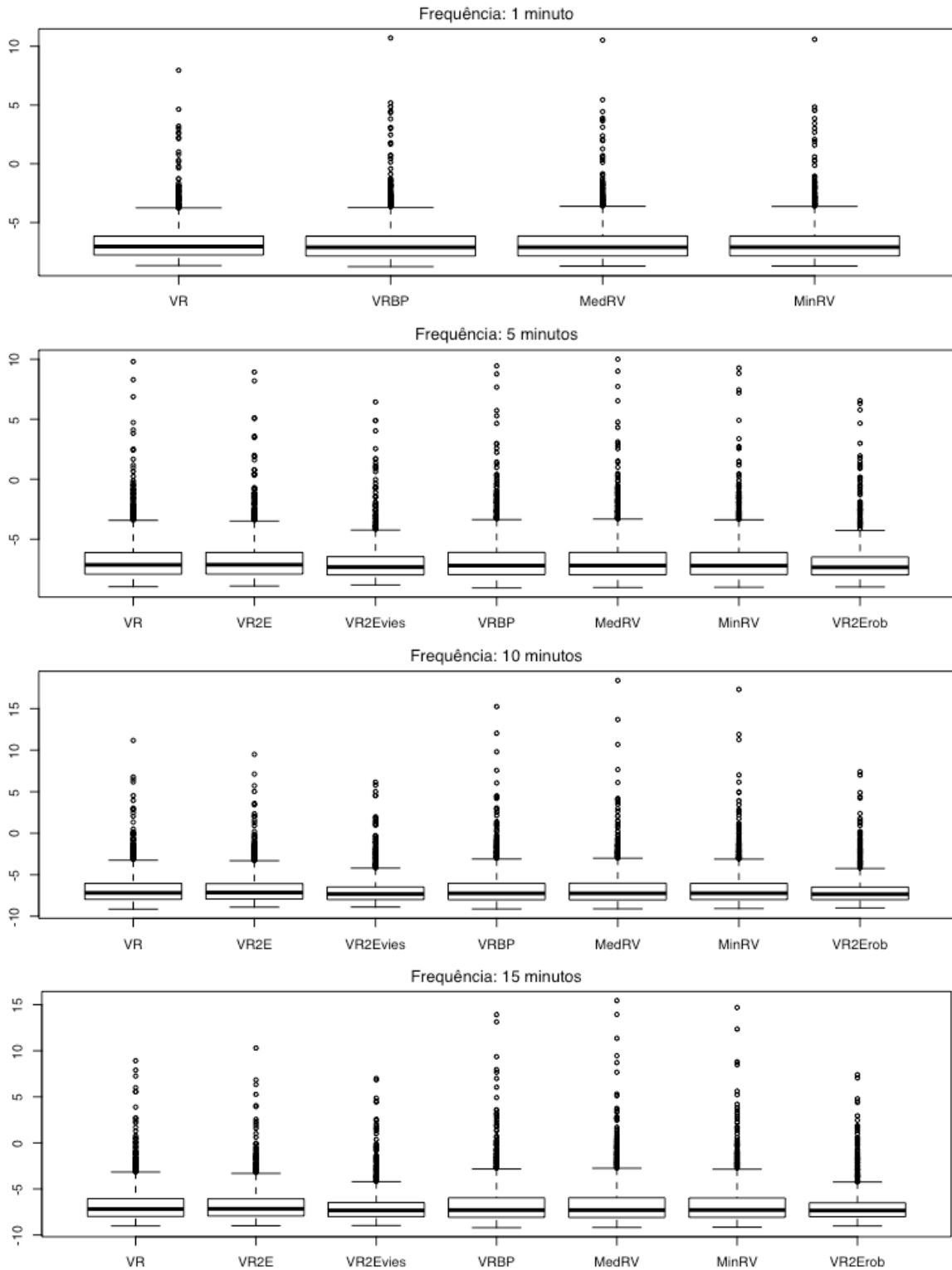
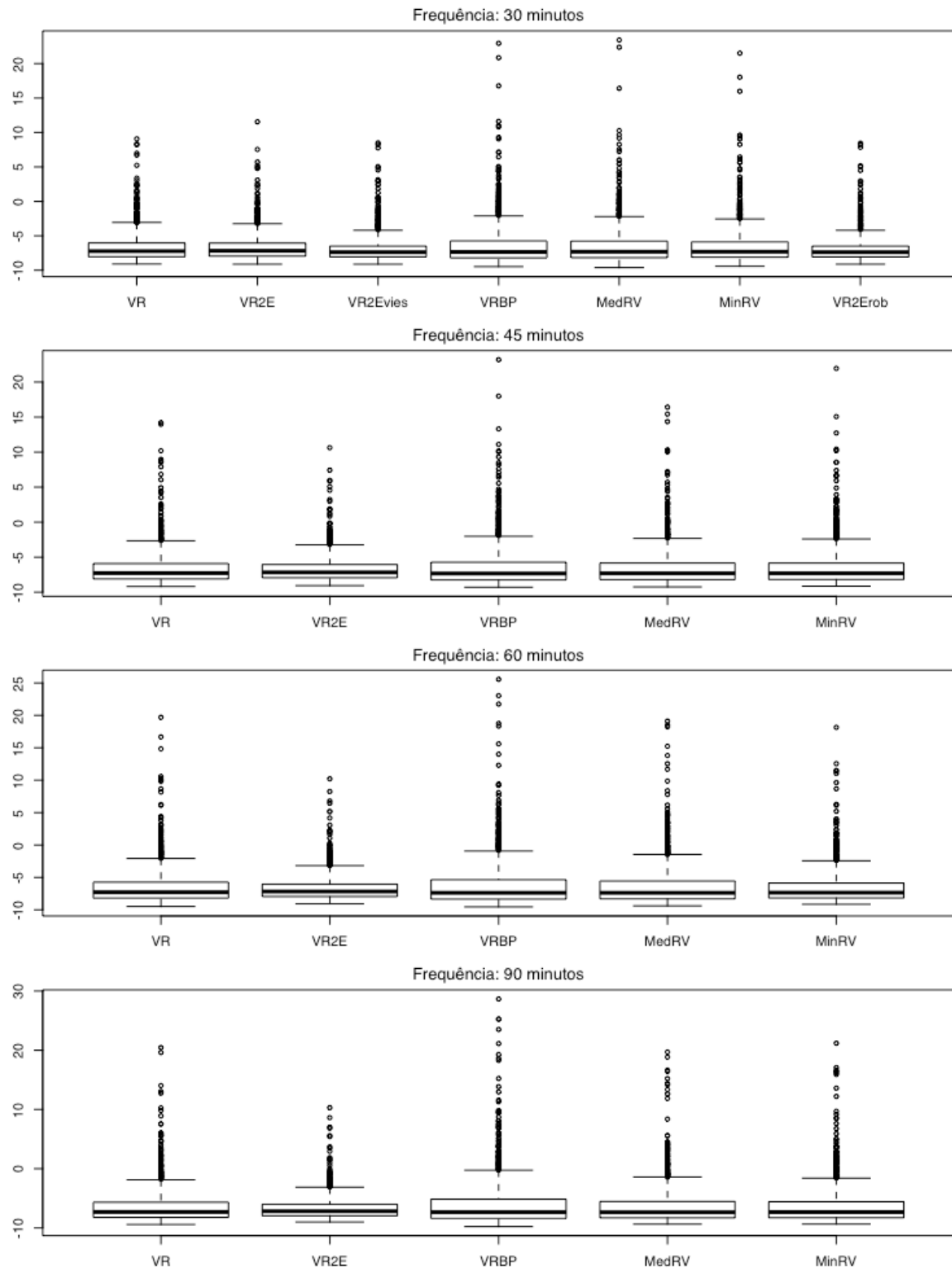


Figura 3 - Boxplot das perdas em relação à proxy de volatilidade para VALE3 agrupado por frequência amostral

(continuação)



Fonte: elaboração própria.