



## Tratamento de dados em alta frequência e estimação de medidas de volatilidade: um estudo de caso para petr4

*High frequency data management and volatility measures estimation: a case study for petr4*

Alcides Carlos Araújo<sup>1</sup>  
Alessandra Montini<sup>2</sup>

### Resumo

O artigo tem objetivo de analisar o tratamento de dados em alta frequência para a estimação de medidas de volatilidade percebida (*realized volatility* - RV). Para atingir os objetivos, buscou-se analisar as metodologias para limpeza de *outliers* e agregação dos preços. Para os métodos de agregação, consideraram-se as seguintes formas de amostragem: último preço negociado; preço ponderado pelo volume; preço ponderado pelo logaritmo do volume; preço ponderado pelo número de negociações; mediana dos preços e preços de maior volume associado. Foram estudadas as métricas *RCov* (sensível a problemas de microestrutura), *rOWCov*, *medRV*, *minRV* e *rRTSCov*, consideradas robustas a saltos e ruídos de microestrutura. Quanto aos resultados, observou-se que a remoção de *outliers* não influenciou de maneira significativa o processo de estimação da volatilidade percebida. Em relação à análise de agregação dos preços, por meio de uma simples mudança na metodologia, observaram-se diferenças significativas nas estimativas das volatilidades percebidas. Para a análise dos métodos de agregação, considerando as seis formas de amostragem, verificou-se que todas as medidas foram sensíveis às mudanças na forma de amostragem para agregar os preços. Do ponto de vista prático, gerenciar dados em alta frequência é um desafio devido à necessidade de manipulação de grandes bases. Por esse motivo, a não correção de possíveis problemas nos bancos de dados pode gerar estimativas de variabilidade imprecisas para a gestão de riscos. O artigo contribui por realizar uma revisão dos estimadores da volatilidade percebida mais recentes, buscando comparar a consistência em relação às diferentes formas de agregação e tratamento da série de preços.

**Palavras-chave:** Dados em alta frequência. Microestrutura. *Outliers*. Volatilidade percebida.

### Abstract

*This article aims to analyze the high frequency data management in order to estimate Realized Volatility (RV) measures. Some price aggregations and an outlier cleaning method were analyzed to reach the aims proposed. The sampling schemes of last price, volume weighted average price, logarithmic volume weighted average price, trading weighted average price, price median and maximum volume price were used as aggregation methods. The metrics studied were RCov (sensitive to microstructure problems) and rOWCov, medRV, minRV and rRTSCov metrics, all robust measures of dealing with jumps and microstructure noise. Regarding results, the outliers cleaning process did not significantly influence the Realized Volatility estimations. Regarding methods of price aggregation, there were significant differences between the measures of Realized Volatility when the sampling schemes were changed. When considering the six sampling schemes all metrics were sensitive to the change in price aggregation. In a practical point of view, the management of high frequency data is a challenge due to the need to manipulate big datasets. Therefore, inaccurate variability estimates for risk management may occur if the problems in these datasets were not corrected. This article contributes to a review of the most recent Realized Volatility estimators and compares the consistency of these estimators when different price aggregation and data cleaning methods are used.*

**Keywords:** High frequency data. Microstructure. *Outliers*. Realized Volatility.

<sup>1</sup> Doutorado em Administração pela Universidade de São Paulo – USP. Brasil. Afiliação: Universidade de São Paulo – USP. Lattes: <http://lattes.cnpq.br/6093802850914214> Email: [alcides.carlos@yahoo.com.br](mailto:alcides.carlos@yahoo.com.br)

<sup>2</sup> Doutorado em Administração Universidade de São Paulo – USP. Brasil. Afiliação: Universidade de São Paulo – USP. Lattes: <http://lattes.cnpq.br/8952860161247290> Email: [amontini@usp.br](mailto:amontini@usp.br)

## 1 Introdução

As análises de dados em alta frequência começaram a obter mais adeptos a partir da maior disponibilidade de dados em horizontes de tempo mais curtos, conforme apresentam Andersen et al. (2003). De acordo com Bauwens et al. (2012), há uma ampla disponibilidade de bases de dados provendo preços intradiários em alta frequência de ativos financeiros.

Diante desse cenário, diversos indicadores e medidas puderam ser criados de forma a adaptar a utilização dos dados disponíveis: uma das aplicações mais importantes foi o desenvolvimento das medidas de volatilidade utilizando dados em alta frequência.

As medidas de volatilidade foram desenvolvidas para estimar o risco de mercado, que, conforme Vanderlei e Carmona (2008), está relacionado às situações em que é possível estimar uma distribuição de probabilidade e associá-la aos resultados. Em finanças, portanto, a volatilidade é utilizada na quantificação do risco.

Há vários métodos para medir a volatilidade. Uma das formas de mensuração que cresceu em relevância nos últimos anos foi a medida de volatilidade realizada/percebida (*realized volatility* – RV), proposta por Andersen e Bollerslev (1998). Alguns trabalhos no Brasil, como o de Val et al. (2014), traduziram o termo *realized volatility* como volatilidade realizada. No presente artigo, os termos volatilidade percebida e volatilidade realizada serão sinônimos ao termo original supracitado.

A medida utiliza dados passados com frequência intradiária, caracterizando-a como puramente direcionada pelos dados. Conforme os trabalhos de Zivot (2005) e Andersen et al. (2007), demonstrou-se comparativamente que as estimativas dos modelos de projeção da volatilidade e da correlação condicional apresentaram melhor desempenho.

Apesar da crescente adoção da análise de dados em alta frequência, Yan e Zivot (2003) e Boudt et al. (2013) demonstram que a gestão desses dados possui diversos desafios devido às características das séries. Isso ocorre devido à grande quantidade de observações a serem processadas pelas capacidades computacionais da atualidade, além da necessidade de estimativa de parâmetros em séries com espaço de tempo irregular nos modelos - como os dados *tick by tick*.

Diante desse cenário, Falkenberry (2001) postula que a importância da filtragem de dados cresceu significativamente devido ao aumento da utilização de séries financeiras em alta frequência. Esse cenário foi viabilizado pelo aumento da capacidade de processamento dos computadores, do custo reduzido de armazenagem de dados e da computação em paralelo, que agiliza a tomada de decisões.

Falkenberry (2001) afirma que séries de dados em alta frequência possuem diversas realizações, completamente fora de padrões, que não representam os preços realmente negociados no mercado. Diante disso, além do tratamento dos dados em relação ao processo de ordenação dos horários, verificação dos dias sem negociação e feriados, e correção para horário de verão, faz-se necessária a filtragem de dados. Esse procedimento tem por finalidade equilibrar o tratamento dos pontos aberrantes (*outliers*) de forma a retirar aqueles que não representam a série de dados sem modificação das propriedades estatísticas.

Além do tratamento da série de preços, a metodologia para estimar a volatilidade percebida não é trivial, uma vez que diversas formas de estimar a RV foram propostas através dos anos. No artigo de McAleer e Medeiros (2008) apresenta-se uma revisão da literatura das diversas medidas de RV propostas até então. Os artigos de Aït Sahalia, Mykland e Zhang (2011) e Zhang e Boudt (2013) adicionaram à literatura mais algumas formas de estimativa da RV.

Há diversas metodologias para essa estimativa: podem se basear desde agregações dos tempos, em 5, 10, 15 e 30 minutos, até a utilização de todos os preços negociados. Cada uma dessas medidas visa melhorar as estimativas em relação aos problemas conhecidos, como saltos e ruídos de microestrutura.

Considerando as diversas metodologias, o artigo possui como objetivo geral verificar se o tratamento de dados em alta frequência é capaz de produzir estimativas consistentes da verdadeira volatilidade percebida.

Dentre os objetivos específicos, analisou-se a metodologia proposta por Brownless e Gallo (2006) para limpeza de *outliers* e analisou-se os métodos de agregação dos preços por diversos métodos, tais como: último preço negociado, 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.

As medidas de volatilidade percebida estudadas no presente artigo foram *RCov*, proposta por Andersen e Bollerslev (1998); *rOWCov*, proposta por Boudt et al. (2011); *medRV*, proposta por Andersen et al. (2012), *minRV*, proposta por Andersen et al. (2012), e *rRTSCov*, proposta por Zhang e Boudt (2013).

A medida *RCov* foi escolhida por ser a primeira forma de mensuração proposta na literatura, e as demais por serem as proposições mais recentes. Quanto a seus métodos, as três primeiras medidas (*RCov*, *rOWCov* e *medRV*) são robustas à presença de saltos, enquanto a última é tanto robusta a saltos quanto a ruídos de microestrutura, além de possibilitar o uso dos dados *tick by tick*.

Como contribuição, o presente artigo apresenta uma revisão dos estimadores mais recentes da volatilidade percebida, buscando verificar seu desempenho em relação a diferentes formas de agregação e tratamento da série de preços, estudo comparativo ainda não realizado no mercado brasileiro. O estudo avança em sua área, portanto, por estudar quais medidas estão mais adaptadas ao contexto do Brasil.

Conforme apresentam Daroit e Feil (2016), do ponto de vista prático, a gestão de riscos é imprescindível no gerenciamento de negócios. Analistas e investidores que necessitam estimar a volatilidade percebida podem, assim sendo, ser direcionados aos procedimentos adequados para obter estimativas consistentes a respeito do risco de mercado ao utilizarem dados em alta frequência.

## 2 Revisão da literatura

Para entender os diversos conceitos utilizados na pesquisa, faz-se necessária a apresentação dos principais elementos utilizados neste artigo, os quais foram divididos nos tópicos: Processo de filtragem, agregação das negociações e medidas de volatilidade percebida.

### 2.1 Processo de filtragem dos dados

O conceito base para o processo de filtragem de séries de dados em alta frequência é apresentado no artigo de Falkenberry (2001). Conforme o autor, o objetivo primário é a remoção de *outliers*, garantindo a manutenção das propriedades de séries observadas em tempo real.

Há, por exemplo, saltos nas séries de dados observados em tempo real que não podem ser classificados como *outliers*, pois são um comportamento normal de mercado quando da chegada de novas notícias, ou em eventuais transações de alto volume financeiro.

Falkenberry (2001) demonstra duas abordagens gerais. A primeira é a “busca e reposição/retirada” (*search and replace/delete*) de preços fora do padrão. Nesta abordagem, busca-se encontrar preços fora do padrão e verificar se devem ser retirados, ou trocados, por algum valor próximo, ou alguma média dos valores mais próximos.

A segunda abordagem é gerar uma série sintética por meio da série original. Neste caso, busca-se construir uma série de dados que represente o comportamento básico de uma série de preços. Um procedimento utilizado é gerar a série sintética por meio de um ajuste de médias móveis, como pode ser observado em Wink Junior e Pereira (2011) e Val et al. (2014).

Falkenberry (2001) defende a utilização da primeira abordagem. Conforme o autor, o procedimento de encontrar preços fora do padrão e repor/retirar possui melhor conexão com as séries observadas em tempo real. Adicionalmente, os filtros utilizados nessa abordagem podem ser aplicados sem dificuldade na prática e em tempo real.

O artigo de Brownless e Gallo (2006) discute de forma detalhada os procedimentos para limpeza e gerenciamento de séries financeiras de alta frequência. Deve-se primeiramente detectar e remover observações gravadas de forma errada na base, como preços iguais a zero na base de dados oferecida

pela bolsa de Nova Iorque (NYSE) existe um identificador específico para esses casos - e busca-se também detectar negociações que não representam os preços no momento do pregão (*outliers*).

A detecção de *outliers* ocorre conforme a expressão 1, proposta em Brownless e Gallo (2006):

$$(|p_i - \bar{p}_i(k)| < 3s_i(k) + \gamma) = \begin{cases} \text{verdadeiro, dado é mantido} \\ \text{falso, dado é removido} \end{cases} \quad (1)$$

em que  $p_i$  é o preço observado da série,  $\bar{p}_i$  é a média aparada (*trimmed mean*) das  $k$  observações mais próximas,  $s_i$  é o desvio padrão amostral das  $k$  observações mais próximas e  $\gamma$  é um parâmetro para redução de ruído.

Conforme observado na expressão 1, os parâmetros  $k$  e  $\gamma$  definem a quantidade de *outliers* a serem detectados na série. Brownless e Gallo (2006) demonstraram que, à medida que os valores de  $k$  e  $\gamma$  aumentavam, o número de *outliers* detectado era menor.

No mesmo artigo, Brownless e Gallo (2006) analisaram a aplicação da limpeza e tratamento de dados em alta frequência na estimação de durações financeiras. Os autores observaram que a remoção de *outliers* diminuiu significativamente o número de durações. Portanto, demonstrando que a série não tratada pode superestimar o número de mudanças de preços em estudos sobre durações financeiras.

## 2.2 Agregação das negociações

Diversos **métodos foram propostos** através dos anos para estimar a volatilidade percebida. Os métodos foram baseados no processo de difusão dos preços apresentado em Andersen et al. (2001), Barndorff-Nielsen e Shephard (2004), Macedo (2011) e Bauwens (2012).

O detalhamento passo a passo desse processo está fora do escopo do presente artigo. Apresentam-se aqui somente os princípios básicos para compreensão dos motivos da agregação das negociações e suas estimativas usando todos os dados *tick by tick*.

O processo é baseado na proposta de que existam negociações no período  $T$ , as quais são igualmente espaçadas no tempo. Desse modo, ocorrem realizações na série a cada  $\Delta t = T/M$  unidades de tempo. Diante disso, a distribuição de probabilidade dos retornos foi definida em Bauwens (2012) por meio da expressão 2:

$$r_{t,i} \sim N\left(\int_{t,i-1}^{t,M} \mu_{t,s} ds, IV_{t,M}\right), \quad (2)$$

em que  $IV_{t,s}$  (*integrated variance*) é dada pela expressão 3:

$$IV_{t,s} \equiv \int_{t,i-1}^{t,M} \sigma_{t,s}^2 ds. \quad (3)$$

Como se observa, os valores de  $IV_{t,s}$  precisam ser estimados, dado que é uma variável latente, o que se realiza pelas medidas de volatilidade percebida.

Mediante a necessidade de estimar  $IV_{t,s}$ , observa-se que, à medida que a frequência amostral aumenta, tende-se a diminuir o erro entre a volatilidade percebida estimada e o verdadeiro valor de  $IV_{t,s}$ . Desse modo, quanto maior a amostra, melhor seria a estimação por alguma medida proposta de volatilidade percebida.

Conforme apresentado em Ait-Sahalia, Mykland e Zhang (2011), Macedo (2011) e Zhang e Boudt (2013), entretanto, ao estimar a volatilidade percebida de forma empírica, o erro relativo ao verdadeiro valor de  $IV_{t,s}$  tende a aumentar conforme a frequência amostral cresce.

A explicação para esse problema está relacionada à falta de robustez do estimador da volatilidade percebida quando o intervalo entre cada negociação diminui. Em frequências de tempo muito curtas, o impacto das imperfeições de mercado é evidente.

Diante disso, os pesquisadores buscaram amostrar os preços em intervalos de tempos maiores para obterem medidas consistentes. Conforme Macedo (2011), o tamanho das janelas de estimação escolhidas tencionaram variar entre 5 e 30 minutos.

Por meio de dados da bolsa brasileira, Wink Júnior e Pereira (2011) testaram a utilização de frequências de 1, 5 e 15 minutos, e os resultados demonstraram que a frequência de 5 minutos apresentou o menor intervalo de confiança para a estimação da volatilidade percebida.

Além da frequência de tempo a ser escolhida, ainda é necessário selecionar a forma de agregar os preços. Boudt et al. (2013) apresenta o método do último preço negociado antes do tempo selecionado como o mais popular.

No artigo de Barndorff-Nielsen et al. (2009), os autores propõem cinco formas de agregar os preços. A primeira é utilizar o preço que teve maior volume associado; a segunda é utilizar o preço ponderado pelo volume; utilizar o preço ponderado pelo logaritmo do volume é a terceira; a quarta é utilizar o preço ponderado pelo número de negociações, e a quinta é utilizar a mediana dos preços.

O problema associado a agregar os preços em intervalos de tempos maiores é o desperdício de dados. Em um único dia, uma ação líquida pode possuir mais de 15.000 negociações. Caso a frequência de 5 minutos seja utilizada, uma série em torno de 80 preços será aproveitada para estimação da volatilidade percebida, descartando os demais preços.

A medida  $rRTSCov$ , estudada em Ait-Sahalia, Mykland e Zhang (2011) e Zhang e Boudt (2013), pode ser uma métrica de estimação da volatilidade percebida eficiente e que não possui problema de descartar os dados. A descrição das medidas estudadas no presente artigo é apresentada no tópico a seguir.

### 2.3 Medidas de volatilidade percebida

Conforme a expressão 3, os valores de  $IV_{t,s}$  (*integrated variance*) são estimados pelas medidas de volatilidade percebida por utilização de dados intradiários. Andersen e Bollerslev (1998) propõem o estimador da variância percebida (*realized variance*, expressão 4), denominado no presente artigo por  $RCov$ :

$$RCov_t \equiv \sum_{i=1}^M r_{t,i} r'_{t,i}, \quad (4)$$

em que  $r$  é um vetor de *log* retornos intradiários.

De acordo com Andersen *et al.* (2003), o estimador foi apresentado com o objetivo de integrar dados de negociação em alta frequência e medidas de estimação de volatilidades. Conforme os autores, o estimador é de fácil implementação e melhora o desempenho dos modelos preditivos.

Entretanto, conforme cita Macedo (2011), essa medida é baseada em uma situação ideal, em que preços de compra e venda são iguais e o tamanho de uma negociação é pequeno o bastante para que não haja influência sobre o mercado. Na prática, observa-se a existência de um *spread* de compra e venda, e certas negociações podem influenciar significativamente o sistema.

Características afins, presentes no ambiente de negociação, podem causar saltos e ruídos nas séries. Diante disso, o estimador da volatilidade percebida, apresentado na expressão 4, tende a superestimar o valor da verdadeira volatilidade. Diante desse problema, outras métricas robustas aos problemas citados foram propostas - neste artigo, são apresentadas as medidas  $rOWCov$ ,  $medRV$ ,  $minRV$  e  $rRTSCov$ .

Conforme Boudt et al. (2011), pode-se construir um estimador capaz de detectar os saltos e ajustar o estimador da volatilidade percebida. A medida foi denominada *realized outlyingness weighted covariance* -  $rOWCov$ , expressão 5:

$$rOWCov_t \equiv c_w \frac{\sum_{i=1}^M w(d_{i,t}) r_{t,i} r'_{t,i}}{\frac{1}{M} \sum_{i=1}^M w(d_{i,t})}, \quad (5)$$

em que  $r_{t,i}$  é um vetor de *log* retornos intradiários,  $i, \dots, M$ , é o número de retornos intradiários e  $d_{i,t}$  indica os pontos discrepantes da série a serem ponderados na estimação. Discrepâncias são dadas pela distância quadrada de Mahalanobis (medida pela diferença entre o vetor de retornos e zero).

O valor de  $w$  é o peso adotado para ponderar a discrepância, sendo atribuído a ele o valor de 1 caso o teste multivariado para detecção de saltos seja significativo. E  $c_w$ , por sua vez, é um valor escalar utilizado como fator de correção para o estimador.

O problema da existência de saltos na série financeira também foi analisado no artigo de Andersen et al. (2012), em que os autores apresentam duas medidas para estimação da volatilidade percebida, nomeadas *medRV* e *minRV*, as quais estão expressas nas equações 6 e 7 respectivamente:

$$medRV_t \equiv \frac{\pi}{6 - 4\sqrt{3} + \pi} \left( \frac{M}{M-2} \right) \sum_{i=2}^{M-1} med(|r_{t,i-1}|, |r_{t,i}|, |r_{t,i+1}|)^2, \tag{6}$$

$$minRV_t \equiv \frac{\pi}{\pi - 2} \left( \frac{M}{M-1} \right) \sum_{i=1}^{M-1} \min(|r_{t,i}|, |r_{t,i+1}|)^2, \tag{7}$$

em que  $r_{t,i}$  é o vetor de retornos intradiários e  $i, \dots, M$  é o número de retornos intradiários no período  $t$ . O operador “min” seleciona o menor retorno absoluto entre  $|r_{t,i}|$  e  $|r_{t,i+1}|$ , e o operador “med” estima a mediana dos retornos absolutos entre  $|r_{t,i-1}|$ ,  $|r_{t,i}|$  e  $|r_{t,i+1}|$ .

Nos artigos de Aït-Sahalia, Mykland e Zhang (2011) e Zhang e Boudt (2013), os autores apresentam uma métrica de estimação da volatilidade percebida que utiliza todos os dados *tick by tick*. De acordo com os autores, a medida é robusta à presença de saltos e ruídos de microestrutura. Denominada *Robust two time scale covariance estimator – rRTSCov*, é apresentada na expressão 8:

$$rRTSCov_t \equiv \left( 1 - \frac{\bar{n}_K}{\bar{n}_J} \right)^{-1} \left[ (Y, Y)_T^{(K)*} - \frac{\bar{n}_K}{\bar{n}_J} (Y, Y)_T^{(J)*} \right], \tag{8}$$

em que  $\bar{n}_K = (n - K + 1) / K$ , e  $\bar{n}_J = (n - J + 1) / J$ , sendo  $k$  as subamostras de tempo,  $J = 1$ ,  $n$  o tamanho total da amostra,  $(Y, Y)_T^{(K)*}$  a variância percebida da escala de tempo  $k$  e  $(Y, Y)_T^{(J)*}$  a variância percebida da escala de tempo  $j$ .

Apresenta-se no quadro 1 um resumo com as características de cada estimador da volatilidade percebida.

Quadro 1: Resumo estimadores de volatilidade percebida

Estimador	Autores	Robustez a saltos	Robustez a ruídos	Amostra de estimação
<i>RCov</i>	Andersen et al. (2003)	Não	Não	Agregada
<i>rOWCov</i>	Boudtet al. (2011)	Sim	Não	Agregada
<i>medRV</i>	Andersen et al. (2012)	Sim	Não	Agregada
<i>minRV</i>	Andersen et al. (2012)	Sim	Não	Agregada
<i>rRTSCov</i>	Zhang e Boudt (2013)	Sim	Sim	<i>Tick by tick</i>

Fonte: Elaborado pelos autores.

### 3 Metodologia

O presente artigo possui o objetivo de analisar o tratamento de dados em alta frequência para a estimação de medidas de volatilidade percebida.

Para tanto, foram utilizados 2 casos de séries *tick by tick* do papel PETR4, sendo 2 dias contendo alta volatilidade, divididos em 1 dia com tendência de alta e 1 dia com tendência de baixa. Os dados foram disponibilizados pela Bovespa.

Essa forma de apresentação foi baseada no artigo de Barndorff-Nielsen et al. (2009), que apresentaram casos para um dia regular, um dia heterocedástico, um dia com saltos graduais e um dia lateralizado com saltos rápidos (*puzzling day*). Os cálculos foram viabilizados pelo *software R*.

### 3.1 Formas de agregação

As análises consideraram os cenários de tratamento e não tratamento das séries, agregação dos preços em todo período de 5 minutos (utilização de todas as negociações no período das 10h15min às 10h19min, por exemplo) e agregação dos preços em horários múltiplos de 5 (isto é, utilização de todas as negociações às 10h15min e depois somente às 10h20min, desconsiderando o período das 10h16min às 10h19min) para os 6 métodos de agregação.

No quadro 2 são resumidos todos os cenários das análises realizadas no presente artigo.

Quadro 2: Resumo das análises realizadas

Séries	Tempo	RCov	rOWCov	medRV	minRV	rRTSCov
Série sem tratamento	Todo o período de 5 minutos	6 métodos de agregação	Tick by tick			
	Período múltiplo de 5	6 métodos de agregação	Tick by tick			
Série tratada	Todo o período de 5 minutos	6 métodos de agregação	Tick by tick			
	Período múltiplo de 5	6 métodos de agregação	Tick by tick			

Fonte: Elaborado pelos autores.

Ressalta-se que, no caso do estimador *rRTSCov*, as agregações não foram necessárias, uma vez que o estimador utiliza os dados *tick by tick*.

As formas de agregação estudadas foram: último preço negociado (*last*), preço ponderado pelo volume (*vol*), preço ponderado pelo logaritmo do volume (*logvol*), preço ponderado pelo número de negociações (média dos preços - *mean*), mediana dos preços (*med*) e preços de maior volume associado (*maxvol*). Os métodos são detalhados tomando como referência os dados apresentados na tabela 1.

Tabela 1: Exemplo de série de negociação

Horário da negociação	Preço	Quantidade
10:29:58.581344	18,15	200
10:29:59.263026	18,16	100
10:30:00.997018	18,15	100
10:30:01.000080	18,15	1400

Fonte: Dados obtidos pela Bovespa.

Em relação ao método do último preço negociado (*last*), para o caso de agregação em todo o período de 5 minutos, será utilizado o último preço negociado antes das 10h30min, o qual, pela tabela 1, será R\$18,16. No caso da agregação no horário múltiplo de 5 minutos, será utilizado o último preço negociado no horário das 10h30min, o qual, pela tabela 1, será R\$18,15.

Em relação ao método do preço ponderado pelo volume (*vol*), para o caso de agregação em todo o período de 5 minutos, o preço adotado seria a média ponderada dos preços e quantidades entre os tempos “10:29:58.581344” e “10:29:59.263026”. No caso da agregação no horário múltiplo de 5 minutos, o preço seria a média ponderada entre os tempos “10:30:00.997018” e “10:30:01.000080”.

Em relação ao preço ponderado pelo logaritmo do volume (*logvol*), para o caso de agregação em todo o período de 5 minutos, o preço adotado seria a média ponderada entre preços e *log* volumes entre os tempos “10:29:58.581344” e “10:29:59.263026”. No caso da agregação no horário múltiplo de 5 minutos, o preço é calculado pela média ponderada dos preços e *log* quantidades entre os tempos “10:30:00.997018” e “10:30:01.000080”.

Em relação ao preço ponderado pelo número de negociações (média dos preços - *mean*), para o caso de agregação em todo período de 5 minutos, seria utilizada a média dos preços entre os tempos “10:29:58.581344” e “10:29:59.263026”. No caso da agregação no horário múltiplo de 5 minutos, seria calculada a média dos preços entre os tempos “10:30:00.997018” e “10:30:01.000080”.

Em relação ao método de agregação pela mediana dos preços (*med*), para o caso de utilizar todo o período de 5 minutos, seria utilizada a mediana dos valores entre os tempos “10:29:58.581344” e “10:29:59.263026”. No caso de agregar no decorrer de horário múltiplo de 5 minutos, seria utilizada a mediana dos preços entre os tempos “10:30:00.997018” e “10:30:01.000080”.

Em relação ao método de agregação pelos preços de maior volume associado (*maxvol*), para o caso de utilizar todo o período de 5 minutos, seria utilizado o preço associado ao maior volume dos últimos negócios ocorridos antes das 10h30min, o qual, pela tabela 1 seria, o valor de R\$18,15 associado ao volume 200. No caso de agregar no horário múltiplo de 5 minutos, seria utilizado o preço associado ao maior volume dos negócios que ocorreram às 10h30min, o qual, pela tabela 1, seria o valor de R\$18,15 associado ao volume de 1.400.

Para realizar as comparações entre as metodologias, visando obtenção de medidas consistentes das estimações, utilizou-se a técnica do *bootstrap*, uma vez que as medidas de volatilidade foram baseadas em esquemas amostrais diferentes.

Proposto por Efron (1979), o *bootstrap* é um método de simulação que se baseia na construção de distribuições amostrais por reamostragem. É muito utilizado para estimar o viés e a variância de estimadores, bem como intervalos de confiança. Desse modo, não depende do conhecimento prévio da distribuição original da estatística do parâmetro estudado.

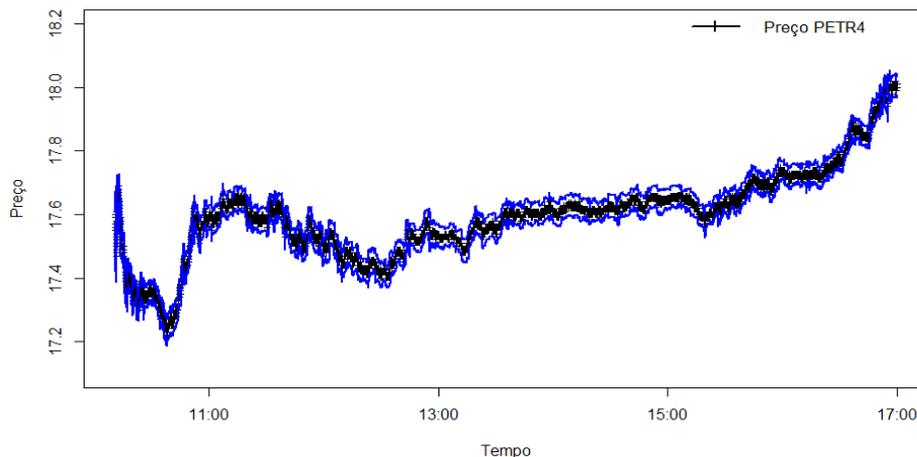
## 4 Análises

As análises dos resultados foram direcionadas pelas perguntas: estimar o RV com dados agregados seria uma melhor solução de que utilizar todos os preços presentes na série? O tratamento dos *outliers* auxiliaria na obtenção de medidas mais consistentes? Portanto, buscou-se analisar a importância da limpeza de *outliers* e a importância dos métodos de agregação.

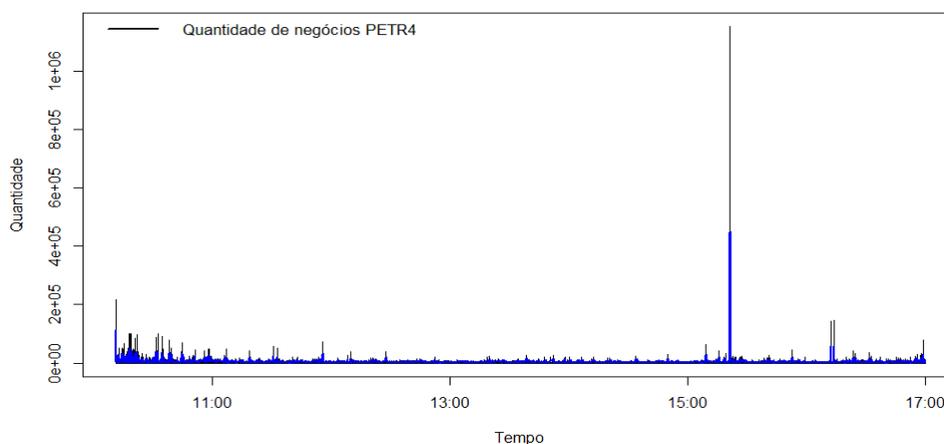
### 4.1 Dia de alta – 06 de março de 2013

Neste dia atípico, a ação da PETR4 apresentou alta de 15,16%, fechando o pregão ao preço de R\$16,41. Na figura 1 são apresentados os gráficos para o comportamento dos preços e volume negociados. As linhas em azul representam os intervalos para limpeza dos *outliers* (expressão 1) conforme proposto em Brownless e Gallo (2006). Os valores utilizados foram  $k = 60$  e  $\gamma = 0,02$ .

Figura 1: Comportamento dos preços e quantidades: caso alta



(a) Comportamento dos preços PETR4: caso alta.



(b) Comportamento das quantidades negociadas PETR4: caso alta. Fonte: Dados da pesquisa.

No horário de negociação, entre 10h e 17h, a série original apresentou 75.587 observações. Após o tratamento de *outliers*, restaram 73.139. Observou-se também que os pontos aberrantes foram encontrados somente na série de volume negociado. Na tabela 2 são apresentados os resultados das análises para cada estimador da volatilidade percebida nas bases sem tratamento e tratamento de *outliers*.

Tabela 2: Comparações série tratada e não tratada: caso alta

Medidas	Série não tratada	Série tratada	Diferença	B. Lim. Inf.	B. Lim. Sup.	Sig.
<i>RCov</i>	0,0406	0,0407	0,0000	-0,0012	0,0011	não
<i>rOWCov</i>	0,0261	0,0261	0,0000	-0,0011	0,0011	não
<i>medRV</i>	0,0346	0,0346	0,0001	-0,0018	w0,0018	não
<i>minRV</i>	0,0331	0,0329	0,0002	-0,0020	0,0023	não
<i>rRTSCov</i>	0,0301	0,0300	0,0001	-0,0022	0,0025	não

Resultados baseados em 10.000 reamostragens.  
 Fonte: Dados da pesquisa.

Observa-se pela tabela 2 que a série tratada demonstrou, em média, menor valor para a volatilidade percebida em relação à série não tratada. Contudo, pela observação dos limites inferiores e superiores estimados pelo bootstrap (B. Lim. Inf. e B. Lim. Sup.) tais diferenças não foram significativas, dado que o valor “zero” esteve contido em todos os intervalos.

Na tabela 3 são apresentadas as comparações dos estimadores para os métodos de agregação dos preços em todo o período de 5 minutos (todo 5 minutos) e agregação dos preços no horário múltiplo de 5 minutos (dentro do minuto 5).

Tabela 3: Comparações tempo para agregação: caso alta

Medidas	Todo 5 minutos	Dentro do minuto 5	Diferença	B. Lim. Inf.	B. Lim. Sup.	Sig.
<i>RCov</i>	0,0421	0,0392	0,0029	0,0019	0,0040	sim
<i>rOWCov</i>	0,0258	0,0263	-0,0005	-0,0016	0,0006	não
<i>medRV</i>	0,0369	0,0323	0,0046	0,0030	0,0062	sim
<i>minRV</i>	0,0368	0,0292	0,0076	0,0062	0,0092	sim

Resultados baseados em 10.000 reamostragens. Medida *rRTSCov* não utiliza agregações de tempo.

Fonte: Dados da pesquisa.

Os valores das volatilidades percebidas foram significativamente maiores neste dia específico quando foi realizada a agregação utilizando todo o período de 5 minutos. Somente a medida *rOWCov* não reproduziu a discrepância.

Os resultados demonstraram que as medidas *RCov*, *medRV* e *minRV* foram sensíveis à maneira como o tempo foi agregado neste dia específico. Esse resultado é encontrado pela observação dos limites inferiores e superiores estimados pelo *bootstrap*, em que o valor “zero” não esteve contido nos intervalos analisados.

Na tabela 4 são apresentadas as estimações das medidas de volatilidade percebida para os 6 métodos de agregação (amostragens). A medida *rOWCov* apresentou os menores valores em todos os métodos de agregação. E método de agregação pelo último preço apresentou menores valores nas medidas *RCov*, *medRV* e *minRV*.

Tabela 4: Estimações por método de agregação: caso alta

Amostragens	<i>RCov</i>	<i>rOWCov</i>	<i>medRV</i>	<i>minRV</i>
Mediana dos preços - <i>med</i>	0,0430	0,0253	0,0356	0,0317
Ponderado pelo logaritmo do volume - <i>logvol</i>	0,0403	<b>0,0248</b>	0,0341	0,0319
Ponderado pelo número de negociações - <i>mean</i>	0,0403	<b>0,0247</b>	0,0340	0,0320
Preço ponderado pelo volume - <i>vol</i>	0,0404	0,0254	0,0348	0,0324
Preço que teve maior volume associado - <i>maxvol</i>	0,0424	0,0285	0,0395	0,0398
Último preço - <i>last</i>	<b>0,0375</b>	0,0278	<b>0,0298</b>	<b>0,0304</b>

Resultados baseados em 10.000 reamostragens. A medida *rRTSCov* não utiliza agregações.

Fonte: Dados da pesquisa.

Quando os métodos foram comparados dentro da medida *RCov*, foram observadas 9 diferenças significantes. Em tais casos, portanto, estimar o *RCov* por algum método específico gerou um valor significativamente maior/menor do que o atingido por outros meios.

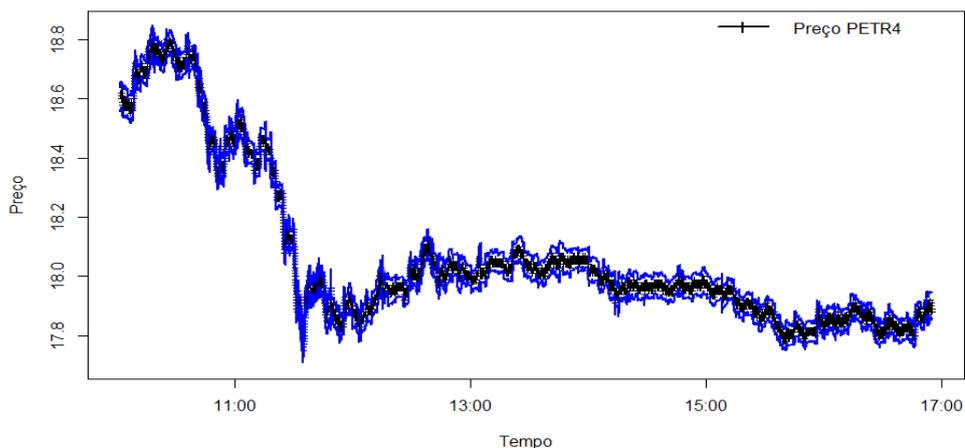
Quando os métodos foram comparados dentro da medida *rOWCov*, foram observadas 7 diferenças significantes. Para a medida *medRV*, 8 diferenças. No caso da *minRV*, apenas 5 diferenças.

Esse resultado demonstra que, para este dia específico, a medida *minRV* demonstrou menor sensibilidade à mudança de amostragem para estimar a volatilidade percebida.

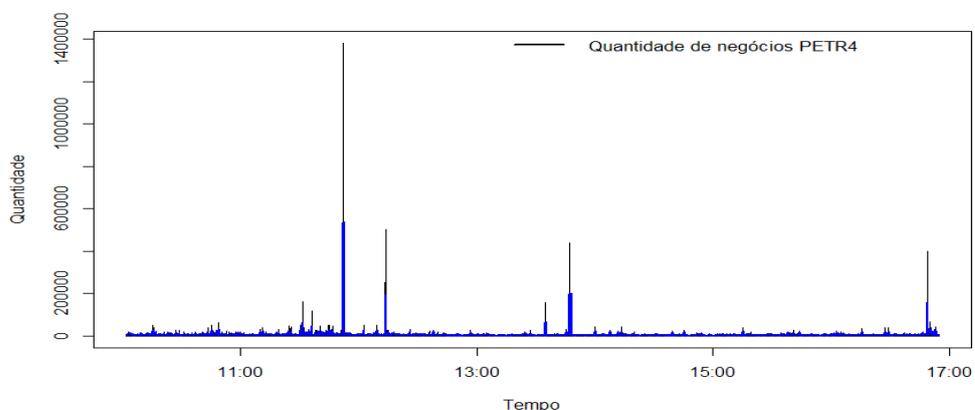
## 4.2 Dia de baixa – 08 de maio de 2014

Neste dia específico, a ação da PETR4 apresentou baixa de -3,77% e fechou o pregão ao preço de R\$17,88. Na figura 2 são apresentados os gráficos para o comportamento dos preços e volume negociados. As linhas em azul representam os intervalos para limpeza dos outliers (expressão) conforme proposta em Brownless e Gallo (2006). Os valores utilizados foram  $k = 60$  e  $\gamma = 0,02$ .

Figura 2: Comportamento preços e quantidades: caso baixa



(a) Comportamento dos preços PETR4: caso baixa.



(b) Comportamento das quantidades negociadas PETR4: caso baixa. Fonte: Dados da pesquisa.

No horário de negociação, entre 10h e 17h, a série original apresentou 58.668 observações, reduzidas para 57.596 após tratamento. Assim como ocorreu na primeira análise, os pontos aberrantes foram encontrados somente na série de volume negociado. Na tabela 6 são apresentados os resultados das análises para cada estimador da volatilidade percebida nas bases sem tratamento e tratamento de outliers.

Tabela 6: Comparações da série tratada e não tratada: caso baixa

Medidas	Série não tratada	Série tratada	Diferença	B. Lim. Inf.	B. Lim. Sup.	Sig.
<i>RCov</i>	0,0799	0,0797	0,0002	-0,0066	0,0070	não
<i>rOWCov</i>	0,0545	0,0545	0,0000	-0,0037	0,0035	não
<i>medRV</i>	0,0921	0,0920	0,0001	-0,0079	0,0079	não
<i>minRV</i>	0,0907	0,0906	0,0001	-0,0079	0,0081	não
<i>rRTSCov</i>	0,0538	0,0538	0,0000	-0,0017	0,0017	não

Resultados baseados em 10.000 reamostragens.

Fonte: Dados da pesquisa.

Conforme a tabela 6, verifica-se pouca diferença entre tratar e não tratar a série neste dia específico. Os valores para as volatilidades foram menores para a série tratada. Esse resultado foi similar ao caso em que o papel demonstrou comportamento de alta.

Na tabela 7 são apresentadas as comparações dos estimadores para os métodos de agregação dos preços em todo o período de 5 minutos (todo 5 minutos) e agregação dos preços no horário múltiplo de 5 minutos (dentro do minuto 5) para o caso de baixa no dia.

Tabela 7: Comparações tempo para agregação: caso baixa

Medidas	Todo 5 minutos	Dentro do minuto 5	Diferença	B. Lim. Inf.	B. Lim. Sup.	Sig.
<i>RCov</i>	0,0715	0,0880	-0,0165	-0,0218	-0,0104	sim
<i>rOWCov</i>	0,0486	0,0603	-0,0117	-0,0145	-0,0089	sim
<i>medRV</i>	0,0788	0,1053	-0,0265	-0,0316	-0,0208	sim
<i>minRV</i>	0,0831	0,0981	-0,0150	-0,0227	-0,0078	sim

Resultados baseados em 10.000 reamostragens. A medida *rRTSCov* não utiliza agregações de tempo.

Fonte: Dados da pesquisa.

Pela observação da tabela 7 os valores das volatilidades percebidas foram significativamente maiores neste dia específico quando foi realizada a agregação mediante utilização do horário múltiplo de 5 minutos - resultado contrário ao caso em que a ação apresentou padrão de alta.

Para este dia específico todas as medidas foram sensíveis à mudança de método sobre como o tempo foi agregado. No caso de alta da ação, 3 das 4 medidas apresentaram resultado igual. Para o caso de baixa, no entanto, todos os resultados das medidas foram alterados mediante utilização de método diverso para agregação de tempo.

Na tabela 8 são apresentadas as estimações das medidas de Volatilidade Percebida para os 6 métodos de agregações (amostragens) para o caso de comportamento de baixa.

Tabela 8: Estimações por método de agregação: caso baixa

Amostragens	<i>RCov</i>	<i>rOWCov</i>	<i>medRV</i>	<i>minRV</i>
Mediana dos preços - <i>med</i>	0,0795	0,0523	0,0912	0,0876
Ponderado pelo logaritmo do volume - <i>logvol</i>	0,0735	0,0504	0,0885	0,0884
Ponderado pelo número de negociações - <i>mean</i>	<b>0,0733</b>	<b>0,0502</b>	0,0882	0,0881
Preço ponderado pelo volume - <i>vol</i>	0,0755	0,0532	0,0905	0,0903
Preço que teve maior volume associado - <i>maxvol</i>	0,0772	0,0530	<b>0,0790</b>	<b>0,0661</b>
Último preço - <i>last</i>	0,0995	0,0677	0,1148	0,1231

Resultados baseados em 10.000 reamostragens. A medida *rRTSCov* não utiliza agregações.

Fonte: Dados da pesquisa.

Assim como no caso da alta, a medida *rOWCov* também apresentou os menores valores em todos os métodos de agregação. Diferente do caso de alta, o método de agregação pelo último preço apresentou os maiores valores das medidas de volatilidade percebida.

Quando os métodos foram comparados dentro das medidas *RCov*, *rOWCov* e *medRV*, foram observadas 5 diferenças significantes, todas observadas no método de agregação pelo último preço. Quando os métodos foram comparados dentro da medida *minRV*, foram 9 diferenças significantes observadas.

Esse resultado demonstra que, para este dia específico, a medida *minRV* implicou maior sensibilidade à mudança de amostragem para estimar a volatilidade percebida. Resultado contrário ao caso de alta, em que a medida *minRV* demonstrou a menor sensibilidade.

### 4.3 Discussão dos resultados

Buscou-se, neste trabalho, relacionar as análises apresentadas às propostas suscitadas pelas perguntas e revisão da literatura. Uma parte dos resultados ocorreu conforme a literatura estudada; contudo, demais resultados apresentaram diferenças em clara indicação de existências de características específicas do mercado brasileiro que devem ser consideradas em trabalhos futuros.

Em relação à remoção de *outliers*, diferentemente do trabalho de Brownless e Gallo (2006), no presente artigo observou-se que a remoção de *outliers* não representou influência relevante no processo de estimação da volatilidade percebida.

Esses resultados foram observados tanto no exemplo do dia de alta quanto no exemplo do dia de baixa no mercado financeiro. Além de não haver diferenças na comparação dos dias, também não ocorreram diferenças significativas tanto para as medidas dependentes de algum método de agregação (RCov, rOWCov, medRV, minRV) quanto para a medida que utiliza os dados tick by tick (rRTSCov).

Ao contrário do mercado brasileiro, cuja negociação dos preços é exclusiva à BMF&BOVESPA, no mercado americano as bolsas de valores são descentralizadas, gerando regras distintas. Desse modo, as transações no Brasil somente ocorrem dentro do spread de compra e venda (bid-ask), fator que impossibilita a ocorrência de algum preço completamente fora do padrão.

Em relação à análise de agregação dos preços, os resultados reforçam a proposição de Boudt et al. (2013), que dá importância específica à metodologia correta para tanto. Conforme pode ser observado nas análises, uma simples mudança na metodologia (todo 5 minutos para dentro do minuto 5) acarreta diferenças significativas nas estimações das volatilidades percebidas.

Para os casos de agregação pelo tempo (todo 5 minutos e dentro do minuto 5) foram encontrados indícios de que as medidas RCov, medRV, minRV são mais sensíveis à mudança de forma de agregação pelo tempo do que a medida rOWCov. Esse indício foi observado somente para o caso de comportamento de alta para o papel. Todavia, torna-se assim necessário analisar mais dias para obter evidências concretas de que esse indício seja verdadeiro.

Para a análise dos métodos de agregação considerando as 6 formas de amostragem, verificou-se que todas as medidas foram sensíveis às mudanças de forma de amostragem para agregar os preços. Para o caso de comportamento de alta, as medidas RCov, rOWCov e medRV foram as que apresentaram maior sensibilidade. Para o caso de comportamento de baixa, a medida minRV apresentou maior sensibilidade.

Os resultados obtidos estão de acordo com o apresentado por Macedo (2011), Boudt et al. (2011), Andersen et al. (2012) e Zhang e Boudt (2013), posto que a medida rCov foi mais sensível às mudanças na metodologia em relação às métricas rOWCov, medRV e minRV.

Contudo, apesar da escolha de métrica de volatilidade ser importante, observou-se a necessidade de atenção à forma como os preços são agregados, pois mesmo utilizando métricas robustas a saltos, conforme proposto por Boudt et al. (2013), a mudança na metodologia alterou de forma significativa os valores das volatilidades percebidas

## 5 Considerações finais

As análises com dados em alta frequência estão em crescimento nas pesquisas da área financeira. Apesar dos pontos positivos como a melhora da performance de modelos de projeção, ainda existem desafios a serem superados. Um exemplo de desafio é como gerenciar e tratar o grande volume de dados que as bases fornecem.

O objetivo geral do artigo foi verificar se o tratamento de dados em alta frequência é capaz de produzir estimativas consistentes da verdadeira volatilidade percebida, atingido mediante análise da metodologia para limpeza de *outliers* e dos métodos de agregação dos preços para estimação do RV.

Para este artigo foram observados os casos em que a série de PETR4 apresentou um dia de alta e um dia de baixa. Para futuros trabalhos, as séries poderão ser analisadas dentro de um período maior e poderão ser observadas séries de outros papéis.

Por meio dos resultados apresentados é possível verificar indícios (a serem analisados em pesquisas futuras) de que o tratamento de *outliers* para preços não é necessário. Diferente da regulação do mercado americano, no mercado brasileiro, todos os negócios precisam ocorrer dentro do *spread* de compra e venda (*bid-ask*).

Caso seja necessário estimar a medida  $RCov$ , há indícios de que a agregação pela utilização do último preço e pela média apresentam valores menos sensíveis aos problemas de microestrutura. Para o caso de estimar as medidas  $medRV$  e  $minRV$ , há indícios de que a agregação pela utilização do último preço e pelo preço de maior volume associado à média apresentam valores menos sensíveis aos problemas de microestrutura.

Finalmente, caso seja necessário estimar a medida  $rOWCov$ , há indícios de que os métodos de agregação pelo preço ponderado pelo logaritmo do volume e preço ponderado pelo número de negociações apresentam valores menos sensíveis aos problemas de microestrutura. O tratamento de *outliers* não causou mudanças significativas na medida  $rRTSCov$ , que utiliza todos os preços da base (*tick by tick*).

Como citado anteriormente, este artigo apresenta indícios que devem ser averiguados em pesquisas futuras. A utilização de mais dias e mais papéis é necessária, e outro ponto a ser analisado é o impacto da manutenção ou retirada de dados dos primeiros e últimos minutos de negociação em forma de variações significativas nos valores das medidas de volatilidade percebida.

## Referências

- AIÏT-SAHALIA, Y.; MYKLAND, P.; ZHANG, L. Ultra high frequency volatility estimation with dependent microstructure noise. **Journal of Econometrics**, [S.l.], v. 160, n. 01, p. 160-175, Jan. 2011.
- ANDERSEN, T. G.; BOLLERSLEV, T. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. **International Economic Review**, [S.l.], v. 39, n. 4, p. 885-905, Nov. 1998.
- ANDERSEN, T. G.; BOLLERSLEV, T.; DIEBOLD, F. X.; LABYS, P. Modeling and forecasting realized volatility. **Econometrica**, [S.l.], v. 71, n. 2, p. 579-625, Mar. 2003.
- ANDERSEN, T. G.; BOLLERSLEV, T.; DIEBOLD, F. X. Roughing it up: including jump components in the measurement, modelling and forecasting of return volatility. **The Review of Economics and Statistics**, [S.l.], v. 89, n.4, p. 701-720, Nov. 2007.
- ANDERSEN, T. G.; DOBREV, D.; SCHAUMBURG, E. Jump-robust volatility estimation using nearest neighbor truncation. **Journal of Econometrics**, [S.l.], v. 169, n. 1, p. 75-93, Jul. 2012.
- BARNDORFF-NIELSEN, O. E.; HANSEN, P. R.; LUNDE, A.; SHEPHARD, N. Realized kernels in practice: Trades and quotes. **The Econometrics Journal**, [S.l.], v.12, n. 3, p. 1-32, 2009.
- BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. Power and bipower variation with stochastic volatility and jumps. **Journal of Financial Econometrics**, Oxford University Press, v. 2, n. 1, p. 1-37, 2004.
- BAUWENS, L.; HAFNER, C.; LAURENT, S. **Volatility models and their applications**. New Jersey: John Wiley & Sons, 2012.
- BOUDT, K.; CORNELISSEN, J.; PAYSEUR, S. Highfrequency: toolkit for the analysis of highfrequency financial data in R. **Recuperado em**, [S.l.], v. 19, p.1-23, 2013.
- BOUDT, K.; CROUX, C.; LAURENT, S. Outlyingness weighted covariation. **Journal of Financial Econometrics**, [S.l.], v. 9, n. 4, p. 657-684, 2011.
- BROWNLESS, C. T.; GALLO, G. M. Financial econometric analysis at ultra-high frequency: Data handling concerns. **Computational Statistics & Data Analysis**, [S.l.], v. 51, n. 1, p. 2232-2245, 2006.
- DAROIT, D. J.; FEIL, A. A. Modelo de gestão de riscos em uma prestadora de serviços. **Revista de Ciências Administrativas**, [S.l.], v. 22, n. 2, p. 637-668, 2016.
- EFRON, B. Bootstrap methods: another look at the jackknife. **The Annals of Statistics**, [S.l.], v. 07, n. 01, p. 1-26, 1979.
- FALKENBERRY, T. N. **High frequency data filtering**: a review of the issues associated with maintaining and cleaning a high frequency financial. [S.l.]:Technical Report Tick Data, 2001.

MACEDO, H. F. **Estimação de volatilidade realizada em alta frequência na presença de microestrutura**: uma abordagem multiescala. 2011. 129f. Dissertação (Mestrado em Métodos Matemáticos em Finanças) - Instituto Nacional de Matemática Pura e Aplicada - IMPA, 2011.

MCALÉER, M.; MEDEIROS, M. M. Realized volatility: a review. **Econometric Reviews**, [S.l.], v. 27, n. 3, p. 10-45, 2008.

VAL, F. F.; PINTO, A. C. F.; KLOTZLE, M. C. Volatilidade e previsão de retorno com modelos de alta frequência e GARCH: evidências para o mercado brasileiro. **Revista Contabilidade & Finanças**, São Paulo, v. 25, n. 65, p. 189-201, 2014.

VANDERLEI, L. O. O.; CARMONA, C. U. M. A teoria das opções reais como ferramental para avaliação de projetos de investimentos sob incertezas. **Revista de Ciências Administrativas**, Fortaleza, v. 14, n. 01, p. 122-139, Ago. 2008.

WINK JÚNIOR, M. V.; PEREIRA, P. L. V. Modeling and Forecasting of Realized Volatility: Evidence from Brazil. **Brazilian Review of Econometrics**, [S.l.], v. 31, n. 2, p. 315-337, 2011.

YAN, B.; ZIVOT, E. **Analysis of High-Frequency Financial Data with S-Plus UWEC-2005-03**. Washington: University of Washington, Department of Economics, 2003.

ZHANG, J.; BOUDT, K. Jump robust two time scale covariance estimation and realized volatility budgets. **Journal Quantitative Finance, Forthcoming**, [S.l.], v 15, n.6: Themed Issue on Equities, p. 1041-1054, 2015.

ZIVOT, E. **Analysis of high frequency financial data: models, methods and software**. Part II: Modeling and forecasting realized variance measures. Disponível em: <<https://pt.scribd.com/document/38505627/Analysis-of-High-Frequency-Financial-Data-Models>>. Acesso em: 06 jan. 2017.

**Data de Submissão:** 25/11/15

**Data de Aprovação:** 3/3/17