

CORRIDA PELOS NEGÓCIOS: ANÁLISE DA VOLATILIDADE PERCEBIDA E A UTILIZAÇÃO DOS PRIMEIROS E ÚLTIMOS MINUTOS DO PREGÃO

ALCIDES CARLOS DE ARAUJO

FEA-SP

alcides.carlos@yahoo.com.br

ALESSANDRA DE AVILA MONTINI

USP - Universidade de São Paulo

amontini@usp.br

CORRIDA PELOS NEGÓCIOS: ANÁLISE DA VOLATILIDADE PERCEBIDA E A UTILIZAÇÃO DOS PRIMEIROS E ÚLTIMOS MINUTOS DO PREGÃO

Resumo: A amplitude de bases de dados provendo preços intradiários em alta frequência de ativos financeiros cresceu substancialmente nos últimos anos. Uma das aplicações mais importantes, para estes tipos de séries, foi o desenvolvimento das medidas de volatilidade utilizando dados em alta frequência. Uma das formas de mensuração que recebeu relevância nos últimos anos foi a medida de Volatilidade Percebida (*Realized Volatility – RV*), proposta por Andersen e Bollerslev (1998). O objetivo geral do artigo é analisar o tratamento de dados em alta frequência para a estimação de medidas de Volatilidade Percebida. Como objetivos específicos, buscou-se analisar a metodologia proposta por Brownless e Gallo (2006) para limpeza de *outliers* e analisou-se a importância de manter ou não os primeiros e últimos minutos do pregão para estimar a Volatilidade Percebida. No artigo foram utilizados dados de negociações *tick by tick* da ação PETR4 no período entre 09/2012 a 10/2014 para estimar as séries de Volatilidades Percebidas para cada dia. Pelos resultados, observou-se que utilizar uma amostra de preços não tratada e mantendo os primeiros e últimos 15 minutos apresentam valores significativamente maiores em relação as Volatilidades Percebidas diárias estimadas com uma amostra de preços tratada ou retirando os primeiros ou últimos minutos.

Palavras-Chave: Dados em alta frequência. Microestrutura. *Outliers*. Volatilidade Percebida.

Abstract: *The range of databases providing high frequency intraday prices grew substantially in recent years. In these type of series, one of the most important applications were the development of volatility metrics using high frequency data. One of the measurement metrics that has received importance in the recent years was the Realized Volatility, proposed by Andersen and Bollerslev (1998). The general objective of this article is to analyze the high frequency data treatment for estimation of Realized Volatility metrics. The specific objectives were to analyze the outliers identification methodology proposed by Brownless and Gallo (2006) and to analyze the importance of maintain or remove the first and last negotiations minutes for the estimation of the Realized Volatility. In the article, the daily Realized Volatility were estimated from tick by tick PETR4 prices covering November 01, 2012 to October 31, 2014. As results, the daily Realized Volatilities estimated by a non treated and with the first and last 15 minutes sample of prices demonstrated higher values when compared with the daily Realized Volatilities estimated with treated and without the first and last 15 minutes sample prices.*

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

INTRODUÇÃO

Conforme Bauwens et al. (2012) a amplitude de bases de dados provendo preços intradiários em alta frequência de ativos financeiros cresceu substancialmente nos últimos anos. Mediante esta disponibilidade, a análise de dados financeiros de alta frequência tendeu a ser mais adotada por analistas e pesquisadores de finanças.

Diante disto, diversos indicadores e medidas puderam ser desenvolvidas de forma a adaptar a utilização destes dados; uma das aplicações mais importantes foi o desenvolvimento das medidas de volatilidade utilizando dados em alta frequência.

Existem vários métodos para medir a volatilidade, uma das formas de mensuração que recebeu relevância nos últimos anos foi a medida de Volatilidade Percebida (*Realized Volatility* – RV), proposta por Andersen e Bollerslev (1998). Alguns trabalhos no Brasil como os de Pontes (2014) e Val et al (2014) traduziram o termo *Realized Volatility* como Volatilidade Realizada, no presente artigo os termos Volatilidade Percebida e Volatilidade Realizada serão sinônimas ao correspondente termo em inglês *Realized Volatility*.

Esta medida utiliza dados passados com frequência intradiária, tornando-se uma medida puramente direcionada pelos dados. Conforme os trabalhos de Zivot (2005) e Andersen et al (2007), demonstrou-se 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 destes dados possui diversos desafios devido às características das séries. Uma vez que existe uma grande quantidade de observações a serem processadas pelas capacidades computacionais da atualidade, além dos modelos precisarem estimar os parâmetros em séries com espaço de tempo irregular, como os dados *tick by tick*.

Diante deste cenário, Falkenberry (2001) apresentou que a importância da filtragem de dados cresceu significativamente. Conforme o autor, séries de dados em alta frequência possuem diversas realizações completamente fora de padrões que não representam os preços que realmente estão sendo negociados no mercado.

Diante disto, além do tratamento dos dados em relação ao processo de ordenação dos horários, verificação dos dias sem negociação/feriados e correção para horário de verão é necessário o procedimento de filtragem dos dados. Neste procedimento se busca equilibrar a forma de tratar os pontos aberrantes (*outliers*), de forma a retirar os pontos que não representam a série de dados sem modificar as 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 o RV foram propostas através dos anos. No artigo de McAleer e Medeiros (2008) é apresentada uma revisão da literatura das diversas medidas de RV propostas até o ano do artigo. Os artigos de Ait Sahalia, Mykland e Zhang (2011) e Zhang e Boudt (2013) adicionaram a literatura mais algumas formas para estimar o RV.

Um ponto pouco tratado na literatura está relacionado a estimação da Volatilidade Percebida utilizando os preços nos períodos de abertura e fechamento do pregão. De acordo com Falkenberry (2001), estes períodos são marcados por forte volatilidade que pode causar viés

em toda a análise. No artigo de Rydberg e Shephard (2003), os autores recomendam retirar estas observações para que as estimações não sofram os efeitos dos leilões de abertura e fechamento das séries.

No Brasil, antes do início das negociações existe o leilão de pré-abertura (*call* de abertura) que ocorre 15 minutos antes do início das negociações e há somente registro de ofertas. No final das negociações existe o leilão de fechamento (*call* de fechamento) que ocorre durante os 5 minutos que antecedem o término as 17:00h.

Em trabalhos que foram estimadas Volatilidades Percebidas como em Wink Júnior e Pereira (2011), Pontes (2014), Val et al. (2014) e Santos e Ziegelmann (2014) não foi detalhado se os períodos de abertura e fechamento foram utilizados para estimação.

Mediante a importância de estimar corretamente a Volatilidade Percebida diária, seria coerente considerar os primeiros minutos de negociação? Seria coerente manter os últimos minutos? O tratamento dos *outliers* auxiliaria para obtenção de medidas mais consistentes?

Diante destas perguntas, o objetivo geral do artigo é analisar o tratamento de dados em alta frequência para a estimação de medidas de Volatilidade Percebida.

Como objetivos específicos, buscou-se analisar a metodologia proposta por Brownless e Gallo (2006) para limpeza de *outliers* e analisou-se a importância de manter ou não os primeiros e últimos minutos do pregão para estimar a Volatilidade Percebida.

2 REVISÃO DA LITERATURA

Para entender os diversos conceitos utilizados na pesquisa, torna-se necessário a apresentação dos principais elementos a serem utilizados neste artigo. Os tópicos a serem apresentados serão processo de filtragem, agregação das negociações e medidas de Volatilidade Percebida.

2.1 Processo de Filtragem dos Dados

No artigo de Falkenberry (2001), o autor apresenta o conceito base para o processo de filtragem de séries de dados em alta frequência. Conforme o autor, o objetivo primário é a remoção de *outliers*, porém mantendo as propriedades de séries que são observadas em tempo real.

Por exemplo, existem saltos nas séries de dados observados em tempo real que não podem ser classificados como *outlier*, mas como um comportamento normal de mercado quando ocorre chegada de novas notícias ou quando alto volume financeiro é transacionado.

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 o mesmo deve ser retirado ou trocado 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. Sendo que os filtros utilizados nesta abordagem podem ser aplicados na prática e em tempo real facilmente.

O artigo de Brownless e Gallo (2006) discute os procedimentos, de forma detalhada, para limpeza e gerenciamento de séries financeiras de alta frequência.

Primeiramente, deve-se detectar e remover observações gravadas de forma errada na base, como preços iguais a zero; na base de dados oferecida pela bolsa de Nova York (NYSE) existe um identificador específico para estes preços. Além disto, busca-se 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 γ é um parâmetro para redução de ruído.

Conforme se observa na expressão 1, os parâmetros k e γ definem a quantidade de *outliers* a serem detectados na série. Brownless e Gallo (2006) demonstraram que a medida que os valores de k e γ aumentavam o número de *outliers* detectados era menor.

No mesmo artigo de Brownless e Gallo (2006) os autores buscaram analisar a aplicação da limpeza e tratamento de dados em alta frequência para o caso de estimar Durações financeiras (tempo estimado em que ocorre uma variação no preço acima de um patamar estabelecido). Os autores observaram que a remoção de *outliers* diminuiu significativamente o número de Durações (quantidade de vezes que o preço variou acima de um patamar estabelecido).

Em relação ao tamanho das Durações, a série tratada demonstrou maior período de Duração, em torno de 10 minutos, enquanto que a série não tratada apresentou aproximadamente 7 minutos. Diante disto, os autores demonstraram que a série não tratada pode superestimar o número de mudanças de preços em estudos sobre Durações financeiras.

2.2 Estimação da Volatilidade Percebida

Durante os anos, diversos métodos foram propostos 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 deste processo está fora do escopo do presente artigo, somente os princípios básicos são apresentados para entendimento do porquê de agregar as negociações ou estimar usando todos os dados *tick by tick*.

O processo é baseado em que existem M negociações no período T , estas negociações são igualmente espaçadas no tempo; deste modo ocorrem realizações na série a cada $\Delta t = T/M$

unidades de tempo. Diante disto, 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, uma vez que é uma variável latente. Estes valores são estimados pelas medidas de Volatilidade Percebida.

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

Entretanto, conforme é apresentado em Aït Sahalia, Mykland e Zhang (2011), Macedo e Zhang e Boudt (2013), ao estimar a Volatilidade Percebida de forma empírica, o erro em relação ao verdadeiro valor de $IV_{t,s}$ tende a aumentar a medida que cresce a frequência amostral.

A explicação para este problema está relacionado a falta de robustez do estimador da Volatilidade Percebida quando o intervalo entre cada negociação é cada vez menor. Em frequências de tempo muito curtas, o impacto das imperfeições de mercado são claramente evidentes. Estas imperfeições são frequentemente citadas na literatura como saltos e ruídos de microestrutura.

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. Conforme, os autores a medida é robusta a presença de saltos e ruídos de microestrutura, a medida denominada *Robust two time scale covariance estimator - rRTSCov* é apresentada na expressão 8,

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

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

2.3 Comportamento das negociações nos primeiros e últimos minutos do pregão

O estudo a respeito do comportamento intradiário de retornos de ativos financeiros possui uma vasta e bem documentada literatura. Os principais trabalhos estão concentrados nos anos

de 1980 e 1990, após este período os principais artigos focaram em estudar os impactos das decisões regulatórias e o desenvolvimento de medidas de mensuração, como medidas de risco e liquidez.

Os principais artigos do tema foram apresentados por Wood, McNish e Ord (1985) e Harris (1986). Em ambas as pesquisas foram analisadas o índice da bolsa de valores de Nova Iorque (NYSE *index*); no primeiro, um período entre 1971 e 1982 foi estudado, no segundo, o período entre 1981 a 1983 foi analisado.

Conforme, Wood, McNish e Ord (1985) para cada dia analisado foi observado o padrão em que nos primeiros e últimos minutos do dia ocorria alta volatilidade. Sendo que após a primeira hora, a volatilidade decaía e permanecia baixa até os últimos minutos de negociação.

Em Harris (1986) foi encontrado padrão similar, diante disto os autores denominaram este padrão como *U-shape*. O autor também demonstrou que a volatilidade no começo das negociações é relativamente maior de que a volatilidade das últimas negociações.

McNish e Wood (1990) ampliam o estudo ao analisar os dados da bolsa de Toronto (*Toronto Stock Exchange* - TSE). Conforme os autores, os retornos e o número de negócios realizados demonstraram o padrão *U-shape*, sendo observada maior volatilidade nos últimos minutos de negociação.

Na pesquisa de Baillie e Bollerslev (1991), os autores buscaram analisar os padrões intradiários dos mercados de moeda do Reino Unido (*British Pound* - BP), Alemanha (*German Deutschmark* - DM), Suíça (*Swiss Franc* - SF) e Japão (YEN). Novamente, o padrão *U-shape* foi observado.

McNish e Wood (1992) buscaram analisar a existência do padrão em formato de U para a diferença entre as melhores ofertas de compra e venda (*bid-ask spread*) nas ações negociadas na bolsa de Nova Iorque. Diferentemente dos resultados para retornos e volume, o *bid-ask spread* apresentou formato de J, isto é, maior diferença no fechamento do pregão.

No artigo de Webb e Smith (1994), os autores estudaram as bolsas de Chicago (*Chicago Mercantile Exchange* - CME), Londres (*London International Financial Futures Exchange* - LIFFE) e Singapura (*Singapore International Monetary Exchange* - SIMEX). Os resultados relacionados aos padrões intradiários foram similares aos estudos anteriores, em que existiu maior volatilidade na abertura das negociações.

No artigo de Docking, Kawaller e Koch (1999) os autores analisaram dados intradiários buscando evidenciar a existência de saltos de volatilidade no meio do dia. Os resultados dos autores também demonstraram a existência do formato em U, entretanto encontraram evidências para aceitar a hipótese de existência de saltos de volatilidade no meio do dia.

Para explicar o padrão encontrado na análise dos retornos intradiários, Gerety e Mulherin (1992) apresentam que este comportamento é resultado da interação entre investidores informados e desinformados. Conforme os autores, o formato de U é explicado pela aversão dos investidores ao risco entre o fechamento e a abertura (*overnight*).

Gerety e Mulherin (1992) explicam que o alto volume na abertura do pregão reflete as intenções antes das negociações começarem. A volatilidade observada durante o fechamento

reflete o risco esperado para o período *overnight*, sendo que as novas informações apresentadas durante este período poderão afetar a atividade de negociação na abertura.

Conforme Webb e Smith (1994), os efeitos de abertura e fechamento são explicados pela iliquidez do mercado. Durante o pregão são ofertadas diversas ordens visando execução imediata, entretanto, os formadores de mercados decidem manter posição até o momento em que estes possuem uma melhor ideia de preços, observados durante o período de abertura e/ou fechamento.

Estas explicações foram testadas em Hong e Wong (2000) para os casos de simetria e assimetria de informação. Para o caso de simetria, os investidores diminuem as posições visando evitar o risco no período *overnight*, sendo que a maior volatilidade na abertura em relação ao fechamento ocorre pela maior sensibilidade dos investidores as novas informações apresentadas no período *overnight*.

Para o caso de assimetria, os autores apresentam que durante o *overnight* o grau de assimetria entre os investidores tende a aumentar. Durante a abertura, os investidores informados revelam parte da informação privada, causando maior negociação pelos investidores desinformados.

Durante o fechamento, os investidores informados diminuem suas posições no mercado, diminuindo o grau de informação privada apresentada. Diante disto, os investidores menos informados decidem modificar suas posições, também causando maior volatilidade neste período.

Durante os anos 90, diversas bolsas em todo o mundo adotaram políticas de pregões de abertura e fechamento (*opening and closing call's*). Diante disto, após este período, os principais artigos focaram em estudar os impactos das decisões regulatórias e o desenvolvimento de medidas de mensuração, como medidas de risco e liquidez.

No Quadro 1, são resumidas as pesquisas a respeito do comportamento intradiário dos preços de ativos financeiros.

Quadro 1 – Trabalhos sobre comportamento intradiário preços

Autor	Dados	Período	Resultados
Wood, McNish e Ord (1985)	NYSE <i>index</i>	1971 a 1972 e 1982	Maior volatilidade na abertura e fechamento
Harris (1986)	NYSE <i>index</i>	Dez. 1981 a Jan. 1983	U-shape, mais volatilidade na abertura
McNish e Wood (1990)	Toronto Stock Exchange (TSE)	Fev. 1987	U-shape para volatilidade e número de negociações
Baillie e Bollerslev (1991)	BP, DM, SF, YEN	Jan.-Jul. 1986	U-shape
McNish e Wood (1992)	Ações da NYSE	Jan.-Jun. 1989	J-shape para o <i>bid-ask spread</i>
Webb e Smith (1994)	CME, LIFFE e SIMEX	Nov. 1987 a Out. 1990	U-shape, mais volatilidade na abertura
Docking, Kawaller e Koch (1999)	Contratos futuros Eurodollars e Deutsche	Set. 1988 a Jun. 1995	U-shape, mais volatilidade na abertura, rápidos picos durante o dia

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. Como objetivos específicos, buscou-se analisar a metodologia proposta por Brownless e Gallo (2006) para limpeza de *outliers*, verificar a importância de manter ou retirar as observações dos primeiros e últimos 15 minutos de negociação.

A escolha dos 15 minutos ocorreu pelo fato de que as pesquisas no Brasil apresentaram diferenças quanto a frequência do tempo para estimação da Volatilidade Percebida diária. No artigo de Wink Júnior e Pereira (2011) foram analisadas frequências de 5, 15 e 30 minutos; em Pontes (2014) e Val et al. (2014) frequências de 5 minutos e Santos e Ziegelmann (2014) frequência de 15 minutos. Deste modo, o período de 15 minutos consegue ser abrangente em todas estas pesquisas.

Neste artigo são utilizados dados de negociações *tick by tick* da ação PETR4 no período entre 09/2012 a 10/2014. Por meio dos dados intradiários foram estimadas as séries de Volatilidades Percebidas diárias, ao todo, foram 8 séries estudadas; buscando compreender as diferenças entre séries tratada/não tratada, com primeiros/sem primeiros 15 minutos e com últimos/sem últimos 15 minutos. No Quadro 2 são identificadas as séries estudadas.

A identificação dos *outliers* seguiu conforme o método proposto em Brownless e Gallo (2006), apresentado na expressão 1, foram analisadas as séries *tick by tick* tanto dos preços quanto do volume. Os parâmetros de observações mais próximas (k) e redução de ruído (γ) foram definidos com $k = 60$ e $\gamma = 0,02$; assim como Brownless e Gallo (2006) utilizaram no respectivo artigo.

A medida de Volatilidade Percebida utilizada foi a *Robust two time scale covariance estimator* - *rRTSCov* (expressão 8) apresentada em Aït Sahalia, Mykland e Zhang (2011) e Zhang e Boudt (2013). Esta medida é estimada utilizando dados *tick by tick* e apresenta robustez a presença de saltos e ruídos de microestrutura.

Mediante os objetivos traçados, as seguintes hipóteses foram testadas:

- $H_{0,1}$ - Não há diferenças de Volatilidades Percebidas entre as metodologias.
- $H_{0,2}$ - Não há diferença de Volatilidades Percebidas entre tratar e não tratar a série de preços em relação a *outliers*.
- $H_{0,3}$ - Não há diferença de Volatilidades Percebidas entre utilizar os primeiros 15 minutos e não utilizar os primeiros 15 minutos.
- $H_{0,4}$ - Não há diferença de Volatilidades Percebidas entre utilizar os últimos 15 minutos e não utilizar os últimos 15 minutos.

Para testar as hipóteses, primeiramente, realizaram-se testes de razão de variâncias para cada dia. Conforme Andersen, Bollerslev e Das (2001) nestes testes busca-se verificar a hipótese nula de não haver mudança da volatilidade dos retornos de um dia específico. O teste é

expresso como V_a/V_b , em que $\hat{V}_a/\hat{V}_b \sim F_{n_a-1, n_b-1}$. Sendo V_a e V_b as volatilidades de uma série a e de uma série b respectivamente, n_a o tamanho da série a e n_b o tamanho da série b .

Conforme os autores, o teste não demonstrou robustez para analisar a razão de variâncias para dados em alta frequência, principalmente pelos problemas observados nestes dados como saltos e ruídos de microestrutura. Contudo, dado que o presente artigo estima a Volatilidade Percebida por meio de um estimador robusto proposto em Aït Sahalia, Mykland e Zhang (2011) e Zhang e Boudt (2013) realizar o teste clássico para razão de variâncias não é um problema.

Após a realização dos testes de razão de variâncias, foram contados os dias em que as diferenças foram significativas ($\alpha < 0,05$). Por meio destas contagens, realizaram-se Testes χ^2 no cruzamento das frequências de razões de variâncias significantes entre as variáveis.

Quadro 2 – Identificação das séries estudadas

Códigos	Variáveis
c_cp15_cu15	Vol. Perc. série não tratada contendo os primeiros e últimos 15 minutos
c_cp15_su15	Vol. Perc. série não tratada com os primeiros e sem os últimos 15 minutos
c_sp15_cu15	Vol. Perc. série não tratada sem os primeiros e com os últimos 15 minutos
c_sp15_su15	Vol. Perc. série não tratada sem os primeiros e sem os últimos 15 minutos
t_cp15_cu15	Vol. Perc. série tratada contendo os primeiros e últimos 15 minutos
t_cp15_su15	Vol. Perc. série tratada com os primeiros e sem os últimos 15 minutos
t_sp15_cu15	Vol. Perc. série tratada sem os primeiros e com os últimos 15 minutos
t_sp15_su15	Vol. Perc. série tratada sem os primeiros e sem os últimos 15 minutos

Fonte: elaborado pelos autores.

4 ANÁLISES

As análises dos resultados foram direcionadas pelas perguntas: Seria coerente considerar os primeiros minutos de negociação? Seria coerente manter os últimos minutos? O tratamento dos *outliers* auxiliaria para obtenção de medidas mais consistentes?

Na Tabela 1 são apresentadas as análises descritivas das séries de Volatilidades Percebidas. Os resultados foram similares aos encontrados na literatura, séries com alta curtose e assimetria a esquerda; conforme pode ser observado nos artigos de Andersen *et al.* (2007), Forsberg e Ghysels (2007) e Santos e Ziegelmann (2014).

Tabela 1 – Análise descritiva

Variáveis	Média de negociações por dia	Média	Desvio Padrão	Assimetria	Curtose
c_cp15_cu15	37975	0,000314528	0,000245431	2,44562	10,089
c_cp15_su15	36728	0,000309597	0,000242324	2,43067	9,968
c_sp15_cu15	36969	0,00029397	0,000233853	2,53557	10,723
c_sp15_su15	35722	0,000288797	0,000230677	2,52464	10,629
t_cp15_cu15	36643	0,000312537	0,000245676	2,45996	10,191
t_cp15_su15	35438	0,000307539	0,000242656	2,44605	10,068
t_sp15_cu15	35726	0,000293477	0,000234061	2,54132	10,766
t_sp15_su15	34521	0,000288373	0,000230953	2,53196	10,687

Fonte: dados da pesquisa

Como contribuição, buscou-se verificar a Volatilidade Percebida (RV) média para cada forma de amostragem. Como esperado, a Volatilidade Percebida média foi maior quando estimada utilizando a série de preços não tratada e mantendo os primeiros e últimos 15 minutos.

Também se observa a menor Volatilidade Percebida média para a série de preços tratada e retirando os primeiros e últimos de cada dia pesquisado.

Observa-se pela análise da Tabela 1 que o número de negociações utilizadas para estimação da RV apresentou maior diminuição quando os primeiros 15 minutos não foram utilizados. Isto demonstra que indícios de maior volume e volatilidade no começo do dia quando comparados com o final do dia. Resultados conforme a revisão da literatura levantada.

Na Tabela 2 são apresentados resultados dos testes de razão de variâncias, conforme explicado na metodologia, para cada dia foram comparadas as Volatilidades Percebidas e realizado um teste F. Depois foram contados os dias que apresentaram razão significativamente maior que 1, ou seja, um nível descritivo menor que a significância adotada no artigo de 0,05.

Na Tabela 3 são apresentadas as diferenças médias entre as Volatilidades Percebidas conforme o tratamento adotado e método de amostragem utilizada. Pela observação de ambas as tabelas, pode-se perceber que conforme são encontradas maiores frequências de Volatilidades Percebidas superiores a 1, observam-se maiores respectivas diferenças médias.

Os maiores valores nas Tabelas 2 e 3 são 517 e 0,1010 que representam o número de razões de variâncias significantes e diferença média respectivamente para a comparação das Volatilidades Percebidas estimadas utilizando a amostra sem tratamento e contendo os primeiros e últimos 15 minutos e amostra com tratamento e não contendo os primeiros e últimos 15 minutos.

Os menores valores nas Tabelas 2 e 3 são 39 e 0,00258 que representam o número de razões de variâncias significantes e diferença média respectivamente para a comparação das Volatilidades Percebidas estimadas utilizando a amostra sem tratamento e não contendo os primeiros e últimos 15 minutos e amostra com tratamento e não contendo os primeiros e últimos 15 minutos.

Estes resultados apresentam indícios de que podem existir diferenças significantes entre as metodologias. Para testar a Hipótese nula 1 ($H_{0,1}$ - Não há diferenças entre as metodologias) realizou-se um Teste χ^2 no cruzamento das frequências de razões de variâncias significantes entre as variáveis c_{cp15_cu15} , c_{cp15_su15} , c_{sp15_cu15} e c_{sp15_su15} contra as variáveis t_{cp15_cu15} , t_{cp15_su15} , t_{sp15_cu15} e t_{sp15_su15} .

O teste apresentou resultados de $\chi^2 = 1445,31$ e $p = 0,000$; demonstrando que há evidências para rejeitar $H_{0,1}$. Deste modo, as Volatilidades Percebidas diárias estimadas utilizando uma amostra de preços não tratada e mantendo os primeiros e últimos 15 minutos apresentam valores significativamente maiores em relação as Volatilidades Percebidas diárias estimadas com uma amostra de preços tratada ou retirando os primeiros ou últimos minutos.

Dado que a Hipótese nula 1 é uma hipótese geral, buscou-se testar as Hipóteses nulas 2, 3 e 4 para verificar em qual aspecto as metodologias para estimação da Volatilidade Percebida diária podem se diferenciar.

Tabela 2 – Frequências de valores abaixo do nível de significância ($\alpha < 0,05$)

Frequências	rv_c_cp15_cu15	rv_c_cp15_su15	rv_c_sp15_cu15	rv_c_sp15_su15	rv_t_cp15_cu15
c_cp15_cu15	-	-	-	-	-
c_cp15_su15	202	-	-	-	-
c_sp15_cu15	505	446	-	-	-
c_sp15_su15	516	503	225	-	-
t_cp15_cu15	120	220	474	509	-
t_cp15_su15	308	118	419	478	199
t_sp15_cu15	500	455	41	230	481
t_sp15_su15	517	501	265	39	513

Fonte: dados da pesquisa

Tabela 3 – Diferenças médias entre as Volatilidades Percebidas

Diferenças	rv_c_cp15_cu15	rv_c_cp15_su15	rv_c_sp15_cu15	rv_c_sp15_su15	rv_t_cp15_cu15
c_cp15_cu15	-	-	-	-	-
c_cp15_su15	0,01843858	-	-	-	-
c_sp15_cu15	0,07578528	0,057095628	-	-	-
c_sp15_su15	0,09809499	0,078302096	0,020787466	-	-
t_cp15_cu15	0,00906214	-0,008593597	-0,060792899	-0,079373833	-
t_cp15_su15	0,02827419	0,009618287	-0,042857058	-0,062404446	0,018985265
t_sp15_cu15	0,07866807	0,0598982	0,002655474	-0,017118915	0,069022688
t_sp15_su15	0,10101897	0,081118824	0,023465902	0,002579379	0,091110041

Fonte: dados da pesquisa.

Na Hipótese Nula 2 ($H_{0,2}$ - Não há diferença entre tratar e não tratar a série de preços em relação a *outliers*) realizou-se um Teste χ^2 para verificar as diferenças entre as colunas “Não Tratada” e “Tratada” na Tabela 4.

Tabela 4 – Não tratada v.s. Tratada

Não Tratada				Tratada			
Variáveis	c_cp15_cu15	c_cp15_su15	c_sp15_cu15	Variáveis	t_cp15_cu15	t_cp15_su15	t_sp15_cu15
c_cp15_su15	202	-		t_cp15_su15	199	-	
c_sp15_cu15	505	446	-	t_sp15_cu15	481	425	-
c_sp15_su15	516	503	225	t_sp15_su15	513	482	223

Fonte: dados da pesquisa.

O teste apresentou resultados de $\chi^2 = 3,0859$ e $p = 0,5435$; demonstrando que não há evidências para rejeitar $H_{0,2}$. Deste modo, os testes de razão de variâncias utilizando as amostras de preços não tratadas não apresentam valores significativamente maiores em relação aos testes com as amostras de preços tratados.

Na Hipótese Nula 3 ($H_{0,3}$ - Não há diferença entre utilizar os primeiros 15 minutos e não utilizar os primeiros 15 minutos) realizou-se um Teste χ^2 para verificar as diferenças entre as colunas “Com Primeiros 15 minutos” e “Sem Primeiros 15 minutos” na Tabela 5.

Tabela 5 – Com primeiros 15 min. v.s. Sem os primeiros 15 min.

Com Primeiros 15 minutos				Sem Primeiros 15 minutos			
Variáveis	c_cp15_cu15	c_cp15_su15	t_cp15_cu15	Variáveis	c_sp15_cu15	c_sp15_su15	t_sp15_cu15
c_cp15_su15	202	-		c_sp15_su15	225	-	
t_cp15_cu15	120	220	-	t_sp15_cu15	41	230	-
t_cp15_su15	308	118	199	t_sp15_su15	265	39	223

Fonte: dados da pesquisa.

O teste apresentou resultados de $\chi^2 = 116,8692$ e $p = 0,000$; demonstrando que há evidências para rejeitar $H_{0,3}$. Deste modo, os testes de razão de variâncias utilizando as amostras de preços com os primeiros 15 minutos apresentam valores significativamente maiores em relação aos testes com as amostras de preços sem os primeiros 15 minutos.

Na Hipótese Nula 4 ($H_{0,4}$ - Não há diferença entre utilizar os últimos 15 minutos e não utilizar os últimos 15 minutos) realizou-se um Teste χ^2 para verificar as diferenças entre as colunas “Com Últimos 15 minutos” e “Sem Últimos 15 minutos.” na Tabela 6.

Tabela 6 – Com últimos 15 min. v.s. Sem os últimos 15 min.

Com últimos 15 minutos				Sem últimos 15 minutos			
Variáveis	c_cp15_cu15	c_sp15_cu15	t_cp15_cu15	Variáveis	c_cp15_su15	c_sp15_su15	t_cp15_su15
c_sp15_cu15	505	-		c_sp15_su15	503	-	
t_cp15_cu15	120	474	-	t_cp15_su15	118	478	-
t_sp15_cu15	500	41	481	t_sp15_su15	501	39	482

Fonte: dados da pesquisa.

O teste apresentou resultados de $\chi^2 = 0,1766$ e $p = 0,9963$; demonstrando que não há evidências para rejeitar $H_{0,4}$. Deste modo, os testes de razão de variâncias utilizando as amostras de preços com os últimos 15 minutos não apresentam valores significativamente maiores em relação aos testes com as amostras de preços sem os últimos 15 minutos.

5 CONSIDERAÇÕES FINAIS

O principal objetivo do presente artigo foi analisar o tratamento de dados em alta frequência para a estimação de medidas de Volatilidade Percebida. Como objetivos específicos, buscou-se analisar a metodologia proposta por Brownless e Gallo (2006) para limpeza de *outliers* e analisou-se a importância de manter ou não os primeiros e últimos minutos do pregão para estimar a Volatilidade Percebida.

Os objetivos apresentados foram traçados mediante as perguntas: Seria coerente considerar os primeiros minutos de negociação? Seria coerente manter os últimos minutos? O tratamento dos *outliers* auxiliaria para obtenção de medidas mais consistentes?

Estas indagações são levantadas mediante a crescente utilização no Brasil de dados em alta frequência nas pesquisas. Como exemplo, podem ser apresentados os artigos de Wink Júnior e Pereira (2011), Pontes (2014), Val et al. (2014) e Santos e Ziegelmann (2014).

Um ponto pouco tratado na literatura está relacionado a estimação da Volatilidade Percebida utilizando os preços nos períodos de abertura e fechamento do pregão. De acordo com Falkenberry (2001), estes períodos são marcados por forte volatilidade que pode causar viés em toda a análise.

No artigo de Rydberg e Shephard (2003), os autores recomendam retirar estas observações para que as estimações não sofram os efeitos dos leilões de abertura e fechamento das séries. Uma vasta literatura de finanças demonstraram o comportamento em formato de U da volatilidade nas séries intradiárias de ativos financeiros, sendo que a volatilidade na abertura tende a ser maior de que no fechamento, conforme é observado no Quadro 1.

Para atingir os objetivos traçados foram utilizados dados *tick by tick* do papel PETR4 no período entre 09/2012 a 10/2014. Estimaram-se 8 séries de Volatilidades Percebidas para cada dia buscando compreender as diferenças entre séries tratada/não tratada, com primeiros/sem primeiros 15 minutos e com últimos/sem últimos 15 minutos.

Pelos resultados, observou-se que utilizar uma amostra de preços não tratada e mantendo os primeiros e últimos 15 minutos apresentam valores significativamente maiores em relação as Volatilidades Percebidas diárias estimadas com uma amostra de preços tratada ou retirando os primeiros ou últimos minutos.

Com objetivo de detalhar as análises, observou-se que as amostras de preços não tratadas apresentaram valores de Volatilidades Percebidas maiores em relação as amostras com preços tratados. Contudo, as diferenças não foram significativas. Este resultado foi similar quando foram comparadas as Volatilidades nas amostras de preços com os últimos e sem os últimos 15 minutos.

Quando foram comparadas as Volatilidades nas amostras contendo os primeiros 15 minutos, observou-se valores maiores e significantes em relação aos testes com as amostras de preços sem os primeiros 15 minutos.

Diante deste resultados, encontram-se indícios de que realizar um tratamento adequado das séries tanto em pesquisas como em consultorias é de extrema importância. Considerar as

recomendações de Rydberg e Shephard (2003) a respeito dos efeitos dos leilões de abertura e fechamento pode levar a estimações mais consistentes da Volatilidade Percebida.

A decisão a respeito de manter ou retirar os primeiros/últimos negócios na estimação de medidas de Volatilidade Percebida é uma decisão individual de cada pesquisador. Para o papel e período utilizado no presente artigo, recomenda-se maior atenção aos primeiros 15 minutos de que aos últimos 15 minutos ou tratar a série para possíveis *outliers* (a medida $rRTSCov$, expressão 8, é robusta para saltos e ruídos na série).

Diante disto, no presente artigo, foram encontradas diferenças significativas ao deletar os primeiros 15 minutos de negociação. Em pesquisas futuras, mais ativos financeiros e diferentes medidas de estimar a Volatilidade Percebida podem ser consideradas. Desta forma, resultados abrangentes poderão ser apresentados.

REFERÊNCIAS

AÏT-SAHALIA, Y.; MYKLAND, P.; ZHANG, L. *Ultra high frequency volatility estimation with dependent microstructure noise*. ***Journal of Econometrics***. Vol. 160, n. 01, p. 160-175, 2011.

ANDERSEN, T. G.; BOLLERSLEV, T. *Answering the skeptics: yes, standard volatility models do provide accurate forecasts*. ***International Economic Review***, v. 39, n. 4, p. 885-905, 1998.

ANDERSEN, T. G.; BOLLERSLEV, T.; DAS, A. *Variance-ratio statistics and high-frequency data: testing for changes in intraday volatility patterns*. ***The Journal of Finance***, v. 56, n. 01, 2001.

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***, v. 89, p. 701-720, 2007.

BAILLIE, R. T.; BOLLERSLEV, T. *Intra-day and inter-market volatility in foreign exchange rates*. ***The Review of Economic Studies***, v. 58, n. 3, p. 565-585, 1991.

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 high frequency financial data in R*. ***Working Paper***, 2013.

BROWNLESS, C. T.; GALLO, G. M. *Financial econometric analysis at ultra-high frequency: Data handling concerns*. ***Computational Statistics & Data Analysis***, v. 51, n. 1, p. 2232-2245, 2006.

DOCKING, D. S.; KAWALLER, I. G.; KOCH, P. D. *Mid-Day volatility spikes in U.S. futures markets. The Journal of Futures Markets*, v. 19, n. 2, p. 195-216, 1999.

FALKENBERRY, T. N. *High frequency data filtering. Technical Report Tick Data*, 2001.

FORSBERG, L.; GHYSELS, E. *Why do absolute returns predict volatility so well? Journal of Financial Econometrics*, v. 5, n. 01, p. 31-67, 2007.

GERETY, M. S.; MULHERIN, J. H. *Trading halts and market activity: an analysis of volume at the open and the close. The Journal of Finance*, v. 47, n. 5, p. 1765-1784, 1992.

HARRIS, L. *A transaction data study of weekly and intradaily patterns in stock returns. Journal of Financial Economics*, v. 16, p. 99-117, 1986.

HONG, H.; WANG, J. *Trading and returns under periodic market closures. The Journal of Finance*, v. 55, n. 01, p. 297-354, 2000.

MACEDO, H. F. **Estimação de Volatilidade Realizada em Alta Frequência na Presença de Microestrutura: Uma Abordagem Multiescala.** 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*, v. 27, n. 3, p. 10-45, 2008.

McINISH, T. H.; WOOD, R. A. *An analysis of transactions data for the Toronto Stock Exchange: return patterns and end-of-the-day effect. Journal of Banking and Finance*, v. 14, p. 441-458, 1990.

McINISH, T. H.; WOOD, R. *An analysis of intraday patterns in bid/ask spreads for NYSE stocks. The Journal of Finance*, v. 47, n. 2, p. 753-764, 1992.

PONTES, T. T. S. *Previsão da volatilidade realizada: o impacto dos saltos na série do Ibovespa.* In: ENCONTRO DA ANPAD, 38., 2014, Rio de Janeiro. **Anais...** Rio de Janeiro: EnANPAD, 2014.

RYDBERG, T. V.; SHEPHARD, N. *Dynamics of trade-by-trade price movements: decomposition and models. Journal of Financial Econometrics*, v. 01, n. 01, p. 2-25, 2003.

SANTOS, D. G.; ZIEGELMANN, F. A. *Volatility forecasting via MIDAS, HAR and their combination: an empirical comparative study for Ibovespa. Journal of Forecasting*, v. 33, n. 04, p. 284-299, 2014.

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

WEBB, R. I.; SMITH, D. G. *The effect of market opening and closing on the volatility of Eurodollar futures prices. The Journal of Futures Markets*, v. 14, n. 1, p. 51-78, 1994.

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

WOOD, R. A.; MCINISH, T. H.; ORD, J. K. *An investigation of transactions data for NYSE Stocks. The Journal of Finance*, v. 40, n. 03, p. 723-739, 1985.

YAN, B.; ZIVOT, E. *Analysis of high-frequency financial data with s-plus. Technical Report UWEC-2005-03*, 2003.

ZHANG, J.; BOUDT, K. *Jump robust two time scale covariance estimation and realized volatility budgets. Quantitative Finance*, n. ahead-of-print, p. 1-14, 2013.

ZIVOT, E. *Analysis of high frequency financial data: models, methods and software. Part II: Modeling and forecasting realized variance measures. Technical Report UWEC-2005-03*, 2005.