BBR
BRAZILIAN BUSINESS REVIEW

ARTIGO

Análise dos Algoritmos *Tick Rule* e *Bulk Volume Classification* no Mercado Acionário Brasileiro

Leonardo Souza Siqueira[1]
leonardosiq91@gmail.com | 0000-0002-7485-8590

Laíse Ferraz Correia[2]
laise@cefetmg.br | 0000-0002-0977-9298

Hudson Fernandes Amaral[3]
hfamaral.cepead@gmail.com| 0000-0001-8455-0285

RESUMO

O objetivo deste artigo foi comparar o desempenho dos algoritmos *Tick Rule* (TR) e *Bulk Volume Classification* (BVC) na classificação de transações de ações negociadas na B3 e, assim, indicar o melhor método como ferramenta de auxílio às decisões de investimento. Os ativos foram separados em três grupos conforme o volume transacionado. Os dados reais foram utilizados para verificar a acurácia dos algoritmos, sendo as informações de 2018 utilizadas para estimar os melhores parâmetros do BVC, e as de 2019, para testar a sua eficiência. Posteriormente, foi calculado o *Volume-Synchronized Probability of Informed Trading* (VPIN) para cada ação utilizando o TR e o BVC. Esses valores foram comparados com o VPIN apurado com os dados reais. Observou-se que o TR apresenta melhor performance em relação ao BVC para todos os três grupos de ações. As análises das propriedades dos métodos revelaram que a base na qual o TR está calcado se sustenta no mercado brasileiro, enquanto a mecânica do BVC não reflete a realidade.

PALAVRAS-CHAVE

Tick Rule, Bulk Volume Classification, VPIN, Microestrutura de mercado

[1]Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

[2]Centro Federal de Educação Tecnológica de Minas Gerais – Programa de Pós-graduação em Administração, Belo Horizonte, Minas Gerais, Brazil

[3]Centro Universitário Unihorizontes, Belo Horizonte, Minas Gerais, Brazil

Recebido: 18/02/2021.
Revisado: 28/09/2021.
Aceito: 14/02/2022.
Publicado Online em: 18/11/2022.
DOI: http://dx.doi.org/10.15728/bbr.2023.20.1.6.pt

CC BY

THIS ARTICLE IS DISTRIBUTED UNDER THE TERMS OF THE CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE

1. INTRODUÇÃO

Na perspectiva tradicional de precificação de ativos, a oferta e a demanda de títulos nos mercados financeiros se igualam, definindo os preços de equilíbrio. Diferentemente, na literatura de microestrutura de mercado, argumenta-se que o processo de formação de preços é mais complexo, uma vez que os atores financeiros não têm o mesmo acesso às informações e não entram simultaneamente no mercado. Assim, pressupostos fundamentais dos modelos tradicionais de precificação, como ausência de custos de transação e informações simétricas, são flexibilizadas de maneira a entender melhor a dinâmica dos preços na perspectiva de microestrutura de mercado. O conteúdo informacional carregado pelos preços dos títulos é, por conseguinte, uma das suas vertentes de estudo.

Em vista do aumento da quantidade de ações negociadas nos mercados de alta frequência e ao simultâneo crescimento na disponibilização de dados *tick-by-tick* pelas plataformas de informações financeiras, a realização de pesquisas sobre a microestrutura dos mercados tem gradativamente se tornado mais viável. De forma a estudar os efeitos da microestrutura no processo de formação do preço, como o da assimetria informacional, são necessárias informações sobre as transações, entre elas o lado que iniciou a negociação.

Como descrevem Easley et al. (2012b), as transações em um mercado financeiro compreendem duas posições: (i) compradora; e (ii) vendedora. A posição que iniciou a negociação do ativo pode indicar assimetria de informações entre os participantes da transação. E a desproporção entre os volumes de negociação das posições de compra e venda pode indicar toxicidade no fluxo de ordens. Porém, determinar a posição que iniciou uma transação não é simples, sobretudo em mercados de alta frequência, cujas informações raramente estão disponíveis. Frente a esse problema, surgiram diversos algoritmos de classificação de transações, entre eles, o *Tick Rule* (TR), *Quote Rule* (QR), Lee-Ready (L-R) e *Bulk Volume Classification* (BVC), que permitem determinar a posição iniciadora da transação a partir de informações de bancos de dados tradicionais.

O objetivo deste artigo foi comparar o desempenho dos métodos TR e BVC na classificação de ordens de compra e venda de títulos no mercado acionário brasileiro. A escolha por testar a acurácia desses métodos deveu-se ao fato de utilizarem, respectivamente, dados *tick-by-tick* e dados comprimidos em intervalos de tempo ou volume. Considerando-se o tipo de mercado (na B3 as negociações são realizadas a partir de ordens enviadas por meio dos sistemas das corretoras – *order-driven market*), a assimetria informacional e a volatilidade dos retornos mais elevadas e o volume de negociação mais reduzido da B3 – característicos de mercados emergentes –, espera-se que informações privadas tendam a chegar ao mercado sequencialmente, em vez de em *bulks* de volume. Com isso, o risco tende a diminuir após uma sequência de ordens de investidores com mais informações (*informed*), alterando o preço negativamente. Nesse caso, algoritmos desenvolvidos para mercados de alta frequência (BVC) podem não ter a mesma acurácia que os tradicionais (TR) na classificação das ordens nesses mercados. Assim, é importante testar qual algoritmo de classificação distingue melhor a negociação informada em um determinado mercado e, assim, indicar o melhor como ferramenta de auxílio às decisões de investimento.

De maneira a verificar a acurácia do BVC, as ações foram divididas em três grupos, segundo seu volume transacionado. Essa abordagem está em linha com outras pesquisas empíricas (Easley et al., 2012b; Panayides et al., 2019) que mostram que, dependendo do volume de negociações, esse algoritmo apresenta performances diferentes. Além disso, foram empregadas as negociações de 2018 para estimar os parâmetros de maior acurácia desse algoritmo para cada grupo de ações. Para averiguar se sua performance se mantinha próxima à observada para o ano de 2018, esses parâmetros foram testados utilizando-se as negociações de 2019.

Em seguida, para analisar o impacto do método de classificação de transações de ações no Brasil, foi utilizado o VPIN de Easley et al. (2011), o qual mensura a probabilidade de negociações privilegiadas para uma dada ação. A escolha pelo VPIN deveu-se à necessidade das informações de volume de compra e venda para o seu cálculo. A partir dos dados reais, foi possível comparar os resultados de VPIN estimado pelo TR e pelo BVC e, assim, tirar conclusões acerca de suas eficiências.

Diferentemente de Easley et al. (2012b), mas em consonância com outras evidências empíricas sobre a performance do BVC (Chakrabarty et al., 2015; Omrane & Welch, 2016; Panayides et al., 2019), neste estudo o BVC apresentou desempenho significativamente inferior ao TR, produzindo estimativas para o VPIN fracamente correlacionadas com o VPIN calculado com os dados reais.

A dificuldade de classificar o lado que inicia as transações levou diversos autores a proporem soluções baseadas em dados disponíveis, sejam *tick-by-tick* ou comprimidos. Algoritmos *tick-by-tick* necessitam da menor granularidade possível – ou seja, transação a transação. Esse aspecto impõe dois desafios aos pesquisadores: (i) acesso aos dados; e (ii) capacidade computacional de tratamento dos dados. A maior parte dos distribuidores de dados os fornecem na forma comprimida, em intervalos de tempo (1 minuto, 5 minutos, 15 minutos; e assim por diante). Por um lado, abordagens que utilizam dados comprimidos são mais acessíveis para a maioria dos pesquisadores – para os dados do presente artigo, ao comprimi-los em intervalos de tempo de 5 minutos, o volume foi reduzido a 12% do seu tamanho original, tornando a sua manipulação mais factível e computacionalmente menos intensa. Por outro, o uso de dados comprimidos tem a desvantagem da perda das informações intrínsecas ao período analisado, uma vez que resume milhares de transações a uma única medida – média, mediana etc.

Dessa forma, este artigo contribui especialmente para aqueles pesquisadores que não possuem acesso a dados reais, na medida em que se torna necessário levar em consideração os vieses da utilização de dados agregados na classificação de transações, por exemplo o volume transacionado das ações e os parâmetros do algoritmo, fatores decisivos na classificação correta do volume de compra e venda dos ativos. Na seção seguinte, serão apresentados os algoritmos de classificação TR e BVC e suas respectivas aplicações empíricas. Posteriormente, serão apresentadas a base teórica e a forma de cálculo do VPIN.

2. REVISÃO DA LITERATURA

2.1. Algoritmos de classificação de transações

2.1.1. Tick Rule

O algoritmo TR utiliza o preço das transações para classificá-las. Quando o preço da operação corrente é maior (menor) do que o preço da operação precedente, ela é classificada como compra (venda). Quando o preço não se altera, repete-se a classificação dada à operação antecedente. Easley et al. (2012b) consideram esse método de classificação vulnerável (suscetível a erros), principalmente em mercados de alta frequência. Para o mercado norte-americano, esses autores reportam uma precisão de 86% do TR ao classificar as transações ocorridas entre os meses de novembro de 2010 e 2011. Ellis et al. (2000), por sua vez, documentam uma precisão de 81% do TR na classificação das transações na NASDAQ. Para o mercado australiano, a precisão do TR foi de 75% (Aikten & Frinos, 1996).

Por outro lado, Chakrabarty et al. (2015) mostram que o desempenho do TR no mercado norte-americano é decrescente ao longo do tempo. Omrane e Welch (2016) constatam que o TR classificou corretamente somente 67% das transações da amostra analisada, corroborando, assim, os resultados de Chakrabarty et al. (2015) de eficácia decrescente do TR – possivelmente por ser um mercado de alta frequência, para o qual as classificações de dados *tick-by-tick* são mais difíceis. Similarmente, Panayides et al. (2019) encontram evidências de redução da precisão do TR para duas amostras de ações, cujas variações foram: (i) de 79% a 92% entre 2007 e 2008; e (ii) de 39% a 65% em 2017.

2.1.2. Bulk Volume Classification

Devido aos problemas do TR para classificar corretamente transações em mercados com altas frequências de negociação, Easley et al. (2012b) desenvolveram o modelo BVC, que se propõe a reduzir os efeitos do desdobramento de ordens de compra e venda (*order splitting*). O BVC utiliza a variação padronizada do preço para classificar os volumes de negociação probabilisticamente. A sua mecânica consiste em agrupar as transações por intervalos de tempo ou de volume, os quais são determinados arbitrariamente ou limitados conforme a estrutura do banco de dados. Após esse agrupamento, as proporções do volume de transações que se iniciaram pelo lado comprador e pelo vendedor são obtidas por meio das equações (1) e (2):

$$V_\tau^B = V_\tau \cdot Z\left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\Delta P}}\right) \tag{1}$$

$$V_\tau^S = V_\tau \cdot \left[1 - Z\left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\Delta P}}\right)\right] = V_\tau - V_\tau^B \tag{2}$$

Em que: V_τ é o volume total do intervalo τ; V_τ^B e V_τ^S são, respectivamente, os volumes de compra e venda para o intervalo τ; Z é a função de distribuição normal padrão acumulada; e $\sigma_{\Delta P}$ é o desvio-padrão estimado da variação de preços entre os intervalos.

Quando não há variações de preço entre o início e o fim do intervalo, o BVC divide o seu volume igualmente em compra e venda (Easley et al., 2012a). Quando o preço aumenta (diminui), maior proporção do volume é classificada como iniciada pelo lado comprador (vendedor) da operação. A proporção de volume classificada como compra ou venda cresce à medida que a variação de preço aumenta.

No método BVC, considera-se que a variação de preços seja independente e identicamente distribuída (i.i.d), com média zero e variância constante ($\sigma_{\Delta P}$). Easley et al. (2012b) consideram que o ideal seria obter a verdadeira função de distribuição acumulada da variação de preços. Todavia, entre outros problemas, esse procedimento diminuiria a precisão na classificação das transações e impossibilitaria a generalização do BVC. Dessa forma, supor a normalidade das variações de preço dos títulos é o mais adequado na estimação dos volumes de compra e venda. Segundo esses autores, o bom desempenho do BVC em classificar transações de ações com elevado volume de negociação se deve ao procedimento de aproximação normal. Para ativos pouco líquidos, consideram que os métodos *tick-by-tick* sejam os mais adequados. Os resultados revelam que o BVC classifica corretamente (i) 86,61% das transações dos contratos futuros do e-mini S&P500, quando utilizados intervalos de 1 minuto, e (ii) 87,35%, quando utilizados intervalos de 5 minutos; enquanto a precisão do TR se mantém em 86,43% em ambos esses intervalos.

Chakrabarty et al. (2015) comparam as eficácias dos algoritmos L-R, TR e BVC na classificação das transações nos mercados à vista. Analisando dados reais de 1.471 ações. os autores encontram

que as precisões do TR e do L-R são superiores à do BVC. O TR e o L-R classificam de forma correta, respectivamente, 90,8% e 92,6% das transações, ao passo que a melhor precisão do BVC é de aproximadamente 80%, quando se utilizam intervalos de 1 minuto. Omrane e Welch (2016) corroboram esses resultados: tanto o TR quanto o QR apresentam resultados melhores do que o BVC para o mercado de moedas estrangeiras.

Após as críticas ao BVC, Easley et al. (2016) argumentam que, nas pesquisas sobre assimetria informacional, o mais adequado seria obter a informação subjacente às transações, que depende de *proxies* por não ser observável. Para comparar a precisão do TR e BVC, eles se utilizam de três *proxies* para a informação subjacente às transações e encontram que o BVC tem um melhor desempenho para duas delas. Concluem, assim, que o BVC permite distinguir as intenções inerentes às transações.

Panayides et al. (2019) corroboram essas evidências empíricas. Eles argumentam que, quando o BVC é calibrado para determinado mercado – com o intervalo de tempo ou volume correto – ele é capaz de transmitir com maior precisão o conteúdo informacional presente nas transações analisadas.

2.2. Cálculo do VPIN

Easley et al. (1996) propuseram a *Probability of Informed Trading* (PIN) como uma forma de quantificar a probabilidade de ocorrência de transações iniciadas por agentes com informações privilegiadas. O método se baseia na disparidade da quantidade de compras e vendas ocorridas para uma ação em dias independentes.

Para resolver o problema de não convergência da função de máxima verossimilhança da *Probability of Informed Trading* (PIN) para dias em que o número de ordens é alto, Easley et al. (2012a) desenvolveram o *Volume-Synchronized Probability of Informed Trading* (VPIN), o qual permite quantificar diretamente o nível de toxicidade das ordens sem a necessidade de estimação de parâmetros por máxima verossimilhança.

A ideia do VPIN consiste na separação das informações de volume, em um dia, em conjuntos iguais (*volume buckets*), tratando cada um equivalente a um período de chegada de informação. O desequilíbrio de transações é estimado pela média sobre n conjuntos de volume *(volume buckets)*. Assim, o VPIN é obtido pela equação (3)

$$VPIN = \frac{\sum_{\tau=1}^{n} V_{\tau}^{S} - V_{\tau}^{B} \vee}{nV} \tag{3}$$

Em que: V_{τ}^{B} e V_{τ}^{S} representam, respectivamente, os volumes de compra e venda em um conjunto de volume t (*buckets*). Seguindo o cálculo padrão na literatura, utilizou-se um n igual a 50, indicando que as transações são agrupadas em 50 conjuntos iguais em volume transacional por dia. Com base nessa agregação, o VPIN é estimado diretamente através da equação (3), ou seja: (i) calcula-se o desbalanceamento absoluto entre as ordens de compra e venda; e (ii) divide-se esse valor pela quantidade de transações observadas para cada conjunto de volumes transacionais.

Por representar uma extensão da PIN, *proxy* bastante testada, o VPIN também vem sendo explorado nos estudos em microestrutura, tendo obtido sucesso na previsão de eventos relevantes como o *Flash Crash* (Wu et al., 2013). Abad e Yagüe (2012) destacam que a toxicidade do fluxo de ordens mensurada pelo VPIN está diretamente associada ao problema de seleção adversa ao qual os *market makers* estão suscetíveis. O desbalanceamento das ordens de compra e venda, em um período curto, está relacionado às informações subjacentes ao VPIN.

3. METODOLOGIA

3.1. População e amostra

A população estudada nesta pesquisa foi formada pelas ações negociadas na B3 entre 02 de janeiro de 2018 e 28 de junho de 2019. Os dados de 2018 foram utilizados para avaliar o conjunto de parâmetros que produz a melhor performance do BVC e os de 2019, para validar esses parâmetros e comparar as performances do BVC com as do TR. No total, foram analisados os ativos negociados todos os dias no período observado: 181 ações. Foi definido o critério de selecionar apenas os ativos com negociação em todos os dias para que a construção dos intervalos de tempo ou volume do BVC não fossem afetados por fatores externos relativos a períodos distantes entre as transações.

A próxima segmentação refere-se ao volume de ações negociadas para cada ativo. Como esse é o insumo utilizado pelos algoritmos, cada ativo foi alocado a uma classe relacionada ao seu volume médio transacionado em 2018. Diferentemente de Panayides et al. (2019), que segmentaram os ativos em três classes com quantidade semelhantes em cada, no presente artigo, optou-se por utilizar o algoritmo de *fisher-jenks* para separar os ativos em três classes: pequeno, médio e grande volume. Esse algoritmo foi escolhido porque ele permite definir os pontos de corte e isolar os ativos dentro de suas respectivas classes, diminuindo a variância entre os ativos de mesma classe, e aumentar a variância em relação aos ativos das demais classes. A quantidade de ativos e o volume médio negociado em 2018 para cada classe são exibidos na Tabela 1.

A média do volume transacionado para as ações pequenas se aproxima do reportado por Panayides et al. (2019) no mercado europeu, porém as de médio e grande porte são 44% e 30% menores do que o reportado por esses autores, indicando assim uma primeira diferença entre a atividade do mercado acionário brasileiro e a de países mais desenvolvidos.

Tabela 1
Número de ativos e volume médio por classe

Classe	Volume médio	Número de ativos
Pequeno	287.688	99
Médio	1.287.740	39
Grande	7.109.882	43

Fonte: dados da pesquisa.

3.2. Coleta de dados

Uma das limitações principais ao se aplicar a PIN e o VPIN é a possibilidade de classificação incorreta das ordens de compra e venda. De forma a contribuir com a análise do desempenho dos algoritmos de classificação, esta pesquisa utilizou dados reais transacionados no mercado brasileiro como base de comparação com os resultados gerados pelos algoritmos TR e BVC. Os dados foram coletados no diretório *market data* da B3, que contém as informações sobre as ordens emitidas de todas as ações transacionadas na B3 nos últimos dois anos, além de horário, preço, quantidade e lado que iniciou a transação.

O volume total de dados utilizados neste artigo foi de cerca de 150 milhões de linhas, em que cada linha representa uma ordem de compra ou venda executada no período referido, com média de 2,6 milhões de ações transacionadas por dia. Em comparação, a amostra utilizada por Panayides et al. (2019) contém uma média de 4 milhões de ações transacionadas diariamente.

Finalmente, para a aplicação do BVC, foram feitas diversas agregações, chegando a reduzir o volume em cerca de 88% quando utilizado um intervalo de 5 minutos, demonstrando uma vantagem em se utilizar dados agregados.

4. ANÁLISE DOS RESULTADOS

Nesta seção, discutem-se resultados para a taxa de precisão dos algoritmos TR e BVC. Comparam-se os valores do $VPIN_{REAL}$ – calculado com a quantidade real de compra e venda – e os de $VPIN_{TR}$ e $VPIN_{BVC}$ – calculados, respectivamente, com os volumes estimados pelo TR e pelo BVC. Finalmente, analisam-se as propriedades dos métodos TR e BVC, de maneira a evidenciar em que momentos esses algoritmos classificam erroneamente as transações.

4.1. Calibração dos parâmetros do BVC

É importante calibrar os parâmetros do BVC – para, subsequentemente, comparar a performance dos algoritmos TR e BVC. Seguindo Panayides et al. (2019), e considerando a ponderação de Easley et al. (2012b) de que as performances do BVC são diferentes para títulos de diferentes volumes de transações, foram testados diversos parâmetros no ano de 2018 para definir o melhor conjunto para cada classe de ativos. Além disso, os parâmetros foram testados com os dados de 2019, de forma a verificar se a performance anterior se mantinha e, assim, atestasse a possibilidade de aplicação do BVC em dados futuros.

A fim de selecionar o melhor conjunto de parâmetros para cada ativo, a acurácia do BVC foi calculada mediante a equação (4).

$$AR = 1 - \frac{\left|\frac{V_B - \hat{V}_B}{max\,(V_B, \hat{V}_B)}\right| + \left|\frac{V_S - \hat{V}_S}{max\,(V_S, \hat{V}_S)}\right|}{2} \qquad (4)$$

Em que: V_B e V_S representam os volumes reais de compra e venda; $\hat{V}_B$ e $\hat{V}_S$ e representam, respectivamente, os volumes de compra e venda estimados pelo BVC. Para cada ativo, foi selecionada a maior acurácia. A Tabela 2 apresenta os percentuais de representatividade de cada parâmetro dentro das três classes analisadas.

O BVC apresentou, para as três classes, uma maior acurácia quando utilizado intervalo de 5 minutos. É interessante destacar que, em linha com estudos anteriores (Easley et al., 2012b), os ativos de menor volume transacionado apresentaram menor consistência em termos de parâmetros gerais, já que os ativos ficaram quase que uniformemente divididos entre os intervalos de tempo de 5 minutos e os intervalos de volume de 75, 100, 200 e 500 mil ações.

Essa primeira evidência levanta uma incerteza em relação à aplicabilidade do BVC como algoritmo de previsão, tendo em vista que, entre os ativos com menor volume transacionado, os parâmetros não apresentam consistência – foram testadas outras formas de clusterização dos ativos e, entre os 80 menores ativos, os parâmetros não se estabilizaram, e isso sugere que esse fenômeno persiste mesmo quando a forma de separação das ações utilizada é o volume transacionado (prática comum na literatura).

Outra característica importante em algoritmos de previsão é a aplicabilidade dos parâmetros em períodos de tempos distintos. Os percentuais de ações em que o parâmetro de maior acurácia se manteve nos anos de 2018 e 2019 diferiu entre os grupos analisados: (i) entre as ações de volume alto (grande), foi de 78%; (ii) entre as ações de volume intermediário (médio), 74%; (iii) entre as ações de volume baixo (pequeno), 35%.

Novamente, verifica-se maior inconsistência em ativos de menor volume, o que alerta para um possível cuidado ao se utilizar o BVC nessa classe de ativos. Tendo estabelecido o melhor conjunto de parâmetros para o BVC – intervalo de 5 minutos –, os resultados subsequentes utilizarão esses valores para a estimativa do volume de compra e venda.

Tabela 2
Percentual de representatividade dos parâmetros utilizados na calibração do BVC

		Classe de ativos		
Agrupamento	Parâmetro	Pequeno	Médio	Grande
Tempo	1 minuto	0,00%	0,00%	0,00%
Tempo	2 minutos	0,00%	0,00%	0,00%
Tempo	3 minutos	1,01%	0,00%	11,63%
Tempo	**5 minutos**	**20,20%**	**66,67%**	**60,47%**
Volume	1.000	0,00%	0,00%	0,00%
Volume	5.000	0,00%	0,00%	0,00%
Volume	10.000	6,06%	0,00%	0,00%
Volume	25.000	3,03%	0,00%	0,00%
Volume	50.000	9,09%	2,56%	0,00%
Volume	75.000	13,13%	2,56%	0,00%
Volume	100.000	12,12%	0,00%	0,00%
Volume	200.000	17,17%	12,82%	4,65%
Volume	500.000	18,18%	15,38%	23,26%

Fonte: dados da pesquisa.

4.2. Desempenho dos algoritmos TR e BVC

A Tabela 3 apresenta os resultados para as taxas de precisão do TR e do BVC. Verifica-se, para ambos os métodos, uma melhora de performance à medida que os ativos crescem em volume transacionado. O desempenho geral do TR foi de 80,82%, valor superior aos relatados por Omrane e Welch (2016) e Chakrabarty et al. (2015).

Tabela 3
Taxas de precisão para o TR e o BVC

	TR			BVC		
Classe	Mínimo	Média	Máximo	Mínimo	Média	Máximo
Pequeno	62,99%	77,71%	92,63%	33,08%	51,60%	67,28%
Médio	74,95%	82,95%	91,48%	37,19%	62,18%	68,69%
Grande	75,10%	86,05%	95,60%	40,89%	64,10%	70,90%
Geral	62,99%	80,82%	95,60%	33,08%	56,85%	70,90%

Fonte: dados da pesquisa.

O desempenho médio do BVC foi de 56,85%, valor inferior ao apresentado por Easley et al. (2012b), os quais analisaram os três contratos de futuros mais ativos do mercado norte-americano; e Omrane e Welch (2016), que analisaram contratos de câmbio.

No geral, percebe-se que o TR apresentou desempenho superior ao BVC. Enquanto o desempenho mais baixo do TR atingiu 62,99% de acurácia, o do BVC foi de 33,08%. Além disso, para todas as classes de ativos, o TR atingiu um desempenho superior a 90%, enquanto o BVC chegou a um máximo de 70,90% entre os ativos de maior volume transacionado.

A Figura 1 mostra que o desempenho do TR se concentra na faixa de 80% sendo que as acurácias entre os ativos apresentam comportamento simétrico em relação à mediana. O BVC apresenta valores os quais giram em torno de 63%, com o primeiro quartil em torno de 45%, e o terceiro quartil em torno de 66%, pouco acima da mediana de 64%. Novamente, essa assimetria é carregada pelos ativos de menor volume transacionado, que apresentam, em geral, desempenho inferior. Esse resultado está em consonância com as evidências de Easley et al. (2012b) de que o BVC possui melhor performance para ações de maior volume transacionado.

Os resultados preliminares indicam que o algoritmo TR apresenta performance superior à do BVC. A próxima seção mostra o resultado da aplicação prática dos dois métodos em um modelo que necessita, como insumos principais, das informações de volumes de compra e venda.

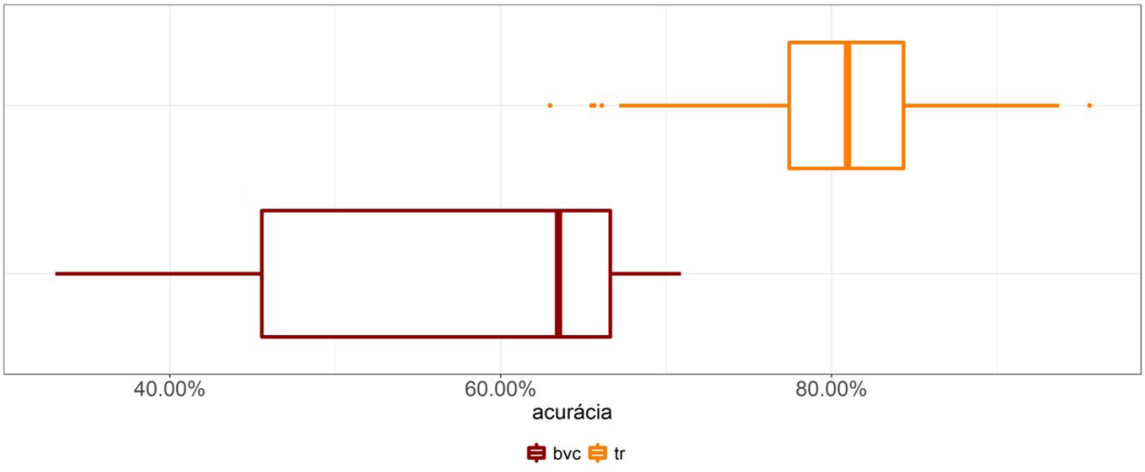

Figura 1. Amplitude das acurácias dos métodos TR e BVC
Fonte: dados da pesquisa.

4.3. Cálculo do VPIN a partir de dados reais, TR e BVC

De maneira a analisar os problemas relacionados à classificação das transações quando aplicado um método que necessita da quantidade de compras e vendas, calculou-se o VPIN utilizando os dados reais das transações realizadas entre 02 de janeiro e 28 de junho de 2019, além dos volumes apontados pelo TR e pelo BVC. Grammig e Theissen (2002) e Hwang et al. (2013) atentam para os problemas relacionados à má classificação das ordens ao se estimar *proxies* de risco informacional.

A Figura 2 mostra a média do VPIN de cada conjunto. De início, verifica-se a disparidade dos VPINs entre as maiores e menores ações. Esse resultado, reportado por diversos autores (Easley et al., 1996; Mohanram & Rajgopal, 2009; Abad & Yagüe, 2012; Wei et al., 2013), indica que há uma correlação negativa entre o VPIN e o valor de mercado da empresa.

A partir desses resultados é possível observar que o $VPIN_{REAL}$ e o $VPIN_{TR}$ não se diferem significativamente, em especial para as ações médias e grandes. Essa evidência é reforçada pelos resultados da Tabela 4 em que a diferença entre o TR e os dados reais flutuou em torno de 2% a 3%.

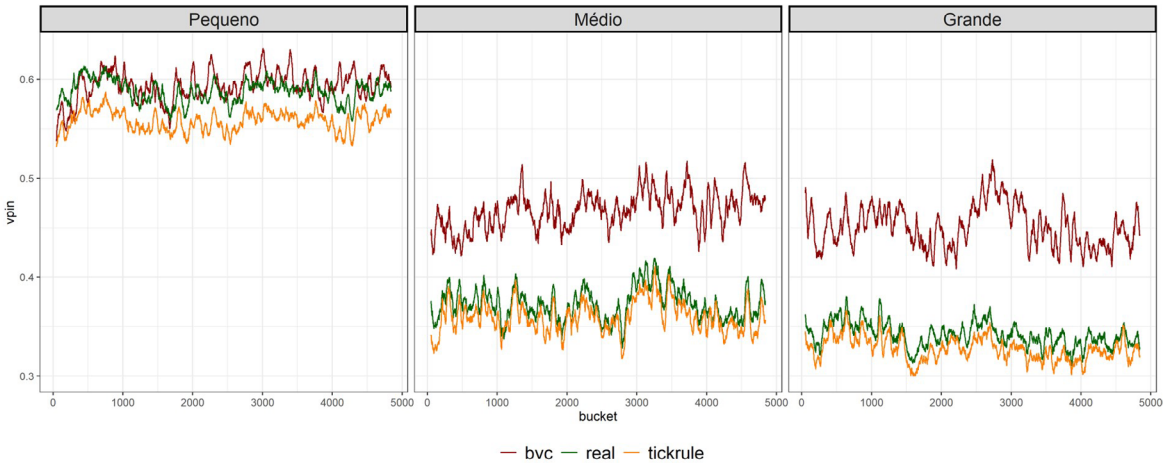

Figura 2. VPINs calculados com base em dados reais, TR e BVC
Fonte: dados da pesquisa.

Um ponto de destaque reside no $VPIN_{BVC}$ calculado para as ações pequenas. Apesar de ter sido evidenciada uma acurácia menor do BVC para essa classe de ação, o seu VPIN foi o mais próximo do real entre as três classes de ações. Esse resultado se deve à particularidade da metodologia do VPIN, em que o desbalanceamento de ordens é levado em consideração. Caso as ordens de compra e venda estejam sendo estimadas de forma incorreta, porém seu desbalanceamento esteja próximo ao real, o VPIN será próximo daquele calculado com os dados reais. Isso pode indicar que mais um cuidado deva ser tomado ao utilizar o BVC, uma vez que ele pode indicar resultados promissores advindos de dados incorretos, podendo torná-lo incapaz de ser aplicado em outros métodos que necessitem dos dados de compra e venda como insumo.

Tabela 4
VPINs médios por classe e por método

Classe	$VPIN_{REAL}$	$VPIN_{BVC}$	$VPIN_{TR}$
Pequeno	58,79%	59,36%	55,82%
Médio	37,32%	46,76%	35,98%
Grande	34,17%	45,11%	32,74%

Fonte: dados da pesquisa.

Objetivando analisar as características dos VPINs estimados, procedeu-se ao cálculo, para cada classe de ações, da correlação entre o $VPIN_{REAL}$ e o $VPIN_{TR}$; e entre o $VPIN_{REAL}$ e o $VPIN_{BVC}$. Calcularam-se também a menor correlação, a média e a maior correlação entre cada classe de ações, bem como a média (Tabela 5).

Percebe-se que, para todas as classes de ações, o $VPIN_{TR}$ apresentou uma forte correlação com o $VPIN_{REAL}$ com média próxima a 80%. Esses números reforçam a acurácia verificada para o TR. Em contraste, ao se analisar a correlação do $VPIN_{BVC}$ com o $VPIN_{REAL}$, verifica-se uma média mais baixa de, no máximo, 50% para as ações da classe média.

As correlações máximas atingidas pelo BVC se aproximam da correlação média do TR. Para as ações pequenas, houve inclusive casos de correlação negativa, indicando que o desbalanceamento informado pelo BVC apresentava sinal contrário ao dos dados reais. Isso significa que, enquanto o $VPIN_{REAL}$ indica momentos de aumento do risco informacional (alertando para o desbalanceamento

das ordens), o $VPIN_{BVC}$ pode indicar o oposto, contrariando, dessa forma, o objetivo do VPIN – que, segundo Easley et al. (2012a), é de alertar os investidores para momentos de desequilíbrio de volume e, assim, evitar eventos de iliquidez que resultem em *crashes* das bolsas, como o *Flash Crash*.

Tabela 5
Correlação entre $VPIN_{REAL}$-$VPIN_{TR}$ e entre $VPIN_{REAL}$-$VPIN_{BVC}$

	TR			BVC		
Classe	Mínimo	Média	Máximo	Mínimo	Média	Máximo
Pequeno	0,3551	0,7817	0,9732	-0,1769	0,2872	0,7630
Médio	0,5733	0,8636	0,9855	0,1018	0,5018	0,8563
Grande	0,6889	0,8544	0,9713	0,2231	0,4508	0,8632

Todas as correlações apresentaram p.valores iguais a zero.
Fonte: dados da pesquisa.

Como forma de mostrar a consequência da utilização do BVC para se classificarem as ordens de transações, as Figuras 3 e 4 mostram o comportamento dos VPINs para as ações que apresentaram, respectivamente, maior e menor aderência em relação aos dados reais para cada classe.

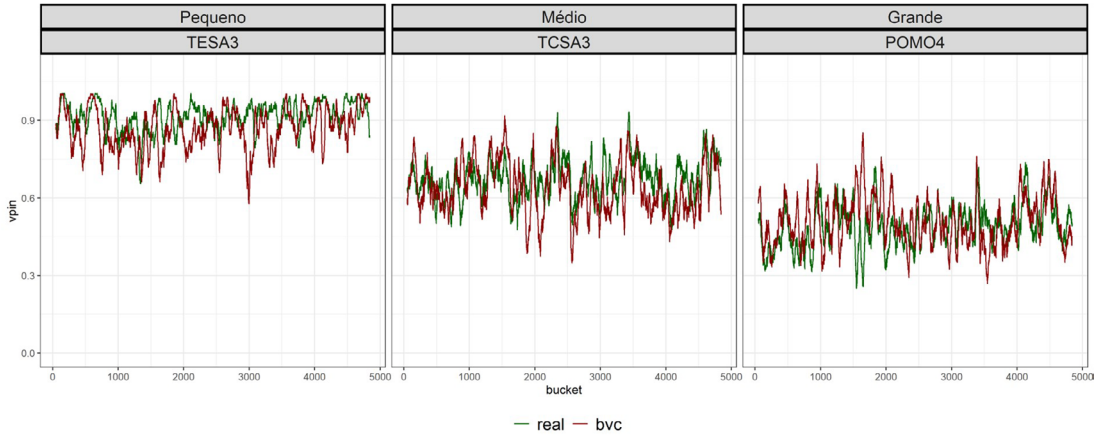

Figura 3. Ações com maior aderência entre o BVC e os dados reais
Fonte: dados da pesquisa.

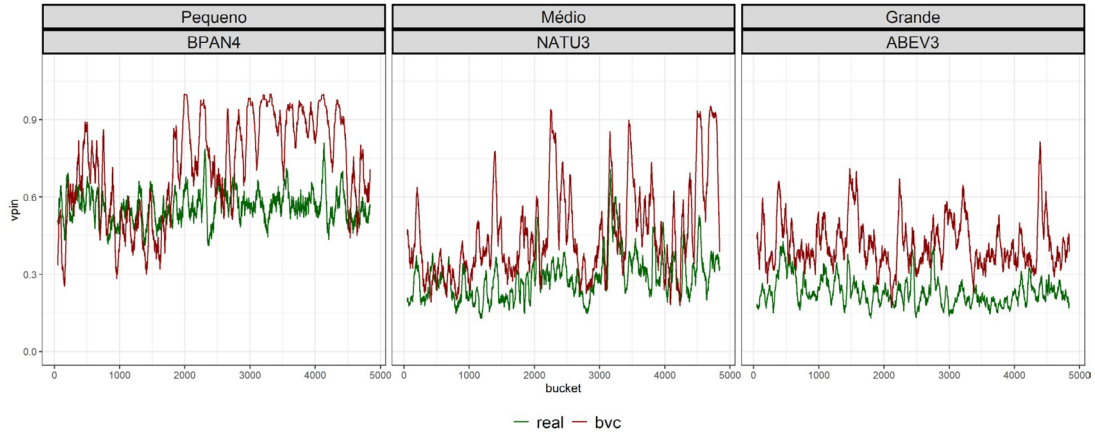

Figura 4. Ações com menor aderência entre o BVC e os dados reais
Fonte: dados da pesquisa.

As figuras mostram que o $VPIN_{BVC}$ apresenta valores mais extremos do que o $VPIN_{REAL}$ mesmo para as ações com maior aderência. Para as ações de menor aderência, o $VPIN_{BVC}$ apresentou em alguns momentos valores próximos a 90% de $VPIN_{REAL}$. Caso fosse utilizado como um indicador de problemas de liquidez, tal método apresentaria diversos falsos positivos quando comparado ao valor real, o que poderia causar problemas em uma utilização prática. Esse comportamento mais desbalanceado do BVC será analisado na seção dedicada à análise de suas propriedades desse método.

Os resultados apresentados nesta seção indicam que o BVC não se apresenta como um algoritmo de classificação de transações eficaz frente aos dados reais. Essa evidência é corroborada pela aplicação do VPIN, que mostra que os valores estimados pelo BVC destoam substancialmente daqueles obtidos a partir dos dados reais. Em contrapartida, o $VPIN_{TR}$ se aproxima do $VPIN_{REAL}$ para todas as ações analisadas nesta pesquisa. Nas próximas seções, analisam-se os momentos em que o TR e o BVC classificam erroneamente as ordens, e isso explicaria as diferenças nas estimações do VPIN.

4.4. Análise das propriedades do *Tick Rule*

O algoritmo TR está calcado no princípio econômico de que uma ordem de compra (venda) aumenta (diminui) a demanda pela ação, o que leva a um acréscimo (decréscimo) no preço. Para verificar em quais situações esse princípio econômico se mantém, foi analisada a frequência dos sinais das ordens dada às mudanças de preço das transações. Ou seja, verificou-se, para cada valor de ΔP_t, qual a quantidade de ordens de compra (para $\Delta P_t>0$) e venda (para $\Delta P_t<0$) em relação ao total. Por fim, foi calculada a quantidade de vezes nas quais a ordem se repetiu para momentos em que $\Delta P_t=0$. Assim, as equações (5), (6) e (7) indicam os cálculos realizados.

$$P(B_t|\Delta P_t = P^+) \tag{5}$$

$$P(S_t|\Delta P_t = P_) \tag{6}$$

$$P(X_t = X_{t-1}|\Delta P_t = 0) \tag{7}$$

Em que: B_t e S_t representam, respectivamente, uma ordem de compra e venda no momento t; e P^+ e $P_$ representam, respectivamente, valores positivos e negativos para as variações de preço entre as transações. Por fim, X_t representa o sinal da ordem lançada no momento t, podendo ser uma ordem de compra (B) ou venda (S). A equação (7) representa o caso em que a mudança de preço é igual a zero e deseja-se verificar qual a frequência de que o lado iniciador da ordem em t seja igual ao lado iniciador da ordem anterior.

Os resultados das equações (5) e (6) estão expostos na Tabela 6. Verifica-se, tanto para as mudanças de preço positivas quanto negativas, uma consistência na frequência do lado da transação. Mesmo em mudanças de preço mais fortes, acima de 0,20 unidade monetária, o percentual de ordens de compra ou venda continuam no mesmo patamar – cerca de 88%.

Os resultados da tabela 6 mostram por que o TR apresenta bom desempenho para a classificação das transações. De forma geral, a base na qual está calcado esse algoritmo se sustenta para a amostra analisada, ou seja, mudanças positivas de preço apontam para ordens de compra, enquanto mudanças negativas indicam ordens de venda.

Para o resultado da equação (7) relacionado às mudanças de preço iguais a zero, verificou-se que $P(X_t = X_{t-1}|\Delta P_t = 0) = 0{,}9531$. Ou seja, para a amostra analisada, em 95,31% dos casos,

quando não houve mudança de preço, a transação no momento *t* foi a mesma do momento *t*-1, como preconizado pelo TR.

Com o objetivo de aprofundar a análise das situações em que o TR classifica incorretamente as transações, foi verificado que cinco variáveis influenciam no desempenho desse método, e são elas: a mudança de preço (ΔP_t); o sinal da ordem que está sendo classificada; o sinal da ordem anterior; a diferença no tempo entre as duas transações; e se as corretoras de compra e venda são as mesmas da transação anterior.

Tabela 6

Frequência dos sinais das ordens dadas as mudanças de preço entre as transações

ΔP_t	$P(B_t \mid \Delta P_t = P^+)$	ΔP_t	$P(S_t \mid \Delta P_t = P^-)$
0,01	88,43%	-0,01	88,74%
0,02	89,15%	-0,02	89,77%
0,03	88,35%	-0,03	89,22%
0,04	87,83%	-0,04	88,85%
0,05	86,94%	-0,05	88,17%
0,06	87,27%	-0,06	88,48%
0,07	87,39%	-0,07	88,60%
0,08	87,40%	-0,08	88,31%
0,09	87,04%	-0,09	87,77%
0,10	85,21%	-0,10	86,79%
0,11	86,31%	-0,11	88,09%
0,12	87,10%	-0,12	87,75%
0,13	87,93%	-0,13	88,76%
0,14	87,96%	-0,14	88,52%
0,15	87,49%	-0,15	87,61%
0,16	87,27%	-0,16	87,86%
0,17	89,05%	-0,17	88,45%
0,18	87,84%	-0,18	89,02%
0,19	87,76%	-0,19	88,61%
0,20	86,23%	-0,20	87,14%
> 0,20	88,11%	< -0,20	89,27%

Fonte: dados da pesquisa.

A Tabela 7 apresenta as situações e as frequências nas quais o TR inicia uma sequência de transações classificadas erroneamente. A maior parte dos erros do TR advém de situações em que a mudança de preço é positiva, porém a ordem sendo classificada é uma venda precedida de outra venda. Nesse caso, a corretora que está envolvida nas vendas em *t* e *t*-1 é a mesma, enquanto a compradora é diferente. Portanto, a situação descrita é aquela na qual: (i) uma corretora lança uma ordem de venda em *t*-1; quando essa ordem é executada, (ii) outra ordem de venda é lançada pela mesma corretora com uma diferença de 0 segundo, sendo executada por um comprador diferente daquele que enviou a ordem. Nesse caso, a segunda ordem (em *t*) apresenta um preço superior em relação ao da transação em *t*-1, preço esse provavelmente decorrente da rápida execução da ordem de venda. Isso indica haver liquidez para a ação naquele momento, e que a sua demanda está alta, o que justifica o acréscimo no preço da venda.

A mesma análise pode ser feita para a situação na qual as ordens atual e anterior são de compra, porém houve uma variação de preço negativa entre as transações (linha 2 da tabela 7). Nesse caso, a corretora que está enviando as ordens de compra é a mesma para as duas transações, porém não é a vendedora. A segunda ordem de compra é executada com rapidez em relação à primeira (0 segundo de diferença), indicando a existência de muitos negociadores interessados na venda da ação (a oferta é alta), provocando uma redução do preço da transação, que é executada mesmo tendo sido a um preço inferior ao da transação anterior.

Os cenários descritos acima foram aqueles que apresentaram maiores frequências de transações classificadas incorretamente pelo TR. Quando uma dessas situações acontece, uma sequência de classificações erradas pode se suceder se não houver mais mudanças de preços. Isso porque, nesse caso, o TR continua a classificar a transação a partir do sinal da ordem anterior, que já estava classificada de forma incorreta.

Tabela 7
Situações em que o TR inicia uma sequência de transações classificadas incorretamente

Diferença de tempo	ΔP_t	Ordem atual	Ordem anterior	Corretora de compra	Corretora de venda	Frequência
0	+	S	S	≠	=	10,97%
0	-	B	B	=	≠	10,70%
0	0	B	S	≠	=	7,82%
0	0	S	B	=	≠	7,80%
+	+	S	S	≠	≠	4,67%
+	-	B	B	≠	≠	4,61%
0	0	B	S	≠	≠	4,32%
0	0	S	B	≠	≠	4,20%
+	0	B	S	≠	≠	4,16%
+	0	S	B	≠	≠	4,08%
0	+	S	S	≠	≠	3,15%
0	0	S	B	=	=	2,94%
0	-	B	B	≠	≠	2,87%
0	0	B	S	=	=	2,75%
+	0	B	S	≠	=	2,40%
+	0	S	B	=	≠	2,25%
0	-	B	B	=	=	2,07%
0	+	S	S	=	=	2,06%
+	+	S	S	≠	=	1,84%
+	-	B	B	=	≠	1,67%
0	0	S	B	≠	=	1,45%
0	0	B	S	=	≠	1,37%

Nota: '+', '-' e '0' significam respectivamente uma variação de preço positiva e negativa (de qualquer magnitude) e nenhuma variação. Os valores das colunas 'Corretora de compra' e 'Corretora de venda', '=' e '≠', representam respectivamente se a corretora é a mesma ou não da transação anterior.
Fonte: dados da pesquisa.

As duas situações seguintes onde houve erros do TR são aquelas em que as ordens em *t* são compras (vendas) e em *t*-1 são vendas (compras), porém com uma variação de preço igual e tempo entre as transações iguais a zero, e a mesma corretora de venda (compra). Como nesse caso o TR repete a classificação, é iniciada uma sequência de erros. Observou-se que o tempo entre as transações e as corretoras nelas envolvidas têm um papel importante na definição do lado que iniciou a compra ou a venda. Como as transações são praticamente instantâneas, dois fenômenos podem influenciar na classificação do TR. O primeiro deve-se à ação de *order splitting*, isto é, uma mesma ordem é dividida em diversas ordens menores a fim de o mercado não perceber que existe um *trader* movimentando um alto volume de ações. Esse tipo de estratégia pode ser detectado ao se analisar o tempo entre as transações, a corretora e o volume transacionado, já que muitas vezes a ordem é dividida em parcelas de volumes iguais. O segundo fenômeno se deve-se ao fato de que, como o tempo entre as transações é praticamente nulo, por ser o tempo entre as transações praticamente nulo, o mercado não se ajusta a tempo para que a variação do preço reflita a oferta e a demanda da ação.

De forma geral, percebe-se por meio dos resultados da tabela 7 que as situações apresentam frequência similar para os sinais das ordens, indicando uma simetria nesse sentido, e mostram que o mercado se comporta de forma similar independente se a transação em questão for iniciada por um comprador ou um vendedor. Por fim, tal análise indica uma oportunidade de construção de um modelo mais complexo que possa capturar a relação entre as variáveis, reduzindo a iniciação da sequência de erros.

Tendo identificado os momentos em que o TR classifica incorretamente os sinais das transações, procede-se na próxima seção à análise das propriedades do BVC.

4.5. Análise das propriedades do *Bulk Volume Classification*

Ao utilizar a distribuição normal para computar os percentuais de compra e venda de cada intervalo de tempo, é possível compará-los com os de compra e venda reais dentro do mesmo intervalo. A Figura 5 mostra como o percentual de compra evolui com a variação do preço em comparação com o assinalado pelo BVC – não foram verificadas grandes divergências entre as classes de ativos, portanto os valores reportados na figura 5 representam toda a amostra analisada (total).

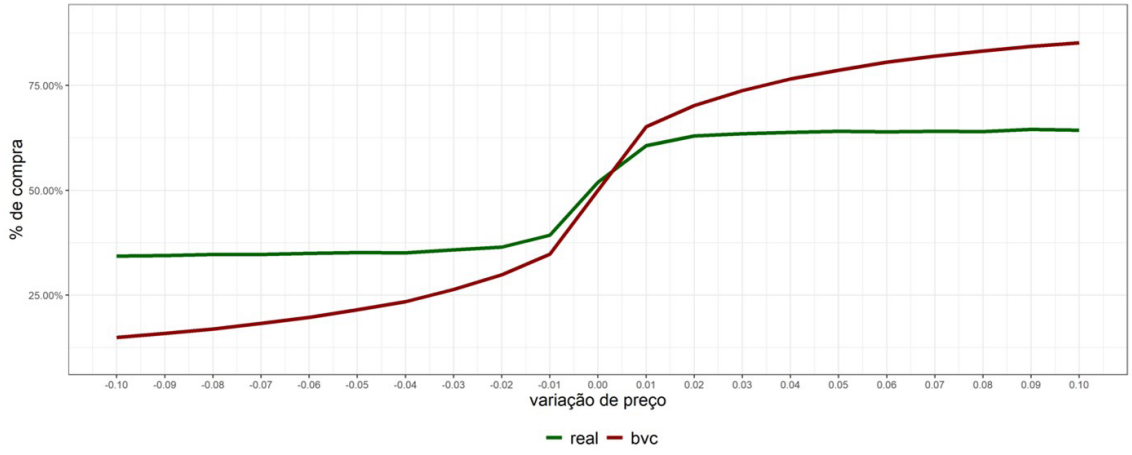

Figura 5. Percentual de compra real e estimado pelo BVC em relação à variação de preço
Fonte: dados da pesquisa.

Tem-se evidência de que uma das principais características do BVC é corroborada: quando não há variação de preço, o percentual de compra e venda dentro do mesmo intervalo fica próxima de 50% (51,88% de volume de compra na amostra analisada). Isso faz com que, na prática, o BVC tenha desempenho satisfatório para intervalos que não apresentem variação de preço (cerca de 22% dos intervalos).

No entanto, à medida que a variação de preço se distancia de zero, o percentual assinalado pelo BVC aumenta de forma mais acelerada do que o verificado na prática. Essa característica decorre da distribuição definida na construção do modelo. Através dos dados reais, verifica-se que, na média, o percentual de compra se estabiliza próximo de uma variação de preço absoluta de cerca de 0,05 unidade monetária. Dada a distribuição escolhida na aplicação do BVC, essa estabilização não ocorre dentro dos primeiros 0,1 unidade monetária de variação absoluta.

Esse comportamento explica por que o $VPIN_{BVC}$ apresenta picos mais frequentes do que o $VPIN_{REAL}$ ou o $VPIN_{TR}$. Como o BVC atribui um percentual maior de compra ou venda inclusive para variações baixas de preço, é natural que o desbalanceamento de volume apresentado por ele seja maior, levando a picos não detectados ao utilizar os dados reais no cálculo da VPIN.

Além disso, como observado na análise das propriedades do TR, quando há variação de preço, em cerca de 88% dos casos, a transação se dá na direção da variação – ou seja, aumento de preço indica compra, e diminuição de preço indica venda. Isso se sustenta quando a análise é realizada transação a transação, o que leva o TR a ter um desempenho de cerca de 80% de acurácia na amostra analisada. Em contrapartida, o BVC agrupa transações em intervalos e utiliza o último preço como indicador de demanda ou de oferta. Isso implica, portanto, que todo o conteúdo informacional presente dentro do intervalo – capturado pelo TR – é descartado na utilização do BVC. Isso explica também por que intervalos calculados com um maior período ou com uma maior quantidade de transações agregadas apresentam desempenho pior, já que o último preço carrega pouca informação acerca das variações ocorridas dentro do intervalo de tempo.

Finalmente, o desempenho do BVC na classificação de ordens de compra e venda de ativos negociados na B3 pode ter sido significativamente inferior ao apresentado em estudos com dados de mercados mais desenvolvidos devido à maior volatilidade do mercado brasileiro. As variações acentuadas de preços não são bem capturadas por esse método, levantando uma possível necessidade de modificação em sua base de cálculo; e não somente uma calibragem dos seus parâmetros.

5. CONSIDERAÇÕES FINAIS

O objetivo deste artigo foi comparar os desempenhos do TR e BVC na classificação de ordens de compra e venda de ações transacionadas na B3. Observou-se, de maneira geral, que o TR apresentou desempenho superior ao BVC. Além disso, os resultados para o VPIN indicaram forte diferença nas estimativas da probabilidade de negociações privilegiadas a partir dos volumes de compra e venda calculados pelo BVC, o que foi confirmado pela baixa – e, por vezes, negativa – correlação entre o $VPIN_{REAL}$ e o $VPIN_{BVC}$.

A despeito da maior facilidade de se aplicar o BVC – maior acesso às bases de dados e menor volume de dados necessários –, ele apresentou performance significativamente inferior em classificar transações no mercado acionário brasileiro – justificando assim a diferença entre o $VPIN_{REAL}$ e o $VPIN_{BVC}$. Ao se analisarem as propriedades do BVC, verificou-se que o seu desempenho inferior decorre da sua mecânica de determinação da parcela de compra e venda. A distribuição normal padrão produz valores extremos à proporção que a variação de preço é maior. Porém, empiricamente, com o BVC, verificou-se que as parcelas de compra e venda não se distanciam do equilíbrio na mesma medida que o valor de Z. O BVC apresenta melhor desempenho quando

não há variação de preço entre os intervalos. À medida que os intervalos de tempo se tornam maiores, a distribuição utilizada pelo BVC não segue a tendência dos dados reais.

Por outro lado, para o TR, verificou-se que as situações nas quais ele inicia uma sequência de classificações incorretas estão relacionadas principalmente à presença das corretoras de compra e venda nas transações e aos momentos de alta atividade no mercado – transações com pouca ou nenhuma diferença de tempo entre elas.

Conclui-se que, devido ao menor volume transacionado no Brasil, em comparação com os de países mais desenvolvidos, e à maior volatilidade nos preços das ações, o TR se revelou um método melhor para classificar transações. A baixa atividade para certas classes de ações, em comparação com as dos mercados para os quais o BVC foi desenvolvido, pode ter contribuído para que a sua mecânica de classificação não seja aplicável com sucesso semelhante no Brasil. Como dados *tick-by-tick* são indisponíveis para muitos pesquisadores, é necessário recorrer a dados agregados. As evidências aqui documentadas sugerem que o BVC deve ser utilizado com cautela, dado que seu desempenho varia muito a depender da natureza da ação. Dessa forma, os pesquisadores (ou investidores do mercado brasileiro) devem se atentar para os problemas do BVC quando não detiverem acesso a dados *tick-by-tick*.

REFERÊNCIAS

Abad, D., & Yagüe, J. (2012). From PIN to VPIN: An introduction to order flow toxicity. *The Spanish Review of Financial Economics*, *10*(2), 74-83. https://doi.org/10.1016/j.srfe.2012.10.002

Aikten, M., & Frino, A. (1996). The determinants of market bid ask spreads on the Australian stock exchange: cross-sectional analysis. *Accounting & Finance*, 36(1), 51-63. https://doi.org/10.1111/j.1467-629X.1996.tb00298.x

Chakrabarty, B., Pascual, R., & Shkilko, A. (2015). Evaluating trade classification algorithms: Bulk Volume Classification versus the Tick Rule and the Lee-Ready algorithm. *Journal of Financial Markets*, *25*, 52-79. https://doi.org/10.1016/j.finmar.2015.06.001

Easley, D., Kiefer, N., O'Hara, M., & Paperman, M. (1996) Liquidity, information, and infrequently traded stocks. *The Journal of Finance*, *51*(4), 1405-1436. https://doi.org/10.1111/j.1540-6261.1996.tb04074.x

Easley, D., Lopéz de Prado, M., & O'Hara, M. (2011). The microstructure of the "Flash Crash": Flow toxicity, liquidity crashes, and the probability of informed trading. *The Journal of Portfolio Management*, *37*(2), 118-128. http://doi.org/10.3905/jpm.2011.37.2.118

Easley, D., Lopéz de Prado, M., & O'Hara, M. (2012a). Flow toxicity and liquidity in a high frequency world. *Review of Financial Studies*, *25*(5), 1457-1493. https://doi.org/10.1093/rfs/hhs053

Easley, D., Lopéz de Prado, M., & O'Hara, M. (2012b). Bulk classification of trading activity. *Johnson School Research Paper Series*, (8), 1-40. http://doi.org/10.2139/ssrn.1989555

Easley, D., Lopéz de Prado, M., & O'Hara, M. (2016). Discerning information from trade data. *Journal of Financial Economics*, *120*(2), 269-286. https://doi.org/10.1016/j.jfineco.2016.01.018

Ellis, K., Michaely, R., & O'Hara, M. (2000). The accuracy of trade classification rules: evidence from NASDAQ. *Journal of Finance and Quantitative Analysis*, *35*(4), 529-551. http://doi.org/10.2307/2676254

Grammig, J., & Theissen, E. (2002). Estimating the probability of informed trading: does trade misclassification matter? *Bonn Econ Discussion Papers*, *37*, 1-21.

Hwang, L., Lee, W., Lim, S., & Park, K. (2013). Does information risk affect the implied cost of equity capital? An analysis of PIN and adjusted PIN. *Journal of Accounting and Economics*, *55*(1-2), 148-167. https://doi.org/10.1016/j.jacceco.2013.01.005

Mohanran, P., & Rajgopal, S. (2009). Is PIN priced risk? *Journal of Accounting and Economics*, *47*(3), 226-243. https://doi.org/10.1016/j.jacceco.2008.10.001

Omrane, W., & Welch, R. (2016). Tick test accuracy in foreign exchange ECN markets. *Research in International Business and Finance*, *37*, 135-152. https://doi.org/10.1016/j.ribaf.2015.10.001

Panayides, M., Shohfi, T., & Smith, J. (2019). Bulk volume classification and information detection. *Journal of Banking and Finance*, *103*, 113-129. https://doi.org/10.1016/j.jbankfin.2019.04.001

Wei, W., Gerace, D., & Frino, A. (2013). Informed trading, flow toxicity and the impact on intraday trading factors. *Australasian Accounting, Business and Finance Journal*, *7*(2), 3-24. http://doi.org/10.14453/aabfj.v7i2.2

Wu, K., Bethel, W., Gu, M., Leiweber, D., & Rübel, O. (2013). A big data approach to analyzing market volatility. *Algorithmic Finance*, *2*(3-4), 241-267. https://doi.org/10.2139/ssrn.2274991

CONTRIBUIÇÕES DE AUTORIA

LSS: conceptualização da pesquisa; análise formal; investigação, metodologia; desenvolvimento do código e escrita. **LFC:** conceptualização da pesquisa; supervisão; metodologia. **HFA:** supervisão; financiamento.

CONFLITO DE INTERESSE

Os autores declaram que não há conflito de interesse na produção deste artigo e que não há relação de nenhum autor, direta, ou indiretamente através das instituições promotoras da pesquisa, com empresas que eventualmente possam se beneficiar dos resultados do estudo.