BBR
BRAZILIAN BUSINESS REVIEW

ARTICLE

Analysis of the Tick Rule and Bulk Volume Classification Algorithms in the Brazilian Stock Market

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

ABSTRACT

This study aimed to compare the performance of Tick Rule (TR) and Bulk Volume Classification (BVC) models in classifying assets traded on the Brazilian stock exchange (B3) and indicate which one performs better as an investment decision tool. The assets were split into three groups based on their volume, and actual data was used to assess the accuracy of both algorithms. Data from 2018 was used to estimate the parameters that best fit BVC, and transactions from 2019 were used to test the algorithm's efficiency. Afterward, the Volume-Synchronized Probability of Informed Trading (VPIN) was calculated for each asset using TR and BVC, and the values obtained were compared against VPIN calculated using real data. In conclusion, the TR algorithm shows betters performance than BVC for all three groups of assets. Analysis of the properties of both methods reveals that the base upon which the TR is built holds up in the Brazilian market, whereas BVC mechanics does not reflect the observed reality

KEYWORDS

Tick Rule, Bulk Volume Classification, VPIN, Market microstructure

[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

Received: 02/18/2021.
Revised: 09/28/2021.
Accepted: 02/14/2022.
Published Online: 11/18/2022.
DOI: http://dx.doi.org/10.15728/bbr.2023.20.1.6.en

1. INTRODUCTION

According to the traditional asset pricing perspective, the supply and demand of securities in financial markets are equal and, therefore, define equilibrium prices. However, the literature on market microstructure argues that the price formation process is far more complex since financial actors do not have access to the same information and do not enter the market simultaneously. Thus, fundamental assumptions of traditional pricing models such as the absence of transaction costs and symmetric information are made flexible to better understand the dynamics of prices from the market microstructure perspective. Therefore, the informational content carried by securities prices is one of the strands of study on this topic.

Given the increasing amounts of stocks traded in high-frequency markets and the simultaneous growth in the availability of tick-by-tick data on financial information platforms, researching the microstructure of markets has gradually become more feasible. To study the effects of microstructure on the price formation process, such as informational asymmetry, information about transactions is needed, including which side initiated the trade.

As Easley et al. (2012b) describe, trades in a financial market comprise the buy and sell positions. The position that initiated the trading of the asset may indicate information asymmetry between the transaction participants. Indeed, the disparity between trading volumes of the buy and sell positions may indicate order flow toxicity. However, determining the position that initiated a transaction is no simple task, especially in high-frequency markets where information is rarely available. Faced with this problem, several trade classification algorithms have emerged, including Tick Rule (TR), Quote Rule (QR), Lee-Ready (L-R), and Bulk Volume Classification (BVC), all of which have allowed determining which position initiated a given transaction from information available in traditional databases.

This paper aimed to compare the performance of the TR and BVC methods in the classification of buy and sell orders of stocks traded on the Brazilian stock market. The choice for testing the accuracy of these methods was made because they use, respectively, tick-by-tick data and compressed data in time or volume intervals. Considering the type of market (in B3, trades are carried out based on orders sent through the brokers' systems, that is, an "order-driven market"), the higher informational asymmetry and volatility of returns and the smaller trading volume at B3 (which is typical of emerging markets), it is expected that privileged information tends to reach the market sequentially, instead of in volume bulks. Therefore, the risk tends to decrease after a sequence of orders from investors having more information (informed traders), thus impacting the price negatively. In this case, algorithms developed for high-frequency markets (BVC) may not have the same accuracy as traditional ones (TR) in classifying orders in these markets. Thus, it is crucial to test which classification algorithm best distinguishes informed trading in a given market and point out which one is the best tool to aid investment decisions.

To test the accuracy of the BVC, stocks were divided into three groups according to their respective traded volume. This approach is in line with other empirical studies (Easley et al., 2012b; Panayides et al., 2019) showing that this algorithm performs differently depending on trading volume. In addition, trades from 2018 were used to estimate the most accurate parameters of the algorithm for each group of stocks. To verify whether its performance remained close to that observed in 2018, these parameters were tested using trades of 2019.

Next, to analyze the impact of the classification method for stock transactions in Brazil, we employed Easley et al. (2011) VPIN, which measures the probability of privileged trades for a given stock. The choice for VPIN was due to the need to collect information on buying and selling volume for its calculation. Based on actual data, it was possible to compare the results of VPIN estimated by the TR and the BVC and, thus, draw conclusions about their efficiency.

In our study, BVC significantly underperformed TR, producing estimates for VPIN that were weakly correlated with VPIN calculated from actual data. This is unlike Easley et al. (2012b), but in line with other empirical evidence on BVC performance (Chakrabarty et al., 2015; Omrane & Welch, 2016; Panayides et al., 2019).

The difficulty of classifying the side that initiates transactions has led several authors to propose solutions based on available data, whether tick-by-tick or compressed. Tick-by-tick algorithms require as little granularity as possible, i.e., transaction by transaction. This aspect imposes two challenges on researchers: (i) access to data; and (ii) computational capacity to handle data. Most data providers offer the data in compressed form, in time intervals (1 minute, 5 minutes, 15 minutes, and so on). On the one hand, approaches that rely on compressed data are more affordable for most researchers. Indeed, as for the data used in this paper, the volume was reduced to 12% of its original size by compressing it in 5-minute time intervals, making its handling more feasible and less computationally intensive. On the other hand, the use of compressed data brings along the disadvantage of causing the loss of information intrinsic to the period analyzed since it summarizes thousands of transactions into a single measure (average, median, etc.).

Therefore, this paper contributes especially to researchers who have limited access to real data, in that it becomes necessary to consider the biases of using aggregated data to classify transactions, such as the volume of stocks traded and the algorithm parameters, which are decisive factors for the correct classification of the buying and selling volume of assets. The following section presents the TR and BVC classification algorithms and their respective empirical applications. Subsequently, the theoretical framework and the VPIN calculation model will be presented.

2. LITERATURE REVIEW

2.1. Trade classification algorithms

2.1.1. Tick Rule

The TR algorithm uses the price of transactions to classify them. When the price of the current transaction is higher (lower) than the price of the preceding transaction, it is classified as a buy (sell). In cases when the price does not change, the classification given to the preceding transaction is repeated. Easley et al. (2012b) consider this classification method vulnerable (susceptible to errors), especially in high-frequency markets. For the US market, these authors identified an 86% accuracy of TR when classifying transactions occurring between the months of November 2010 and 2011. In turn, Ellis et al. (2000) pointed to an 81% TR accuracy in classifying transactions on NASDAQ. For the Australian market, the TR accuracy was 75% (Aikten & Frinos, 1996).

On the other hand, Chakrabarty et al. (2015) showed that TR performance in the US market declines over time. Omrane and Welch (2016) found that TR correctly classified only 67% of the transactions in the sample analyzed, corroborating the results of Chakrabarty et al. (2015), who pointed to the decreasing effectiveness of TR – possibly because it is a high-frequency market, where tick-by-tick data classifications are more challenging. Similarly, Panayides et al. (2019) found evidence of reduced TR accuracy for two sample stocks, ranging (i) from 79% to 92% between 2007 and 2008; and (ii) from 39% to 65% in 2017.

2.1.2. Bulk Volume Classification

Because of the problems of TR in correctly classifying transactions in markets with high-frequency trading, Easley et al. (2012b) developed the BVC model, which proposes to reduce the effects of order splitting. BVC relies on standardized price variation to classify trading volumes probabilistically. Its mechanics consists of grouping transactions by time or volume intervals, which are arbitrarily determined or limited according to the structure of the database. After the transactions are grouped, the ratios of the volume of transactions that were initiated by the buyer and seller sides are obtained through equations (1) and (2):

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

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

Where: V_τ is the total volume of the interval τ; V_τ^B and V_τ^S are, respectively, the buying and selling volumes for the interval τ; Z is the cumulative standard normal distribution function; and $\sigma_{\Delta P}$ is the estimated standard deviation of the price variation between intervals.

When prices do not change between the start and end of the interval, BVC divides its volume equally into buy and sell (Easley et al., 2012a). When the price increases (decreases), a higher ratio of the volume is classified as initiated by the buy (sell) side of the transaction. The volume ratio classified as buy or sell grows as the price change increases.

The BVC method treats the price variation as independent and identically distributed (i.i.d), with a mean of zero and a constant variance ($\sigma_{\Delta P}$). Easley et al. (2012b) consider that it would be ideal to obtain the true cumulative distribution function of the price variation. However, among other problems, this procedure would decrease the accuracy in classifying transactions and render it impossible to generalize BVC. Thus, assuming normality in the price variations of securities is the most appropriate way to estimate the buying and selling volumes. According to these authors, BVC's satisfactory performance in classifying stock transactions with high trading volume is due to the normal approximation procedure. As for illiquid assets, they consider tick-by-tick methods to be the most appropriate. The results revealed that BVC correctly classified (i) 86.61% of the trades of the e-mini S&P500 futures contract when using 1-minute intervals, and (ii) 87.35% when using 5-minute intervals, whereas the TR accuracy remained at 86.43% for both intervals.

Chakrabarty et al. (2015) compared the efficiencies of the L-R, TR, and BVC algorithms for classifying trades in spot markets. By analyzing actual data relative to 1,471 stocks, the authors found that the TR and L-R accuracies are higher than the BVC accuracy. TR and L-R correctly classified 90.8% and 92.6% of the trades respectively, whereas the best accuracy rate of BVC was approximately 80% when using 1-minute intervals. Omrane and Welch (2016) corroborate these results as they found that both TR and QR showed better results than BVC for the foreign currency market.

Following the criticism of BVC, Easley et al. (2016) argued that, in research on informational asymmetry, the most appropriate approach would be to obtain the underlying information of trades, which depends on proxies because it is unobservable. To compare the accuracies of TR and BVC, they employed three proxies for the information underlying the trades and found that BVC outperformed TR in two of them. Thus, they concluded that BVC allows them to distinguish the intentions inherent in transactions.

Indeed, Panayides et al. (2019) corroborate this empirical evidence. They argue that when BVC is calibrated for a given market with the correct time interval or volume, it can convey the informational content present in the analyzed trades with a higher degree of accuracy.

2.2. VPIN Calculation

Easley et al. (1996) proposed the Probability of Informed Trading (PIN) as a methodology to quantify the probability of occurrence of trades initiated by insiders. The method is based on the disparity of buy and sell trades of a given stock on independent days.

To solve the problem of non-convergence of the maximum likelihood function of the Probability of Informed Trading (PIN) on days when the number of orders is high, Easley et al. (2012a) developed the Volume-Synchronized Probability of Informed Trading (VPIN), which allows for the direct quantification of the level of order toxicity without needing to estimate parameters by maximum probability.

The idea behind VPIN is to separate the volume information for a given day into equal sets (volume buckets) and treat each one as a unit equivalent to one information arrival time. The transaction imbalance is estimated by the average over n volume buckets. Thus, VPIN is obtained from equation (3)

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

Where V_τ^B and V_τ^S are, respectively, the buying and selling volumes in a given t volume bucket. Following the standard calculation employed in the literature, a value of n equal to 50 was adopted, indicating that trades are grouped into 50 buckets of equal transactional volume per day. Based on this aggregation, VPIN is estimated directly through equation (3), that is: (i) the absolute imbalance between buy and sell orders is calculated; and (ii) this value is divided by the number of transactions observed for each set of trading volumes.

Since it represents an extension of PIN, which is a well-tested proxy, VPIN has also been explored in microstructure studies and has had success in predicting relevant events such as the Flash Crash (Wu et al., 2013). Abad and Yagüe (2012) point out that the order flow toxicity measured by VPIN is directly associated with the adverse selection problem to which market makers are susceptible. The buy and sell order imbalance over a short period is related to the information underlying the VPIN.

3. METHODOLOGY

3.1. Population and sample

The population studied in this paper consisted of the stocks traded on the B3 between January 02, 2018, and June 28, 2019. The year 2018 was used to evaluate the set of parameters that produce the best performance of BVC, and 2019 was used to validate these parameters and compare the performances between BVC and TR. Only the 181 stocks traded every day over the period studied were included in the sample. We adopted the criterion of selecting only those assets that were traded on all days so that the construction of the BVC time or volume intervals would not be affected by external factors relating to long periods between transactions.

The next segmentation refers to the volume of shares traded for each asset. As this is the input used by the algorithms, each asset was allocated to a class referring to its average volume traded in 2018. Unlike Panayides et al. (2019), who segmented the assets into three classes with a similar amount in each, in the present paper, we chose to use the Fisher-Jenks algorithm to separate the assets into three volume classes: small, medium, and large. This algorithm was chosen because it allows for defining the threshold points and isolating the assets within their respective classes. This, in turn, decreases the variance between assets of the same class and increases the variance in relation to the assets of the other classes. The number of assets and the average volume traded in 2018 for each class of assets are displayed in Table 1.

The average volume traded for small stocks is close to that reported by Panayides et al. (2019) in the European market; however, medium and large stocks were 44% and 30% lower than what was reported by these authors, which, in turn, points to the first difference between the Brazilian stock market activity and that in more developed countries.

Table 1
Number of assets and average volume per class

Class	Average volume	Number of assets
Small	287,688	99
Medium	1,287,740	39
Large	7,109,882	43

Source: Research data.

3.2. Data collection

One of the primary limitations of the application of PIN and VPIN is the possibility of misclassifying buy and sell orders. To contribute to the analysis of the classification algorithms' performance, this study relied on actual data traded on the Brazilian market as a basis for comparison with the results generated by the TR and BVC algorithms. The data was collected from B3's market data directory, which contains information about the issued orders of all stocks traded in B3 in the last two years, as well as the time, price, amount, and side that initiated the transaction.

The volume of data used in this paper totaled about 150 million rows, where each row represents a buy or sell order executed in the referenced period, averaging 2.6 million shares traded per day. In comparison, the sample used by Panayides et al. (2019) contained an average of 4 million shares traded daily.

Finally, several aggregations were performed for the application of BVC, reducing the volume by about 88% when using a 5-minute interval, which, in turn, shows the advantage of using aggregated data.

4. ANALYSIS AND RESULTS

In this section, the results of the accuracy rate of the TR and BVC algorithms are discussed. We compare the values of $VPIN_{ACTUAL}$ – calculated from the actual buy and sell amount – and those of $VPIN_{TR}$ and $VPIN_{BVC}$ – calculated from the volumes estimated by TR and BVC, respectively. Finally, the properties of the TR and BVC methods are investigated to highlight at what points these algorithms misclassify trades.

4.1. Calibration of the BVC parameters

It is important to calibrate the BVC parameters to subsequently compare the performance of the TR and BVC algorithms. Following Panayides et al. (2019), and considering Easley et al. (2012b) ponderation that BVC performs differently for securities with different transaction volumes, several parameters were tested in 2018 to define the best set for each asset class. In addition, the parameters were tested with 2019 data to check whether the previous performance held and thus attest to the possibility of applying BVC to future data.

To select the best set of parameters for each asset, the BVC accuracy was calculated using equation (4).

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

Wherein: V_B and V_S are the actual buying and selling volumes; and $\hat{V}_B$ and $\hat{V}_S$ are the buying and selling volumes estimated by BVC, respectively. For each asset, the highest accuracy was selected. Table 2 shows the representativity of each parameter within the three analyzed classes, in percent.

For all three classes, BVC showed higher accuracy when using a 5-minute interval. Interestingly, in line with previous studies (Easley et al., 2012b), assets with lower traded volume showed less consistency in terms of overall parameters, as assets were almost evenly split between the 5-minute time intervals and the volume intervals of 75, 100, 200 and 500 thousand shares.

This first evidence creates uncertainty regarding the applicability of BVC as a forecasting algorithm, considering that, among the assets with lower traded volume, the parameters do not present consistency. Indeed, other forms of clustering of assets were tested, and, among the 80 smallest assets, the parameters did not stabilize. This suggests that this phenomenon persists even when the traded volume is used to separate the stocks (a customary practice in the literature).

Another important feature in forecasting algorithms is the applicability of the parameters in various periods. The percentages of stocks for which the most accurate parameter was maintained in the years 2018 and 2019 differed between the groups analyzed, as follows: (i) 78% among high-volume (large) stocks; (ii) 74% among intermediate volume (medium) stocks; (iii) and 35% among low volume (small) stocks.

As before, inconsistency was more significant in lower-volume assets, which advises caution when using BVC for this asset class. Having established the best set of parameters for BVC (i.e., the 5-minute interval), subsequent results will use these values for estimating the buying and selling volumes.

Table 2
Percentage of representativeness of the parameters used for BVC calibration

		Asset class		
Grouping	Parameter	Small	Medium	Large
Time	1 minute	0.00%	0.00%	0.00%
Time	2 minutes	0.00%	0.00%	0.00%
Time	3 minutes	1.01%	0.00%	11.63%
Time	**5 minutes**	**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%

Source: Research data.

4.2. Performance of the TR and BVC algorithms

Table 3 presents the results of the TR and BVC accuracy rates. It can be seen, for both methods, an improvement in performance as the assets grow in traded volume. The overall performance of TR was 80.82%, a higher value than those found by Omrane and Welch (2016) and Chakrabarty et al. (2015).

Table 3
TR and BVC accuracy rates

	TR			BVC		
Class	Minimum	Mean	Maximum	Minimum	Mean	Maximum
Small	62.99%	77.71%	92.63%	33.08%	51.60%	67.28%
Medium	74.95%	82.95%	91.48%	37.19%	62.18%	68.69%
Large	75.10%	86.05%	95.60%	40.89%	64.10%	70.90%
Total	62.99%	80.82%	95.60%	33.08%	56.85%	70.90%

Source: Research data.

The average performance of BVC was 56.85%, lower than the percentage found by Easley et al. (2012b), who analyzed the three most active futures contracts in the US market; and Omrane and Welch (2016), who analyzed foreign exchange contracts.

Overall, we can see that TR outperformed BVC. While the lowest accuracy rate of TR was 62.99%, BVC achieved a minimum rate of 33.08%. In addition, for all asset classes, TR performed above 90%, whereas BVC reached a maximum of 70.90% for the assets with the highest trading volume.

Figure 1 shows that the performance rate of TR is concentrated in the 80% range and the accuracies among the assets present symmetric behavior in relation to the median. BVC values are around 63%, the first quartile is around 45% and the third quartile is around 66%, slightly above the median (64%). As before, this asymmetry stems from the assets with lower trading volume, which typically deliver inferior performance. This result corroborates the evidence found by Easley et al. (2012b), that BVC performs better for stocks with higher trading volumes.

The preliminary results indicate that the TR algorithm outperforms the BVC algorithm. The next section shows the result of the practical application of the two methods based on a model that requires information on buying and selling volumes as its primary input.

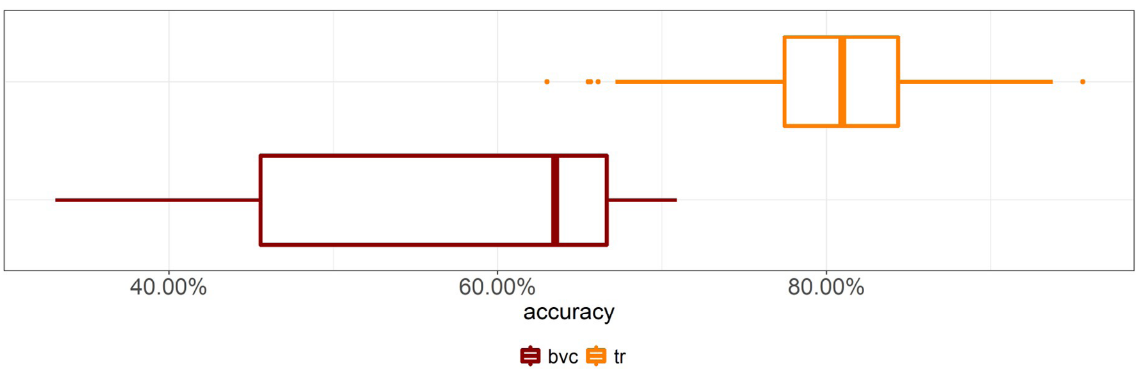

Figure 1. Accuracy range of the TR and BVC methods
Source: Research data.

4.3. Calculating VPIN from actual data, TR and BVC

To analyze the problems regarding the classification of trades when applying a method that requires the number of buy and sell trades, we proceeded to calculate VPIN using the actual data of transactions performed between January 02 and June 28, 2019, in addition to the volumes determined by TR and BVC. Grammig and Theissen (2002) and Hwang et al. (2013) draw attention to the problems concerning the misclassification of orders when estimating informational risk proxies.

Figure 2 shows the average VPIN of each set. At first, the disparity of VPINs between the largest and smallest stocks can be identified. This result has been reported by several authors (Easley et al., 1996; Mohanram & Rajgopal, 2009; Abad & Yagüe, 2012; Wei et al., 2013) and points to the existence of a negative correlation between VPIN and the company's market value.

From these results, we can see that $VPIN_{ACTUAL}$ and $VPIN_{TR}$ do not differ significantly, particularly for medium and large stocks. This evidence is reinforced by the results shown in Table 4, in which the difference between TR and the actual data varied around 2% or 3%.

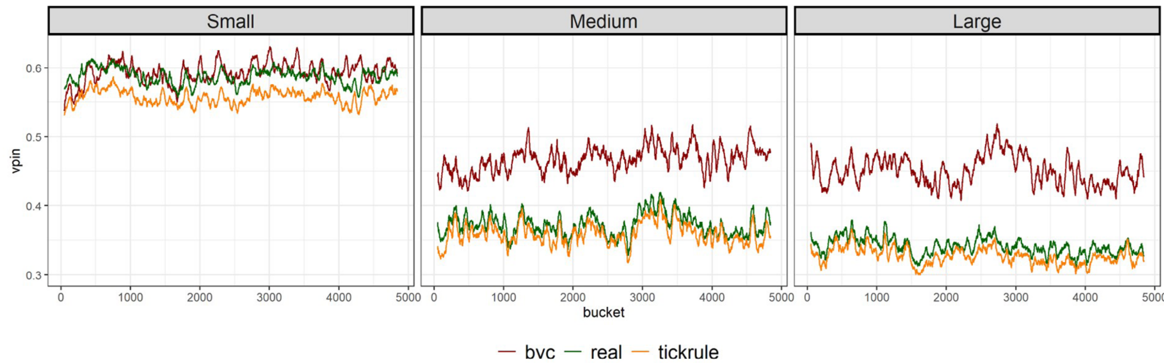

Figure 2. VPINs calculated from actual data, TR and BVC
Source: Research data.

The $VPIN_{BVC}$ calculated for small stocks must be highlighted. Although the lower accuracy of BVC was evidenced for this asset class, its VPIN was the closest to the actual one among the three classes. This result is explained by the particularity of the VPIN methodology, in which the imbalance of orders is taken into consideration. If the buy and sell orders are incorrectly estimated but their imbalance is close to the actual estimate, the VPIN will be close to the one calculated with the actual data. This may advise additional caution when using BVC, as it may indicate promising results from incorrect data; this, in turn, may render it inapplicable to other methods that require the buying and selling data as an input.

Table 4
Average VPINs by class and method

Class	$VPIN_{ACTUAL}$	$VPIN_{BVC}$	$VPIN_{TR}$
Small	58.79%	59.36%	55.82%
Medium	37.32%	46.76%	35.98%
Large	34.17%	45.11%	32.74%

Source: Research data.

To analyze the characteristics of the estimated VPINs, we calculated, for each asset class, the correlation between $VPIN_{ACTUAL}$ and $VPIN_{TR}$, and between $VPIN_{ACTUAL}$ and $VPIN_{BVC}$. The weakest correlation, the average, and the strongest correlation between each asset class, as well as the mean, were also calculated (Table 5).

For all asset classes, $VPIN_{TR}$ presented a strong correlation with $VPIN_{ACTUAL}$, averaging roughly 80%. These numbers reinforce the accuracy verified for TR. In contrast, analysis of the correlation of $VPIN_{BVC}$ with $VPIN_{ACTUAL}$ points to a lower average, of 50% for medium-class assets at most.

The maximum correlations achieved by BVC are close to the average correlation of TR. For small stocks, there were even cases of negative correlation, which indicates that the imbalance reported by BVC showed an opposite sign to the actual data. This means that, while $VPIN_{ACTUAL}$ indicates moments of increased informational risk (alerting to the orders' imbalance), $VPIN_{BVC}$

may indicate the opposite. Therefore, this contradicts the very purpose of VPIN, which, according to Easley et al. (2012a), is to alert investors of moments of volume imbalance and thus avoid illiquidity events that result in stock market crashes, such as the Flash Crash.

Table 5
Correlation between $VPIN_{ACTUAL}$-$VPIN_{TR}$ and between $VPIN_{ACTUAL}$-$VPIN_{BVC}$

	TR			BVC		
Class	Minimum	Mean	Maximum	Minimum	Mean	Maximum
Small	0.3551	0.7817	0.9732	-0.1769	0.2872	0.7630
Medium	0.5733	0.8636	0.9855	0.1018	0.5018	0.8563
Large	0.6889	0.8544	0.9713	0.2231	0.4508	0.8632

All correlations showed p-values equal to zero.
Source: Research data.

As a way of showing the consequence of using BVC to classify trade orders, Figures 3 and 4 show the behavior of VPINs for the stocks that presented, respectively, stronger and weaker correlation with the actual data for each class.

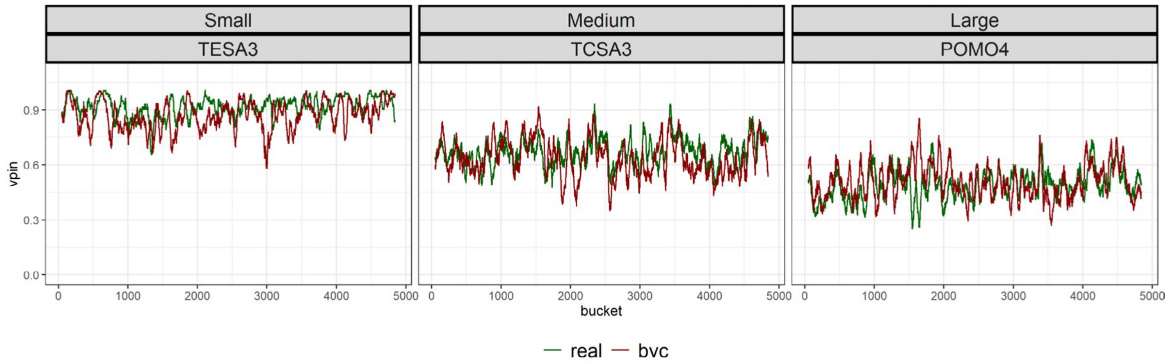

Figure 3. Stocks with a stronger correlation between BVC and actual data
Source: Research data.

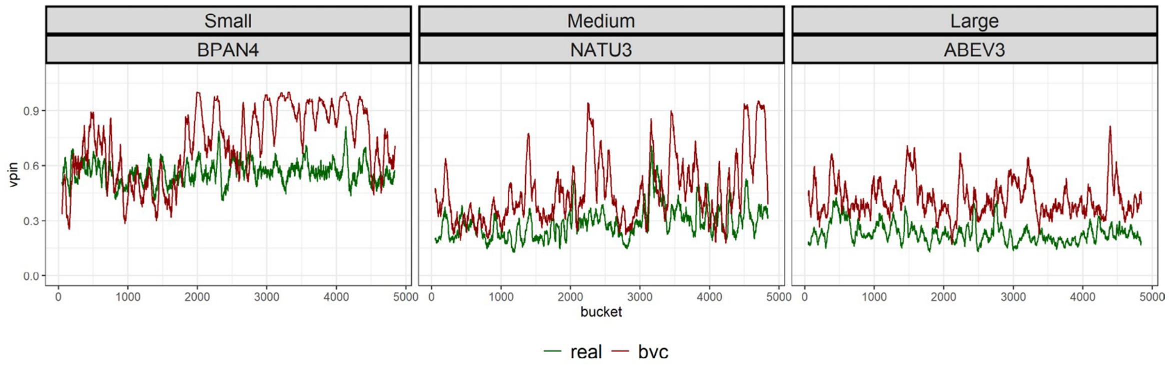

Figure 4. Stocks with a weaker correlation between BVC and actual data
Source: Research data.

The figures show that $VPIN_{BVC}$ presents more extreme values than $VPIN_{ACTUAL}$ even for the stocks with a stronger correlation. As for the stocks with weaker correlation, $VPIN_{BVC}$ reached values close to 90% of $VPIN_{ACTUAL}$ on occasion. If used as an indicator of liquidity problems, this method would present multiple false positives when compared to the actual value, which could lead to problems in its practical use. The unbalanced behavior of BVC will be analyzed in the section dedicated to the analysis of the properties of this method.

The results presented in this section indicate that BVC is not an effective trade classification algorithm compared to actual data. This evidence is corroborated by the application of VPIN, which shows that the values estimated by BVC differ substantially from those obtained from actual data. On the other hand, $VPIN_{TR}$ showed similar values to $VPIN_{ACTUAL}$ for all stocks analyzed in this study. In the next sections, we will identify and analyze the points when TR and BVC misclassify orders, which would explain the differences in VPIN estimates.

4.4. Analysis of the Tick Rule properties

The TR algorithm is based on the economic principle that a buy (sell) order increases (decreases) the demand for a given stock, which leads to an increase (decrease) in its price. To verify in which situations this economic principle holds, we analyzed the frequency of the order signs given the price changes of the trades. That is, we checked, for each value of ΔP_t, the number of buy (for $\Delta P_t>0$) and sell orders (for $\Delta P_t<0$) in relation to the total. Finally, we calculated the number of times the order was repeated for moments when $\Delta P_t=0$. Thus, equations (5), (6), and (7) represent the calculations performed.

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

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

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

Where: B_t and S_t are, respectively, a buy and sell order at the moment *t*; and P^+ and $P\_$ are the positive and negative values for the price variations between transactions, respectively. Lastly, X_t represents the sign of the order posted at the moment *t*, and it can be either a buy (B) or sell order (S). Equation (7) represents the case where the price change equals zero and we intend to check how often the side that initiated the order in *t* is equal to the side that initiated the previous order.

The results of equations (5) and (6) are shown in Table 6. There is consistency in the frequency of the transaction side, both for positive and negative price changes. Even in more pronounced price changes, above 0.20 currency units, the percentage of buy or sell orders remained at the same level (around 88%).

The results in Table 6 show why TR performs well for classifying trades. In general, the basis of this algorithm holds for the sample analyzed, that is, positive price changes point to buy orders, while negative changes indicate sell orders.

For the result of equation (7) related to price changes equal to zero, it was found that $P(X_t = X_{t-1}|\Delta P_t = 0) = 0{,}9531$. That is, for the sample analyzed, in 95.31% of the cases, when there was no price change, the transaction at the moment *t* was the same as at the moment *t*-1, as recommended by TR.

To further analyze the situations in which TR misclassifies trades, we verified that five variables influence the performance of this method, namely the price change (ΔP_t); the sign of the order being sorted; the sign of the previous order; the difference in time between the two trades; and whether the buying and selling brokers are the same as in the previous transaction.

Table 6
Frequency of order signs given the price changes between transactions

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%

Source: Research data.

Table 7 presents the situations and frequencies in which TR initiates a sequence of misclassified transactions. Most of the TR errors derive from situations where the price change is positive, but the order is classified as a sale preceded by another sale. In this case, the brokerage firm involved in the sales at *t* and *t*-1 is the same, whereas the buyer is different. Therefore, the situation described is one in which: (i) a given brokerage firm places a sell order at *t*-1; when this order is executed, (ii) another sell order is placed by the same brokerage firm with a difference of zero (0) seconds, executed by a different buyer than the one who sent the order. In this case, the second order (at *t*) has a higher price than the trade at *t*-1, which is probably due to the fast execution of the sell order. This suggests that there is liquidity for the stock at that moment and that its demand is high, which explains the increase in the sale price.

This is also true for the situation where the current and previous orders are both buy orders, but the price varied negatively between trades (row 2 of Table 7). In this case, the broker sending the buy orders is the same for both transactions but is not the seller. The second buy order is executed faster than the first (zero-second difference), indicating that many traders are interested in selling the stock (i.e., the supply is high). This, in turn, leads to a reduction in the price of the transaction, which is executed even at a lower price than the previous transaction.

The scenarios described above were those that presented the highest frequencies of trades misclassified by TR. When one of these situations occurs, a sequence of misclassifications can follow if there are no further price changes. This is because, in this case, TR continues to misclassify the transaction from the sign of the previous order that had been misclassified before.

Table 7
Scenarios in which TR initiates a sequence of misclassified trades

Time difference	ΔP_t	Current order	Previous order	Buying Broker	Selling Broker	Frequency
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%

Note: ‘+,’ ‘-’ and ‘0’ respectively mean positive and negative price changes (of any magnitude) and no changes. The values in the ‘Buying Broker’ and ‘Selling Broker’ columns, ‘=’ and ‘≠,’ represent whether the broker is the same or different from the previous transaction.
Source: Research data.

The next two situations involving TR errors are those in which the orders at *t* are buys (sells), and in which the orders at *t*-1 are sells (buys), but with an equal price change and time between transactions equal to zero, and the same selling (buying) broker. As in this case the TR repeats the classification, a sequence of errors is initiated. Indeed, the time between transactions and the brokers involved in them play a significant role in defining which side initiated the buy or sell order. Since trades are practically instantaneous, two phenomena can influence the TR classification. The first has to do with order splitting, that is, the same order is divided into several smaller orders so that the market does not notice that a trader is moving a high volume of shares. This type of strategy can be detected by analyzing the time between transactions, the broker executing the order, and the volume traded, since many times the order is split into multiple orders of equal volume. The second phenomenon happens because since the time between trades is practically zero, the market does not adjust in time for the price variation to reflect the supply and demand of the stock in question.

In general, the results in Table 7 show that the scenarios present similar frequency for the orders' signals, therefore pointing to symmetry, besides revealing that the market behaves equivalently regardless of whether the trade in question is initiated by a buyer or a seller. Finally, this analysis points to the opportunity to build a more complex model that can capture the relationship between the variables, thus reducing the initiation of the error sequence.

Having identified the scenarios in which TR misclassifies trade signs, we proceed to analyze the properties of BVC in the next section.

4.5. Analysis of Bulk Volume Classification properties

By using the normal distribution to compute the buy and sell percentage for each time interval, it is possible to compare them with the actual buy and sell percentages within the same interval. Figure 5 shows how the percentage of buying evolves with price variation compared to the percentage signaled by BVC. There were no major disparities between asset classes, so the values reported in Figure 5 represent the complete (total) analyzed sample.

Evidence shows that one of the primary characteristics of BVC is corroborated; that is, when price does not vary, the percentage of buying and selling within the same interval is roughly 50% (51.88% of buying volume in the sample analyzed). In practice, this makes BVC perform satisfactorily for intervals that do not present price variation (about 22% of the intervals).

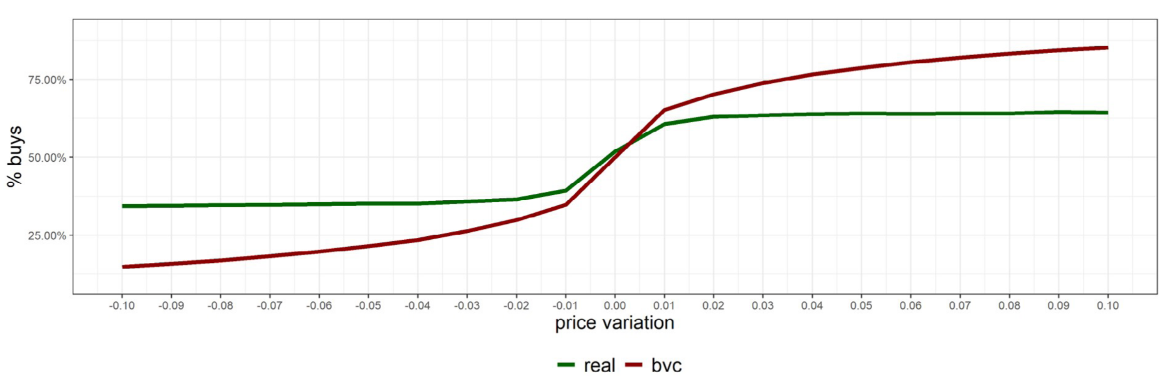

Figure 5. Actual and estimated buy percentage by BVC in relation to price variation
Source: Research data.

However, as the price variation moves away from zero, the percentage signaled by BVC increases more rapidly than is found in practice. This characteristic derives from the distribution defined in the model design. Through the actual data, we can see that, on average, the percentage of buying stabilizes near an absolute price variation of about 0.05 currency units. Given the distribution chosen in the application of BVC, this stabilization does not occur within the first 0.1 currency unit of absolute change.

This behavior explains why $VPIN_{BVC}$ has more frequent peaks than $VPIN_{ACTUAL}$ or $VPIN_{TR}$. Since BVC assigns a higher percentage of buying or selling even for low price variations, it is natural that the volume imbalance presented by this model will be higher as well, leading to undetected peaks when relying on actual data to calculate VPIN.

Moreover, as observed in the analysis of the TR properties, when the price varies, in about 88% of the cases the trade occurs in the direction of the variation; that is, the price increase indicates a buy, and the price decrease indicates a sell. This holds when the analysis is performed transaction by transaction, which led TR to achieve about 80% accuracy in the sample analyzed. In contrast, BVC groups transactions into intervals and uses the last price as a supply or demand indicator. This implies that all the information content present within the interval (captured by TR) is discarded when BVC is used. This also explains why intervals calculated with longer periods or a larger number of aggregate transactions have inferior performance since the last price contains scarce information about the variations occurring within the time interval.

Finally, the performance of BVC in classifying buy and sell orders for assets traded on B3 may have been significantly inferior to that found in studies with data collected from more developed markets, due to the higher volatility of the Brazilian market. Sharp price variations are not properly captured by this method, which suggests the need to modify its calculation basis in addition to the calibration of its parameters.

5. FINAL REMARKS

This paper aimed to compare the performance of TR and BVC in classifying buy and sell orders for stocks listed on the B3 stock exchange. In general, TR outperformed BVC. Also, the VPIN results pointed to a significant difference in the probability estimates of insider trades from the buy and sell volumes calculated by BVC, which was confirmed by the low – and occasionally negative – correlation between $VPIN_{ACTUAL}$ and $VPIN_{BVC}$.

Despite the greater ease of applying BVC, due to its expanded access to databases and the smaller volume of data required, it showed significantly inferior performance in classifying transactions in the Brazilian stock market, which in turn explains the difference between $VPIN_{ACTUAL}$ and $VPIN_{BVC}$. By analyzing the properties of BVC, we concluded that its underperformance stems from its mechanics of determining the buying and selling ratios. The standard normal distribution produces extreme values as the price variation increases. However, in the case of BVC, we empirically verified that the buying and selling parcels do not move away from equilibrium to the same extent as the value of Z. BVC performs best when there is no price variation at all between intervals. As the time intervals become longer, the distribution used by BVC does not follow the trend of the actual data.

On the other hand, in the case of TR, the situations in which the algorithm initiates a sequence of misclassifications are primarily related to the presence of the buying and selling brokers in the trades and moments of high activity in the market, that is, trades with little or no time difference between them.

We conclude that, due to the lower trading volume in Brazil, compared to those of more developed countries, and the higher volatility in stock prices, TR proved to be a more efficient method for classifying trades. The low activity for certain asset classes, compared to those listed in markets for which BVC was developed, may have contributed to its classification mechanics not being applicable with equivalent success in Brazil. Since tick-by-tick data are unavailable to many researchers, they must resort to aggregate data. The evidence documented here suggests that BVC should be used with caution, given that its performance varies greatly depending on the nature of the stock. Thus, researchers (or investors in the Brazilian market) should be aware of BVC problems when they have no access to tick-by-tick data.

REFERENCES

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

AUTHOR'S CONTRIBUTION

LSS: research conception and design; formal analysis; investigation; methodology; code development and writing. **LFC:** research conception and design; supervision; methodology. **HFA:** supervision; funding.

CONFLICTS OF INTEREST

The authors state that there is no conflict of interest in the production of this article and none of the authors have any direct or indirect relationship through the institutions promoting the research with companies that may eventually benefit from the results of the study.