Journal of Financial Markets 25 (2015) 52–79

Contents lists available at ScienceDirect

Journal of Financial Markets

journal homepage: www.elsevier.com/locate/finmar

Evaluating trade classification algorithms: Bulk volume classification versus the tick rule and the Lee-Ready algorithm ☆

Bidisha Chakrabarty [a,*], Roberto Pascual [b], Andriy Shkilko [c]

[a] Saint Louis University, St. Louis, MO, USA
[b] University of the Balearic Islands, Spain
[c] Wilfrid Laurier University, Ontario, Canada

ARTICLE INFO

Article history:
Received 13 August 2013
Received in revised form 1 June 2015
Accepted 3 June 2015
Available online 11 June 2015

JEL classification:
G10
G14
G29

Keywords:
Trade classification
Bulk volume classification
Tick rule
Lee and Ready
VPIN

ABSTRACT

We compare the accuracy of the bulk volume classification (BVC) to that of the tick rule (TR) and the Lee-Ready (LR) algorithm for a large sample of equities. TR and LR produce significantly better classifications than the BVC. This result applies to stocks of all sizes, including the most frequently traded. Iteratively optimizing the BVC improves its performance, but the conventional rules still outperform. TR and LR produce more accurate estimates of the volume-synchronized probability of informed trading. Order imbalances computed using TR and LR are comparable to those computed using the BVC in explaining returns, liquidity, and trading costs.

☆ We thank Marcos López de Prado for answering a number of our questions about the BVC methodology. We are grateful to Maureen O'Hara for insightful comments on our results. We thank two anonymous referees, Amit Goyal (the editor), Robert Battalio, Oleg Bondarenko, Tarun Chordia, Stephen Christophe (FMA Europe discussant), Francesco Consonni, Jagjeev Dosanjh, Amit Goyal, Joel Hasbrouck, Frank Hatheway, Craig Holden, Pankaj Jain, Rebeca Méndez-Durón, Pam Moulton, Andreas Park, Gideon Saar, Heather Tookes, David Veredas, Mao Ye, Bart Zhou Yueshen (SSFMR discussant), and audiences at the FMA European Conference and the Swiss Society for Financial Market Research Conference for helpful discussion and comments. We thank Michael Markes for generous help with ITCH data. Pascual acknowledges the financial support of the Spanish Ministry of Education DGICYT project ECO2010-18567. This paper was partly written while Pascual was a visiting fellow at the International Center for Finance at the Yale School of Management. This research was supported by the Social Sciences and Humanities Research Council of Canada, Canada Foundation for Innovation, and Canada Research Chairs program.

* Corresponding author.
E-mail addresses: chakrab@slu.edu (B. Chakrabarty), rpascual@uib.es (R. Pascual), ashkilko@wlu.ca (A. Shkilko).

http://dx.doi.org/10.1016/j.finmar.2015.06.001
1386-4181/

1. Introduction

Most trades in continuous markets have an active side that takes liquidity and a passive side that provides liquidity. The active side is referred to as the trade initiator, and a trade is classified as a buy (sell) if it is buyer- (seller-) initiated. Although identifying the trade initiator is important for empirical research,[1] most public databases do not contain initiator flags, forcing researchers to infer the trade initiator using trade classification algorithms.

Traditional classification algorithms assign the initiator trade by trade and require processing of large amounts of granular data. In contemporary markets characterized by high speeds, fragmentation, and unprecedented rates of order submission and cancellation, such processing may be quite taxing on a researcher's time and computational resources and, more importantly, result in sub-par classification accuracy. To mitigate these issues, Easley, López de Prado, and O'Hara (2012) propose an alternative classification algorithm: the bulk volume classification (BVC).[2] The BVC focuses on fixed time, volume, or trade intervals called *bars*. Applying probabilistic analysis to price changes between bars, BVC splits the aggregated volume in each bar into the buyer- and seller-initiated volume. Analyzing data on index and commodity futures, ELO conclude that the BVC algorithm is superior to the tick-based algorithm (the tick rule) in resource requirements and provides similar classification accuracy.

Will researchers benefit from switching to the new volume classification paradigm? Are there any trade-offs in such a switch, particularly in the equity markets, where market structure research has been most active? In this study, we offer answers to these questions by extending Easley, López de Prado, and O'Hara's work in several ways.

First, using true trade classification for equity transactions derived from the NASDAQ's TotalView-ITCH (hereafter, ITCH) order book, we show that both the tick rule (TR) and the popular Lee-Ready (LR) algorithm are more accurate than the BVC across the board. More specifically, misclassification increases by 7.4–18.1% when we switch from the TR to the BVC and by 10.3–19.0% when we switch from the LR algorithm to the BVC. For example, the BVC is most accurate when applied to time bars of approximately one-hour (3,900 seconds) duration. For these bars, the BVC correctly classifies 79.7% of volume, while the TR correctly classifies 90.8% of volume and the LR correctly classifies 92.6% of volume.[3] Notably, the BVC accuracy is considerably lower in our equity data than in ELO's futures data. ELO report the highest attained BVC accuracy of 94.5%, which is noticeably higher than the accuracy that we obtain for equities even when we iteratively search for the bars in which BVC performs the best. It appears that the structural differences between equity and futures markets negatively affect the accuracy of bulk volume classification. We emphasize that our goal is to examine BVC applicability to equities rather than to dispute its usefulness for the futures. With this goal in mind, our results tend to suggest that the BVC is not the most optimal classification approach for equities.

Our ITCH order data allow us to compare the results of the bulk and conventional classification techniques to true classification on the NASDAQ, but they come with a limitation; we do not observe orders and trades that occur elsewhere. The reader may therefore wonder if our classification accuracy

---

[1] Some uses of trade classification are to compute order imbalances (Chordia and Subrahmanyam, 2004), measure the costs of market making (Huang and Stoll, 1997), evaluate the informational content of trades (Hasbrouck, 1991), gauge the presence of informed traders (Easley, Kiefer, O'Hara, and Paperman, 1996), predict short-run volatility and impending market crashes (Easley, López de Prado, and O'Hara, 2012), etc.

[2] The BVC algorithm is proposed in Easley, López de Prado, and O'Hara (2012) and further developed in Easley, López de Prado, and O'Hara (2013). Hereafter, we use the abbreviation ELO to refer to the results in both of these papers.

[3] The TR and LR accuracy rates reported by the earlier studies (e.g., Ellis, Michaely, and O'Hara, 2000; Finucane, 2000; Odders-White, 2000; Chakrabarty, Li, Nguyen, and Van Ness, 2007; Chakrabarty, Moulton, and Shkilko, 2012) are in the 75–85% range, which is notably lower than the accuracy rates reported here. The discrepancy arises from the estimation specifics in the bulk volume framework, where misclassified buys and sells offset each other as the length of the estimation interval increases. We discuss this issue in detail in Section 3.3.