Comparing Trade Flow Classification Algorithms in the Electronic Era:

The Good, the Bad, and the Uninformative

Marios Panayides[†]
University of Pittsburgh

Thomas Shohfi
University of Pittsburgh

Jared Smith
Clemson University

This Draft: September 2014

Abstract

We use recent low-latency data from Euronext Paris for which we can identify the true trade initiator to test the performance of the Lee and Ready (1991) and tick test classification algorithms, as well as the newly developed bulk volume classification method (Easley et al. (2013)). We find that, despite the use of quote data, Lee and Ready underperforms the other methods, particularly during intervals of high trade and/or quote frequency. The bulk volume algorithm (BVC) demonstrates superiority with respect to data efficiency, accuracy, and the ability to capture informative trade flow. Consistent with Chakrabarty et al. (2013), we find that the BVC benefits from the netting of misclassified trades and underperforms the "bulk" tick test at comparable bar sizes. Nevertheless, we show that the magnitude of this accuracy underperformance is driven by systematic biases related to both volume and time bar choices, as well as the characteristics of the price change distribution. We explore several BVC calibrations that significantly mitigate, and in some cases alleviate, the accuracy advantage of the bulk tick test while capturing the data efficiency and informational advantage of the BVC.

[†] Corresponding author. E-mail: mpanayides@katz.pitt.edu

1 Introduction

The landscape of markets across the world has experienced dramatic change in the past ten years. Proliferation of high-speed computers and the spread of the electronic limit order book have taken trading off the floors, out of the pits, and onto computers operated by professional trading firms. This has produced an explosion in trading volume and an even larger increase in the speed with which traders trade (Jain (2005); Hendershott and Moulton (2011)). The rapid growth of algorithmic low-latency trading (including high frequency trading (HFT), characterized by rapid order cancellation and order splitting) has called into question the efficacy of traditional methods to identify the trade initiator of each trade (Holden and Jacobsen (2014)). The effectiveness of these trade classification algorithms crucially affects the ability of both researchers and investors to detect informative and/or toxic order flow, estimate trading costs, and describe investor clientele behavior.

This paper examines how the (1) Lee and Ready (1991; hereafter LR), (2) tick test (Holthausen, Leftwich and Mayers (1987)), and, (3) Easley, Lopez de Prado, and O'Hara (2013) bulk volume classification (hereafter BVC) algorithms perform in an equities sample that contains HFT. All three methods are assessed in terms of the accuracy of signing trades, data efficiency, and ability to capture informative trade flow. We use NYSE Euronext equities data ("NextHistory") from April 2007, February 2008 and April 2008 to perform our analysis. These data have several advantages that make them ideal for our study. First, the period of our sample makes the data well suited to testing the effects of low-latency trading on the performance of trade classification algorithms.[1] Second, because we have order-level data, we are able to classify the true trade initiators of an ex-

[1] We discuss algorithmic trading's presence in our data in Section 2.

tremely high percentage of trades executed on the NYSE Euronext.[2] Third, European markets, specifically Euronext, are more consolidated than the rapidly fragmenting U.S. equity markets (see Figure 1 of Menkveld (2013)). The low level of fragmentation, combined with our high level of true trade classification, means that our study uses a large percentage of Euronext equity trading volume. Studies that rely on one U.S. market cannot classify the true trade initiator of a large percentage of the trading volume and thus use a small portion of overall trade volume. This affects their ability to effectively test the performance of trade classification algorithms, particularly the BVC, which relies on using volume and price changes. Our data therefore allow for more unbiased tests and better comparisons of the performance of the trade classification algorithms.

We begin the analysis by investigating the performance of the LR algorithm. Widely used among both academics and practitioners, the impact of the LR algorithm is substantial: as of August 2014, LR has 2,189 citations in Google Scholar, 607 citations in Web of Science, and is the 32nd most frequently cited article throughout the history of the Journal of Finance.[3] Consistent with the notion that more and faster trading impairs LR, we find that the LR classification accuracy is far lower than rates reported in prior literature at only 78.67%.[4] Further, when we isolate fast-paced quote and trade activity (a proxy for low-latency traders), LR performs poorly. In particular, when a given trade is executed in the same second as other trades, it is more than 9% less likely to be classified correctly. When a trade is associated with multiple quotes in the same second, it is 17% less likely

[2] We can classify true trade initiators for 94% of our trade volume. This is a much larger percentage than older studies measuring the Lee-Ready and tick test performance (e.g., Odders-White (2000) classifies 74.9% of trades). Our sample's true trade initiation percentage is comparable to contemporaneous work (see, for example, Chakrabarty et al. (2012)); it has the added advantage of low fragmentation.

[3] Journal of Finance citation rankings are available at http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1540-6261/homepage/top_cited_articles_of_all_time.htm

[4] In Odders-White (2000) the LR algorithm correctly classifies 85% of sample trades; in Peterson and Sirri (2003) the algorithm classifies 90% of trades correctly.

to be correctly classified. With multiple trades and quotes in a single second, the accuracy rate drops nearly 21% vis-à-vis those seconds with only one trade and quote. The result that multiple quotes seem to penalize, rather than provide more information for the trade classification algorithm, is consistent with the model in Baruch and Glosten (2013), in which frequent, flickering quotes contain more randomization than useful information. The authors show that such a quote randomization strategy is necessary for low-latency liquidity suppliers in order to manage risk related to predatory algorithmic trading. Our result is also consistent with Hasbrouck's (2013) finding that HFT-driven quote volatility degrades information within quotes.

We also investigate the performance of the tick test. Overall, we find that the tick test is slightly more accurate than LR across our sample (1.7% better performance). Similarly to LR, trade volume misclassification increases for the tick-test when investors interact in a low-latency environment. However, the decline in accuracy of the tick test is only 4.62% in comparison to 10.69% for LR. In sum, our results suggest that the tick test is a more effective trade level algorithm than LR in fast-paced electronic markets. Importantly, since it requires only traded prices, the tick test is, by construction, more data efficient than LR.

Next, we examine the performance of the newly developed BVC algorithm.[5] The BVC involves putting trades into blocks, or bars, by either volume or time.[6] A percentage of the block is then classified as buys (the remainder is sells) based upon the movement of prices around the bars. By construction the BVC algorithm is highly data efficient as it uses aggregate bar-size trading volume and prices (less than 1% of the trade data points).

[5] Along with the contemporaneous Chakrabarty et al. (2013) paper, our study represents the first equities test of the newly developed BVC algorithm.

[6] ELO use trade bars in addition to volume and time bars when they investigate BVC performance in the futures market. Our study does not include trade bars since the trade size distribution is highly concentrated and discrete in the futures market but significantly less so in the equities market.

With respect to accuracy, we find that the volume (time) bar BVC, offers an overall improvement of up to 11.61% (11.93%) and 13.34% (13.66%) relative to the tick test and LR, respectively. Further, the BVC suffers from none of the systematic issues that make the LR (tick test) algorithm less accurate (e.g., many quotes and trades in a second, seller versus buyer initiated trades) and performs consistently well across varying levels of market volatility.

Thus, our initial tests indicate that the BVC is superior to trade level algorithms not only with respect to data efficiency but also accuracy. Chakrabarty et al. (2013), however, find that putting trades into bars gives the BVC an advantage because misclassified buys and sells offset each other (i.e., trades are netted within the bar). Therefore, to run "netted-neutral" comparisons, we create bulk tick and bulk LR measures by aggregating both the tick test and LR classifications across comparable BVC bar sizes. We find that both the bulk tick test and bulk LR outperform the BVC. In particular, following Chakrabarty et al. (2013) analysis, we find that for time-bar equivalent sizes, the accuracy of the bulk tick test (bulk LR) outperforms BVC on a range from 8.29% to 18.39% (8.08% to 17.54%); similarly for volume bar the bulk tick test (bulk LR) outperforms BVC on a range from 8.28% to 17.16% (7.93% to 16.69%). While the bulk tick test's accuracy advantage over the BVC appears substantial, suboptimal application of the BVC seems to be driving the outperformance. First, there is a systematic bias using volume bars that will lower the accuracy of the BVC compared to the bulk tick test when trades are truncated to make bar sizes exact. In fact, with small volume bars, this bias can be as large as 25%. We also show how the use of flexible minimum, rather than exact, volume bar sizes eliminates this bias. Second, we note the differing relationships between bar size and accuracy for the BVC versus the bulk tick test/LR. The relationship between bar size, both volume and time, and accuracy is strictly increasing for the bulk tick test/LR. For the BVC, however, the effect of bar size is less clear. This greater sensitivity of BVC performance to bar size makes matched size comparisons with the bulk tick test/LR inappropriate and emphasizes the importance of identifying the appropriate bar size. We find that, when using ex-post calibrated bar sizes based on stock and trading characteristics (i.e., trade size, stock trading frequency and stock size), the accuracy advantage of the best overall bulk tick test (bulk LR) relative to the best overall BVC falls to as low as 1.72% (1.39%) for volume bars and 1.49% (0.86%) for time bars, results that are lower than the 7.4% to 16.3% range reported in Chakrabarty et al. (2013).

Increasingly, in today's fast, electronic markets, researchers and practitioners are interested in identifying buying or selling pressure that may destabilize markets and understand whether this is driven by informed or uninformed market participants. Although netting appears to greatly benefit the bulk tick test, resulting in the documented slightly higher accuracy relative to the BVC, it is important to also identify whether the algorithms can correctly identify underlying informed order flow. To do that, we follow the Easley, Lopez de Prado, and O'Hara (2013; hereafter ELO) approach: If an algorithm successfully estimates the underlying informed order flow, then a larger estimated order imbalance should be directly associated with a larger high-low trade price range in a given bar. We thus run regressions of the high-low trading range in a bar on the order imbalance estimated by the BVC and the tick test.[7] We find that while the BVC order imbalance is positively related to the high-low trading range, the tick test order imbalance is negatively related, indicating poor performance in estimating the underlying order flow. Overall, consistent with ELO, we find that the BVC is the only algorithm that can detect the underlying information behind any price pressure.

[7] In an unreported analysis we run these regressions using the bulk LR imbalance and find that it performs strictly worse than the tick test. Combined with its inferior accuracy, we choose not to report the results.

This paper makes several contributions to the nascent literature on low-latency trading and trade classification algorithms (including the BVC), as well as adding to the long list of papers investigating the performance of the LR and tick test algorithms. First, our study compares the newly-developed BVC methodology to existing methodologies using *equities—rather than futures market—data*. This is important because of the institutional differences, clientele effects, and differences in investors' trading behavior (especially for informed traders) that exist between futures and equities markets. For example, there is evidence that block purchases and sales have differential price impact in equities markets but not the futures market (see, for example, Chan and Lakonishok (1993), Berkman, Brailsford and Frino. (2005)). In addition, our paper explores improvements to the BVC's accuracy. Given the data efficiency advantage of the BVC versus trade level algorithms, it is worthwhile to document how the BVC performs in equities, and whether its performance can be improved. The tick test cannot become more data efficient, but it is possible to improve the BVC's accuracy.[8]

Second, this study is one of the first to examine the ability of the commonly used LR algorithm to correctly classify trades in a high frequency trading environment. Our results indicate that the LR algorithm appears weaker when there are multiple trades and quotes for the same reported time stamp (i.e., second).[9] Further, our findings show that LR performs weaker than in prior studies, due to the fact that the methodology does not perform

---

[8] The importance of accuracy improvements for these algorithms is emphasized by Andersen and Bondarenko (2013) who note that "the accuracy of the underlying order imbalance measure is an ever-present source of ambiguity for VPIN."

[9] As of June 2011, Wharton Research Data Services (WRDS) offers NYSE millisecond data as part of the Daily Trade and Quote (TAQ) data set. This would allay some concerns going forward, but it would clearly not be of help when conducting tests on data prior to millisecond time-stamping. Additionally, as trading continues to get faster, the same problem would occur. In fact, the Securities and Exchange Commission has already acknowledged trading faster than in milliseconds: "For example, the speed of trading has increased to the point that the fastest traders now measure their latencies in microseconds" (SEC (2010) page 3605).

well in periods of low latency trading. This is an important point, because misclassifying trades can be costly for investors' strategies, as order flow is an important measure of trading costs (Odders-White (2000)) and order flow toxicity.

The remainder of the paper proceeds as follows. Section 2 describes our data in greater detail. Section 3 describes the methodology used to detect true trade initiators, followed by a review of the Lee and Ready and bulk volume classification algorithms. Section 4 presents the main empirical results. Section 5 discusses calibration and other methodology refinements for improving the accuracy of the BVC. Section 6 specifically examines the ability of each algorithm to detect underlying information in trade flow. Finally, section 7 provides a brief conclusion.

2 Data

2.1 Data Set

Our data come from NYSE Euronext "NextHistory" files. These proprietary data contain all trades and quotes, and also almost all orders submitted for the period 2007 and 2008. The data files do not include iceberg orders that did not participate in trades. The data are time-stamped at the second-level.[10] We focus exclusively on trades on the Paris Bourse. Many different types of securities besides stocks are included in this dataset, but we drop instruments other than common stock from the analysis. Additionally, we focus only on continuously traded French equities, as opposed to those that use once- or twice-daily call auctions. Option market activity, SBF-120 index membership, and abnormal equity return data, used in cross-sectional and regression analyses, are obtained from Bloomberg.

[10] Recent data, including the contemporaneous work of Chakrabarty et al. (2013) include millisecond trade stamps. See Holden and Jacobsen (2014) for more information regarding the impact of timestamp granularity on trade level classification algorithms.

We believe that the quality of this Euronext data set is high due to the relatively greater consolidation of French equity trading at the Euronext exchange. In particular, we capture 82.9% of all trading volume by the Euronext listed equities in our data sample in April 2007. This compares to 63.7% of NYSE listed equities volume occurring on the NYSE exchange in the second quarter of 2007. In April 2008, trades on the Euronext in our sample represented 53.8% of overall volume versus 46.8% for NYSE listed equities.[11] While this margin does decrease over the timeframe of these data, given the large number of shares traded, the on-exchange traded volume of Euronext listed equities is significantly higher than NYSE listed firms. Further, when compared to the 23.8% to 25.6% range of volume share within 2005 INET data used in Chakrabarty et al. (2013), our data exhibit far less market fragmentation. The higher volume share of the Euronext Paris market for French listed stocks makes our performance findings of the three trade initiator classification algorithms less biased to market fragmentation especially when we investigate informed order flow and low-latency trading.

2.2 Sample

In total, our Euronext data files span over a period of 19 months (Jan 2007-Jul 2008) and cover all stocks traded on Euronext Paris. Due to their large sizes, we make several choices regarding which portions of the overall data to include. First, we choose sample periods. As 2008 was an especially volatile period for the world stock market, we have an opportunity to construct an "implicit" test of the three algorithms by examining how they perform across different periods of market volatility. Specifically, we use April 2007, February 2008, and April 2008. These months represent periods of stable-low, stable-high, and dropping periods of volatility, respectively. This is seen clearly in the volatility and

[11] NYSE listed on-exchange trading as a percentage of overall volume is taken directly from NYSE Euronext 2nd quarter 2007 (http://www1.nyse.com/press/1185968434184.html) and 2008 (http://www1.nyse.com/press/1217498932808.html) operating data within the exchange's reported financial results.

volume graph of the CAC-250 index, or "French VIX-equivalent," contained in Figure 1. Additionally, our sample period is well-suited to testing HFT's effect on the classification algorithms. Recent literature has identified that European markets (Hendershott and Riordan (2012)) and in particular Euronext (Menkveld (2013)) has attracted HFT activity during the period that we investigate. Indeed, our sample period is toward the end of the world-wide rise in message traffic associated with the rise of algorithmic trading (Boehmer et al. (2014));

<Insert Figure 1>

Second, in order to choose our sample of stocks we focus on the 469 continuously-traded French stocks common to all three time periods and then form a random, representative sample of 100 stocks. Thirty-four of these are small-cap stocks, 33 are mid-cap, and 33 are large-cap, which we define as those companies less than €700 million, more than €700 million but less than €7 billion, and those more than €7 billion, respectively. Summary statistics shown in Table 1 indicate that the sample includes stocks of varying liquidity and volatility levels among capitalization groups. The list of included companies is available upon request.

<Insert Table 1>

Consistent with prior literature, our analysis excludes trades in the first 15 minutes of the daily trading period to avoid inclusion of opening call auctions in our continuously traded sample (Odders-White (2000)). Since Euronext stocks also utilize closing call auctions, we also exclude trades executed during the last 5 minutes of the daily trading period. Therefore, our sample includes only trades executed between 09:15:00 and 17:25:00

CEST.[12] We also impose standard trade and quote filters on the data, such as positive price, volume, and quote size, and the bid must be weakly lower than the ask (though this last requirement can be violated when we take best quotes, this is further explained in the next section).

Interestingly, our sample suggests a flight to quality effect demonstrated by Acharya and Pedersen (2005): in an unreported table we find a 23.4% average decrease in the number of intraday trades of small capitalization stocks from the April 2007 low-volatility period to the higher-volatility periods in February and April of 2008. Mid- and large-capitalization stocks experience trade count increases of 22.5% and 51.5%, respectively. Differences in mid- and large-capitalization order counts are even more pronounced.

2.3 True Trade Initiator

In order to identify whether each trade in our sample is a buy or a sell, we follow the definition of initiator used in Odders-White (2000).[13] She defines the trade initiator based on chronological order arrival, that is, the order that arrives second is the order that actually "initiates" the trade. For example if a market buy order comes in at 11:15AM and hits a limit sell order that had been standing in the book since 11:00AM, that trade would be classified as a buy for our purposes. To determine the true trade initiator in our sample, we first classify fully-executed orders into active and passive categories. An active order is executed at the same date and time as it is submitted to the marketplace, and is, essentially, a market order. In other words, the submission of an active order leads directly to a trade and the trade initiator will take the *same* buy or sell direction as that trig-

[12] According to Euronext rules, from 07:15 until 09:00, orders accumulate in the order book, at 09:00 orders in the central book are matched and an opening price is set. Stocks are then to trade continuously starting at 09:01 so we are being conservative in deleting the first 15 minutes. This process occurs at the end of the day as well, with orders accumulating in the book starting at 17:25.

[13] Ellis et al. (2000) note that the Odders-White (2000) method is preferred when a researcher has access to the order book, as we do. Current work on LR uses this classification scheme (see, for example, Chakrabarty et al. (2012)).

gering active order. A passive order is a non-market order whose execution time is always later than its submission time. In this case, the initiator of a trade will be the *opposite* buy or sell direction of a matching passive order. Active orders account for 97.59% of true trade initiator matches across our sample.

Given order and trade data, to ascertain the true trade initiator for our sample, we constructed a six stage procedure, the details of which, including marginal accuracy improvement for each step, are available upon request. Overall, untabulated results suggest that our procedure performs very well at identifying true trade initiators. Only 6.02% of intraday trades in our sample have unknown true trade initiators as compared to 25.1% of transactions examined in Odders-White (2000). Theissen (2001) uses a sample of stocks on the Frankfurt Stock Exchange for which he finds that 10.3% of trades have no true trade initiators. However, on an equal weighted true trade representation among stocks in his sample, the average percentage of trades without true trade initiators increases to 18.52%. At only 5.23% equal weighted unknown true trade initiators, our sample and true trade initiator methodology are highly accurate. Recent INET data used by Chakarabarty et al. (2013) have only 3% unknown true trade initiators. However, Table 2 shows that INET data represent 23.8% to 25.6% of total NASDAQ trade flow in 2005 (Chakrabarty, et al. (2012)). Comparatively, our 2007 and 2008 time period represents 53.8% to 82.9% of total Euronext trade flow. While the difference in percentage true trade initiators between our study and Chakarabarty et al. (2012, 2013) is small (2.23%), our superior coverage of trade flow makes our results robust and less likely to be affected by biases related to market fragmentation.

3 Methodology

3.1 LR and Tick Test Algorithms

3.1.1 Overview

The Lee and Ready (1991) trade classification process has become one of the most widely used classification algorithms in market microstructure. Lee and Ready improved upon existing algorithms by combing the quote and tick methods. In their method the quote rule is used when trades are not at the midpoint, such that any trade price above the midquote is a buy and any trade below is a sell. At the midquote, LR uses the tick rule. For the tick rule the current trade price is compared to the previous price. When the price is higher (lower) than the previous price, the trade is classified as a buy (sell). If the price of the current trade is the same as the prior trade, the closest unequal lagged price is used for comparison. We also use the tick rule as a standalone algorithm (which we will refer to as the tick test method) since it only uses trade data (level 1 data) and can be directly compared to the newly developed BVC algorithm, which also uses similar level data.

We use recent Euronext Paris data to test trade-level classification algorithms. Importantly, our paper marks one of the first tests of trade level algorithms using low-latency equities trading data. If trade level classification algorithms are unable to handle high-frequency data, then this presents a challenge for all microstructure research—and investment practitioners—going forward.

3.1.2 Implementation

Since the Euronext data do not provide sub-second timestamps, we collapse trades at the second level using volume-weighted average price (VWAP; similar to Boehmer and Kelly (2009)) when implementing trade-level LR.[14] We do this to simplify the trade flow

[14] An alternative is to aggregate trade prices at the price-second level. In an unreported analysis, we use this method and find overall LR accuracy rate of 80.42%, a slight increase. The advantage of VWAP is that it does not assume an order for same-second trades that occur at different prices, but at a cost of reduced granularity, while the price-second approach requires assuming an order for trades in some situations using the tick rule. Unfortunately, because we do not know trade ordering within a second, we cannot implement the interpolated time method that Holden and Jacobsen (2014) find to approximate ordering within a millisecond, and are forced to aggregate to the second-level.

because we cannot observe the exact ordering of the trades within a second. Although one potential criticism of our trade level algorithm implementations is this lack of millisecond timestamps, we do not view it as a major concern to our results for several reasons. First, most available databases report information at the second timestamp; those data that have millisecond time stamps are only for very recent time periods. Second, if trading continues to get faster, it is possible that soon millisecond timestamps will not be adequate, requiring microsecond stamps, and then microsecond stamps may in turn become inadequate (see Budish, Cramton, and Shim (2013) and footnote 9 above). Therefore, our conclusions are likely to remain the same regardless of our timestamps, indicating the need for a method that avoids this particular problem entirely.[15]

Relatedly, we also have to establish a prevailing quote to be in force in a given second when there are multiple quotes in that second. We treat quotes in the same second with the same bid and ask prices but different sizes (approximately 49% of sample quotes) as one quote. For any multiple-quote seconds that remain (approximately 28% of sample quotes), we take the best bid and offer (BBO) during that second to create the quote that in force in the market. In the rare cases (0.4% of sample quotes, or just less than 130,000 quotes) in which we "create" a quote that crosses (i.e., the bid is less than the offer) we just take a midpoint and set both the bid and ask equal to it.[16] This process establishes a single prevailing BBO quote for each firm-second in the sample, which allows us to sign trades using the LR classification algorithm.[17]

[15] In a low-latency environment, the BVC algorithm is an example of such a method. Our findings suggest that BVC performs well and is indeed a superior classification method.

[16] Dropping the trades matched to crossed quotes from the analysis does not significantly impact our LR transaction or volume level accuracy rates.

[17] Holden and Jacobsen (2014) interpolated method estimates millisecond timestamps for quotes marked with identical second-level timestamps. Because our data do not identify the order of quotes within a second, we use an alternative method.

Finally, trades are matched to quotes and signed according to the quote and tick rules to implement LR. These LR classified trades are then matched to sample trades for which a true trade initiator could be established (i.e., the LR classifications are matched to the disaggregated trade file). All LR results below are out of this subset of sample trades. In Table 2, to be consistent with prior studies, the accuracy of LR is calculated based on transactions, i.e., the number of trades classified correctly over the total number of trades.

3.2 The Bulk Volume Classification (BVC) Algorithm

3.2.1 Overview

The bulk volume classification procedure was developed in ELO for use in the Easley et al. (2012) volume-synchronized probability of informed trading (VPIN) calculation. It is designed to classify bars of trades (i.e., trades put in blocks either by time or volume) as a percentage of buys and sells, rather than classifying each individual trade. It was implemented this way in order to find large order imbalances, which would point to "flow toxicity." It represents an attractive alternative to traditional classification methods, particularly in situations where a researcher need only know the percentage of buys and sells in the data (rather than the direction of individual trades) or if the number of aggregate trades to be analyzed is extremely large. One notable situation is the calculation of VPIN introduced by Easley et al. (2012). By putting the trades into volume blocks, the algorithm is able to mitigate any impact from order splitting and economize on the number of data points used for classification. For instance, using time or volume bars in our analysis, the best accuracy is achieved using only 0.21% of the individual trade data points. This represents an incredible difference in computing storage resources. Whereas the tick test may take days to run, the BVC with a large, appropriately chosen block size can be implemented in a matter of minutes (even the bulk tick method, discussed later, must run the standard tick rule prior to aggregation).

3.2.2 Implementation

We apply the BVC algorithm using the Perl programming language and directly adapted from the example Python code provided by ELO.[18] First, we use aggregated trade data to the second. Bars are filled with consecutive trade seconds until the specified bar size is met or exceeded,[19] then the working bar data is stored in a MySQL database and construction of the next bar begins if additional trade second data is available.[20] Each bar record contains beginning/ending timestamps, share volume, actual buy initiated volume, volume with high frequency and other cross-sectional trade/bar level characteristics (multiple quotes, trades, or both within a second), last price in the bar, and last price in the previous bar (if applicable). Next, for each stock-month combination, we calculate the volume-weighted standard deviation of price changes between consecutive bars as shown in formula (2). With these available data points, we can then use formula (1) of ELO to calculate the BVC's buy volume for each bar:

$$\hat{V}_{i,\tau}^{Buy} = V_{i,\tau} \cdot t\left(\frac{P_{i,\tau} - P_{i,\tau-1}}{\sigma_{\Delta P_i}}, df\right)$$
$$\hat{V}_{i,\tau}^{Sell} = V_{i,\tau} - \hat{V}_{i,\tau}^{Buy} = V_{i,\tau} \cdot \left[1 - t\left(\frac{P_{i,\tau} - P_{i,\tau-1}}{\sigma_{\Delta P_i}}, df\right)\right] \tag{1}$$

[18] We thank David Easley, Marcos Lopez de Prado, and Maureen O'Hara for making this code available.

[19] Volume bar size can be exceeded if the final added trade second contains more volume than the specified bar size. The benefits of this volume bar construction methodology are further examined in Section 5. Time bars, on the contrary, can never exceed their specified bar size. Only the final volume (time) bar in a stock-month may have lower volume (duration) than the specified bar size since the stock-month may terminate before the final bar is completely filled.

[20] We create the bars continuously throughout a stock-month, meaning that if a volume bar is unfilled at the end of trading (5:25 PM) on 15 April it will continue to fill with trades from after 9:15 AM on 16 April. This does not apply to time bars, which are constructed using clock time (rather than market time) and are therefore truncated at the end of the trading day. Further, unlike Chakrabarty et al. (2013), our implementation does not use "fixed" time bar beginning and ending timestamps. That is, the "dynamic" timestamps for a time bar begin when the first trade occurs and ends with the last trade within the specified bar size. Time bar construction is discussed in greater detail in Section 5.

where $V_{i,\tau}$ is the actual volume of shares traded of stock-month i during the time or volume bar τ which is decomposed into the buy ($\hat{V}_{i,\tau}^{Buy}$) and sell ($\hat{V}_{i,\tau}^{Sell}$) volume estimate components. $\Delta P_{i,\tau} = P_{i,\tau} - P_{i,\tau-1}$ is the price change between two consecutive bars. The price associated with each bar, $P_{i,\tau}$, is the price of the last trade within that particular bar and t is simply the cumulative density function of Student's t distribution with df degrees of freedom. Following ELO, we perform our baseline analysis using df of 0.25.[21]

$$\sigma_{\Delta P_i} = \sqrt{\frac{\sum_{\tau=1}^{n} V_{i,\tau}\left(\Delta P_{i,\tau} - \overline{\Delta P_i}\right)^2}{\sum_{\tau=1}^{n} V_{i,\tau}}} \tag{2}$$

Finally, to calculate accuracy ratios for high frequency and other trade-level cross sections, actual and estimated volume sums are adjusted by proportional categorical representation within each bar. For example, in the case of a 50,000 size volume bar with half of its volume having a particular high frequency indicator, the denominator of the accuracy ratio would increase by 25,000 while the numerator would reflect scaled quantities, in this case by a half, of buy and sell values and BVC estimates. To demonstrate this further, we first calculate the volume within the cross section c that is correctly classified by the BVC for each stock-month i within each bar τ:

$$S_{i,\tau,c} = \min\left(V_{i,\tau,c}^{Buy}, \hat{V}_{i,\tau,c}^{Buy}\right) + \min\left(V_{i,\tau,c}^{Sell}, \hat{V}_{i,\tau,c}^{Sell}\right) \tag{3}$$

where $V_{i,\tau,c}^{Buy}$ and $V_{i,\tau,c}^{Sell}$ are the actual volumes scaled by the cross-section c for bar τ of stock-month i while $\hat{V}_{i,\tau,c}^{Buy}$ and $\hat{V}_{i,\tau,c}^{Sell}$ are the cross-sectional BVC estimates produced by using formula (1) with the amount of volume in the bar represented by the cross section ($V_{i,\tau,c}$) rather than the total volume in the bar ($V_{i,\tau}$). We then sum the above measure across all bars for the stock-month i and divide it by the total volume to produce a cross-sectional accuracy ratio:

[21] We also replicate our analysis using normal distribution. Like ELO, we find that Student's t distribution with 0.25 degrees of freedom offers a substantial improvement in performance. See Table 5 for a direct comparison.

$$AR_{i,c} = \frac{\sum_{\tau=1}^{n_i} S_{i,\tau,c}}{\sum_{\tau=1}^{n_i} V_{i,\tau,c}} = \frac{\sum_{\tau=1}^{n_i} S_{i,\tau,c}}{\sum_{\tau=1}^{n_i} \left(V_{i,\tau,c}^{Buy} + V_{i,\tau,c}^{Sell}\right)} \tag{4}$$

where n_i is the number of applicable bars in the stock-month i.[22] The above measure can be aggregated at the cross-section by using:

$$AR_c = \frac{\sum_{i=1}^{k} \sum_{\tau=1}^{n_i} S_{i,\tau,c}}{\sum_{i=1}^{k} \sum_{\tau=1}^{n_i} V_{i,\tau,c}} = \frac{\sum_{i=1}^{k} \sum_{\tau=1}^{n_i} S_{i,\tau,c}}{\sum_{i=1}^{k} \sum_{\tau=1}^{n_i} \left(V_{i,\tau,c}^{Buy} + V_{i,\tau,c}^{Sell}\right)} \tag{5}$$

where k is the number of stock-months to be analyzed. This accuracy ratio calculation is the cross-sectional equivalent to that of equation 3 in Chakrabarty et al. (2013).

4 Performance Results

4.1 LR and Tick Test Algorithms

Table 2 displays the results for the LR algorithm on the trades for which we have a true trade initiator, as defined above. The overall trade level accuracy of the LR algorithm in our data is 78.67%. The accuracy is lower than that of other studies: Odders-White (2000) reports an overall accuracy near 85%, while Peterson and Sirri (2003) report close to 90%. The overall drop in accuracy of the LR classification is likely due to the much higher frequency of trades and quotes present in our sample, both intra-second and overall. For example, Peterson and Sirri (2003) use NYSE Superdot system trades for twenty trading days in 1997, from which they get over 3.3 million trades, roughly 70 trades per company per trading day.[23] In contrast, our final sample of 100 Euronext companies over three months contains over 11.7 million trades, which is an average of approximately 1,900 trades per company per trading day. Odders-White shows a decrease in the accuracy of LR for high volume firms (those with little time between transactions and

[22] There are $n_i + 1$ *total* bars in each stock-month. Since the BVC uses price differences to estimate buy and sell volume, the first bar in each stock-month is not included in these accuracy ratio calculations since it has no prior bar.

[23] The average number of trades per firm per day is calculated assuming that Peterson and Sirri retained 90% of US companies listed on the NYSE in 1997. The number of US listings is taken from http://www.nyxdata.com/factbook.

those with a large number of transactions in the sample) and given that our sample contains more of such firms, we are likely to see a dip in LR accuracy. The performance of LR in these high frequency settings is the further focus of subsequent tests below.

<Insert Table 2>

First, in Panel A, the accuracy of our LR results based on trade position in the spread is similar to other studies (Odders-White (2000) and Peterson and Sirri (2003)) in that it is most accurate at the bid or ask and less accurate outside or inside the spread. However, our overall results inside the spread are different in that, when inside the spread, the tick rule component of LR is superior to the quote method (unreported). Interestingly, correct classification of trades at the ask is 90.43%, whereas classification for trades at the bid is 81.24%.

We next investigate the performance of LR in various subsamples, including some that are more likely to include HFT. Panel B displays the overall results from running LR separately for each sample month. This is a test of the effect of volatility in the market as a whole and its effect on the accuracy of the algorithm. There is only a small difference between the accuracy rates: 80.54% for April 2007, with a dip to 77.74% in February 2008, the most volatile month. Our expectation was that market volatility would create an environment in which many trades would be executed in short windows of time as traders tried to prevent price-swings from destroying portfolio value. Further, we assume that volatile periods are periods in which the market has an inability to establish an intrinsic value for an asset (i.e., a midpoint), and quote swings would hamper the LR algorithm. Overall, these results suggest that market volatility has a small effect on the accuracy rates. Therefore, we do not split the next subsamples by month.

Panel C creates subsamples based on whether or not there is more than one concurrent trade in a second. For the whole sample the accuracy for "lone" trades is 85.61%, while the corresponding number for seconds with multiple trades is only 76.54%. This is a meaningful drop in accuracy, translating to over 600,000 misclassified transactions (the vast majority of trades cluster in seconds).

In Panel D, we test whether quote frequency impacts the performance of the LR algorithm. If trades are happening in milliseconds and quotes are being correspondingly adjusted, because we only view the data at the second level, we cannot assign trades to the "right" quotes. If taking best quotes creates a representative quote that correctly signs the trades, we should see no change in accuracy. However, in the table we see that if a trade was executed in a second with more than one quote, the accuracy of LR drops by 17.33%, which corresponds to over 800,000 additionally misclassified trades relative to single quote accuracy. Further, we see that again this result is due in large part to lower accuracy on the sell side. The accuracy differential for buys and sells is 11.06% and 23.45%, respectively. This large decrease in multi-quote accuracy supports the theoretical model of Baruch and Glosten (2013) in which quotes are frequently updated at random rather than with information (i.e., flickering), which should lead to lower LR accuracy. The result demonstrates the possibility that high frequency traders can overwhelm LR, if their rapid order submission and cancellation (see Figure 2 in Hasbrouck and Saar (2013)) in part generates the rapid quote updating.

In Panel E we look at the accuracy rates for trades that occur in the highest (those seconds with many trades and many quotes) and lowest (those seconds with only a single trade and quote) frequency settings. There is a 20.93% difference in classification accuracy, corresponding to nearly 800,000 misclassified transactions.

In Panel F, we investigate whether quote swings lowered our LR accuracy. We use a measure of intraday volatility to classify trades as high- or low-volatility. Insofar as the midpoint in the quotes represents the market's (there is no market maker on Euronext) best estimate of the "true" value of an asset, the variation in the midquote throughout the day would be informative as to the underlying volatility of the asset, at least at a given time on a given day. If this creates more quote movement and more trading activity, both of these would contribute to a lower rate of accuracy for LR. For our estimate of volatility, we use the hourly standard deviation of $\ln(midquote_t) - \ln(midquote_{t-1})$ for each firm. If an hour was above the median volatility of the firm for that day, then all trades during that hour are classified as volatile. Surprisingly, there is only a 2.07% difference between the accuracy of trades classified during high and low periods of intraday volatility.

In Panel G we investigate whether firm size categories produce differences in the accuracy of LR. We use our previous definitions for large-, mid-, and small-capitalization stocks. Brogaard et al. (2014) find that in their sample the vast majority of HFT trading volume occurs in large capitalization firms. Therefore, we expect that more trading volume, and HFT volume specifically, will decrease LR accuracy in large stocks. Consistent with this, the accuracy of LR is decreasing in market capitalization and the difference between large- and small-cap accuracy rates is 4.1%.

Volume-weighted accuracy results for both LR and the tick test algorithm are presented in the first two rows of Table 3.[24] There are slight declines in accuracy during periods of increasing volatility for both algorithms; they perform best during the April 2007 period of low and stable volatility. Large capitalization stocks see slightly larger declines for the tick test (6.98%) than LR (4.80%) relative to small cap stocks. Most notable are those cross sections in which multiple trades or quotes are present in a single second, where we see that the accuracy of the LR algorithm degrades to 75.41% and 67.20%, respectively. When both trades and quotes are present intra-second, the accuracy of LR falls to

[24] Since the BVC is designed to sign percentages of bar volume and not individual trades, for most of our analysts we compare the BVC performance to volume-weighted LR and tick test results (as in Table 3).

66.55%. The volume weighted consistency of the simple tick test in high frequency data, which actually improves slightly from 78.97% to 79.70% during seconds of multiple trades, emphasizes the inferior performance of LR, in terms of both accuracy and data efficiency, in high frequency data.

Overall, the evidence in Tables 2 and 3 suggests that, at least without millisecond data, LR is not well-equipped to handle any recent data with a high volume of trades and that likely contain high frequency trading. Further, despite the use of trade and quote data, the LR is generally less accurate than the tick test. Odders-White (2000) investigates the consequences of misclassification, and shows that it overestimates the cost of trading and produces a trading anomaly around earnings announcements. Similarly, Boehmer et al. (2007) find that misclassifications lead to downward-biased probability of informed trading (PIN) estimates. Therefore, in undertaking any microstructure study that necessitates signing individual trades, researchers should proceed with caution when using LR in a high frequency setting as the potential for misclassification is large.

4.2 BVC Algorithm

4.2.1 BVC Performance versus LR and Tick Test Algorithms

In this section we test the BVC algorithm in our equities data. Given its design, our expectation is that it will outperform trade level algorithms in our data, particularly in the cross-sections that are aimed at finding high frequency data. Table 3 contains these results from the BVC algorithm using Student's t-distribution with 0.25 degrees of freedom. The bold type in the body of the table designates accuracy rates that are greater than the tick test's rate and the boxed rates represent the peaks in the BVC accuracy rates (over the range of bars we test).

<Insert Table 3>

The first important note is that the accuracy rates are lower than those reported in ELO; their accuracy rates top 98% versus 90% in our analysis. We expect this lower accuracy in our equities sample because of greater heterogeneity in stock trading characteristics as well as the general disparity in price impact between blocks of buys and sells in equities market that does not exist in the futures market ("more informative buys than sells") documented by Chan and Lakonishok (1993) and Berkman, Brailsford, and Frino (2005). This difference in price impact would serve to lower the overall accuracy rates of the BVC because the algorithm uses symmetric distributions (the t and z) to estimate buy volume in a bar. That is, the algorithm assumes that a given price movement in either direction was caused by an equal number of trades (buys for a price increase, sells for a price decrease). For example, consider a 50,000 share volume bar composed of a buy and sell, both of 25,000 shares. Asymmetric price impact suggests there will be a price change, which means the BVC will not classify buy volume as 50%, and will thus be expected to have lower accuracy.

Despite this challenge, the BVC performs surprisingly well in equities. In Panel A of Table 3, the overall accuracy ranges from 62.63% to 90.58% with volume bars of between 1,000 and 800,000 shares, respectively. The overall accuracy peaks at 90.58% with a bar size of 600,000, compared to 78.97% trade level accuracy for the tick test and 77.24% for the LR algorithm. The optimal volume bar size BVC also outperforms the tick test in all of the subsamples displayed, including different months (changes in volatility), and between purchases and sales (unreported). Importantly, the accuracy rates of the volume bar BVC are consistent on the high frequency cross sections (multiple trades and quotes), suggesting that the BVC does indeed perform extremely well in high-frequency data.

In unreported results, we also calculate accuracy ratios by time of day. While these are largely consistent, there is a small decline in the final hour of the day (the 25 minutes after 5 PM) that is likely due to market closure. In our BVC implementation, volume bars of larger sizes may extend from one trading day into the next, which again could result in lower accuracy. Finally, it is important to note that throughout Panel A accuracy ratios for the volume bar BVC are not monotonically increasing with bar size. This suggests that the choice of bar size is important for the BVC to perform optimally.

Results for the BVC using time bars are presented in Panel B of Table 3. The overall results slightly surpass those for volume bars, peaking at 90.90% at the 18,000 second bar size. Largely, results are mirrored between time and volume bars, including the results across the multiple intra-second high frequency cross sections. Unreported time of day results also exhibit a drop off in the 5 PM cross section which is likely driven by both the time bar truncation at market close and the market closure itself. Similar to Panel A, Panel B does not exhibit a monotonic relationship between BVC accuracy ratios and time bar size. In terms of market capitalization, however, medium and large capitalization stocks seem to outperform small stocks in time bars. Accuracy ratios for large capitalization stocks can exceed those for small stocks by as much as 13.5% in some bars, though at their peaks, large stocks outperform small by only 2.32%. This lower accuracy is likely due to the presence of unequally spaced time bars,[25] which contributes to a price "staleness" effect. Additionally, lower volume per time bar can reduce the netting benefit of the BVC. We explore both of these issues in a later subsection. We also must consider

[25] This a variation of what Chakrabarty et al. (2013) refer to as "zero volume" time bars in a static implementation.

the possibility that the time bar based BVC is simply not appropriate for less liquid, small capitalization stocks for reasons related to data efficiency.[26]

The final column in each piece of Table 3 indicates by how much the BVC can reduce the size of the input data set, that is, how data efficient it is relative to the standard tick rule. The data efficiency advantage of using the BVC is simply because when trades are grouped into bars there are far fewer data points to process when running the algorithm. Comparing only to the tick test data points, a researcher could move from processing over 5.7 million records down to under 10,000. The volume (time) bars displayed in Table 3 represent data compression ranging from 68.36% to 99.92% (19.05% to 99.81%). Running LR requires quote data which, even considering only best quotes (that is, after filtering invalid quotes and creating a best quote for every second), requires the handling of almost 19 million quote observations, vastly increasing the rate of compression the BVC offers. This has enormous benefits with respect to data efficiency for research applications.[27] This is a clear advantage in working with the BVC if the researcher does not need to sign each individual trade. Further, the algorithm suffers no drop in accuracy when working through high frequency data. Clearly, the potential ability of the BVC to drastically increase data efficiency without sacrificing accuracy compared to the tick rule or other methods makes it a very promising development in the literature.

4.2.2 BVC Results versus Bulk LR and Bulk Tick Test Algorithms

[26] In unreported results, we find that small capitalization stock time bars of 60 seconds have compression ratios under 50%. Compression ratios for small capitalization stocks are lower than large (mid) capitalization stocks by up 42.8% (22.1%). This result suggests using volume bars over time bars for small, illiquid stocks since volume bars see little difference in compression across market capitalization groups. In contrast to time bars, the largest difference in volume bar compression ratios between small and large (mid) capitalization stocks is 0.07% (0.04%).

[27] Chakrabarty et al. (2013) find that algorithm CPU time between the tick test and BVC are comparable. However, the data efficiency advantage of the BVC has benefits in practical implementation. For example, several market data providers can transmit bar, as opposed to trade, level data. This will reduce network bandwidth utilization. Similarly, necessary algorithmic back-testing of the BVC requires less archival data storage capacity.

While the results of Table 3 suggest that the BVC outperforms both the LR and tick test algorithms, Chakrabarty et al. (2013) argue that the netting of misclassified buy and sell trades within bars drives the BVC outperformance. Aggregation of trades into bars changes the goal of an algorithm from individual trade signing to identifying the proportion of buys and sells in a given bar (order imbalance). With netting benefit, incorrectly identified individual trades can be offset within a given bar leading to a more accurate order imbalance. As a result, the authors recommend comparing the BVC with the bulk tick test which aggregates the tick test classifications into proportion of buys and sells in a given bar size (we also include a bulk LR in our analysis). We repeat this method of comparison and present the results in Table 4. Trades are first signed individually using the tick test and LR algorithms and are then placed into time and volume bars. We calculate accuracy ratios using formulas (4) and (5). Formula (3), rather than using the BVC estimate, will use the volume classified as buys or sells using either the tick test or LR.

Panel A of Table 4 shows that accuracy ratios for the bulk tick test increase monotonically with volume bar size, reaching levels over 95% within each cross section starting at a bar size of 400,000. Similar to the trade level tick test results in Table 3, the volume bar bulk tick test performs best for small capitalization stocks. It also performs consistently across all of the tested cross sections.

<Insert Table 4>

Bulk tick test results for time bars are presented in Panel B of Table 4. In a similar manner to Panel A, accuracy ratios for time bars increase monotonically to values above 90% in each cross section. Bulk tick accuracy ratios for small capitalization stocks, however, can lag large caps by up to 3.2%. This small capitalization stock result is analogous to that in Panel B of Table 3 as lower volume per time bar is likely reducing the netting benefit. In the first rows of Panels A and B of Table 4, we also include the highest accuracy ratio for bulk LR, which is at the largest bar sizes (800,000 and 25,200 for volume and time bars, respectively). This, like the bulk tick test, is due to a monotonically increasing relationship between bar size and LR accuracy ratio. The bulk tick test almost uniformly outperforms the bulk LR, with an overall advantage in volume (time) bars of 0.46% (0.85%). However, these results suggest that the choice of the trade level algorithm is not nearly as important a contributor to the greater bulk accuracy ratios as the netting process itself.

<Insert Table 5>

To further examine the difference in accuracy ratios between the bulk tick test and the BVC, we compare our results using Euronext data to those based on INET data from Table 1 of Chakrabarty et al. (2013). The results for volume bars are presented in Panel A of Table 5. For the same volume bars and BVC definition,[28] our overall results exhibit lower differences and less variation than the 2005 and 2011 Chakrabarty findings. For our Euronext sample, volume bar differences range from 8.28% to 13.24% compared to 9.7% to 15.7% for the INET results. At volume bar sizes of 5,000 or more, the advantage of the bulk tick test in the Euronext data is lower than any of those in the INET results. Using ELO recommended Student's t distribution, rather than normal, BVC yields substantial improvement with increased bar size (for volume bars of 100,000 shares the difference in accuracy between the BVC and bulk tick-test is reduced by 12.02%). Panel B shows similar results for time bars. The accuracy advantage of the bulk tick test in our Euronext results is much greater for very small time bars. However, for time bar sizes greater of 300 seconds or more, the accuracy ratio difference falls below 6% and is lower than for any of the INET time bar results. While lower market fragmentation in Euron-

[28] Chakrabarty, et al. (2013) only provide BVC results using the normal distribution.

ext versus INET is likely contributing to these differences, the variation in our results also suggests that “not all bar sizes are the same.” This is particularly true for the BVC since, unlike the bulk tick test and LR, its relationship between the specified bar size and accuracy ratio is non-monotonic. We investigate this in more detail throughout Section 5 below.

5 Calibration and Refinements to the BVC

Chakrabarty et al. (2013) state the following: “clearly, BVC and TR [bulk tick test] offer a tradeoff between accuracy and computational efficiency when applied to equities. We believe that researchers should be aware of this tradeoff.” Our results to this point echo that note. However, it is not unreasonable to assume that the accuracy of the BVC can be improved, which, along with its insurmountable data efficiency, makes a simple tradeoff between the two methods far less clear. For example, one particularly effective improvement, demonstrated in the first two columns of Table 5, is the replacement of the normal distribution by the Student’s t-distribution in the BVC calculation. We subsequently examine several “enhancements” to the BVC in this section: (1) choice of bar size and price change t-distribution degrees of freedom (df) parameter (2) time spacing and weighting considerations, and (3) exact versus minimum volume bar sizes.

5.1 Netting and Bar Size/Distribution Calibration

The choice of bar size impacts the BVC and the bulk tick test in very different ways. In the latter, the tick test is performed on individual trades which are then aggregated into volume or time bars. The aggregation step will serve to have misclassified trades net out thereby increasing the accuracy in the bar. Indeed, this is the likely reason why the accuracy results for the bulk tick test exhibit a monotonically increasing relationship with bar size, consistent with Chakrabarty at al. (2013) using INET data. On the other hand, the relationship of BVC accuracy with bar size is less clear: the choice of bar size influences the distribution of price changes across bars. In other words, as bar size changes so do both the numerator (price changes) and denominator (weighted standard deviation of price changes) of the parameter for the CDF in formula (1). For this reason, it is more appropriate to compare BVC and bulk tick results after controlling for the distribution of bar price changes per stock. This will determine the appropriate bar size for running the BVC algorithm and test its accuracy. As an example, we present BVC accuracy ratio results that are ex-post calibrated for bar size and df parameter at the stock-sample level in the second row of Table 4. We also present the differences of these calibrated BVC results with the bulk tick test at various bar sizes in Table 5 (the last two columns). As Panel A of Table 5 shows, when the BVC is ex-post calibrated to select the best volume bar size for each security, the accuracy discrepancy between the algorithms shrinks; the bulk tick test outperforms the calibrated BVC by only 3.93% to 5.39% for volume bar sizes of 100,000 through 500,000.[29] Calibration of the BVC Student's t-distribution in the final column of Table 5 shrinks these differences further to between 1.72% and 2.51%. This additional distribution parameter calibration, motivated by the return distribution analysis of Bakshi, Kapadia, and Madan (2003), lowers degrees of freedom from 0.25 to 0.05 (0.1) for large (mid) cap stocks.[30] Results appear similar in Panel B when we calibrate BVC with respect to time bars of 7,200 seconds and larger. Overall, our findings suggest that when the bar size of the BVC is properly calibrated for the underlying trading instrument, the accuracy advantage of the bulk tick test can decline substantially.

In order to further investigate the effect of bar size choice on the BVC accuracy, we consider scenarios in which bar size can either be "too small" or "too large," given the dis-

[29] The majority of stock months are calibrated at volume (time) bar sizes of 100,000 shares (7,200 seconds) or larger.

[30] Bakshi et al. (2003) identify that the returns distribution kurtosis across stocks decreases with market capitalization. Thus, to account for these fatter tails, we follow their return distribution analysis by reducing the degrees of freedom in the large and mid-cap group Student's t-distributions to 0.05 and 0.1 respectively.

tribution of trade sizes. In particular, with respect to time bars, if the bar size is too small, the bar will not contain enough trades to benefit from netting misclassified trades. This reduced netting will impact both bulk tick/LR and BVC algorithms. At the same time, however, a greater number of smaller bars, , will lead to a smaller standard deviation of price changes which will only impact the BVC and *not* bulk trade level algorithms. To better see this effect, we present the mean volume within time bars in Table 6.

<Insert Table 6>

For the smallest time bar size (2 seconds), the average bar volume is no more than 200 shares greater than the average trade second. Not surprisingly, for small capitalization stocks average bar volume does not increase with greater time bar size as much as for large capitalization stocks. For example, mean volume in two second bars is similar for small and large capitalization stocks, 688 and 789 shares, respectively. However, in a 20 minute bar mean volume is 4,303 and 62,283 shares for small and large capitalization stocks, respectively. Because of this small increase in mean volume, for small cap stocks there is a greater proportion of bars that will have very small price changes. As a result, we will see bars that are more evenly weighted between buys and sells using BVC. Indeed, we believe that this is why average bar volume which is close to average trade size will lead to substantially lower results for the BVC versus the bulk tick test.[31] The small bar results (fewer than 5,000 shares or 10 seconds) of close to 50% for both volume and time bars in the first few rows in Panels A and B of Table 3 support this conclusion.

If bars are "too large" and thus consecutive bars are too far apart in time, the price of the previous bar can become "stale" and much of the useful price variation within the current bar can be lost. This will affect the BVC more than the bulk tick-test since a

[31] The rationale for volume bars is similar but we do not report average volume for volume bars since the actual volume (using minimum rather than truncated volume bar sizes) exhibits far less cross-sectional variation.

fundamental difference between the two algorithms is the duration of time that elapses between consecutive data points. In the case of the tick test, this time period is determined only by the arrival distribution of trades. Duration between time points for the BVC, however, is influenced by both the distribution of trades and the choice of bar size. Slower trade arrival will result in longer time periods between volume bars since they will take more time to fill completely. Holding trade arrival constant, increasing volume bar size will have a similar effect. For time bars, since the duration of the bar is predefined, this temporal effect is driven by the choice of bar size as well as the presence of empty (zero volume) time bars which increase the temporal distance between bars that generate price differences. Since less liquid stocks see longer and more frequent periods of trading inactivity, they are more likely to have time bars that are unequally spaced and/contain very low volume. This increases the duration between consecutive bars, providing a possible explanation for the lower BVC accuracy of small capitalization stocks in Panel B of Table 3.

To assess the potential impact of "staleness" in BVC implementation, we present the time elapsed between trades and volume/time bars for our sample in Table 7. In Panel A, time elapsed between consecutive volume bars is much longer for small capitalization stocks, due in part to lower liquidity and trading frequency. For a volume bar size of 50,000, small capitalization stocks see an average of 17,902 seconds between consecutive bars. Considering that there are 29,400 seconds within each trading day in our sample, this represents only 1.64 volume bars per day. A standard deviation for 50,000 bar size time differential of 44,933 seconds suggests that some small cap stocks may have as few as two volume bars in a given week. This reemphasizes the need for appropriate choice of bar size.

<Insert Table 7>

Time bar results are reported in Panel B. Unsurprisingly, the difference between time bar size and elapsed time between consecutive time bars is lower for more liquid, large capitalizations tocks. In fact, for time bars of size 3,600 and greater, the mean distance between consecutive non-zero time bars is less than the specified bar size, a result that reflects both the presence of truncated time bars at the end of the trading day and otherwise equally spaced time bars.[32] Small and medium capitalization stocks exhibit a greater likelihood of "staleness" as differences in time between consecutive bars are generally larger than the specified time bar size.

Addressing the potentially negative impact of "staleness" on the BVC's accuracy requires a separate investigation, though a prime consideration is the length of time to include price changes in the volume weighted standard deviation ($\sigma_{\Delta P_i}$ in formula (2)) calculation. In this study, we calculate each $\sigma_{\Delta P_i}$ at the stock-month level. The key question here is: how far back in the price change history should one go? Given that $\sigma_{\Delta P_i}$ is volume weighted, it does not seem reasonable that an extremely large trade 11 months before a bar to be signed should have the same impact on sign as a similar large trade 11 hours prior. A simple way to prevent this is to impose a limit, likely based on the trade distribution of the stock in question, on the time window of the data points used to calculate $\sigma_{\Delta P_i}$. Large, more liquid stocks would have shorter windows (e.g., weekly or monthly), while small, less liquid stocks would require longer windows (e.g., quarterly). To more rigorously approach this issue, one could use multi-dimensional weighted standard deviation of price changes, weighting on both time elapsed and volume. We leave the development of such an approach to future research.

5.2 Bias in Truncated versus Minimum Volume Bar Sizes

[32] Note that the elapsed time in the body of the table refers to market time rather than clock time. Clock time is used to mark the beginning and end of each time bar.

A potential source of bias in the volume bar BVC applied to equities arises not from the choice of bar size but how that bar size is applied to the data. The BVC proposed in ELO does not specify whether volume bars should contain the exact volume as the specified volume bar size or if the specified size is a minimum amount of volume that each bar should contain.[33] In the former case, if the last trade in the bar causes the volume in the bar to be greater than the specified size then the trade will be truncated and the remainder applied to the next bar. Figure 2 shows how this application of the BVC can cause a systematic bias. Suppose that $Trade_L$ is a large true buy trade executed at a price higher than that of both the previous trade and volume bar ($P_1 > P_0$). First, note that the bulk trade level algorithms do not suffer from the bias. The tick test will correctly classify the trade as a buy since the price of $Trade_L$ is greater than the previous trade price. Whether the volume of this trade is then inserted within one volume bar or many, each part of the trade will be correctly signed.

<Insert Figure 2>

Illustration (a) of Figure 2 shows how such a trade can decrease the accuracy of the volume bar BVC. If the remainder of $Trade_L$ not in bar 1 exceeds the specified volume bar size then both bar 1 and 2 in (a) use the price of $Trade_L$, resulting in a price change of zero for bar 2. The volume within bar 2 will then be incorrectly classified as only half buy volume. The signing of the remaining trade volume that is pushed into bar 3 will depend upon P_3, which is unknown. Illustration (b) shows the BVC allowing for minimum volume bar sizes (i.e., no trade truncation) in which the entire volume of $Trade_L$ is contained in bar 1. If we assume that P_1 is substantially greater than P_0 relative to the vol-

[33] This bias is likely to be much lower for futures than equities since futures exhibit far less variation in trade size. For example, Figure 1 of ELO shows that more than half of E-mini S&P 500 futures trades are of size one. The proportion of single contract trades for WTI Crude Oil and Gold futures in Figure 3 is even higher. In contrast, no single trade size in our Euronext data accounts for more than 3.7% of trades (100 share trade) and 84.41% of trades include trade sizes of 1,000 shares or less.

ume weighted standard deviation of volume bar price changes, then then the majority of $Trade_L$ should be correctly classified as buy volume, reducing the bias in (a).

<Insert Table 8>

Using flexible minimum, rather than truncated, bar sizes allows large bar ending trades, which are less likely to have been initiated by split orders, to contribute more volume to a given bar. This form of volume bar aggregation more effectively mitigates the effects of order splitting. To better quantify this bias, we estimate the potential bias that may have appeared in our results if we had used truncated, rather than minimum, volume bars in Table 8. This bias is estimated by summing the volume in each volume bar over the specified size ("excess" volume), multiplying the sum by one half, and dividing the product by the total volume in the respective cross section. The bias can be as large as 25% for volume bar sizes as small as 1,000 shares and monotonically declines with volume bar size. Not surprisingly, the bias is larger when average trade size is large relative to volume bar size. Therefore, smaller average trade size, either from a market with more order splitting or from millisecond data, may well result in smaller bias. These results suggest that minimum volume bar sizes should be used rather than forcing all volume bars to contain the same amount of trade volume. Additionally, since these large trades are more likely to convey underlying information (Easley and O'Hara (1987)), truncating volume bars will reduce the efficacy of the BVC at detecting that information.

Overall, we reemphasize that the BVC is sensitive to the choice of bar size. Researchers should choose bar size, price change distribution, and additional construction methodologies (e.g., flexible minimum volume bars, parameter estimation window length, etc.) keeping in mind a given security's liquidity and trading frequency, as this influences how quickly bars fill (price "staleness") as well as the bias from using truncated volume bars.

6 Can the Algorithms Detect Informed Order Flow?

Although accuracy in classifying the aggressor side of trades and trade bars is important, which we have focused on thus far, arguably as or more important is the ability of these algorithms to detect underlying information. Informed traders are increasingly relying on passive orders, i.e., limit orders (which can be exacerbated by order splitting), to disguise themselves in the market (see, for example, Bouchaud et al. (2009); Zhang (2013) estimates the probability of informed liquidity provision to be 85% post-decimalization). Therefore it is not necessarily true that correctly classifying trade aggressors captures informed order flow. In order to determine how well the BVC and bulk tick algorithms capture this underlying information, we replicate the regression used in ELO,

$$HL_\tau = \alpha_0 + \alpha_1[HL_{\tau-1}] + \gamma\left|\widehat{OI}_\tau\right| + \varepsilon_\tau, \tag{6}$$

where the estimated order imbalance is

$$\left|\widehat{OI}_\tau\right| = \left|\frac{\hat{V}_\tau^B - \hat{V}_\tau^S}{V_\tau}\right| = \left|2\frac{\hat{V}_\tau^B}{V_\tau} - 1\right|. \tag{7}$$

ELO argue that if an algorithm is capturing the underlying information, then the absolute value of the estimated order imbalance (from either the BVC or bulk tick) should be positively related to the high-low trading range over bar τ. This is because if informed traders are driving an order imbalance then market makers should widen spreads to protect themselves, increasing the difference between high and low prices in the bar. Table 9 contains the results of these regressions in our Euronext data.

<Insert Table 9>

Panel A contains the regression results for the BVC order imbalance for both volume and time bars. The coefficient of interest is the order imbalance coefficient γ (gamma). Across all regressions for different bar-sizes the γ coefficients are positive and significant. In addition, coefficients and model R-squared are increasing with bar size. The positive coefficients suggest that when the high-low range is increasing in a bar representing market makers widening the spreads, the BVC estimated order imbalance is also increasing. These results are consistent with the notion that the BVC captures the underlying intentions of informed traders.

Panel B contains the regressions for order imbalance estimated by the bulk tick. In the volume bars, there are only two positive and significant γ coefficients, two insignificant coefficients, and the rest are significantly negative. In the time bars every γ coefficient is negative and significant. These results suggest that when the high-low trading range is narrower the bulk tick rule estimates a larger order imbalance. If the high-low range is a reasonable proxy for price impact or underlying trading intentions, this is opposite of what the theory would predict.

In Table 10, we focus on subsamples with small and large absolute returns. We expect the large absolute return subsamples, which we define as bars with returns in the first or fourth quartile of non-absolute bar returns, to contain more informed trading due to the large price movements. In Panel A of Table 10, we find that the BVC OI estimate relates much more strongly to the high-low range for the large absolute return subsample (0.369 mean effect versus 0.031). This, however, is expected by construction, because the BVC uses returns to identify the percentage of buys and sells. But it does reinforce the need to define bars with adequate price movement for each security. In Panel B, we find that the OI estimated by the bulk tick performs worse (is more negative) for the large absolute return sample. The mean effect sizes across bars are -0.051 and -0.033 for large and small returns, respectively, and again both are highly statistically significant.

<Insert Table 10>

To test whether it is the presence of informed, passive orders that render the tick test unable to detect underlying information, we run the high-low regressions in subsamples that are likely to contain informed trading using limit orders. Baruch et al. (2013) find that when borrowing costs are high for investors (likely when firms are not index members or there is no options market) informed traders tend to use passive orders more. This is because when informed traders cannot sell short, competition among them is reduced creating less aggressive trading. If simply correctly classifying the aggressor side of a trade is no longer sufficient in detecting the underlying trading information, then the tick test's inverse relation between estimated order imbalance and the high-low trading range should be exacerbated here. In Table 11, we run the regressions separately for firms that are and are not members of the SBF-120 stock index and for stocks that do and do not have an active options market. We define active options market as *any* options volume in Bloomberg for that stock month. In Panel A, we report γ coefficients where the BVC OI is used as a right-hand side variable whereas in Panel B we report similar statistics for the bulk tick test OI as a right-hand side variable. The γ coefficients are positive for all regressions in Panel A. In Panel B, we find that all γ coefficients estimated using the tick test are again negative and significant. However, for firms that have no options markets or are not members of the SBF-120, the estimated γ are more strongly negative than those for firms in an index or with an options market. Specifically, the mean effect sizes across bars for stock-months (not) in the index and with (without) an options market are (-0.258) -0.042 and -0.041 (-0.071), respectively, all of which are statistically significant.[34] That is, for firm-months likely to contain informed limit trading, the tick test implies an even narrower trading range for a given estimated order imbalance. The γ estimated parameters when BVC is used show no large differences between stock months with and without an index, or with and without an options market. Overall, the results of Table 11 show the superiority in the BVC algorithm over tick test in capturing informed order

[34] In Tables 10 and 11, we estimate the mean effect sizes by weighting each regression coefficient by the reciprocal of the squared standard error. We estimate cross-bar size dependence using the ratio of the true to estimated standard errors found in Chordia et al. (2000, 2005) ($[1 + 2(N-1)\rho]^{1/2}$).

flow. When informed traders trade passively, the bulk tick test's inability to capture informed order flow activity appears stronger.

<Insert Table 11>

Overall, the results in the last three tables suggest that while the BVC does well at detecting underlying information through order imbalance, the tick test does not. This is exacerbated in subsamples likely to contain informed trading (large returns) and subsamples in which informed traders are likely to use passive orders. Correctly classifying the aggressor side of the trade no longer adequately captures informative buying and selling activity and thus renders the tick test ineffective in detecting informed order flow.

6.1 Accuracy Classification Accuracy and Information Detection

In Section 5 we showed that the accuracy of the BVC improves dramatically when we control for issues related bar size. Do the improvements of the BVC accuracy related to bar size affect information detection? To formally investigate this, we select the bar sizes for each stock-month for which the BVC was most accurate, and re-run the OI regressions using only these bars for both the OI estimated by the BVC and the bulk tick. If accuracy is important in capturing underlying information, we expect that the gap between the BVC and tick OI coefficients will be wider in these regressions than in previous.

<Insert Table 12>

Table 12 displays the results of these regressions for volume bars. Models 1 and 3 contain no fixed effects, while models 2 and 4 contain stock, bar size, and month fixed effects. Comparing the models with fixed effects, the coefficient on the BVC-estimated OI is 1.967 while that on the tick-estimated OI is -0.125. This result indicates that the most accurate bars are the ones in which OI is most strongly related to the high-low trading range for the BVC, and it greatly outperforms the bulk tick with these bars. This suggests that for the BVC, accuracy is important in detecting underlying information, i.e., whether or not you care about information, maximizing accuracy through calibrating the BVC is important.[35]

7 Conclusion

The ability to correctly identify a trade initiator in data which do not contain orders or even quotes is a critical part of much of the market microstructure literature. The most common process for researchers to use is the Lee and Ready trade classification algorithm if quotes are available or the tick test when they are not. In our sample of recent Euronext trades (from April 2007 and 2008 and February 2008) for which we can assign a true trade initiator, we find that low-latency trading has a detrimental impact on the accuracy of the LR algorithm. The decrease in accuracy can be substantial: for seconds with multiple trades and quotes versus seconds with a single trade and quote, LR accuracy drops 20.97 percentage points (a 23.6% drop from the single trade and quote rate). This is an important contribution as the volume that high-frequency traders are responsible for is large (Brogaard (2011) estimates that high-frequency traders are involved in 68.5% of the dollar volume in his sample; using our simple definition of high-frequency trading, 77% of the trades in the sample are high-frequency). Therefore, our result indicates the caution that will be needed for any future research on trade data that uses the LR algorithm for signing individual trades.

The uniqueness of our data plays an important part in the contribution of our paper. European markets have been slower to fragment than the U.S. markets; and therefore a large scale test of the LR, tick, and BVC algorithm is much more feasible than attempting to aggregate all trades for a given listing across many exchanges that are executing them. This first look into BVC's performance in equities is important because of the disparate

[35] In an unreported test, we find that the BVC-OI that uses the normal distribution does a poorer job at detecting information than the BVC using the t distribution. This further suggests that accuracy matters. Additionally, if we use the most accurate bulk tick bars, its OI estimate still does not positively relate to the high-low range.

price impact of buys and sells in equities that does not affect the futures market. Our study is thus a valuable addition to existing literature on signing trades and establishes that going forward the LR algorithm is not appropriate for use in high-frequency data.

However, if the research question does not necessitate signing each trade we find that the BVC can be used without any loss in accuracy due to high-frequency trading. The BVC outperforms the trade-level LR and tick test algorithms in nearly all cross-sections that we explore. With an overall calibrated accuracy rate of 94.40% it offers a significant improvement over the (volume based) LR and tick methods, 77.24% and 78.96%, respectively. We do note that because the BVC is a very new method, trading instrument specific calibration is required in its optimal implementation.

Chakrabarty et al. (2013) find that the netting of trades into bars contributes much of the increase in trade classification accuracy and the bulk tick test can greatly outperform the BVC. However, we find that the accuracy of the BVC is very sensitive to its implementation. In the bulk tick test, trades are classified individually using the tick test, and the aggregation into bars only serves to monotonically increase accuracy as individual misclassifications are offset, or netted. The BVC algorithm requires appropriate bar sizes because the bars (and price changes between them) are actually used in the classification of order flow. In exploring calibration of the BVC we find that bar size must be appropriate for the underlying instrument because bars that are "too small" can lead to price changes that are "too small," and bar sizes that are "too big" can lead to price changes that are from too far in the past. Similarly, poor fitting price change distributions can also limit the accuracy of the BVC versus bulk tick. Bar size and price change distribution are intimately related inputs to the BVC that drive its ability to perform. When these BVC inputs are chosen appropriately, the accuracy advantage of the bulk tick method substantially declines and for some instances is alleviated. Finally, the advantage of the BVC in capturing information rather than just trade aggressor suggests that the choice between the BVC and tick test ultimately is not simply a trade-off between the slightly higher accuracy of the tick test and the computational efficiency of the BVC.

References

Acharya, V.V., and L.H. Pedersen, 2005. Asset pricing with liquidity risk. Journal of Financial Economics 77, 375-410.

Andersen, T. G., and O. Bondarenko, 2013. Assessing Measures of Order Flow Toxicity via Perfect Trade Classification. Working Paper. Northwestern University.

Bakshi, G., N. Kapadia, and D. Madan, 2003. Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options. Review of Financial Studies 16, 101-143.

Baruch, S., and L. R. Glosten, 2013. Flickering Quotes. Working paper. University of Utah.

Baruch, S., M. Panayides, and K. Venkataraman, 2013. Informed Trading Before Unscheduled Corporate Events: Theory and Evidence, Working Paper.

Berkman, H., T. Brailsford, and A. Frino, 2005. A note on execution costs for stock index futures: Information versus liquidity effects. Journal of Banking & Finance 29, 565-577.

Blume, M., and M. Goldstein, 1997. Quotes, order flow, and price discovery. Journal of Finance 52, 221-244.

Boehmer, E., J. Grammig, and E. Theissen, 2007. Estimating the probability of informed trading – does trade misclassification matter? Journal of Financial Markets 10,1 26-47.

Boehmer, E., and E. K. Kelley, 2009. Institutional Investors and the Information Efficiency of Prices. Review of Financial Studies 22, 3563-3594.

Boehmer, E., Kinglsey Fong, and Julie Wu, 2014. International Evidence on Algorithmic Trading. Working paper, Singapore Management University.

Bouchaud, J.P., J. Farmer and F. Lillo, 2009, How markets slowly digest changes in supply and demand, in Handbook of Financial Markets: Dynamics and Evolution, North-Holland, Elsevier.

Brogaard, J., 2011. High Frequency Trading and its Impact on Market Quality. Working paper, Northwestern University.

Brogaard, J., Hendershott, T., and R. Riordan, 2014. High Frequency Trading and Price Discovery. Review of Financial Studies 27,8 2267-2306.

Budish, E., P. Cramton, and J. Shim, 2013. The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response. Working paper, University of Chicago.

Chakrabarty, B., P. Moulton, and A. Shkilko, 2012. Short Sales, Long Sales, and the Lee-Ready Trade Classification Algorithm Revisited. Journal of Financial Markets 15, 467-491.

Chakrabarty, B., R. Pascual, and A. Shkilko, 2013. Trade Classification Algorithms: A Horse-race Between the Bulk-based and Tick-based Rules. Working Paper.

Chan, L. K. C., J. Lakonishok, 1993. Institutional trades and stock price behavior. Journal of Financial Economics 33, 173-199.

Chordia, T., R. Roll, and A. Subrahmanyam, 2000. Commonality in liquidity. Journal of Financial Economics 56, 3-28.

Chordia, T., R. Roll, and A. Subrahmanyam, 2005. Evidence on the speed of convergence to market efficiency. Journal of Financial Economics 76, 271-292.

Easley, D., and M. O'Hara, 1987. Price, Trade Size, and Information in Securities Markets. Journal of Financial Economics 19, 69-90.

Easley, D., N. Kiefer, M. O'Hara, J. Paperman, 1996. Liquidity, information, and infrequently traded stocks. Journal of Finance 51, 1405-1436.

Easley, D., M. M. Lopez de Prado, and M. O'Hara, 2012. Flow toxicity and liquidity in a high-frequency world. Review of Financial Studies 25, 1457-1493.

Easley, D., M. M. Lopez de Prado, and M. O'Hara, 2013. Bulk classification of trading activity. Working paper, Cornell University.

Ellis, K., R. Michaely, and M. O'Hara, 2000. The Accuracy of Trade Classification Rules: Evidence from Nasdaq. Journal of Financial and Quantitative Analysis 35,4 529-551.

Hasbrouck, J., 2013. High Frequency Quoting: Short-Term Volatility in Bids and Offers. Working paper. New York University.

Hasbrouck, J., and G. Saar, 2013. Low-latency trading. Journal of Financial Markets 16, 646-679.

Hendershott, T. and P. Moulton, 2011. Automation, speed, and stock market quality: The NYSE's Hybrid. Journal of Financial Markets 14, 568-604.

Hendershott, T. and R. Riordan, 2013. Algorithmic Trading and the Market for Liquidity. Journal of Financial and Quantitative Analysis 48, 1001-1024.

Holden, Craig W., and Stacey Jacobsen., 2014. Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions. Journal of Finance 69,4 1747-1785.

Holthausen, Robert W., Richard W. Leftwich, and David Mayers, 1987. The Effect of Large Block Transactions on Security Prices: A Cross-Sectional Analysis. Journal of Financial Economics 19,2 237-267.

Jain, Panaj K., 2005. Financial Market Design and the Equity Premium: Electronic versus Floor Trading. Journal of Finance 60,6 2955-2985.

Lee, C., and M. Ready, 1991. Inferring trade direction from intraday data. Journal of Finance 46, 733–746.

Menkveld, Albert J., 2013. High-Frequency Trading and The New-Market Makers. Journal of Financial Markets 16, 712-740.

Odders-White, E., 2000. On the occurrence and consequences of inaccurate trade classification. Journal of Financial Markets 3, 259–286.

Peterson, M. and E. Sirri, 2003. Evaluation of the biases in execution cost estimates using trade and quote data. Journal of Financial Markets 6, 259-280.

Theissen, E., 2001. A test of the accuracy of the Lee/Ready trade classification algorithm. Journal of International Financial Markets, Institutions and Money 11,147-165.

U.S. Securities and Exchange Commission, 2010, Part III: Concept Release on Equity Market Structure; Proposed Rule, 17 CFR Part 242, *Federal Register*, Vol. 75, No. 13, January 21, 2010, Proposed Rules, pages 3594-3614. http://www.sec.gov/rules/concept/2010/34-61358fr.pdf.

Zhang, Z., 2013. Informed Liquidity Provision and Adverse Selection Measures. Working paper, Indiana University.

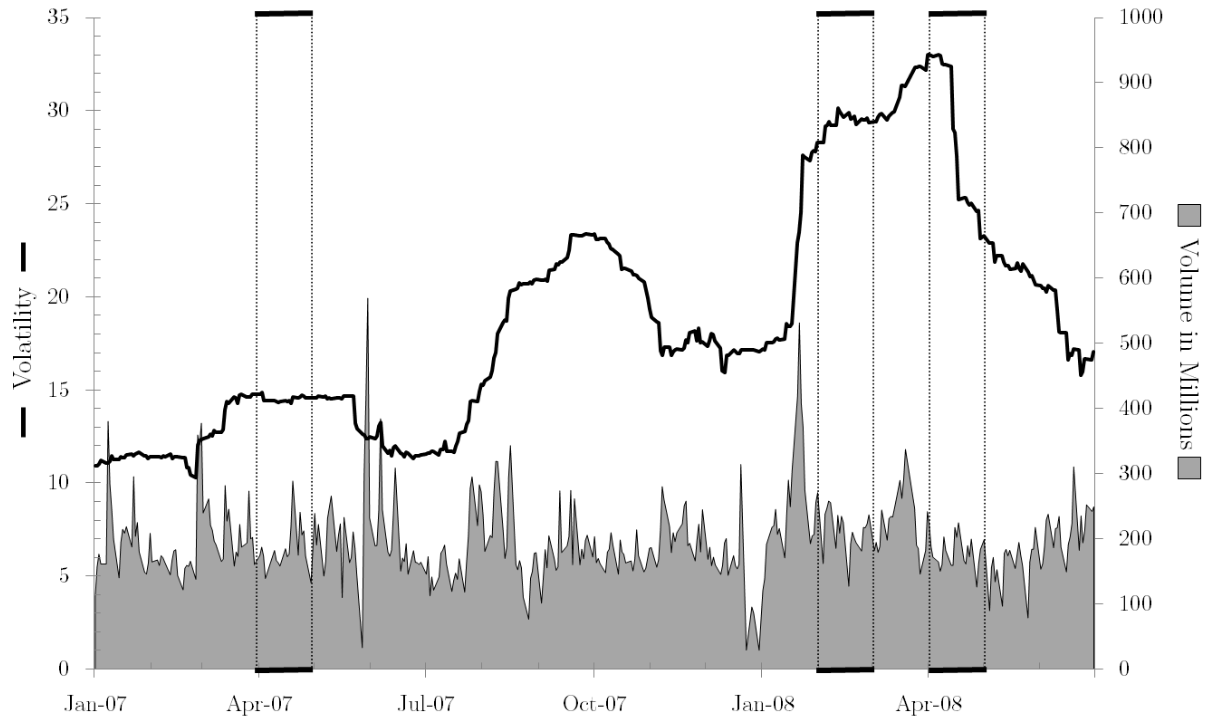
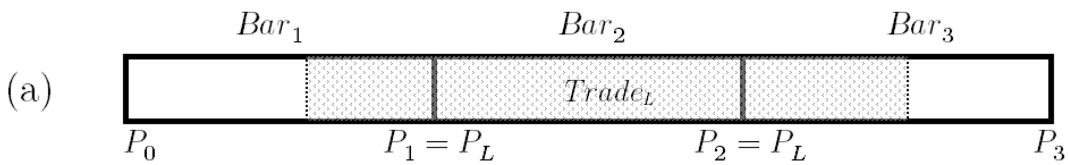


Figure 1 - Graph of volatility and volume of the CAC-250 Index from January 2007 through June 2008. 60 day average volatility of the index is represented by the black line and measured by the left vertical axis. Volume of the CAC-250 index components, in millions, is represented by the gray area at the bottom of the chart and measured by the right axis. The time periods of our sample, April 2007, February 2008 and April 2008, are highlighted.



$$\Delta P_1 = P_1 - P_0 = P_L - P_0$$
$$\Delta P_2 = P_2 - P_1 = P_L - P_L = 0 \quad \rightarrow \quad \hat{V}_2^{Buy} = \hat{V}_2^{Sell} = \frac{1}{2} Size$$
$$\Delta P_3 = P_3 - P_2 = P_3 - P_L$$

$$\Delta P_1 = P_1 - P_0 = P_L - P_0$$
$$\Delta P_2 = P_2 - P_1 = P_2 - P_L$$

Figure 2 - The illustrations above demonstrate how using truncated bar sizes can create a systematic bias in the BVC's volume bar accuracy. Bars are separated by thick black lines and labeled accordingly. Consider a large trade $Trade_L$ (represented by the shaded area) that is larger than the specified volume bar size and is initiated by a buyer (seller) which moves the price P_L up (down) from the previous bar's price P_0. Illustration (a) indicates that using truncated volume bar sizes will incorrectly classify Bar_2 as being made up of half buy volume and half sell volume. In illustration (b), the use of minimum specified bar size prevents $Trade_L$ from being split across bars. By allowing the bar size to extend beyond the minimum, Bar_1 includes all of the volume in $Trade_L$. In this case, the bias is corrected and the contribution to price change by $Trade_L$ will be applied to its entire volume.

Table 1
Sample Summary Statistics

Panel A displays the mean, standard deviation, median, minimum and maximum market capitalization for each capitalization group, small, medium, and large, which we define as companies worth less than €700 million, between €700 million and €7 billion, and above €7 billion, respectively. All market capitalization numbers are in millions of euros. These summary statistics are taken over all three months of our sample, April 2007 and 2008 and February 2008. Panels B, C, and D display analgous statistics for daily traded volume, volume per second, and best bid-offer to trade frequency ratio respectively for each capitalization group.

Panel A: Market Capitalization

	N	Mean	Standard Deviation	Median	Min	Max
Small Cap	34	279.49	174.46	240.51	79.50	683.30
Mid Cap	33	2264.42	1,701.70	1,533.15	703.92	6,834.06
Large Cap	33	25740.33	26,632.00	16,668.87	7,376.94	119,823.84
Total	100	9336.59	19,098.20	1,478.49	79.50	119,823.84

Panel B: Daily Share Volume

	N	Mean	Standard Deviation	Median	Min	Max
Small Cap	34	119,256	405,141	11,370	72	2,279,730
Mid Cap	33	433,730	781,140	79,104	239	3,557,854
Large Cap	33	2,098,083	2,354,407	1,286,111	843	10,352,700
Total	100	876,045	1,674,335	106,174	72	10,352,700

Panel C: Volume Per Second (VWAP Trade Size)

	N	Mean	Standard Deviation	Median	Mode	Mode / Total
Small Cap	34	649.40	1,680.61	211	100	5.2389%
Mid Cap	33	672.77	1,709.15	250	100	4.7573%
Large Cap	33	626.68	1,266.65	265	100	3.4649%
Total	100	634.18	1,353.92	261	100	1.8518%

Panel D: BBO/Trade Ratio

	N	Mean	Standard Deviation	Median	Min	Max
Small Cap	34	4.71	3.33	3.84	1.30	14.24
Mid Cap	33	3.78	2.39	2.98	1.23	10.79
Large Cap	33	1.83	0.93	1.55	0.93	5.95
Total	100	3.45	2.70	2.67	0.93	14.24

Table 2

Breakdown of Lee and Ready Misclassification

This table provides numbers and percentages of Lee and Ready trade (mis)classifications along a number of trade and firm characteristics described throughout. Each panel contains a different cross-section of the data. Summing across a row yields the total number of trades (transactions) in a given sample-category. The denominator for percentages of incorrect and correct classification is the total number of trades in a sample-category (a row). Summing down a column for a full sample or within subsets yields the total number of correctly or incorrectly classified trades in the whole sample or subset. All analyses are at the transaction level for consistency with past studies and are only performed on the trades for which a true trade initiator could be established.

Sample	Category	Correct		Incorrect	
		Number	Percent	Number	Percent
	Overall	9,274,315	78.67%	2,514,726	21.33%
Panel A: Trade price position relative to the quotes					
Full sample	Above the ask	958,885	67.15%	469,098	32.85%
	At the ask	2,656,292	90.43%	281,075	9.57%
	Inside the best quotes	2,204,301	74.60%	750,625	25.40%
	At the bid	2,494,728	81.24%	575,939	18.76%
	Below the bid	960,109	68.67%	437,989	31.33%
Panel B: Market volatility - separate results for each month					
Full Sample	April 2007	2,444,327	80.54%	590,560	19.46%
	Februrary 2008	3,360,459	77.74%	962,070	22.26%
	April 2008	3,469,529	78.29%	962,096	21.71%
Panel C: Trade Frequency - high frequency trades					
Full sample	Many trades in 1s	6,907,835	76.54%	2,116,927	23.46%
	Single trade in 1s	2,366,480	85.61%	397,799	14.39%
Buys	Many trades in 1s	3,535,839	79.55%	909,158	20.45%
	Single trade in 1s	1,152,521	86.60%	178,283	13.40%
Sells	Many trades in 1s	3,371,996	73.63%	1,207,769	26.37%
	Single trade in 1s	1,213,959	84.69%	219,516	15.31%
Panel D: Quote Frequency - high frequency quoting					
Full sample	Many quotes in 1s	3,356,614	68.54%	1,540,627	31.46%
	Single quote in 1s	5,917,701	85.87%	974,099	14.13%
Buys	Many quotes in 1s	1,809,662	74.75%	611,345	25.25%
	Single quote in 1s	2,878,698	85.81%	476,096	14.19%
Sells	Many quotes in 1s	1,546,952	62.47%	929,282	37.53%
	Single quote in 1s	3,039,003	85.92%	498,003	14.08%
Panel E: Trade and Quote Frequency - Multiple trades or quotes					
Multiple Trades per Second	Many quotes in 1 s	2,908,719	67.70%	1,387,915	32.30%
	Single quote in 1 s	3,999,116	84.58%	729,012	15.42%
Single Trade per Second	Many quotes in 1 s	447,895	74.57%	152,712	25.43%
	Single quote in 1 s	1,918,585	88.67%	245,087	11.33%
Panel F: Trade Volatility					
Full Sample	High volatility	3,927,726	77.49%	1,140,781	22.51%
	Low volatility	5,346,589	79.56%	1,373,945	20.44%
Panel G: Market Capitalization					
Full Sample	Large Cap	7,741,142	78.21%	2,156,251	21.79%
	Mid Cap	1,324,570	80.85%	313,653	19.15%
	Small Cap	208,603	82.31%	44,822	17.69%

Table 3

Bulk Volume Classification Results

This table details the results from implementing the Bulk Volume Classification (BVC) from Easley, Lopez de Prado, and O'Hara (2013) using the t-distribution with 0.25 df. The algorithm is implemented so that the unit of observation is monthly trade data. Results are shown for the overall accuracy, as well as accuracy in the monthly, size, and multiple trades and/or quotes subsamples. Panel A displays results for volume bar aggregation and Panel B shows the results for time bars. The first column in Panel A (B) shows the size of the volume (time) bar used and the other columns show corresponding accuracy. The bordered boxes show the peaks in BVC accuracy and the accuracy ratios higher than the tick rule are in bold-italic type. The Lee and Ready results using volume (as opposed to transactions as in Table 2) are included for reference.

	Overall	Sample Period			Market Capitalization			Multiple Intra-Second			# Data	Compression
		April 2007	Feb 2008	April 2008	Small	Medium	Large	Quotes	Trades	Both	Points	Ratio
L&R	77.24%	78.85%	76.40%	76.88%	81.35%	80.11%	76.55%	67.20%	75.41%	66.55%	--	
Tick Test	78.97%	81.56%	77.46%	78.55%	85.16%	81.99%	78.18%	74.49%	79.70%	75.35%	5,702,246	--
Panel A: Volume Bars												
1,000	62.63%	62.25%	62.44%	63.11%	63.03%	62.53%	62.64%	62.80%	61.97%	62.20%	1,804,147	68.36%
2,500	68.88%	68.03%	68.87%	69.53%	68.59%	68.22%	69.02%	69.60%	68.22%	68.97%	974,275	82.91%
5,000	74.00%	72.82%	74.16%	74.73%	72.89%	72.97%	74.24%	***75.12%***	73.49%	74.62%	568,965	90.02%
10,000	78.81%	77.61%	***79.05%***	***79.49%***	77.28%	77.55%	***79.11%***	***80.07%***	78.48%	***79.73%***	314,709	94.48%
15,000	***81.26%***	80.10%	***81.55%***	***81.84%***	79.76%	79.83%	***81.58%***	***82.48%***	***81.01%***	***82.23%***	218,341	96.17%
25,000	***83.89%***	***82.88%***	***84.16%***	***84.39%***	82.70%	***82.59%***	***84.19%***	***84.94%***	***83.74%***	***84.78%***	135,908	97.62%
30,000	***84.72%***	***83.81%***	***84.99%***	***85.12%***	83.60%	***83.45%***	***85.00%***	***85.69%***	***84.58%***	***85.55%***	114,341	97.99%
50,000	***86.65%***	***85.90%***	***86.87%***	***87.00%***	***85.79%***	***85.76%***	***86.85%***	***87.38%***	***86.60%***	***87.31%***	70,099	98.77%
75,000	***87.86%***	***87.21%***	***88.07%***	***88.15%***	***87.53%***	***87.05%***	***88.03%***	***88.42%***	***87.83%***	***88.38%***	47,303	99.17%
100,000	***88.55%***	***87.90%***	***88.75%***	***88.83%***	***88.43%***	***87.94%***	***88.67%***	***88.99%***	***88.53%***	***88.97%***	35,716	99.37%
150,000	***89.28%***	***88.80%***	***89.49%***	***89.43%***	***89.41%***	***88.78%***	***89.37%***	***89.58%***	***89.29%***	***89.58%***	24,002	99.58%
200,000	***89.71%***	***89.13%***	***89.91%***	***89.95%***	***90.20%***	***89.42%***	***89.75%***	***89.92%***	***89.72%***	***89.92%***	18,084	99.68%
250,000	***89.95%***	***89.48%***	***90.17%***	***90.07%***	***90.84%***	***89.84%***	***89.94%***	***90.11%***	***89.95%***	***90.11%***	14,534	99.75%
300,000	***90.13%***	***89.67%***	***90.41%***	***90.20%***	***91.07%***	***90.09%***	***90.12%***	***90.27%***	***90.14%***	***90.27%***	12,161	99.79%
400,000	***90.38%***	***89.92%***	***90.62%***	***90.48%***	***90.85%***	***90.90%***	***90.28%***	***90.47%***	***90.39%***	***90.47%***	9,180	99.84%
500,000	***90.45%***	***89.88%***	***90.63%***	***90.71%***	***91.61%***	***90.71%***	***90.37%***	***90.51%***	***90.46%***	***90.51%***	7,386	99.87%
600,000	***90.58%***	***90.19%***	***90.79%***	***90.65%***	***91.52%***	***90.94%***	***90.49%***	***90.62%***	***90.59%***	***90.63%***	6,194	99.89%
700,000	***90.48%***	***89.97%***	***90.69%***	***90.63%***	***90.70%***	***90.98%***	***90.38%***	***90.54%***	***90.50%***	***90.55%***	5,345	99.91%
800,000	***90.54%***	***90.19%***	***90.60%***	***90.73%***	***91.14%***	***91.24%***	***90.40%***	***90.52%***	***90.55%***	***90.53%***	4,696	99.92%

Table 3 - Continued

Bulk Volume Classification Results

		Sample Period			Market Capitalization			Multiple Intra-Second			# Data	Compression
	Overall	April 2007	Feb 2008	April 2008	Small	Medium	Large	Quotes	Trades	Both	Points	Ratio
L&R	77.24%	78.85%	76.40%	76.88%	81.35%	80.11%	76.55%	67.20%	75.41%	66.55%	--	
Tick Test	78.97%	81.56%	77.46%	78.55%	85.16%	81.99%	78.18%	74.49%	79.70%	75.35%	5,702,246	--
Panel B: Time Bars												
2 seconds	58.66%	58.10%	59.03%	58.70%	57.50%	58.06%	58.81%	59.96%	59.07%	60.16%	4,615,890	19.05%
5	62.51%	61.21%	63.52%	62.45%	58.51%	60.35%	63.06%	64.66%	62.92%	64.80%	3,503,071	38.57%
10	65.74%	64.02%	67.04%	65.70%	59.68%	62.41%	66.58%	68.20%	66.11%	68.29%	2,708,430	52.50%
15	67.87%	65.95%	69.31%	67.87%	60.53%	63.89%	68.89%	70.48%	68.22%	70.54%	2,254,136	60.47%
30	71.78%	69.64%	73.29%	71.85%	62.42%	66.86%	73.04%	74.45%	72.09%	74.49%	1,571,037	72.45%
60	75.88%	73.65%	77.41%	76.00%	64.74%	70.30%	77.32%	***78.49%***	76.14%	***78.50%***	1,032,884	81.89%
120	***79.82%***	77.81%	***81.14%***	***79.99%***	67.60%	74.13%	***81.32%***	***82.12%***	***80.04%***	***82.13%***	643,606	88.71%
300	***84.22%***	***82.65%***	***85.25%***	***84.35%***	72.17%	79.24%	***85.57%***	***85.94%***	***84.39%***	***85.94%***	326,094	94.28%
600	***86.68%***	***85.49%***	***87.39%***	***86.85%***	76.09%	***82.76%***	***87.78%***	***87.91%***	***86.82%***	***87.92%***	188,261	96.70%
900	***87.84%***	***86.78%***	***88.47%***	***88.00%***	78.14%	***84.63%***	***88.77%***	***88.80%***	***87.97%***	***88.81%***	135,220	97.63%
1,800	***89.20%***	***88.24%***	***89.68%***	***89.42%***	81.63%	***87.13%***	***89.84%***	***89.83%***	***89.29%***	***89.84%***	76,188	98.66%
3,600	***90.12%***	***89.39%***	***90.41%***	***90.38%***	84.27%	***89.08%***	***90.51%***	***90.52%***	***90.18%***	***90.53%***	42,925	99.25%
7,200	***90.69%***	***89.90%***	***91.00%***	***90.97%***	***86.71%***	***90.37%***	***90.87%***	***90.97%***	***90.72%***	***90.97%***	24,635	99.57%
9,000	***90.79%***	***90.08%***	***91.19%***	***90.93%***	***86.87%***	***90.62%***	***90.95%***	***91.04%***	***90.83%***	***91.03%***	20,937	99.63%
10,800	***90.86%***	***90.18%***	***91.19%***	***91.04%***	***87.58%***	***90.93%***	***90.95%***	***91.03%***	***90.88%***	***91.03%***	16,646	99.71%
14,400	***90.82%***	***89.98%***	***91.24%***	***91.03%***	***88.56%***	***91.02%***	***90.85%***	***91.04%***	***90.84%***	***91.03%***	14,797	99.74%
18,000	***90.90%***	***90.34%***	***91.28%***	***90.93%***	***88.63%***	***91.29%***	***90.89%***	***91.02%***	***90.91%***	***91.01%***	11,504	99.80%
25,200	***90.55%***	***90.30%***	***90.67%***	***90.61%***	***88.15%***	***91.29%***	***90.47%***	***90.61%***	***90.54%***	***90.59%***	10,977	99.81%

Table 4

Bulk Tick Rule Classification Results

This table details the results from the bulk tick rule as in Chakrabarty, et al (2013). The algorithm is implemented so that the unit of observation is monthly trade data. Results are shown for the overall accuracy, as well as accuracy in the monthly, size, and multiple trades and/or quotes subsamples. Panel A displays results for volume bar aggregation and Panel B shows the results for time bars. The first column in Panel A (B) shows the size of the volume (time) bar used and the other columns show corresponding accuracy. The bordered boxes show the peaks in tick rule accuracy and the accuracy ratios greater than the optimized BVC are in bold-italic type. The best bulk Lee and Ready results using volume (as opposed to transactions as in Table 2) are included for reference, though they are nearly uniformly lower than the bulk tick results. We also include bar size and distribution parameter calibrated BVC accuracy ratios for comparison purposes.

		Sample Period			Market Capitalization			Multiple Intra-Second		
	Overall	April 2007	Feb 2008	April 2008	Small	Medium	Large	Quotes	Trades	Both
Bulk L&R	95.79%	96.37%	96.01%	95.13%	96.43%	96.11%	95.71%	95.65%	95.76%	95.63%
Calibrated BVC	94.40%	93.94%	94.57%	94.57%	92.92%	94.07%	94.51%	94.64%	94.41%	94.61%
Panel A: Volume Bars										
1,000	79.85%	82.17%	78.48%	79.48%	85.84%	82.69%	79.10%	76.30%	80.17%	76.64%
2,500	81.17%	83.12%	80.01%	80.89%	86.67%	83.70%	80.51%	78.56%	81.26%	78.64%
5,000	82.80%	84.35%	81.84%	82.62%	87.59%	84.96%	82.23%	80.99%	82.76%	80.95%
10,000	84.93%	85.95%	84.21%	84.89%	88.79%	86.58%	84.48%	83.82%	84.82%	83.72%
15,000	86.31%	87.07%	85.74%	86.31%	89.63%	87.65%	85.94%	85.56%	86.19%	85.45%
25,000	88.07%	88.52%	87.66%	88.15%	90.87%	89.05%	87.79%	87.64%	87.95%	87.53%
30,000	88.69%	89.04%	88.34%	88.77%	91.27%	89.53%	88.44%	88.35%	88.57%	88.26%
50,000	90.35%	90.52%	90.13%	90.45%	92.50%	90.86%	90.18%	90.19%	90.26%	90.12%
75,000	91.57%	91.60%	91.42%	91.71%	***93.41%***	91.86%	91.46%	91.50%	91.50%	91.44%
100,000	92.36%	92.29%	92.26%	92.52%	***93.93%***	92.50%	92.29%	92.35%	92.29%	92.29%
150,000	93.32%	93.18%	93.32%	93.41%	***94.59%***	93.41%	93.26%	93.33%	93.27%	93.29%
200,000	93.98%	93.85%	93.99%	94.08%	***95.06%***	94.03%	93.94%	94.00%	93.94%	93.96%
250,000	***94.41%***	***94.20%***	94.50%	94.47%	***95.61%***	***94.39%***	94.37%	94.44%	94.37%	94.40%
300,000	***94.77%***	***94.57%***	***94.84%***	***94.84%***	***95.71%***	***94.78%***	***94.74%***	***94.81%***	***94.73%***	***94.77%***
400,000	***95.25%***	***95.00%***	***95.33%***	***95.36%***	***96.09%***	***95.35%***	***95.21%***	***95.28%***	***95.22%***	***95.25%***
500,000	***95.59%***	***95.25%***	***95.78%***	***95.68%***	***96.39%***	***95.61%***	***95.56%***	***95.66%***	***95.57%***	***95.63%***
600,000	***95.86%***	***95.60%***	***95.99%***	***95.94%***	***96.68%***	***95.90%***	***95.83%***	***95.91%***	***95.84%***	***95.88%***
700,000	***96.03%***	***95.74%***	***96.18%***	***96.08%***	***96.65%***	***96.14%***	***95.98%***	***96.08%***	***96.00%***	***96.05%***
800,000	***96.26%***	***95.93%***	***96.39%***	***96.36%***	***96.90%***	***96.24%***	***96.24%***	***96.32%***	***96.24%***	***96.29%***

Table 4 - Continued

		Sample Period			Market Capitalization			Multiple Intra-Second		
	Overall	April 2007	Feb 2008	April 2008	Small	Medium	Large	Quotes	Trades	Both
Bulk L&R	95.73%	96.35%	95.97%	95.01%	93.83%	95.33%	95.86%	95.70%	95.74%	95.71%
Calibrated BVC	94.54%	93.66%	95.14%	94.58%	89.22%	93.25%	94.95%	94.98%	94.61%	95.00%
Panel B: Time Bars										
2 seconds	79.32%	81.77%	77.94%	78.88%	85.23%	82.17%	78.58%	75.48%	79.94%	76.18%
5	79.94%	82.14%	78.78%	79.46%	85.27%	82.41%	79.29%	76.84%	80.49%	77.42%
10	80.60%	82.56%	79.66%	80.07%	85.34%	82.67%	80.04%	78.08%	81.06%	78.54%
15	81.12%	82.89%	80.34%	80.59%	85.42%	82.87%	80.65%	79.00%	81.55%	79.40%
30	82.32%	83.68%	81.81%	81.82%	85.58%	83.42%	82.00%	80.84%	82.64%	81.14%
60	83.92%	84.85%	83.71%	83.44%	85.93%	84.23%	83.80%	83.05%	84.17%	83.26%
120	85.87%	86.39%	85.88%	85.47%	86.37%	85.34%	85.96%	85.52%	86.06%	85.67%
300	88.65%	88.64%	88.92%	88.38%	87.28%	87.25%	88.96%	88.72%	88.79%	88.81%
600	90.66%	90.41%	91.00%	90.50%	88.39%	88.96%	91.06%	90.89%	90.78%	90.97%
900	91.71%	91.37%	92.06%	91.60%	88.99%	90.01%	92.13%	91.96%	91.81%	92.02%
1,800	93.24%	92.85%	93.58%	93.20%	***90.42%***	91.58%	93.66%	93.52%	93.33%	93.57%
3,600	94.49%	***94.03%***	94.83%	94.50%	***91.90%***	93.10%	94.84%	94.75%	94.56%	94.79%
7,200	***95.50%***	***94.98%***	***95.85%***	***95.55%***	***93.15%***	***94.30%***	***95.81%***	***95.76%***	***95.56%***	***95.79%***
9,000	***95.69%***	***95.20%***	***96.03%***	***95.70%***	***93.34%***	***94.53%***	***95.99%***	***95.92%***	***95.74%***	***95.94%***
10,800	***96.03%***	***95.48%***	***96.37%***	***96.10%***	***93.75%***	***94.92%***	***96.32%***	***96.26%***	***96.08%***	***96.28%***
14,400	***96.27%***	***95.77%***	***96.56%***	***96.33%***	***94.28%***	***95.25%***	***96.53%***	***96.49%***	***96.31%***	***96.51%***
18,000	***96.49%***	***95.95%***	***96.83%***	***96.54%***	***94.41%***	***95.48%***	***96.75%***	***96.71%***	***96.53%***	***96.73%***
25,200	***96.57%***	***96.00%***	***96.87%***	***96.70%***	***94.54%***	***95.61%***	***96.83%***	***96.78%***	***96.61%***	***96.80%***

Table 5

Comparison of BVC and Bulk Tick Test

This table compares the accuracy of the bulk tick test to that of the BVC for our Euronext sample and the INET sample of Chakrabarty et al. (2013). Each number represents tick test accuracy minus BVC accuracy at the same bar size, so that a positive (negative) number means the bulk tick (BVC) is outperforming the BVC (bulk tick). The first two columns show differences between normal and Student's t distribution BVC implementations in our Euronext data. The second two columns display differences from Chakrabarty, et al. (2013) using the normal distribution over two different time periods. The last two columns show differences for bar calibrated along with bar and distribution parameter (market capitalization based) calibrated Student's t distribution BVC implementations and the quantity of sample stock-months calibrated at the specified bar size in parentheses. Panel A shows the results for a selection of volume bars and Panel B shows a selection of time bars.

Bar Size	Euronext Data (Normal)	Euronext Data	Chakrabarty 2011 Data	Chakrabarty 2005 Data	Euronext (Bar Calibrated)		Euronext (Bar&Dist Calib)	
Panel A: Volume Bars								
1,000	13.24%	17.22%	10.20%	12.80%	4.15%	(15)	2.44%	(15)
2,500	10.48%	12.29%	9.70%	11.45%	1.34%	(8)	0.12%	(7)
5,000	8.89%	8.80%	10.40%	11.50%	3.39%	(9)	4.21%	(9)
7,500	8.39%	7.11%	11.10%	12.00%	2.42%	(8)	2.06%	(9)
10,000	8.28%	6.11%	11.60%	12.40%	2.16%	(17)	4.59%	(13)
30,000	9.37%	3.97%	14.20%	14.50%	4.57%	(21)	3.75%	(23)
50,000	10.54%	3.70%	15.70%	15.50%	5.39%	(36)	5.45%	(33)
100,000	12.49%	3.82%	N/A	N/A	4.42%	(39)	2.51%	(23)
250,000	15.05%	4.46%	N/A	N/A	3.93%	(63)	1.72%	(37)
500,000	17.16%	5.14%	N/A	N/A	4.35%	(79)	1.86%	(126)
Panel B: Time Bars								
2 seconds	15.58%	20.66%	12.80%	15.30%	N/A		N/A	
5	13.67%	17.43%	12.00%	14.80%	N/A		N/A	
10	12.08%	14.86%	11.20%	14.30%	N/A		N/A	
30	9.47%	10.54%	9.40%	13.30%	N/A		N/A	
60	8.29%	8.04%	8.40%	12.30%	N/A		N/A	
300	8.49%	4.43%	7.40%	10.40%	N/A		N/A	
1,800	12.06%	4.05%	9.50%	11.00%	-16.46%	(5)	10.30%	(2)
3,600	13.63%	4.37%	11.10%	12.20%	-2.11%	(13)	14.92%	(3)
7,200	15.29%	4.81%	12.90%	13.50%	4.60%	(54)	1.49%	(10)
25,200	18.39%	6.03%	16.30%	16.30%	4.81%	(231)	1.71%	(288)

Table 6
Time Bar Volume

This table shows the mean and standard deviation of the volume in various time bars. The first column shows the overall mean by stock-month, while the last three show the stock-month averages split by market capitalization. Small, medium, and large capitalizations are defined as companies worth less than €700 million, between €700 million and €7 billion, and above €7 billion, respectively. Our sample is the trading of 100 firms listed on Euronext for April 2007 and 2008 and February 2008.

	Overall		Firm Size					
			Small		Medium		Large	
Mean (Std Dev) Volume within Time Bars with Non-Zero Volume								
Sub-second	306	(643)	365	(834)	351	(819)	298	(602)
1 second	627	(1,267)	649	(1,681)	673	(1,709)	627	(1,267)
2 seconds	783	(1,650)	688	(1,795)	773	(1,960)	789	(1,576)
3	880	(1,843)	699	(1,834)	828	(2,098)	898	(1,784)
5	1,031	(2,139)	732	(1,935)	905	(2,289)	1,074	(2,110)
10	1,335	(2,732)	800	(2,116)	1,058	(2,671)	1,439	(2,770)
15	1,604	(3,244)	854	(2,272)	1,186	(2,981)	1,776	(3,352)
20	1,846	(3,708)	901	(2,412)	1,295	(3,263)	2,088	(3,884)
30	2,289	(4,563)	980	(2,628)	1,507	(3,751)	2,673	(4,884)
60	3,496	(6,896)	1,172	(3,238)	2,025	(5,013)	4,397	(7,689)
90	4,601	(9,051)	1,330	(3,703)	2,509	(6,154)	6,058	(10,334)
120	5,619	(11,102)	1,467	(4,138)	2,943	(7,219)	7,665	(12,895)
180	7,540	(14,994)	1,708	(4,902)	3,753	(9,155)	10,847	(17,856)
300	11,090	(22,344)	2,110	(6,158)	5,252	(12,762)	17,099	(27,437)
600	19,209	(39,437)	2,939	(9,172)	8,734	(20,842)	32,497	(50,347)
900	26,743	(55,440)	3,658	(11,818)	12,040	(28,516)	47,552	(72,129)
1200	33,890	(70,472)	4,303	(14,163)	15,241	(36,036)	62,283	(92,668)
1800	47,464	(99,800)	5,515	(18,903)	21,388	(50,279)	90,930	(133,260)
2700	67,663	(143,429)	7,126	(25,008)	30,300	(70,181)	137,715	(195,554)
3600	84,245	(181,181)	8,640	(31,712)	38,843	(90,792)	169,913	(246,829)
5400	117,785	(246,248)	11,318	(41,306)	53,500	(121,159)	251,276	(338,408)
7200	146,792	(328,105)	13,952	(53,471)	69,529	(162,784)	304,030	(455,012)

Table 7
Market Time Between Filled Bars

This table shows the mean and standard deviation of the time (in seconds) between bars. Panel A (B) shows the seconds between bars for volume (time) bars. The first column in each panel shows the overall mean by stock-month, while the last three show the stock-month averages split by market capitalization. Small, medium, and large capitalizations are defined as companies worth less than €700 million, between €700 million and €7 billion, and above €7 billion, respectively. Our sample is the trading of 100 firms listed on Euronext for April 2007 and 2008 and February 2008.

	Overall		Firm Size					
			Small		Medium		Large	
Panel A: Volume Bars								
Mean (Std Dev) Time in Seconds Elapsed Between Volume Bars								
Trade Level	30.8	(342.2)	414.4	(1,840.9)	67.7	(408.4)	12.4	(60.3)
500	66.0	(951.2)	908.6	(4,923.7)	146.6	(1,318.1)	26.9	(232.5)
1000	94.0	(1,295.4)	1,266.0	(6,573.7)	208.4	(1,825.9)	38.9	(333.7)
2500	167.4	(1,846.0)	2,213.6	(9,788.3)	366.4	(2,143.1)	71.5	(599.3)
5000	276.7	(2,476.0)	3,490.7	(13,407.9)	601.6	(2,882.0)	121.4	(477.5)
10000	480.4	(3,465.2)	5,694.3	(18,043.1)	1,040.4	(4,283.1)	218.8	(669.2)
15000	675.1	(4,447.2)	7,710.1	(23,294.4)	1,457.0	(5,242.2)	315.4	(896.4)
20000	865.1	(5,385.8)	9,653.2	(28,168.9)	1,853.2	(6,209.6)	411.2	(1,120.0)
25000	1,046.5	(6,062.2)	11,350.8	(31,098.0)	2,248.3	(7,353.3)	505.9	(1,229.7)
30000	1,223.0	(6,809.3)	12,829.3	(34,485.9)	2,636.3	(8,437.9)	600.5	(1,421.3)
40000	1,567.6	(8,103.9)	15,677.0	(39,985.8)	3,394.8	(10,482.1)	789.8	(1,855.8)
50000	1,892.3	(9,263.1)	17,902.0	(44,933.4)	4,130.7	(12,340.9)	977.9	(2,207.6)
75000	2,645.7	(10,771.2)	22,254.0	(49,162.7)	5,813.8	(15,253.8)	1,445.7	(3,076.0)
100000	3,338.0	(12,330.0)	25,356.2	(54,069.4)	7,371.6	(18,317.7)	1,906.8	(3,857.1)
Panel B: Time Bars								
Mean (Std Dev) Time Elapsed Between Time Bars with Non-Zero Volume								
Trade Level	30.8	(342.2)	414.4	(1,840.9)	67.7	(408.4)	12.4	(60.3)
3	42.7	(404.7)	452.8	(1,931.5)	83.5	(453.9)	17.7	(71.9)
5	50.1	(437.9)	472.7	(1,971.1)	91.4	(474.1)	21.2	(78.3)
10	64.8	(497.1)	511.9	(2,046.4)	106.5	(510.4)	28.4	(89.9)
20	89.7	(583.3)	575.2	(2,160.9)	131.2	(563.9)	41.2	(107.4)
60	169.8	(796.6)	748.2	(2,439.3)	204.7	(696.7)	86.8	(153.2)
90	223.2	(908.9)	849.2	(2,583.4)	252.6	(769.1)	119.6	(178.2)
120	272.6	(1,000.5)	936.7	(2,699.8)	296.3	(829.0)	151.4	(200.1)
300	538.2	(1,379.0)	1,348.4	(3,167.4)	528.9	(1,085.4)	337.7	(289.4)
600	932.5	(1,774.5)	1,879.6	(3,646.7)	879.7	(1,370.0)	641.8	(386.6)
900	1,298.6	(2,069.9)	2,340.9	(3,990.3)	1,212.8	(1,629.2)	939.2	(461.1)
1200	1,645.9	(2,298.6)	2,754.6	(4,263.4)	1,535.5	(1,812.2)	1,230.1	(528.2)
3600	4,094.1	(3,490.6)	5,542.3	(5,710.1)	3,914.4	(2,785.7)	3,354.4	(1,224.7)
5400	5,724.3	(4,083.1)	7,263.6	(6,396.2)	5,390.2	(3,386.0)	4,959.4	(1,495.7)
7200	7,129.7	(4,825.3)	8,956.0	(7,015.0)	7,001.8	(4,032.2)	5,992.3	(2,828.4)

Table 8

Bias from Exact Volume Bar Implementation of BVC

This table documents the potential bias from implementing the BVC algorithm using exact volume bar sizes. It also shows the mean and standard deviation of trade size, as this will affect the size of the potential bias. The bias is shown for a range of bar sizes as well as on the cross-sectional cuts of sample month and firm capitalization. Figure 2 explains the bias in volume bar implementation. Our sample is for three months of daily trading on Euronext Paris. We have the complete order book time-stamped to the second.

	Overall	Sample Period			Firm Size		
		April 2007	Feb 2008	April 2008	Small	Medium	Large
Trade Size (Mean)	626.7	650.0	648.2	609.5	649.4	672.8	626.7
Trade Size (Std Dev)	1,266.6	1,523.1	1,322.1	1,252.8	1,680.6	1,709.1	1,266.6
Bias from Exact Volume Bars							
200	39.63%	40.36%	39.67%	39.04%	40.26%	40.26%	39.49%
500	31.96%	33.10%	31.96%	31.10%	33.12%	33.10%	31.70%
1000	25.06%	26.42%	25.01%	24.07%	26.36%	26.59%	24.72%
2500	16.33%	17.76%	16.18%	15.40%	18.35%	18.21%	15.89%
5000	10.68%	11.83%	10.52%	9.96%	12.44%	12.50%	10.26%
10000	6.51%	7.39%	6.34%	6.01%	7.90%	8.03%	6.17%
15000	4.75%	5.45%	4.63%	4.34%	5.98%	6.07%	4.45%
20000	3.71%	4.28%	3.59%	3.41%	4.62%	4.87%	3.46%
25000	3.07%	3.55%	2.98%	2.81%	4.03%	4.05%	2.85%
30000	2.64%	3.10%	2.52%	2.42%	3.47%	3.51%	2.44%
40000	2.02%	2.40%	1.94%	1.82%	2.68%	2.73%	1.86%
50000	1.65%	1.93%	1.58%	1.50%	2.11%	2.31%	1.50%
75000	1.12%	1.34%	1.06%	1.02%	1.65%	1.52%	1.03%
100000	0.84%	0.99%	0.80%	0.77%	1.27%	1.16%	0.77%
150000	0.58%	0.68%	0.56%	0.53%	0.81%	0.84%	0.52%
200000	0.44%	0.52%	0.41%	0.39%	0.68%	0.64%	0.39%

Table 9

High Low Trading Range Regressions on Tick and BVC Order Imbalance

This table shows the regression results for the regression of $HL_\tau = \alpha_0 + \alpha_1[HL_{\tau-1}] + \gamma|\widehat{OI}_\tau| + \varepsilon_\tau$ with the addition of firm and month fixed effects. Please see the text for variable definitions. Panel A shows the volume and time bar regression results for the BVC estimates of order imbalance and Panel B shows the estimates for bulk tick order imbalance. The t-statistics are generated from standard errors clustered by firm.

Bar Size	Adj. R^2	Coeff(α_0)	Coeff(α_1)	Coeff(γ)	t(α_0)	t(α_1)	t(γ)
Panel A: BVC OI Estimate Results							
1,000	0.654	-0.002	0.336	0.143	-0.42	8.84	6.58
2,500	0.693	-0.001	0.332	0.208	-0.11	8.16	6.66
5,000	0.728	0.000	0.345	0.277	-0.04	12.74	6.58
7,500	0.750	0.003	0.310	0.333	0.31	14.11	6.61
10,000	0.760	0.004	0.310	0.376	0.37	12.86	6.48
30,000	0.804	0.003	0.355	0.616	0.12	5.90	6.53
50,000	0.822	0.010	0.340	0.767	0.41	6.31	6.45
100,000	0.832	0.058	0.182	1.124	1.47	4.72	6.33
250,000	0.804	0.076	0.230	1.648	1.22	5.71	6.35
2 seconds	0.023	0.001	0.032	0.013	3.28	3.29	10.17
3	0.039	0.001	0.053	0.023	2.88	4.33	10.33
5	0.067	0.002	0.093	0.037	2.48	7.77	10.39
10	0.116	0.003	0.124	0.065	2.50	9.52	9.99
30	0.243	0.006	0.218	0.136	1.87	10.96	8.31
60	0.323	0.008	0.256	0.204	1.59	6.96	7.67
300	0.473	0.020	0.351	0.427	1.20	3.97	6.93
1,800	0.472	0.067	0.311	0.998	1.45	2.43	7.91
3,600	0.476	0.093	0.257	1.540	1.67	2.88	6.93
7,200	0.484	0.138	0.179	2.280	1.90	2.88	6.73

Table 9 - Continued

High Low Trading Range Regressions on Tick and BVC Order Imbalance

Bar Size	Adj. R^2	Coeff(α_0)	Coeff(α_1)	Coeff(γ)	t(α_0)	t(α_1)	t(γ)
Panel B: Bulk Tick OI Estimate Results							
1,000	0.632	0.048	0.271	-0.032	16.70	3.76	-16.05
2,500	0.683	0.049	0.346	-0.024	24.62	10.08	-13.33
5,000	0.710	0.059	0.358	-0.023	32.19	16.57	-10.81
7,500	0.730	0.071	0.323	-0.021	38.21	19.21	-9.58
10,000	0.742	0.080	0.328	-0.022	33.82	18.89	-8.91
30,000	0.787	0.121	0.371	-0.019	11.53	6.99	-4.79
50,000	0.805	0.157	0.352	-0.009	13.43	8.00	-1.92
100,000	0.801	0.244	0.294	0.000	18.64	8.80	-0.03
250,000	0.784	0.398	0.240	0.065	13.82	5.81	1.98
2 seconds	0.060	0.039	0.031	-0.038	9.20	3.32	-8.43
3	0.075	0.043	0.053	-0.041	9.23	4.38	-8.11
5	0.094	0.047	0.094	-0.043	9.49	7.84	-7.84
10	0.130	0.053	0.127	-0.046	9.34	9.42	-7.08
30	0.234	0.061	0.227	-0.047	10.12	12.13	-5.88
60	0.304	0.071	0.267	-0.049	12.15	7.87	-5.40
300	0.450	0.116	0.365	-0.067	12.40	4.24	-5.36
1,800	0.451	0.286	0.321	-0.182	6.49	2.52	-3.72
3,600	0.449	0.436	0.262	-0.316	9.57	2.94	-3.81
7,200	0.452	0.656	0.178	-0.515	13.29	2.90	-4.27

Table 10

High Low Trading Range Regressions on Return Subsamples

This table shows the regression results for the regression of $HL_\tau = \alpha_0 + \alpha_1[HL_{\tau-1}] + \gamma|\widehat{OI}_\tau| + \varepsilon_\tau$. Please see the text for variable definitions. Panel A (B) displays results for the regression using the order imbalance estimated using the BVC (bulk tick test). In both panels, the regressions are split by large (first or fourth quartile of returns) and small (second or third quartile of returns) magnitude of returns. The distribution of returns that defines the quartiles is estimated for each bar. The t-statistics are generated from standard errors clustered by firm. The aggregated coefficients are weighted by the reciprocal of the squared standard error for each regression specification and the aggregate standard errors are corrected using the method from Chordia et al. (2000).

	Large |Returns|			Small |Returns|		
Bar Size	Adj. R^2	Coeff(γ)	t(γ)	Adj. R^2	Coeff(γ)	t(γ)
Panel A: BVC OI Estimate Results						
1,000	0.566	0.222	7.45	0.347	0.029	2.41
2,500	0.605	0.335	7.33	0.408	0.032	2.55
5,000	0.641	0.444	7.24	0.474	0.032	2.58
7,500	0.662	0.542	7.54	0.531	0.030	2.17
10,000	0.670	0.609	7.20	0.508	0.026	1.79
30,000	0.740	1.088	7.28	0.625	0.039	1.53
50,000	0.776	1.357	7.54	0.775	0.022	0.90
100,000	0.750	1.992	7.14	0.835	0.139	2.65
250,000	0.761	2.937	6.02	0.631	0.038	0.80
Mean Effect Size		0.369	8.58		0.031	2.79
Panel B: Bulk Tick OI Estimate Results						
1,000	0.448	-0.066	-6.37	0.390	-0.030	-10.61
2,500	0.592	-0.046	-8.31	0.439	-0.030	-8.59
5,000	0.626	-0.046	-8.09	0.497	-0.032	-5.91
7,500	0.641	-0.048	-6.93	0.561	-0.036	-6.62
10,000	0.654	-0.054	-6.26	0.533	-0.041	-6.43
30,000	0.725	-0.079	-4.73	0.631	-0.053	-3.74
50,000	0.756	-0.074	-2.80	0.760	-0.084	-5.05
100,000	0.732	-0.139	-3.51	0.819	-0.079	-2.49
250,000	0.744	-0.353	-4.22	0.633	-0.214	-4.03
Mean Effect Size		-0.051	-7.41		-0.033	-7.93

Table 11

High Low Trading Range Regressions on Order Imbalance Index and Option Subsamples

This table shows the regression results for the regression of $HL_\tau = \alpha_0 + \alpha_1[HL_{\tau-1}] + \gamma|\widehat{OI}_\tau| + \varepsilon_\tau$. Please see the text for variable definitions. The estimated order imbalance in this table comes from the BVC (bulk tick test) in Panel A (B). Both panels contain results for SBF-120 members and non-members and those firms with and without active options markets. The t-statistics are generated from standard errors clustered by firm. The aggregated coefficients are weighted by the reciprocal of the squared standard error for each regression specification and the aggregate standard errors are corrected using a method similar to Chordia et al. (2000).

	SBF-120 Member			SBF-120 Non-Member			Options Market			No Options Market		
Bar Size	Adj. R^2	Coeff(γ)	t(γ)	Adj. R^2	Coeff(γ)	t(γ)	Adj. R^2	Coeff(γ)	t(γ)	Adj. R^2	Coeff(γ)	t(γ)
Panel A: BVC OI Estimate Results												
1,000	0.285	0.123	6.40	0.625	0.749	2.33	0.260	0.122	5.84	0.603	0.212	3.21
2,500	0.386	0.176	6.41	0.654	1.055	2.34	0.358	0.176	5.80	0.639	0.280	3.20
5,000	0.443	0.232	6.35	0.691	1.374	2.42	0.415	0.234	5.73	0.677	0.359	3.23
7,500	0.471	0.278	6.31	0.713	1.624	2.54	0.443	0.279	5.69	0.698	0.424	3.26
10,000	0.484	0.314	6.19	0.718	1.753	2.43	0.454	0.316	5.57	0.705	0.465	3.23
30,000	0.546	0.532	6.23	0.796	2.781	2.58	0.510	0.535	5.58	0.779	0.764	3.34
50,000	0.575	0.694	6.18	0.824	2.862	3.04	0.533	0.702	5.53	0.808	0.901	3.48
100,000	0.600	0.995	6.09	0.812	4.958	2.05	0.568	1.008	5.44	0.797	1.320	3.06
250,000	0.640	1.492	6.39	0.807	8.673	1.55	0.592	1.538	5.76	0.799	1.921	2.71
Mean Effect Size		0.201	7.91		1.270	2.25		0.199	7.14		0.340	2.99
Panel B: Bulk Tick OI Estimate Results												
1,000	0.262	-0.042	-11.19	0.479	-0.276	-2.23	0.227	-0.040	-10.64	0.476	-0.105	-3.86
2,500	0.345	-0.037	-9.43	0.636	-0.197	-2.45	0.300	-0.037	-9.07	0.629	-0.065	-4.59
5,000	0.398	-0.039	-7.79	0.676	-0.227	-2.11	0.353	-0.039	-7.50	0.668	-0.062	-3.91
7,500	0.425	-0.041	-7.12	0.690	-0.227	-2.00	0.379	-0.041	-6.90	0.682	-0.063	-3.89
10,000	0.438	-0.045	-6.59	0.702	-0.267	-2.32	0.391	-0.044	-6.30	0.695	-0.067	-3.83
30,000	0.497	-0.066	-5.31	0.786	-0.317	-2.42	0.444	-0.065	-5.06	0.774	-0.076	-3.28
50,000	0.523	-0.074	-4.70	0.806	-0.530	-2.34	0.462	-0.074	-4.54	0.794	-0.089	-2.79
100,000	0.548	-0.110	-4.28	0.795	-0.492	-1.57	0.496	-0.101	-3.85	0.786	-0.120	-2.52
250,000	0.589	-0.133	-2.66	0.792	-1.838	-1.78	0.525	-0.125	-2.33	0.792	-0.238	-2.10
Mean Effect Size		-0.042	-7.81		-0.258	-2.19		-0.041	-7.17		-0.071	-3.61

Table 12

High Low Trading Range Regressions using Most Accurate Volume Bars

This table shows the regression results for the regression of $HL_\tau = \alpha_0 + \alpha_1[HL_{\tau-1}] + \gamma|\widehat{OI}_\tau| + \varepsilon_\tau$. Please see the text for variable definitions. Models 1 and 2 (3 and 4) use the order imbalance estimated using the BVC (bulk tick test). In all regressions, we select the stock-month-bar size for which the BVC is most accurate and use only the most accurate bars for each stock-month as data points. The standard errors clustered by firm are below each coefficient estimate in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

	BVC OI		Tick Test OI	
Variable	(1)	(2)	(3)	(4)
Order Imbalance Estimate	1.820***	1.967***	-0.075*	-0.125**
	(0.251)	(0.277)	(0.040)	(0.056)
High-Low Trading Range $_{t-1}$	0.745***	0.311	0.754***	0.328***
	(0.027)	(0.745)	(0.028)	(0.053)
Constant	-0.201***	-0.471*	0.169***	0.001
	(0.054)	(0.283)	(0.030)	(0.275)
Stock, Bar Size, Month FEs	No	Yes	No	Yes
Number of Observations	23,885	23,885	23,885	23,885
Adj. R-Squared	0.672	0.772	0.6412	0.7369