DISCERNING INFORMATION FROM TRADE DATA

David Easley[1]
dae3@cornell.edu

Marcos López de Prado[2]
marcos.lopezdeprado@
guggemheimpartners.com

Maureen O'Hara[3]
mo19@cornell.edu

March 2015

ABSTRACT

How best to discern trading intentions from market data? We examine the accuracy of three methods for classifying trade data: bulk volume classification (BVC), Tick Rule and Aggregated Tick Rule. We develop a Bayesian model of inferring information from trade executions, and show the conditions in which tick rules or bulk volume classification will predominate. Empirically, we find that Tick rule approaches and BVC are relatively good classifiers of the aggressor side of trading, but bulk volume classifications are better linked to proxies of information-based trading. Thus, BVC would appear to be a useful tool for discerning trading intentions from market data.

Keywords: Trade Classification, Bulk Volume Classification, flow toxicity, volume imbalance, market microstructure.

JEL codes: C02, D52, D53, G14.

---

* The authors have applied for a patent on 'Bulk Volume Classification' and have a financial interest in it.

[1] Scarborough Professor, Department of Economics, Cornell University.
[2] Senior Managing Director, Guggenheim Partners, and Research Affiliate, Computational Research Division, Lawrence Berkeley National Laboratory.
[3] Purcell Professor of Finance, Johnson Graduate School of Management, Cornell University.

1. Introduction

Microstructure models attach a key role to trade data because of its signal value for underlying private information: traders informed of good news will profit by buying while traders informed of bad news will profit by selling. Trade imbalance between buys and sells can also signal liquidity pressure in markets, leading to subsequent price movements. Discerning information from trade data, however, has never been straightforward, and there are a variety of trade classification algorithms in the literature devoted to this task.[4]

The advent of high frequency markets complicates this endeavor in two fundamental ways. First, the mechanics of trading are radically different than in times past. Trading now takes place largely in electronic markets where designated liquidity providers need not be present, and practices such as hidden orders make drawing inferences from market data problematic. In U.S. equity markets, trading is fragmented across 13 exchanges and 40 or more alternative trading venues, each reporting trades to the consolidated tape, but at different latencies. Thus, trades on the tape are out of order, compromising trade classification rules based on up-ticks or down-ticks.[5] Order cancellation rates of 98% or more complicate knowing actual quotes, so trade classification algorithms based on proximity to bid and ask quotes are also severely compromised.[6]

A second, and potentially more serious, problem is that the trading process is fundamentally different. Algorithms chop "parent orders" into numerous "child" orders, so it is order flow rather than individual orders that relate to trade motivation. Trading is also done dynamically, with DMA (direct market access) allowing participants strategically to place

---

[4] See, for example, Lee and Ready (1991), Ellis, Michaely and O'Hara (2000), and Chakrabarty, Moulton and Shilko [2012].

[5] This problem is also particularly acute in the new swap trading markets. Dodd Frank currently requires reporting of non-block trades to the Swap Data Repository but current reporting rules allow a 30 minute delay. So there is no way to determine the correct order of trades.

[6] See Hasbrouck (2013) for an excellent analysis of quote volatility and its implications.

multiple orders at various price levels in the book, monitor the progression of their limit orders in the queue, and cancel and replace orders at different levels. To see why this matters, consider, for example, a trader informed of good news who dynamically trades via limit orders. Rather than hit the ask price to buy, this trader posts an order at the bid. When that order is hit, the trade appears to be a sell as it is taking place at the lower bid price. To continue to buy via limit orders, the informed trader has to post a higher bid or due to time priority his order will simply sit behind the orders at the existing bid. This "persistent bidder" strategy means that prices will be forced up even though the active side of the trade is always the uninformed seller.

To the extent that informed traders use limit orders, the notion of the "active" side of the trade signaling underlying information is undermined because informed orders are not actually crossing the spread. A variety of research [see Bloomfield et al (2005), Hasbrouck and Saar (2009), Bouchard et al (2009); Eisler et al (2012); Kim and Stoll (2013); Collin-Dufresne and Fos [(2014); O'Hara (2014)] suggests that equating "informed traders" with "aggressive traders" is no longer accurate. This has the important implication that buy and sell trades, or imbalances in these trades, may not be good indicators of underlying information.

In this paper, we investigate how best to discern information from trade data. Microstructure research often relies on simple trade classification algorithms to accomplish this task, and we investigate the efficacy of two such approaches for classifying trade data: the bulk volume classification (BVC) methodology, and Tick Rules. The tick rule approaches use simple movements in trade prices (upticks or downticks) to classify a trade as either a buy or a sell. The bulk volume technique, which was first applied in Easley et al (2011), aggregates trades over short time or volume intervals and then uses a standardized price change between the beginning and end of the interval to approximate the percentage of buy and sell order flow. Each of these

techniques maps observable data into proxies for information, but how well any of these approaches works in the new high frequency world is unclear.

To understand the differences between these approaches, it is useful to start from the conceptual framework that there is some underlying unobservable information we care about (buys or sells, or information events) and that we observe data (trade prices) that are correlated with the unobservable information. A Bayesian statistician would start with a prior on the unobservable information, observe the data, and use a likelihood function to update his or her prior to form a posterior on the underlying information. This is not what a tick rule does; instead, it classifies a trade as a buy if the previous price is below the current price, and a sell if it is above. The bulk volume approach, by contrast, can be thought of as assigning a probability to a trade being a buy or sell, an approach closer conceptually to Bayes Rule.

Using a statistical model, we investigate the errors that arise from a tick rule approach and the bulk volume approach, relative to a Bayesian approach. We show that when the noise in the data is low, tick rule errors can be relatively low, and over some regions the tick rule can perform better than the bulk volume approach. When there is substantial noise, however, the bulk volume approach can outperform a tick rule, and permit more accurate sorting of the data. Moreover, our model shows that when order flow is imbalanced (as would be the case, for example, when trades are motivated by private information), tick rules based on noisy data will lead to biased estimates of buy or sell probabilities whereas bulk volume classification will be unbiased.

The underlying information about trading intentions that we seek is not observable, but microstructure theory suggests a variety of proxies for that information. In our empirical work, we test the accuracy of the bulk volume and tick rule approaches using three such proxies: the

4

aggressor side of trading (as given by buy-sell indicator flags in the data), an estimate of spreads, and the permanent price effect of trades. In particular, we show that the BVC and tick rule approaches all do reasonably well in matching the aggressor side of trading, with tick rules generally being more accurate. When we consider spread effects using the Corwin and Shultz methodology, however, we find that the BVC approach dominates tick rule measures, which can have perverse correlations with this information proxy. Similarly, using daily price changes as a reflection of the permanent price effect, we show that here, too, the BVC methodology dominates tick-based measures. We conclude that BVC appears to be a useful tool for market researchers interested in discerning trading intentions.

An interesting upshot of these results is that the aggressor side of trading appears little related to any underlying information – a decoupling we argue that arises from how trading transpires in modern high frequency markets. Our findings here complement recent work by Collin-Dufresne and Fos [2014] who find that standard measures of adverse selection which rely on estimates of the persistent price effects of trades do not reveal the presence of informed trading, where this trading is proxied by Schedule 13D filers. Our analysis shows this difficulty to be more fundamental in that trade information (buys and sells) is also unable to capture informed trading. Both papers highlight the need to consider carefully how well the standard tools of microstructure work in contemporary markets.

As with any research, our analysis has limitations that should be kept in mind when evaluating our results. We use futures data in our empirical analysis because the nature of futures trading is more easily related to the results of our statistical model. The implications of our model should apply to other markets, but implementation issues for the bulk volume approach may be non-trivial in some market settings. Moreover, because the information

motivating trading intentions is unobservable, we have to rely on proxies to capture this underlying information. Microstructure theory provides the basis for these proxies, but both BVC and tick rules rely on variants of price changes which, in turn, can be influenced by market structure.

Our analysis also has strengths that may not be apparent when viewed through the lens of prior research. In particular, high frequency markets today are different and, as discussed in O'Hara (2014), fundamental concepts such as information (is it about the asset's fundamental value? its order flow? the behavior of correlated assets?), orders (are they parent orders, child orders?), and trades (are they independent? correlated from algorithmic programs? arise from pinging?) can have very different meanings. The BVC approach relies on order flows, not individual orders, is agnostic about what the underlying information has to be, and its statistical basis is more forgiving with respect to the data difficulties (i.e. time stamp issues, orders out of sequence, massive data bases) characteristic of modern markets. As such, BVC can be a useful addition to the microstructure tool kit.

The paper is organized as follows. Section 2 provides a statistical analysis of the problem of drawing inferences from trade data, with a particular focus on the accuracy of tick rules and the bulk volume approach. Section 3 sets out the E-mini S&P 500 futures, gold futures, and oil futures data we use in our study, and discusses its characteristics. Section 4 presents our empirical analysis of the accuracy of the tick rule, the aggregate tick rule, and the bulk volume methodology in classifying three proxies for underlying information: the aggressor side of trades, high-low spreads, and daily signed price movements. Section 5 discusses the limitations and applications of our analysis. Section 6 summarizes our conclusions and discusses

the implications of our results for microstructure analyses of high frequency markets. The Appendix provides the BVC algorithm in Python Language.

2.   Discerning information from trade data

Suppose there is some underlying information that we care about and that we observe data correlated with it. A Bayesian approach to the problem of inferring the underlying information from the data requires that we specify the data generating process for the underlying unobservable variables (trades or information events) and then specify the data generating process for whatever we observe (prices), conditional on realizations of the underlying unobservable data. If we knew all of these distributions (or, in principle, even had priors over them), then we could compute the conditional probability of the underlying unobservable variables given the data. Unless the observable data is perfectly informative (or we have a point mass prior), Bayesian calculations would yield probabilities on the underlying unobservable data rather than point estimates. Of course, actually specifying these distributions may prove a daunting task, and even if we had them, computing closed form solutions for conditional probabilities could be complex.

In addition to these potential knowledge and computational difficulties, two other issues arise in applying a Bayesian approach to trading data. First, we may want to know something about the underlying information that generated trades rather than just the trades themselves. To do this using Bayes, we would have to specify the set of possible information processes, how information generates trade, and then how trade generates prices. Second, it is likely that the observations we have are not independent given the underlying unobservable data. Our observations are prices or price changes, and the correlation structure here, too, may be complex.

These difficulties suggest that any practical solution to this problem will involve an approximation. The simplest approximation is the tick rule, which assigns a trade to be a buy if the trade price was an uptick relative to the previous trade and to be a sell if it was a downtick (in the case of a zero-tick the signing relies on the movement relative to the last price change). This approach eschews any distributional assumptions and relies instead on the basic notion that buys raise prices and sells lower them. But how well this approximation works to infer trades, or underlying information, is debatable, particularly in light of the trading practice and market structure issues raised earlier.

If there is noise in the data, the Bayesian approach does not provide a point prediction (of a buy or sell, for example) but rather a posterior probability. This is the intuition that underlies the bulk volume classification algorithm. Our approximation aggregates trades over short time or volume intervals and then uses the standardized price change between the beginning and end of the interval to approximate the percentage of buy and sell order flow. Thus, this approach can be interpreted as assigning probabilities to buys and sells given the observable data. Intuitively, we say that the underlying trade was more likely to have been buyer-initiated the larger, and more positive, is the price change and more likely to have been seller-initiated the smaller, and more negative, is the price change, relative to the distribution of past price changes.[7]

A (time or volume) bar τ is assigned the price change $P_\tau - P_{\tau-1}$, where P_τ is the last price included in bar τ, and $P_{\tau-1}$ the last price included in bar $\tau - 1$.[8] To define the *bulk volume* procedure, let

---

[7] See also Easley, et al. (2012a) where we apply this technique in estimating VPIN measures, and Gollapulli and Bose (2013) who use this approach to estimate order imbalances in swap markets.

[8] We start the first bar with the second transaction in our sample, so that the algorithm has a P_0 for initialization.

$$\hat{V}_\tau^B = V_\tau \cdot t\left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\Delta P}}, df\right)$$ (1)

$$\hat{V}_\tau^S = V_\tau \cdot \left[1 - t\left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\Delta P}}, df\right)\right]$$

where V_τ is the volume traded during (time or volume) bar τ which we wish to classify in terms

of buy and sell volume $\hat{V}_\tau^B$ and $\hat{V}_\tau^S$, and t is the CDF of Student's t distribution, with df degrees of

freedom.[9] $P_\tau - P_{\tau-1}$ is the price change between two consecutive bars and $\sigma_{\Delta P}$ is our estimate of

the standard derivation of price changes between bars. Our procedure splits the volume in a bar

equally between buy and sell volume if there is no price change from the beginning to the end of

the bar. Alternatively, if the price increases, volume is weighted more toward buys than sells

depending on how large the price change in absolute terms is relative to the distribution of price

changes.

2.1 Statistics for Bulk Volume Classification and the Tick rule

Comparing the performance of the tick rule and the BVC approaches is not

straightforward as they do not produce the same type of output. The tick rule produces a list of

buy and sell classifications, one for each trade, whereas BVC produces a list of fractions of buys

and sells, one for each bar (time, volume or trade) to which it is applied. Even on a single bar

(with multiple trades in the bar) they produce different output: the tick rule provides a list of

buys and sells and BVC provides fractions of buys and sells. To compare the two approaches we

consider two transformations. The first, and most obvious, is that we compare the actual tick

rule with an application of BVC to a bar containing a single trade. Applying BVC on a single

trade makes sense if we interpret BVC as assigning to any observation a probability that the

---

[9] We use the t-distribution because the parameters of the true distribution are unknown. Other distributions, such as the Normal or the actual empirical distribution of the data, can be used, but in empirical testing we found no improvement over results from the t-distribution. Based on calibration, we used $df, = 0.25$ to account for the fat tails present in the data.

underlying trade was a buy. In the second transformation we create an aggregate version of the tick rule and compare it to the BVC procedure which is already in a bulk form.

We first show that, even in the single trade case, whether BVC or the tick rule does a better job of trade classification depends on how informative the trade is about the underlying data we want to infer. BVC does better if the observation is very noisy, which seems more likely if we interpret the underlying data as information (good or bad news) and less likely if we interpret the underlying data as a trade (Buy or Sell).

2.2. Observations Classified One-By-One

Suppose that we observe a price change ϵ where the distribution of ϵ differs if the (unobservable) trade type was a Buy or Sell.[10] Assume that $\epsilon \sim d(\bar{\epsilon}, \sigma^2)$ if the trade was a Buy, and $\epsilon \sim d(-\bar{\epsilon}, \sigma^2)$ if the trade was a Sell, where $\bar{\epsilon} > 0$. We denote the prior probability that the unobservable trade was a Buy by $PR(Buy) = p$, where $0 < p < 1$.

[margin note: probability that the news (unobservable) was buy or sell: The price change must be different than the predicted price change (the probability that the price chane e was buy or sell differ from the probability of the news being buy or sell)]

We consider three methodologies to assign a probability that the underlying trade type was a buy or a sell given the observation of a single draw of ϵ : Bayes rule, the tick rule and BVC specialized to a single observation. The tick rule assigns probability one or zero to the trade having been a Buy. BVC when applied to one observation can be interpreted as assigning the probability of a Buy. Bayes rule, of course, actually assigns a probability of the trade having been a Buy. For each methodology, the formula for the conditional probability of a Buy is:

1. Bayes: $B(\epsilon) = \dfrac{pPR(\epsilon \mid Buy)}{PR(\epsilon)}$ where $PR(\epsilon) = pPR(\epsilon \mid Buy) + (1-p)PR(\epsilon \mid Sell)$

2. Tick: $T(\epsilon) = 1$ if $\epsilon > 0$ and $T(\epsilon) = 0$ if $\epsilon < 0$

---

[10] We will refer to trade types and Buy or Sell, but the analysis also applies if we interpret the unobservable event as information which can be good or bad news.

3. BVC: $F(\epsilon)$ where $F(\bullet)$ is the CDF of the unconditional distribution of ϵ which in this example is $D = pd(\bar{\epsilon},\sigma^2) + (1-p)d(-\bar{\epsilon},\sigma^2)$

Bayes is the statistically correct approach, but it requires knowledge of the process by which observations are generated. For each of tick rule and BVC we want to compute the absolute value of the error relative to Bayes for any observation of ϵ. We then compute the expected absolute error using the unconditional distribution of observations. Finding closed form solutions for these errors is not straightforward, so we provide illustrative calculations using a uniform distribution for ϵ.

For Buys we assume that the distribution of ϵ is Uniform on $[-a,b]$ and for Sells that it is Uniform on $[-b,a]$, where $b \geq a \geq 0$ and $b+a=1$. So, if a=0, then observations identify Buys and Sells perfectly; if $a=b=1/2$, then observations contain no information about Buys and Sells. This structure is illustrated in Figure 1.

To simplify the presentation of the calculations, we also assume that $p=1/2$, i.e. Buys and Sells are equally likely. The classification rules for each approach yield the following probabilities that the trade was a buy:

1. Bayes, $B(\epsilon)$:
   a. For $\epsilon \in [-a,a]$, $B(\epsilon)=1/2$
   b. For $\epsilon < -a$, $B(\epsilon)=0$
   c. For $\epsilon > a$, $B(\epsilon)=1$
2. Tick, $T(\epsilon)$:
   a. For $\epsilon < 0$, $T(\epsilon)=0$
   b. For $\epsilon > 0$, $T(\epsilon)=1$
3. BVC[11], $F(\epsilon)$:

---

[11] Note that in our application of BVC to this example of a single trade, we use the actual distribution of price changes. Here the distribution of price changes is an equally weighted mixture of Uniform distributions on [-a,b] and [-b,a].

a. For $\epsilon \in [-b, -a]$, $F(\epsilon) = \dfrac{b + \epsilon}{2}$

b. For $\epsilon \in [-a, a]$, $F(\epsilon) = \epsilon + \dfrac{1}{2}$

c. For $\epsilon \in [a, b]$, $F(\epsilon) = \dfrac{1 + a + \epsilon}{2}$

The errors of each approach relative to Bayes (measured as the absolute value of the difference between the estimate and Bayes) are:

1. Tick

    a. 0 if $\epsilon \in [-b, -a] U [a, b]$

    b. $1/2$ if $\epsilon \in [-a, a]$

2. BVC

    a. $\dfrac{b + \epsilon}{2}$ if $\epsilon \in [-b, -a]$

    b. $|\epsilon|$ if $\epsilon \in [-a, a]$

    c. $\dfrac{b - \epsilon}{2}$ if $\epsilon \in [a, b]$

We are interested in the expectation of these errors under the unconditional distribution of ϵ. For the tick rule, the error is one-half in the interval $[-a, a]$ and this interval has probability $2a$ under the unconditional distribution, so the expected error for the tick rule is a. Note that if $a = 0$, then the expected error for the tick rule is 0 as observations perfectly identify buys and sells. The tick rule error is maximized at $a = 1/2$ when observations have no information content and the tick rule is correct one-half of the time because one-half of the trades are buys.

Remark 1: The expected tick rule error increases with a, the noise in the data, from 0 at $a = 0$ to $1/2$ at the maximum value of $a = 1/2$.[12]

---

[12] The probability of a positive price change given a sell and the probability of a negative price change given a buy are both a.

For BVC the expected error is $\frac{(1-2a)^2}{4}+a^2$. If $a=0$, the bulk volume expected error is

$1/4$. So, if there is no noise in the data, BVC preforms worse than the tick rule. At $a=1/2$ the

BVC expected error is again $1/4$, so in this case it dominates the tick rule.

Remark 2: The expected BVC error is a strictly positive, convex function of the noise in

the data, a, which is minimized at $a=1/4$.

Let $f(a)$ be the expected tick rule error minus the expected BVC error. Calculation

shows that this difference in errors is

$$f(a)=\frac{8a(1-a)-1}{4}$$

This function, defined on $[0,1/2]$ is increasing and concave with $f(0)=-1/4$ and

$f(1/2)=1/4$. The function is 0 at a approximately equal to 0.15. So for $a>0.15$ the BVC

procedure does better than the tick rule, and for $a<0.15$ it does worse.

Remark 3: If the data are not too noisy ($a<0.15$) then the tick rule preforms better than

BVC on a trade-by-trade basis. Alternatively, if the data are noisy ($a>0.15$) then BVC

performs better than the tick rule on a trade-by-trade basis.

2.3 Multiple observations classified by Tick or BVC

The BVC methodology was designed to estimate the fraction of buys in a group of trades;

the tick rule was designed to identify trades on a one-by-one basis. Directly aggregating the

accuracy of the tick rule applied trade-by-trade over a group of trades is possible, and doing so

yields the fraction of trades in the group correctly identified by the tick rule. But comparing this

measure of accuracy of the tick rule with the accuracy of BVC over a group of trades is

problematic as BVC does not identify specific trades as buys or sells. To facilitate comparability,

we create an aggregate version of the tick rule and compare it to BVC.

13

We now extend the simple model to include multiple buckets each containing N observations generated from N underlying Buys and Sells. We suppose that the observations in any bucket are generated independently, given some probability on Buys and Sells, using the distributions from Section 2.1. This assumption does not accurately capture the true trading process which surely features dependence across trades, but we choose it for two reasons. First, as the tick rule works trade-by-trade, in contrast to BVC which applies to groups of trades, it favors the tick rule over BVC. Second, it simplifies the statistical analysis. If, instead, we consider dependence we would have to specify the underlying process in more detail than we know. We begin by considering a bucket with N observations, $\epsilon = (\epsilon_1, \ldots, \epsilon_N)$.

The Tick rule classifies the observation at n as coming from a Buy if $\epsilon_n > 0$ and from a Sell if $\epsilon_n < 0$. The aggregate tick rule counts the number of observations classified by the Tick rule as coming from a Buy for the given ϵ and divides by N to compute the fraction of Buys assigned by the rule. This is an obvious modification of the tick rule, but it is important to remember that its output is a fraction rather than a string of Buy-Sell classifications. We denote this measure $BT(\epsilon)$ and we refer to it as the tick rule since the context will make it clear whether we are using the tick rule trade-by-trade or in aggregate.

BVC assigns a fraction of buys based on the price change from the beginning to the end of the bucket. This is just the sum, $\hat{\epsilon} = \sum_{n=1}^{N} \epsilon_n$. For a given probability of Buys p, each ε_n is distributed as $pd(\bar{\epsilon}, \sigma^2) + (1-p)d(-\bar{\epsilon}, \sigma^2)$. So as N $\to \infty$, $\hat{\epsilon}/N \to p\bar{\epsilon} + (1-p)(-\bar{\epsilon})$ almost surely. Define $o = p\bar{\epsilon} + (1-p)(-\bar{\epsilon})$. For large bucket sizes, i.e. large N, we approximately observe o. If $\bar{\epsilon}$ were known, we could compute an estimate of the probability of Buys, p.

14

The motivation for BVC is to approximate p without reference to distributions of the underlying probabilities of Buys and Sells and without knowledge of the parameters. In our approximation, we use the CDF of a standardized version of the price change and approximate p by the CDF at this standardized value.

To formally define the aggregate version of the tick rule, let o_n be 1 if $\epsilon_n > 0$ and 0 otherwise. Then $BT \equiv \sum_n o_n / N$ is the fraction of buys predicted by the (aggregate version of the) tick rule. If all trades generated price changes of the same absolute size, varying between negative for sells and positive for buys, then the tick rule and bulk volume would be based on monotonic transformations of the same data. For example, suppose that buys yield a price change of Δ and sells yield a price change of $-\Delta$ and that there are B buys and S sells in a bar of size N=B+S. The aggregate tick rule would correctly predict that the fraction of buys is $B/(B+S)$ while the bulk volume statistic would be based on where the standardized value of $\Delta(B-S)$ lies in its CDF. If trade is balanced then both measures predict ½ as the fraction of buys. In our statistical model, price changes are of different sizes so we get differing fractions. In reality, some price changes are zero, some are small and some are large, each associated with different trade sizes, and so the two procedures will yield differing estimates.[13]

Remark 4: If all trades yield price changes of the same absolute size, varying between negative for sells and positive for buys, then the aggregate tick rule and bulk volume statistic both reveal the actual numbers of buys and sells in a bar.

Note that given some probability of Buys p, o_n is iid and its mean is $pq + (1-p)(1-q)$ where $q = PR(\epsilon > 0 | Buy) = PR(\epsilon < 0 | Sell)$ and $1-q = PR(\epsilon > 0 | Sell) = PR(\epsilon < 0 | Buy)$. By

---

[13] We thank Craig Holden for prompting us to think about this point.

the Strong Law of Large Numbers, $BT \to_{a.s.} pq + (1-p)(1-q)$. This limit is not p unless

$p = 1/2$ or $q = 1$. Otherwise, the aggregate tick rule, BT, tends to underestimate p if $p > 1/2$,

and overestimate p if $p < 1/2$. If q is fixed and p varies, the estimates generated by the tick

rule are correct on average (the mean of BT is the mean of p) but they do not vary sufficiently

with the actual value of p.[14]

Remark 5: The aggregate tick rule underestimates the probability of buys when buys are more likely than sells, and overestimates it when buys are less likely than sells.

Remark 5 is useful to think of in the context of market trading. Buys will be more likely than sells when there is new underlying good information (and less likely when there is new bad information). But these are exactly when the aggregate tick rule is biased, suggesting that Bulk Volume classification will be more accurate in times of new information. We now turn to testing how well and when these classification algorithms work in discerning trading information.

3. Data

Testing the empirical accuracy of the bulk volume and Tick rule approaches requires

market data. As we discuss in the next section, we evaluate alternative proxies for the

underlying information we are trying to discern, with one of those proxies being the aggressor

side of the trade. Individual equity data bases such as Nasdaq ITCH data have buy-sell

identifiers for each trade, but standard equity databases such as TAQ do not. Moreover, because

---

[14] If we knew q we could correctly estimate p from BV by computing the quantity $\dfrac{BT - q}{1 - q}$. Of course, using this procedure to estimate the fraction of buys would be contrary to the spirit of the Tick Rule which actually assigns individual trades as Buys or Sells and so produces a direct estimate of the fraction of buys.

equity trading is fragmented across multiple markets and signed data is not available for the vast majority of these market settings, determining the accuracy of classification algorithms in the equity market is a daunting task.[15]

Futures markets also have signed data (for a price), and futures have several advantages relative to equities. Futures trading is not fragmented, with each contract trading only on one market. Our statistical analysis demonstrates how noise affects classification accuracy, and noise in actual markets is greatly influenced by factors such as order matching protocols, book dynamics and liquidity. Observing all trading in a contract allows us to characterize this noise and so test the implications of our model. Another advantage of futures is that all trades must occur either at the best bid or the best offer. This provides the tick rule approaches with the best possible setting as trades between the spread are not allowed.

We chose for our sample three futures contracts: the E-mini S&P 500 future, the Gold future, and the WTI Crude Oil future. These contracts trade in different markets with differing trading volume levels and order book activity. Specifically, the E-mini S&P 500 Futures trades on the Chicago Mercantile Exchange (CME) and is the most actively traded index futures contract, with an average daily volume of 2.2 million contracts. Gold Futures trade on the Commodities Exchange (COMEX) and, while active, their trading volume is approximately one-fifth that of the E-mini. The WTI Crude Oil Futures trades on the New York Mercantile Exchange (NYMEX) and is the most actively traded commodities contract.

3.1 E-mini S&P500 futures

We acquired tick data for the CME E-mini S&P500 Futures contract from November 7th 2010 to November 6th 2011. The database *DataMine Market Depth* provides all messages

---

[15] See Chakrabarty et al (2012) who use INET data to test the accuracy of the Lee-Ready algorithm, and Chakrabarty et al (2013) who test bulk and tick rules using Nasdaq data.

needed to recreate the book and trade data for any CME GLOBEX trade product, time-stamped to the millisecond, following the FIX/FAST protocol.[16]

There are a variety of challenges in working with this data. The data come in a highly irregular format in which a single line can contain an arbitrary number of messages. Among these messages, we find anywhere between 1 and 19 trades per line. Most messages relate to requests to modify or cancel quotes. A trade cannot be identified by any particular FIX tag, but only by a combination of them (for example, when tag 269=2 after another tag 279=0, then tag 270 contains the price, tag 271 the traded size, tag 5797 the aggressor side, tag 52 the UTC transmission time and tag 107 the instrument). Files mix messages from all E-mini S&P500 Futures contracts trading at that time, not only the front contract, requiring care in separating trades from the different expirations. Exchanges also do not always report trades in the sequence they occurred, particularly when their networks are overloaded with dense traffic. Book updates are incremental, so losing or misplacing a message within the sequence of events means that the researcher will not reconstruct the book correctly, a particular problem for Tick-based algorithms. Finally, some reported trades are fictional, and the only way to tell the difference with real trades is by checking the trade time (they are time stamped during periods when the Exchange was actually closed).[17] In short, a complex data handler needs to be programmed to extract the fields we need: Time, Price, Volume, Aggressor and Instrument.

Our final data set contains 128,579,415 e-mini trades. Most trades are small, averaging 4.50 contracts per reported fill. Approximately 52% of trades are for one contract. Because the

---

[16] This protocol receives frequent updates and modifications. In the context of this paper, we will always refer to version 2.19, dated 12/09/11. This level 3 data was purchased directly from the CME, and was delivered as 357 zip files containing 2272 flat files. This represents about 21.6GB of compressed data, and about 220GB uncompressed. We mention these numbers to signal the difficulty of working with this data using standard commercial packages.

[17] Fictional trades can arise as part of the algorithm testing process. Another oddity in the data is that 27,419 trades (or 0.0213% of the total) reported at 4.30pm EST on weekdays and 6pm on Sundays are matched in the opening auction, and therefore have no aggressor flag. These were deleted from our study.

CME applies a FIFO (First In, First Out) matching algorithm for E-mini S&P500 futures, reducing the size of an order does not place it lower in the queue.[18]

3.2 WTI Crude Oil futures

WTI Crude Oil Futures are the most liquid of all crude contracts, and the futures product with the largest volume among all physical commodities. We acquired Level 3 data from NYMEX, from November 28th 2010 to November 27th 2011. Our sample size for the Oil Futures is 78,630,179 signed trades. Trading in the WTI Oil futures contracts has some important differences with the E-mini S&P 500 Futures contract. For example, book dynamics of the WTI contract are quite different, with frequent modifications and cancellations of orders making the WTI book much more volatile than the E-mini's. In our particular sample, there is an average of 17.91 BBO updates for each WTI trade, which is strikingly greater than 3.8 BBO updates for the E-mini. Trade sizes are also different. The larger number of quote updates means that there is greater noise in the trading book, a factor our statistical model says should affect trade classification. The average trade size for the WTI Oil is only 1.9 contracts per reported fill, and 83% of the trades are of size one. The WTI Crude's contract value is typically about 50% more expensive than the e-mini S&P500's, as well as 49% more volatile, so this smaller trade size may reflect the greater costs of transacting in the WTI contract.

3.3 Gold futures

Gold Futures trade at the Commodity Exchange (COMEX). Level 3 tick data was acquired from COMEX, from November 28th 2010 to December 20th 2011. The number of trades in Gold futures is smaller than either the E-mini or oil futures, with a sample size of

---

[18] That is not the case for all CME products. The CME reports the matching algorithm through FIX tag 1142. For instance, CME matches Eurodollar short futures following an Allocation algorithm. This is an enhanced pro-rata algorithm that incorporates a priority (TOP order) to the first incoming order that betters the market. CME follows a Pro-Rata algorithm to match orders on FX Futures Spreads. The CME applies 10 different matching algorithms, depending on the product.

27,960,542 signed trades for Gold Futures. However, the order book in Gold futures is more active with an average of approximately 40 BBO updates per trade. Trade sizes are also smaller, with an average trade size for the Gold contract of 1.6 contracts. Although the Gold contract is about as volatile as the E-mini S&P500, its contract value is typically 50% more expensive than crude's, and almost three times as expensive as E-mini's.

Figure 2 provides plots of the frequency of trades per trade size for each of the three futures contracts. The frequency line quickly decays as a function of trade size, with the exception of round trade sizes (5, 10, 20, 25, 50, 100, 200, etc.). That round trade sizes are much more common than their neighbors may be attributed to so-called 'mouse' or 'GUI' traders, i.e. human traders who send orders by clicking buttons on a GUI (Graphical User Interface). In the e-mini, for example, size 10 is 2.9 times more frequent than size 9. Size 50 is 10.9 times more likely than size 49. Size 100 is 16.8 times more frequent than size 99. Size 200 is 27.2 times more likely than size 199. Size 250 is 32.5 times more frequent than size 249, and size 500 is 57.1 times more frequent than size 499. Such patterns are not typical of 'silicon traders', who usually are programmed to randomize trades to disguise their footprint in markets.

[FIGURE 1 HERE]

3.4 Time or Volume Groupings

One final data issue has to do with how we group trades for empirical analysis. The standard approach to analyzing trades is via time, looking at trades throughout a day or over particular time intervals. A time-based approach will result in an unequal number of trades in each group (a particular amount of elapsed time). Alternatively, we could consider groups based on equal volume of trade, an approach we have argued elsewhere is particularly appropriate for high frequency market settings [see Easley et al (2012)]. Whether time or volume grouping is most

20

useful for estimating the underlying buying pressure or for understanding price movements is an empirical question we consider in Section 4.

A related issues has to do with data size and compression. Researchers using tick algorithms must use data on every trade to determine buys and sells even if they are using the aggregate tick variation. By contrast, the BVC approach needs only data on the volume (or number of trades) in an interval and the beginning and ending price of the interval. Such data is available for researchers using Bloomberg, Tick Writer, and many other commercial data packages. From a computational perspective, using bulk volume with volume or trade bars results in data compression of upwards of 99%, which can greatly reduce both the expense and time involved in processing the data.

4.   **Testing the accuracy of classification approaches**

How well do these algorithms work in discerning information from trade data? As our statistical model demonstrates, any implementable solution to the inference problem confronting traders involves an approximation, so all of these approaches will have classification errors. In this section, we investigate the nature of these errors, using our statistical model to guide our analysis of the comparative accuracy of the bulk and tick approaches.

An immediate issue in designing empirical tests is that accuracy must be determined relative to some benchmark. In times past, this was fairly straightforward – traders wishing to buy (perhaps because they knew good news about the asset) would enter a buy order and it would execute against a passive liquidity provider. The congruity of orders and trades meant that the aggressive side of the trade would also then signal the underlying information motivating the trade. Given this, prior literature measured accuracy as the gap between the buys and sells

identified by the classification algorithm and the actual buys and sells identified by aggressor flags.

Trading now is more complex, and the congruity between trading intentions, orders, and trades is compromised. As noted earlier, algorithmic trading takes parent orders and chops them into child orders, some of which will then turn into actual trades. So orders and trades are no longer congruent. Equally important, these trades may not even appear to have the same underlying trading intention as the parent order. This is because traders using dynamic trading strategies place limit orders to minimize transactions costs, and so avoid crossing the spread. Using 2013 data from VWAP algorithmic orders, O'Hara (2014) shows that 87% of the executed child orders were passive – meaning that a parent order to buy would largely turn into trades classified by the aggressor flag as sales. Thus, trading intentions (and their linkage to new underlying information) may not be well captured by the aggressor flag.

Microstructure theory suggests alternative approaches to capture this linkage. Trade imbalance should impact the level at which market makers are willing to provide liquidity. Due to its linkage to underlying new information, a greater order flow imbalance should lead to greater impacts on trade prices as high frequency market makers adjust their bids and offers.[19] These price impacts can be captured by the high-low spread, so a relevant metric to determine the accuracy of an order flow imbalance classification is to estimate how well it correlates with these spreads. We use the Corwin-Schulz estimator to find these spreads. An advantage of this approach is that it directly controls for the fundamental variance of the asset, allowing us to examine the effects of order imbalance on illiquidity spreads.

---

[19] Note that in high frequency markets "information" is a broad concept and can relate to information on the order flow in the stock, the liquidity and order flow in related stocks or markets, or to fundamental information on the asset value as in times past. This expansion reflects in part that silicon market makers use pre-programmed inventory limits to control risk, so trade imbalances of any sort can have price effects. See O'Hara (2014); Haldane [2012] for more discussion.

Microstructure theory also predicts that new information related to fundamental asset value should have a more permanent effect on prices. This suggests that trade imbalances linked to new information should also be linked to longer-term price changes. So another measure of accuracy it to estimate how well the order flow imbalances identified by the tick rule approaches and by the bulk volume methodology correlate with signed subsequent daily price movements. In our analysis, we use each of these approaches (aggressor flags, spreads, and permanent price impacts) as a proxy for the unobservable underlying information.

4.1 Classification accuracy with respect to aggressor flags

We first consider tick rule accuracy relative to "true" trades identified by the aggressor flag. Suppose, for example, that we have a sequence of buy (B) and sell (S) trades, such as BBSS. If the Tick rule classifies these trades as BSBS, we define its trade-by-trade accuracy ratio to be 50%. Table 1 provides the trade-by-trade accuracy rates for the three contracts.[20]

[TABLE 1 HERE]

We find that the tick rule determines the aggressor's side correctly for 86.43% of E-mini S&P 500 Futures volume, but it is less accurate in gold futures (78.95%), and in oil futures (67.18%). Remark 1 in our statistical model shows that the tick rule should work less well as the data becomes noisier, and we attribute this degradation of performance across contracts to the noise resulting from smaller trading volume, lower liquidity (i.e. thinner books) and greater dynamicity (number of quote changes per fill) in oil and gold futures.

Bulk accuracy is defined as the fraction of overall volume correctly classified within bars. A bar is a collection of trades that occur within a given time period (e.g., one minute) or volume exchanged (e.g. 1,000 contracts). Following our previous example of a sequence of trades

---

[20] We are measuring the amount of volume correctly classified. This is a number close to, but different from, the fraction of trades correctly classified (i.e., ignoring trade size). For example, in the case of the E-mini S&P500 futures, the latter would amount to 86.14% of trades correctly classified by the Tick rule.

BBSS, if these trades occur in one bar and the Tick rule classifies them as BSBS, the bulk

accuracy ratio would be 100% due to the offsetting of classification errors within each bar.

More precisely, we define the bulk accuracy ratio as

$$Ar = \frac{\sum_\tau \left[min\left(\widehat{V_\tau^B}, V_\tau^B\right) + min\left(\widehat{V_\tau^S}, V_\tau^S\right) \right]}{\sum_\tau V_\tau} \tag{2}$$

where (V_τ^B, V_τ^S) is the actual buy and sell-initiated volume in bar τ and $\left(\widehat{V_\tau^B}, \widehat{V_\tau^S}\right)$ is the estimated

buy and sell-initiated volume in bar τ. We use this accuracy measure both for BVC and for the

aggregate version of the tick rule.

BVC uses the standardized price change between the two consecutive intervals to estimate the

percentage of buy and sell volume in an interval. Any sigmoid function, such as the logistic

function or the Gaussian function could be used to standardize price changes [see Franses and

Van Dijk (2000) for multiple financial examples where similar non-linear transformations have

proven to be useful]. Our preference for Student's t rests on practical grounds: It is well-known

and available in most numerical packages. Also, its *df* parameter provides flexibility to model

various levels of non-linearity. Figure 3 illustrates how the smaller the *df* parameter in Student's

t distribution, the less order flow imbalance we associate with small price changes. In the

analysis reported below, we use 0.25 degrees of freedom as it leads to good accuracy rates

(similar results can be obtained, however, for larger or smaller *df*). For each contract, and each

type of bar, we compute the empirical standard derivation of price changes between bars from

the entire sample and use that number in our BVC construction.

We report in Tables 2, 3 and 4 the accuracy ratios for the aggregate tick rule and BVC for

each futures contracts using time bars and volume bars of various sizes. Recall that Remark 3 in

our statistical model showed that BVC should be more accurate than the Tick Rule if noise in the data is sufficiently high.

We highlight three findings in the data. First, accuracy ratios, as expected, are higher for the aggregate tick rule than the tick rule on a trade-by-trade basis. This is due to the offsetting that occurs within bars, and the larger the bar size the greater is the accuracy rate. The improvement in accuracy even at aggregations as small as a few seconds or 1,000 trades illustrates how much noise arises in the data over very short intervals. Second, accuracy ratios are higher for volume bars than for time bars. Evaluating accuracy over volume removes some of the noise that arises from varying trading rates in the market. Early researchers [see, for example, Mandelbrot (1968) or Clark (1973)] argued for using a volume clock to investigate markets, and in more recent work [Easley et al (2012)] we show its importance for high frequency markets. These results suggest that, particularly for intra-day studies, volume aggregation may be a preferred approach.

Third, consistent with the model, BVC accuracy rates are generally higher than the trade-by-trade accuracy of the tick rule. For example, the tick rule rate for Crude Oil Futures is 67.18%, while BVC accuracy rates for Crude Oil range from 65.94% to 81.81% for time bars, and from 86.6% to 91.15% for volume bars (so for every specification except the smallest time bar BVC accuracy beats tick). The results for the Gold accuracy rates are similar in that the accuracy rate for BVC always exceed the tick accuracy rate for volume bars, and BVC beats tick for time bars longer than 10 seconds. For shorter time bars, the advantage shifts to the tick rule. Finally, for the e-mini, all but one BVC volume specification is more accurate, but the time bar results are split: for time bars above 20 seconds BVC is more accurate, while tick rule is more accurate for shorter time bars.

Comparing the accuracy of BVC and the aggregate tick rule, we find that BVC has slightly lower accuracy rates for the e-mini contract and gold. For oil futures, there are specifications in which BVC is more accurate (i.e. for time bars greater than 20 seconds) but for other specifications the aggregate tick rule is more accurate. These results are not surprising given our statistical analysis: when noise is high, BVC can have smaller errors than a tick rule; when noise is low tick rules can be more accurate.

What does seem clear is that BVC can produce reasonably accurate trade classifications relative to the aggressor flag, and it is most useful when there is substantial noise in the data. BVC also requires much less data, and so may be particularly useful for actively traded securities. As the Compression column in the Tables show, bulk volume classification using volume bars requires less than 1% of the data needed to implement tick rules. For many research applications, this data advantage can be a substantial benefit.

This comparability of methods in determining buy and sell accuracy does not mean, however, that the BVC and tick approaches classify the same buy and sell volume for each bar. Consequently, the information in these outputs can be very different. Our statistical model demonstrates why this is the case. When order flow is imbalanced, Remark 5 shows that the Aggregate Tick approach will be biased (understating buys when buys are more likely than sells and overstating buys when sells are more likely). As imbalanced order flow is more likely when there is new information, this suggests that BVC classifications may be better correlated with underlying information. Our second testing approach to characterizing accuracy investigates this possibility.

3.2. Classification accuracy and Spread

In standard microstructure models, a greater probability of information-based trading results in a larger spread. In high frequency markets, however, quoted spreads are problematic as measures of information (or even of prevailing prices) due to a variety of problems such as excessive cancellation rates, binding minimum tick levels, and hidden orders [see Hasbrouck (2013)]. An alternative approach to estimate the presence and impact of information-based trade is to focus on High-Low trading ranges. However, period-by-period High-Low prices are affected by both the illiquidity of the asset and the fundamental variance of the asset. We are interested in isolating the illiquidity component as it is determined by the spread and by price pressure arising from order imbalance.

The Corwin and Schultz (2012) high-low spread estimator was designed to isolate this component of High-Low prices. The Corwin and Schultz estimator is based on their observation (page 719) that "the component of the high-low price ratio that is due to volatility increases proportionately with the length of the trading interval while the component due to bid-ask spreads does not." Corwin and Schultz create high-low spread estimates by sequentially comparing the sum of high-low price ranges over two consecutive intervals with the high-low price ranges over these intervals treated as one (two-period) interval. They then show that these high-low spread estimates correlate well with a variety of spread measures in a variety of settings.

In some futures contracts, like the E-mini S&P500, the bid-ask spread is almost always 0.25, its tick size, even during the illiquid night session. As Corwin and Schultz note (page 721), "the high-low spread estimator captures liquidity more broadly than just the bid-ask spread." It is also affected by price changes due to large orders or to sequences of buys or sells. For our purposes, this is a desirable feature. We are interested in separating underlying volatility from all

of the market microstructure components of high-low prices as we want to use these microstructure components to evaluate various order imbalance measures. A measure of order imbalance that accurately reflects actual order imbalance should at least be positively correlated with the high-low spread estimate regardless of whether they primarily reflect spreads or other liquidity effects.

We begin by estimating high-low spreads using the Corwin and Schultz (2012) technique applied to overlapping bars of size 10,000 for our E-mini S&P500 Futures data. Denote the estimated (Corwin and Schultz) high-low spread for bar τ by S_τ. These estimated high-low spreads are depicted in Figure 4 and summary statistics for them are reported in Table 5. There are 8,283 estimated spreads with a mean spread of 0.0023.

The estimated (absolute value of) order flow imbalance in bar τ arising from an estimate of buy and sell volumes $(\widehat{V_\tau^B}, \widehat{V_\tau^S})$ in bar τ is

$$|\widehat{OI_\tau}| = |\frac{\widehat{V_\tau^B} - \widehat{V_\tau^S}}{V_\tau}| = |2\frac{\widehat{V_\tau^B}}{V_\tau} - 1|. \tag{3}$$

If the estimated buy and sell volumes are generated by the aggregate version of the tick rule we denote the order flow imbalances by $|\widehat{OI_{BT,\tau}}|$; if they are from BVC we use $|\widehat{OI_{BV,\tau}}|$. Summary statistics for these order flow imbalances for the same overlapping bars of size 10,000 for our E-mini S&P500 Futures data as were used to create the Corwin and Schultz high-low spread estimates are provided in Table 5. There are 8,283 estimates for each order imbalance measure and the mean order flow imbalances are 0.0316 for BVC and 0.0414 for BT.

The relationship between order flow imbalance and the high-low spread for the E-mini S&P500 Futures is illustrated in Figure 5. For each decile of the BVC order flow imbalance measure we computed the average high-low spread over the bars in that decile; we did the same

computation for Aggregate Tick Rule order flow imbalance deciles. For comparison, we also computed the order flow imbalance using the aggressor flag data. The figure provides graphs of these average high-low spreads against order flow imbalance deciles.

It is obvious from this graph that these imbalance measures have different relationships with high-low spreads. Most importantly, average spread is approximately constant over aggregate tick rule order flow imbalance deciles one to six, and once aggregate tick rule order flow imbalance becomes large (in the seventh or higher decile) average spread is slightly declining. This suggests a negative relationship between these two variables, and in fact the simple correlation between high-low spread and Aggregate Tick Rule order flow imbalance is -0.086. The relationship between average high-low spread and BVC order flow imbalance is strikingly different. Average spread consistently increases with increasing BVC order flow imbalance deciles. This effect is particularly pronounced for deciles seven and above, which is the opposite of the relationship between spread and tick rule order flow imbalance. As this figure suggests the simple correlation between high-low spread and BVC order flow imbalance is positive (0.068). The relationships illustrated in Figure 5 are consistent with Result 5 from our simple model of trade classification which showed that the aggregate tick rule is biased when true order imbalance is large. This is because tick rule estimates of the fraction of trade based on buying or selling pressure do not vary sufficiently with price change data; the tick rule estimates are too nearly constant when large variations in the probability of buying or selling pressure occur. This would lead to low correlations between estimated Hi-Lo Spreads and Aggregate Tick Rule order flow imbalances----and this is what we see in Figure 5.

What can account for this puzzling behavior? One possibility is that while the tick rule provides a reasonably accurate classification of the aggressor side of a trade, the aggressor side is

not a good indicator of order flow informativeness.[21] The behavior of the aggressor side

imbalances in Figure 5 demonstrates exactly this point. As was the case with the tick rule

imbalances, spreads are not increasing in the aggressor side imbalances, and for very large

imbalances they actually go the "wrong way" (spreads being smaller when imbalances are

larger). This is not a surprising outcome: As large and sophisticated investors rely on smart

execution algorithms that minimize market impact, a burst of aggressive trades within a short

period of time is more likely to come from small or uninformed traders. Such trading, for

example, may arise from one of the most unsophisticated of all execution algorithms: TWAP

(Time Weighted Average Price algorithms) which generally leaves a large "footprint" in the

market [see Easley et al. (2012b) for discussion]. Market makers may be aware of this tendency

for uniformed traders to use more aggressive orders, and as they are more concerned with

persistently imbalanced order flow coming from passive sophisticated traders, they do not widen

spreads in response to it. What matters for our purposes is that that order imbalance created from

bulk volume works, in the sense that it is positively related to the high-low spread, but that order

imbalance created from the tick rule (or even derived directly from the aggressor flag) does not

work. The tick rule is successful in creating a measure of the aggressor side of trading, but it is

not successful in creating a measure of informed trading because the aggressor side of trading

itself is not a good measure of informed trading.

　　　Order imbalances and the high-low spread are time series and they all are positively

serially correlated. To be sure that the seemingly obvious conclusions from Figure 5 and the

simple correlations are not being created by times series effects, we consider the regression

---

[21] In fact, it would seem that aggressors tend to be uninformed since market makers do not react to market imbalances created by aggressors by widening the high-low range.

below with the bar τ -1 spread, $S_{\tau-1}$, included as a regressor to reduce the positive serial correlation of the residuals:

$$S_\tau = \alpha_0 + \alpha_1 S_{\tau-1} + \gamma \left| \widehat{OI}_\tau \right| + \varepsilon_\tau \qquad (4)$$

where $\left| \widehat{OI}_\tau \right|$ is the order imbalance from either the aggregate tick rule or BVC. We expect γ to be positive as it measures the contribution to the change in the high-low spread from τ-1 to τ due to order imbalance at τ. Even after adding the lagged regressor, the estimated residual may not be serially uncorrelated and homoscedastic. For this reason we compute the Newey-West HAC estimates of the regressors in order to determine statistical significance. The Newey-West procedure allows us to determine the optimal number of lags to include in analyzing the correlations structure in the residuals. In particular, following Newey and West (1994), we apply a Bartlett kernel on a number of lags equal to $Int\left[4 \left(\frac{n}{100} \right)^{2/9} \right]$, where n is the total number of observations. For our analysis, we use volume bars of size 10,000 yielding a sample size of 8283. This translates (given the formula above) to an optimal lag structure of 10 lags for the residuals, and this is what we estimate and report in Table 6.

Table 6 reports the result of these regressions using order imbalance created by the aggregate tick rule and by BVC for E-mini S&P500 Futures. If order imbalance is related to underlying information, then spreads should be positively related to order imbalance. Despite the high accuracy ratios achieved by the aggregate tick rule, the estimated coefficient associated with the $\left| \widehat{OI}_{BT,\tau} \right|$ regressor ($\hat{\gamma}$) are significant and negative. In contrast, the coefficients on BVC order imbalance are positive and significant. Thus, aggregate tick order imbalance renders results inconsistent with market microstructure theory, while BVC imbalances are consistent with this predicted information linkage.

Table 6 also reports the result of a regression including lagged high-low spread and both order imbalance measures. Again the coefficients on tick rule and BVC order imbalances have opposite signs. This confirms our hypothesis that, once we take into account the combined market impact from aggressive and passive trades, we have a much better (and consistent) explanatory model of changes in liquidity. To the extent that the high-low spread is a good proxy for the portion of trade arising from informed traders, these results suggest that the tick rule order flow imbalance fails to detect the presence of informed traders, while the BVC order flow imbalance succeeds in doing so.

3.3 Order imbalance and daily price changes

Over short intervals, underlying new information should affect the willingness of market makers to provide liquidity. Over a longer interval, however, market efficiency dictates that new information should affect market prices. This linkage between trading and market efficiency of prices is a fundamental insight of market microstructure research, and it sets the stage for our third accuracy test. If order imbalances are signals of informed trade, then these imbalances should be correlated with price movements.

Testing this proposition using futures data requires careful consideration of some particular features of futures market microstructure. Unlike equity markets which have defined closing times, futures trade on a 24 hour cycle. Trading volume, however, can vary wildly over this interval. In the case of the e-mini S&P future, for example, trading can be frenetic around the open and close of the U.S. equity markets, and somnolent during U.S. overnight hours. Such time patterns might be expected to influence tests relating order imbalances measured using time intervals, but should be of much less importance for imbalances measured over volume increments.

Using data from the e-mini S&P futures, we found that in our sample period the average number of trades per day was approximately 329,000. We calculated order imbalances over volume buckets of size 10,000 using the aggregate tick rule and BVC approaches. For each of the aggregate tick rule and BVC measures we then create a "daily" order imbalance measure by averaging order imbalances over these 30 buckets. We create a "daily" price change (the price change over the volume in an average day) by measuring the price change from the beginning of the 1st bucket to the end of the 30th bucket. We then regressed these "daily" price changes on the "daily" order imbalance measures. Thus we regress "day" t price change on "day" t order imbalance.

The graphs in Figure 6 show the respective performance of the aggregate tick rule order imbalances and BVC imbalances relative to price movements. What is apparent is that BVC imbalances have greater correlation with daily price changes than tick rule imbalances have with daily price changes. The R^2 of the BVC regression is 0.88 while that of the tick rule regression is 0.54. It may seem odd that tick rule order imbalances have any relationship with daily price movements given that they have little correlation to short-term price movements. We believe these results highlight the changing dynamics of high frequency markets. Information still gets into market prices through trading. BVC order imbalances capture the information in order flows, while tick rule imbalances are measures of individual orders. Over time, both matter for the adjustment of market prices, but it is order flows that are more indicative of underlying new information.

5. Limitations, extensions, and applications

Our analysis suggests that bulk volume classification is an accurate and useful technique for discerning trading intentions from market data. As with any approximation, however, BVC

33

has its limitations and in this section we discuss these in more detail. We also consider the broader issues related to implementing this approach over asset classes more generally.

One issue has to do with price changes. Both tick rule approaches and bulk volume classification techniques rely on price changes to classify trades. For tick rule approaches, these price ticks are computed for each trade; for BVC, it is the net price change between the beginning and end of a time bar or volume bar that matters. As discussed in Section 2, under certain conditions these approaches will be identical – the cumulative sum of the individual ticks will equal the net change in prices over the interval (see Result 4). In this case, BVC would be expected to provide no added benefit over using a tick rule approach. But, generally, these two approaches will not be the same and it is useful to understand why this is the case. If trades are of different sizes, or have different price impacts over the day, or if sequences of trades elicit different price responses, then BVC and tick rule estimates will differ. They will also differ if changes in the book can signal new information as this can result in large price changes arising from little trading volume. All of these are features characteristic of high frequency markets. In such settings, adding up individual up ticks and down ticks will not provide the same information as is captured in the price change used in BVC.

Why does this matter? We are interested in this trade data because it tells us about the underlying information motivating trade, and that information is unobservable. So we must look for its reflection in markets – and that generally involves some aspect of price behavior. This is a standard approach in microstructure where tick-signed order imbalance have been shown to relate well to bid and ask price movements, or tick-signed trade price movements decomposed into temporary and permanent price effects are related to inventory and information respectively. The non-linear transformation we apply in BVC is another example of how price-based proxies

suggested by microstructure theory can help discern the presence of information in markets. But it is not tautological – information affects prices and it affects trades, but these effects are not identical as the divergent behavior of tick-based imbalances and BVC imbalances can attest.

There may also be better proxies than buy and sell identifications to capture the information content of trading data. Trade sequences and time patterns, cancellations and additions to the book, and cross market data all seem likely candidates from which to extract useful information. To do so, however, we need to develop models capable of linking these variables to underlying information, and as yet that remains an elusive goal.

Another issue is the applicability of the BVC approach to other markets. While the implications of our model should apply generally to other markets, implementation of the bulk volume methodology may be non-trivial in some market settings. In particular, BVC is a constructed variable requiring estimation of the distribution of price changes over specified intervals. The optimal interval employed will also differ depending upon factors such as trading activity and noise in the data. For some markets, such as futures or FX, where trading is active, implementation is unlikely to be a problem. For equities, however, BVC will have to be implemented thoughtfully, recognizing that the optimal interval is unlikely to be uniform across stocks with disparate trading activity.

Chakrabarty, Pascual, and Shkilko (2013) (CPS) provide some evidence of these difficulties in their analysis of trade classification using Nasdaq ITCH data. They find that BVC is more accurate than the standard tick rule in matching the aggressor flag of trades, the same result we find in futures data. The Aggregated tick rule appears to work better than BVC in matching the aggressor flag, and this difference is particularly pronounced in small stocks, a not unexpected result given the differences we found across futures contracts. The authors also find

that tick rule accuracy declines in the number of trades over short intervals, that tick rule approaches suffer more from hidden volume than does BVC, and that tick rule accuracy has declined over time- all of these factors speak to the growing challenges posed by high frequency trading. These authors do not test for accuracy using any of the other information proxies considered in this paper.

What is not clear is how well tick rules or BVC do when applied to trading in the equity market as a whole where latency issues arise, and aggressor flags are not generally available.[22] Because trades on the consolidated tape are time-stamped when they are received at the SIP and not when they occurred in the market, trades on the tape are out of order. This problem is also particularly acute in the new swap trading markets. Dodd Frank requires reporting of non-block trades to the Swap Data Repository but current reporting rules allow up to a 30 minute delay. So there is also no way to determine the correct order of trades – and no way to use tick rules based on movements from prior trade prices. BVC does not require individual trades, is robust to trades being out of sequence or reported with different latencies, and works better than tick rules when there is noise in the data. For many market settings, it seems likely that BVC will be a useful technique for researchers looking to discern information from trade data.

6. Conclusions

Much of market microstructure analysis is built on the concept that traders learn from market data. Some of this learning is prosaic, such as inferring buys and sells from trade execution. Other learning is more complex, such as inferring underlying new information from

---

[22] CPS attempt to match trades in the ITCH data to the TAQ data, but to do so they assume a 5 second lag between INET trades and TAQ trades– a relative lifetime in today's high frequency markets! The results of Holden and Jacobsen (2013) showing large errors when times stamps are not precise suggests the futility of such a 5 second approach.

trade executions. In this paper, we investigate the general issue of how to discern underlying information from trading data. We examined the accuracy and efficacy of three methods for classifying trades, the tick rule, the aggregated tick rule, and the bulk volume classification methodology. Our results indicate that the tick rule is a reasonably good classifier of the aggressor side of trading, both for individual trades and in aggregate. Bulk volume is shown to also be reasonably accurate for classifying buy and sell trades, but unlike the tick-based approaches, it can also provide insight into other proxies for underlying information.

Our results have a variety of important implications for researchers. For research problems requiring specific identification of individual buy and sell trades, tick rules will be most useful in settings where market microstructure noise is limited. Data bases that provide buy-sell indicators may resolve this problem, but these tend to be expensive and are not available for all markets. For research requiring indications of buy and sell volume imbalance, bulk tick approaches can work reasonably well. Time aggregation, however, leads to lower accuracy rates than does volume aggregation. BVC is generally accurate and has the advantage of requiring substantially less data to implement. For markets such as the newly established swaps trading markets where accurate time-stamped trade data do not exist, BVC can provide a valuable research tool.

For research focusing on discerning underlying information from trading data, bulk volume classification produces consistently better results for trading ranges than tick rule approaches. This reflects the new reality that informed trading is not well captured by the aggressor side of the trade due to advances in algorithmic and high frequency trading strategies. The tick rule attempts to measure buying (or selling) pressure originated by aggressive buyers, but there are at least two additional sources of buying pressure ignored by the tick rule: Persistent

bidders and offer cancellations. Although the tick rule achieves a high accuracy in terms of classifying the side that initiated the trade, we find that the tick rule fails to explain the dynamics of spreads or prices. This suggests that the other sources of buying (or selling) pressure ignored by the tick rule are so important that the tick rule does not succeed in detecting the informational content carried by the order flow. Because BVC uses the market makers' aggregated response to order flow to infer its informational content, it overcomes these limitations of the tick rule and provides a new tool for discerning the presence of underlying information from market data.

REFERENCES

D. Abad and A. Yague, 2012. From PIN to VPIN: An Introduction to Order Flow Toxicity. The Spanish Review of Financial Economics, forthcoming.

Aitken, M. and A. Frino, 1996. The accuracy of the tick test: Evidence from the Australian Stock Exchange. Journal of Banking and Finance, 20, 1715-1729.

Almgren, R. and G. Burghardt, 2011. A window into the world of futures market liquidity. CME Group Research Note, March 30th.

Andersen, T. and O. Bondarenko, 2013. On the Measurement of Order Flow Toxicity: VPIN under Perfect Trade Classification. Working paper.

Ané, T., and H. Geman, 2000. Order Flow, Transaction Clock and Normality of Asset Returns. Journal of Finance 55:2259–84.

Asquith, P., R. Oman and C. Safaya, 2010. Short Sales and Trade Classification Algorithms. Journal of Financial Markets, 13, 157-173.

Bechler, K. and M. Ludkovski (2014): Optimal Execution with Dynamic Order Flow Imbalance, Working Paper, available in arXiv, http://arxiv.org/abs/1409.2618

Bethel, W., D. Leinweber, O. Ruebel and K. Wu, 2011. Federal Market Information Technology in the Post Flash Crash Era: Roles for Supercomputing. Journal of Trading, 7(2), 9-25. Available at http://ssrn.com/abstract=1939522

Bloomfield, R., M. O'Hara, and G. Saar, 2005. The 'Make or Take' Decision in an electronic market: evidence on the evolution of liquidity. Journal of Financial Economics, 75, 165-199.

Boehmer, E., G. Saar, and L. Yu, 2005. Lifting the Veil: An Analysis of Pre-Trade Transparency at the NYSE. Journal of Finance, 60, 763-815.

Bohn, S., 2011. The Slippage Paradox. Centre National de la Recherche Scientifique, working paper. Available at http://arxiv.org/abs/1103.2214

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.

Bouchaud, J.P., Y. Gefen, M. Potters and M. Wyart, 2004. Fluctuations and response.in financial markets: The subtle nature of "random" price changes. Quantitative Finance, 4(2), 176–190.

Bouchaud, J.P., J. Kockelkoren and M. Potters, 2008. Random Walks, liquidity molasses and critical response in financial markets. Quantitative Finance, 6(2), 115-123.

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

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

Cheung, W., R. Chou, A. Lei (2015): "Exchange-Traded Barrier Option and VPIN", Journal of Futures Markets, forthcoming.

Clark, P. K., 1973. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Econometrica 41:135–55.

Chan Wung Kim, Timothy T. Perry, and Manjeet Dhatt (2014): Informed Trading and Price Discovery Around the Clock, Journal of Alternative Investments, http://www.iijournals.com/doi/abs/10.3905/jai.2014.17.2.068

Easley, D., R. F. Engle, M. O'Hara and L. Wu, 2008. Time-Varying Arrival Rates of Informed and Uninformed Traders. Journal of Financial Econometrics.

Easley, D., M. López de Prado and M. O'Hara, 2011. The Microstructure of the Flash Crash. Journal of Portfolio Management, 37(2), winter, 118-128.

Easley, D., M. López de Prado and M. O'Hara, 2012a. Flow Toxicity and Liquidity in a High Frequency World. Review of Financial Studies, 25(5), 1547-1493.

Easley, D., M. López de Prado and M. O'Hara, 2012b. The Volume Clock: Insights into the High Frequency Paradigm. Journal of Portfolio Management.

Easley, D., M. López de Prado and M. O'Hara, 2012c. Optimal Execution Horizon. Mathematical Finance, forthcoming.

Eisler, Zoltan, J.-P. Bouchaud and J. Kockelkoren, 2012. The Impact of order book events: Market orders, limit orders and cancellations. Quantitative Finance, 12(9), 1395-1419.

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), December.

Franses, P. and D. Van Dijk, 2000. Non-Linear Time Series Models in Empirical Finance. Cambridge University Press.

Hasbrouck, J., 2007. Empirical Market Microstructure. Oxford University Press.

Hasbrouck, J., 2013. High Frequency Quoting: Short Term Volatility in Bids and Offers. NYU Working Paper.

Hasbrouck, J. and G. Saar, 2009. Technology and Liquidity Provision: The Blurring of Traditional Definitions. Journal of Financial Markets.

Holden, C. and S. Jacobsen, 2013. Liquidity Measurement Problems in Fast, Competitive Markets: Expensive and Cheap Solutions. Journal of Finance, forthcoming.

Kim, S. and H. Stoll, 2013. Are Order Imbalances Related to Information? Working Paper, Vanderbilt University.

Lee, C. and M. Ready, 1991. Inferring Trade Direction from Intraday Data. Journal of Finance, 46, 733-746.

Mandlebrot, B., and M. Taylor, 1967. On the Distribution of Stock Price Differences. Operations Research 15(5):1057–62.

Menkveld, A. J., and B. Z. Yuesheng, 2013. Anatomy of the Flash Crash. Working Paper, VA University of Amsterdam.

Mike, S. and J. Farmer, 2008. An Empirical Behavioral model of Liquidity and Volatility. Journal of Economic Dynamics & Control, 32, 200-234.

Mosteller, F., 1965. Fifty challenging problems in probability with solutions. Dover Publications, pp.79-80.

Newey, W. and K. West, 1994. Automatic Lag Selection in Covariance Matrix Estimation. Review of Economic Studies, 61, 631-653.

Odders-White, E., 2000. On the Occurrence and Consequences of Inaccurate Trade Classification. Journal of Financial Markets, 3, 259-286.

O'Hara, M., 1995. Market Microstructure Theory. Blackwell.

O'Hara, M., 2014, High Frequency Market Microstructure, Working Paper.

Parkinson, M., 1980. The Extreme Value Method for Estimating the Variance of the Rate of Return. Journal of Business, 53, 61-65.

Song, J, K. Wu and H. Simon: Parameter Analysis of the VPIN (Volume Synchronized Probability of Informed Trading) Metric. Constantin Zopounidis, Editor. Quantitative Financial Risk Management: Theory and Practice. 2014. Wiley. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2427086

Wei, W., D. Gerace and A. Frino, 2013. Informed Trading, Flow Toxicity and the Impact on Intraday Trading Factors. Australasian Accounting Business and Finance Journal, 7(2).

Yildiz, S., R. Van Ness and B. Van Ness, 2013. Analysis of the Determinants of VPIN, HFTs' Order Flow Toxicity and Impact on Stock Price Variance. Working Paper, University of Mississippi. Available at http://faculty.bus.olemiss.edu/rvanness/Speakers/Presentation%202013-2014/Yildiz-VPIN.pdf

APPENDICES

A.1. TICK-RULE IMPLEMENTATION

Here we present a simple implementation of the tick rule in Python language. More

efficient implementations exist, but we believe the one outlined below is the clearest. *queryCurs*

is assumed to contain the output of a SQL query such as

queryCurs.execute('SELECT Price, Volume, VolBuy FROM ' + tablename + ' ORDER BY Instrument, Time')

VolBuy is the field that stores the Volume from traders initiated by an aggressive buyer,

as reported by the Exchange. The tick list variable will accumulate the amount matched over the

entire volume. The rest of the code is self-explanatory.

```
a =queryCurs.fetchone()
flag, price, tick=1, a[0], [0,0]
while True:
   try:
     a=queryCurs.fetchone()
     if a==None:break
     # tick rule
     if a[0]>price:
       flag=1
     elif a[0]<price:
       flag=2
     if flag==1:
       tick[0]+=a[2]  #correctly classified as buy
     else:
       tick[0]+=a[1]-a[2]  #correctly classified as sell
     tick[1]+=a[1]  #volume to be classified
     # reset price
     price=a[0]
```

A.2. BULK VOLUME CLASSIFICATION IMPLEMENTATION

An equivalent codification of the bulk volumeC algorithm would be as follows. *stDev* is

a real variable storing the volume weighted Standard Deviation of price changes across bars.

The amount matched over the entire volume is stored in the list variable *bulk*.

```
a =queryCurs.fetchone()
price, bulk=a[0], [0,0]
while True:
    a=queryCurs.fetchone()
    if a==None:break
    # bulk classification
    z=float(a[0]-price)/stDev
    z=scipy.stats.t.cdf(z,df)
    bulk[0]+=min(a[1]*z,a[2])  #correctly classified as buy
    bulk[0]+=min(a[1]*(1-z),a[1]-a[2])  #correctly classified as sell
    bulk[1]+=a[1]  #volume to be classified
    # reset price
    price=a[0]
```

Table 1
Individual Accuracy Rates Delivered by the tick rule

Contract	# Trades	Ind. Accuracy
E-Mini S&P500	128,579,415	86.43%
WTI Crude Oil	78,630,179	67.18%
Gold	27,960,542	78.95%

Table 1 reports the number of trades and the trade-by-trade accuracy rates of the Tick rule for the three futures contracts we consider: the E-Mini S&P 500 futures, the WTI Crude Oil futures and the Gold futures contract.

Table 2
Classification Accuracy Rates for E-Mini S&P500 Futures
for Time Bars and Volume Bars

Time bar size	Aggregate Tick Accuracy	BVC Accuracy	# Points	Compression
1	87.01%	70.22%	8,732,609	93.21%
2	87.40%	73.52%	5,978,503	95.35%
3	87.75%	75.64%	4,658,935	96.38%
5	88.35%	78.49%	3,312,454	97.42%
10	89.49%	82.38%	1,994,549	98.45%
20	90.94%	85.92%	1,145,907	99.11%
30	91.86%	87.75%	813,401	99.37%
40	92.53%	88.90%	634,046	99.51%
50	93.04%	89.71%	521,566	99.59%
60	93.45%	90.32%	443,659	99.66%

Volume bar size	Aggregate Tick Accuracy	BVC Accuracy	# Points	Compression
1000	89.88%	82.58%	579,315	99.55%
2500	92.37%	88.28%	231,724	99.82%
5000	94.19%	91.22%	115,861	99.91%
7500	95.10%	92.40%	77,240	99.94%
10000	95.68%	93.08%	57,929	99.95%
12500	96.08%	93.53%	46,342	99.96%
15000	96.38%	93.80%	38,619	99.97%
17500	96.61%	94.07%	33,102	99.97%
20000	96.83%	94.24%	28,963	99.98%
25000	97.13%	94.47%	23,170	99.98%

Table 2 reports the aggregate trade classification accuracy percentages using the E-Mini S&P 500 futures contract for the Aggregate Tick rule and the Bulk Volume procedure for various sizes of time and volume bars. This table also reports the number of data points used by the Bulk Volume procedure and the compression (in percentages) of the original data set represented by this number of data points.

Table 3
Classification Accuracy Rates for WTI Crude Oil Futures
for Time Bars and Volume Bars

Time bar size	Aggregate Tick Accuracy	BVC Accuracy	# Points	Compression
1	70.40%	65.94%	15,163,245	82.78%
2	71.36%	68.43%	11,813,733	86.58%
3	72.09%	70.00%	10,019,471	88.62%
5	73.22%	72.08%	7,993,451	90.92%
10	75.07%	74.97%	5,685,630	93.54%
20	77.23%	77.80%	3,892,042	95.58%
30	78.61%	79.38%	3,068,056	96.52%
40	79.60%	80.40%	2,573,556	97.08%
50	80.40%	81.20%	2,237,372	97.46%
60	81.05%	81.81%	1,990,779	97.74%

Volume bar size	Aggregate Tick Accuracy	BVC Accuracy	# Points	Compression
1000	87.42%	86.86%	162,095	99.82%
2000	89.83%	88.82%	81,037	99.91%
3000	91.04%	89.62%	54,016	99.94%
4000	91.81%	90.12%	40,509	99.95%
5000	92.36%	90.42%	32,398	99.96%
6000	92.77%	90.68%	26,999	99.97%
7000	93.12%	90.82%	23,139	99.97%
8000	93.37%	91.00%	20,246	99.98%
9000	93.61%	91.09%	17,991	99.98%
10000	93.81%	91.15%	16,189	99.98%

Table 3 reports the aggregate trade classification accuracy percentages using the Crude Oil futures contract for the Aggregate Tick rule and the Bulk Volume procedure for various sizes of time and volume bars. This table also reports the number of data points used by the Bulk Volume procedure and the compression (in percentages) of the original data set represented by this number of data points.

Table 4
Classification Accuracy Rates for Gold Futures for Time Bars and Volume Bars

Time bar size	Aggregate Tick Accuracy	BVC Accuracy	# Points	Compression
1	79.34%	64.95%	7,109,395	75.65%
2	79.90%	67.92%	5,486,413	81.21%
3	80.37%	69.83%	4,564,913	84.37%
5	81.13%	72.38%	3,503,755	88.00%
10	82.45%	75.88%	2,323,971	92.04%
20	83.99%	79.09%	1,474,103	94.95%
30	84.94%	80.73%	1,116,926	96.17%
40	85.60%	81.80%	915,253	96.87%
50	86.12%	82.60%	784,174	97.31%
60	86.53%	83.16%	691,196	97.63%

Volume bar size	Aggregate Tick Accuracy	BVC Accuracy	# Points	Compression
1000	91.03%	88.87%	47,588	99.84%
2000	92.32%	90.62%	23,786	99.92%
3000	93.00%	91.51%	15,853	99.95%
4000	93.48%	91.90%	11,887	99.96%
5000	93.78%	92.31%	9,508	99.97%
6000	94.07%	92.56%	7,920	99.97%
7000	94.29%	92.66%	6,785	99.98%
8000	94.50%	92.72%	5,935	99.98%
9000	94.68%	92.95%	5,277	99.98%
10000	94.75%	93.00%	4,748	99.98%

Table 4 reports the aggregate trade classification accuracy percentages using the Gold futures contract for the Aggregate Tick rule and the Bulk Volume procedure for various sizes of time and volume bars. This table also reports the number of data points used by the Bulk Volume procedure and the compression (in percentages) of the original data set represented by this number of data points.

Table 5

Corwin-Schultz High-Low Spreads and Order Imbalances

for Bulk Volume and Bulk Tick

STATS	Corwin-Schultz Spreads	OI_BV	OI_BT
# Points	8283	8283	8283
Min	8.580E-08	1.752E-07	1.220E-05
Max	1.770E-02	2.086E-01	1.682E-01
Mean	2.324E-03	3.164E-02	4.142E-02
1Q	1.065E-03	1.270E-02	1.773E-02
2Q	2.011E-03	2.661E-02	3.566E-02
3Q	3.173E-03	4.505E-02	6.006E-02
StDev	1.705E-03	2.445E-02	2.991E-02
Skew	1.507E+00	1.217E+00	8.629E-01
Kurt	4.913E+00	2.199E+00	4.154E-01

This table provides summary statistics for estimated spreads (using the Corwin and Schultz procedure) and for order imbalances computed from bulk volume and from the bulk tick rule. The data set used is our E-mini S&P500 Futures data using trade bars of size 10,000.

Table 6
Order Imbalance and Spread for the E-Mini S&P500 Futures

A. Aggregate Tick Rule (Adjusted $R^2 = 0.191$)

$$S_\tau = \alpha_0 + \alpha_1[S_{\tau-1}] + \gamma|\widehat{OI}_{BT,\tau}| + \varepsilon_\tau$$

Variable	Coefficients	tStats	NW_StDev	NW_tStats
Intercept	0.001374	36.89408	5.18E-05	26.5124165
logTrOI	-0.00155	-2.75918	0.0005739	-2.7069934
Spread_1	0.43641	44.16817	0.02403119	18.1601326

B. Bulk Volume Order Imbalance (Adjusted $R^2 = 0.196$)

$$S_\tau = \alpha_0 + \alpha_1[S_{\tau-1}] + \gamma|\widehat{OI}_{BVC,\tau}| + \varepsilon_\tau$$

Variable	Coefficients	tStats	NW_StDev	NW_tStats
Intercept	0.001165	34.18198	5.35E-05	21.7648921
logBvOI	0.005238	7.565329	0.00082849	6.32187338
Spread_1	0.427476	43.04452	0.02466698	17.3298908

C. Bulk Volume and Aggregate Tick Rule Order Imbalance (Adjusted $R^2 = 0.201$)

$$S_\tau = \alpha_0 + \alpha_1[S_{\tau-1}] + \gamma_1|\widehat{OI}_{BVC,\tau}| + \gamma_2|\widehat{OI}_{BT,\tau}| + \varepsilon_\tau$$

Variable	Coefficients	tStats	NW_StDev	NW_tStats
Intercept	0.001284	33.65692	5.30E-05	24.2134749
logBvOI	0.007593	9.849579	0.00094606	8.02593829
logTrOI	-0.00429	-6.86749	0.00065649	-6.5389723
Spread_1	0.420701	42.27098	0.02438932	17.2494049

This table reports the result of regressions of the Corwin-Schultz spread on its lagged value and on order imbalance created by the aggregate tick rule (panel A), the Bulk Volume procedure (panel B) and on both variables for the E-mini S&P500 Futures using volume bars of 10,000 shares. The sample size is 8230 bars, and the optimal number of lags for the error terms was determined from applying a Bartlett kernel using the Newey-West (1994) procedure.

Figure 1

Price Change Model for Uniform Distributions
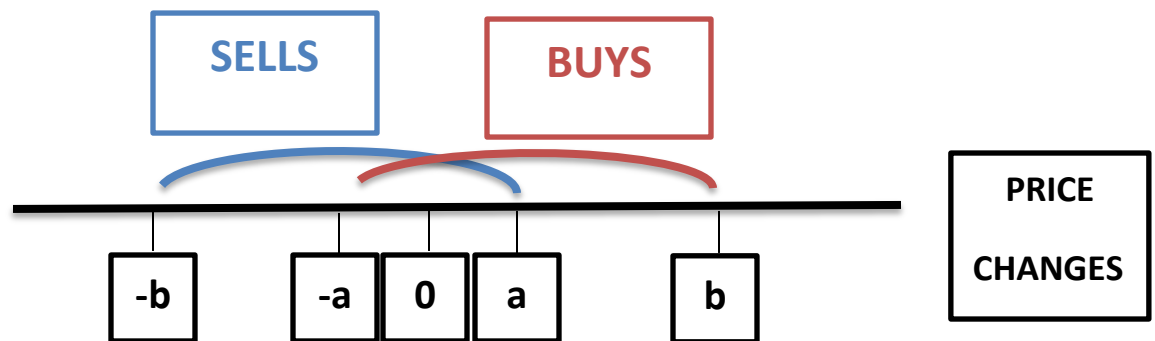
If a=0 then price changes are perfectly informative about trades.
If a=b then price changes have no information content.

Figure 2
Percentage of trades as a function of the trade size (in log scale)
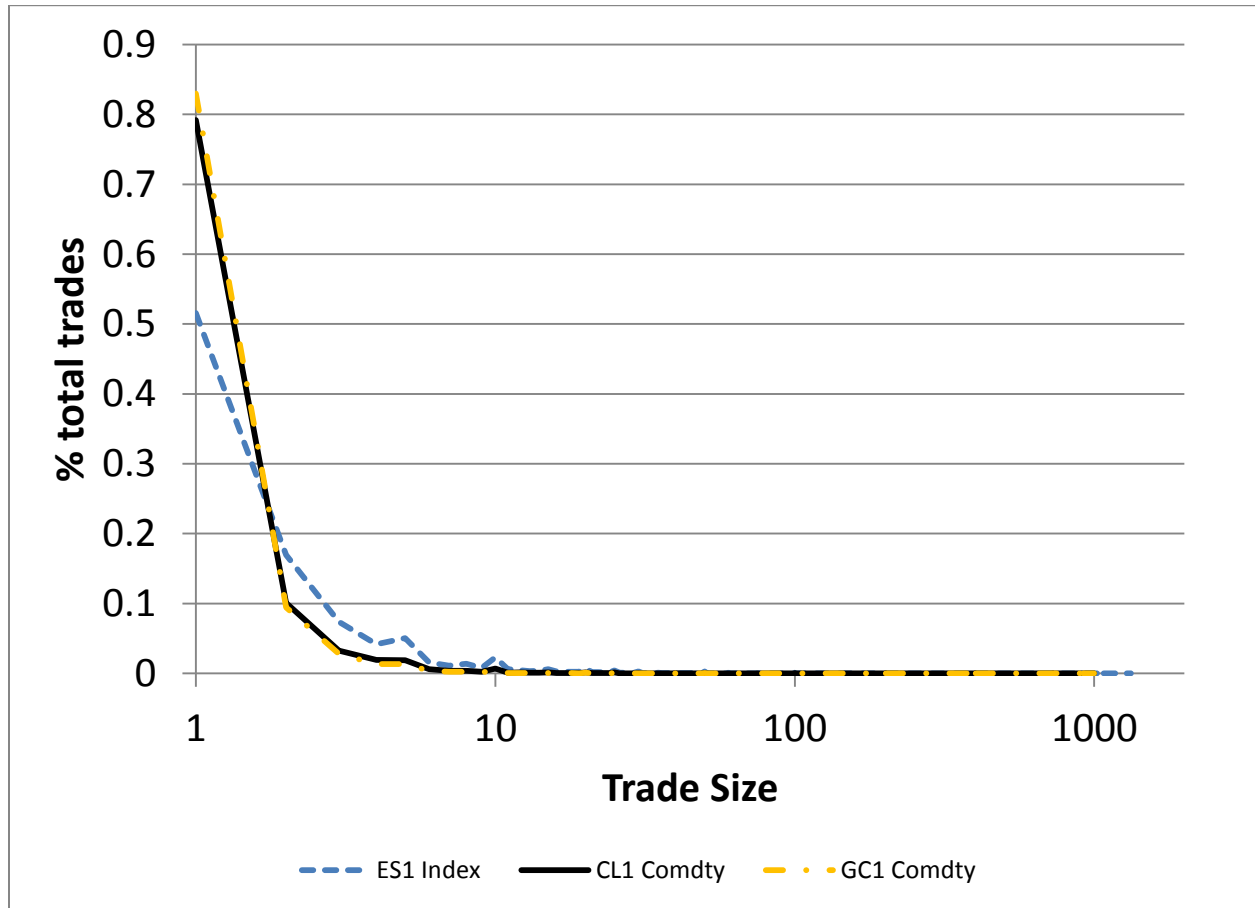**for E-mini S&P500 futures, WTI Oil Futures, and Gold Futures**

Figure 2 plots the frequency of trades per trade size for each of the three futures contracts: the E-Mini S&P 500 futures, the WTI Crude Oil futures and Gold futures. The frequency c quickly decays as a function of trade size, with the exception of round trade sizes (5, 10, 20, 25, 50, 100, 200, etc.). For all contracts at least one-half of trades are for only one contract.

Figure 3
Non-Linear transformation performed through a t-Distribution,
for various degrees of freedom
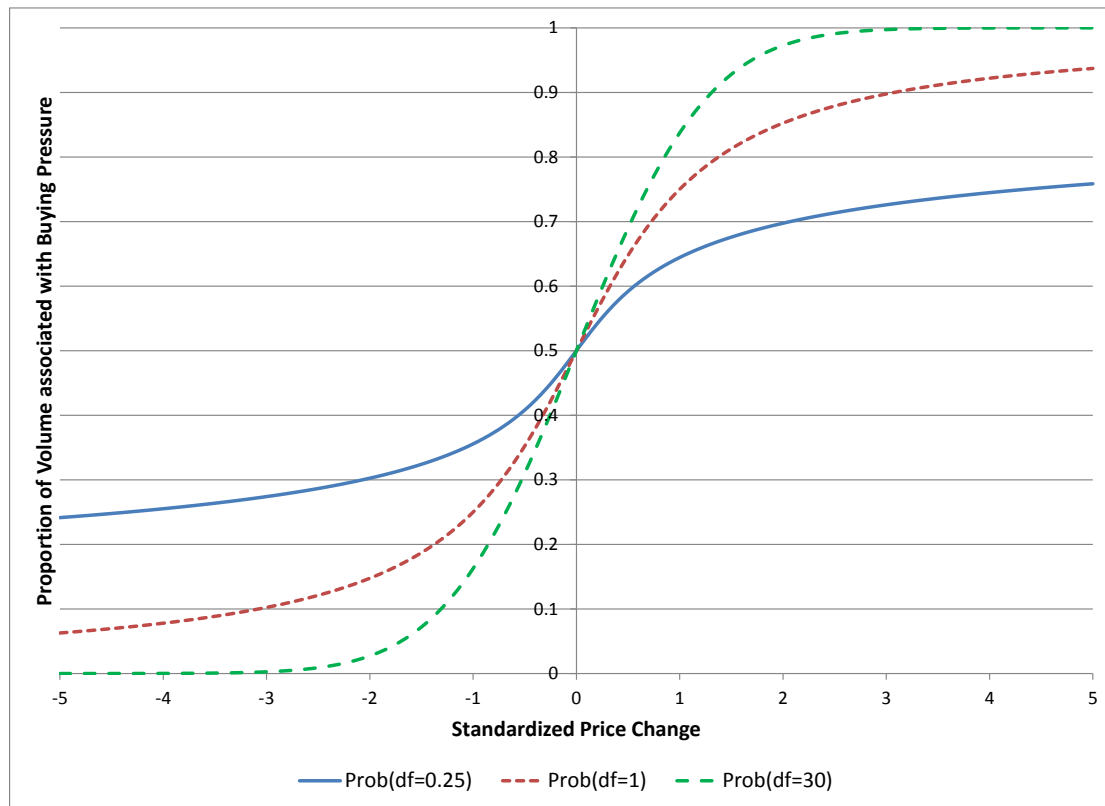
Figure 3 illustrates the effect of degrees of freedom (df) on the CDF of the Student's t distribution. The smaller the *df* parameter in Student's t distribution, the lower order flow imbalance we associate with small price changes.   Our empirical work uses df=0.25.

Figure 4

High-Low Spreads Calculated by the Corwin-Schultz Technique
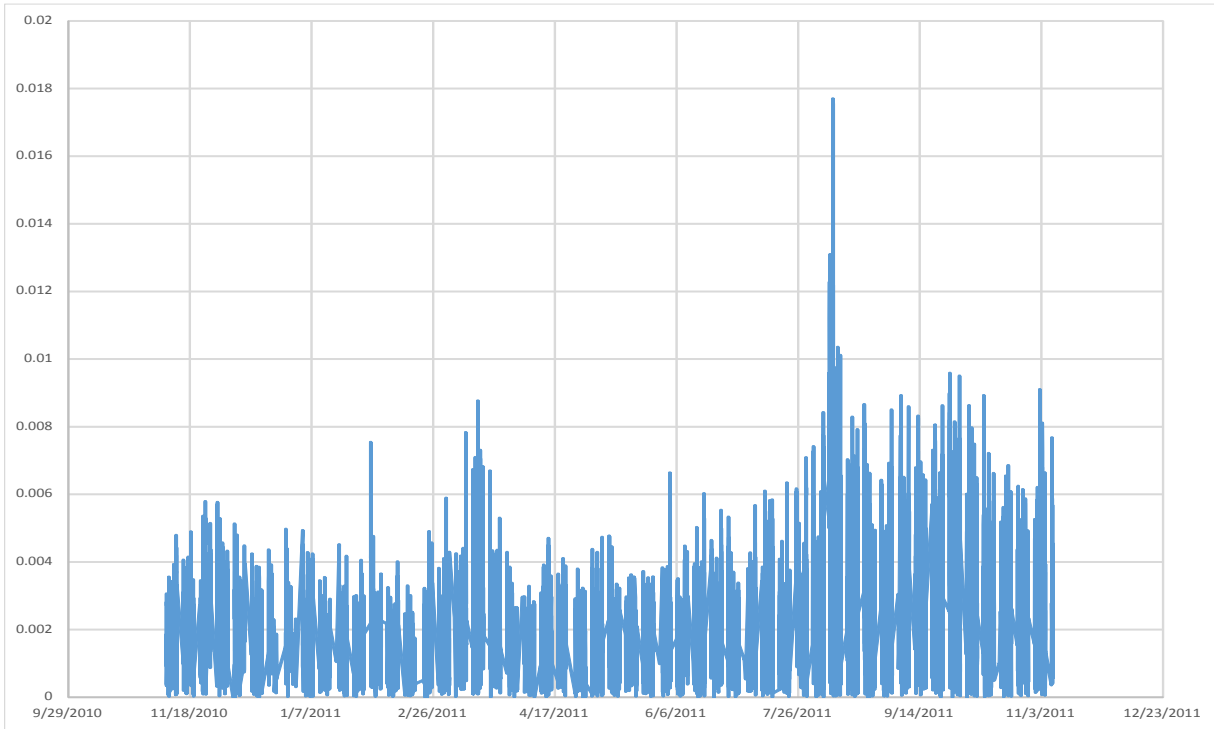
This figure depicts the Corwin and Schultz high-low spreads for volume bars of size 10,000 for our E-mini S&P500 Futures data. There are 8,283 estimated spreads with a mean spread of 0.0023.

Figure 5
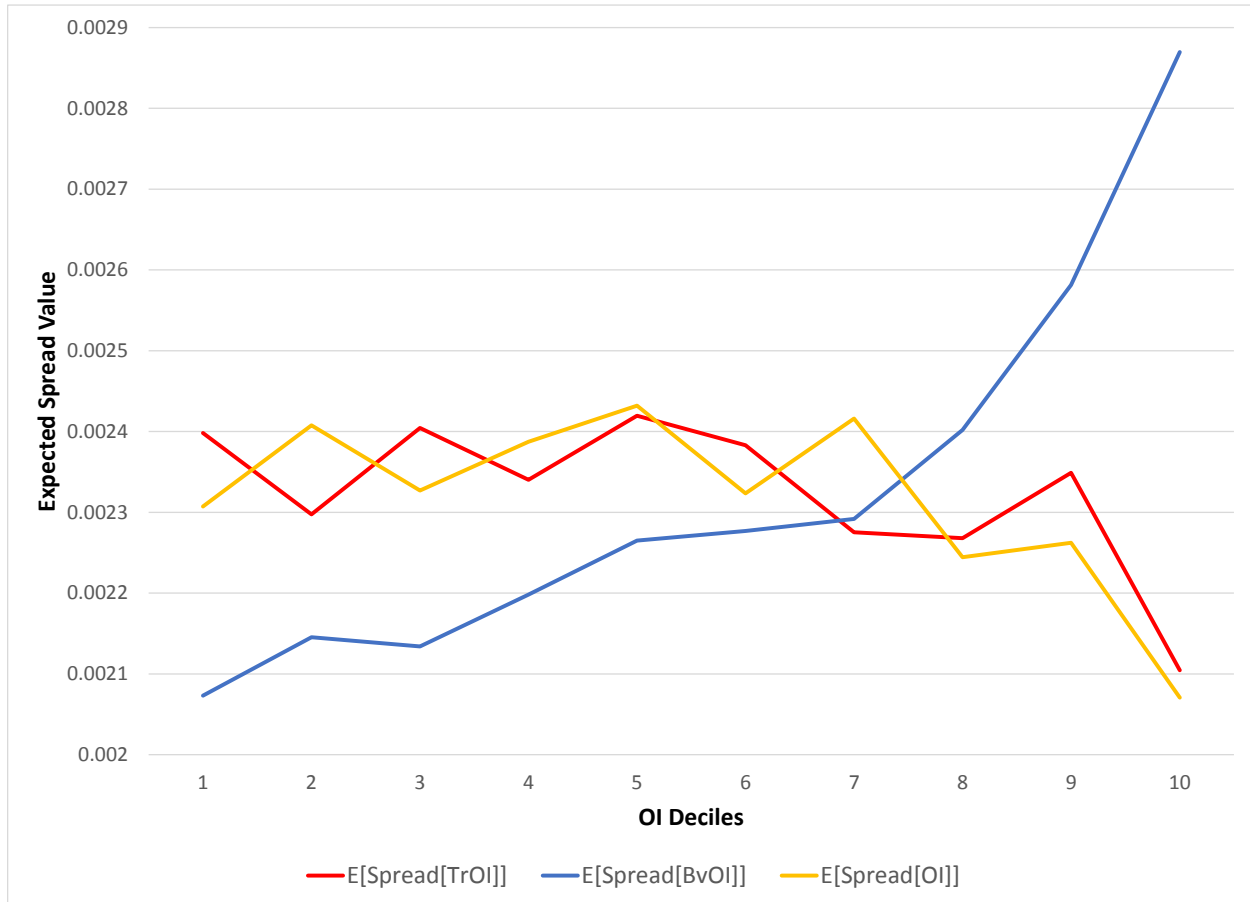**High-Low Spread Graphed Against Tick Rule and BVC Order Flow Imbalances**

Figure 5 illustrates the relationship between order flow imbalance and the estimated Corwin-Schultz spread for the E-Mini S&P futures. The red line shows the estimated spread for each Tick-rule order imbalance decile. The yellow line plots the estimated spread for each actual order imbalance decile (as reported by the exchange's aggressor flag). The blue line shows the estimated spread for each Bulk-Volume order imbalance decile.

Figure 6
Daily Price Changes and Order Flow Imbalances in E-mini S&P Futures

A. BVC Order Imbalance and "Daily" Price Changes
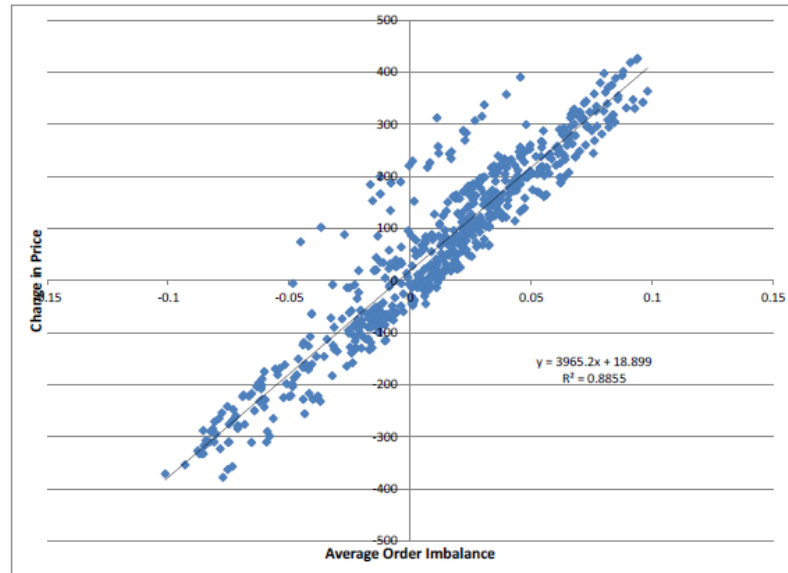
B. Tick Rule Order Imbalance and "Daily" Price Changes
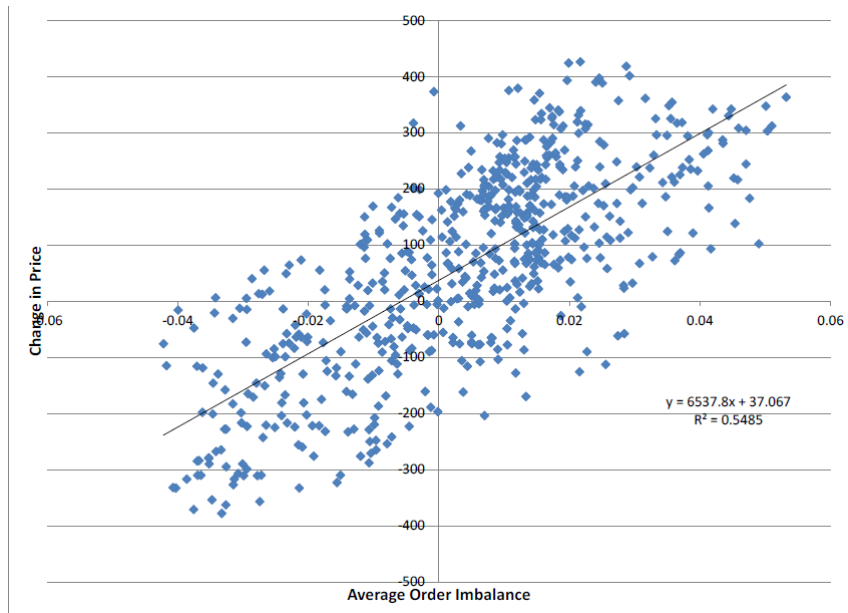

Figure 6 illustrates the relationship between daily price changes for the E-Mini S&P 500 futures contract and order imbalances computed using bulk volume (panel A) and the aggregate tick rule (panel B). In both graphs volume buckets of size 10,000 were used.