VPIN and the Flash Crash*

Torben G. Andersen† and Oleg Bondarenko‡

May 2013 (first version: February 2011)

Abstract

The Volume-Synchronized Probability of Informed trading (VPIN) metric is introduced by Easley, López de Prado, and O'Hara (2011a) as a real-time indicator of order flow toxicity. They find the measure useful in monitoring order flow imbalances and conclude it may help signal impending market turmoil, exemplified by historical high readings of the metric prior to the flash crash. More generally, they show that VPIN is significantly correlated with future short-term return volatility. In contrast, our empirical investigation of VPIN documents that it is a poor predictor of short run volatility, that it did not reach an all-time high prior, but rather after, the flash crash, and that its predictive content is due primarily to a mechanical relation with the underlying trading intensity. We also investigate a later incarnation of VPIN, stemming from Easley, López de Prado, and O'Hara (2012a), and reach similar conclusions. In general, we stress that adoption of any specific metric for order flow toxicity should be contingent on satisfactory performance relative to suitable benchmarks, exemplified by the analysis we undertake here.

JEL Classification: G01; G12; G14; G17; and C58

Keywords: VPIN; PIN; High-Frequency Trading; Order Flow Toxicity; Order Imbalance; Flash Crash; VIX; Volatility Forecasting

*We are indebted to the Zell Center for Risk at the Kellogg School of Management, Northwestern University, for financial support. We are grateful to the referee, Pete Kyle, for inquisitive comments on the earlier version of the paper that strengthened the empirical evidence, and to the editor, Tarun Chordia, for the help in obtaining details of the BV-VPIN implementation. We also thank Craig Furfine, Kathleen Hagerty, Andrei Kirilenko, Robert McDonald, Maureen O'Hara, Mark Ready, and seminar participants at the University of Illinois at Chicago, the Commodity Futures Trading Commission, the Federal Reserve Bank of Chicago, the High-Frequency Trading Leaders Forum, 2011, the Duke University Economics Department Brown Bag and the Kellogg Finance Department Brown Bag for discussions on this topic. Andersen also acknowledges support from CREATES funded by the Danish National Research Foundation. Finally, we are grateful to the CME Group for providing access to data from the CME DataMine system.

†Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208; NBER, and CREATES; Phone: (847) 467-1285; e-mail: t-andersen@kellogg.northwestern.edu

‡Department of Finance (MC 168), University of Illinois at Chicago, 601 S. Morgan St., Chicago, IL 60607; Phone: (312) 996-2362; e-mail: olegb@uic.edu

1. Introduction

In a series of articles, Easley, López de Prado, and O'Hara, henceforth ELO, (2011a, 2011b, 2011c, 2012a) develop the "Volume-Synchronized Probability of Informed trading" (VPIN) metric as a proxy for the imbalance or "toxicity" of order flow. The construction of VPIN relies on an underlying trade classification scheme, and this choice has implications for the properties of the measure. In the initial papers, ELO use a version of the tick rule to classify trades into buy and sell volume, and we denote any metric based on this procedure TR-VPIN for "Tick Rule-VPIN." In ELO (2012a), they instead advocate a "bulk volume" classification strategy, and we refer to the associated metric as BV-VPIN. Another important feature is that VPIN captures the market dynamics in event time, i.e., equal increments of trading volume rather than calendar time. Hence, their analysis uses a transformed time scale where the basic unit is a fixed volume bucket rather than a constant stretch of calendar time. They find their VPIN implementation to produce a set of striking empirical results, using one-minute observations for the order flow on the E-mini S&P 500 futures contract at the Chicago Mercantile Exchange.

ELO (2011a) focus on the events surrounding the "flash crash" on May 6, 2010. First, they note that the TR-VPIN measure was unusually high in the week preceding May 6, 2010, and the situation worsened in the hours prior to the crash. In fact, they observe that the TR-VPIN metric for the E-mini S&P 500 futures contract reached an all-time historical high by 13:30 local Chicago Time, and the crash began at 13:32 according to the time line established by CFTC-SEC (2010). Second, they find that the TR-VPIN measure leads the Volatility Index (VIX) for the S&P 500 index, both prior, during, and following the dramatic events of May 6, 2010. As such, they suggest TR-VPIN provides a superior and more timely indicator of future short-term volatility, or emerging turmoil, than the option-implied volatility measure, VIX, which is otherwise often referred to as the "market fear" gauge.

The findings reported by ELO raise the prospect that TR-VPIN may serve as a reliable indicator of stress in the financial markets, thus providing regulators, brokers, and traders alike with a real-time warning signal of market malfunction. To allow the broader public access to this information in a timely fashion, they advocate introducing an exchange-traded futures contract written on TR-VPIN.

In this article, we take an in-depth look at the empirical performance of TR- and BV-VPIN applied to the E-Mini S&P 500 futures contract. We initially focus on first variant, TR-VPIN, and develop an empirical framework for assessing the properties of this metric.[1] Even within this set of measures, there are numerous alternative classification strategies. We document that the results hinge critically on the choice among those. We reach four main conclusions that, on key points, diverge from ELO. One, TR-VPIN is not a useful predictor for future return volatility. Traditional forecast variables, including the VIX index, are generally vastly superior to TR-VPIN, even for very short horizons. Two, TR-VPIN is, by construction, mechanically related to the underlying trading intensity and its predictive content is subsumed by that of the trading pattern. Three, TR-VPIN did not attain a historical high *prior* to the flash crash, but only after it subsided. In fact, reconstructing the real-time information available prior to the crash, we find little evidence that TR-VPIN would have alerted an observer of a sharply rising probability of an impending market collapse. Four, TR-VPIN is subject to considerable idiosyncratic sampling noise due to dependence on the point at which the volume clock is initiated.

Although sampled according to a volume clock, the TR-VPIN metric of ELO (2011a, 2011b, 2011c) is highly correlated with trading intensity. This stems from the use of time bars in aggregating individual transactions into blocks of volume. Within a time bar all trades are jointly classified as (active) buys or sells so, effectively, they are treated as a single transaction. When trading is intense,

[1] The algorithm for computing TR-VPIN, detailed in ELO (2011c), was submitted to the U.S. Patent and Trademark Office. For a discussion of potential contract design for a TR-VPIN futures contract, see ELO (2011b).

each time bar contains a lot of volume and the number of time bars used for constructing TR-VPIN shrinks. This, in turn, inflates the order imbalance measure, independently of the actual order flow imbalance. Quantifying this effect, we find almost all systematic variation in TR-VPIN to be explained by the heterogeneity in the trading pattern. Trade classification plays a minimal role.

Of course, the trading pattern is endogenous and may respond to, and reflect, the underlying order flow imbalances so the above does not necessarily imply that TR-VPIN is uninformative. However, trading intensity has long been known to covary with, and contain predictive content for, a variety of other activity variables, including return volatility. Moreover, the trading intensity is readily observed so, to assess the incremental contribution of TR-VPIN, it is critical to disentangle the information conveyed by the metric from that associated with the trading pattern. Towards this end, we explore a couple of alternative measures, designed to neutralize the confounding impact of the time bar and to identify the dependence on the trade classification scheme. For instance, using a *fixed* volume bin rather than a time bar in measuring order imbalances provides a pure trading time based metric for VPIN, while *randomizing* the trade classification annihilates the effect of systematic order imbalances.

For fixed volume bin VPIN measures, we find a negligible link with trading volume and, strikingly, a pronounced *negative* association with volatility. Thus, once we annihilate the mechanical link between volume and TR-VPIN, the correlation with volatility reverses sign. Likewise, using an alternative control for the trading pattern, the contribution of TR-VPIN to short-term volatility prediction, over and beyond the component explained by the trading pattern, is negative. Moreover, we reach similar conclusions using a VPIN measure based on the standard tick rule from (non-aggregated) transaction data. Finally, on the day of the flash crash, both the transaction-VPIN and fixed volume bin VPIN measures rise prior to the crash, but do not reach extreme values around the crash. Hence, our study raises serious questions about the reliability of TR-VPIN for assessing order flow toxicity. In particular, the fixed volume bin approach is, theoretically, more in line with the volume clock advocated by ELO, so it is troubling that associated metric reverses all main findings obtained via TR-VPIN.

We also experiment with a *signed* VPIN measure which allows order flow imbalances to offset over time. This approach reduces the idiosyncratic noise and appears helpful in detecting the momentum in the order flow around the flash crash. Nonetheless, we caution against the adoption of any specific metric, unless it retains significance in formal tests which incorporate readily observable real-time market activity measures, such as trading intensity, implied volatility measures, and the like. The latter is necessary to gauge the incremental information content of any new metric.

Our analysis should be useful in rationalizing the behavior of VPIN measures more generally. As mentioned, ELO (2012a) propose a different trade classification strategy, but otherwise compute VPIN as before, generating the BV-VPIN metric. We explore whether our findings regarding TR-VPIN – coupled with the features introduced by the shift to the BV scheme – provide insights into the properties of this newer metric.[2] We confirm that the qualitative behavior of BV-VPIN is consistent with our analytical framework, further corroborating our empirical findings.

The remainder of the paper is structured as follows. Section 2 verifies that we obtain results comparable to ELO (2011a) when exploiting their TR-VPIN metric. Section 3 introduces alternative ways of constructing TR-VPIN style measures. Section 4 explores the properties of TR-VPIN and identifies the source of mechanical correlation with trading intensity. Section 5 presents empirical results based on the full sample. Section 6 introduces BV-VPIN and explores the properties of this measure. Section 7 revisits the flash crash and reviews the evidence through the lens of alternative TR- and BV-VPIN measures. Section 8 concludes.

[2]Our analytic framework was developed, and put in writing, prior to the appearance of the initial working paper introducing the BV-VPIN strategy. In this respect, the exploration of BV-VPIN is an "out-of-sample" exercise.

2. A first look at VPIN and the flash crash

2.1 Data

Our study is based on transaction data for the E-mini S&P 500 futures contract over the January 2008 through July 2010 period. The E-mini contract is traded exclusively in a fully electronic limit order book market on the Chicago Mercantile Exchange (CME) Globex trading platform. Our data were extracted from the Time & Sales series obtained from CME DataMine and include the full sequence of trades consummated over the given period, along with the time (in seconds), the price, and the number of contracts exchanged for each transaction. This series covers a period similar to the one explored by ELO (2011a), although it does not include the last few months of their sample. It is important that the underlying transaction data are comprehensive. We have confirmed that the number of contracts traded during the regular trading hours on May 6, 2010, match the figure reported by Kirilenko et al. (2011).[3] Moreover, our series contains a slightly larger trading volume than what is employed by ELO, who rely on a different source for their data. Hence, our transaction series provides a comprehensive account of the trading activity in the E-mini S&P 500 futures contract.

2.2 One-minute TR-VPIN and the flash crash

Before proceeding, we first confirm that our data series is compatible with the one used by ELO and, in particular, that we obtain similar evidence regarding TR-VPIN on May 6, 2010. Hence, we construct TR-VPIN according to the algorithm in ELO (2011c). For this purpose, we aggregate our transaction series into one-minute observations, or "time bars," containing the last recorded price and the cumulative trading volume. This generates our version of the one-minute bars used by ELO.[4]

Figure 1 depicts the S&P 500 futures price, the VIX, the daily maximum value of the one-minute time-bar TR-VPIN measure, and, for later reference, the corresponding daily maximum value of the one-minute time-bar BV-VPIN metric. It is evident that TR-VPIN spikes to an all-time high on May 6, 2010. The only other day displaying a similar type of spike is June 6, 2008, but it does not attain the level reached on the day of the "flash crash." We also note that VIX jumps on May 6, 2010, but it remains well below the values observed during the financial crisis of 2008-2009.

Figure 2 offers a detailed look at May 6, 2010. It shows the extremely rapid drop of the equity index level during the flash crash and the equally dramatic recovery. This development was accompanied by an escalation in trading activity and a quick run-up in the VIX measure. Finally, we confirm that TR-VPIN had been rising steadily throughout the day, increasing from below 0.40 in early trading to about 0.53 just prior to the crash – largely mimicking the evolution portrayed in Exhibit 5 of ELO (2011a). Comparing Figure 2 to the graphs in ELO (2011a), there is a close correspondence between all main qualitative features. Hence, we are able to replicate the primary characteristics of the ELO study for the overall sample, as well as for this critical day.

Importantly, however, there is one key result, cited in ELO (2011a), that we cannot confirm. We summarize this point as our first "finding":

Finding 1: The level of TR-VPIN just prior to the flash crash is elevated, but it is not at a historical high. The TR-VPIN metric only achieves a historical high after the crash has subsided.

This result is corroborated by the bottom left panel of Figure 1 where the higher dashed horizontal line indicates the level 0.53 – the value TR-VPIN attains just prior to the flash crash. The TR-VPIN

[3] This study explores features of the flash crash using a more detailed audit-trail data set for the transactions in the E-mini S&P 500 futures contract during regular trading hours over May 3-6, 2010.

[4] We do not provide details about the construction of TR-VPIN here, as they are covered at length below.

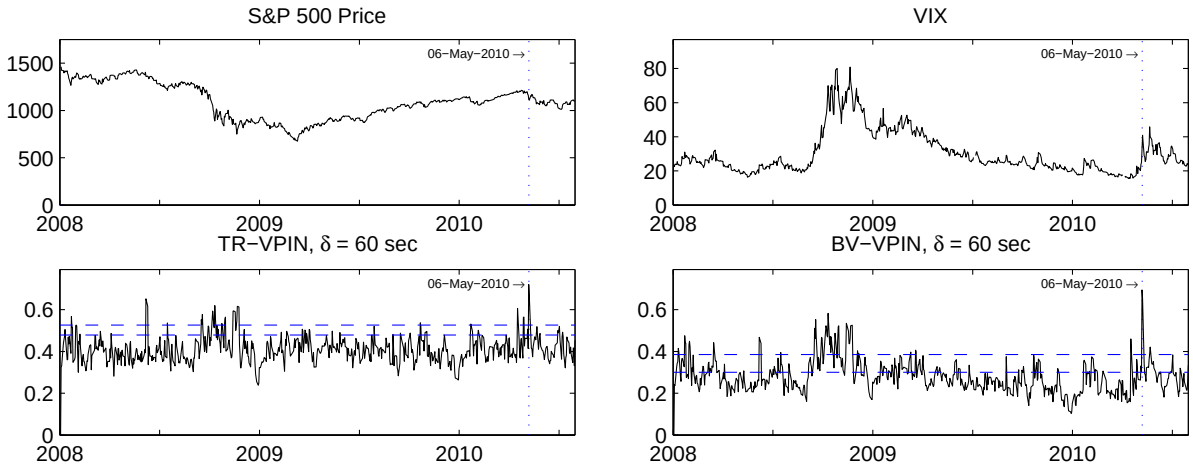

Figure 1: **The evolution of S&P 500, VIX, TR-VPIN, and BV-VPIN.** The figure depicts daily values of the S&P 500 index, VIX, TR- and BV-VPIN for $\delta = 60$ seconds from January 2008 through June 2010. The horizontal dashed lines show the level of VPIN measures on the Flash Crash day at 12:30 (the lower one) and 13:30 (the higher one). Prior to May 6, 2010, TR-VPIN (VB-VPIN) exceeds the 13:30 dashed line on 26 (49) separate days, constituting 4.3% (8.1%) of the days prior to the crash.

series exceeds this level on 26 separate days, prior to May 6, 2010, during our sample, constituting 4.3% of the days prior to the crash. In other words, according to our, admittedly short, historical series, one would expect to observe the value, attained by TR-VPIN prior to the flash crash, about once every month. Since one main motivations behind the development of the VPIN metric is to provide a warning signal for an impending market disruption, this point is important.[5]

In light of this observation, the behavior of TR-VPIN prior to the crash may be noteworthy, but it is not exceptional. Even more importantly – as we document later – other market variables were behaving in an even more unusual fashion prior to and during the crash. The key question will be what incremental information is conveyed by the TR-VPIN metric. However, before turning to our empirical investigation, we need to explain how TR-VPIN is constructed.

3. Constructing the TR-VPIN metric

3.1 Data aggregation and trade classification

Our study is based on transaction data extracted from the Time & Sales files for the E-mini S&P 500 futures contract, covering a sample period we denote $[0,T]$. Each transaction, or tick, is represented by the triplet (t_i, p_i, s_i), where t_i indicates the time of transaction i, p_i denotes the price at which the contracts were traded, and s_i denotes the size of transaction i, expressed in terms of the number of futures contracts exchanged. Transaction times are measured in seconds and form a non-decreasing sequence $0 \leq t_1 \leq t_2 \ldots \leq T$. While many trades may occur within the same second, we know the order

[5]ELO (2011a) state (using Eastern Time): "By 2:30 p.m., the VPIN metric reached its highest level in the history of the E-mini S&P 500. At 2:32 p.m., the crash began, according to the CFTC-SEC Report time line." Similar statements made headlines in a number of prominent media outlets. However, whether the claim refers to the maximum of VPIN for the full sample or to the running maximum prior to the flash crash, the claim is misguided.

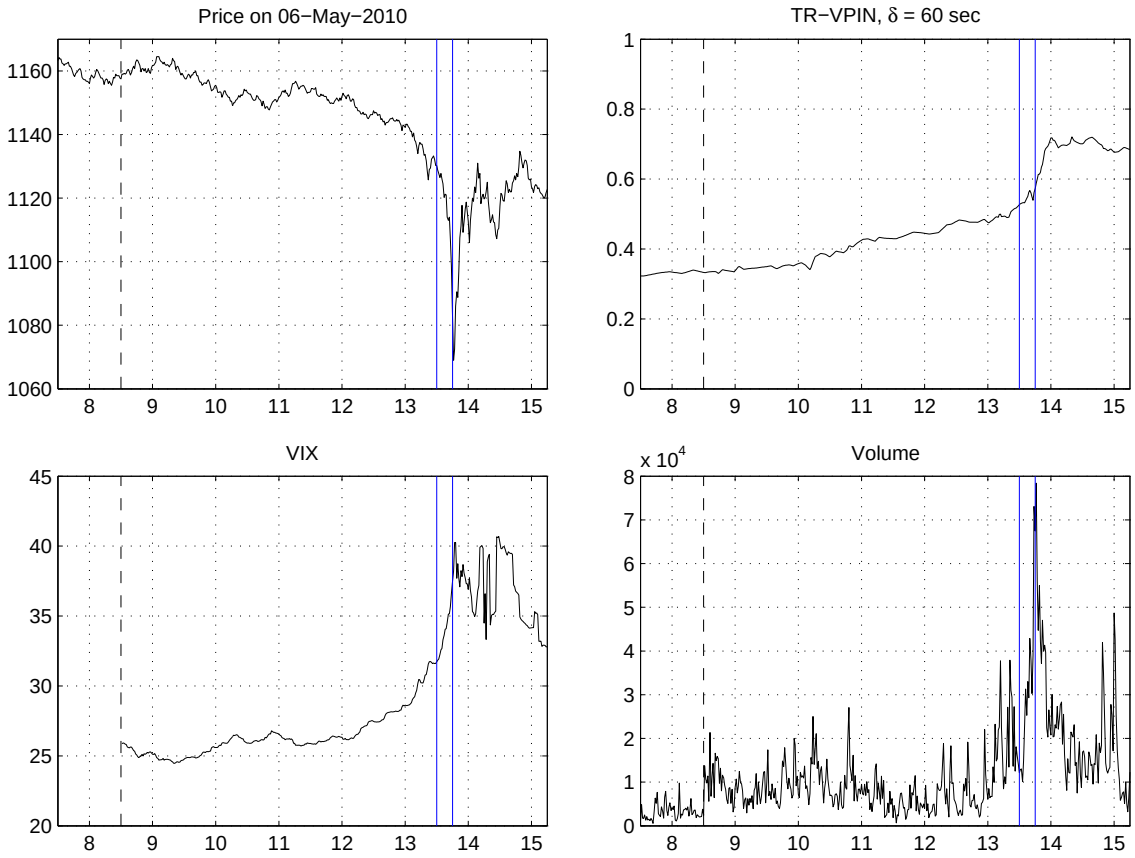

Figure 2: **S&P 500, TR-VPIN, VIX, and trading volume on May 6, 2010.** The figure depicts minute-by-minute data for the S&P 500 futures index level, the TR-VPIN measure constructed from one-minute data, the volatility index, VIX, and the volume of traded contracts of the S&P 500 E-mini futures on the CME, for May 6, 2010. The dashed vertical line shows the start of the regular trading hours, while the solid vertical lines indicate the timing of the "flash crash."

in which they were executed via an associated transactions sequence indicator. As such, we have a complete transactions history for the contract over the sample period.

ELO (2011c) argue that trade classification is error prone at the transaction level. Hence, they focus on transaction data aggregated into time bars. However, there are many different ways in which to aggregate transactions data and we also explore a procedure that combines the underlying transactions into blocks with an equal trading volume (the same number of traded contracts). Consequently, we adopt the notation to accommodate such different types of aggregation.

First, we define the time bars. We let $0 = T_0 < T_1 < T_2 \ldots$ represent an *equally*-spaced calendar time grid with fixed time step δ, so that $T_j = \delta \cdot j$. The empirical analysis in ELO exploits one-minute bars, or $\delta = 60$ seconds, although they consider alternative values. Correspondingly, we focus on the one-minute time bars, but we explore the sensitivity of the results to variation in the degree of time aggregation. Once the time bars are defined, the data may be represented by the triplets (T_j, P_j, N_j), where P_j is the last transaction price prior to time T_j, and N_j is the total number of contracts traded over the time interval $[T_0, T_j)$, for $j = 1, \ldots, J^\delta = T/\delta$. Note that N_j denotes the *cumulative* volume traded by the end of the j^{th} time bar. Sometimes, it might be more convenient to represent the aggregated data

by the triplets (T_j, P_j, n_j), where $n_j = N_j - N_{j-1}$ denotes the volume within the j^{th} bar.

Following ELO, we sign the entire order flow in each time bar using a classification scheme akin to the tick rule. The binary variable $b_j = \pm 1$ indicates whether the contracts exchanged within the j^{th} time bar are labeled *buyer* or *seller* initiated. For $j = 2, \ldots, J^{\delta}$, this indicator is defined as,

$$b_j = \begin{cases} 1, & \text{if } P_j > P_{j-1}, \text{ or } P_j = P_{j-1} \text{ and } b_{j-1} = 1 \\ -1, & \text{otherwise.} \end{cases} \tag{1}$$

This rule ascribes a price increase (decrease) over a given time bar to buying (selling) pressure and classifies the full transaction volume during this bar as active buying (selling) volume. While this will mis-classify some transaction whenever a time bar contains both active buys and sells, ELO deem this approach, based on aggregated order flow, superior to assigning trade direction based on an actual tick rule, where each individual transaction is classified according to the price change from one tick to the next. If we need to be explicit about the choice of time grid, we can write $(T_j^{\delta}, P_j^{\delta}, N_j^{\delta})$ and b_j^{δ}.

Second, we define the volume bins. We let $0 = N_0 < N_1 < N_2 \ldots$ represent an *equally*-spaced grid with the fixed volume step ν, so that $N_k = \nu \cdot k$ and $n_k = N_k - N_{k-1} = \nu$. Our data are now represented by triplets (T_k, P_k, N_k), where T_k and P_k are, respectively, the calendar time and trade price associated with the last contract included in volume bin k for $k = 1, \ldots, K^{\nu}$, and K^{ν} is the number of complete non-overlapping volume bins of size ν in the sample. As for classification rule (1), the binary variable b_k classifies the entire order flow within the k^{th} volume bin, $k = 2, \ldots, K^{\nu}$. If we need to be explicit about the volume grid, we write $(T_k^{\nu}, P_k^{\nu}, N_k^{\nu})$ and b_k^{ν}. Furthermore, transaction data may be seen as a limiting case with a bin size of only one contract, i.e., $\nu = 1$.

In summary, the classification rule (1) assigns a buy or sell indicator to each transaction throughout the sample, but they are bundled into sequences of unidirectional buys and sells according to the time bar or volume bin they reside in. Whether this rule dominates alternative classification schemes for constructing proxies for order flow imbalances over longer time spans is in part an empirical question. We provide evidence for alternative values of δ or ν below.

3.2 The volume clock, the OI measure, and TR-VPIN

Following ELO, we now introduce a volume-based time-scale transformation. Rather than monitoring the market dynamics in calendar time, we employ a volume bucket, V, defined as a fixed number of traded contracts. Thus, each bucket represents an equidistant increment to trading volume, but potentially highly varying periods of calendar time. Throughout our analysis, we set $V = 40{,}000$ futures contracts, corresponding roughly to $(1/50)^{\text{th}}$ of the average daily trading volume. This choice mimics the leading case adopted by ELO.

Each volume bucket comprises a set of aggregate transaction triplets, each representing a different block of trading volume. For underlying triplets defined from one-minute bars, we end up splitting bars, that comprise transactions straddling adjacent volume buckets, into separate pieces so that each fraction belongs to a unique bucket. We generically denote the number of blocks within a volume bucket by Q, and we define the relative size of each trade block in the bucket as $w_q = n_q/V$ for $q = 1, \cdots, Q$, where $n_q = N_q - N_{q-1}$ indicates the number of contracts traded in block q. We obviously have, $0 \leq w_q \leq 1$ and $w_1 + \cdots + w_Q = 1$. Moreover, since every trade is classified as a buy or sell, we may define V^B and V^S as the number of contracts classified as bought and sold, respectively, over the volume bucket, so that, $V^B + V^S = V$.

Utilizing the binary trade indicator, we construct the *signed* order imbalance measure, SOI,

$$SOI \;=\; w_1 b_1 + \cdots + w_Q b_Q \;=\; \frac{V^B - V^S}{V} \;=\; \frac{V^B - V^S}{V^B + V^S}. \tag{2}$$

The focus of ELO (2011a) is on the *absolute* order flow imbalance relative to the total volume for the given bucket. They define their order imbalance measure as,

$$OI \;=\; |SOI| \;=\; \frac{|V^B - V^S|}{V} \;=\; \frac{|V^B - V^S|}{V^B + V^S}\,. \tag{3}$$

ELO construct the TR-VPIN metric as the *moving average* of the order imbalance for the preceding L volume buckets of size V, so the computation exploits the last $L \cdot V$ contracts traded. Formally, let $\tau_0 \leq \tau_1 \leq \ldots \leq \tau_L = t$ denote the sequence of times corresponding to the endpoints of the relevant volume buckets prior to time t, and let OI_ℓ denote the order imbalance measure for the volume bucket that ends at time τ_ℓ. Then, making the dependence on the underlying time bar explicit, we define,

$$TR\text{-}VPIN_t^\delta \;=\; \frac{1}{L}\sum_{\ell=1}^{L} OI_\ell^\delta\,. \tag{4}$$

This TR-VPIN metric constitutes the "toxicity" measure constructed according to the ELO (2011c) algorithm. It is obtained for a given set of transactions over a specific sample and reflects underlying choices of the volume bucket, V, the time bar, δ, the trade classification indicator b, and the length of the moving average, L. These parameters interact in a complex manner to determine both the level and the dynamic behavior of the metric. The following section provides a more detailed analysis of the basic properties of TR-VPIN.

4. Basic properties of TR-VPIN measures

ELO (2011a) emphasize that trade time, not calendar time, is the relevant metric for sampling the information set. For example, return volatility across volume buckets is more homogeneous than across calendar time intervals. Hence, we maintain a fixed volume bucket, V, in computing TR-VPIN throughout. In addition, they rely on a rather long moving average to smooth the series and enhance the signal relative to any transitory noise. Thus, we abstain from any inquiry into this aspect of the TR-VPIN definition and fix $L = 50$. Instead, we focus on the critical building block for TR-VPIN, namely the behavior of the OI measure for given bucket size, V.

Since our transactions are ordered, the set of trades belonging to a specific bucket is fixed once we decide where to initiate the sampling. For now, we take this starting point as given. Consequently, the variation in the OI measure stems solely from how we assign the buy–sell indicators to individual trades. As explained earlier, ELO (2011a) use one-minute time bars, computing the measure as if all trades within the same bar operate on the identical side of the market.[6]

The main task of this section is to shed light on the consequences of adopting a time bar as the basic ingredient for trade classification within a volume clock scheme. This design feature induces an extreme degree of heterogeneity into the order imbalance measure, driven by the trading intensity. An increase in trading implies there are more trades per time bar and, thus, a smaller number of bars, Q, involved in computing the OI measure within a volume bucket. This raises the expected value of the OI measure, *independent* of any characteristic of the underlying trades. As a result, the OI measure is mechanically correlated with trading volume and, thus, also with return volatility, irrespective of the actual order imbalances. In contrast, using tick or fixed volume bin sampling generates homogeneous distributions for the quantities of interest. We first illustrate this point in a simple example.

[6]One rationale for using one-minute time bars is that commercial vendors often make data available to customers in this format as part of their regular data subscription services.

4.1 A simple illustration

Imagine we operate with a volume bucket of $V = 1{,}000$. The last price prior to 9:37 is 10.01, and the last observed trade indicator is a buy. For the trading period which has its first time bar recorded at 9:38, the volume bucket may contain six separate one-minute time bars. For illustration, using $\delta = 60$ seconds, we assume the triplets $(T_q^\delta, P_q^\delta, n_q^\delta) = (T_q, P_q, n_q)$ take the form,

(I) $\{(T_q, P_q, n_q)\}_{q=1...6} \;=$

$\{\,(9{:}38{:}00, 10.01, 100); (9{:}39{:}00, 10.02, 200); (9{:}40{:}00, 10.02, 200);$
$(9{:}41{:}00, 10.01, 300); (9{:}42{:}00, 10.01, 100); (9{:}43{:}00, 10.00, 100)\,\}$

The order imbalance for this bucket is $OI = |\,100 + 200 + 200 - 300 - 100 - 100\,| \,/\, 1{,}000 \;=\; 0.$

Now, assume the trading is twice as intensive, meaning that the time between trades shrinks by a factor of two, so that the same underlying trades now form the following time bars,

(II) $\{(T_q, P_q, n_q)\}_{q=1,2,3} \;=\; \{\,(9{:}38{:}00, 10.02, 300); (9{:}39{:}00, 10.01, 500); (9{:}40{:}00, 10.00, 200)\,\}$

The order imbalance measure then becomes $OI = |\,300 - 500 - 200\,| \,/\, 1{,}000 \;=\; 0.40.$

Next, imagine the trading is three times as intensive as in the original case, so the trades now combine to form only two one-minute bars,

(III) $\{(T_q, P_q, n_q)\}_{q=1,2} \;=\; \{\,(9{:}38{:}00, 10.02, 500); (9{:}39{:}00, 10.00, 500)\,\}$

The order imbalance measure is then $OI = |\,500 - 500\,| \,/\, 1{,}000 \;=\; 0.$

Now, increasing the original trading intensity fourfold, the first four trade blocks are combined into the first time bar, containing 800 contracts. The next 200 contracts will belong to a second time bar. Assuming the last trade price within this bar remains at 10.00 or below, we obtain,

(IV) $\{(T_q, P_q, n_q)\}_{q=1,2} \;=\; \{\,(9{:}38{:}00, 10.01, 800); (9{:}38{:}30, 10.00, 200)\,\}$

The order imbalance measure is $OI = |\,800 - 200\,| \,/\, 1{,}000 \;=\; 0.60.$

If the trading intensity reaches fivefold the original level, the first five trade blocks form the first new time bar, containing 900 contracts. The last 100 contracts will belong to a second time bar. Under the assumptions above, we have,

(V) $\{(T_q, P_q, n_q)\}_{q=1,2} \;=\; \{\,(9{:}38{:}00, 10.01, 900); (9{:}38{:}12, 10.00, 100)\,\}$

The order imbalance measure now becomes $OI = |\,900 - 100\,| \,/\, 1{,}000 \;=\; 0.80.$

Finally, if the trading intensity is sixfold or more than in the original scenario, then we only have one time bar in the volume bucket,

(VI) $\{(T_1, P_1, n_1)\} \;=\; \{\,(9{:}38{:}00, 10.00, 1000)\,\}$

We now have $OI = |-1{,}000|\,/\,1{,}000 \;=\; |-1.0| \;=\; 1$. This trading intensity may appear unrealistic, but it is actually typical of turbulent market conditions, including the flash crash, when trading is highly elevated for prolonged periods of time. This mechanically generates a sequence of OI measures taking the value of unity. The only moderating effect during such periods is when the time bars are split across adjacent volume buckets. The OI measure then becomes a volume-weighted average of the trade direction indicators for the relevant fractions of the two bars.

Consequently, for the sequence of trades above, the SOI measure varies from -1 to 0.8 across the scenarios, while OI fluctuates from 0 to 1. Obviously, the order imbalance measure associated with a given bucket can be extraordinarily noisy – for the identical set of transactions, OI may take on values in the full range between 0 and 1, depending on the trade intensity and how the boundary of the volume buckets interact with the time bars. Moreover, it illustrates how the OI measure inflates as the trading intensity rises. As the speed of trading grows, the number of time bars in the bucket declines and there is less diversification of buy and sell indicators. In the limit, it becomes unity, irrespective of the actual price path and the proportion of active buy and sell transactions in the bucket: OI degenerates into a pure trading intensity measure.

What happens if we instead exploit a fixed bin (FB) size for volume, using, say, $v = 200$? Assuming, for simplicity, that the original sequence of aggregate trades in scenario (I) constitutes actual trades, and each occur in the last second of the time bar, we obtain the following scenario,[7]

(FB) $\{(T_q^v, P_q^v, n_q^v)\}_{q=1\ldots5} \;=$

$\{\,(9{:}39{:}00, 10.02, 200); (9{:}40{:}00, 10.02, 200); (9{:}41{:}00, 10.01, 200);$
$(9{:}41{:}00, 10.01, 200); (9{:}43{:}00, 10.00, 200)\,\}$

The order imbalance for the bucket based on this bin size is then $OI = |200 + 200 - 200 - 200 - 200|\,/\,1{,}000 \;=\; |-0.2| \;=\; 0.2$.

Notice that, in the fixed bin approach, the SOI (OI) statistic depends only on the given transaction sequence. In our illustration, it remains -0.2 (0.2), *irrespective* of the intensity of trading, whereas time bar OI may attain any value between zero and unity. Obviously, it also avoids any mechanical correlation with trading volume, while TR-VPIN tends to rise as trading intensifies. In other words, while TR-VPIN *is* updated in trade time, it *is not* computed according to a trade clock – calendar time remains a critical determinant of the measure. In this sense, FB-VPIN is preferable as it stays true to the notion that markets evolve in transaction time.

Of course, our illustration may be overly simplistic. The variation in trading intensity may be informative in and of itself, and TR-VPIN may extract information about the state of the market from such patterns. This could potentially render the awkward idiosyncratic variation in the time bar SOI less prominent. We turn to a more formal analysis to address such questions.

[7] If the aggregate volume within each minute stems from numerous smaller trades, as is the case for the E-mini contract, the splitting of individual large transactions into adjacent volume bins is much less of an issue. The improved granularity would allow for more variability in the trade indicators, and thus help diversify the signed imbalances across volume bins, typically resulting in a lower IO measure.

4.2 Benchmark VPIN measures

We have seen that the OI measure may attain values across the entire $[0, 1]$ range for a fixed set of underlying transactions. The effect arises from an interaction of the trading intensity with the time bars. Since the VPIN is obtained as a moving average of the OI measure, it is evident that TR-VPIN may be highly sensitive to variation in the trading pattern – even if the underlying order flow imbalance (in trading time) is unchanged. This is a problem to the extent we cannot reliably associate movements in TR-VPIN with shifts in order flow toxicity.

We address the above issue in several distinct ways. First, we seek to determine how the TR-VPIN measure is expected to evolve given the observed variation in the trading pattern, but *absent* any systematic order flow imbalance. We label such benchmark series "uninformed" VPIN (U-VPIN) measures. Below, we develop two such benchmark measures which, by construction, are void of systematic order flow imbalances, yet capture the effect of time variation in the trading pattern. We will use these U-VPIN measures as controls for the impact of the trading pattern on VPIN in the absence of toxic order flow. The evolution of the actual TR-VPIN metric relative to the benchmarks should then reflect the component of TR-VPIN that is attributable to the systematic order flow imbalances, or toxicity. Ultimately, however, it is possible that toxicity is also impacting the speed and size of individual transactions. That is, the variation in the trading process may not be exogenous relative to order flow toxicity. This motivates our second alternative VPIN metric, namely the fixed bin VPIN (FB-VPIN).[8] This measure exploits fixed volume bins within each volume bucket, as exemplified in Section 4.1. It annihilates the impact of variation in the trading intensity on the individual OI statistics, while retaining the feature that VPIN is updated in event time. In combination, the U-VPIN and FB-VPIN measures allow us to gauge the relative impact of different facets of the market dynamics on the TR-VPIN metric, as implemented in ELO (2011a, 2011b, 2011c). Second, we seek to establish whether the TR-VPIN measures carry incremental information regarding future market disruptions beyond what is implied by market activity variables known to convey information regarding future return volatility, such as the raw volume and implied volatility, VIX, series.

Before introducing our U-VPIN measures, it is useful to portray a setting, entirely void of market microstructure features, for which volume and volatility measures possess forecast power for return volatility. Standard asset pricing models assume the price process constitutes a semi-martingale with respect to the natural filtration generated by the past history of prices and trades. This ensures the absence of arbitrage in a frictionless setting. It implies that any predictability, or drift, in the price over short horizons is trivial relative to the size of the (unpredictable) return innovations. At the same time, the setting is consistent with a large degree of predictability in trading volume and return volatility. In fact, both series are often viewed as driven by a serially correlated latent information flow variables so that they naturally become positively correlated.[9] Moreover, since volatility and volume both are highly persistent, lagged values of either series will be significant predictors of future volatility. Obviously, this framework is void of microstructure foundations. It implicitly assumes that fundamentals are incorporated into prices instantaneously, while trading reflects idiosyncratic liquidity or saving needs that induce random buying and selling. This is evidently not a sensible model at the ultra high-frequency level, but it does provide a benchmark with no role for systematic order flow toxicity. The point is, of course, that we should require any useful crash predictor, including the VPIN metric, to embed auxiliary information regarding short-term return fluctuations beyond what is attained

[8] In principle, we should refer to this measure as TR-FB-VPIN. However, we do not explore any BV-FB-VPIN measures in this paper, so we retain the shorter abbreviation throughout.

[9] This account echoes the "mixture-of-distributions hypothesis" explored by, for example, Clark (1973), Epps and Epps (1976), Tauchen and Pitts (1983), and Andersen (1996).

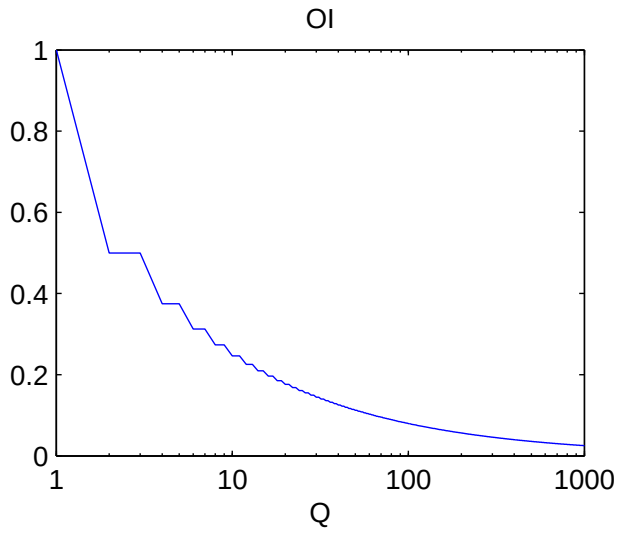

Figure 3: **Expected order imbalance versus number of time bars.** The figure plots the expected order imbalance as a function of the number of components, Q, used in the computation. The expectation is obtained under the assumption of purely random and uninformed order flow and homogeneous trade sizes.

by volume or implied volatility variables.

4.2.1 Systematic variation in TR-VPIN due to trading intensity

We first explore the pure impact of variation in the trading intensity on the OI measure. For this purpose, we assume, counter-factually, that the trading process induces an i.i.d. sequence of buy–sell indicators, $\{b_q\}$, taking on the values +1 and -1 with probability 1/2 each. In this setting, the expected value of any future trade indicator is zero, i.e., $E[b_q] = 0$. Moreover, to annihilate any heterogeneity in trade size, we assume each time bar contains an equal amount of volume, i.e., $w_q = 1/Q$. The only variable systematically impacting OI is then Q, which varies directly with the trading intensity.

Within this simplified setting, we obtain an analytical expression for the expected OI measure as a function of the number of time bars within a volume bucket. Letting this mapping be denoted $F(Q) = E(|SOI(Q)|) = E[OI(Q)]$, we document, in the Appendix, that:

$$F(Q) = E[OI(Q)] = \frac{(2q)!}{2^{2q}q!q!}, \qquad \text{if} \quad Q = 2q, \quad \text{or} \quad Q = 2q+1\,. \tag{5}$$

Figure 3 plots this *expected* order imbalance function for different values of Q, with the latter displayed on a log scale.

Intuitively, the order imbalance is non-increasing in Q, starting at unity for $Q = 1$. Moreover, we note that $F(2q) = F(2q+1)$ so, for example, $F(2) = F(3) = 0.5$, $F(4) = F(5) = 0.375$, and even $F(10) = F(11) = 0.246$. Finally, as Q grows large, $F(Q)$ approaches zero at the rate $Q^{-\frac{1}{2}}$, as is also evident in Figure 3. In fact, for Q large, we formally have,

$$F(Q) \;\sim\; \sqrt{\frac{2}{\pi Q}}\,. \tag{6}$$

These expected OI measures, derived assuming no information asymmetry and no systematic order imbalance, provide an *extremely* conservative benchmark for the OI measures. In reality, the volume within individual time bars vary in size, and the expected order imbalance is strictly minimized when volume is evenly distributed across the bars. For example, if $Q = 2$, but $w_1 = 0.9$ and $w_2 = 0.1$, the expected order imbalance is 0.9, much higher than $F(2) = 0.5$, even in the absence of any systematic order imbalance. In addition, the actual trade classification sequence, $\{b_q\}$, will not be independent, for a variety of reasons, which also increases the value of $F(Q)$.

In order to contrast the $F(Q)$ measure to TR-VPIN, it must be smoothed in a comparable manner. Hence, as for the TR-VPIN definition in equation (4), we let $\tau_1 \leq \ldots \leq \tau_L = t$ denote the sequence of times corresponding to the endpoints of the relevant volume buckets prior to time t, and let Q_ℓ denote the number of time bars included in the volume bucket that ends at time τ_ℓ. Then we define,

$$\text{U1-VPIN}_t^\delta \;=\; \frac{1}{L}\sum_{\ell=1}^{L} F(Q_\ell)\,. \tag{7}$$

This measure is constructed assuming a *homogeneous* volume distribution and *random* order flow. These assumptions are surely violated and either feature inflates the measure. Hence, U1-VPIN will invariably be significantly smaller than TR-VPIN. Nonetheless, we use it to gauge, qualitatively, the time series variation in the TR-VPIN measure we may expect solely due to changes in the trading intensity. The label U1-VPIN indicates it is our initial "uninformed" VPIN measure. It does not exploit any information about the price path or trade size distribution.

Finding 2: All else equal, the expected value of the OI measure is decreasing in Q. This implies, importantly, that as trading intensifies, OI and the associated TR-VPIN measure will tend to rise in concert with the decline in Q.

One corollary to Finding 2 is that persistent time variation in the trading intensity will induce prolonged swings in the level of OI and the associated TR-VPIN, which are correlated with overall trading volume, even in the absence of any systematic order imbalances or heterogeneity in trade size.

We again caution against the conclusion that observed time variation in trading intensity is unrelated to order imbalances. It is, indeed, likely that a significant proportion of this variation is related to the general market environment. The relevant question is rather whether the trade classification scheme provides *incremental* information beyond what we can infer from the observable trading pattern.

Finding 3: All else equal, the level of OI, and hence the associated TR-VPIN, is monotonically related to the length of the time bar, δ. In particular, adopting longer time bars (a larger δ) leads to higher TR-VPIN measures, as Q declines.

This result is no surprise. ELO (2011c) note one should compare only the relative size of TR-VPIN measures over time, and not their levels. Our finding explicates the mechanism behind this feature. It has one noteworthy implication: TR-VPIN should be calibrated to existing market conditions to allow for meaningful intertemporal or cross-market comparisons. In particular, if the volume is trending, the "effective" Q is shifting, and a corresponding change in the time bar is required to render the TR-VPIN series stationary. Specifically, in the absence of any adjustment, a (positive) drift in volume will elevate the average order imbalance and, inadvertently, signal an increasingly turbulent market environment over time. Thus, if the time bar is fixed, it is implicitly assumed that volume has no time trend.

4.2.2 Systematic variation in VPIN attributable to the trading pattern

The U1-VPIN measure captures only the impact of the number of terms used in computing OI. Any variation in volume across the bars will increase the measure. Since the trading intensity varies strongly

over time, we develop an alternative benchmark that captures the impact of volume heterogeneity on the expected OI measure. Direct analytic expressions are no longer feasible. However, we note that the expected OI measure represents an L^1 norm applied to the signed order imbalance measure,

$$E[OI] \;=\; E\left[\,|\,SOI\,|\,\right] \;=\; E\left[\,|\,w_1 b_1 + \cdots + w_Q b_Q\,|\,\right]. \tag{8}$$

It turns out that the corresponding L^2 norm is tractable under the maintained assumption that the $\{b_q\}$ sequence is i.i.d. and symmetric. That is, we can instead exploit the measure,

$$\sqrt{E\left[SOI^2\right]} \;=\; \sqrt{E\left[\,(w_1 b_1 + \cdots + w_Q b_Q)^2\,\right]} \;=\; \sqrt{w_1^2 + \cdots + w_Q^2}\,. \tag{9}$$

This expression is tantalizingly simple. Given the observed volume distribution, $w = (w_1, \ldots, w_Q)'$, for a given volume bucket, this expected order imbalance measure equals the Euclidean norm of w. Due to Jensen's inequality, we have the relation,

$$E[OI(w)] \;=\; ||SOI||_1 \;\leq\; ||SOI||_2 \;=\; |\,w\,| \;\equiv\; \sqrt{w_1^2 + \cdots + w_Q^2}.$$

Hence, this metric provides an upper bound on the impact of the trading pattern on expected OI, provided the trade indicator follows an i.i.d. process. In reality, we expect positive serial correlation in the trade indicator, implying that OI may be smaller or larger than $|\,w\,|$, depending on the size of the Jensen inequality bias versus the serial correlation in $\{b_q\}$.

In summary, the time variation in the L^2 norm provides a gauge for the variation in the order imbalance that is attributable to the characteristics of the trading process and not directly associated with the trade indicator sequence, $\{b_q\}$. By the same token, any systematic residual variation in the OI measure is likely due to asymmetries in the active order flow, as captured by the $\{b_q\}$ indicators.

As for the other order imbalance measures, we convert the $|w|$ measure into a TR-VPIN style metric by computing a backward-looking moving average. Letting w_ℓ denote the volume weight vector for the bucket that terminates at time τ_ℓ, we define,

$$\text{U2-VPIN}_t^{\delta} \;=\; \frac{1}{L}\sum_{\ell=1}^{L} |\,w_\ell\,|\,. \tag{10}$$

In summary, U1-VPIN is based on an L^1 norm for the expected order imbalance measure, while U2-VPIN is based on a related L^2 norm. Moreover, like U1-VPIN, U2-VPIN is "uninformed" about the actual price path, and thus the trade indicator sequence, utilized by TR-VPIN.

For a given distribution of Q_i across buckets, U2-VPIN is minimized when the volume in the bars of each bucket is identical, while it is maximized when volume is heavily unbalanced, i.e., one volume weight is near unity and the remainder are negligible. Thus, the measure captures variation in the heterogeneity of the observations in the volume buckets. At the same time, U2-VPIN is constructed assuming a random trade direction, so its variation cannot be attributed directly to order flow imbalances.

Finding 4: All else equal, the expected value of OI, as well as the associated tick rule VPIN measure, is increasing in the degree of volume heterogeneity across time bars.

Finding 4 rationalizes the relatively large discrepancy in level between our two U-VPIN measures.

4.2.3 Removing the impact of trading intensity via fixed volume bin VPIN

The ELO (2011a, 2011b, 2011c) TR-VPIN metric relies on time bars which renders the measure highly sensitive to variation in volume. One way to avoid this mechanical dependence on the trading intensity is to compute the OI measure from equally-sized volume bins instead. In fact, as argued above, this approach is natural given the notion that activity within a high-frequency setting should be measured in trading time rather than calendar time.

Following the notation in Section 3.1, the basic data triplets now consist of $(T_k, P_k, N_k) = (T_k^v, P_k^v, N_k^v)$, where T_k and P_k are, respectively, the calendar time and trade price associated with the last contract included in volume bin k, and the cumulative volume within each bin is $N_k = v \cdot k$. Hence, each volume bin contains, $v = N_k - N_{k-1}$, traded contracts, for $k = 1, \ldots, K$. In our empirical work, $v = 1$, $v = 1{,}000$ or $v = 5{,}000$. The first case corresponds to transaction data, while the latter two cases imply that each volume bucket with $V = 40{,}000$ contains either 40 or 8 separate bins. These figures span the average number of time bars included in the ELO OI measure across the typical trading day.

Each bucket now comprises $Q = V/v$ bins, each containing an equal fraction, $w_v = v/V = 1/Q$, of the overall volume within the bucket. We apply the classification scheme (1) to each bin, and define the "fixed bin" *signed* order imbalance, SOI^v, for a volume bucket as follows,

$$SOI^v \;=\; w_1 b_1 + \cdots + w_Q b_Q \;=\; \frac{b_1 + \cdots + b_Q}{Q} \;=\; \frac{V^B - V^S}{V}. \tag{11}$$

We let $\tau_1 \leq \ldots \leq \tau_L = t$ be the endpoints of the volume buckets prior to time t and SOI_ℓ^v be the signed order imbalance for the bucket terminating at τ_ℓ. We define the "Fixed-Bin" VPIN as,

$$\text{FB-VPIN}_t^v \;=\; \frac{1}{L} \sum_{\ell=1}^{L} \; |SOI_\ell^v \,|. \tag{12}$$

This FB-VPIN measure retains the dependence on the classification scheme, but it breaks the mechanical relationship with trading intensity. As such, it is useful in identifying the variation in TR-VPIN that stems from order imbalances measures based on a genuine trading time scale.[10]

5. Empirical results for the full sample

5.1 Summary statistics for the trading in E-mini S&P 500 futures

Table 1 provides descriptive summary statistics for the trading of the E-mini S&P 500 futures contract over our sample.

Evidently, the market is extremely liquid. There were on average more than 185,000 trades and in excess of 2,165,000 contracts exchanged per day. This implies an average transaction size of around 11.7 contracts. During the regular trading hours, there were about 390 transactions per minute or 6.5 transactions per second. Although the numbers are much lower outside regular trading hours, the activity is still impressive, with a trade consummated about once every two seconds.

[10] In principle, one can use the approach of the previous subsection to construct "uninformed" versions of FB-VPIN, which not only annihilate the mechanical relationship with trading intensity but also remove the dependence on price information. However, the corresponding measures, U1-FB-VPIN and U2-FB-VPIN, will be constants, because each volume bucket now consists of the same number (Q) of equal size bins, implying,

$$\text{U1-FB-VPIN}_t^v \equiv F(Q), \qquad \text{and} \qquad \text{U2-FB-VPIN}_t^v \equiv \frac{1}{\sqrt{Q}}.$$

Table 1: **Descriptive trading statistics for the E-mini S&P 500 futures contract**

	All	Regular	Overnight	Holiday
Volume (1 day), 000s	2166.53	1949.48	263.72	137.94
# Trades (1 day), 000s	185.45	156.18	33.07	20.56
Volume (1 min), 000s	1.52	4.81	0.26	0.10
# Trades (1 min), 000s	0.13	0.39	0.03	0.01
Trade Size	11.68	12.48	7.98	6.71
# Days	667.00	652.00	652.00	15.00

Order size: average daily percentiles

	Min	10%	50%	75%	90%	99%	99.9%	Max
All	1.0	1.0	2.1	6.3	23.6	187.5	539.0	1654.9

Notes: The table reports summary statistics for the trading in E-mini S&P 500 futures contract over the period January 2008 - July 2010. The data are reported separately for Regular Trading Hours (Regular, 8:30-15:15), Overnight Trading Hours (Overnight, 15:15-8:30), Holiday Trading Hours (Holiday, exchange holidays), and combined hours (All).

Our choices of $L = 50$ and $V = 40{,}000$ contracts ensure a close approximation to the setting of ELO (2011c). In particular, V represents about $(1/50)^{\text{th}}$ of the average daily volume (42,320 contracts), and TR-VPIN is computed based on a number of transactions close to that of a typical trading day.[11]

Another critical dimension is the number of transactions per time bar. It is a major factor in determining the reliability of the trade classification scheme and the number of blocks used in computing the OI measure. A time bar of one minute, on average, encompasses 390 separate transactions during regular trading hours. These will all be classified as active buys or active sells depending on the price change over the one-minute period. Clearly, many of these trades are misclassified, as small sequences of active buys and sells often alternate in rapid succession.[12] Nonetheless, it is an empirical question whether the classification scheme provides useful insights into the effective order flow imbalance, and we explore this issue below. However, it also motivates us to analyze the behavior of the TR-VPIN metric across alternative choices of time bars for which the degree of trade misclassification will vary. Second, there will be an average of 8–9 time bars in each volume bucket during regular trading hours. Moreover, depending on the trading intensity, this number fluctuates anywhere from 30 down to 1. Inspecting Figure 3, it is evident that the expected OI measure is highly sensitive to the number of bars in the volume bucket, Q, as reflected in the steepness of the curve. In other words, TR-VPIN will, by construction, vary dramatically over time in response to persistent variation in volume, independently of whether or not the transactions are evenly balanced across buys and sells.

[11] ELO (2001c) have fewer transactions in their sample as they use $V = 39{,}351$ contracts. We let V be a multiple of 5,000 to guarantee that buckets constructed from bins of $v = 1{,}000$ and 5,000 contracts contain an integer number of bins.

[12] The first order tick return autocorrelation is around -0.41, while the next four autocorrelations also are negative, albeit small. This suggests a rapid alternation between transactions consummated at the bid and ask quote. While it is not fool-proof to associate a down-tick with an active sell and an up-tick with an active buy, it should not induce a dramatic bias as the order book depth typically is much deeper than the volume associated with individual transactions. Consequently, the best bid and ask quotes are generally relatively stable compared to the oscillation of transactions between the bid and ask. Hence, sequences of hundreds of transactions typically involve a large number of both buys and sells.

Finally, we note the extreme right-skewed trade size distribution. In a typical time bar containing 390 transactions, about 200 will involve the exchange of only one or two contracts. However, about 40 involve more than 20 contracts each, and a few entail exchanges of hundreds of contracts in a single transaction. While this size heterogeneity will not have a strong bearing on one-minute time bar OI, it will impact measures computed from smaller time bars or volume bins.

5.2 Comparing alternative tick rule VPIN measures

TR-VPIN is subject to distortions stemming from the time variation in volume, inhomogeneity of the volume weights across bars, and random variation in the trade classification scheme. We assess the significance of these factors by comparing TR-VPIN computed over alternative time bars with our U1-, U2-, and FB-VPIN measures, as each convey information about the forces impacting TR-VPIN.

Figure 4 displays TR-VPIN for $\delta = 10$, $\delta = 60$, and $\delta = 300$ plotted alongside the corresponding U1-VPIN and U2-VPIN. The latter series represent the expected values of TR-VPIN, assuming random trade classification and conditional on the observed trading pattern. Finally, the bottom panels provide the FB-VPIN series, computed for volume bins of $v = 1$, i.e., tick data, $v = 1000$, and $v = 5000$. Recall that the FB-VPIN series eliminate any mechanical time series correlation with trading intensity.

Figure 4 highlights a number of features. First, as expected, TR-VPIN increases significantly as the bar grows, and FB-VPIN rises as the bins lengthen. In both cases, the effective Q drops. Second, it is evident that aggregated TR-VPIN measures are strongly correlated with U1-VPIN and U2-VPIN, and the association grows stronger with the aggregation level, from $\delta = 10$, through $\delta = 60$ to $\delta = 300$. In fact, for the latter two cases, TR-VPIN and U2-VPIN almost coincide and have near identical trends and spikes. Moreover, U1-VPIN displays the same qualitative time series variation even if, by construction, it attains lower values. In contrast, the FB-VPIN series are much less aligned with TR-VPIN. Third, the extreme values attained by the U-VPIN and TR-VPIN series on May 6, 2010, are striking, since none of the FB-VPIN series attain an unusual value. The discrepancy occurs even if the number of volume bins with $v = 1,000$ and $v = 5,000$ closely match that used, on average, by TR-VPIN with $\delta = 10$ and $\delta = 60$, respectively.

Finding 5: TR-VPIN is highly correlated with U1-VPIN and U2-VPIN and the correlation increases with the length of the time bar. Hence, the trade classification rule is largely negligible as a determinant for the time series variation of TR-VPIN.

Finding 6: TR-VPIN behaves dramatically differently from FB-VPIN, even if the identical volume buckets are used for computation. In particular, only TR-VPIN attains an exceptional value on May 6, 2010.

Complementary evidence is provided by Table 2. It reports the sample correlations between the TR-VPIN measures for different time bars as well as correlations with FB-VPIN, the daily trading volume, and VIX. We already noted that large time bars tend to induce a mechanical correlation with trading volume. The VIX index is incorporated as a reference for the subsequent discussion of TR-VPIN as a predictor for future return volatility.

Table 2 confirms the main conclusions from Figure 4 and brings out new features. First, there is strong and generally increasing correlation between the TR-VPIN and U-VPIN measures as δ grows: 0.67–0.74 for $\delta = 10$, 0.77-0.80 for $\delta = 60$, and around 0.84 for $\delta = 300$.

Second, there is a dramatic break in the VPIN-volume correlation as we move from TR-VPIN to FB-VPIN. For TR-VPIN, the positive correlation is not surprising, even if the magnitude might be. It is a manifestation of Finding 2. The main insight is rather that the FB-VPIN measures computed from relatively small bins are *negatively* correlated with volume. Moreover, these (absolute) correlations

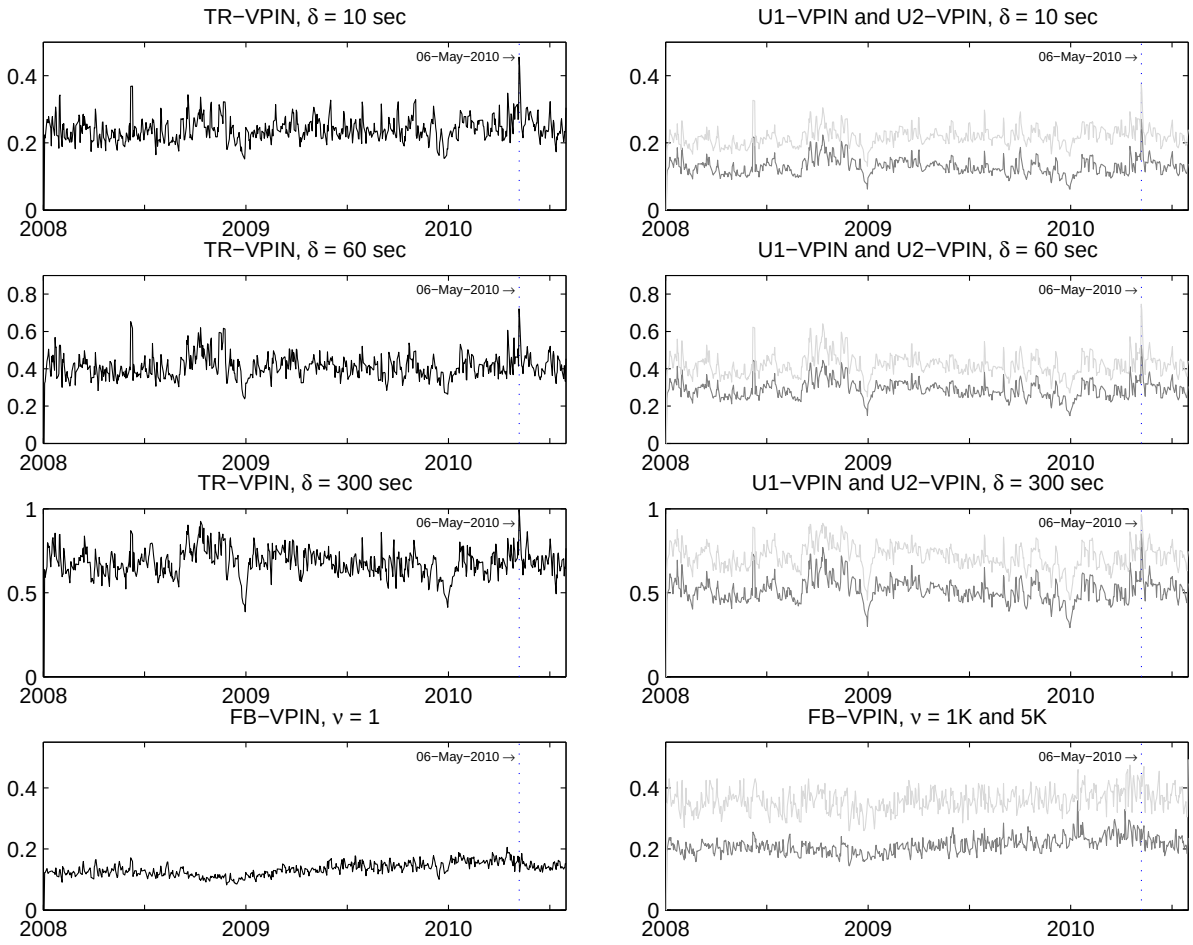

Figure 4: **The evolution of alternative VPIN measures.** The top three rows plot the daily maximum values of VPIN (black), U1-VPIN (L^1 metric, dark gray), and U2-VPIN (L^2 metric, light gray) for $\delta = 10$, 60, and 300 seconds. The bottom panels depict FB-VPIN for $\nu = 1$ (black), 1000 (dark gray), and 5000 (light gray).

drop in a monotone fashion as the size of the volume bins grow. Since FB-VPIN annihilates the impact of the trading intensity, this suggests that the trade indicators are largely unrelated to overall trading volume. That is, to the extent TR-VPIN captures order imbalances, it is almost exclusively due to the size distribution, while the trade indicators are near irrelevant. This is, of course, also consistent with the very high degree of correlation between the TR-VPIN and U-VPIN measures, and serves to corroborate Finding 5. In contrast, TR-VPIN is only weakly correlated with FB-VPIN, as is evident from the entries in the left part of the bottom panel of Table 2.

Third, TR-VPIN displays positive correlation with VIX, although it is noticeably weaker than the TR-VPIN-volume correlation. Given that volume and volatility are known to be related, as exemplified by our sample correlation of 0.52, the correlation between TR-VPIN and VIX is anticipated. In contrast, FB-VPIN is *strongly* negatively related to VIX. This feature is surprising! It suggests that TR-VPIN, measured consistently in trading time, is *inversely* related to the (long run) volatility level. Moreover, the effect is again strongest for the smaller bins and dissipates as the transactions are aggregated. The relationship is highly significant for transaction VPIN, suggesting a systematic interaction between

Table 2: **Correlations between various VPIN and market activity variables**

	$\delta = 10$				
	TR-VPIN	U1-VPIN	U2-VPIN	Volume	VIX
TR-VPIN	1.00			0.50	0.06
U1-VPIN	0.67	1.00		0.83	0.47
U2-VPIN	0.74	0.86	1.00	0.61	0.14

	$\delta = 60$				
	TR-VPIN	U1-VPIN	U2-VPIN	Volume	VIX
TR-VPIN	1.00			0.62	0.24
U1-VPIN	0.77	1.00		0.85	0.47
U2-VPIN	0.80	0.98	1.00	0.82	0.40

	$\delta = 300$				
	TR-VPIN	U1-VPIN	U2-VPIN	Volume	VIX
TR-VPIN	1.00			0.71	0.37
U1-VPIN	0.83	1.00		0.83	0.44
U2-VPIN	0.85	0.96	1.00	0.82	0.47

	TR-VPIN				
FB-VPIN	$\delta = 10$	$\delta = 60$	$\delta = 300$	Volume	VIX
v = 1	0.33	0.16	-0.03	-0.24	-0.58
v = 1K	0.33	0.19	0.04	-0.09	-0.40
v = 5K	0.34	0.29	0.14	0.04	-0.21

Notes: The table reports correlations of TR-VPIN, U1-VPIN, and U2-VPIN for $\delta = 10$, 60, and 300 sec, one-day trading volume, VIX, and FB-VPIN for $v = 1$, 1K, and 5K. The sample period is January 2008 - July 2010.

uncertainty or turmoil in the market and the trade classification scheme in equation (1). The slow dissipation of this relation with the aggregation level may indicate that the same phenomenon remains at work, but the trade indicators become less reliable, as transactions are bundled into larger blocks. We are led to the surprising hypothesis that volume-clock based measures of VPIN are *negatively* related to trading volume and return volatility. By the same token, TR-VPIN, in general, displays positive correlation with VIX only because it, by construction, covaries strongly with the trading intensity.

Finding 7: Time bar TR-VPIN is strongly correlated with volume and weakly correlated with VIX. In contrast, FB-VPIN is weakly, and predominantly negatively, correlated with volume, and strongly negatively correlated with VIX. Thus, the TR-VPIN–volatility correlation largely reflects a mechanical association between TR-VPIN and trading intensity.

In summary, FB-VPIN displays an entirely different dynamic behavior than TR-VPIN, raising the question of whether these VPIN variants capture fundamentally distinct features of the trading process. We explore the negative association between FB-VPIN and VIX in more detail later.

Table 3: **Correlations of future absolute returns with various measures**

	TR-VPIN			U1-VPIN			FB-VPIN				
	δ=10	δ=60	δ=300	δ=10	δ=60	δ=300	v=1	v=1K	v=5K	Volume	VIX
$\lvert r_{t,t+1}\rvert$	0.07	0.15	0.21	0.27	0.27	0.26	-0.30	-0.20	-0.08	0.31	0.46
$\lvert r_{t,t+10}\rvert$	0.04	0.14	0.22	0.27	0.28	0.27	-0.34	-0.24	-0.11	0.32	0.49
$\lvert r_{t,t+50}\rvert$	-0.01	0.09	0.18	0.23	0.24	0.22	-0.35	-0.24	-0.15	0.29	0.47
$\lvert r_{t,t+250}\rvert$	0.06	0.17	0.25	0.30	0.30	0.30	-0.32	-0.24	-0.08	0.37	0.54

Notes: Volume is one-day trading volume. The sample period is January 2008 - July 2010.

5.3 VPIN as a forecast of future short-term return volatility

ELO (2011a) suggest that TR-VPIN may be useful for predicting impending turmoil in financial markets. They provide evidence that TR-VPIN with $\delta = 60$ is correlated with future volatility, and that the correlation is stronger for more extreme volatility realizations. These observations are consistent with the findings above, where we also document a positive, albeit somewhat weak, correlation between the TR-VPIN metric and concurrent VIX. However, as suggested by ELO (2011c), it may be important to distinguish short-term order imbalance or toxicity induced volatility, which may occur within the current trading day and only last, say, a few hours, versus the broader volatility expectations for the coming month, as reflected in the VIX index. Consequently, this section explores how well TR-VPIN performs as a forecast variable relative to other standard volatility predictors. Moreover, we seek to shed additional light on the mechanism that generates a correlation between TR-VPIN and future volatility.

Table 3 provides a first overview of the evidence. It tabulates the correlations between alternative predictor variables and future cumulative absolute returns over four different horizons, ranging from 1 to 250 volume buckets, corresponding to an average of a few minutes to five full trading days. The candidate forecast measures include TR-VPIN and U1-VPIN obtained from different bar lengths, FB-VPIN obtained from different bin sizes, along with lagged daily volume, and the VIX index.

The raw correlations suggest that TR-VPIN provides a comparatively poor forecast. The TR-VPIN correlations with future volatility, at all horizons, are uniformly below those associated with U1-VPIN, even though the only difference between the two is the trade classification scheme. Taken at face value, it implies that the classification rule induces variation in TR-VPIN that *lowers* its correlation with future volatility, relative to the random classification associated with U1-VPIN. One explanation is that lagged daily volume is more highly correlated with future volatility than TR-VPIN. Since U1-VPIN, in turn, is more strongly correlated with volume than TR-VPIN, this corroborates the hypothesis that TR-VPIN largely predicts future return volatility due to its mechanical correlation with trading volume. Furthermore, we note that VIX, by far, is the variable most strongly correlated with future volatility. Finally, the strong negative association between FB-VPIN and future volatility is striking. The relation is strongest for small bins and declines as volume is aggregated into larger bins. These findings are consistent with the negative correlation between transaction VPIN and VIX, discussed in Section 5.2.

To more formally assess the association between TR-VPIN and future realized volatility, we quantify the predictive performance within a regression setting where we control for the impact of auxiliary variables. We focus on TR-VPIN measures with $\delta = 60$, but the qualitative findings are identical for other time bars. Table 4 summarizes the evidence for a representative set of regressions.

The results are clear cut. From regression one in the first column of either panel of Table 4, we see that there, indeed, is a highly significant relationship between TR-VPIN and future volatility, although the predictive power is limited as reflected in the adjusted R^2 of about 2% and 8%, respectively,

Table 4: **Forecast regressions for absolute return**

Panel A: One-period forecast

	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
Const.	-0.01	-0.16	0.02	-0.02	-0.12	0.06	-0.08	-0.10	-0.05	-0.07
	(-0.38)	(-6.12)	(1.92)	(-4.21)	(-4.49)	(2.91)	(-5.68)	(-6.67)	(-6.91)	(-4.23)
TR-VPIN	0.49				-0.44	-0.15	0.17	-0.01		-0.01
	(6.97)				(-5.64)	(-2.38)	(4.95)	(-0.33)		(-0.18)
U1-VPIN		1.22			1.69			0.38		0.11
		(12.18)			(14.68)			(5.45)		(1.28)
Vol$\times 10^{-7}$			0.63			0.69			0.20	0.16
			(12.12)			(12.49)			(6.83)	(3.81)
VIX$\times 10^{-2}$				0.60			0.58	0.55	0.54	0.53
				(33.66)			(35.99)	(32.03)	(34.31)	(32.42)
$\bar{R}^2$	2.41	7.58	8.93	21.25	8.36	9.07	21.53	21.76	21.90	21.92

Panel B: 50-period forecast

	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
Const.	0.00	-0.15	0.02	-0.02	-0.11	0.08	-0.06	-0.09	-0.04	-0.04
	(0.15)	(-6.14)	(2.16)	(-4.08)	(-4.41)	(4.03)	(-5.72)	(-6.83)	(-7.23)	(-3.61)
TR-VPIN	0.46				-0.47	-0.20	0.13	-0.06		-0.05
	(7.15)				(-6.67)	(-3.50)	(4.85)	(-1.70)		(-1.47)
U1-VPIN		1.18			1.69			0.38		0.08
		(12.62)			(15.32)			(6.39)		(1.07)
Vol$\times 10^{-7}$			0.62			0.71			0.19	0.18
			(12.67)			(13.07)			(7.49)	(5.03)
VIX$\times 10^{-2}$				0.59			0.58	0.55	0.53	0.53
				(37.14)			(38.81)	(36.71)	(37.64)	(36.84)
$\bar{R}^2$	7.97	27.23	33.64	78.35	30.71	34.57	78.97	79.82	80.50	80.54

Notes: The figures represent OLS regression coefficients; t-statistics based on HAC-standard errors, constructed with 50 lags, are reported in parentheses. TR-VPIN and U1-VPIN are for $\delta = 60$ seconds; "Vol" is the one-day backward trading volume. The forecast horizon is one volume bucket ("one period") and fifty volume buckets ("50 periods"), respectively. The sample period is January 2008 - July 2010.

for the two forecast horizons. The second regression presents the corresponding regression results for U1-VPIN. The explanatory power rises by a factor of more than 3! Given the high correlation between the two measures, this once more suggests that the variation in trading activity, as reflected in U1-VPIN, is the underlying source of volatility predictability: the modification of the statistic to also reflect the trade classification scheme, as done in TR-VPIN, but not U1-VPIN, is detrimental to forecast performance. This brings us to regression three, which documents another improvement from simply forecasting future volatility with the one-day lagged trading volume. The explanatory power is now about four times that obtained with TR-VPIN. Along similar lines, column six demonstrates that TR-VPIN has no – even negative – auxiliary explanatory power in forecasting return volatility once we control for the trading activity. Obviously, the two regressors are strongly correlated, rendering

the point estimates somewhat unreliable, but the minimal increase in the overall explanatory power, relative to the univariate volume regression in column three, confirms the lack of incremental predictive content of TR-VPIN. Finally, regression four shows that the VIX index provides superior forecasts relative to trading volume.[13] Completing the picture, regression five shows that U1-VPIN also crowds out TR-VPIN as a predictor for future volatility. The minimal increase in explanatory power relative to regression two again confirms the lack of incremental predictive power in TR-VPIN. Finally, regressions 7-10 reveal that the trading activity variables do contain useful information for future return volatility over and above the VIX measure, although the improvements in explanatory power is marginal. As before, the TR-VPIN metric is the one with the weakest predictive power, and it is insignificant when included in regressions containing other trading activity related variables, e.g., regressions 8 and 10.

Finding 8: The TR-VPIN metric is, in general, much less robustly correlated with future short-term realized volatility than regular volatility predictor variables. Moreover, it is less correlated with future return volatility than the corresponding U1-VPIN measure, suggesting that the trade classification rule actively degrades the volatility forecast content of VPIN.

Finding 9: The evidence is consistent with the hypothesis that the TR-VPIN metric is weakly correlated with future return volatility because of its correlation with trading volume. Once we control for trading volume, TR-VPIN is, if anything, negatively related to future return volatility.

5.4 On the behavior of transactions VPIN

Perhaps our most striking finding is that transaction and FB-VPIN produce opposite conclusions of those obtained via TR-VPIN. The effect of moving from time bars to tick or volume bin data is so pronounced that there *must* be a rational explanation. We provide initial observations on the issue, but do not pursue this at length as it would take us outside the scope of the present work.

ELO (2011c) argue that the results obtained from tick data are so counterintuitive that, a priori, they should be disregarded. They suggest the seemingly perverse results stem from massive misclassification at the high-frequency level. This is not obvious, however. In some respect, tick data provide the best opportunity to minimize the effects of trade misclassification, given that the market, most of the time, operates with a well-defined spread of one to two ticks. In this scenario, it is usually correct to associate an up-tick with an active buy and a down-tick with an active sell. The classification of zero tick change transactions, which constitute a large proportion of the observations, is more dubious. As the best bid and ask prices shift over time, the classification rule will mislabel a number of these trades. However, given the oscillation between up- and down-tick transactions, these mistakes will typically not produce long sequences of errors, and those that do occur will be mitigated through diversification via the averaging across a huge set of transactions. For example, with $V = 40{,}000$, there are thousands of transactions in a typical bucket and random misclassifications largely wash out.

The use of transaction data has additional benefits. First, it avoids assigning the same trade direction to hundreds of transactions within large blocks, and thus likely misclassifying close to half of them. Second, the use of transaction data eliminates the dependence on initial conditions, as the classification of any given trade now is independent of the location of the volume buckets.

Given the apparent advantages of using tick data, what is behind the curious empirical results obtained from transaction- and FB-VPIN measures? In particular, why is transaction VPIN strongly negatively correlated with volume and volatility, and why do these variants of VPIN drop rather than soar during episodes like the flash crash? The following related facts provide a partial answer.

[13] This does not imply that VIX is an ideal predictor of future volatility. It is feasible to construct even better forecast measures using different aspects of the option data, see, e.g., Andersen and Bondarenko (2007).

One, the trade size is negatively related to return volatility. In fact, the correlation between VIX and the average trade size is -0.38 for one volume bucket, it is -0.69 when assessed over 10 buckets, and it is -0.86 when measured over 50 buckets. This relation likely reflects the lower depth of the limit order book when volatility, and economic uncertainty, increases. For a given bucket, we thus have a larger set of transactions as volatility increases, implying even better diversification of the trades, and a tendency for transactions VPIN to drop. Nonetheless, this effect may be minor as transactions already should be well diversified within buckets. Moreover, the drop in trade size does not explain why the more highly aggregated FB-VPIN measures inherit similar features, as the volume bins are not directly impacted by this effect. Thus, there is at least one more key to the puzzle.

Two, and importantly, for volume bins of 1,000 and 5,000 contracts, we find the daily probability of a "continuation" in the trade indicator across consecutive bins to be 67% and 58%, respectively. These persistence measures turn out to be strongly negatively related to the VIX index, with correlations of -0.67 and -0.39, respectively. This is consistent with the hypothesis that an increase in volatility reduces the proportion of zero tick change buckets and raises the frequency of oscillation between positive to negative tick changes. If the tick changes are close to symmetric over short horizons, this leads to a lower degree of persistence in the trade classification indicator, and hence a drop in the associated VPIN measure. However, it is clear that other forces may be at play, and it requires a detailed study at the transaction level to more systematically sort out the relevant factors.

Finding 10: Transaction- and FB-VPIN are negative correlated with volatility because the trade classification produces less serially correlated trade indicators during volatile market conditions. This leads to lower order imbalance measures as the volume bins (transactions) diversify more effectively within the buckets. Nonetheless, the reason behind this negative relation between volatility and trade classification is worthy of additional study.

6. Bulk volume VPIN

ELO (2012a) invoke a "bulk volume" (BV) classification strategy, using probabilistic assignment of buys and sells from aggregated (bulk) volume. Specifically, the proportion of buy volume over the bar of size δ is determined as a function of the price change. Letting $Z(\cdot)$ denote the cumulative distribution function of a standard normal variate, for time bar j, the procedure takes the form,

$$V_j^B \;=\; Z\left(\frac{\Delta P_j}{\overline{\sigma}}\right)\cdot V_j \qquad \text{and} \qquad V_j^B - V_j^S \;=\; \left[2\cdot Z\left(\frac{\Delta P_j}{\overline{\sigma}}\right) \;-\; 1\right]\cdot V_j\,,$$

where $\overline{\sigma}$ is the sample (unconditional) standard deviation of the transaction price change between adjacent bars, V_j denotes the trading volume over time bar j, while V_j^B and V_j^S indicate the assigned buy and sell volume, respectively. Converting the BV trade classification into trade indicators, as defined in equation (1), we have,

$$b_j = 2\cdot Z\left(\frac{\Delta P_j}{\overline{\sigma}}\right) \;-\; 1\,.$$

The trade indicator, b_j, still maps the time bars into the interval $[-1,1]$ but, unlike for the tick rule, it now attains interior values and not just -1 and 1. In fact, the BV approach interprets no price change as balanced trading ($b_j = 0$), while a large positive (negative) price change is translated into a proportionally large (small) buy volume. For a given bucket, the signed order imbalance measures SOI and the associated order imbalance measure OI are defined as in equations (2) and (3). The BV-VPIN is then computed as in equation (4) but, of course, using BV rather than tick rule trade

classification. At first glance, this seems sensible. Surely, large price changes mostly occur when order flow is unbalanced. Moreover, the extreme allocation of all transactions in a time bar to the same side of the market is avoided. Consequently, for a given time bar, the BV procedure may indeed produce a more accurate trade classification than the (bulk) tick rule. In turn, this could render the VPIN measure more reliable and informative.

Nonetheless, the BV classification represents an additional step away from the traditional approach. No element of the BV procedure references features of the individual trades – rather the classification *asserts* that price changes over time bars provide useful indicators for the underlying order imbalances. We have already documented that the TR-VPIN approach, exploiting aggregate blocks of trading volume, has a dramatic impact on the properties of the order imbalance measures relative to the standard tick rule identification using individual trades. The BV methodology adds another critical determinant to the mix by letting the size of price changes determine order imbalances. It is thus pertinent to reflect on the properties of the BV-VPIN measure. We now turn to that task.

6.1 Sources of Variation in BV-VPIN

The new element introduced by BV-VPIN stems from the trade classification. We explore how this procedure operates across different market conditions. We first formalize the notion that BV produces less extreme order imbalance measures for any given time bar than TR. For a given bar size δ, let b_j^{TR} and b_j^{BV} refer to the trade indicators for bar j obtained via TR and BV classification. It is readily confirmed that,

$$|\, b_j^{\mathrm{BV}}\,| \;\leq\; |\, b_j^{\mathrm{TR}}\,|.$$

This fact might suggest that the BV-OI will be smaller than TR-OI for volume buckets, and therefore also that BV-VPIN will be lower than TR-VPIN. However, it is easy to construct counterexamples to this prediction. It does not hold realization-by-realization, as the individual time bar TR-SOI measures may cancel each other out at the bucket level.

To establish a general result, we resort to the type of assumptions invoked in Section 4.2.2. If the BV trade indicator process is mean zero, independent of the volume weights, and i.i.d., then we have, conditional on the volume pattern across the time bars, the extension of equation (9),

$$\sqrt{E\,[SOI^2]} \;=\; \sqrt{E\left[\,(w_1\,b_1+\cdots+w_Q\,b_Q\,)^2\,\right]} \;=\; \sqrt{E\left[\,w_1^2\,b_1^2+\cdots+w_Q^2\,b_Q^2\,\right]} \;=\; |w|\cdot\sigma(b)\,, \quad (13)$$

where $\sigma(b)$ is the standard deviation of a trade indicator. For BV classification, $\sigma(b)<1$, implying that

$$E\left[\,BV\text{-}SOI^2\,\right] \;<\; E\left[\,TR\text{-}SOI^2\,\right].$$

Equation (13) suggests that the main sources of variation in the BV-OI and BV-VPIN measures are linked to directly observable market variables, namely volatility and trading intensity. Clearly, $\sigma(b)$ is driven by the time variation in absolute price changes, or volatility. We comment on this specific mechanism further below. The second factor is the L^2 norm of the volume weight vector. It is tied to the number of time bars in the volume bucket. For example, if the volume is homogeneous across the bars, we have $w_q = 1/Q$ and $|w| = Q\cdot w_q^2 = 1/Q$. That is, $|w|$ is inversely related to Q. Effectively, $\gamma = 1/Q$ is a measure of trade intensity, as it is governed (inversely) by the average trading volume within the time bars of the bucket, V/Q.

We now briefly consider the relation between $\sigma(b)$ and price change volatility. The BV approach normalizes the price change over each time bar by (an estimate of) its unconditional standard deviation. As is well known, and readily confirmed in the present sample, one-minute price changes are extremely

far from being normally distributed. That is, the standard normal CDF is merely a devise for converting price changes via a monotone transformation into units that fall in the prescribed interval. In fact, ELO (2012a) make no direct assumption of Gaussianity, and other CDF functionals could be used for this purpose. To shed light on the implications of their use of the normal CDF, we adopt an assumption which provides a much improved approximation to the actual distribution of price changes across time.

Consider the case, where the price changes over each one-minute interval is indeed Gaussian, but volatility varies across time bars, so the non-Gaussian nature of the unconditional price change is generated by a normal-mixture distribution. At any point in time, the price change is Gaussian but volatility is high in some part of the sample and low in others. Likewise, the volatility is systematically much lower overnight than during regular trading hours. Using this simple, yet descriptive, data generating process for asset price changes, we seek to understand how the relevant features of the BV-trade classification and the associated BV-OI measure are impacted by volatility fluctuations.

Formally, we denote as by σ_j the volatility over time bar j. Our mean-zero Gaussian mixture assumption implies that

$$b_j \;=\; 2 \cdot Z\left(\frac{\Delta P_j}{\sigma_j} \cdot \frac{\sigma_j}{\bar{\sigma}} \right) \;-\; 1\,. \tag{14}$$

Importantly, the standard normal CDF is applied to a standard normal variate *scaled* by $\sigma_j/\bar{\sigma}$. Thus, the normalized price change is a mean-zero Gaussian, but with variance $\sigma_j^2/\bar{\sigma}^2$. Given these assumptions, we may exemplify the effect of time-varying price change volatility. If $\sigma_j = 2\bar{\sigma}$, then the probability of $|b_j|$ exceeding 0.50 is 73.6% rather than the 50% for $\sigma_j = \bar{\sigma}$, and the probability of $|b_j|$ exceeding 0.90 is 41.1% rather than the 10% for $\sigma_j = \bar{\sigma}$. Likewise, for the more extreme – but still empirically relevant – scenario of $\sigma_j = 5\bar{\sigma}$, the probability thresholds of 0.50 and 0.90 are exceeded 89.3% and 74.2% of the time compared to the 50% and 10% benchmarks associated with $\bar{\sigma}$. The point is that the average size of the b_j realizations starts approaching the sequence of alternating -1 and 1 associated with the indicators of the TR procedure as price volatility increases.

In summary, our analysis suggests that the trade intensity and the price volatility are systematic drivers of the BV order imbalance measure. Moreover, since the effect in equation (13) is given by the product of these two activity indicators, the strong correlation between volume and volatility will further magnify the impact during volatile market conditions. As such, we would expect BV-VPIN to be even more sensitive to concurrent market activity than TR-VPIN which primarily is governed only by the trading intensity. Of course, the key question remains whether the BV-VPIN metric provides significant incremental information regarding future market conditions relative to such directly observable activity variables.

6.2 Rationalizing the empirical behavior of BV-VPIN

We now revisit our prior findings regarding the behavior of TR-VPIN and seek to infer whether we expect BV-VPIN to display similar traits or deviate in specific ways due to the modification of the trade classification scheme. Moreover, we seek to verify whether the qualitative results are borne out over our sample. As we postpone considerations of the events surrounding the flash crash to the subsequent section, we start from Finding 2.

The positive correlation between the level of TR-VPIN and trading volume as well as the length of the time bar, noted in Findings 2 and 3, carries over to BV-VPIN. The relation is driven by the average number of separate time bars within the volume bucket and this dependence is preserved in the BV-VPIN metric, as evidenced by equation (13). The second right most column of Table 5 verifies that BV-VPIN is strongly correlated with trading volume. Moreover, the lower right panel of Figure 1 portrays the BV-VPIN series for $\delta = 60$ seconds. It averages about 0.3 across the sample. Computing

Table 5: **Correlations of BV-VPIN with various activity variables**

			FB-VPIN				
BV-VPIN	U1-VPIN	U2-VPIN	v=1	v=1K	v=5K	Volume	VIX
$\delta = 10$	0.85	0.68	-0.13	-0.01	0.16	0.73	0.46
$\delta = 60$	0.89	0.86	-0.26	-0.11	0.10	0.80	0.56
$\delta = 300$	0.84	0.86	-0.40	-0.23	-0.00	0.81	0.64

Notes: The table reports correlations of BV-VPIN for $\delta = 10$, 60, and 300 sec, with U-VPIN, FB-VPIN, one-day trading volume and the VIX. The sample period is January 2008 - July 2010.

the average values of BV-VPIN for $\delta = 10$ and 300 seconds produces average values of about 0.2 and 0.45, respectively, confirming Finding 3. The related Finding 4 also trivially will apply to BV-VPIN, but it is mostly relevant for understanding our benchmark U-VPIN metrics.

Moving to Finding 5, Table 5 confirms that BV-VPIN is also highly correlated with our U1- and U2-VPIN benchmark measures. This is consistent with the trading intensity being an important determinant of BV-VPIN. Moreover, the interaction with volatility may actually help mitigate the noise in the BV-VPIN measure and render the correlation stronger than for TR-VPIN. This is where we may see the first noticeable impact of the shift in trade classification as the built-in volatility correlation starts having an impact. Of course, as noted in Finding 6, BV-VPIN also has dramatically different properties than FB-VPIN and transaction VPIN. The latter are not, in contrast to BV-VPIN, directly related to price volatility and they are constructed to annihilate any direct volume dependence so – not surprisingly – the middle panel of Table 5 shows that the BV- and FB-VPIN measures are largely unrelated. Thus, while ELO (2011c) assert that FB-VPIN is a viable alternative to TR-VPIN, it continues to produce diametrically opposite results to the approaches employed in ELO (2011a, 2011c, 2012a). These observations also relate to Finding 7. BV-VPIN is expected to be more strongly correlated with VIX than TR-VPIN, as the price volatility is one of the direct forces behind the former, as demonstrated by equations (13) and (14). Indeed, as may be verified by comparing Tables 5 and 2, over our sample, BV-VPIN is more strongly associated with VIX than is the case for TR-VPIN.[14] It is also evident from Figure 1 that BV-VPIN accentuates the volatile sample periods more strongly than TR-VPIN – notice in particular the more pronounced elevation of the BV-VPIN measure throughout the financial crisis.

The above remarks have relevance for Findings 8 and 9. Most importantly, because the trade classification scheme actively imbues BV-VPIN with some short-term realized volatility information, it should contain predictive power for future realized volatility beyond what is captured by our U-VPIN metrics, which are solely governed by the trading pattern. On the other hand, the monotonic, yet highly nonlinear, transformation of volatility in equation (14) is clearly not an efficient realized volatility indicator. Thus, it is still likely to underachieve relative to a more direct volatility indicator like the VIX index in terms of forecasting future volatility.

Table 6 reports on a set of predictive regressions for future average absolute returns over one and fifty volume buckets. While TR-VPIN is poor, even relative to the U1-VPIN metric, the BV-VPIN metric improves on the uninformed benchmark and even beats volume in terms of predictive power for future volatility. Nonetheless, is falls far short of the VIX measure which roughly doubles the R^2 obtained by BV-VPIN. Moreover, the results are remarkably consistent over both forecast horizons.

[14] One word of caution, however, as BV-VPIN clearly responds to current price volatility while the VIX represents longer term (one month) volatility expectations. In this sense, a more suitable benchmark is a realized volatility estimator.

Table 6: **Forecast regressions for absolute return**

	One-period forecast					50-period forecast				
	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5
Const.	-0.01	-0.16	-0.05	0.02	-0.02	0.00	-0.15	-0.04	0.02	-0.02
	(-0.38)	(-6.12)	(-4.00)	(1.92)	(-4.21)	(0.15)	(-6.14)	(-3.51)	(2.16)	(-4.08)
TR-VPIN	0.49					0.46				
	(6.97)					(7.15)				
U1-VPIN		1.22					1.18			
		(12.18)					(12.62)			
BV-VPIN			0.88					0.85		
			(15.77)					(15.94)		
Vol$\times 10^{-7}$				0.63					0.62	
				(12.12)					(12.67)	
VIX$\times 10^{-2}$					0.60					0.59
					(33.66)					(37.14)
$\bar{R}^2$	2.41	7.58	11.09	8.93	21.25	7.97	27.23	39.26	33.64	78.35

Notes: The figures represent OLS regression coefficients; t-statistics based on HAC-standard errors, constructed with 50 lags, are reported in parentheses. TR-VPIN, U1-VPIN, and BV-VPIN are for $\delta = 60$ seconds; "Vol" is the one-day backward trading volume. The forecast horizon is one volume bucket ("one period") and fifty volume buckets ("50 periods"), respectively. The sample period is January 2008 - July 2010.

These findings are fully consistent with the empirical results in ELO (2012a). They present extensive evidence for a highly significant correlation between BV-VPIN and future volatility, as do we. However, the interpretation differs greatly. ELO (2012a) argue this verifies that order flow toxicity predicts future return volatility in a unique fashion, not directly related to other forecast variables. We find instead that the correlation – beyond what is explained purely by the trading pattern and already conveyed by our U-VPIN metrics – arises from the modification of the trade classification strategy from TR-OI to BV-OI. This shift lets the size of the concurrent price change – a realized volatility measure – directly impact the buy–sell indicator. Effectively, it is a distorted volatility measure which combines trading intensity and price volatility in a nonlinear fashion, while the relationship between the BV-VPIN metric and the true underlying order imbalance remains unclear.[15] Thus, our interpretation of this aspect of the evidence is that BV-VPIN constitutes an imperfect realized volatility metric which, by construction, will have forecast power, due to the persistence in the volatility process. However, this predictability stems exclusively from the incorporation of volume and price volatility into the measure. If the purpose is to forecast future return volatility, we have well-known and much superior measures available.[16] The proof that BV-VPIN captures salient features of order flow toxicity, not embodied in regular real-time volatility and volume measures, and provides a superior indicator of future market conditions *must* be

[15] The evidence in ELO (2012b) that BV classification is more accurate than traditional trade classification from individual transactions using the tick rule is erroneous. Andersen and Bondarenko (2013) document that the relative accuracy is reversed, i.e., the traditional tick rule is more precise than the BV classification. This finding is also consistent with Chakrabarty, Pascual and Shkilko (2012), who evaluate the two classification rules on individual stock trades.

[16] As documented in Section 6.1, BV-VPIN is, by construction, responsive to volume and, in particular, realized volatility innovations. In line with our prior evidence for TR-VPIN, we document, in AB (2013), using a longer sample, that all predictive content of BV-VPIN for future return volatility is subsumed by a contemporaneous realized volatility measure.

based on evidence that goes beyond generic volatility forecasting power.

When conducting forecasting exercises, ELO (2012a) first transform the VPIN values using the empirical cumulative distribution function (CDF). While the CDF transformation facilitates comparison of VPIN levels across alternative implementations, when we fix a specific implementation and look at the time-series properties of VPIN, the rationale for this transformation is less clear. First, since the CDF transform is monotone, it does not change rankings. Hence, all our observations in this regard remain valid, even if judged by the CDF of VPIN rather than VPIN itself, including the point that TR-VPIN (VB-VPIN) exceeds the May 6, 2010, value at 13:30 during 26 (49) of the preceding days in the sample, constituting 4.3% (8.1%) of the days prior to the crash. Second, since the CDF transform dampens the extreme readings of VPIN, it tends to lose predictive power. In appendix C, we report the OLS regressions from Table 6, but using the CDF of VPIN in lieu of VPIN itself. We find that the corresponding R^2 statistics decrease marginally, but uniformly.[17]

7. Revisiting the flash crash

A potential concern about the preceding analysis is that it does not allow for the possibility that VPIN is relatively uninformative during benign times, but may become highly informative when order flow turns decidedly toxic. Indeed, one raison d'etre for VPIN is the ability to signal impending turbulence. This is a harder hypothesis to test from a short sample, so we rely on a descriptive account of the behavior of TR-VPIN and BV-VPIN across the dramatic events surrounding the flash crash. As above, we first present evidence relating to the TR-VPIN metric, exploited in ELO (2011a, 2011c), and subsequently check whether the conclusions are altered if we instead construct the BV-VPIN measure.

7.1 Replication and verifiability

The illustration in Section 4.1 shows that the TR-VPIN order imbalance measure can be sensitive to minor changes in the trading process. This raises the possibility that TR-VPIN itself may not be robust to small perturbations in the transaction record. Moreover, it is clearly dependent on the initial conditions as the point at which we start cumulating the trading volume determines the location of the buckets. It is not evident, however, if this is a major concern or whether the long moving average, used in computing VPIN, suffices to minimize the impact and stabilize the measure.

To assess the magnitude of such effects during critical scenarios, we compile 400 different versions of one-minute TR-VPIN for March 6, 2010, with each trajectory corresponding to a different location of the volume buckets. Using a bucket of $V = 40{,}000$, the first trajectory is initiated with the first transaction of the day, the second is initiated after the first 100 contracts have been traded, the third after 200 contracts are traded, and so on until the 400^{th} trajectory is initiated after 39,900 contracts have been traded. This provides a simple way of approximating the distribution of TR-VPIN on the day of the flash crash, reflecting the dependence on the initiation of the volume bucketing. In reality, the starting point of the sample, years earlier, determines the exact location of the volume buckets on this day. Hence, any shift in the initial condition alters the placement of the buckets on May 6, 2010, in much the same manner as in the experiment described above.

Figure 5 summarizes the effects of shifting the buckets on May 6, 2010. The upper left panel shows that TR-VPIN can fluctuate within a band approaching 20% of its median level across the trajectories, while the average width of the band is near 10%. The upper right panel displays only five of the

[17] We have likewise repeated all our other exercises using the VPIN CDF, and the case for the toxicity metric weakens slightly in all cases. These results are available upon request.

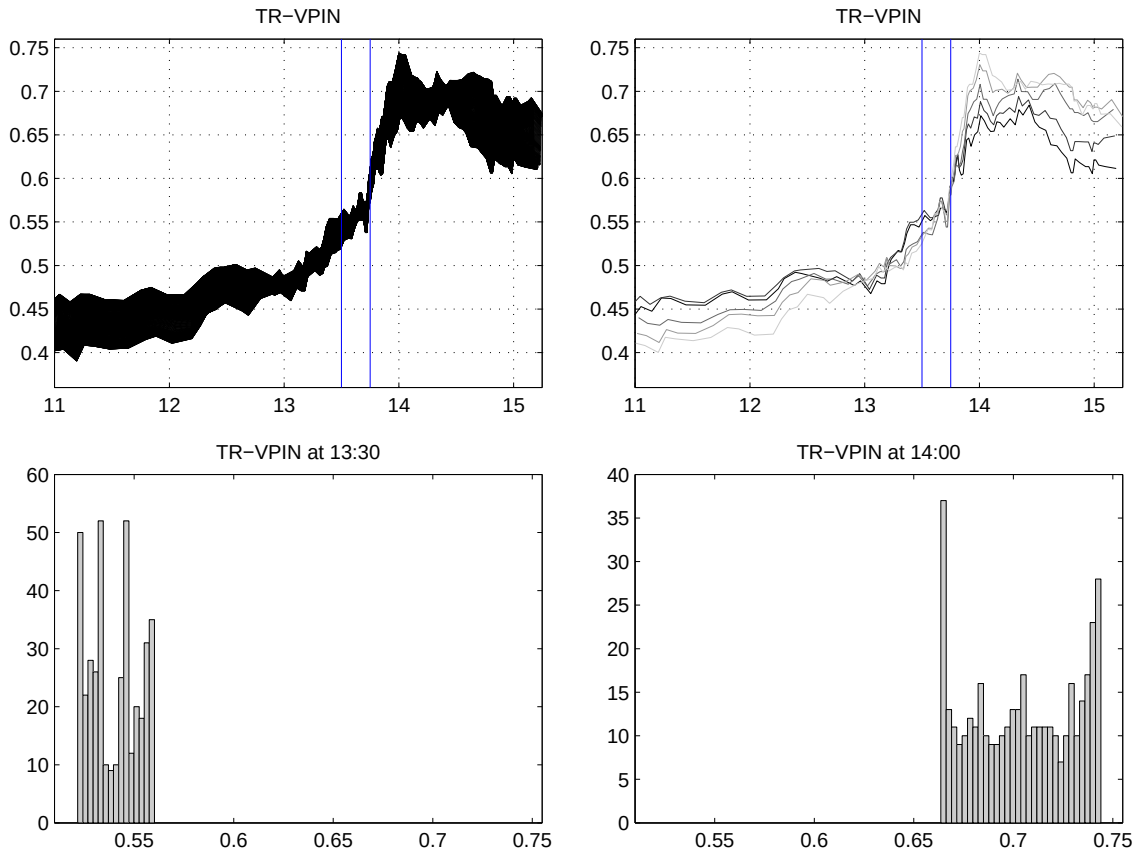

Figure 5: **Distribution of TR-VPIN values on May 6, 2010.** The figure illustrates the distribution of TR-VPIN values ($\delta = 60$ seconds) for different starting points of the volume grid. The bottom panels show the histograms of TR-VPIN values at 13:30 and 14:00. The top left panel shows the TR-VPIN trajectories generated by shifting the starting point every 100 contracts. The top right panel shows only five trajectories, which are selected based on the TR-VPIN values at 14:00 (min, 25%, 50%, 75%, and max).

trajectories to allow closer inspection. It is evident that the trajectories behave quite differently, even if they are correlated. For example, the black trajectory is the highest at 11:00, and it enters the crash period at the second highest level. However, this TR-VPIN trajectory actually drops during the crash and its subsequent rise is less pronounced than that associated with, say, the lightest shaded trajectory. In fact, the latter evolves similarly to the trajectory used in our earlier depictions of this trading day. Finally, the bottom panels display the range of values attained by the different TR-VPIN trajectories at the start of the crash (0.52-0.56) and at their maximum level (0.66-0.74). The left panel is relevant for judging whether VPIN attained a historically high prior to the crash while the right panel reflects the surge in trading intensity associated with the crash, which inevitably forces the metric to spike.

In light of these results, our overview of the TR-VPIN dynamics during the flash crash in Section 2 may be problematic as it is based on only one path among thousands of potential candidates. However, our conclusions, including Finding 1, remain unaffected. Specifically, the pre-crash value of 0.53 is typical of the scenarios depicted in the lower left panel of Figure 5. Moreover, while there is uncertainty about the TR-VPIN values, as reflected in the width of bands in the top panels, the qualitative features

are quite similar across the paths.[18]

Finding 11: The TR-VPIN metric at any specific point in time on a given trading day is sensitive to the exact sequence of trades recorded prior to that trading day. It implies that any change in the starting point of the sample or any removal of potentially invalid trades early in the sample will exert a potentially significant impact on the metric throughout the entire sample.

The observations above bring the issue of replication to the forefront. Unless there is agreement concerning the starting point as well as the recording and status of all transactions across two alternative data sources, competing computations of TR-VPIN will produce divergent behavior across critical periods in the sample. This presents practical problems for using the metric as an indicator of market stress or as the basis for a futures contract. While an exchange may take on the task of monitoring the relevant transaction series, it may become contentious that realized TR-VPIN depends on inclusion or exclusion of specific trades.[19]

Another concern is the potential lack of robustness of empirical work involving TR-VPIN. At a minimum, some standard for robustness should be adopted to ensure that conclusions do not hinge on idiosyncratic features of the design, such as the original starting point of the sample, the exact source of data, and the criterion for excluding "unusual" trades.

Finally, we reiterate Finding 1. Relative to our original TR-VPIN series, not a single one of our 400 trajectories for March 6, 2010, reach a historical high prior to the flash crash.

7.2 Signed TR-VPIN measures

The documented dispersion in TR-VPIN values may be surprising. One might expect random fluctuations in OI to diversify across the 50 moving average terms used to compute the metric. However, since the OI measures represent absolute values, there is no opportunity for positive and negative values to cancel each other out so, instead, random outliers tend to cumulate.

Figure 6 illustrates this feature. The right panels display the SOI measures associated with two of the TR-VPIN trajectories from Figure 5. Evidently, the black trajectory, by chance, has many more moderate SOI observations, falling within the range of [-0.50, 0], than the gray one during the period 12:00-14:00. This translates into a significantly different evolution of TR-VPIN, as depicted in the top left panel. The gray trajectory starts out well below the black one at 12:00 but ends up at a higher level at 14:00. However, if we average the signed – not the absolute – SOI measures, we obtain alternative smoothed signed order imbalance indicators. These are plotted in the lower left panel. Now, the black and gray trajectories basically coincide, showing that diversification across 50 observations is highly effective once we retain the sign of SOI. Hence, the signed TR-VPIN measure is far less sensitive to the positioning of the buckets and avoid stark dependencies on initial conditions. Moreover, it send a clear message: there was a growing dominance of active selling from noon until the end of the crash, and then an immediate reversal towards restoration of the cumulative order imbalance, which is complete before 15:00. At a more detailed level, we note that the truly dramatic collapse in price over the last 2-3 minutes of the crash period coincides with a huge jump in the (negative) order imbalance. Likewise, the

[18] We also investigated the robustness of our key findings by repeating the analysis for the average values of TR-VPIN, U1-VPIN, and U2-VPIN over the full sample across the 400 sample paths described above. The averaging reduces the noise in the series, so the TR-VPIN measures become even more highly correlated, and their correlations with volume and VIX go up marginally. Thus, these more precise measurements only serve to reinforce all major results. In particular, the regression evidence in Table 4 is strengthened further. The results for the averaged VPIN measures are provided in the Appendix. We thank the referee for suggesting this approach.

[19] As noted by the referee, a VPIN futures contract would also need to grapple with the issue of being robust to manipulation, as the innovations to VPIN will be sensitive to trades consummated at the end of each time bar or volume bin.

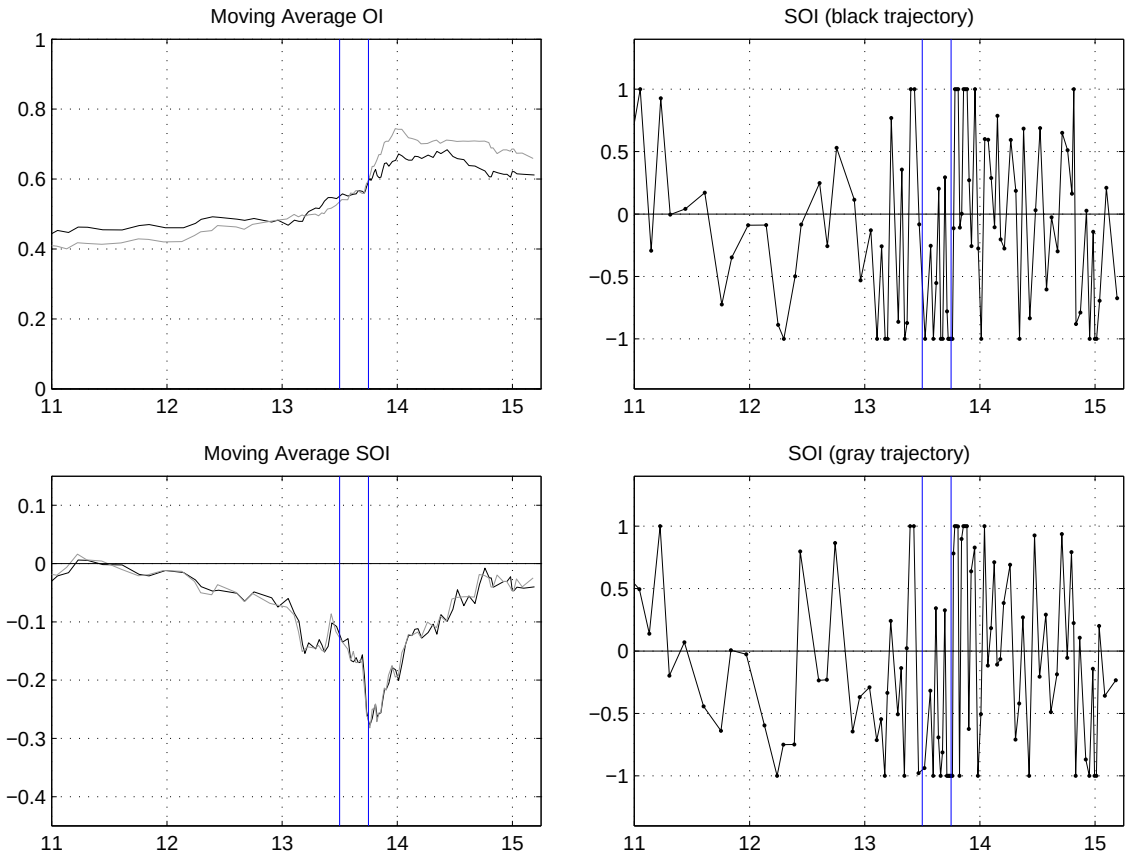

Figure 6: **Alternative TR-VPIN measures on May 6, 2010, using signed order imbalance.** The right panels depict TR-SOI ($\delta = 60$ seconds) for two different starting points of the volume grid, corresponding to the minimum (black) and maximum (gray) values of TR-VPIN at 14:00. The top left panel displays TR-VPIN, while the bottom left panel portrays the corresponding VPIN measures based on the *signed* order imbalance.

significant (negative) increase in the order imbalance just after 13:00 is striking and could, in retrospect, be seen as a forewarning of the ensuing turbulence.

We deem these preliminary observations intriguing. It is certainly feasible to construct cumulative (signed) order imbalance measures in real time. Nonetheless, one concern is that they may be extremely highly correlated with the realized price path and thus not provide much independent information. It is beyond the scope of this article to present a thorough analysis of signed VPIN style metrics.

Finding 12: The use of absolute (OI), rather than signed, order imbalance (SOI) measures in constructing TR-VPIN inflates the idiosyncratic variation of the metric. In contrast, the cumulative SOI is measured accurately and is helpful in identifying the trading activity that is driving the concurrent price changes. Nonetheless, it is unclear if the cumulative SOI can be used in *predicting* price changes and volatility rather than simply rationalizing them ex post.

7.3 A closer look at the dynamics of trading activity

The trading volume on May 6, 2010, exceeded 250% of the average for the sample, but the trade size distribution was close to that of a regular trading day, reported in Table 1. Figure 7 complements

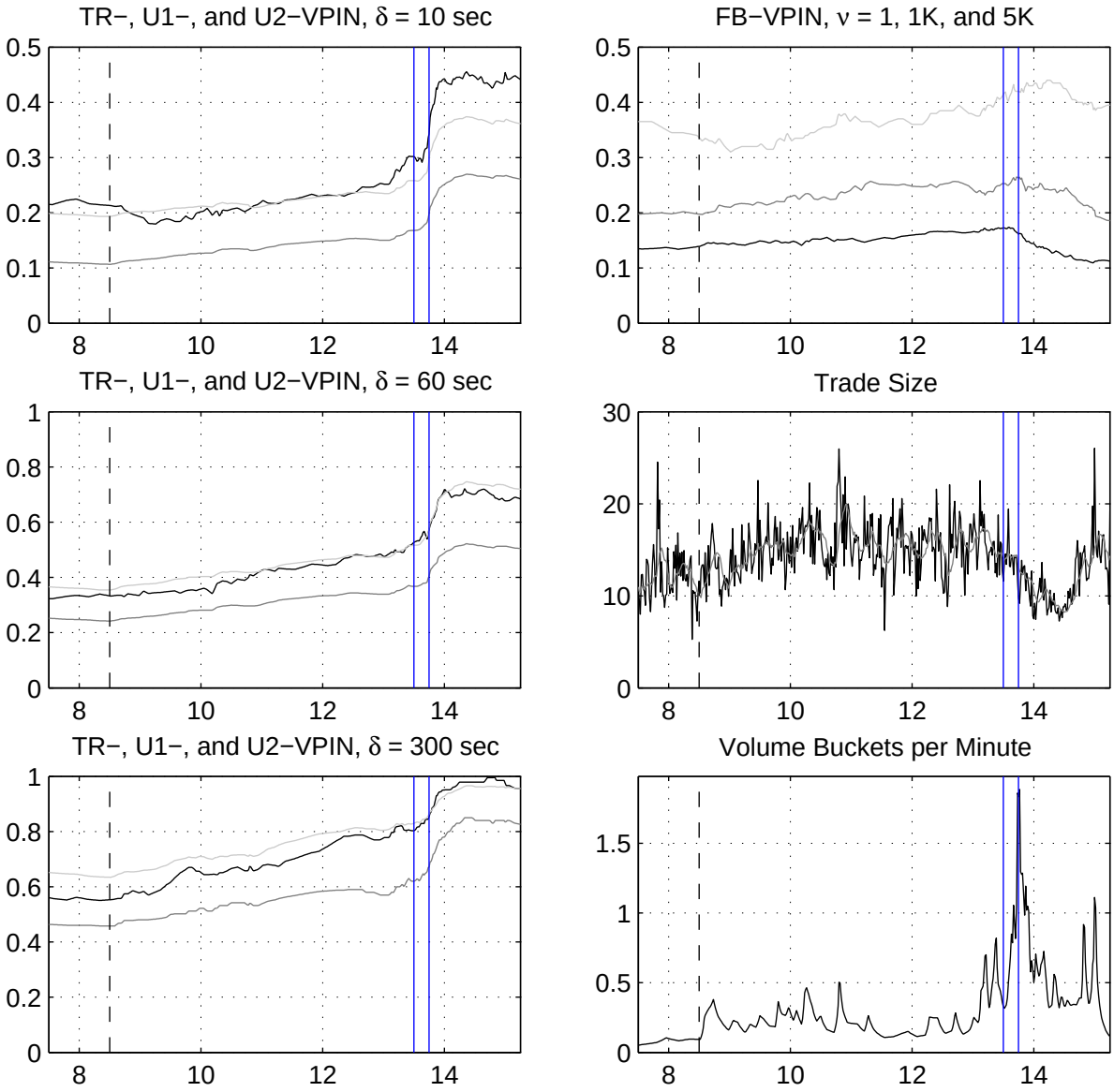

Figure 7: **Evolution of various VPIN and market activity variables on May 6, 2010.** The figure plots market statistics at the one-minute frequency from 7:30 to 15:15. The dashed vertical line shows the start of the regular trading hours, while the solid vertical lines indicate the timing of the "flash crash." Left panels: (1) TR-VPIN (black), U1-VPIN (dark gray), and U2-VPIN (light gray) for $\delta = 10$ sec, (2) same for $\delta = 60$ sec, (3) same for $\delta = 300$ sec. Right panels: (1) FB-VPIN for $v = 1$ (black), $v = 1K$ (dark gray), and $v = 5K$ (light gray); (2) the average trade size, (3) the fraction of $V = 40{,}000$ contracts traded per one-minute time bar.

Figure 2. It displays the evolution of alternative VPIN measures, the VIX, and the trading intensity, as given by the fraction of 40,000 contracts (one volume bucket) traded per minute, i.e., per time bar, throughout the day of the flash crash.

Focusing initially on the bottom right panel in Figure 7, we note that the trading intensity around the flash crash rose to levels which imply that each volume bucket between 13:10-15:00 contains no more

than three time bars, and, even more strikingly, from 13:40-13:55, the buckets were filled in less than one minute. Tautologically, this produces a string of TR-OI measures near unity, irrespective of the underlying order imbalances. This is confirmed by the qualitatively similar increase in U1-VPIN and U2-VPIN over this period for both $\delta = 60$ and $\delta = 10$. In other words, TR-VPIN *must* rise sharply, for mechanical reasons, due to the elevated trading activity. While it is possible we would never observe such a trading pattern in the absence of order flow toxicity, the dynamics of TR-VPIN simply cannot shed light on the issue. For VPIN to signal the impending chaos, it must attain extreme values *prior* to the crash. Turning towards this crucial timing dimension, we observe that the TR-VPIN and U-VPIN series for $\delta = 10$ and $\delta = 60$ reach their maximum after 14:00, and thus some time after the flash crash cycle has played itself out. As observed in Finding 1, the evidence for VPIN reaching an extreme level prior to the crash is much less compelling, but we pursue this question further in the subsequent section.

Finding 13: The explosive increase in TR-VPIN during and following the flash crash is fully explained by the underlying trading pattern, as evidenced by the qualitatively identical behavior of the U-VPIN series. In particular, the trading intensity rose to such levels that the OI measures mechanically were attaining the maximum value of unity, irrespective of the actual order flow imbalances.

A few additional observations on Figure 7 are warranted. First, transaction VPIN, or FB-VPIN with $v = 1$, again depicts an entirely different evolution than TR-VPIN. It is flat during the crash and then drops off sharply as the prices rebound, even as TR-VPIN continues to soar. FB-VPIN for $v = 1{,}000$ displays similar features, while the FB-VPIN for $v = 5{,}000$ represents an intermediate case between TR- and transaction VPIN. Second, the VIX index also remains elevated for a lengthy period and appears to reach a maximum more than an hour after the crash. However, as documented by Andersen, Bondarenko, and Gonzalez-Perez (2011), this is an artifact of dramatic swings in the liquidity of the S&P 500 options market, which distorts the computation of the index. If one applies a coherent range of option strikes within the VIX index formula across May 6, 2010, the index attains its maximum value exactly at the nadir of the S&P 500 index. Third, the extremely steep increase in all the TR- and U-VPIN series during the last couple of minutes during the actual crash period coincides with the downward jump in the SOI measure on Figure 6, suggesting this is the point in time when the depth (on the bid side) of the order book truly evaporates. The key question is whether any of our metrics reliably can be used to predict such events in advance.

7.4 TR-VPIN as crash predictor

The issue of whether TR-VPIN provided a clear signal indicating sharply rising order toxicity prior to the flash crash cannot be conclusively answered from our limited historical sample. However, we can summarize the evidence as it would appear at the time just prior to the crash and seek to infer whether the prevailing real-time value of TR-VPIN was exceptional compared to the recent history.

To this end, we construct three separate figures with scatter plots depicting pairwise observations of alternative volatility predictors (on the horizontal axis) versus subsequent realized volatility (on the vertical axis) over different forecast horizons. The associated regression lines convey the predictions from a simple linear model exploiting available historical evidence. The plots cover the beginning of our sample through the start of the flash crash, thus reflecting the real-time perspective an observer may have acquired at that juncture by constructing the candidate volatility predictors on a high-frequency basis. The gray dots indicate the predictor-volatility pairs constructed from data more than five days before the flash crash, while the black dots refer to the forecast-realization pairs obtained within the last five days prior to the crash. Note that we exploit non-overlapping forecast horizons to avoid excessive cluttering of the displays.

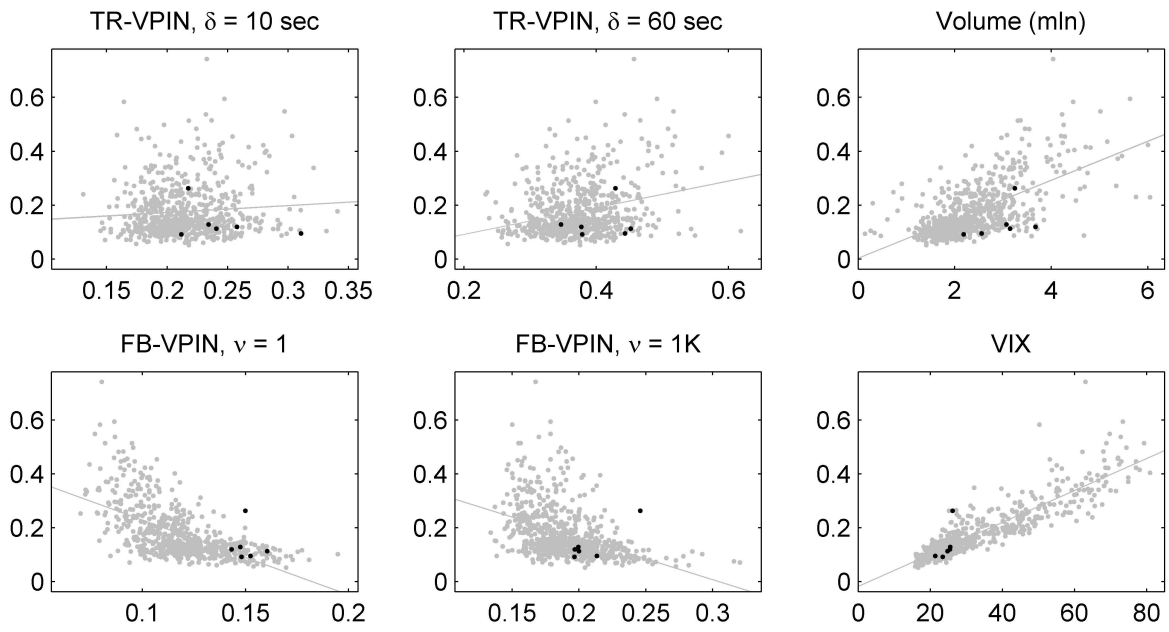

Figure 8: **Average absolute 50-day returns versus TR-VPIN, volume and volatility measures.** The figure shows six scatter plots of average absolute return $AAR(t,t+50)$ versus TR-VPIN with $\delta = 10$ and 60 sec, versus FB-VPIN with $\nu = 1$ and 1K, volume, and VIX. Volume is one-day trading volume, in millions. $AAR(t,t+T) = \frac{1}{T}\sum_{i=1}^{T}|r_{t+i}|$, where $r_t = 100\ln(P_t/P_{t-1})$ is the log return of the S&P 500 futures over one volume bucket. Black dots indicate the five days preceding 13:30 on May 6, 2010. The sample period is January 1, 2008 - May 6, 2010.

Figure 8 depicts results for the 50-volume-bucket-ahead forecast horizon. For all predictors, it is hard to detect any unusual pre-crash pattern: even if TR-VPIN clearly is elevated it is not exceptional, and the realized volatilities are fairly subdued compared with the values attained during the financial crisis. We also note the pronounced negative relation between transaction and FB-VPIN and the future realized volatility as well as the more well-defined positive association between the volume and VIX series and future realized absolute returns. In fact, the generally limited explanatory power of TR-VPIN relative to a number of the other candidate predictor variables is evident by the diffuse shape of the TR-VPIN scatter plots. Nonetheless, it is also clear that none of the other variables provide any clear indication that volatility is about to erupt prior to the crash. For none of the displays, we detect a noticeable disconnect between the grey and black dots.

One problem with the relatively long forecast horizon employed in Figure 8 is that it limits the number of pre-crash observations quite severely, given our use of non-overlapping forecasts. We therefore turn to shorter horizons for a more rich set of data points. The disadvantage is that the scatter plots will tend to more dispersed as the realized absolute returns over shorter horizons provide noisier measures of the underlying volatility, see, e.g., Andersen and Bollerslev (1998).

The plots in Figure 9, reflecting a forecast horizon of ten buckets, reveal qualitatively similar patterns. Likewise, moving to the extreme end of the spectrum in Figure 10 – forecasting only the absolute return over the next volume bucket – does not alter the impression that TR-VPIN fluctuates within a mildly elevated, but fairly standard, range prior to the crash. The main difference is that all "clouds" become more diffuse, reflecting the use of very imprecise volatility proxies. In summary,

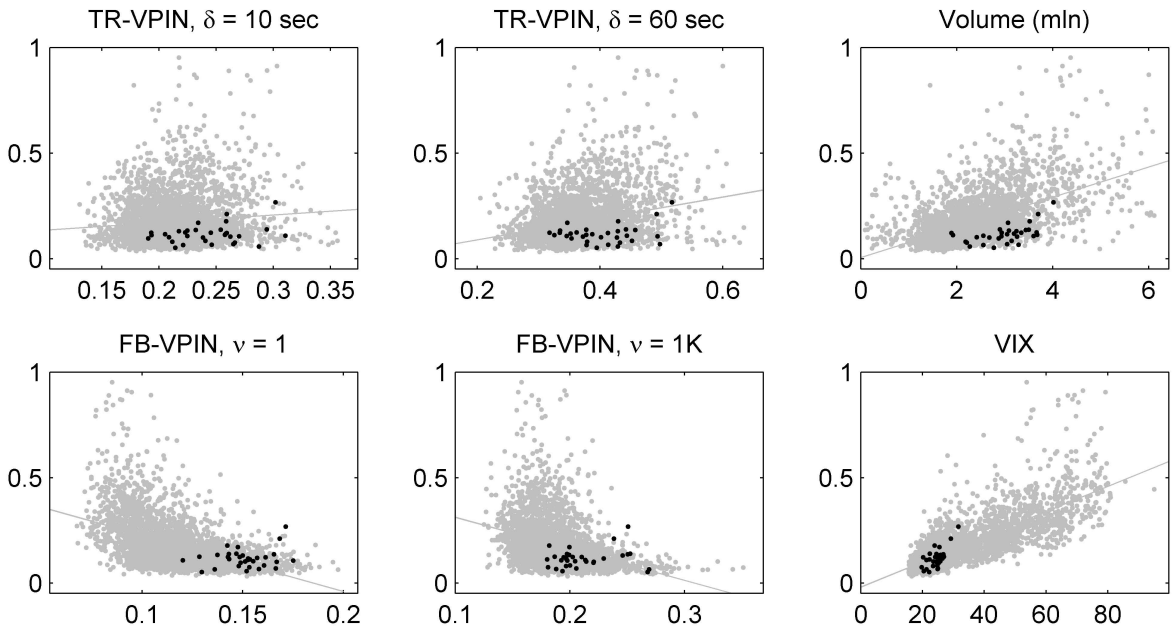

Figure 9: **Average absolute 10-day returns versus TR-VPIN, volume and volatility measures.** The figure shows six scatter plots of average absolute return $AAR(t,t+10)$ versus TR-VPIN with $\delta = 10$ and 60 sec, versus FB-VPIN with $\nu = 1$ and 1K, Volume, and VIX. Volume is one-day trading volume, in millions. $AAR(t,t+T) = \frac{1}{T}\sum_{i=1}^{T}|r_{t+i}|$, where $r_t = 100\ln(P_t/P_{t-1})$ is the log return of the S&P 500 futures over one volume bucket. Black dots indicate the five days preceding 13:30 on May 6, 2010. The sample period is January 1, 2008 - May 6, 2010.

when focusing on the pre-crash sample, there is no indication that real-time TR-VPIN measures would have forewarned an observer of the impending turmoil.[20]

Finding 14: The evolution of TR-VPIN series prior to the flash crash does not appear genuinely remarkable along any dimension. In particular, the level of the series was elevated relative to the average day, but it was not reaching values close to historical extremes.

7.5 BV-VPIN and the flash crash

In ELO (2012a), the implementation of VPIN is based on so-called bulk volume classification, leading to the BV-VPIN metric. We now assess whether any of the critical issues surrounding the behavior of the VPIN measure on the day of the flash crash is sensitive to this change in metric.

First, we reproduce the analysis of Section 7.1 using BV-VPIN in lieu of TR-VPIN. Figure 11 provides the analogue of Figure 5.

As expected, given the analysis in Section 6.1, the level of BV-VPIN is significantly lower than for TR-VPIN, especially in the early parts of the day. In addition, the dispersion across the 400 trajectories is slightly lower, which combines to generate a degree of dispersion relative to the median value which is roughly of the magnitude we found for TR-VPIN. For this particular day, the issue regarding replication of the actual trajectory seems to be qualitatively similar to our findings for TR-VPIN.

[20] We have confirmed that this conclusion is robust to the choice of sample period and, in particular, that it is not dependent on the inclusion of the highly volatile period from August 2008 to May 2009.

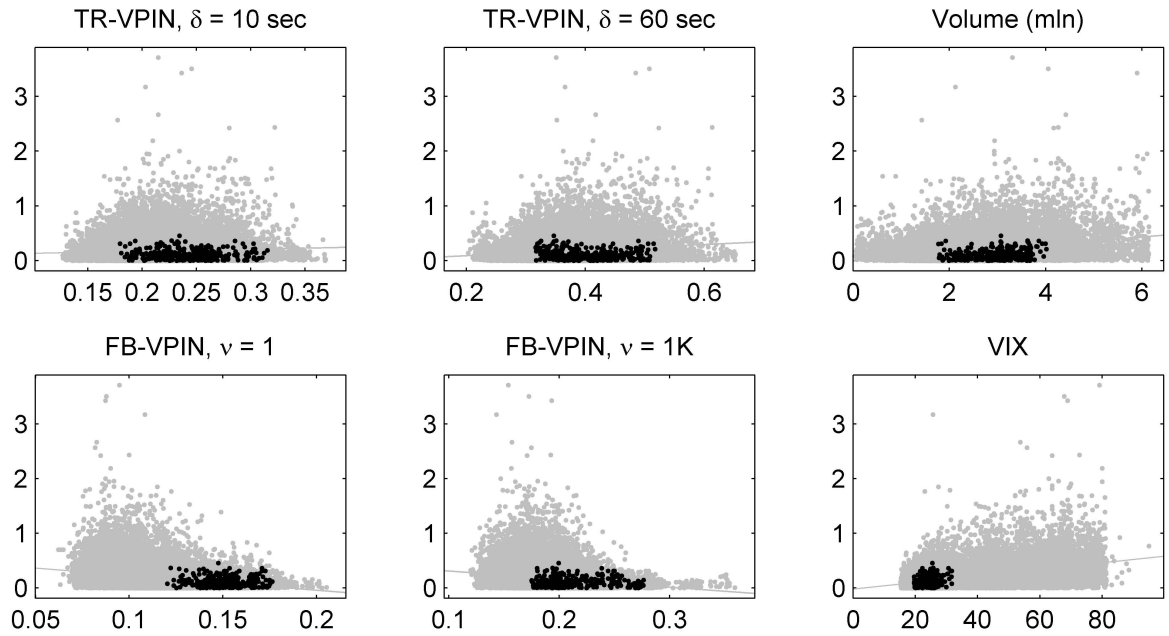

Figure 10: **Average absolute one-day returns versus TR-VPIN, volume and volatility measures.** The figure shows six scatter plots of average absolute return $AAR(t,t+1)$ versus TR-VPIN with $\delta = 10$ and 60 sec, versus FB-VPIN with $\nu = 1$ and 1K, Volume, and VIX. Volume is one-day trading volume, in millions. $AAR(t,t+T) = \frac{1}{T}\sum_{i=1}^{T}|r_{t+i}|$, where $r_t = 100\ln(P_t/P_{t-1})$ is the log return of the S&P 500 futures over one volume bucket. Black dots indicate the five days preceding 13:30 on May 6, 2010. The sample period is January 1, 2008 - May 6, 2010.

To obtain a more detailed perspective, we also reproduce Figure 6 using BV-VPIN rather than TR-VPIN.

The left panels of Figure 12 are remarkably consistent with the corresponding panels in Figure 6. One trajectory is steadily increasing throughout the trading day, while the other is quite stable up to about 13:00. At that point, both display a fairly steep ascent, but the path that started out at a lower level in the morning ends up significantly above the other path by 14:00. Thus, the discrepancy between the trajectories can be quite striking. Nonetheless, the associated *signed* order imbalance measure in the lower left panel again shows remarkable consistency across the two realizations, confirming the findings from Section 6.2.

The right panels of Figure 12 are also quite informative. Prior to 13:00, the SOI trajectories are muted relative to the corresponding paths on Figure 6. However, from 13:00 and, in particular, during and following the crash, the BV-SOI measures amplify greatly to resemble those from Figure 6 in all respects. That is, once the trading intensity and price volatility rise sharply, the innovations in the TR- and BV-OI measures become near indistinguishable. As discussed in Section 6.1, this implies that the *relative* increase in BV-VPIN is more dramatic than for TR-VPIN during volatile episodes. As such, the crash event provides a nice illustration of the effect induced by a simultaneous escalation of trading intensity and price volatility, highlighted theoretically in equation (13).

Of course, the ultimate question is whether the level of the BV-VPIN metric was foreshadowing the crash event. On this point, the evidence seems clear and non-confirmatory. The level of BV-VPIN at 13:00, when the trajectories in Figure 12 largely coincide, was nowhere close to historical highs, as

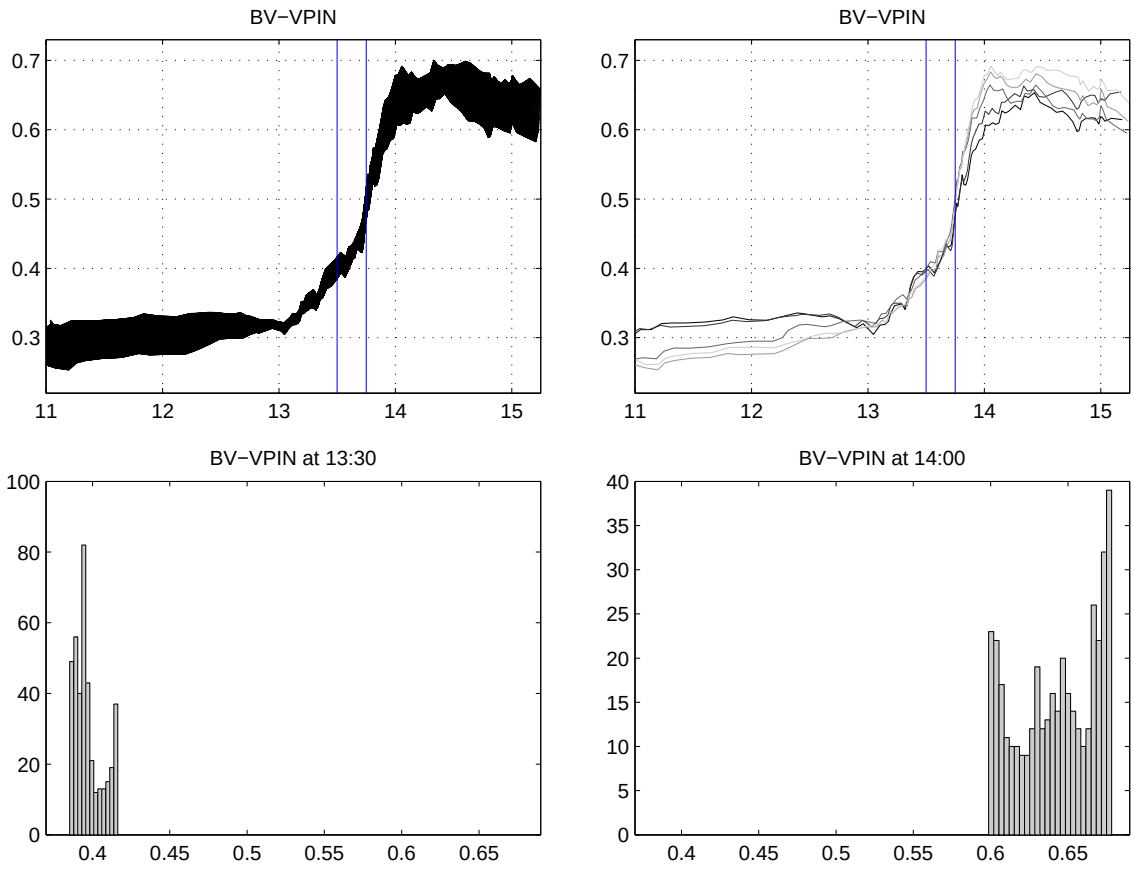

Figure 11: **Distribution of BV-VPIN values on May 6, 2010.** The figure depicts the distribution of BV-VPIN values ($\delta = 60$ sec) for different starting points of the volume grid. The bottom panels show the histograms of BV-VPIN values at 13:30 and 14:00. The top left panel shows the BV-VPIN trajectories generated by shifting the starting point every 100 contracts. The top right panel shows only five trajectories, which are selected based on the BV-VPIN values at 14:00 (min, 25%, 50%, 75%, and max).

can be seen from Figure 1. Even after the relatively steep increase from 13:00 to 13:30, the maximum value attained across the 400 trajectories in Figure 11 falls well within the range of values observed previously in our sample. In fact, from this perspective, the level of the BV-VPIN metric at either 13:00 or 13:30 seems to provide less of a signal about impending turmoil than the TR-VPIN metric.

In summary, we do not find any indication that BV-VPIN outperforms TR-VPIN in predicting potential flash crashes. The metric is, by construction, more highly correlated with price volatility than TR-VPIN, but this does not translate into much predictive power for future volatility relative to standard volatility forecast measures, as documented in Table 6. Moreover, the level of the BV-VPIN metric failed to provide a clear signal of potential trouble ahead of the flash crash.

8. Conclusion

Inspired by the striking empirical evidence of ELO (2011a), we examine the TR-VPIN measure and its ability to forecast short-term price volatility and signal impending market turmoil. Our results are largely non-confirmatory. First, TR-VPIN, almost by construction, is highly correlated with trading

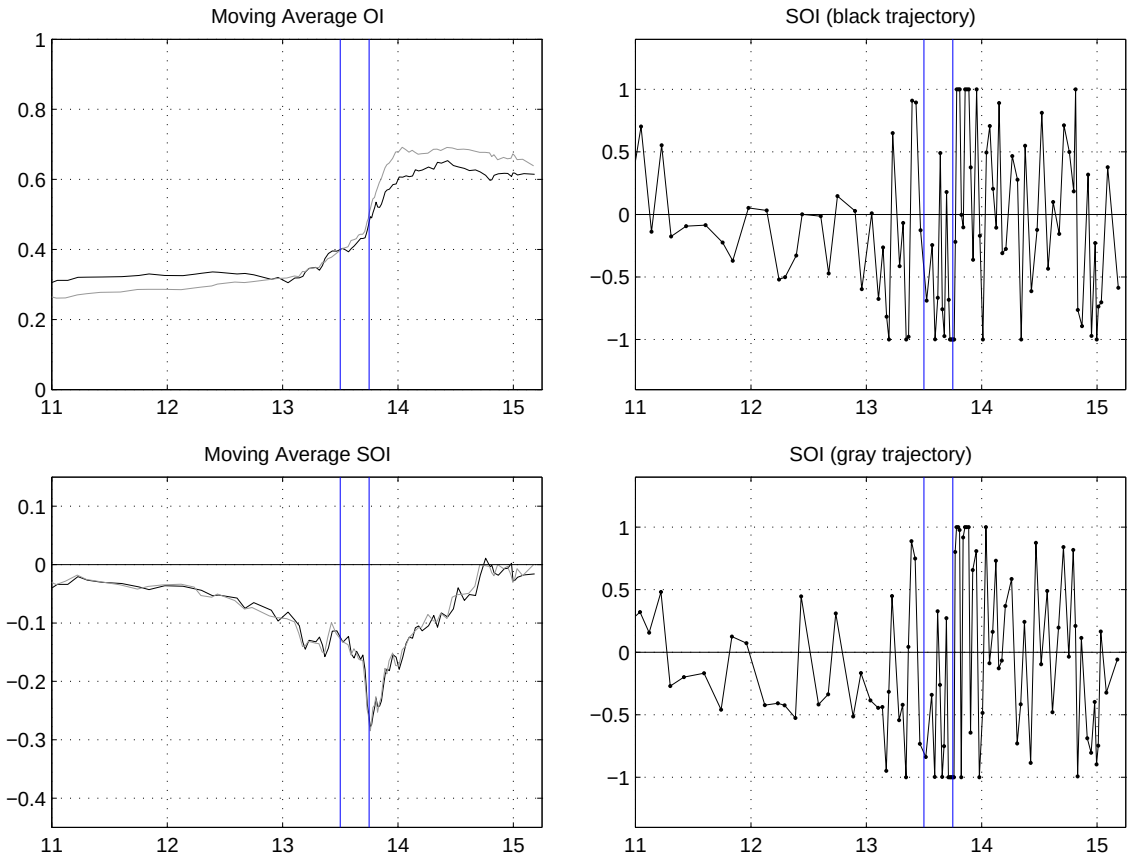

Figure 12: **Alternative BV-VPIN measures on May 6, 2010, using signed order imbalance.** The right panels depict BV-SOI ($\delta = 60$ sec) for two different starting points of the volume grid, corresponding to the minimum (black) and maximum (gray) values of BV-VPIN at 14:00. The top left panel displays BV-VPIN, while the bottom left panel portrays the corresponding VPIN measures based on the *signed* order imbalance.

volume and thus will tend to covary also with the current and future volatility level. However, once we control for the component of VPIN that is driven by the volume dynamics, we find no incremental information in TR-VPIN concerning future short-term volatility. Similarly, constructing TR-VPIN style measures from fixed volume bins or transaction data, avoiding the mechanical correlation with trading volume, we find VPIN to be negatively correlated with future volatility – a feature we tentatively ascribe to a negative correlation between the volatility level and the trade classification rule. This may potentially arise from a drop in the depth of the limit order book and smaller transaction sizes as volatility and economic uncertainty increases. Second, we found the TR-VPIN metric prior to the crash on May 6, 2010, to be elevated but less than extraordinary. Moreover, the VPIN dynamics throughout this day is readily accounted for by the trading pattern. As such, the identification of order flow imbalances via the trade classification scheme has no visible impact on TR-VPIN over this period.

Next, we extend the analysis to cover the recent modification of VPIN through the so-called bulk volume trade classification scheme. The resulting BV-VPIN metric is constructed by smoothing the absolute value of monotone transformations of one-minute price changes, and thus inherits features usually associated with a realized volatility measure, along with the dependence on the trading intensity embedded in the TR-VPIN metric. We confirm that BV-VPIN is strongly correlated with both volatility

and trading intensity. Unfortunately, its predictive power regarding impending market volatility falls well short of what can be obtained by existing real-time volatility indicators. Moreover, there is no evidence that it improves upon the TR-VPIN metric in terms of signaling the onset of the flash crash.

Our findings suggest that the TR- and BV-VPIN metrics do not offer the most fruitful ways of monitoring order flow imbalances and market tensions, so additional experimentation with alternative trade classification schemes may be warranted. For example, the signed version of VPIN is less sensitive to initial conditions and may be more suitable for direct monitoring of cumulative (signed) order imbalances. Nonetheless, any procedure adopting a VPIN metric solely by modifying the trade classification rule, including the signed VPIN measure, would still be subject to a number of the potential distortions analyzed in this paper, so in-depth analysis is warranted prior to adoption.

A different and useful approach to assess the potential of the basic VPIN approach would be to establish how the procedure works under perfect classification of active buys and sells. This would enable direct analysis of the precision of any given classification rule, and it would serve as the natural benchmark for results obtained via TR-, BV-, and FB-based measures of order imbalance. In principle, this is feasible for a centralized electronic order book market such as the E-mini S&P 500 futures, as trades are typically consummated when an existing bid or ask price is hit. However, this identification requires access to an even more detailed data set, including limit order book information and reliable sequencing of trades versus quote revision events.

In conclusion, while a real-time statistic for gauging the prevailing order imbalance and predicting episodic market stress scenarios is in high demand, we conclude that current incarnations of the VPIN metric are not ideal. The search for such a metric is high on the research agenda. We hope the type of analysis undertaken in this article will be helpful in identifying the robust and promising market stress indicators among the large set of candidate measures that may be proposed.

Appendix

A Proof of equations (5) and (6)

Consider Q independent binary random variables $b_1, b_2, \ldots, b_Q$, where $b_i = \pm 1$. Let $S_Q = \sum_{i=1}^{Q} b_i$ denote their sum. The function $F(Q)$ can be computed using the expectation of the absolute value $|S_Q|$ as:

$$F(Q) = \frac{E\left[|S_Q|\right]}{Q} = \frac{1}{2^Q Q} \sum_{i=0}^{Q} C_Q^i \cdot |Q - 2 \cdot i|,$$

where

$$C_n^k = \frac{n!}{k!(n-k)!}$$

is the binomial coefficient, sometimes also denoted as $C(k,n)$, or $\binom{n}{k}$. We are going to use the method of mathematical induction to prove that:

$$G(Q) := \sum_{i=0}^{Q} C_Q^i \cdot |Q - 2 \cdot i| = \begin{cases} (2q) C_{2q}^q, & \text{if } Q = 2q \\ 2(2q+1) C_{2q}^q, & \text{if } Q = 2q+1 \end{cases} \tag{15}$$

It is easy to verify that the relationship in (15) is true for $Q = 1$, 2, 3, for which $G(Q) = 1$, 4, 12. Next we prove in two separate cases that (1) if the relationship is true for some $Q = 2q$, then it is also true for $Q = 2q+1$, and (2) if the relationship is true for $Q = 2q+1$, then it is also true for $Q = 2q+2$. In both cases, we rely on two basic properties of the binomial coefficients:

(i) $C_n^0 = C_n^n = 1$,

(ii) $C_{n+1}^{k+1} = C_n^k + C_n^{k+1}$, for all $0 \le k \le n-1$.

Case 1: Suppose that $G(2q) = (2q) C_{2q}^q$. We can re-write:

$$G(2q+1) := \sum_{i=0}^{2q+1} C_{2q+1}^i \cdot |2q+1-2 \cdot i|$$

$$= C_{2q+1}^0 \cdot (2q+1) + C_{2q+1}^1 \cdot (2q-1) + \ldots + C_{2q+1}^q \cdot 1 + C_{2q+1}^{q+1} \cdot 1 + \ldots + C_{2q+1}^{2q+1} \cdot (2q+1).$$

The above expression is the sum of $(2q+2)$ products, for which the first factor takes values $C_{2q+1}^0, C_{2q+1}^1, \ldots C_{2q+1}^{2q+1}$ and the second factor takes values $(2q+1), (2q-1), \ldots, 3, 1, 1, 3, \ldots, (2q-1), (2q+1)$. Using properties (i) and (ii), we substitute $C_{2q+1}^0 = C_{2q}^0$, $C_{2q+1}^i = C_{2q}^{i-1} + C_{2q}^i$ for $1 \le i \le 2q$, and $C_{2q+1}^{2q+1} = C_{2q}^{2q+1}$. Re-arranging terms, we obtain:

$$G(2q+1) = C_{2q}^0 \cdot (4q) + C_{2q}^1 \cdot (4q-4) + \ldots + C_{2q}^{q-1} \cdot 4 + C_{2q}^q \cdot 2 + C_{2q}^{q+1} \cdot 4 + \ldots + C_{2q}^{2q} \cdot (4q)$$

$$= 2G(2q) + C_{2q}^q \cdot 2 = C_{2q}^q \cdot (4q+2) = 2(2q+1) C_{2q}^q.$$

Case 2: Suppose that $G(2q+1) = 2(2q+1) C_{2q}^q$. Proceeding similarly to the previous case,

$$G(2q+2) := \sum_{i=0}^{2q+2} C_{2q+2}^i \cdot |2q+2-2 \cdot i|$$

$$= C_{2q+2}^0 \cdot (2q+2) + C_{2q+2}^1 \cdot (2q) + \ldots + C_{2q+2}^q \cdot 2 + C_{2q+2}^{q+2} \cdot 2 + \ldots + C_{2q+2}^{2q+2} \cdot (2q+2)$$

$$= C_{2q+1}^0 \cdot (4q+2) + C_{2q+1}^1 \cdot (4q-2) + \ldots + C_{2q+1}^q \cdot 4 + C_{2q+1}^{q+1} \cdot 4 + \ldots + C_{2q+1}^{2q+1} \cdot (4q+2)$$

$$= 2G(2q+1) = 4(2q+1) C_{2q}^q = \frac{(2q)!}{q!q!} \cdot (2q+1) \frac{(2q+2)(2q+2)}{(q+1)(q+1)}$$

$$= (2q+2)\frac{(2q+2)!}{(q+1)!(q+1)!} = (2q+2)C_{2q+2}^{q+1}.$$

The proof of (5) now obtains from (15) because:

$$F(Q) = \frac{G(Q)}{2^Q Q} = \frac{1}{2^{2q}} C_{2q}^{q} \qquad \text{if} \quad Q = 2q, \quad \text{or} \quad Q = 2q+1.$$

The proof of (6) follows from (5) and Stirling's approximation for large factorials:

$$Q! \sim \sqrt{2\pi Q}\left(\frac{Q}{e}\right)^Q.$$

■

B Average VPIN measures

We explore how much of the time series variation in the VPIN measures is due to the noise associated with the initial conditions. For each aggregation level, $\delta = 10$, $\delta = 60$, and $\delta = 300$, we compute 400 alternative versions of VPIN, U1-VPIN, and U2-VPIN, corresponding to different starting points, each 100 contracts apart. Then we compute the average across the 400 trajectories. Tables 7-8 repeat the results of Tables 2 and 4, but using the averaged versions of VPIN.

Table 7: **Correlations between various average VPIN and market activity variables**

	$\delta = 10$				
	TR-VPIN	U1-VPIN	U2-VPIN	Volume	VIX
TR-VPIN	1.00			0.57	0.08
U1-VPIN	0.75	1.00		0.84	0.47
U2-VPIN	0.81	0.86	1.00	0.61	0.14

	$\delta = 60$				
	TR-VPIN	U1-VPIN	U2-VPIN	Volume	VIX
TR-VPIN	1.00			0.67	0.24
U1-VPIN	0.83	1.00		0.85	0.48
U2-VPIN	0.86	0.99	1.00	0.82	0.40

	$\delta = 300$				
	TR-VPIN	U1-VPIN	U2-VPIN	Volume	VIX
TR-VPIN	1.00			0.75	0.39
U1-VPIN	0.88	1.00		0.84	0.46
U2-VPIN	0.89	0.97	1.00	0.83	0.48

Notes: VPIN measures are computed as averages of 400 alternative versions, corresponding to different starting point. The table reports correlations of average TR-VPIN, U1-VPIN, and U2-VPIN for $\delta = 10$, 60, and 300 sec, one-day trading volume, and VIX. The sample period is January 2008 - July 2010.

Table 8: **Forecast regressions for absolute return, average VPIN**

Panel A: One-period forecast

	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
Const.	-0.04	-0.16	0.02	-0.02	-0.10	0.08	-0.09	-0.10	-0.05	-0.07
	(-1.33)	(-6.18)	(1.92)	(-4.21)	(-3.87)	(3.52)	(-6.22)	(-6.67)	(-6.91)	(-4.24)
TR-VPIN	0.57				-0.70	-0.21	0.21	0.01		0.01
	(7.34)				(-8.30)	(-3.11)	(5.57)	(0.12)		(0.28)
U1-VPIN		1.23			1.99			0.36		0.09
		(12.22)			(16.14)			(4.92)		(1.01)
Vol$\times 10^{-7}$			0.63			0.72			0.20	0.16
			(12.12)			(12.60)			(6.83)	(3.69)
VIX$\times 10^{-2}$				0.60			0.58	0.55	0.54	0.53
				(33.66)			(35.94)	(32.08)	(34.31)	(32.48)
$\bar{R}^2$	2.76	7.72	8.93	21.25	9.05	9.15	21.62	21.77	21.90	21.92

Panel B: 50-period forecast

	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9	Reg 10
Const.	-0.02	-0.15	0.02	-0.02	-0.09	0.10	-0.07	-0.08	-0.04	-0.04
	(-0.75)	(-6.19)	(2.16)	(-4.08)	(-3.63)	(5.08)	(-5.96)	(-6.61)	(-7.23)	(-3.41)
TR-VPIN	0.52				-0.78	-0.29	0.16	-0.08		-0.07
	(7.44)				(-9.76)	(-4.65)	(5.19)	(-2.03)		(-1.85)
U1-VPIN		1.19			2.03			0.41		0.12
		(12.63)			(16.71)			(6.26)		(1.51)
Vol$\times 10^{-7}$			0.62			0.74			0.19	0.18
			(12.67)			(13.23)			(7.49)	(4.99)
VIX$\times 10^{-2}$				0.59			0.58	0.54	0.53	0.53
				(37.14)			(38.93)	(36.35)	(37.64)	(36.54)
$\bar{R}^2$	8.85	27.75	33.64	78.35	33.99	35.16	79.11	79.86	80.50	80.56

Notes: The figures represent OLS regression coefficients; t-statistics based on HAC-standard errors, constructed with 50 lags, are reported in parentheses. TR-VPIN and U2-VPIN are for $\delta = 60$ seconds and are computed as averages of 400 alternative versions, corresponding to different starting points; "Vol" is the one-day backward trading volume. The sample period is January 2008 - July 2010.

C Forecast regressions for absolute return using VPIN CDF

This appendix explores the predictive power of the CDF transformation of VPIN measures for future short run return volatility. Table 9 reports on the same type of OLS regressions as in Table 6, except that the raw VPIN metrics are replaced with the corresponding VPIN CDF measures. The regressions for volume and VIX are repeated to facilitate comparison.

Table 9: **Forecast regressions for absolute return using VPIN CDF**

	One-period forecast					**50-period forecast**				
	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5
Const.	0.13	0.09	0.07	0.02	-0.02	0.13	0.09	0.08	0.02	-0.02
	(23.45)	(18.83)	(17.84)	(1.92)	(-4.21)	(24.91)	(20.53)	(19.39)	(2.16)	(-4.08)
TR-VPIN CDF	0.09					0.09				
	(7.25)					(7.34)				
U1-VPIN CDF		0.17					0.17			
		(13.97)					(14.54)			
BV-VPIN CDF			0.20					0.20		
			(17.23)					(17.86)		
Vol$\times 10^{-7}$				0.63					0.62	
				(12.12)					(12.67)	
VIX$\times 10^{-2}$					0.60					0.59
					(33.66)					(37.14)
$\bar{R}^2$	1.97	7.16	9.70	8.93	21.25	6.83	26.61	35.80	33.64	78.35

Notes: The figures represent OLS regression coefficients; t-statistics based on HAC-standard errors, constructed with 50 lags, are reported in parentheses. Unlike Table 6, this table focuses on the CDF transformation of VPIN measures, instead of raw VPIN measures. TR-VPIN, U1-VPIN, and BV-VPIN are for $\delta = 60$ seconds; "Vol" is the one-day backward trading volume. The regressions for volume and VIX are the same as in Table 6 and are repeated to facilitate direct comparison. The forecast horizon is one volume bucket ("one period") and fifty volume buckets ("50 periods"), respectively. The sample period is January 2008 - July 2010.

References

Andersen, T.G., 1996. Return volatility and trading volume: an information flow interpretation of stochastic volatility. Journal of Finance 51, 169–204.

Andersen, T.G., Bollerslev, T., 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

Andersen, T.G., Bondarenko, O., 2007. Construction and interpretation of model-free implied volatility. In: Nelken, I. (Ed.), Volatility as an Asset Class. Risk Books, London, pp. 141–181.

Andersen, T.G., Bondarenko, O., 2013. Assessing VPIN measurement of order flow toxicity via perfect trade classification. Working Paper, Northwestern University and University of Illinois at Chicago.

Andersen, T.G., Bondarenko, O., Gonzalez-Perez, M.T., 2011. A corridor fix for VIX: constructing a coherent model-free option implied volatility measure. Working Paper, Northwestern University and University of Illinois at Chicago.

CFTC-SEC, 2010. Preliminary findings regarding the market events of May 6, 2010. Joint Commodity Futures Trading Commission (CFTC) and the Securities and Exchange Commission (SEC) Advisory Committee on Emerging Regulatory Issues, May 18, 2010.

Chakrabarty, B., R. Pascual and A. Shkilko, 2012. Trade classification algorithms: a horse race between the bulk-based and the tick-based rules. Working Paper, SSRN, December 2012.

Clark, P., 1973. Liquidity, information, and infrequently traded stocks. Econometrica 41, 135–156.

Epps, T.W., Epps, M.L., 1976. The stochastic dependence of security price changes and transactions volumes: implications for the mixture of distributions hypothesis. Econometrica 44, 305–321.

Easley, D., López de Prado, M., O'Hara, M., 2011a. The microstructure of the "flash crash": flow toxicity, liquidity crashes, and the probability of informed trading. Journal of Portfolio Management 37 (2), 118–128.

Easley, D., López de Prado, M., O'Hara, M., 2011b. The exchange of flow toxicity. Journal of Trading 6 (2), 8–13.

Easley, D., López de Prado, M., O'Hara, M., 2011c. Flow toxicity and volatility in a high frequency world. Working Paper, SSRN, February 2011.

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

Easley, D., López de Prado, M., O'Hara, M., 2012b. Bulk classification of trading activity. SSRN, March 2012.

Kirilenko, A., Kyle, A.S., Samadi, M., Tuzun, T., 2011. The flash crash: the impact of high frequency trading on an electronic market. Working Paper, SSRN, May 2011.

Tauchen, G.E., Pitts, M., 1983. The price variability–volume relationship on speculative markets. Econometrica 51, 485–505.