VPIN and the Flash Crash*

Torben G. Andersen† Oleg Bondarenko‡

July 2011 (first version: February 2011)

Abstract

The VPIN, or Volume-Synchronized Probability of INformed trading, metric is introduced by Easley, López de Prado and O'Hara as a real-time indicator of order flow toxicity. They find the measure useful in measuring evolving order flow imbalances which 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. Our empirical investigation of VPIN documents that it is a relatively 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. Our preliminary experimentation suggests that modified variants of VPIN contain superior information which may be more effective as real-time market stress indicators. Nonetheless, we caution against the adoption on any such specific metric until it has been shown to be robust to the type of analysis and tests that we undertake in the present paper.

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 thank Craig Furfine, Kathleen Hagerty, Robert McDonald, Maureen O'Hara, Mark Ready, and seminar participants at University of Illinois at Chicago 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; e-mail: t-andersen@kellogg.northwestern.edu

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

Electronic copy available at: http://ssrn.com/abstract=1881731

1 Introduction

In a recent series of articles, Easley, López de Prado and O'Hara, henceforth ELO, (2011a, 2011b, 2011c) develop the "Volume-Synchronized Probability of INformed trading" (VPIN) metric as a proxy for the imbalance or "toxicity" of order flow. ELO (2011a) focus on the events surrounding the "flash crash" on May 6, 2010. An important feature of their definition of VPIN is that it captures the market dynamics in event time, as measured in equal increments of trading volume, rather than in calendar time. Hence, their analysis relies on a transformed time scale where the basic unit is a fixed volume bucket rather than a constant stretch of calendar time. They find that their implementation of VPIN provides a couple 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.

First, they note that the VPIN measure was unusually high in the week preceding May 6, 2010, and the situation worsened in the hours just prior to the crash. In fact, they observe that the VPIN metric for the E-mini S&P 500 futures contract reached an all-time historical high by 1:30pm CT, and the crash began at 1:32pm CT according to the CFTC-SEC Report time line, detailed in CFTC-SEC (2010). Second, they show that the VPIN measure leads the celebrated Volatility Index, or VIX, for the S&P 500 index, both prior, during and following the dramatic events of May 6, 2010. As such, they suggest VPIN may provide 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 are quite extraordinary. As they note, they raise the possibility that 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. In order to make the measure available to the broader public in a timely fashion they also suggest the introduction of an exchange traded futures contract written on VPIN.[1]

In this paper, we take an in-depth look at the empirical performance of this VPIN metric, applied to the E-Mini S&P 500 futures contract. We reach conclusions that, on key points, diverge from those of ELO. One, VPIN is not a useful predictor for future return volatility. Traditional forecast variables, including the VIX index, are generally vastly superior to VPIN, even for very short horizons. Two, VPIN did not attain a historical high *prior* to the flash crash, but only after the crash subsided. Moreover, reconstructing the real-time information available just prior to the crash, we find little evidence that VPIN would have alerted an observer of a sharply rising probability of impending market collapse. Three, VPIN is, by construction, mechanically related to the underlying trading intensity and its predictive content is largely subsumed by that of the trading pattern. Four, 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 trading scale governed by a volume clock, VPIN is highly correlated with trading intensity. This occurs because of the critical role of the (one-minute calendar) time bar in aggregating individual transactions into volume blocks that, effectively, are treated as single transactions. In particular, all transactions within a time bar are jointly classified as all active buys or all active sells. When trading is active, each time bar contains a lot of volume, the number of time bars used for constructing an update to VPIN becomes small, and the fraction of trades classified jointly as unidirectional rises. This, in turn, increases the

[1] The algorithm for VPIN, detailed in ELO (2011c), has been submitted to the U.S. Patent and Trademark Office, and a patent is pending. For detailed prescriptions on potential contract design, see ELO (2011b).

Electronic copy available at: http://ssrn.com/abstract=1881731

expected value of the order imbalance measures, independently of the order flow imbalance. Quantifying this effect, we find that most of the time variation in VPIN is explained by the heterogeneity in the trading pattern, while trade classification plays a minimal role.

Of course, the trading pattern is endogenous and may respond to, and to a large extent reflect, the underlying order flow imbalances so the above does not imply that VPIN is uninformative. However, trading intensity is well known to correlate strongly with, and contain predictive content for, a variety of market activity variables, including return volatility. Hence, to assess the incremental contribution of VPIN, it is critical to disentangle the information conveyed by the VPIN metric from that associated with the trading pattern. Towards this end, we introduce a few novel 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.

Relying on fixed volume bin VPIN measures, we find only a negligible link with trading volume but, more strikingly, a pronounced *negative* association with return volatility. Thus, once we annihilate the source of mechanical correlation between the trading intensity and VPIN, the relationship between VPIN and volatility reverses sign. Likewise, using an alternative approach to control for the impact of the trading pattern, the contribution of VPIN to short term volatility prediction, over and beyond the component of VPIN explained by the observed trading pattern, is negative. Moreover, we obtain similar conclusions from a VPIN measure computed from (non-aggregated) transaction data. Finally, on the day of the flash crash, both the transaction and fixed volume bin VPIN measures rise prior to the crash, confirming that order flow imbalances likely were operative on this day, but the measures never reach truly extreme values neither before nor after the crash. Hence, overall, our study raises questions about the reliability of this VPIN metric for assessing order flow toxicity.

Finally, we experiment with the computation of a *signed* VPIN measure, allowing signed order imbalance measures to offset in the computation of VPIN. This approach seems to drastically reduce the idiosyncratic noise and appears helpful in detecting the direction and momentum in the order flow imbalance around the flash crash. Thus, it may be fruitful to explore modifications of the VPIN measures along such lines. Nonetheless, we caution against the adoption and reliance on any specific metric before it has been carefully scrutinized and shown to be robust to the type of analysis and tests that we undertake in the present paper.

Our main results are summarized throughout the paper in brief paragraphs labeled "findings." The reader wanting a quick overview may skim the paper and focus on the text surrounding any "finding" of particular interest for supporting evidence.

The remainder of the paper is structured as follows. Section 2 verifies that we obtain results comparable to ELO (2011a) when exploiting their VPIN metric. Section 3 introduces alternative ways of constructing VPIN. Section 4 explores the properties of VPIN and notices the potential for a mechanical correlation with trading intensity. Section 5 presents empirical results based on the full sample. Section 6 revisits the flash crash and reviews the evidence through the lens of alternative VPIN measures. Section 7 provides some concluding general reflections on VPIN, including issue of reproducing the measure, the use of a signed version of VPIN, and reasons behind the surprising negative correlation between fixed bin VPIN and volatility. Section 8 concludes.

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 period January, 2008 through July 2010. 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 although it does not include the last few months of their sample. It is important that the underlying transaction data are comprehensive so 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).[2] Moreover, our series seems to contain a slightly larger trading volume than what is employed by ELO who rely on a different source for their data. Hence, our transaction series appears to provide a comprehensive account of the trading activity in the E-mini S&P 500 futures contract.

2.2 One-Minute VPIN and the Flash Crash

Before proceeding with our more detailed empirical analysis, we confirm that our data series is compatible with the one used by ELO and, in particular, that we are able to obtain similar evidence regarding the behavior of VPIN on May 6, 2010. Hence, we construct VPIN following the prescriptions in ELO (2011c). We obtain comparable data by aggregating our transaction series into one-minute observations, or "time bars," containing the last recorded price and the cumulative trading volume over the minute. This series represents our version of the one-minute bars used by ELO. Since the two series are not identical it is important to check whether they produce qualitatively similar results.[3]

Figure 1 depicts the daily maximum values achieved across our sample for the S&P 500 futures price, the one-minute based VPIN measure and the VIX. It is evident that 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 immediately thereafter. This development was accompanied by an escalation in trading activity and a quick run-up in the VIX measure. Finally, we confirm that VPIN had been rising steadily throughout the day, increasing from around 0.30 in early trading to about 0.56 just prior to the crash – quite closely mimicking the values of ELO (2011a), Exhibit 5. Moreover, the ascent of the VPIN measure was particularly marked in the last 25 minutes before the crash. Comparing Figure 2 to the various 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.

[2] 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.

[3] We do not provide details about the construction of VPIN here, as it is covered at length below.

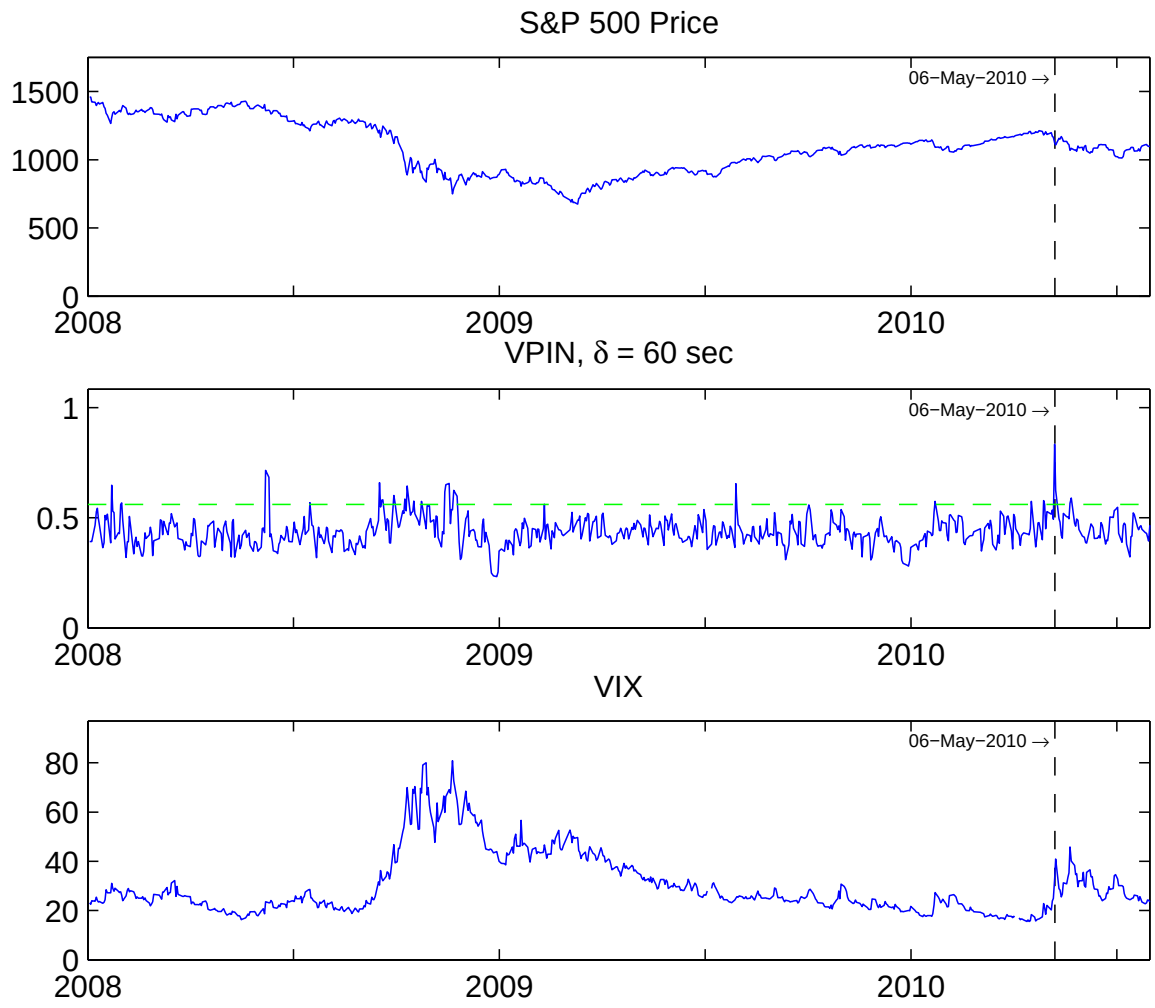

Figure 1: This figure depicts daily values of the S&P 500 index, VPIN for $\delta = 1$ min, and the VIX from January 2008 through June 2010.

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 VPIN just prior to the flash crash is elevated, but it is not at a historical high. The VPIN metric only achieves a historical high after the crash has subsided.

This result is corroborated by the middle panel of Figure 1 where the dashed horizontal line indicates the level 0.56 – the value VPIN attains just prior to the flash crash. The VPIN series exceeds this level on 26 separate days, prior to May 6, 2010, during our sample, constituting more than 4% of the days prior to the crash. In other words, according to our, admittedly short, historical series, one would expect to observe a value, corresponding to the level attained by

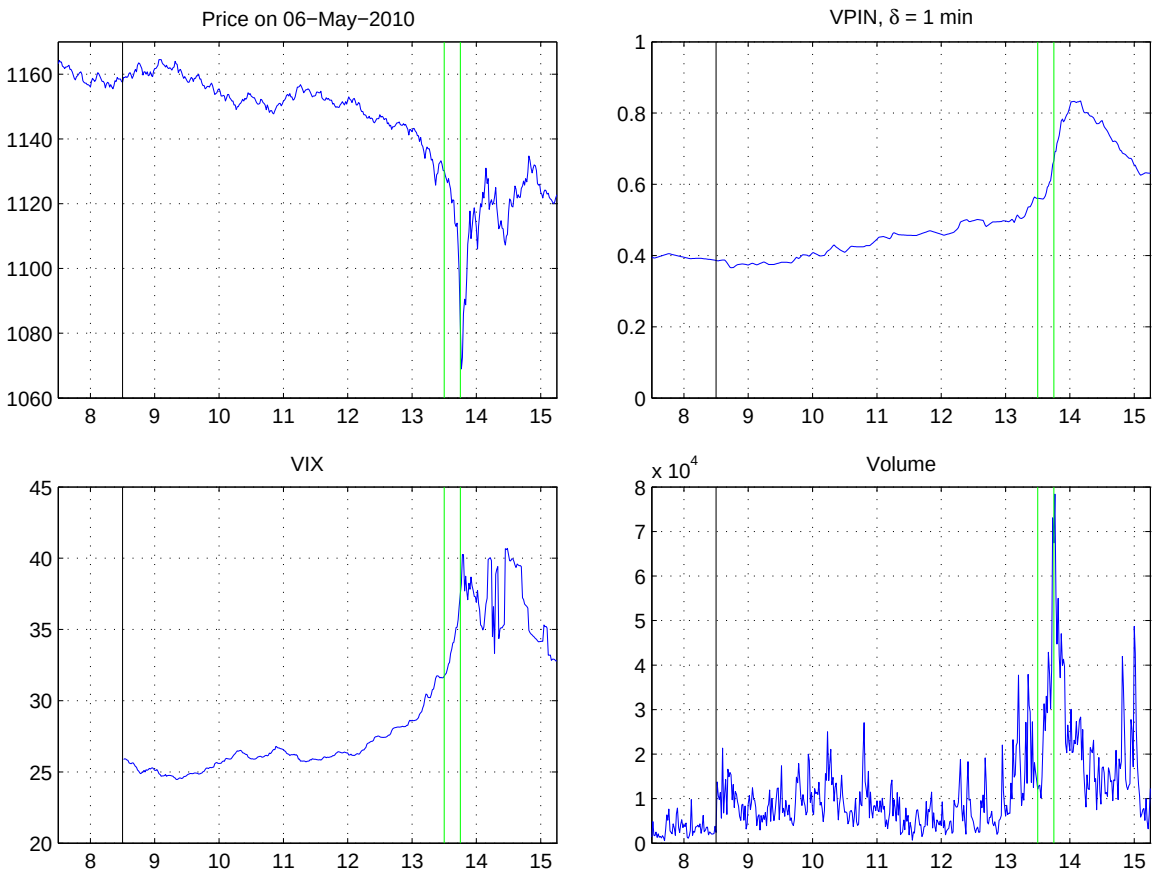

Figure 2: This figure depicts minute-by-minute data for the S&P 500 futures index level, the VPIN measure constructed from one-minute data, the S&P 500 volatility index, VIX, and the volume of traded contracts of the S&P 500 E-mini futures on the CME, for May 6, 2010. The green vertical lines indicate the timing of the "flash crash."

VPIN prior to the flash crash, about once a month.[4] In light of this observation, the behavior of VPIN prior to the crash, arguably, remains noteworthy but not truly exceptional. We now turn towards a more detailed investigation of potential sources of predictive power in VPIN. However, first we need to explain how the VPIN metric is computed.

3 Constructing the 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 given sample period that we denote $[0, T]$. Each transaction, or tick, is represented by 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

[4]This also appears consistent with Exhibit 2 in ELO (2011a), where the VPIN metric exceeds 0.60 on many distinct days before May 6, 2010. Hence, the statement in ELO (2011a) on this point may reflect incorrectly attributing the maximum value of VPIN, attained **after** the crash, to the value prior to the crash.

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 at 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 forcefully 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 shall also explore a procedure that combines the underlying transactions into blocks with an equal trading volume, i.e., the same number of traded contracts. Consequently, we adapt the notation to accommodate such different forms 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 in detail. 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 jth 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 jth bar.

Following ELO, we sign the entire order flow within each time bar using a classification scheme, akin to the so-called tick rule. Let the binary variable $b_j = \pm 1$ indicate whether the contracts exchanged within the jth 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 clearly will misclassify 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 may write $(T_j^\delta, P_j^\delta, N_j^\delta)$ and b_j^δ. Furthermore, we may think of the use of transaction data as a limiting case where we perform no time aggregation so that, with slight abuse of notation, we have $\delta = 0$.

Second, we define the volume bins. We let $0 = N_0 < N_1 < N_2 \ldots$ represent an *equally*-spaced volume 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 may now be represented by triplets (T_k, P_k, N_k), where T_k and where P_k are, respectively, the calendar time and trade price associated with the last contract included in volume bin k for $k = 1, \ldots, K^\nu$, where K^ν is the number of complete non-overlapping volume bins of size ν in the sample. Similarly to the classification rule (1), the binary variable b_k classifies the entire order flow within the kth volume bin, for $k = 2, \ldots, K^\nu$. If we need to be explicit about the choice of volume grid, we write $(T_k^\nu, P_k^\nu, N_k^\nu)$ and b_k^ν.

In summary, the trade classification rule (1) assigns an active buy or sell indicator to each transaction throughout the sample, but they are bundled into sequences of unidirectional buys

and sells according to the length of the time bar or volume bin they reside in. Whether this rule dominates alternative classification schemes in terms of providing a useful input 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 when presenting our findings.

3.2 The Volume Clock, the OI Measure and VPIN

Following ELO, we now introduce a volume-based time-scale transformation. Thus, rather than monitor the market dynamics in calendar time, we employ a fixed volume bucket, V, expressed in terms of the number of traded contracts. Hence, each time increment in the subsequent analysis represents equidistant trading volumes, but potentially highly varying periods of calendar time. Throughout our analysis, we set $V = 35{,}000$ traded futures contracts, corresponding roughly to $(1/50)^{\text{th}}$ of the average daily trading volume. This choice mimics the leading case adopted by ELO. It is well documented that such "tick time" sampling alleviates the non-normality of returns stemming from time-varying volatility because volatility is highly, albeit imperfectly, correlated with trading volume. Consequently, this "event time" transformation reduces the heterogeneity of the return distribution and helps render the information content of each incremental observation comparable.

Each volume bucket comprises a set of aggregate transaction triplets, each representing a different block of trading volume. For underlying triplets defined from time bars, we will inevitably need to sever those time bar, which comprise transactions straddling adjacent volume buckets, into separate pieces so that each new fraction of the time bar falls squarely within only one bucket. We generically denote the number of transaction blocks within a volume bucket Q, and we define the relative size of each trade block within 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, $0 < w_q \leq 1$ 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,

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

Motivated by the market microstructure theory, detailed in ELO (2011c), the focus of ELO (2011a) is on the *absolute* order flow imbalance relative to the total volume for the given bucket. Hence, 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 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,

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

This VPIN metric constitutes the informed trading, or "toxicity", measure constructed according to the ELO procedure. It is obtained for a given set of observed transactions over a specific sample period and reflects underlying choices of the volume bucket, V, the time bar, δ, the trade classification indicator b, and the moving average parameter, L. These parameters interact in a complex manner to determine both the level and the dynamic behavior of the metric. The evolution of VPIN on the day of the flash crash must be interpreted with these features in mind. The following section provides a more detailed analysis of the basic properties of VPIN.

4 Basic Properties of VPIN

ELO (2011a) emphasize that trade time is the relevant metric to use in sampling the information set as it is volume ranges that produce comparable amounts of information. This is illustrated in ELO (2011c) by showing that return volatility across equal volume buckets is much more homogenous than across equal calendar time intervals. Hence, we maintain a fixed volume bucket, V, in computing VPIN throughout. In addition, recognizing that order imbalance measures over short intervals may be somewhat noisy, the moving average parameter L serves to smooth the series and thus enhance the signal content relative to transitory noise components. We also abstain from any inquiry into this particular aspect of the VPIN definition. Instead, we focus squarely on the critical building block for VPIN, namely the behavior of the OI measure for a given volume bucket of size, V.

Since all transactions within the sample are ordered, the set of trades falling within a given volume bucket is fixed. The only variation in the OI measure stems from how we assign the buy-sell indicator to the individual trades within the bucket. As explained in the preceding section, ELO (2011a) use one-minute time bars for this purpose, thus computing the measure as if all trades within the same time bar operate on the identical side of the market. In other words, depending on the classification rule in equation (1), all trades within a given time bar are jointly classified as either buyer- or seller-initiated.[5]

The adoption of a calendar time interval as a critical part of the trade classification device within a volume clock scheme is unconventional. One important implication of this design feature of VPIN is the introduction of an extreme degree of heterogeneity in the distribution of the order imbalance measure as a function of the underlying trading intensity. This arises from the fact that an increase in trading volume implies there are more trades per time bar and thus a smaller number of time bars, Q, involved in the computation of the OI measure within a volume bucket. This systematically increases the expected value of the OI measure, independent of any other 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 any actual trading imbalances. This is in sharp contrast to the outcome of tick or fixed volume bin sampling which serves to generate homogeneous sampling distributions for the quantities of interest. We first illustrate the implications of this point within a simple example.

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

4.1 A Simple Illustration

For simplicity, imagine that we have a volume bucket size of $V = 1{,}000$. The last recorded price prior to 9:37 is 10.01, and the last observed trade indicator is a buy. For a moderately busy trading period starting at 9:38, the volume bucket may contain six separate one-minute time bars. Thus, for $\delta = 60$ seconds, we assume that the triplets $(T_q^\delta, P_q^\delta, n_q^\delta) = (T_q, P_q, n_q)$ take the following form,

(I) $\{(T_q, P_q, n_q)\}_{q=1\ldots6} \;=$

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

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 $OI = |\,500 - 500\,|\,/\,1,000 \;=\; 0.$

Increasing the trading intensity to four times the original level, the first four original trade blocks are combined into the first time bar. It will contain 800 traded contracts. The next 200 contracts will belong to a second time bar. Assuming the last trade price within this time bar remains at 10.00 or below, we have the following situation,

(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 is now assumed to reach fivefold the original level, the first five original trade blocks form the first new time bar, containing 900 traded 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 within the volume bucket,

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

The order imbalance measure now becomes $OI = |-1,000|\,/\,1,000 \;=\; |-1.0| \;=\; 1$. This trading intensity may appear excessive, but it is typical of truly chaotic market conditions, including the period around the flash crash when trading was extremely elevated for long stretches of time. The result is the mechanical production of a sequence of OI measures that take the value of unity. The only moderating effect during such periods is when the time bars are split across two adjacent volume buckets and the OI measure becomes a volume-weighted average of the trade direction indicators for the two bars.

Hence, the SOI measure varies from -1 to 0.8 across the scenarios, while OI fluctuates from 0 to 1. This makes it clear that the order imbalance measure associated with a given volume bucket can be extraordinarily noisy – for a given set of underlying transactions, OI may take on any possible value between 0 and 1, depending on the trade intensity and the way in which the boundary of the volume buckets interact with the time bars. Moreover, it also shows how the OI measure tends to inflate in concert with the trading intensity. As the speed of trading grows, the number of time bars within the volume bucket declines and there is less room for diversification of the buy and sell indicators. In the limit, the measure becomes identically unity, irrespective of both the actual price path and the proportion of active buy and sell transactions within the volume bucket. The OI measure degenerates into a pure trading intensity measure.

What would happen if we instead were using a FB-VPIN methodology, relying on a bin size of $\nu = 200$? Unfortunately, this requires a more detailed knowledge of the actual underlying transactions. However, assuming simply that the original sequence of aggregate trades within the minute bars in scenario (I) constitutes the actual trades, and each occur within the last second of the time bar, we obtain the following scenario,[6]

$$\textbf{(FB)} \quad \{(T_q^{\nu}, P_q^{\nu}, n_q^{\nu})\}_{q=1\ldots5} \quad =$$

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

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

The point of the fixed volume bin approach is that this SOI statistic depends only on the sequence of trades and thus will remain at -0.2, irrespective of the intensity of trading. Thus, it is arguably a more suitable measure for capturing the volume-weighted order imbalances in pure trading time.

[6] If there were several trades within each minute, as there almost inevitably would be, we would avoid the rough splitting of individual large transactions into adjacent volume bins. This 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.

Of course, our illustration may be overly simplistic and may not capture the systematic effects across the sample appropriately. In particular, it is likely that the trading pattern will change in systematic ways as volatility and uncertainty increases and this may drive VPIN in particular directions that distinguish these episodes from the VPIN measures observed during more benign trading conditions. In other words, the variation in trading intensity may be informative in and of itself, and VPIN may extract additional information about the state of the market from such patterns. This could also potentially help render the time bar OI more compatible with the corresponding FB-OI measures, avoiding any systematic impact of the awkward idiosyncratic variation in time bar SOI observed in this example. We turn to a more formal analysis to address such questions.

4.2 Systematic Variation in the Order Imbalance and VPIN Measure

To assess the significance of the temporal variation in the OI measure, it is useful to benchmark it against a simple reference measure that is void of explicit market microstructure features indicating the presence of trade imbalances. One standard approach in asset pricing is to assume the price process constitutes a semi-martingale with respect to the natural filtration generated by the past history of prices and trades. This assumption is necessary to ensure the absence of arbitrage within a frictionless market. It implies that any predictability, or drift, in the price over short horizons is trivial and may be ignored relative to the size of the (unpredictable) return innovations. At the same time, this setting is consistent with a large degree of predictability in trading volume and return volatility. In fact, it is often alleged that volatility and volume both tend to respond to a positively serially correlated information flow series that is latent or unobserved. Consequently, a strong positive correlation is expected between return volatility and trading volume.[7] Moreover, since volatility and volume both are highly persistent series, either may be readily predicted by its own lagged values. The illustration above suggest that OI mechanically will be strongly correlated with volume, so it is useful to disentangle the various factors governing the dynamics of the OI measure and, ultimately, the VPIN metric to assess their relative information content.

4.2.1 Systematic Variation in VPIN due to Trading Intensity

We first explore the pure impact of time variation in trading intensity on the OI measure. For this purpose, assume counterfactually that the trade classification scheme produces an i.i.d. sequence of buy-sell indicators, $\{b_q\}$, taking on the values of +1 and -1 each with probability 1/2. In this setting, the expected value of any future trade indicator is zero, i.e., $E[b_q] = 0$. Moreover, to annihilate any impact from heterogeneity in order size, let each time bar represent an equal amount of trading volume so that $w_q = 1/Q$. The only variable impacting the value of OI is now 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 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

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

[7] This account is closely related to the so-called "mixture-of-distributions hypothesis" which was explored by, e.g., Clark (1973), Epps and Epps (1976), Tauchen and Pitts (1983), and Andersen (1996).

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

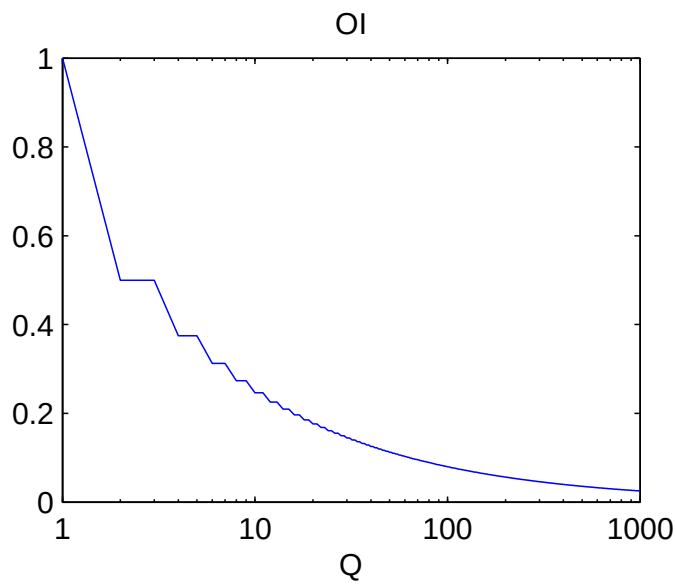

Figure 3: The figure plots the expected order imbalance as a function of the number of components, Q, used in the computation. The expectation is computed under the assumption of purely random and uninformed order flow and homogeneous trade sizes.

Intuitively, the order imbalance is non-increasing in Q, starting at unity for $Q = 1$. Secondly, we observe that $F(2q) = F(2q+1)$, for example, $F(2) = F(3) = 0.5$, $F(4) = F(5) = 0.375$, and even $F(10) = F(11) = 0.246$. Thirdly, 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 a *very* conservative benchmark for the observed OI measures. In reality, the volume contained within individual time bars vary in size, and $F(Q)$ 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, which is much higher than $F(2) = 0.5$, even in the absence of any systematic order imbalance. In addition, for a variety of reasons, the trade classification sequence, $\{b_q\}$, will not be independent which also tends to increase the value of OI.

In summary, the values indicated in Figure 3 provide a conservative lower bound for the expected OI measure for a given value of Q. The result has important implications.

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 VPIN, mechanically, tend to rise, because Q declines.

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

We hasten to caution against the conclusion that any observed time variation in trading intensity is unrelated to order imbalances. It is possible, indeed likely, that a significant proportion of the variation in trading intensity is related to the general market environment. The relevant question is then whether the trade classification scheme provides additional information beyond what we can infer from the observable trading pattern. We explore such issues in the empirical section below.

FINDING 3

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

This last result is no surprise. ELO (2011c) discuss related issues and emphasize the need to compare only the relative size of OI and VPIN over time, and not the actual levels. The finding does, however, have the practical implication that the design of the OI and VPIN variables must be calibrated to the existing market conditions to allow for comparisons across time or markets. In particular, if there is a trend in the liquidity, or average trading intensity, of the market, then the "effective" Q has shifted, and a corresponding change in the size of the time bar is required to render the VPIN series stationary over the full period. That is, the time bars must vary in size to accommodate systematic shifts in overall liquidity across time. For example, a positive trend in trading volume should be accompanied by a shrinkage of the time bar. Otherwise, the drift in trading volume would mechanically be interpreted as an increase in the average order imbalance and seen as signaling a more turbulent market setting. One response is to only compute or compare VPIN measures over relatively short sample periods where the overall trading environment is deemed sufficiently stable.

In order to contrast the $F(Q)$ measure to VPIN, it must be smoothed in a comparable manner. Hence, as for the 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,

$$\text{U-VPIN1}_t^\delta \;=\; \frac{1}{L}\sum_{\ell=1}^{L} F(Q_\ell)\,. \tag{7}$$

This measure is constructed under the assumption of a purely random order flow and a homogeneous volume distribution. But these assumptions are likely to be violated and either feature will inflate the measure. Hence, it is an extremely conservative measure that will tend to produce values much smaller than the realized VPIN values. The label U-VPIN1 refers to the fact that it is our first measure computed along the lines of VPIN, and it is "uninformed" about the underlying price path and order size distribution. As such, we intend to use it for gauging, qualitatively, the type of variation in the VPIN measure we may expect to experience due merely to changes in the underlying trading intensity.

4.2.2 Systematic Variation in VPIN Attributable to the Trading Pattern

The $F(Q)$ and VPIN measures capture only the impact of the number of elements used for computing OI. Any variation in volume across the time bars will increase the measure. Since the actual trading process involves transactions of highly varying size, we now develop an alternative benchmark that reflects the impact of volume heterogeneity on the expected OI measure. Unfortunately, direct analytic expressions are no longer feasible. However, observe that the expression for the expected OI measure is,

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

This is an L^1 norm applied to the signed order imbalance measure. It turns out that the corresponding L^2 norm is highly tractable under the maintained assumption that the $\{b_q\}$ sequence is i.i.d. and symmetric. That is, we can instead exploit the measure,

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

This measure is tantalizingly simple. Given the observed volume weight vector, $w = (w_1, \ldots, w_Q)'$, for a given volume bucket, the expected order imbalance measure is simply given by 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 observed trading pattern on the expected OI measure, provided the trade indicator follows an i.i.d. process. However, we expect positive serial correlation in the trade indicator sequence, implying that the OI measure may be smaller or larger than the $|\,w\,|$ measure, depending on the strength of the Jensen inequality bias versus the serial correlation in $\{b_q\}$.

In summary, the time variation in this L^2 norm should provide a gauge on the variation in the order imbalance measure 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 VPIN style metric by computing a backward looking moving average. Exploiting the identical timing notation as for VPIN, and letting w_ℓ denote the volume weight vector for the bucket that terminates at time τ_ℓ, we define,

$$\text{U-VPIN2}_t^\delta \;=\; \frac{1}{L}\sum_{\ell=1}^{L}\,|\,w_\ell\,|\,. \tag{10}$$

While U-VPIN1 is based on an L^1 norm for the expected order imbalance measure, U-VPIN2 is based on a related L^2 norm. Moreover, like U-VPIN1, U-VPIN2 is "uninformed" about the actual price path, and thus the trade indicator sequence, utilized by VPIN.

For a given distribution of Q_i across buckets, U-VPIN2 is minimized when the trading volume associated with each time bar within each bucket is identical, while it is maximized when the volume within one time bar dominates the volume associated with the other time

bars, i.e., if one volume weight is near unity and the remainder are tiny. Hence, this measure captures variation in the heterogeneity of the observations within the volume buckets. At the same time, U-VPIN2 is constructed assuming a purely random trade direction, so its variation cannot be directly attributed to order flow imbalances. Again, we caution that the trading pattern, surely, is not independent of market conditions, and thus U-VPIN2 is not unrelated to tensions in the market. It rather reflects the component of VPIN that is directly related to the observed trading pattern.

FINDING 4

All else equal, the expected value of the OI and VPIN measures is increasing in the degree of volume heterogeneity across time bars.

Finding 4 implies that any systematic relationship between volume heterogeneity across time bars and the state of the market will create corresponding correlations between VPIN and overall market conditions. We document below that this feature is empirically relevant for VPIN and helps explain its otherwise seemingly anomalous behavior as the time bar shrinks.

4.2.3 Removing the Impact of Trading Intensity via Fixed Volume Bin VPIN

The ELO VPIN metric is tied to the measurement of volume-weighted order imbalances across time bars. As noted in Finding 2, this renders the measure highly sensitive to the trading intensity. One way to avoid this mechanical dependence on trading intensity is to compute the OI measure from equally-sized volume bins instead. In fact, this approach may be seen as the natural response to the observation that activity within a high-frequency market 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 $(T_k, P_k, N_k) = (T_k^\nu, P_k^\nu, N_k^\nu)$, 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 traded by the end of each volume bin is $N_k = \nu \cdot k$. Hence, each volume bin contains a fixed number, $N_k - N_{k-1} = \nu$, of traded contracts, for $k = 1, \ldots, K^\nu$. In our empirical work, we choose $\nu = 1{,}000$ or $\nu = 5{,}000$. These values imply that each volume bucket with $V = 35{,}000$ contains either 35 or 7 separate volume bins. These figures span the average number of time bars included in the ELO OI measure across the typical trading day while, by construction, eliminating the heterogeneity in volume across the bins used for computing OI.

Each volume bucket now comprises a fixed set of $Q = V/\nu$ bins, each containing an equal fraction, $w_\nu = \nu/V = 1/Q$, of the overall trading volume within the bucket. We apply the ELO trade classification scheme (1) to each bin. Hence, we may define the "fixed bin" *signed* order imbalance measure, SOI^ν, for a given volume bucket as follows,

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

As previously, 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. Moreover, we let SOI_ℓ^ν denote the fixed bin signed order imbalance measure for the bucket terminating at time τ_ℓ. We then define

the "Fixed-Bin" VPIN, or FB-VPIN, as,

$$FB\text{-}VPIN_t^{\nu} \;=\; \frac{1}{L}\sum_{\ell=1}^{L} \;|SOI_{\ell}^{\nu}|\;. \tag{12}$$

This FB-VPIN measure retains the dependence on the trade classification scheme, but it breaks the mechanical relationship with trading intensity. As such, this measure will be useful in identifying the variation in the ELO VPIN metric that stems from order imbalances measures based on a genuine trading time scale.[8]

5 Empirical Results for the Full Sample

5.1 Summary Statistics for the Trading in E-mini S&P 500 Futures

We first present the general features of the trading volume series for the E-mini S&P 500 futures across our sample period, with a view towards the critical aspects discussed in the Section 4. Table 1 provides descriptive summary statistics.

It is evident that the E-mini S&P 500 futures market generally is extremely liquid. There were more than 185,000 trades per day on average, and the daily volume averaged in excess of 2,166,000 contracts. That 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 the regular trading hours, the activity is still impressive, with a trade being consummated about once every two seconds.

Our choice of $L = 50$ and $V = 35,000$ contracts for computing VPIN must be assessed in view of these numbers. As indicated, they were calibrated to ensure a close approximation to the setting of ELO. In particular, V represents almost $(1/50)^{\text{th}}$ of the average daily volume, so the amount of trades used to compute VPIN corresponds to about one typical trading day. This makes it comparable to the corresponding choice in ELO although it is infeasible to obtain a perfect match as their data set covers a slightly lower number of transactions.

Another critical dimension is the number of transaction per time bar, as this is a major factor both in determining the reliability of the trade classification scheme and in determining the average number of blocks used in computing the OI measure. First, we note that a time bar of one minute will, on average, encompass 390 separate transactions during the 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 series of active buys and sells often alternate in rapid succession.[9] Nonetheless, it is an empirical

[8] 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, U-FB-VPIN1 and U-FB-VPIN2, will be constants, because each volume bucket now consists of the same number (Q) of equal size bins, implying,

$$U\text{-}FB\text{-}VPIN1_t^{\nu} \equiv F(Q), \qquad \text{and} \qquad U\text{-}FB\text{-}VPIN2_t^{\nu} \equiv \frac{1}{\sqrt{Q}},$$

[9] The first order tick return autocorrelation across the sample is around -0.41, while the next four tick return 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

Table 1: **Descriptive Trading Statistics for the E-mini S&P 500 Futures Contract**

	All	RTH	OTH	HTH
Volume (1 day), 000s	2166.53	1952.20	261.00	137.94
# Trades (1 day), 000s	185.45	156.42	32.83	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
Order Size	11.68	12.48	7.95	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: This table reports summary statistics for the trading in E-mini S&P 500 futures contract over the period 01/2008-07/2010. The data are reported separately for Regular Trading Hours (RTH, 8:30-15:15 CT), Overnight Trading Hours (OTH, 15:15-8:30 CT), Holiday Trading Hours (HTH, exchange holidays), and combined hours (All).

question whether the trade classification scheme provides useful insights into the effective order flow imbalance in the market, and we explore this issue further below. However, it also motivates us to analyze the behavior of the VPIN metric across alternative choices of time bars for which the degree of trade misclassification will vary. Second, there will be an average of 7-8 time bars within each volume bucket during regular trading hours. Moreover, depending on the trading intensity of any given trading period, this number fluctuates anywhere from 25 down to 1. Inspecting Figure 3, it is evident that the expected OI measure is highly sensitive to the number of time bars in the volume bucket, Q, as reflected in the steepness of the curve in the relevant range. In other words, VPIN will, by construction, vary dramatically over time in response to persistent variation in the underlying trading volume, independently of whether the transactions are evenly balanced across buys and sells or not.

Finally, we note the extreme right-skewed order size distribution. In a typical time bar containing 390 transactions, about 200 will involve the exchange of only one or two futures contracts. However, about 40 will involve more than 20 contracts each, and a few will entail active trading of hundreds of contracts in a single transaction. While this order size heterogeneity is not likely to have a strong bearing on the behavior of the OI measures obtained from one-minute bars, they will impact measures computed from a smaller number of transactions, e.g., from transaction data, or from smaller time bars or volume bins.

deeper than the volume associated with individual transactions. Consequently, the best bid and ask quotes are generally quite stable relative to the oscillation of transactions between the bid and ask. Hence, we expect that sequences of hundreds of transactions typically involve a large number of both buys and sells.

5.2 Comparing Alternative VPIN Measures

We have argued that VPIN potentially is subject to a multitude of distortions stemming from the variation in trading intensity, inhomogeneity of the volume weights across time bars, and random variation in the trade classification scheme. One way to assess the significance of these factors is to compare the behavior of the VPIN metric computed over alternative time bars and contrast them with our U-VPIN1, U-VPIN2, and FB-VPIN measures, as each of these convey information about the strength of the underlying forces impacting VPIN.

Figure 4 displays VPIN across the full sample computed using time bars of $\delta = 0$, i.e., tick data, $\delta = 10$, and $\delta = 60$, with each series plotted against the corresponding U-VPIN1 and U-VPIN2. The latter series represent the expected values of VPIN, assuming random trade classification and computed conditional on specific observed patterns of trading. Finally, the bottom panel provides the FB-VPIN series, computed for fixed volume bins of $\nu = 5,000$ and $\nu = 1,000$. By computing the metric in pure trade time, the FB-VPIN series eliminate any mechanical time series correlation between trading intensity and VPIN.

Figure 4 highlights a number of interesting features. First, as expected, the VPIN measure increases very significantly as the time bar grows longer. Likewise, the FB-VPIN values rise as the volume bins grow lengthier. This simply reflects the bundling of ever larger sets of transactions into volume blocks which each are classified as consisting of unidirectional trades. As a result, the effective Q decreases and the expected VPIN grows. Second, it is evident that VPIN is strongly correlated with U-VPIN1 and U-VPIN2, especially for the time-aggregated series with $\delta = 10$ and $\delta = 60$. In fact, for $\delta = 60$, U-VPIN2 almost coincides with VPIN, with both series having very similar trends and spikes. Moreover, U-VPIN1 displays the same qualitative time series variation even if, by construction, it attains somewhat lower values. In contrast, the transactions ($\delta = 0$) and FB-VPIN series appear much less consistent with VPIN computed from time bars. Third, the upward spikes in the 10-second and 60-second (U-)VPIN series on May 6, 2010, are striking. However, neither the transaction nor the FB-VPIN series attain an unusually high value on May 6, 2010. The discrepancy between the time bar and volume bin based VPIN series is remarkable, given that the volume bins of $\nu = 1,000$ and $\nu = 5,000$ compute the underlying OI measures from a set of volume blocks that mirror the number of blocks utilized, on average, by VPIN computed with $\delta = 10$ and $\delta = 60$, respectively.

FINDING 5

The ELO VPIN is highly correlated with U-VPIN1 and U-VPIN2 and the correlation increases with the length of the time bar. Hence, the trade classification rule is close to negligible in determining the time series variation of VPIN.

FINDING 6

VPIN computed from time bars behave dramatically differently from VPIN computed from fixed volume bins or transaction data, even if the identical volume buckets are used for computations. In particular, only time bar VPIN attains an exceptional value on May 6, 2010.

Complementary evidence is provided by Table 2. It reports the sample correlations between the VPIN measures for different time bars along with correlations with the daily trading volume and VIX. In addition, a separate correlation table for FB-VPIN, volume and VIX is included. We have already noted the tendency for large time bars to induce a mechanical correlation with

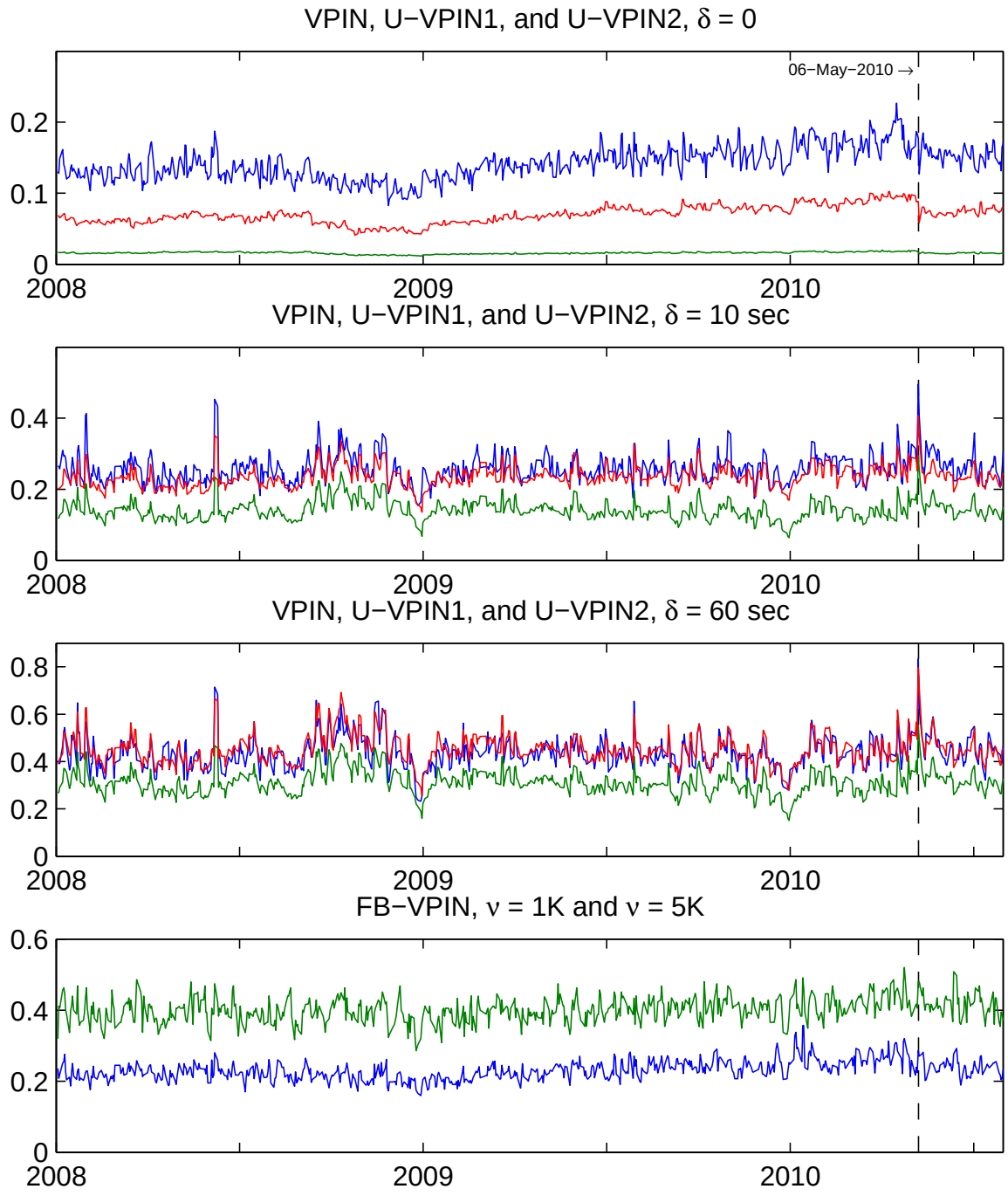

Figure 4: The top three panels of the figure plots the daily maximum values of VPIN (blue), U-VPIN2 (L^2 metric, red), and U-VPIN1 (L^1 metric, green) for $\delta = 0$, and $\delta = 1$ min. The bottom panel depicts the FB-VPIN for $\nu = 1,000$ (blue) and $\nu = 5,000$ (green)

trading volume. The VIX index is incorporated as a reference for the subsequent discussion of VPIN as a predictor for future short term return volatility.

Table 2 confirms the main conclusions from Figure 4 and brings out some new features.

Table 2: **Correlations**

$\delta = 0$

	VPIN	U-VPIN1	U-VPIN2
VPIN	1.00		
U-VPIN1	0.66	1.00	
U-VPIN2	0.80	0.85	1.00
Volume	-0.23	-0.37	-0.41
VIX	-0.57	-0.87	-0.75

$\delta = 10$

	VPIN	U-VPIN1	U-VPIN2
VPIN	1.00		
U-VPIN1	0.70	1.00	
U-VPIN2	0.76	0.87	1.00
Volume	0.51	0.81	0.59
VIX	0.08	0.45	0.14

$\delta = 60$

	VPIN	U-VPIN1	U-VPIN2
VPIN	1.00		
U-VPIN1	0.78	1.00	
U-VPIN2	0.81	0.98	1.00
Volume	0.60	0.82	0.79
VIX	0.21	0.46	0.38

$\delta = 300$

	VPIN	U-VPIN1	U-VPIN2
VPIN	1.00		
U-VPIN1	0.84	1.00	
U-VPIN2	0.86	0.94	1.00
Volume	0.69	0.81	0.78
VIX	0.37	0.43	0.45

FB-VPIN

	$\nu = 1K$	$\nu = 5K$	Volume	VIX
Volume	-0.09	0.05	1.00	
VIX	-0.42	-0.17	0.52	1.00

Notes: This table reports correlations of VPIN, U-VPIN1, U-VPIN2, 1-day trading volume, and VIX, for different cases of δ =0, 10, 60 and 300 sec and ν = 1K and 5K. The sample period is 01/2008-07/2010.

First, we note the strong and generally increasing correlation between VPIN and the U-VPIN measures as δ grows. It is 0.70-0.75 for $\delta = 10$, about 0.80 for $\delta = 60$, and around 0.85 for $\delta = 300$. Even at the transaction level, this correlation is very robust which is hard to discern from Figure 4 due to the different scaling of the series.

Second, there is a dramatic break in the correlation between volume and VPIN measures as we move from computing VPIN by time bars to ticks or volume bins. For time bar VPIN, the positive correlation is not surprising, even if the magnitude might be. Trading volume is strongly serially correlated, so Findings 2 and 3 suggest that volume will be correlated with time bar VPIN. The main insight is rather that FB-VPIN covaries only weakly with trading volume. The use of volume bins annihilate any impact from the variation in trading intensity on FB-VPIN, so the result suggests that the trade direction indicators are largely unrelated to overall trading volume. That is, to the extent VPIN captures order imbalances, this stems almost exclusively from the impact of the trading pattern, while the trade indicators are largely irrelevant. This is, of course, also consistent with the very high degree of correlation between VPIN and the U-VPIN measures, and serves to corroborate Finding 5.

Third, the time bar VPIN measures display positive correlation with VIX, although it is noticeably weaker than the VPIN-volume correlations. Given the well-known strong positive association between volume and volatility, exemplified by the current sample correlation of 0.52, the correlation between these VPIN measures and VIX is expected. In contrast, FB-VPIN and transaction VPIN are strongly negatively related to VIX. This feature is remarkable given the ELO findings. It suggests that VPIN, obtained at the transaction level or measured consistently in trading time, is *inversely* related to the general long-run volatility level. Moreover, the effect is strongest at the transaction level and slowly dissipates as the transactions are aggregated into larger volume bins. The relationship is highly significant for transaction VPIN, suggesting a interaction between uncertainty or turmoil in the market and the trade classification scheme in equation (1). The slow dissipation of this relation with the aggregation level suggests that the same phenomenon remains at work but that the trade indicators become less reliable as transactions are bundled into larger blocks with unidirectional classifications. The surprising hypothesis arising from these observations is that VPIN measured in trading time is *negatively* related to overall trading volume and longer run return volatility. By the same token, time bar based VPIN measures, in general, display positive correlation with VIX only because they are constructed in a manner that mechanically makes them covary strongly with the concurrent trading intensity.

FINDING 7

Time bar VPIN is strongly correlated with trading volume and weakly correlated with VIX. In contrast, VPIN, computed from fixed volume bins or transaction data, is weakly, and predominantly negatively, correlated with trading volume, and strongly negatively correlated with VIX. Hence, the correlation between ELO VPIN and general market volatility largely reflects the mechanical association between time bar VPIN and trading intensity.

In summary, when VPIN is computed in trading time it displays an entirely different dynamic behavior than when extracted using time bars. At a minimum, this raises the question of whether the two variants of VPIN capture fundamentally distinct features of the trading process. To help address this issue, we explore the pronounced negative correlation between transaction and fixed volume bin VPIN vis-a-vis VIX in further detail later.

5.3 VPIN as a Forecast of Future Short-Term Return Volatility

ELO suggest that VPIN may be useful for predicting impending turmoil in financial markets. They provide evidence that 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 uncover a positive, albeit somewhat weak, correlation between the ELO VPIN metric and concurrent VIX. However, as argued 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 VPIN performs as a forecast variable relative to other standard volatility predictors. Moreover, we seeks to shed additional light on the mechanism that generates a correlation between VPIN and future volatility.

Table 3 provides a first overview of the evidence. It tabulates the correlations between alter-

Table 3: **Correlations of Future Absolute Returns with Various Measures**

	VPIN				U-VPIN2					
	$\delta = 0$	$\delta = 10$	$\delta = 60$	$\delta = 300$	$\delta = 0$	$\delta = 10$	$\delta = 60$	$\delta = 300$	Volume	VIX
$\lvert r_{t,t+1} \rvert$	-0.30	0.08	0.14	0.20	-0.38	0.10	0.23	0.25	0.30	0.46
$\lvert r_{t,t+10} \rvert$	-0.33	0.07	0.13	0.22	-0.41	0.10	0.24	0.28	0.32	0.49
$\lvert r_{t,t+50} \rvert$	-0.35	0.01	0.07	0.18	-0.42	0.04	0.18	0.23	0.28	0.47
$\lvert r_{t,t+250} \rvert$	-0.35	0.05	0.16	0.24	-0.44	0.11	0.25	0.27	0.38	0.54

Notes: "Volume" is one-day trading volume. The sample period is 01/2008-07/2010.

native predictor variables and future cumulative absolute returns over four different horizons, ranging from a single to 250 volume buckets, corresponding to an average of a few minutes to five full trading days. The candidate forecast measures include VPIN and U-VPIN2 obtained from different time bars, along with lagged daily trading volume and the VIX index.

The raw correlations indicate that VPIN generally provides a comparatively poor volatility forecast. For $\delta > 0$, the VPIN correlations with future volatility, at all horizons, are uniformly lower than those associated with U-VPIN2, even though the only difference between them is the trade classification scheme. Taken at face value, this implies that the trade classification rule induces variation in the VPIN metric that lowers its correlation with future return volatility, relative to the random classification associated with U-VPIN2. One potential explanation is the fact that lagged daily volume is more highly correlated with the future realized return volatility than VPIN. Since U-VPIN2 is more strongly correlated with volume than VPIN, this suggests that the source of the VPIN-volatility correlation is the volume-volatility correlation. Hence, one may hypothesize that the VPIN measures predict future return volatility solely due to their mechanical correlation with trading volume. We further note that the correlations suggest VIX, generally, is a vastly superior predictor of future realized volatility than any of the other forecast variables. Finally, we observe that the strikingly strong negative association between transaction VPIN and future realized volatility is consistent with the negative contemporaneous correlation between transaction VPIN and VIX, discussed in the preceding section.

In order to more formally assess the hypothesis concerning the general relationship between VPIN and future short-term realized volatility, we now compare the predictive performance within a regression setting where we can control directly for the impact of auxiliary variables. We focus on VPIN measures obtained from time bars with $\delta = 60$, but the qualitative findings are identical for other choices of (strictly positive) time bars. Table 4 summarizes the evidence for a representative subset of the regressions we have explored.

The results are clear cut. From the first row in either panel, we see that there, indeed, is a highly significant relationship between VPIN and future return volatility although the degree of predictive power is limited as reflected in the adjusted R^2 of about 2% and 6%, respectively, for the two forecast horizons. The second row of the panels present the corresponding regression results for U-VPIN2. The improvement in forecast performance is striking, with the explanatory power rising more than twofold. Given the high correlation between the two alternative measures, it supports the hypothesis that the variation in trading activity, as reflected in U-VPIN2, is the underlying source of volatility predictability: the inclusion of order imbalance

Table 4: **Forecast Regressions for Absolute Return**

Panel A: 1-Period Forecast

	Const.	VPIN	U-VPIN2	Vol$\times 10^{-7}$	VIX$\times 10^{-2}$	$\bar{R}^2$
Reg 1	0.01	0.39				1.98
	(0.32)	(7.34)				
Reg 2	-0.14		0.72			5.27
	(-5.35)		(10.88)			
Reg 3	0.02			0.57		8.64
	(2.27)			(13.10)		
Reg 4	-0.02				0.55	20.87
	(-4.44)				(36.35)	
Reg 5	0.07	-0.16		0.64		8.84
	(3.86)	(-3.09)		(13.00)		
Reg 6	-0.07	0.15			0.54	21.15
	(-6.92)	(5.80)			(37.11)	
Reg 7	-0.04			0.18	0.50	21.45
	(-7.75)			(7.40)	(34.02)	

Panel B: 50-Period Forecast

	Const.	VPIN	U-VPIN2	Vol$\times 10^{-7}$	VIX$\times 10^{-2}$	$\bar{R}^2$
Reg 1	0.02	0.36				6.28
	(1.17)	(7.34)				
Reg 2	-0.13		0.69			18.61
	(-5.16)		(11.10)			
Reg 3	0.02			0.57		33.31
	(2.45)			(13.62)		
Reg 4	-0.02				0.55	78.49
	(-4.41)				(40.63)	
Reg 5	0.08	-0.21		0.66		34.75
	(5.46)	(-4.54)		(13.42)		
Reg 6	-0.06	0.11			0.54	79.05
	(-6.38)	(5.07)			(40.96)	
Reg 7	-0.04			0.17	0.49	80.52
	(-8.00)			(8.05)	(38.91)	

Notes: The figures represent OLS regression coefficients. t-statistics based on HAC-standard errors, constructed with 50 lags, are reported in parentheses. VPIN and U-VPIN are for $\delta = 60$ seconds; "Vol" is the one-day backward trading volume. The sample period is 01/2008-07/2010.

information, as reflected in VPIN, but not U-VPIN2, is detrimental to forecast performance. This brings us to row three of the panels which document another dramatic improvement from simply using the lagged one-day trading volume as a forecast variable for future volatility. The explanatory power is now fivefold that obtained with VPIN. Along similar lines, row five demonstrates that VPIN has no – or even negative – auxiliary explanatory power in forecasting return volatility once we control for overall trading activity. Obviously, the two regressors are strongly correlated, rendering the point estimates somewhat unreliable, but the minimal increase in overall explanatory power relative to the univariate trading volume regression in row three confirms the lack of independent predictive content of the VPIN metric. Finally, row four shows that the VIX index provides superior predictive power relative to trading volume.[10] Nonetheless, as evidenced by rows six and seven, the trading activity variables do contain useful information for future return volatility over and above the VIX measure although the improvements in explanatory power over using VIX alone is marginal.

FINDING 8

The ELO 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 U-VPIN2 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 ELO VPIN metric is weakly correlated with future return volatility largely because of its correlation with trading volume. Once we control for trading volume, ELO VPIN is, if anything, negatively related to future return volatility.

One concern about the above analysis is that it utilizes the entire sample and thus mixes evidence across all market regimes. Consequently, it does not allow for the possibility that VPIN is relatively uninformative during benign times, but may become highly informative about market conditions as order flow turns decidedly toxic. This is a harder hypothesis to test using a relatively short historical sample, so we turn to a more informal descriptive account of the behavior of VPIN across the dramatic events associated with the flash crash.

6 Revisiting the Flash Crash

6.1 A Closer Look at the Dynamics of Trading Activity

Table 5 provides summary statistics for the trading process on May 6, 2010. Comparing the statistics with the corresponding numbers in Table 1, we note that the trading volume was more than 250% of the daily average for the sample. However, there is no significant difference in the order size distribution for May 6, 2010, relative to a regular day.

Next, Figure 5 complements our original depiction of the trading activity on this day in Section 2.2. It displays the evolution of numerous alternative VPIN measures and the dynamics

[10] Moreover, we note that VIX is not an ideal predictor of future return volatility, since it is feasible to construct better performing forecast measures using different aspects of the option data, see, for example, Andersen and Bondarenko (2007) for evidence.

Table 5: **Descriptive Statistics for the E-mini S&P 500 Futures, May 6, 2010**

	All	RTH	OTH
Volume (1 day), 000s	5672.59	5097.10	575.49
# Trades (1 day), 000s	439.28	383.09	56.19
Volume (1 min), 000s	3.98	12.55	0.56
# Trades (1 min), 000s	0.31	0.94	0.06
Order Size	12.91	13.31	10.24

Order Size Percentiles

	Min	10%	50%	75%	90%	99%	99.9%	Max
All	1.0	1.0	2.0	6.0	25.0	200.0	559.0	2334.0

Notes: This table reports summary statistics for the trading in the E-mini S&P 500 futures contract on May 6, 2010. The data are reported separately for Regular Trading Hours (RTH, 8:30-15:15 CT), Overnight Trading Hours (OTH, 15:15-8:30 CT), and combined hours (All).

of the order size as well as the trading intensity, as given by the number of minutes it takes to trade 35,000 contracts (one volume bucket), throughout the day.

Focusing initially on the bottom right panel, we note that the trading intensity around the flash crash rose to levels which imply that each volume bucket between 13:10-15:00 CT contains no more than three one-minute time bars, and, even more strikingly, from 13:40-13:55 CT, the volume buckets were filled in less than one minute. This tautologically implies a string of volume buckets producing OI measures of near unity, irrespective of the underlying order imbalances. This is confirmed by the qualitatively similar increase in U-VPIN1 and U-VPIN2 over this period for both $\delta = 60$ and $\delta = 10$. In other words, the one-minute time bar VPIN *must* rise sharply, for purely mechanical reasons, due to the acceleration in trading activity, and independently of the degree of toxicity in the market. While it is likely we would never observe such a trading pattern in the absence of rapidly rising order flow toxicity, the dynamics of VPIN simply cannot shed light on the issue. For VPIN to signal the impending market chaos, it must attain extreme values *prior* to the crash. Turning towards this crucial timing dimension, we observe that the VPIN and U-VPIN series for $\delta = 10$ and $\delta = 60$ reach their maximum after 14:00 CT, and thus some time after the flash crash cycle has played itself out. Again, this may be viewed as a mechanical consequence of the construction of the metric. VPIN is computed from a long moving-average of order imbalance measures which induces extreme short run momentum in response to a rapid and maintained escalation in the trading activity. As observed in Finding 1, the evidence for VPIN reaching an extreme level prior to the crash is less compelling, but we pursue this question further shortly.

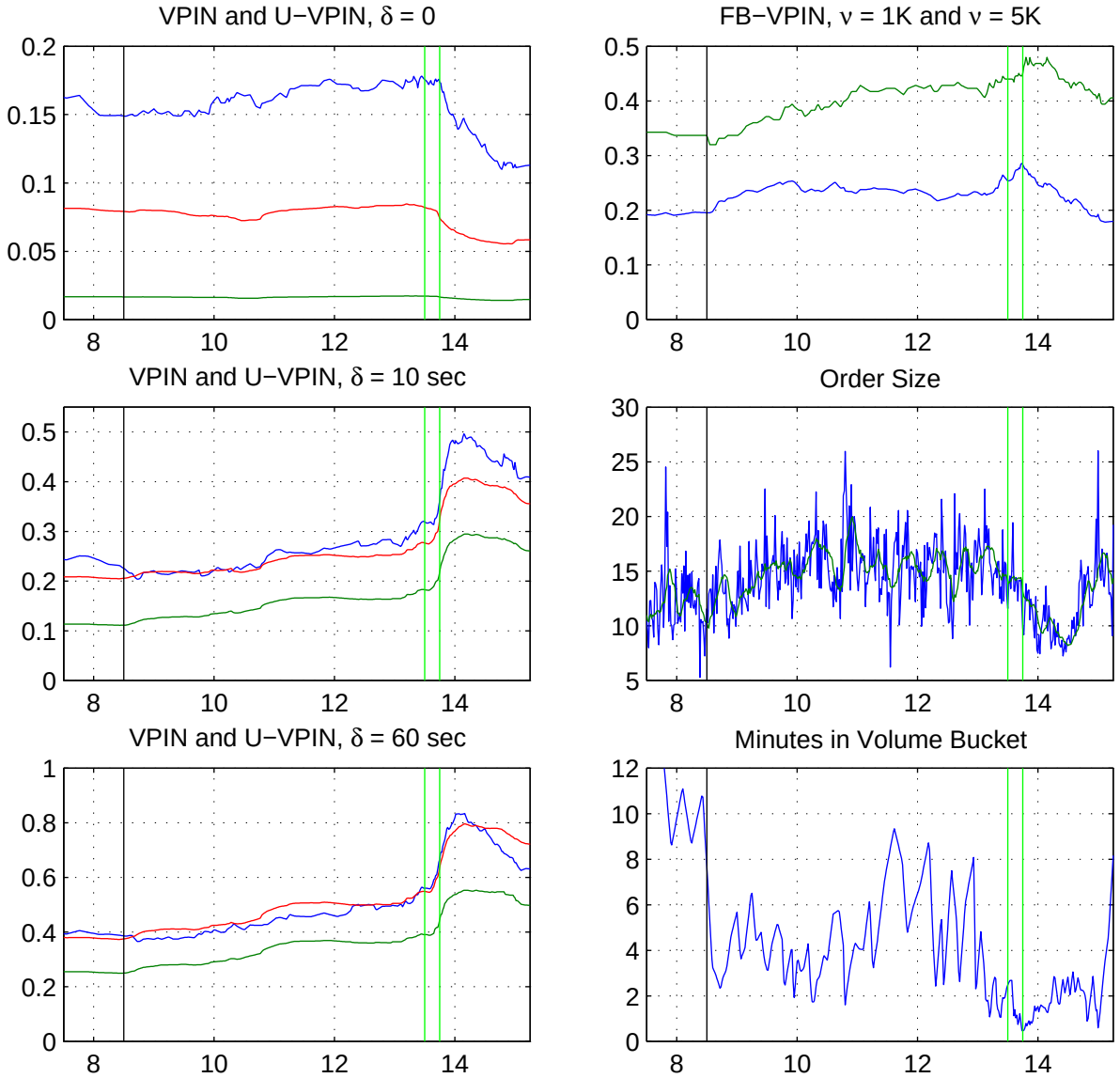

Figure 5: For May 6, 2010, this figure plots market statistics at 1-minute frequency from 7:30-15:15 CT. The black vertical line shows the start of the regular trading hours (RTH). Left panels: (1) VPIN (blue), U-VPIN1 (green), and U-VPIN2 (red) for $\delta = 0$, (2) same for $\delta = 10$ sec, (3) same for $\delta = 60$ sec. Right panels: (1) VPIN for fixed volume bins with $\nu = 1K$ (blue) and $\nu = 5K$ (green); (2) the average trade size and its 10-minute moving average, (2) the estimated number of 1-minute intervals in volume bucket of size V.

FINDING 10

The explosive increase in one-minute 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 many of the underlying OI measures mechanically were attaining their maximum value of unity during this period, independent of the actual order flow imbalances in the market.

A few additional observations on Figure 5 are warranted. First, transaction VPIN again behaves very differently from ELO VPIN. It is flat during the crash itself and then drops off sharply as the prices start rebounding following the crash, even as one-minute VPIN continues to soar. The FB-VPIN for $\nu = 1{,}000$ displays features quite similar to transactions VPIN, while the FB-VPIN for $\nu = 5{,}000$ represents an intermediate case between ELO and transaction VPIN. Second, we observe that the VIX index also was elevated for a relatively long period and seems to reach a maximum over 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 for the day. Third, we note that the dramatic increase in trading activity around the flash crash may have received a minor boost from a drop in the average order size, presumable resulting from a decline in the depth of the order book.

6.2 VPIN as Crash Predictor

The issue of whether VPIN provided a signal indicating a sharply elevated degree of order toxicity and a high risk of impending market turmoil 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 VPIN was exceptional compared to the past history.

To this end, we construct three separate figures with scatter plots depicting pairwise observations of alternative volatility predictors versus subsequent realized volatility over different forecast horizons. The associated linear regression lines convey the predictions from a simple linear model exploiting available historical evidence. The sample covers the beginning of our sample through the start of the flash crash, thus reflecting the real-time perspective that an observer may have acquired at that juncture by constructing the various candidate volatility predictors on a high-frequency basis. Note that we exploit non-overlapping forecast horizons to avoid excessive cluttering of the displays.

Figure 6 provides the results for the 50-volume-bucket-ahead forecast horizon. The upper right panel refers to the ELO version of VPIN, and the red dots indicate the VPIN values observed over the five days preceding the flash crash. It is hard to detect any unusual pre-crash pattern: even if 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 ELO VPIN relative to a number of the other candidate predictor variables is evident by the relative diffuse shape of the scatter plot. 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.

Unfortunately, one problem with the relatively long forecast horizon employed in Figure 6 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.,

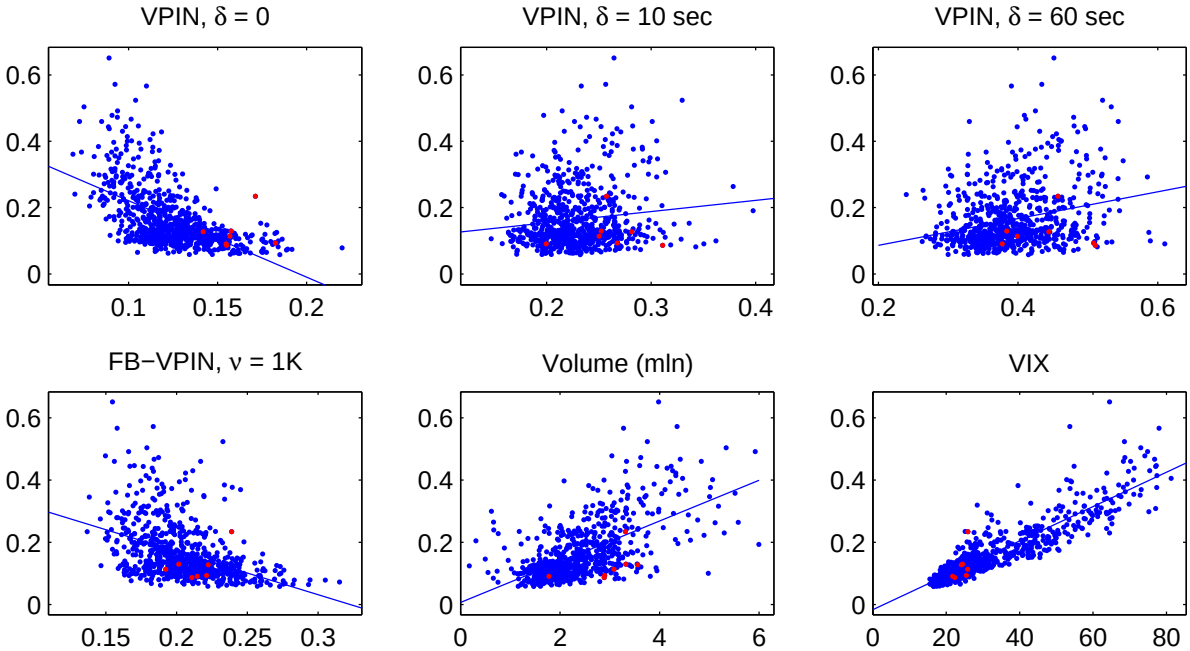

Figure 6: This figure shows 6 scatter plots of average absolute return $AAR(t, t+50)$ versus VPIN with $\delta = 0$, 10, and 60 sec (top panels), versus VPIN with fixed volume bins $B = 1000$, Volume, and VIX (bottom panels). "Volume" is one-day trading volume, in millions. $AAR(t, t+T) = \frac{1}{T}\sum_{i=1}^{T}|r_{t+i}|$, where $r_t = 100 \ln P_t/P_{t-1}$ is the log return of the S&P 500 futures over one volume bucket. Red dots indicate the 5 days preceeding 13:30 on May 6, 2010. The sample period is 01/01/2008-05/06/2010.

Andersen and Bollerslev (1998).

The plots in Figure 7, reflecting a forecast horizon of ten volume buckets, produce qualitatively similar results to those discussed above. Likewise, moving to the extreme end of the spectrum, and only forecasting the absolute return over the next volume bucket, does not alter the impression that one-minute VPIN moves around within an elevated, but nonetheless fairly standard range prior to the crash on May 6, 2010. The main difference is that all the "clouds" become more diffuse, reflecting the use of very imprecise volatility proxies. In summary, when focusing on the pre-crash sample, there is no indication that real-time data on VPIN would have forewarned an observed of the impending market turmoil.[11]

FINDING 11

The evolution of the one-minute 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 the historical extremes, and it was not otherwise displaying any truly extraordinary behavior.

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

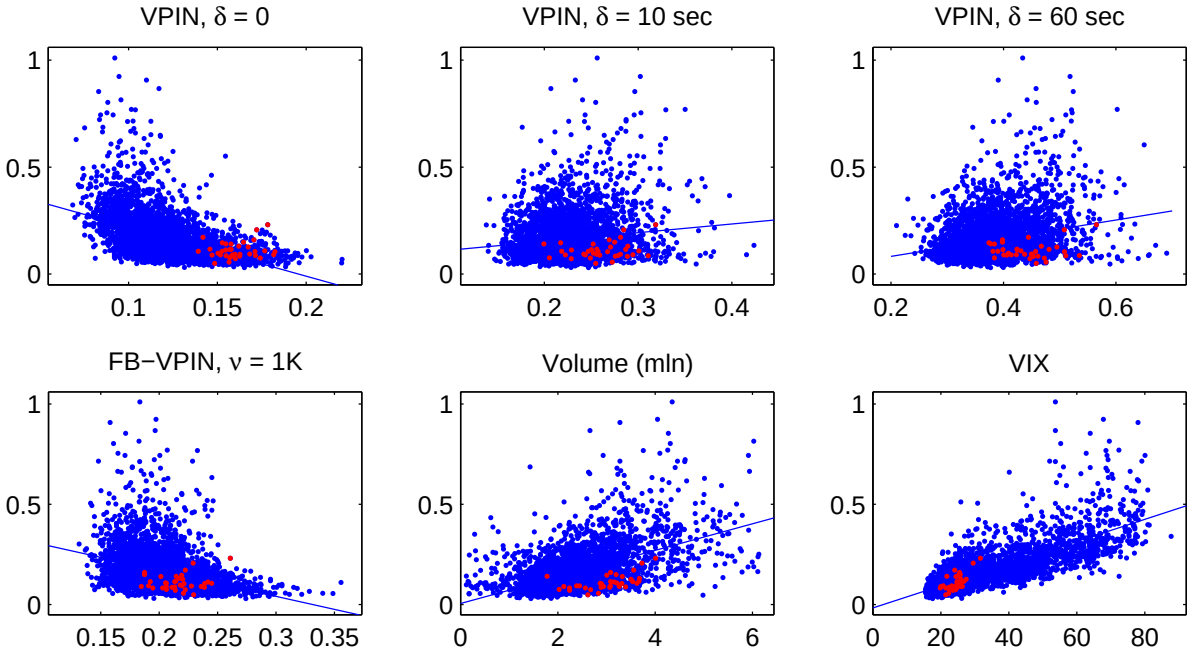

Figure 7: This figure shows 6 scatter plots of average absolute return $AAR(t, t+10)$ versus VPIN with $\delta = 0$, 10, and 60 sec (top panels), versus VPIN with fixed volume bins $B = 1000$, Volume, and VIX (bottom panels). "Volume" is one-day trading volume, in millions. $AAR(t, t+T) = \frac{1}{T}\sum_{i=1}^{T} |r_{t+i}|$, where $r_t = 100 \ln P_t/P_{t-1}$ is the log return of the S&P 500 futures over one volume bucket. Red dots indicate the 5 days preceeding 13:30 on May 6, 2010. The sample period is 01/01/2008-05/06/2010.

7 Some Final Reflections on VPIN and Potential Modifications

7.1 The Issue of Replication and Verifiability

The illustration in Section 4.1 shows that VPIN can be very sensitive to minor changes in the trading process. This suggests that the measure is not robust to small errors or changes in the transaction record. Moreover, it is clearly dependent on the initial conditions as, given the recorded trade sequence, the point at which we start cumulating the trading volume determines the location of the volume buckets for any day throughout the sample. It is not clear, however, whether this is a major concern or whether the use of a long moving average for computing VPIN helps diversify and mitigate such effects so as to render them immaterial.

To assess the magnitude of such effects, and their potential impact on inference regarding the events during the flash crash, we compile 350 different version of one-minute VPIN for March 6, 2010, with each trajectory corresponding to a different location of the volume buckets. Using a bucket of $V = 35{,}000$, 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 350th trajectory is initiated after 34,900 contracts have been traded. This provides an empirical bootstrap of the distribution of VPIN on the day of the flash crash. In reality, the starting point of the sample, years earlier, determines the exact location of the volume buckets on this day, and any shift in this starting date will alter the placement of the buckets on May 6, 2010.

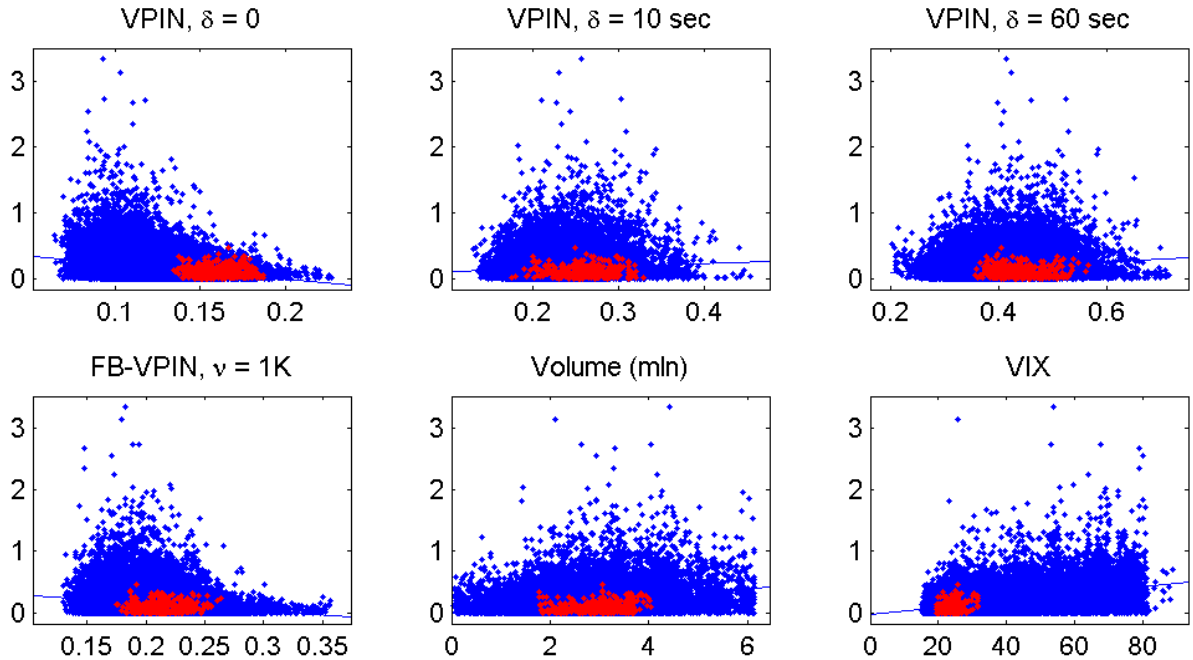

Figure 8: This figure shows 6 scatter plots of average absolute return $AAR(t, t+1)$ versus VPIN with $\delta = 0$, 10, and 60 sec (top panels), versus VPIN with fixed volume bins $B = 1000$, Volume, and VIX (bottom panels). "Volume" is one-day trading volume, in millions. $AAR(t, t+T) = \frac{1}{T}\sum_{i=1}^{T} |r_{t+i}|$, where $r_t = 100 \ln P_t/P_{t-1}$ is the log return of the S&P 500 futures over one volume bucket. Red dots indicate the 5 days preceeding 13:30 on May 6, 2010. The sample period is 01/01/2008-05/06/2010.

Figure 9 summarizes the effects of shifting the location of the volume buckets for the critical part of the trading day on May 6, 2010. The upper left panel shows that VPIN often fluctuates within a band exceeding 20% of the median level across the trajectories. The upper right panel displays only five of the trajectories to allow closer inspection. It is evident that the individual trajectories behave quite differently, even if they are correlated. For example, the dark blue trajectory is at the highest level of the five at 11am CT, and it enters the crash period at the second highest level. However, during the crash period, this VPIN trajectory actually drops and its subsequent increase in much less pronounced than that associated with, say, the violet trajectory. In fact, the latter evolves quite 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 VPIN trajectories at the start of the crash (0.55-0.60) and at their highest level during the day (0.67-0.82). The left panel is relevant for judging whether VPIN attained a historical 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.

FINDING 12

The one-minute 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 significant impact on the VPIN metric throughout the entire sample.

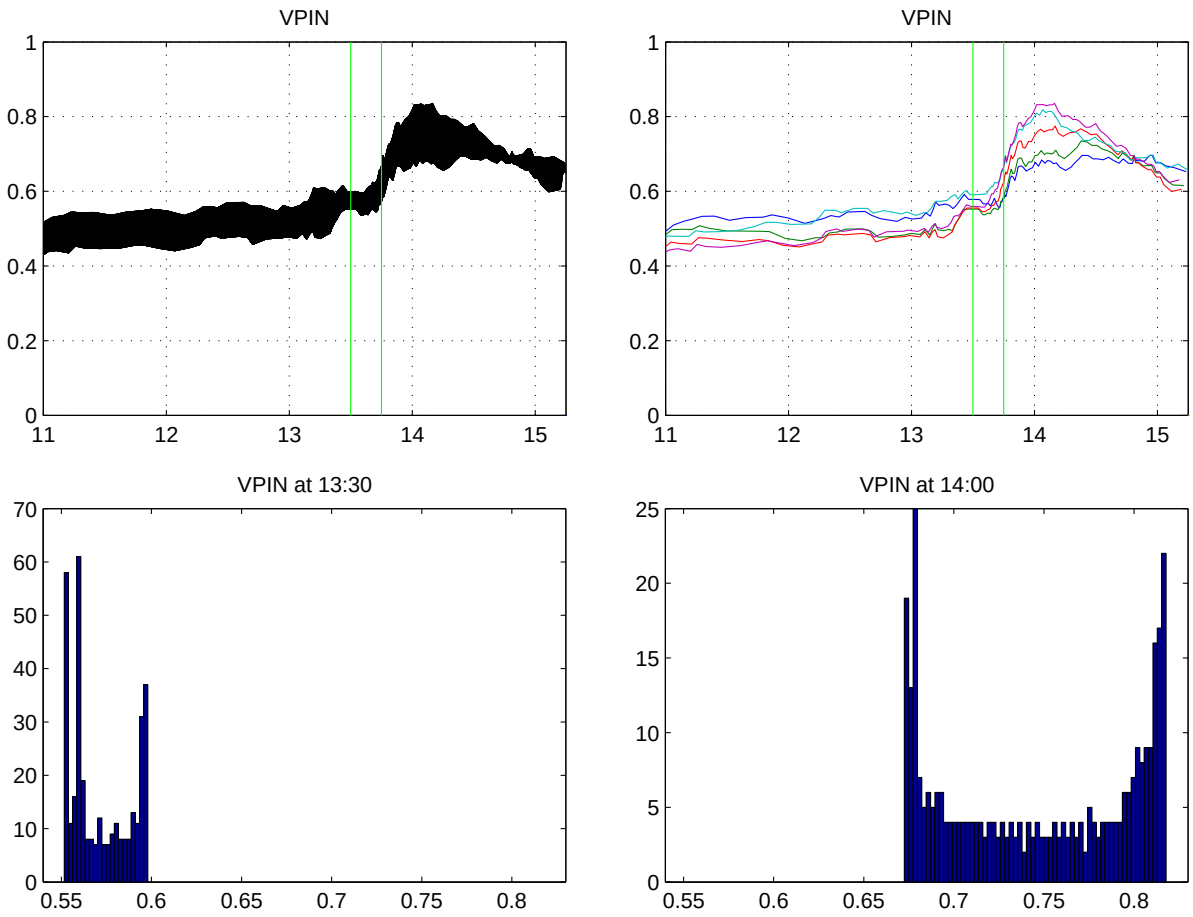

Figure 9: For May 6, 2010, this figure illustrates the distribution of VPIN values ($\delta = 60$ sec) for different starting points of the volume grid. The bottom panels show the histograms of VPIN values at 13:30 and 14:00. The top left panel shows the VPIN trajectories generated by shifting the starting point every 100 contracts. The top right panel shows only 5 trajectories, which are selected based on the VPIN values at 14:00 (min, 25%, 50%, 75%, and max).

This brings up the issue of replication of the VPIN metric. Unless there is exact agreement concerning the starting point and the recording and status of all transactions across two alternative data sources, alternative computations of VPIN will produce significantly diverging behavior across critical periods in the sample. This presents practical problems for the use of the metric as an indicator of market stress and for the use of VPIN as the basis for a futures contract. While an exchange may take on the task of monitoring the evolution of the relevant transaction series, it is likely to remain an issue of contention that the realized VPIN series will depend critically on inclusion or exclusion of specific trades.

In our case, we had to calibrate the starting date of our sample carefully to obtain a sequence of VPIN observations that mirror those obtained by ELO across the flash crash. This led us to adopt a slightly earlier starting date for our construction of the VPIN metric than ELO, namely December 27, 2007, rather than the initial trading day of 2008.

Another concern is the potential lack of robustness of empirical work involving VPIN. The series cannot be readily replicated and the conclusions may hinge critically on rather idiosyncratic features of the VPIN design such as the original starting point of the sample, the exact source of the data, and the criterion adopted for excluding "unusual" trades from the series. At a minimum, some standard for robustness should be adopted, documenting the validity of inference across the range of possible alternative configurations of the volume buckets across the relevant part of the sample.

7.2 Constructing Signed VPIN Measures

The dispersion in VPIN values documented above may be surprising. One might expect unusual random fluctuations in the OI measure to diversify away across the fifty terms used to compute the VPIN metric. However, since the OI measures represent absolute values, there is no opportunity for large positive and negative values to cancel so, instead, the random outliers will tend to cumulate across the moving average statistic.

Figure 10 illustrates this feature of the VPIN construction. The right panels display the SOI measures associated with two of the VPIN trajectories from Figure 9. It is evident that the blue trajectory, by chance, has many more moderate SOI observations, falling largely within the range of [-0.50, 0], than the green trajectory during the period from 1-2pm CT. This translates into significant difference in the behavior of the corresponding one-minute VPIN series depicted in the top left panel. The green trajectory starts out well below the blue at 1pm but ends up at a substantially higher level at 2pm. However, if we average the signed SOI measures, rather than the absolute SOI measures, we obtain alternative signed, smoothed order imbalance indicators. These are plotted in the lower left panel. Now, the blue and green trajectories nearly coincide, showing that diversification across fifty observations is effective once we retain the sign of the SOI measures. Apparently, these signed VPIN measures are far less sensitive to the positioning of the volume buckets and may thus avoid the stark dependence on initial conditions that plague the regular VPIN metric. Moreover, they provide a clear message: there was a growing dominance of active selling in the market from noon until the end of the crash, and then an immediate reversal towards restoration of the order imbalance which is complete before 3pm.

We find these preliminary observations intriguing. They raise the possibility that constructive use of this type of real-time order imbalance indices is feasible in practice. Obviously, one concern is that they will turn out to be extremely highly correlated with the realized price path and thus not provide much independent information. However, it is far beyond the scope of this paper to present a thorough analysis of such signed VPIN metrics and related measures.

FINDING 13

The use of absolute (OI), rather than signed, order imbalance (SOI) measures in the construction of one-minute VPIN tends to inflate the idiosyncratic variation of the metric. The use of signed order imbalance measures, along with their dynamic evolution, may improve robustness and help identify the direction of developing order imbalances in the market.

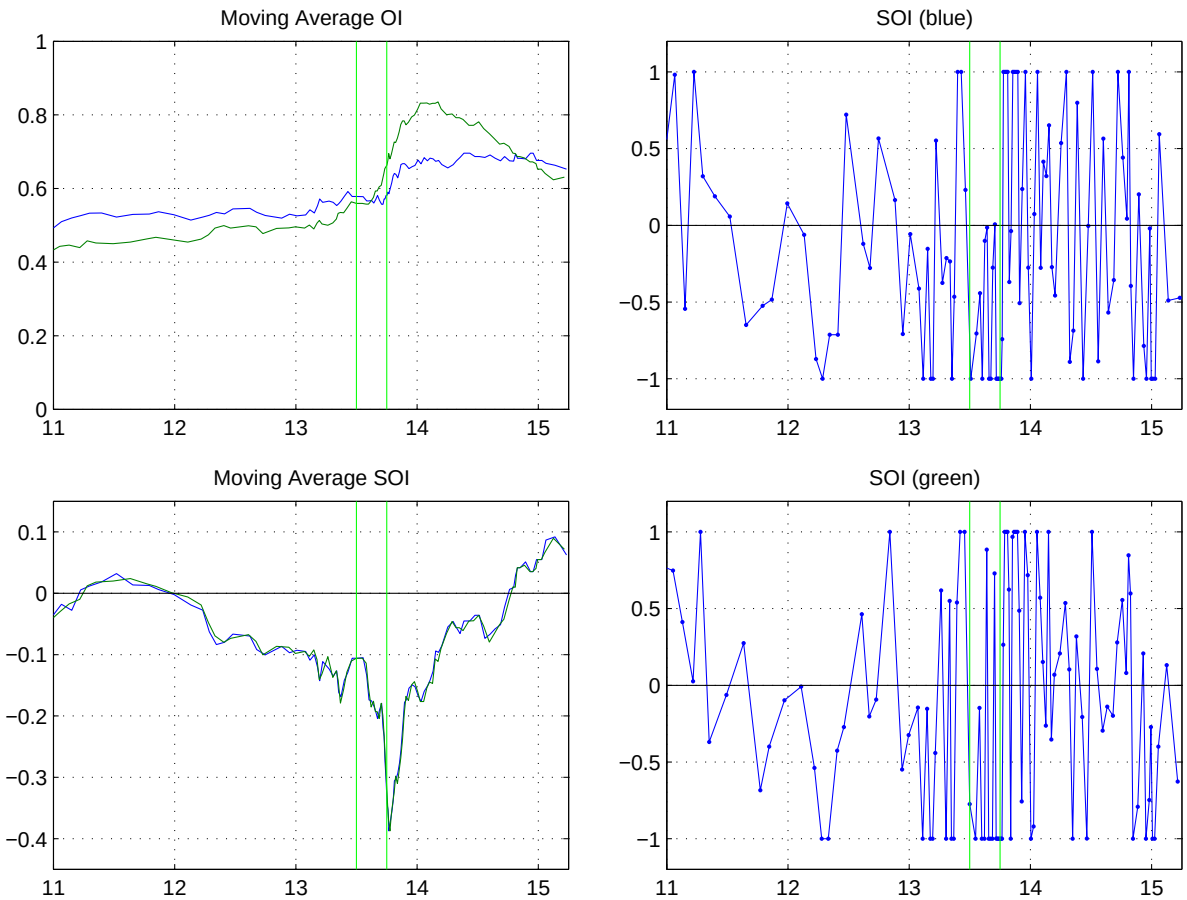

Figure 10: For May 6, 2010, this figure illustrates the signed OI ($\delta = 60$ sec) for two different starting points of the volume grid, corresponding to the minimum (blue) and maximum (green) values of VPIN at 14:00. Note that for this plot, "OI" stands for the *signed* open interest.

7.3 On the Behavior of Transactions VPIN

Perhaps our most striking finding is the fact that transaction and fixed volume bin VPIN produce opposite conclusions of those obtained via time-bar VPIN. The effect of moving from time bars to tick or fixed 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 will take us outside the scope of the present work.

ELO (2011c) argue that the results obtained from tick data are so counterintuitive that they, a priori, should be disregarded. They suggest the seemingly perverse results stem from a massive amount of trade misclassification at the high-frequency level. This is not entirely 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 bid-ask 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 very long sequences of errors, and those that do occur will tend to be alleviated through diversification via the averaging across a huge set of transactions. For example, with a volume bucket of $V = 35{,}000$, there are thousands of transactions in the 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 blocks defined by volume bins or time bars, and thus likely misclassifying close to half them. If these blocks of trades are large, the errors from using high aggregation levels do not readily diversify away. 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 the transaction VPIN measure as well as fixed volume bins? In particular, why is transaction VPIN strongly negatively correlated with trading volume and return volatility, and why do these variants of VPIN drop rather than soar during episodes like the flash crash? We suggest that the following closely related facts provide part of the explanation.

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

Second, and importantly, for volume bins consisting of 1,000 and 5,000 traded contracts, we find an average daily probability of a "continuation" in the trade indicator across consecutive bins to be 67% and 58%, respectively. These trade persistence measures turn out to be strongly negatively related to the VIX index, with correlations of -0.67 and -0.39, respectively. These results are consistent with the hypothesis that an increase in volatility shrinks the proportion of zero tick change buckets and thus raises the frequency of oscillation between positive to negative tick changes. If the tick changes are close to symmetric over short horizons, as the semi-martingale property implies, then this leads to a lower degree of persistence in the trade classification indicator, and hence a drop in the associated VPIN measures. However, it is clear that many 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 14

Transaction and fixed volume bin VPIN have negative correlation with volatility because the trade classification produces less serially correlated trade indicators during volatile market conditions. All else equal, this leads to lower order imbalance measures, for a given volume bucket, as the volume bins (transactions) diversify more effectively when successive bins (transactions) more frequently attain opposite signs. The reason behind this negative correlation between return volatility and the trade classification rule is worthy of additional study.

8 Conclusion

Inspired by the striking empirical evidence of ELO (2011a), we have undertaken a detailed study of the VPIN measure and its ability to forecast short term return volatility and signal impending market turmoil. Our results are largely non-confirmatory. First, we note that VPIN almost by construction is highly correlated with trading 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 VPIN concerning short term future volatility. Similarly, constructing VPIN measures from volume bins or transaction data, avoiding the mechanical correlation with trading volume, we find VPIN 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 as volatility and economic uncertainty increases. Second, we found the 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 underlying trading pattern. As such, the identification of order flow imbalances via the trade classification scheme has no visible impact on VPIN over this period.

Our findings suggest that the ELO construction of VPIN may not be the most suitable way of monitoring order flow imbalances and market tensions. This skepticism is further bolstered by the idiosyncratic variation in VPIN associated with the location of the volume buckets which renders it hard to replicate empirically on a day-by-day basis. A major reason for this non-robustness is the computation of VPIN from a moving average of absolute order imbalance measures. Seemingly, a signed version of the metric would be less sensitive to such features and it would also allow for more direct monitoring of cumulative (signed) order imbalances. Nonetheless, a signed VPIN measure would still be subject to a number of the potential distortions analyzed in this paper, so a more in-depth analysis is warranted prior to adoption. By the same token, it would be of interest to further study the empirical performance of the signed transaction and FB-VPIN measures.

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 the ELO procedure for computing the VPIN metric is not ideal. The search for a suitable alternative is high on the research agenda. We hope the type of analysis performed and the battery of tests undertaken in this paper ultimately will be helpful in identifying the more suitable and robust market stress indicators among the very 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 denote their sum

$$S_Q = \sum_{i=1}^{Q} b_i.$$

The function $F(Q)$ can be computed using the expectation of the absolute value $|S_Q|$ as

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

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

$$G(Q) := \sum_{i=0}^{Q} C_Q^i \cdot |Q - 2 * i| = \begin{cases} (2q)C_{2q}^q, & \text{if } Q = 2q \\ 2(2q+1)C_{2q}^q, & \text{if } Q = 2q+1 \end{cases} \tag{13}$$

It is easy to verify that the relationship in (13) 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) $C_{n+1}^{k+1} = C_n^k + C_n^{k+1}, \qquad \text{for all} \quad 0 \le k \le n-1.$

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

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

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

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,

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

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

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

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

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

The proof of (5) now obtains from (13) 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:

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

□

References

Andersen, T.G., 1996, "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance*, 51, pp. 169-204.

Andersen, T.G., and O. Bondarenko, 2007, "Construction and Interpretation of Model-Free Implied Volatility," *Volatility as an Asset Class*, London: Risk Books, I. Nelken (editor), pp. 141-181.

Andersen, T.G., and T. Bollerslev, 1998, "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, pp. 885-905.

Andersen, T.G., O. Bondarenko, and M.T. Gonzalez-Perez, 2011, "A Corridor Fix for VIX: Constructing a Coherent Model-Free Option Implied Volatility Measure," Working Paper, Kellogg School of Management, Northwestern University and College of Business Administration, 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.

Clark, P., 1973, "Liquidity, Information, and Infrequently Traded Stocks," *Econometrica*, 41, pp. 135-156.

Epps, T.W., and M.L. Epps, 1976, "The Stochastic Dependence of Security Price Changes and Transactions Volumes: Implications for the Mixture of Distributions Hypothesis," *Econometrica*, 44, pp. 305-321.

Easley, D., N.M. Kiefer, M. O'Hara, and J.B. Paperman, 1996,"Liquidity, Information, and Infrequently Traded Stocks," *Journal of Finance*, 51, pp. 1405-1436.

Easley, D., M. López de Prado, and M. O'Hara, 2011a,"The Microstructure of the "Flash Crash": Flow Toxicity, Liquidity Crashes, and the Probability of Informed Trading," *Journal of Portfolio Management*, 37 (2), pp. 118-128.

Easley, D., M. López de Prado, and M. O'Hara, 2011b,"The Exchange of Flow Toxicity," *Journal of Trading*; forthcoming.

Easley, D., M. López de Prado, and M. O'Hara, 2011c,"Measuring Flow Toxicity in a High-Frequency World," Working Paper, SSRN, February, 2011.

Kirilenko, A., A.S. Kyle, M. Samadi and T. Tuzun, 2011,"The Flash Crash: The Impact of High Frequency Trading on an Electronic Market," *Journal of Trading*; Working Paper, SSRN, May 2011.

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

Tauchen, G.E., and M. Pitts, 1983,"The Price Variability-Volume Relationship on Speculative Markets," *Econometrica*, 51, pp. 485-505.