Flow Toxicity and Liquidity in a High Frequency World

David Easley
Scarborough Professor and Donald C. Opatrny Chair
Department of Economics
Cornell University
dae3@cornell.edu

Marcos M. Lopez de Prado
Head of High Frequency Futures Trading at Tudor Investment Corporation
and Postdoctoral Fellow of RCC at Harvard University
marcos.lopezdeprado@tudor.com

Maureen O'Hara
Purcell Professor of Finance
Johnson Graduate School of Management
Cornell University
mo19@cornell.edu

February, 2012
Forthcoming, **Review of Financial Studies**

We thank the editor of the Review of Financial Studies, Matthew Spiegel, and an anonymous referee for helpful comments. We also thank Robert Almgren, Torben Andersen, Oleg Bondarenko, John Campbell, Ian Domowitz, Robert Engle, Andrew Karolyi, Mark Ready, Riccardo Rebonato, Luis Viceira, and seminar participants at the CFTC, the SEC, Cornell University, Harvard University, the QFF seminar at the Chicago Mercantile Exchange, the University of Notre Dame, Complutense University, Risk Magazine's Quant Congress 2011, the University of Piraeus and ITG for helpful comments. We acknowledge David Leinweber, Kesheng Wu and the CIFT group at the Lawrence Berkeley National Laboratory for confirming our VPIN calculations on the 'flash crash'. We are grateful to Sergey Kosyakov and Steven Jones for their research assistance.

* VPIN is a trademark of Tudor Investment Corp. The authors have applied for a patent on VPIN and have a financial interest in it.

Flow Toxicity and Liquidity in a High Frequency World

ABSTRACT

Order flow is toxic when it adversely selects market makers, who may be unaware they are providing liquidity at a loss. We present a new procedure to estimate flow toxicity based on volume imbalance and trade intensity (the *VPIN* toxicity metric). VPIN is updated in volume-time, making it applicable to the high frequency world, and it does not require the intermediate estimation of non-observable parameters or the application of numerical methods. It does require trades classified as buys or sells, and we develop a new bulk volume classification procedure that we argue is more useful in high frequency markets than standard classification procedures. We show that the VPIN metric is a useful indicator of short-term, toxicity-induced volatility.

Keywords: Flash crash, liquidity, flow toxicity, volume imbalance, market microstructure, probability of informed trading, VPIN.

JEL codes: C02, D52, D53, G14.

Flow Toxicity and Liquidity in a High Frequency World

High frequency (HF) trading firms represent approximately 2% of the nearly 20,000 trading firms operating in the U.S. markets, but since 2009 they have accounted for over 70% of the volume in U.S. equity markets and are fast approaching 50% of the volume in futures markets (Iati [2009], CFTC [2010]). These HF firms typically act as market makers, providing liquidity to position-takers by placing passive orders at various levels of the electronic order book. A passive order is defined as an order that does not cross the market, thus the originator has no direct control on the timing of its execution. HF market makers generally do not make directional bets, but rather strive to earn tiny margins on large numbers of trades. Their ability to do so depends on limiting their position risk, which is greatly affected by their ability to control adverse selection in the execution of their passive orders.

Practitioners usually refer to adverse selection as the *"natural tendency for passive orders to fill quickly when they should fill slowly and fill slowly (or not at all) when they should fill quickly"* (Jeria and Sofianos [2008]). This intuitive formulation is consistent with market microstructure models (see Glosten and Milgrom [1985], Kyle [1985], and Easley and O'Hara [1987, 1992]), in which informed traders take advantage of uninformed traders. Order flow is regarded as toxic when it adversely selects market makers who may be unaware that they are providing liquidity at a loss.

This paper develops a new framework for measuring order flow toxicity in a high frequency world. A fundamental insight of the microstructure literature is that the order arrival process is informative for subsequent price moves in general, and flow's toxicity in particular. Extracting this information from order flow, however, is complicated by the very nature of

trading in high frequency markets. We argue that in the high frequency world, trade-time, as captured by volume, is a more relevant metric than clock-time. Information is also different, relating now to an underlying event that induces unbalanced or accelerated trade over a relatively short horizon. Information events can arise for a variety of reasons, some related to asset returns, but others reflecting more systemic or portfolio-based effects. Our particular application is to futures contracts, where information is more likely to be related to systemic factors, or to variables reflecting hedging or other portfolio considerations.

We present a new procedure to estimate flow toxicity directly and analytically, based on a process subordinated to volume arrival, which we name *Volume-Synchronized Probability of Informed Trading*, or the *VPIN* flow toxicity metric. The original PIN estimation approach (see Easley, Kiefer, O'Hara and Paperman [1996]) entailed maximum likelihood estimation of unobservable parameters fitted on a mixture of three Poisson distributions of daily buys and sells on stocks. That static approach was extended by the Easley, Engle, O'Hara and Wu [2008] GARCH specification, which models a time-varying arrival rate of informed and uninformed traders. The approached based on the VPIN toxicity metric developed in this paper does not require the intermediate numerical estimation of non-observable parameters, and is updated in stochastic time which is calibrated to have an equal volume of trade in each time interval. Thus, our methodology overcomes the difficulties of estimating PIN models in highly active markets and provides an analytically tractable way to measure the toxicity of order flow using high frequency data.

We provide empirical evidence on the statistical properties of the VPIN metric. We show how volume bucketing (time intervals selected so that each has an equal volume of trade) reduces the impact of volatility clustering in the sample.[1] Because large price moves are

associated with large volumes, sampling by volume is a proxy for sampling by volatility.[2] The resulting time series of observations follows a distribution that is closer to normal and is less heteroskedastic than it would be if it were sampled uniformly in clock-time.

We illustrate the usefulness of the VPIN metric by estimating it for the E-mini S&P 500 futures (CME) and the WTI crude oil futures contract (NYMEX). We also demonstrate that VPIN has important linkages with future price variability. Because toxicity is harmful to liquidity providers, high levels of VPIN should presage high volatility. We show that VPIN predicts short-term toxicity-induced volatility, particularly as it relates to large price moves.

An incidental contribution of this paper is a new approach for classifying buy and sell volume. The speed and volume of trading in high frequency markets challenges traditional classification schemes for assigning trade direction. We propose a new "bulk volume" classification algorithm in which we aggregate trades over short time or volume intervals (respectively denoted time bars and volume bars) and then use the standardized price change between the beginning and end of the interval to approximate the percentage of buy and sell volume. We believe this new approach will be useful for a wide variety of applications in high frequency markets.

Estimates of the toxicity of order flow have a number of immediate applications. Market makers, for example, can use the VPIN metric as a real-time risk management tool. In other research (see Easley, López de Prado, and O'Hara [2011a]), we presented evidence that order flow as captured by the VPIN metric was becoming increasingly toxic in the hours before the May 6th 2010 'flash crash', and that this toxicity contributed to the withdrawal of many liquidity providers from the market.[3] Tracking the VPIN metric would allow market makers to control their risk and potentially remain active in volatile markets. Regulators and exchanges could use

the VPIN metric to monitor the conditions under which liquidity is provided, and pro-actively restrict trading or impose market controls if conditions deteriorate to the point that liquidity provision is threatened.[4] In a high frequency world, effective regulation needs to be done on an ex-ante basis, anticipating problems before, and not after, they lead to market breakdowns. Monitoring VPIN metric levels can signal when liquidity provision is at risk and allow for market halts, slowdowns or other regulatory actions to forestall crashes. Furthermore, this will limit the success of predatory algorithms that attempt to profit from a failure of the liquidity provision process. Traders can also use measures based on the VPIN metric in designing algorithms to control execution risks. Microstructure models have long noted (see, for example, Admati and Pfleiderer [1988]) that intra-day seasonalities can reflect the varying participation rates of informed and uninformed traders. Designing algorithms to delay or accelerate trading depending on the VPIN metric may reduce the so- called implementation shortfall.

Our analysis of order toxicity and its effects in high frequency markets is related to a growing body of recent research looking at high frequency trading in a multiplicity of markets. Hendershott and Riordan [2009] present evidence on HF trading on the Deutsche Borse; Brodegaard [2010] and Hasbrouck and Saar [2010] analyze the role and strategies of high frequency traders in U.S. equity markets. Kirilenko, Kyle, Samadi and Tuzun [2010] extensively characterize the behavior of HF traders and other market participants in the S&P 500 futures market. There is also a developing literature looking at the more normative effects of computerized or HF trading on liquidity. Hendershott, Jones and Menkveld [2011] study the empirical relationship between algorithmic trading and liquidity, finding that algorithmic trading improves liquidity for large stocks. Chaboud, Chiquoine, Hjalmarsson, and Vega [2009] provide a similar analysis of the effects of computerized trading in foreign exchange. These analyses

complement recent theoretical research looking at the relation between liquidity and market fragility (see Brunnermeir and Pedersen [2009] and Huang and Wang [2011]).

Methodologically, this paper is related to research by Engle and Lange [2001] and Deuskar and Johnson [2011]. Engle and Lange proposed a market depth measure, VNET, which is calculated using order imbalance measured over price change increments. Our analysis is calculated over volume increments (or buckets), but both their analysis and ours depart from standard time-based approaches to analyze the effects of asymmetric information in dynamic market environments. Deuskar and Johnson also analyze order flow imbalance in futures markets. These authors estimate the flow-driven component of systematic risk and its dynamic properties. Our focus is not on asset pricing issues, but their finding that flow-driven risk accounts for over half of the risk in the market portfolio underscores our argument that order flow imbalance (a source of toxicity) has important effects on market behaviour and performance.

This paper is organized as follows. Section 1 discusses the theoretical framework and shows how PIN impacts the bid-ask spread. Section 2 presents our procedure for estimating the VPIN metric. Section 3 evaluates the robustness of the VPIN metric. Section 4 provides estimates of the VPIN metric for equity indices and oil futures. Section 5 discusses predictive properties of the VPIN metric for volatility. Section 6 summarizes our findings. Technical Appendices (available on the RFS web site) present the pseudocode for computing the VPIN toxicity metric, and its Monte Carlo accuracy.

1. THE MODEL

In this section we describe the basic model that allows us to infer the toxicity of order flow. We begin with a standard microstructure model in which we derive our measure of flow toxicity, PIN, and we then show how to modify PIN to apply it to high frequency markets. Readers conversant with the standard PIN approach can proceed directly to Section 2.

In a series of papers, Easley et. al. demonstrate how a microstructure model can be estimated for individual stocks using trade data to determine the probability of information-based trading, PIN. This microstructure model views trading as a game between liquidity providers and traders (position takers) that is repeated over trading periods $i=1,\ldots,I$. At the beginning of each period, nature chooses whether an information event occurs. These events occur independently with probability α. If the information is good news, then informed traders know that by the end of the trading period the asset will be worth $\overline{S}_i$; and, if the information is bad news, that it will be worth $\underline{S}_i$, with $\overline{S}_i > \underline{S}_i$. Good news occurs with probability (1-δ) and bad news occurs with the remaining probability, δ. After an information event occurs or does not occur, trading for the period begins with traders arriving according to Poisson processes throughout the trading period. During periods with an information event, orders from informed traders arrive at rate μ. These informed traders buy if they have seen good news, and sell if they have seen bad news. Every period, orders from uninformed buyers and uninformed sellers each arrive at rate ε.[5]

The structural model relates observable market outcomes (i.e. buys and sells) to the unobservable information and order processes that underlie trading. The previous literature focuses on estimating the parameters determining these processes via maximum likelihood. Intuitively, the model interprets the normal level of buys and sells in a stock as uninformed trade, and it uses that data to identify the rate of uninformed order flow, ε. Abnormal buy or sell

volume is interpreted as information-based trade, and it is used to identify μ. The number of periods in which there is abnormal buy or sell volume is used to identify α and δ.

A liquidity provider uses his knowledge of these parameters to determine the price at which he is willing to go long, the *Bid*, and the price at which he is willing to go short, the *Ask*. These prices differ, and so there is a Bid-Ask Spread, because the liquidity provider does not know whether the counterparty to his trade is informed or not. This spread is the difference in the expected value of the asset conditional on someone buying from the liquidity provider and the expected value of the asset conditional on someone selling to the liquidity provider. These conditional expectations differ because of the adverse selection problem induced by the possible presence of better informed traders.

As trade progresses, liquidity providers observe trades and are modeled as if they use Bayes rule to update their beliefs about the toxicity of the order flow, which in our model is described by the parameter estimates. Let P(t) = (P_n(t), P_b(t), P_g(t)) be a liquidity provider's belief about the events "no news" (n), "bad news" (b), and "good news" (g) at time t. His belief at time 0 is P(0) = (1-α, αδ, α(1-δ)).

To determine the Bid or Ask at time *t*, the liquidity provider updates his beliefs conditional on the arrival of an order of the relevant type. The time *t* expected value of the asset, conditional on the history of trade prior to time *t*, is

$$E[S_i|t] = P_n(t)S_i^* + P_b(t)\underline{S_i} + P_g(t)\overline{S}_i \qquad (1)$$

where $S_i^* = \delta\underline{S_i} + (1-\delta)\overline{S}_i$ is the prior expected value of the asset.

The Bid is the expected value of the asset conditional on someone wanting to sell the asset to a liquidity provider. So it is

$$B(t) = E[S_i|t] - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}\left(E[S_i|t] - \underline{S_i}\right) \tag{2}$$

Similarly, the Ask is the expected value of the asset conditional on someone wanting to buy the asset from a liquidity provider. So it is

$$A(t) = E[S_i|t] + \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}\left(\overline{S_i} - E[S_i|t]\right) \tag{3}$$

These equations demonstrate the explicit role played by arrivals of informed and uninformed traders in affecting quotes. If there are no informed traders (μ = 0), then trade carries no information, and so the Bid and Ask are both equal to the prior expected value of the asset. Alternatively, if there are no uninformed traders (ε=0), then the Bid and Ask are at the minimum and maximum prices, respectively. At these prices no informed traders will trade either, and the market, in effect, shuts down. Generally, both informed and uninformed traders will be in the market, and so the Bid is less than $E[S_i|t]$ and the Ask is greater than$E[S_i|t]$.

The Bid-Ask Spread at time *t* is denoted by Σ(t) = A(t) – B(t). This spread is

$$\Sigma(t) = \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)}\left(\overline{S_i} - E[S_i|t]\right) + \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)}\left(E[S_i|t] - \underline{S_i}\right) \tag{4}$$

The first term in the spread equation is the probability that a buy is an information-based trade times the expected loss to an informed buyer, and the second is a symmetric term for sells. The spread for the initial quotes in the period, Σ, has a particularly simple form in the natural case in which good and bad events are equally likely. That is, if δ = 1-δ then

$$\Sigma = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}\left(\overline{S_i} - \underline{S_i}\right) \tag{5}$$

The key component of this model is the probability that an order is from an informed trader, which is called PIN. It is straightforward to show that the probability that the opening trade in a period is information-based is given by

$$PIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \tag{6}$$

where $\alpha\mu + 2\varepsilon$ is the arrival rate for all orders and $\alpha\mu$ is the arrival rate for information-based orders. PIN is thus a measure of the fraction of orders that arise from informed traders relative to the overall order flow, and the spread equation shows that it is the key determinant of spreads.

These equations illustrate the idea that liquidity providers need to correctly estimate PIN in order to identify the optimal levels at which to enter the market. An unanticipated increase in PIN will result in losses to those liquidity providers who don't adjust their prices.

2. THE VPIN METRIC AND THE ESTIMATION OF PARAMETERS

The standard approach to computing the PIN model uses maximum likelihood to estimate the unobservable parameters $(\alpha, \mu, \delta, \varepsilon)$ driving the stochastic process of trades and then derives PIN from these parameter estimates. In this section, we propose a direct analytic estimation of toxicity that does not require intermediate numerical estimation of non-observable parameters. We update our measure in volume-time in an attempt to match the speed of arrival of new information to the marketplace. This volume-based approach, which we term VPIN, provides a simple metric for measuring order toxicity in a high frequency environment. First, we begin with a discussion of the role of information and time in high frequency trading.

2.1. The Nature of Information and Time

Information in the standard sequential trade model is generally viewed as data that is informative about the future value of the asset. In an equity market setting it is natural to view information as being about future events such as the prospects of the company or the market for its products. In an efficient market, the value of the asset should converge to its full information value as informed traders seek to profit on their information by trading. Because market makers can be long or short the stock, future movements in the value of the asset affect their profitability and so they attempt to infer any underlying new information from the patterns of trade. It is their updated beliefs that are impounded into their bid and ask prices.

In a high frequency world, market makers face the same basic problem, although the horizon under which they operate changes things in interesting ways. A high frequency market maker who anticipates holding the stock for minutes is affected by information that influences its value over that interval. This information may be related to underlying asset fundamentals, but it may also reflect factors related to the nature of trading in the overall market or to the specifics of liquidity demand over a particular interval. For example, in a futures contract, information that induces increased hedging demand for a contract will generally influence the futures price, and so is relevant for a market maker. This broader definition of information means that information events may occur frequently during the day, and they may have varying importance for the magnitude of future price movements. Nonetheless, their existence is still signaled by the nature and timing of trades.

The most important aspect of high frequency modeling is that trades are not equally spaced in terms of time.[6] Trades arrive at irregular frequency, and some trades are more important than others as they reveal differing amounts of information. For example, as Figure 1

shows, trading in E-mini S&P 500 futures (the blue curve and scale on the left side of the graph) and EUR/USD futures (the red curve and scale on the right side of the graph) exhibit a different intraday seasonality. The arrival of new information to the marketplace triggers waves of decisions that translate into volume bursts. Information relevant to different products arrives at different times, thus generating distinct intraday volume seasonalities.

[FIGURE 1 HERE]

In this study, rather than modeling clock-time, we work in volume-time. Easley and O'Hara [1992] developed the idea that the time between trades was correlated with the existence of new information, providing our basis for looking at trade time instead of clock time.[7] It seems reasonable that the more relevant a piece of news is, the more volume it attracts. By drawing a sample every occasion the market exchanges a constant amount of volume, we attempt to mimic the arrival to the market of news of comparable relevance. If a particular piece of news generates twice as much volume as another piece of news, we will draw twice as many observations, thus doubling its weight in the sample.

2.2. Volume Bucketing

In the example above, if we draw one E-mini S&P 500 futures sample every 200,000 traded contracts, we will draw on average about 9 samples per day. On very active days, we draw a large multiple of 9, while inactive days contribute fewer data points. Since the EUR/USD futures contract trades about 1/10 of E-mini S&P 500 futures' daily volume on average, targeting 9 draws per day will require reducing the volume-distance between two observations to about 20,000 contracts. Because of their differing intra-day patterns of trade, by the time we draw our

first E-mini S&P 500 futures observation of the day, we are about to draw our fourth observation of the day for the EUR/USD futures.

To implement this volume dependent sampling, we group sequential trades into *equal volume buckets* of an exogenously defined size V. A volume bucket is a collection of trades with total volume V. If the last trade needed to complete a bucket is for a size greater than required, the excess size is given to the next bucket. We let $\tau=1,\ldots,n$ be the index of equal volume buckets. A detailed algorithm for this volume packaging process is presented in the Appendix. Sampling by volume buckets allows us to divide the trading session into periods of comparable information content over which trade imbalances have a meaningful economic impact on the liquidity providers.

2.3. Buy Volume and Sell Volume Classification

An issue we have not yet addressed is how to distinguish buy volume and sell volume.[8] Recall that signed volume is necessary because of its potential correlation with order toxicity. While the overall level of volume signals the possible presence of new information, the direction of the volume signals its implications for the direction of price changes. Thus, a preponderance of buy (sell) volume would suggest toxicity arising from the presence of good (bad) news.

Microstructure research has relied on tick-based algorithms to sign trades. Trade classification, however, has always been problematic. One problem is that reporting conventions in markets could treat orders differently depending upon whether they were buys or sells. The NYSE, for example, would report only one trade if a large sell block crossed against multiple buy orders on the book, but would report multiple trades if it were a large buy block crossing against multiple sell orders. Similarly, splitting large orders into multiple small orders meant that

trades occurring in short intervals were not in fact independent observations. Aggregating trades on the same side of the market over short intervals into a single observation was the convention empirical researchers used to deal with these problems.

A second difficulty is that signing trades also requires relating the trade price to the prevailing quote. Traders taking the market maker's bid (ask) were presumed to be sellers (buyers), and trades falling in between were signed using a tick-based algorithm. The Lee-Ready [1991] algorithm also suggested using a 5 second delay between the reported quote and trade price to reflect the fact that the mechanism reporting quotes to the tape was not the same as the trade-reporting mechanism. Even in the simpler world of specialist trading, trade classification errors were substantial.

In a high frequency setting, trade classification is much more difficult. In the futures markets we investigate, there is no specialist, and liquidity arises from an electronic order book containing limit orders placed by a variety of traders. In this electronic market, a trader could hit the book at the same level as the last trade or could submit a limit order that improves the last traded price, and the tick rule can assign the wrong side to the trade.[9] Additionally, order splitting is the norm, cancellations of quotes and orders are rampant, and the sheer volume of trades is overwhelming.[10] Using E-mini S&P 500 futures data from May 2010, we found that an average day featured 2,650,391 quote changes due to order additions or cancellations, and 789,676 quote changes due to trades. Because the Best-Bid-or-Offer (BBO) changes several times between trades, many contracts exchanged at the same price in fact occurred against the bid and the offer. In this high frequency world, applying standard algorithms over individual transactions is problematic.

In our analysis, we aggregate trades over short intervals and then use the standardized price change between the beginning and end of the interval to determine the percentage of buy and sell volume.[11] Aggregation mitigates the effects of order splitting, and using the standardized price change allows volume classification in probabilistic terms (which we call *bulk classification*). In this particular paper, we calculate buy and sell volumes (V_τ^B and V_τ^S) using one-minute time bars (we show later that our analysis works equally well with other time aggregations), but the analysis can also be done using volume bars.[12] Let

$$V_\tau^B = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right) \tag{7}$$

$$V_\tau^S = \sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot \left[1 - Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right)\right] = V - V_\tau^B$$

where $t(\tau)$ is the index of the last time bar included in the τ volume bucket, Z is the CDF of the standard normal distribution and $\sigma_{\Delta P}$ is the estimate of the standard derivation of price changes between time bars. Our procedure splits the volume in a time bar equally between buy and sell volume if there is no price change from the beginning to the end of the time bar. Alternatively, if the price increases, the volume is weighted more toward buys than sells and the weighting depends on how large the price change is relative to the distribution of price changes.

A key difference between bulk classification and the Lee-Ready algorithm is that the latter signs volume as either buy *or* sell, whilst the former signs a fraction of the volume as buys *and* the remainder as sells.[13] In other words, the Lee-Ready algorithm provides a discrete classification, while the bulk algorithm is continuous. This means that even in the extreme case that a single time bar fills a volume bucket, volume may still be perfectly balanced according to bulk classification (contingent on $\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}$).

Our primary use of volume is to compute order imbalance. Let $OI_\tau = |V_\tau^B - V_\tau^S|$ be the order imbalance in volume bucket τ. Our measure is, of course, an approximation to actual order imbalance as it is based on our probabilistic volume classification. We first ask how $E[OI_\tau]$ relates to the rate of trading by showing that it is unaffected by a simple rescaling of trading. Suppose that each time bar's volume is rescaled by a factor of $\beta > 0$, $V_i^* = \beta V_i$, and that volume imbalance is homogeneously distributed within the bucket.[14] Then the expected number of time bars required to fill a bucket will be inversely proportional to β, $\frac{t(\tau)-t(\tau-1)}{\beta}$. From Eq. (7), this leaves the expected order imbalance, $E[OI_\tau]$, unaltered,

$$E[OI_\tau^*] = E[|V_\tau^{*B} - V_\tau^{*S}|] = \frac{1}{\beta} E[|\beta V_\tau^B - \beta V_\tau^S|] = E[OI_\tau] \tag{8}$$

Second, we ask whether, within reasonable bounds, the amount of time contained in a time bar effects our measure of order imbalance? To determine this, we computed order imbalance for the E-mini S&P 500, for the period January 1, 2008 to August 1, 2011, using time bars ranging from 1 to 240 minutes per time bar. For each specification of a time bar we use 50 volume buckets per day and compute the ratio of order imbalance to the bucket size as measured by the volume in each bucket. We found that the ratio of order imbalance to bucket size (as a function of the average number of time bars per bucket) increases gradually as the number of time bars per bucket declines, never approaching one and eventually levels off as we use (unreasonable) long time bars.[15] Thus, for time bars of reasonable length the amount of time in a time bar is of little consequence in measuring the order imbalance.

This methodology will misclassify some volume. Our goal is not to classify correctly each individual trade (a hopeless exercise in any case), but rather to develop an indicator of overall trade imbalance that is useful for creating a measure of toxicity. We use time bars to

allow time for the market price to adjust to the trade direction information that we attempt to recover through bulk classification. Later in the paper we present evidence that our bulk classification procedure leads to more useful results for the purposes of estimating flow toxicity than those based on the itemized classification of raw transaction data.[16]

2.4. Volume-Synchronized Probability Of Informed Trading (the VPIN Flow Toxicity Metric)

The standard PIN model looks only at the number of buys and sells to infer knowledge about the underlying information structure; there is no explicit role for volume. In the high frequency markets we analyze, the number of trades is problematic. Going back to the theoretical foundation for PIN, what we ultimately want is information about trading intentions that arise from informed or uninformed traders. The link between these trading intentions and transactions data is very noisy as trading intentions may be split into many pieces to minimize market impact, one order may produce many trade executions, and information-based trades may be done in various order forms. For these reasons, we treat each reported trade as if it were an aggregation of trades of unit size (i.e. a trade for five contracts at some price *p* is treated as if it were five trades of one contract each at price *p*). This convention explicitly puts trade intensity into the analysis.

We know from Easley, Engle, O'Hara and Wu [2008] that for each period the expected trade imbalance is $E[|V_\tau^S - V_\tau^B|] \approx \alpha\mu$ and that the expected total number of trades is $E[V_\tau^B + V_\tau^S] = \alpha\mu + 2\varepsilon$. Volume bucketing allows us to estimate this specification very simply. In particular, recall that we divide the trading day into equal-sized volume buckets and treat each volume bucket as equivalent to a period for information arrival. That means that $V_\tau^B + V_\tau^S$ is

constant, and it is equal to *V*, for all τ. We then approximate expected trade imbalance by average trade imbalance over *n* buckets.

From the values computed above, we can write the *Volume-Synchronized Probability of Informed Trading*, the VPIN flow toxicity metric, as[17]

$$VPIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} \approx \frac{\sum_{\tau=1}^{n} |V_\tau^S - V_\tau^B|}{nV} \quad (9)$$

Estimating the VPIN metric requires choosing *V*, the volume in every bucket, and *n*, the number of buckets used to approximate the expected trade imbalance. As an initial specification, we focus on *V* equal to one-fiftieth of the average daily volume. If we then choose *n=50*, we will calculate the VPIN metric over 50 buckets, which on a day of average volume would correspond to finding a daily VPIN. Our results are robust to a wide range of choices of *V* and *n* as we discuss in Section 5.

The VPIN metric is updated after each volume bucket. Thus, when bucket 51 is filled, we drop bucket 1 and calculate the new VPIN based on buckets 2 – 51. We update the VPIN metric in volume-time for two reasons. First, we want the speed at which we update VPIN to mimic the speed at which information arrives to the marketplace. We use volume as a proxy for the arrival of information to accomplish this goal. Second, we would like each update to be based on a comparable amount of information. Volume can be very imbalanced during segments of the trading session with low participation, and in such low-volume segments it seems unlikely that there is new information. So updating the VPIN metric in clock-time would lead to updates based on heterogeneous amounts of information.

As an example, consider the trading of the E-mini S&P 500 futures on May 6th 2010. Volume on this date (remembered for the "flash crash" that took place) was extremely high, so

our procedure produces 137 estimations of the VPIN metric, compared to the average 50 daily estimations. Because our sample length (n) is also 50, the time range used for some estimations of the VPIN metric on May 6th 2010 was only a few hours, compared to the average 24 hours.

[FIGURE 2 HERE]

Figure 2 illustrates the way time ranges become "elastic", contingent on the trade intensity (a proxy for speed of information arrival). At 9:30 am (EST), the data used to compute VPIN covered almost an entire day. But as the New York Stock Exchange opened on May 6th 2010, our algorithm updated the VPIN metric more frequently and based its estimates on a shorter interval of clock-time. By 12:17pm, VPIN was being computed looking back only one-half of a day. Note how reducing the time period covered by the sample did not lead to noisier estimations. On the contrary, the VPIN metric kept changing following a continuous trend. The reason is that time ranges do not contain comparable amounts of information. Instead, it is volume ranges that produce comparable amounts of information per update.[18]

GARCH specifications provide an alternative way to deal with the volatility clustering typical of high frequency data sampled in clock-time. Working in volume-time reduces the impact of volatility clustering, since we produce estimates based on samples of equal volume. Because large price moves are associated with large volumes, sampling by volume can be viewed as a proxy for sampling by volatility. The result is a collection of observations whose distribution is closer to normal and is less heteroskedastic than it would be if we sampled uniformly in clock-time. Thus, working in volume-time can be viewed as a simple alternative to employing a GARCH specification.

To see how this transformation allows a partial recovery of normality, we consider an E-mini S&P 500 futures 1-minute bars sample from January 1, 2008 to August 1, 2011. We draw

an average of 50 price observations a day, equally spaced by time in the first case (chronological time), and equally spaced by volume in the second (volume time). Next, we compute first order differences and standardize each sample. Both samples are negatively skewed and have fat tails, but the volume-time sample is much closer to normal, exhibits less serial correlation, and is less heteroskedastic. This becomes more obvious as the sampling frequency increases (compare tables for 50 and 100 draws per day). Table 1 gives the resulting statistics and Figure 3 provides a graphical illustration of the normalized price changes.

[TABLE 1 HERE]

[FIGURE 3 HERE]

3. THE STABILITY OF VPIN ESTIMATES

Estimating the VPIN volatility metric involves a variety of specification issues. In this section, we show robustness of the VPIN measure to two of the most important of these issues, namely, alternative volume classification schemes and changes in the transaction record. Our calculations are based on the E-mini S&P 500 futures series of 1-minute bars, from January 1st 2008 to August 15th 2011, for a bucket size consistent with an average of 50 volume buckets per day, and a sample length of 50 buckets.

3.1. Stability Under Different Volume Classification Schemes

The choice of how to classify volume has an important effect on the estimate of VPIN. In particular, because VPIN involves looking at trade imbalance and intensity, aggregating over time bars would be expected to reduce the noise in this variable as well as to rescale it. In Easley, López de Prado, and O'Hara [2011a], we argued that this necessitates looking at the

relative levels of VPIN as captured by their cumulative distribution function rather than at absolute levels of VPIN.[19]

This point can be illustrated by looking at the behavior of VPIN on a given day for different volume classification algorithms. The "flash crash" on May 6 is of particular importance in futures (and equity) markets and so we illustrate the behavior of VPIN on this day using three alternative volume classification schemes. Figure 4(a) shows VPIN calculated using bulk classification of 1-minute time bars; 4(b) uses bulk classification of 10-second bars; and 4(c) uses Lee-Ready trade-by-trade classification.

[FIGURE 4 (a), (b), (c) HERE]

The three methods concur in signaling an extreme level for the VPIN flow toxicity metric at least two hours before the crash (see the CDF(VPIN) dashed line crossing the 0.9 threshold).[20] The 1-minute and 10-second time bars produce qualitatively similar stories about the relative level of VPIN. In both cases, it rises before the crash and stays high throughout the rest of the day. The trade-by-trade classification results are different. Here the estimated VPIN increases before the crash, although not as dramatically as it does with time bars, but it falls to unusually low levels immediately after the crash and it remains low for the rest of the day. This inertia, however, is over a period in which the market rose by more than 4%. It seems highly unlikely that volume was, in fact, balanced over this period. We suspect that this is more a result of trade misclassification than anything else, and so we do not use trade-by-trade classification. Instead, we report results for 1-minute time bars while noting that the relative results with 10-second time bars are very similar.

3.2. Stability to Changes in the Transaction Record

A second stability test concerns the impact that changes in the trading record have on the VPIN estimate. Small discrepancies in the trading record, such as missing trades, produce two effects. First, missing trades could alter the volume imbalance. But since we use an amount of volume equivalent to an entire trading session, this impact is expected to be negligible (the typical trade is for a few contracts, compared to the daily average of more than two million contracts traded on E-mini S&P 500 futures in 2010). The second effect of missing trades comes in the form of a shift in the VPIN trajectory. This impact can be evaluated by shifting the starting point of VPIN trajectories and calculating the cross-sectional standard deviations over time.

To assess the influence of different starting times, we compute 1,000 alternative VPIN trajectories for the E-mini S&P 500 futures, each starting one time bar after the previous one. We then take those 1,000 VPIN trajectories and align them for each time bar. Because VPIN is computed at the completion of each bucket, and buckets are not completed simultaneously, to each time bar we assign the last computed VPIN. We can then estimate a cross-sectional standard deviation on the difference between the first trajectory and the following 999. This estimate is likely to be greater than the true value of the standard deviation as a result of the asynchronicity in the completion of buckets.

[FIGURE 5 HERE]

Figure 5 shows that the cross-sectional standard deviation of differences on the 1,000 VPIN trajectories is negligible. That time series has a mean of 0.015 while VPIN's average value is much greater (around 0.23). But as we argued earlier, the fact that buckets for each of the 1,000 trajectories are not completed simultaneously means that the true cross-sectional standard deviation is even smaller than this 0.015 average value. The spikes in cross-sectional standard

deviations that are apparent in Figure 5 coincide with spikes in VPIN values. For example, on May 6th 2010 the cross-sectional standard deviation was just above 0.02, on a day when VPIN reached a level close to 0.5.

In conclusion, differences in the trading record due to missing transactions or alternative starting points do not seem to significantly impact VPIN estimates. To see this final point, consider two estimates of VPIN, one of 0.45, the other 0.5. This difference is well beyond what we see in Figure 5. Although the difference between these two VPIN estimates may seem large, they are not significantly different as the two VPIN estimates are at approximately the same point on the CDF of VPIN (0.991 compared to 0.995).

4. ESTIMATING THE VPIN METRIC ON FUTURES

Having established the robustness of our estimation procedures for the VPIN toxicity metric, we now illustrate its application to two of the most actively traded futures contracts: the E-mini S&P 500 (trading on the CME) and the WTI crude oil future (trading on the NYMEX). Our sample period is January 1st 2008 to June 6th 2011, using at each point in time the expiration with highest daily volume, rolled forward. We use a bucket size equal to 1/50th of the average daily volume in our sample period (V). Parameters are estimated on a rolling window of sample size $n = 50$ (equivalent to 1-day of volume on average). We use the entire sample period to determine the cumulative distribution function of the estimated parameters. Table 2 provides basic statistics of the VPIN metric estimates for these contracts.

[TABLE 2 HERE]

A natural concern arising from our estimates of VPIN is that its AR(1) coefficients are very close to 1. This might suggest that VPIN has a unit root, with its consequent implications that VPIN is unstable and so its CDF over any sample period would not be useful in evaluating the likelihood of future values. Fortunately, this is not the case. We have already shown that VPIN is not affected by the shifts of the starting date of our sample considered in Figure 5. Table 3 also shows that the CDFs of VPIN obtained by splitting our sample into a before and after Flash Crash (May 6, 2011) date are nearly identical. Thus, the VPIN is, in fact, highly stable. The high AR(1) coefficient and the stability of VPIN occur for the same reason. Our VPIN measure is computed using 50 buckets. When it is updated, the first of the existing 50 buckets is dropped and the latest one is added. This averaging makes VPIN highly auto-correlated, but also insures that the process does not have a long memory as at each point the current value of VPIN cannot depend on the value that VPIN took on 51 buckets earlier.[21]

[TABLE 3 HERE]

4.1. S&P 500 (CME)

Figure 6 shows the evolution of the E-mini S&P 500 futures contract (red line, expressed in terms of market value) and its VPIN metric value (green line). The VPIN metric is generally stable, although it clearly exhibits substantial volatility. We note that the VPIN metric reached its highest level for this sample on May 6th 2010, the day of the flash crash. Such high levels of the VPIN metric are consistent with the largest part of the order flow being one-sided for the equivalent of one day of transactions. As discussed in Easley, López de Prado and O'Hara [2011a], this excessive toxicity led some market makers to become liquidity consumers rather than liquidity providers as they shut down their operations.[22]

[FIGURE 6 HERE]

A more recent episode of extreme toxicity occurred in the aftermath of the Japanese earthquake. Although the major Tohoku earthquake and tsunami took place in the early morning of March 11th 2011, the S&P 500 did not experience a large move until the subsequent Fukushima nuclear crisis unfolded on March 14th 2011. That day the S&P 500 registered another extreme level of order flow toxicity. Unlike on May 6th 2010, the March 14th 2011 crash occurred with light volume, during the night session (from 6pm to 11pm EST). After only 287,360 contracts had been traded, the index had lost approximately 2.5% of its value. Figure 7 shows that CDF(VPIN) was at its 0.97 threshold as early as 3pm, illustrating that flow toxicity also occurs in instances of reduced trade intensity.

[FIGURE 7 HERE]

4.2. WTI Crude Oil (NYMEX)

Crude oil is the most heavily traded commodity, and its strategic role in the world economy makes it ideal for placing geopolitical and macroeconomic wagers. Energy futures are also a venue in which market makers face extreme volatility in order flows. As shown in Figure 8, the highest flow toxicity reading for this contract occurred on May 6th 2010. Such behavior is consistent with the fact that while the problems on May 6th were not energy related, oil futures were affected by the contagion of liquidity and toxicity across markets. Other than the day of the flash crash, the next highest toxicity levels for this contract occurred on May 5th 2011.

[FIGURE 8 HERE]

In early May 2011, the CFTC reported the largest long speculative position among crude traders in history.[23] The New York Times attributed these large positions to traders believing that

energy prices would ramp up, fueled by the violence sweeping through North Africa and the Middle East.[24] Some of these traders decided to take profits on May 5th 2011.[25] The unwinding of their massive positions led them to seek liquidity, but as market makers realized that the selling pressure was persistent, they started to withdraw, which in turn increased the concentration of toxic flow in the overall volume. Figure 9 shows that by 9:53am CDF(VPIN) crossed the 0.9 threshold, remaining there for the rest of the day. During those few hours, WTI lost over 8%.[26]

[FIGURE 9 HERE]

5. TOXICITY AND FUTURE PRICE MOVEMENTS

In a high frequency market, market makers can use the VPIN metric derived and estimated above to measure the toxicity of order flow. Because toxicity affects market makers' profits, toxicity should also affect market maker behavior, and by extension liquidity in the market. In this section, we address in more detail the linkage between toxicity and future price movements.

As noted earlier in the paper, time is not a particularly meaningful concept to a high frequency market maker. Since market makers are passive traders who must wait for the order flow to come to them, it is volume rather than time that is the operative metric. This same volume metric is also relevant for considering the future linkage of toxicity and price movements. A market maker needs to know how toxicity will influence price behavior while he or she is holding a position.[27] Market makers seek to turn over their positions multiple times a day, but how frequently they are able to do so depends on the volume of trade. Thus, for the market maker, two questions are relevant. First, how does high toxicity affect price behavior

over this holding period? And, second, how does the persistence of high toxicity affect price behavior?

While these questions are straightforward, answering them is not. One difficulty is that standard microstructure models are not well-suited for capturing behavior in the new high frequency world. We simply do not have models of multiple, competing market makers who face information and inventory constraints, and who manage risk by moving across and between (or even completely out of) markets in microseconds. Thus, theory does not provide the exact linkage between toxicity, liquidity, and volatility that we seek. A second difficulty is that the econometrics of analyzing liquidity and volatility in such a world are embryonic, reflecting the many distinctive features of high frequency data already noted throughout this paper.

To address these questions, therefore, we draw on basic relationships to examine the linkages between toxicity, liquidity and volatility. We first look at the relationship between toxicity and price movements defined over the subsequent volume bucket, which we argue is the relevant interval from the perspective of the market maker. We then consider how the persistence of toxicity influences return behavior over longer intervals. In general, we know that as toxicity increases, market makers face potential losses and so may opt to reduce, or even abandon market making activities. This decrease in liquidity, in turn, suggests that high levels of VPIN should presage greater price variability.[28]

5.1. Correlation Surface

We begin by asking a simple question: Is the VPIN volatility metric correlated with future price movements? To measure this relationship, we use Pearson's correlation between the natural logarithm of VPIN and the absolute price return over the following bucket,

$\rho\left(Ln(VPIN_{\tau-1}), \left|\frac{P_\tau}{P_{\tau-1}} - 1\right|\right)$, where τ indexes volume buckets. Because VPINs can be estimated using various combinations of the number of volume buckets per day and the sample length, we examine how these estimation parameters affect the VPIN metric's relationship with future price movements.

For E-mini S&P 500 futures, VPINs are positively correlated with future price volatility. This relationship is demonstrated in Figure 10 where each point on the graph was computed using large samples, some of over 44,000 observations. The figure shows that the correlation between VPIN and the following absolute future returns varies smoothly across different parameter values. In general, increasing the sample length increases the correlation as does, to a lesser degree, increasing the numbers of buckets per day.

[FIGURE 10 HERE]

The (*50,250)* combination seems a reasonably good pair for E-mini S&P 500, and it has a simple interpretation as "one week" of data (50 volume buckets per day and 5 trading days per week). For this combination, we obtain a correlation $\rho\left(Ln(VPIN_{\tau-1}), \left|\frac{P_\tau}{P_{\tau-1}}\right|\right) = 0.400$ on 44537 observations.[29] Figure 10 suggests that there is little advantage in "over-fitting" these parameters, as there is a wide region of parameter combinations yielding similar predictive power.[30]

[FIGURE 11 HERE]

Figure 11 provides a plot of the return on the E-mini S&P 500 futures over the next volume bucket (1/50 of an average day's volume using the (*50,250)* combination) sorted by the previous VPIN level. The graph spreads out vertically as VPIN rises, illustrating that higher toxicity levels, as measured by higher VPIN, lead to greater absolute returns.

These findings in terms of correlation are suggestive, but simple correlation is a narrow criterion of dependency. In fact, VPIN exhibits significant serial correlation, which makes

drawing conclusions from a simple correlation problematic. One alternative would be to estimate a time series model of the joint VPIN-returns process. We do not pursue this approach here as our hypothesis is not that high VPIN levels lead to high absolute returns in a particular time period. Instead, our hypothesis is that persistently high VPIN levels have implications for market maker behavior, which in turn have implications for absolute returns measured using a volume clock. To shed light on this hypothesis, we employ a model-free framework based on conditional probabilities. We then ask two fundamental questions: (1) When VPIN is high, what is the subsequent behavior of absolute returns? (2) When absolute returns are high, what was the preceding level of VPIN?

5.2. Conditional Probabilities

To display these conditional probabilities we need to compute the joint distribution of VPIN and absolute returns. We do this by first grouping VPINs in 5%-tiles and absolute returns in bins of size 0.25% so that we can display discrete distributions. We then compute the joint distribution of $\left(VPIN_{\tau-1}, \left|\frac{P_\tau}{P_{\tau-1}} - 1\right|\right)$. From this joint distribution we derive two conditional probability distributions.

We first examine *the distribution of absolute returns* over the subsequent volume bucket conditional on VPIN being in each of our twenty 5%-tile bins. This results in twenty conditional distributions, one for each VPIN-bin, which are displayed in Table 4a. Each row of this table represents a conditional distribution of absolute returns, conditioned on the prior level of VPIN.

[TABLE 4(a), 4(b) HERE]

There are three important results to note. First, when VPIN is low, subsequent absolute returns are also low. In particular, when VPIN is in its bottom quartile, subsequent absolute

returns are in the 0% to 0.25% range more than 90% of time.. Second, when VPIN is high, the conditional distribution of subsequent returns is much more dispersed. In particular, high absolute returns (above 1.5%) sometimes occur during the subsequent volume bucket, while they never occur following very low VPINs. Third, even for high levels of VPIN, absolute returns over the next volume bucket are most often not large. We elaborate on this point in the next subsection where we argue that it takes persistently high levels of VPIN to reliably generate large absolute returns.

Next, in Table 4b we examine the distribution of *VPIN* in bucket τ-1 conditional on absolute returns between buckets τ-1 and τ. Each column of this table provides the distribution of prior VPINs conditional on absolute returns in each bin of size 0.25%. The important result here is that when absolute returns are large the immediately preceding VPIN was rarely small. In particular, the upper quartile of this distribution contains over 84% of all absolute returns greater than 0.75%. This fact suggests that VPIN has some insurance value against extreme price volatility.To compute these conditional probabilities, we used our standard *(50,250)* parameter combination. This choice of parameters maximized the correlation between VPIN and absolute returns, but it does not necessarily maximize a particular "risk scenario", say for example, $Prob\left(\mathrm{CDF}(VPIN_{\tau-1}) > \frac{3}{4} \middle| \left|\frac{P_\tau}{P_{\tau-1}} - 1\right| > 0.75\%\right)$. The following figure shows the effect of parameter choices on VPIN's insurance value against absolute returns greater than 0.75%.

[FIGURE 12 HERE]

Figure 12 shows that as long as the number of buckets per day is not extremely large, and the sample length is not extremely small, the probability that VPIN was in the upper quartile one volume bucket prior to a 0.75% or greater absolute return will be between 80% and 90%. This

result not only means that VPIN anticipates a large proportion of extreme volatility events, but also that toxicity-induced volatility seems to be a significant source of overall volatility.

5.3. Is Extreme Volatility Always Preceded by High VPIN?

Table 4a shows that for the E-mini S&P 500 futures contract, large absolute returns, greater than say 2%, are very unlikely if the preceding VPIN was low. But they are possible and searching across contracts certainly provides examples in which they occur. A recent example of extreme price volatility in the natural gas futures vividly illustrates this possibility. According to The Financial Times[31], on June 8th 2011: *"The New York Mercantile Exchange floor had been closed for more than five hours when late on Wednesday Nymex July natural gas dropped 39 cents, or 8.1 per cent, to $4.510 per million British thermal units. After a few seconds, it bounced back up"*. An explanation was offered in The Financial Times on June 9th 2011: *"Some market watchers attributed the decline to a 'fat finger' error, when a trader mistakenly types an extra zero on the end of an order, increasing its size by a factor of 10. Others blamed it on a glitch in computer algorithms that trade futures. Volume was light, meaning any big order would have had an outsize impact and potentially triggered automated selling"*.

If the FT's explanation is correct, then VPIN should not have been high before the price decline. Figure 13 shows that this is what occurred. There was a sudden decline in prices followed by an immediate recovery, all of which occurred at relatively low toxicity levels. This example illustrates two useful points: Not all volatility is due to toxicity; and, it may be helpful to regulators to know when a price drop is due to toxicity or arises from other potential causes.

[FIGURE 13 HERE]

5.4. Does Extreme Volatility Always Occur Once VPIN is High?

We know from Table 4b that large absolute returns do not necessarily occur in the next volume bucket when VPIN is large. In fact, most of the absolute returns immediately following a high VPIN observation are not large. This observation, however, is consistent with our hypothesis that persistently high levels of VPIN lead to volatility. To examine our hypothesis, we need to quantify the maximum amount of volatility that a market maker is exposed to once VPIN reaches some critical level and stays at or above that critical level. The flash crash provides a stark illustration of why the distinction between immediate volatility and ensuing volatility is important. During the flash crash, VPIN remained high from before the crash began until well after it ended and prices began their partial recovery. If we focus on the time period beginning when VPIN reached some high level and ending when it fell, there is only a small price decline. This fact would be of little comfort to the many market makers who left the market during the crash. They were affected by the large intermediate volatility that occurred while VPIN was high.

To examine the maximum intermediate volatility experienced by a market maker while VPIN is high, we compute volatility events while VPIN remains within any 5%-tile. In particular, every time VPIN moves from one 5%-tile to another, we compute the largest absolute return that occurs between any two intermediate (not necessarily consecutive) buckets, until VPIN moves to another 5%-tile. For example, suppose that the VPIN metric moves into the 85^{th} percentile and 4 volume buckets later it moves to the 90^{th} percentile. We would calculate the absolute return between buckets 1 and 2, 1 and 3, 1 and 4, as well as between 2 and 3, 2 and 4, and 3 and 4. We are interested in the maximum of these absolute returns, which we view as the maximum volatility the market maker faces. We do this for every such VPIN crossing.

To be precise, we let *i* be an index that is updated by one every time that VPIN crosses from one percentile into another, and $\tau(i)$ the bucket associated with the *i*th cross. This means that VPIN remained in the same percentile for $\tau(i+1) - \tau(i)$ buckets; in our example above this was the 4 bucket interval over which VPIN remained in the 85th percentile. The largest absolute return between any two intermediate buckets while VPIN remained in a particular percentile (i.e., from the bucket at which it made the *i*th crossing until it made the *i+1*th crossing) is then

$$\max_{\substack{\tau(i)\leq m<l \\ \tau(i)<l\leq\tau(i+1)}} \left|\frac{P_l}{P_m} - 1\right| \tag{10}$$

In our example above this is the maximum of the six intermediate returns over the interval in which VPIN remained in the 85th percentile.

This analysis is richer than the conditional probabilities illustrated in Table 4a in two respects. First, we are not just considering the absolute return that occurs in the bucket immediately following VPIN's crossing into a bin, but rather the maximum absolute return that occurs while VPIN remains in a bin. This is consistent with our microstructure theory that volatility appears once toxicity has reached a saturation point that exceeds market makers' tolerance. Second, it captures price volatility across all sequences of intermediate buckets, thus incorporating the effect of price drifts as well as price recoveries. This is important, because slow price drifts and price recoveries may both hide extreme volatility. But this analysis is also limited in that it does not capture sustained increases in toxicity spanning multiple VPIN percentiles. In particular, we do not capture what happens when toxicity increases from the 75th to the 80th and then on to the 85th, 90th, etc. Thus, the hurdle we set here will surely underestimate the effect of toxicity-induced volatility.

[FIGURE 14 HERE]

Figure 14 plots the probabilities of the largest absolute return being in excess of 0.75% while VPIN remains in any 5%-tile in the upper quartile of its distribution (i.e. VPIN is in the 0.75-0.80, 0.80-0.85, 0.85-0.90, 0.90-0.95, or 0.95-1.00 bin) for various combinations of buckets per day and sample length. For our standard combination of *(50,250)*, we find that 51.84% of the times that VPIN enters a 5%-tile within the upper quartile there is at least one intermediate return in excess of 0.75% before VPIN leaves that 5%-tile. The parameter combination that maximizes VPIN's predictive power is *(10,350)*, i.e. 10 buckets per day for a sample length of 350 (about 1.6 months). For this parameter combination, Table 5 shows that 78.57% of the times that VPIN enters a 5%-tile within the upper quartile of its distribution, toxicity-induced volatility often will be substantial (absolute returns in excess of 0.75%) before VPIN leaves that 5%-tile.

[TABLE 5 HERE]

6. CONCLUSIONS

This paper presents a new procedure to estimate the *Volume-Synchronized Probability of Informed Trading*, or the *VPIN* flow toxicity metric. An important advantage of the VPIN toxicity metric and its associated estimation procedure is that it is updated intraday with a frequency attuned to volume in order to match the speed of information arrival. It is also a direct analytic estimation procedure that does not rely on intermediate estimation of unobservable parameters, or numerical methods. We have shown that the VPIN metric has significant forecasting power over toxicity-induced volatility, and that it offers insurance value against future high absolute returns.

It is this latter property that we believe makes VPIN a risk management tool for the new world of high frequency trading. Liquidity provision is now a complex process, and levels of toxicity affect both the scale and scope of market makers' activities. High levels of VPIN signify a high risk of subsequent large price movements, deriving from the effects of toxicity on liquidity provision. This liquidity-based risk is important for market makers who directly bear the effects of toxicity, but it is also significant for traders who face the prospect of toxicity-induced large price movements.[32] Developing algorithms that would vary the execution pattern of orders depending upon toxicity would allow traders to mitigate this risk.[33] Exchanges could also apply VPIN to provision machine resources in a way that speeds up trading on the side with greater liquidity while slowing down trading on the side under attack (a sort of dynamic circuit-breaker), thus allowing market makers to remain active. We believe this is an important area for future research.

FOOTNOTES

[1] See Clark [1973] or Ané and Geman [2000] for a primer on subordinated stochastic processes.

[2] See Tauchen and Pitts [1983], DeGennaro and Shrieves [1995], Jones et al. [1994].

[3] Kirilenko, Kyle, Samadi and Tuzun [2010] give empirical evidence on market maker behavior during the "flash crash".

[4]Bethel, Leinweber, Rubel and Wu [2011] discuss the use of the VPIN metric to monitor liquidity in equity markets.

[5] The literature has more complex models of the arrival process, but to illustrate our ideas we stay with the simplest model. The simple model has an advantage over more complex models in that it yields a simple expression for the probability of information-based trade which is easy to compute. In spite of its simplicity and its obvious abstraction from the reality of the trading process, this expression has proven useful in a variety of settings.

[6] Mandelbrot and Taylor [1967] noted that this was true of equity trading even in the 1960s.

[7] A variety of authors have further developed this notion of time as an important characteristic of trading. Of particular importance, Engle [1996] and Engle and Russell [2005] develop the role of time in a new class of autoregressive-conditional duration (ACD) models.

[8] Publically available data generally do not distinguish buys and sells, so an algorithm is necessary to infer buys and sells. The algorithm we use seems to work well, but the algorithm itself is independent of the VPIN metric. Any algorithm (or even better data about buys and sells) could be used to provide input to the estimation of VPIN.

[9] Informed and uninformed trading in an electronic limit order market has been examined by numerous authors, for example Bloomfield, O'Hara and Saar [2005], and Foucault, Kadan and Kandel [2009].

[10] Over the last three years of E-mini S&P 500 futures data, in an average 10-minute period there were 2,200 trades encompassing 21,000 contracts traded.

[11] We thank Mark Ready for helpful advice on trade classification in high frequency markets.

[12] We focus on time bars because data vendors (such as Bloomberg) provide such information and so it is a more familiar concept to market practitioners. However, both conventions (Time and Volume bars) have their own merits, and we hope to investigate these alternatives further in future work.

[13] Readers familiar with *unsupervised machine learning* techniques will not be surprised with the logic behind our bulk classification procedure.

[14] This assumption may of course not reflect the empirical characteristics of the data, but we are at this point simply discussing the properties of OI_τ as they relate to bulk classification.

[15] As the average number of time bars per bucket increases from 1 to 30, order imbalance as a fraction of bucket size decreases from 0.52 to 0.25, with standard deviations ranging from 0.33 to 0.22.

[16] This is hardly a surprising result as it corresponds to the principle that bulk counting can be more useful than item counting under measurement error.

[17] This metric uses an approximation because the arrival rate of information-based trades is approximated by the expected order imbalance. A more accurate estimator is presented in Appendix A.1.3. However, Appendix A.2 shows through Monte Carlo simulations that this simpler expression produces an acceptable estimation error.

[18] At a daily level the relationship between volume and price change (which is a proxy for information arrival) has been explored by many authors including Clark [1973], Tauchen and Pitts [1983], Harris [1986], and Easley and O'Hara [1992].

[19] For predicting toxicity-induced volatility what matters is whether the level of VPIN at any time is unusual relative to its distribution for the asset in question. The actual level of VPIN, which is sensitive to the choice of bucket size, sample length, trade classification, and so on, is not our primary concern. Of course, if making different choices of these parameters does more than rescale VPIN it could affect relative VPINs. The results in this section and those in later estimations strongly suggest that this is not the case.

[20] For the 1-minute and 10-second time bars VPIN continues to increase during the flash crash and it remains high for the remainder of the flash crash day. Thus, according to our interpretation, order flow was highly toxic throughout the flash crash and it remained toxic during the price recovery following the crash. This is consistent with the very high price volatility observed during and after the crash---prices first fell and then rose. VPIN is not a directional indicator; it only indicates toxicity-induced volatility without predicting the sign of the price changes.

[21] There are, however, two means by which dependence can enter into VPIN. First, the exact timing of buckets does depend on the entire sample. Our results on variations in the starting date of the sample show that this timing issue is not important over our fairly long sample. Second, VPIN is based on order imbalance which is auto-correlated. But the auto-correlation of order imbalance for the E-mini S&P 500 contract over our sample is only 0.2146.

[22] A video of this event can be found at http://youtu.be/IngpJ18AhWU

[23] *"Crude oil traders trim bets on price rise, CFTC data shows"*. Bloomberg News, May 6th 2011.

[24] *"Price of crude oil falls again, but analysts warn it will remain at lofty levels"*. The New York Times, May 7th 2011.

[25] *"Nervy investors dump commodities"*. Financial Times, May 7th 2011.

[26] A video of this event can be found at http://youtu.be/ifW-apeHeI0

[27] Of course, market makers affect prices through their own trading. Our analysis best applies to markets in which there are many participants, which is clearly the case for the E-mini S&P 500 futures contract.

[28] Deuskar and Johnson [2011] also look at the effects of order imbalance (which they term flow-driven risk) on market returns and volatility. Their analysis uses chronological time and so it is not directly comparable to what we do here. They find significant flow-driven effects on returns but note that their measure does not appear to be consistent with measuring the degree of asymmetric information. We believe this dimension is better captured by our volume-based analysis which incorporates the arrival of new information. However, both their model and ours provide strong evidence that order imbalance has important implications for market liquidity.

[29] Statisticians often prefer to look at correlation using the Fisher transform, which for our analysis is given by $\sigma_{arctanh(\rho)} = 0.004742$, and 95% confidence bands given by $\rho\left(Ln(VPIN_{\tau-1}), \left|\frac{P_\tau}{P_{\tau-1}}\right|\right) \in [0.392, 0.408]$. More details regarding the statistical properties of correlation coefficients can be found in Fisher (1915).

[30] In the context of high frequency trading, such correlation is very significant. According to the Fundamental Law of Active Management, $IR = IC\sqrt{BR}$, where IR represents the Information Ratio, IC the Information Coefficient (correlation between forecasts and realizations), and BR the breadth (independent bets per year). Although a correlation of 40% may seem relatively small, the breadth of high frequency models is large, allowing these algorithms to achieve high Information Ratios. For example, an IR of 2 can be reached through a monthly model with IC of

0.58, or a weekly model with IC of 0.28, or a daily model with IC of 0.13. High frequency models produce more than one independent bet per day, thus a correlation of over 0.4 is very significant.

[31] Reported by Gregory Meyer at the *Financial Times* of June 9th 2011.

[32] We stress that, in our view, VPIN is not a substitute for VIX, but rather a complementary metric for addressing a different risk. VIX captures the markets' expectation of future volatility, and hence is useful for hedging the effects of risk on a portfolio's return. VPIN captures the level of toxicity affecting liquidity provision, which in turn affects future short-run volatility when this toxicity becomes unusually high. For more discussion, see Easley, López de Prado and O'Hara [2011b].

[33] The Waddell and Reed trader who submitted a large sell order in S&P 500 futures precipitating the flash crash would surely have been well advised to avoid trading in a market that was exhibiting record high levels of toxicity.

APPENDICES

A.1. ALGORITHM FOR COMPUTING THE VPIN METRIC[1]

In this appendix, we describe the procedure to calculate *Volume-Synchronized Probability of Informed Trading*, a measure we called the *VPIN* flow toxicity metric. Similar results can be reached with more computationally efficient algorithms, such as re-using data from previous iterations, performing fewer steps or in a different order, etc. The algorithm described below is illustrative of the general idea.

One feature of this algorithm is that we apply a probabilistic approach to classify the volume exchanged within each 1-minute time bars. We cannot expect users to have data which unequivocally identifies trades as buyer-initiated or seller-initiated so some classification procedure is necessary. One could classify each trade separately or one could classify trades in aggregates of an alternative size (based on either time, number of trades or volume bars). Different schemes will lead to different levels of VPIN. We have used a variety of schemes and conclude that data aggregation leads to better flow toxicity estimates than working on raw transaction data. Data granularity has an impact on the magnitude of VPIN levels, however our focus is on how rare a particular VPIN is relative to the distribution of VPINs derived from any classification scheme, and this is unaffected by the classification schemes we have examined (including trade-by-trade as well as groups based on one-tenth of a bucket). We focus on 1-minute time bars as this data is less noisy, more widely available and easier to work with.

A.1.1. INPUTS

1. Time series of transactions of a particular instrument (T_i, P_i, V_i):
    a. T_i: Time of the trade.
    b. P_i: Price at which securities were exchanged.
    c. V_i: Volume exchanged.
2. V: Volume size (determined by the user of the formula).
3. n: Sample of volume buckets used in the estimation.

P_i, V_i, V, n are all integer values. T_i is any time translation, in integer or double format, sequentially increasing as chronological time passes.

A.1.2. PREPARE EQUAL VOLUME BUCKETS

1. Sort transactions by time ascending: $T_{i+1} \geq T_i, \forall i$.
2. Compute $\Delta P_i, \forall i$.
3. Expand the number of observations by repeating each ΔP_i as many times as V_i. This generates a total of $I = \sum_i V_i$ observations ΔP_i.
4. Re-index ΔP_i observations, $i=1,\ldots,I$.
5. Initiate counter: $\tau = 0$.
6. Add one unit to τ.
7. If $I < \tau V$, jump to step 11 (there are insufficient observations).

[1] Patent pending. U.S. Patent and Trademark Office.

8. $\forall i \in \left[(\tau-1)V+1, \tau V\right]$, split volume between *buy* and *sell initiated*:
   a. $V_\tau^B = \sum_{i=(\tau-1)V+1}^{\tau V} Z\left(\frac{\Delta P_i}{\sigma_{\Delta P}}\right)$
   b. $V_\tau^S = \sum_{i=(\tau-1)V+1}^{\tau V} \left[1 - Z\left(\frac{\Delta P_i}{\sigma_{\Delta P}}\right)\right] = V - V_\tau^B$
9. Assign to variable V_τ^B the number of observations classified as *buy* in step 8, and the variable V_τ^S the number of observations classified as *sell*. Note that $V = V_\tau^B + V_\tau^S$.
10. Loop to step 6.
11. Set $L = \tau - 1$ (last bucket is always incomplete or empty, thus it will not be used).

A.1.3. APPLY VPIN's FORMULA

If $L \geq n$, there is enough information to compute $VPIN_L = \frac{\sum_{\tau=L-n+1}^{L} |V_\tau^S - V_\tau^B|}{\sum_{\tau=L-n+1}^{L} (V_\tau^S + V_\tau^B)} = \frac{\sum_{\tau=L-n+1}^{L} |V_\tau^S - V_\tau^B|}{nV}$.

Alternatively, the following formula could be used:

$VPIN_L = \frac{E[|V_\tau^S - V_\tau^B|]}{V}$, where:

- $E\left[\left|V_\tau^S - V_\tau^B\right|\right] = \sigma\sqrt{\frac{2}{\pi}} e^{\left(\frac{-\left(E\left[V_\tau^S - V_\tau^B\right]\right)^2}{2\sigma^2}\right)} + E\left[V_\tau^S - V_\tau^B\right]\left[1 - 2Z\left(-\frac{E\left[V_\tau^S - V_\tau^B\right]}{\sigma}\right)\right]$.
- $E\left[V_\tau^S - V_\tau^B\right] \approx \frac{1}{n}\sum_{\tau=L-n+1}^{L}\left(V_\tau^S - V_\tau^B\right)$.
- $\sigma^2 = E\left[V_\tau^S - V_\tau^B - E\left[V_\tau^S - V_\tau^B\right]\right]^2 \approx \frac{1}{n}\sum_{\tau=L-n+1}^{L}\left(V_\tau^S - V_\tau^B - \frac{1}{n}\sum_{\tau=L-n+1}^{L}\left(V_\tau^S - V_\tau^B\right)\right)^2$.
- $Z(\cdot)$ is the cumulative standard normal distribution.

A.2. MONTE CARLO VALIDATION

Although three parameters appear in the formulation of the VPIN metric, only two need to be simulated. This observation follows from the facts that $\varepsilon = \frac{V - \alpha\mu}{2}$ and that in our methodology $V = \frac{1}{n}\sum_{\tau=1}^{n}(V_\tau^B + V_\tau^S)$ is constant. Therefore we can measure the accuracy of our VPIN metric estimate by running a bivariate Monte Carlo on only the parameters (α, μ).

MEAN ESTIMATION ERROR

(α,μ)	0	50	100	150	200	250	300	350	400	450	500
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1	0.0000	-0.0001	0.0001	-0.0001	0.0009	0.0008	0.0011	0.0007	0.0016	0.0024	-0.0012
0.2	0.0000	-0.0002	0.0002	0.0001	0.0008	0.0009	0.0006	0.0012	0.0022	0.0025	-0.0002
0.3	0.0000	-0.0001	0.0000	0.0001	0.0004	0.0015	0.0010	0.0018	0.0034	0.0035	0.0023
0.4	0.0000	0.0001	0.0003	-0.0005	0.0009	0.0000	0.0021	0.0013	0.0032	0.0034	0.0035
0.5	0.0000	0.0000	0.0002	0.0002	0.0006	0.0024	0.0007	0.0010	0.0022	0.0026	0.0028
0.6	0.0000	0.0003	-0.0002	0.0003	0.0009	0.0009	0.0012	-0.0004	0.0005	0.0025	0.0025
0.7	0.0000	0.0000	0.0000	0.0005	0.0006	-0.0003	0.0008	0.0004	0.0012	0.0002	0.0013
0.8	0.0000	0.0002	0.0003	0.0000	0.0002	0.0006	0.0007	0.0013	0.0012	0.0009	0.0006
0.9	0.0000	0.0001	-0.0001	-0.0003	0.0001	0.0017	0.0001	-0.0007	0.0001	-0.0011	0.0001
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0001	0.0000	-0.0001	0.0000

TRUE VALUE

(α,μ)	0	50	100	150	200	250	300	350	400	450	500
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1	0.0000	0.0100	0.0200	0.0300	0.0400	0.0500	0.0600	0.0700	0.0800	0.0900	0.1000
0.2	0.0000	0.0200	0.0400	0.0600	0.0800	0.1000	0.1200	0.1400	0.1600	0.1800	0.2000
0.3	0.0000	0.0300	0.0600	0.0900	0.1200	0.1500	0.1800	0.2100	0.2400	0.2700	0.3000
0.4	0.0000	0.0400	0.0800	0.1200	0.1600	0.2000	0.2400	0.2800	0.3200	0.3600	0.4000
0.5	0.0000	0.0500	0.1000	0.1500	0.2000	0.2500	0.3000	0.3500	0.4000	0.4500	0.5000
0.6	0.0000	0.0600	0.1200	0.1800	0.2400	0.3000	0.3600	0.4200	0.4800	0.5400	0.6000
0.7	0.0000	0.0700	0.1400	0.2100	0.2800	0.3500	0.4200	0.4900	0.5600	0.6300	0.7000
0.8	0.0000	0.0800	0.1600	0.2400	0.3200	0.4000	0.4800	0.5600	0.6400	0.7200	0.8000
0.9	0.0000	0.0900	0.1800	0.2700	0.3600	0.4500	0.5400	0.6300	0.7200	0.8100	0.9000
1	0.0000	0.1000	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000

MONTE CARLO STANDARD ERROR

(α,μ)	0	50	100	150	200	250	300	350	400	450	500
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.1	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.2	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.3	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.4	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.5	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.6	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.7	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
0.8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	0.0000	0.0000
0.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table A1 – Monte Carlo results for n=50

We generate nodes of (α,μ) scanning the entire range of possible parameters, $\alpha \in [0,1]$ and $\mu = [0,V]$, with *V=100*, and *n=50* (where *n* is the number of volume buckets used to approximate the expected trade imbalance and intensity). The arrival rate of uninformed trade ε is then determined at each node, using the aforementioned expression. Ten equidistant partitions are investigated per dimension (11^2 nodes), carrying out 100,000 simulations of flow arrival per node (11^2 x 10^5 simulations in total). This yields a Monte Carlo accuracy of the order of 10^{-4}.

The Table shows that there is a negligible bias, on the order of 10^{-3}, towards underestimating the true value of the VPIN metric. These results motivate our use of *n=50* in our estimation of VPIN.

Next, we present the Monte Carlo algorithm that we use to simulate order flow based on known parameters *(α,μ,ε)*.[2] The VPIN metric can be estimated on that simulated order flow, and compared with the PIN values derived from the actual parameter values, *(α,μ,ε)*.

1. Set Monte Carlo Parameters:
    a. Number of Simulations: *S*

[2] δ can be arbitrarily set, as it does not affect the value of PIN or VPIN This can be easily seen by re-running this Monte Carlo with different values of δ.

   b. VPIN: *(V,n)*
   c. PIN: *(α,μ)*
2. Set $\varepsilon = \frac{V-\alpha\mu}{2}$, *s=0, j=0*
3. *j=j+1*
4. Draw three random numbers from a *U(0,1)* distribution: u_1, u_2, u_3
5. If $u_1<\alpha$:
   a. If $u_2<\delta$:
      i. $V_j^B = F^{-1}(u_3, \varepsilon)$
      ii. $V_j^S = F^{-1}(u_3, \mu + \varepsilon)$
   b. If $u_2\geq\delta$:
      i. $V_j^B = F^{-1}(u_3, \mu + \varepsilon)$
      ii. $V_j^S = F^{-1}(u_3, \varepsilon)$
6. If $u_1\geq\alpha$:
   a. $V_j^B = F^{-1}(u_3, \varepsilon)$
   b. $V_j^S = V_j^B$
7. If *j=n*:
   a. *j=0*
   b. *s=s+1*
   c. $VPIN_s = \frac{\sum_{j=1}^{n}\left|V_j^S - V_j^B\right|}{\sum_{j=1}^{n}\left(V_j^S + V_j^B\right)}$
8. If *s<S*, loop to Step 3
9. Compute results:
   a. $E[VPIN] = \frac{1}{S}\sum_{s=1}^{S} VPIN_s$
   b. $V[VPIN] = \frac{1}{S-1}\sum_{s=1}^{S} VPIN_s^2 - \frac{1}{S(S-1)}\left(\sum_{s=1}^{S} VPIN_s\right)^2$

TABLES

Table 1: Results of sampling by Chronological time versus Volume time

Stats (50)	Chrono time	Volume time
Mean	0.0000	0.0000
StDev	1.0000	1.0000
Skew	-0.0878	-0.4021
Kurt	34.2477	20.5199
Min	-23.5879	-25.1189
Max	20.8330	12.2597
L-B*	40.7258	24.8432
White*	0.0983	0.0448
J-B*	40.6853	12.8165

Stats (100)	Chrono time	Volume time
Mean	0.0000	0.0000
StDev	1.0000	1.0000
Skew	-0.1767	-0.4352
Kurt	48.1409	29.2731
Min	-30.5917	-34.1534
Max	26.5905	17.2191
L-B*	138.9320	48.7241
White*	0.0879	0.0278
J-B*	84.9096	28.7930

We use data from 1-minute time bars from the E-mini S&P futures contract from January 1, 2008-August 1, 2011. We draw an average of 50 (left panel) and 100 (right panel) price observations a day in two samples, one equally spaced in time (denoted chrono time) and the other equally spaced in volume (denoted volume time). We compute first differences in mean returns and standardize each sample. The table gives the resulting statistical properties, where $L-B^{*}=T(T+2)\sum_{i=1}^{10}\frac{\rho_{\tau,\tau-i}^{2}}{T-i}$, where $\rho_{\tau,\tau-i}$ is the sample autocorrelation at lag *i*. Both samples have the same number of observations (*T*); *White** is the R^2 of regressing the squared series against all cross-products of the first 10-lagged series; and $J-B^{*}=\frac{1}{6}\left(s^{2}+\frac{k^{2}}{4}\right)$, where *s* is skewness and *k* is excess kurtosis.

Table 2: The VPIN toxicity metric for S&P 500 E-mini Futures and WTI Crude Oil Futures

Stat	S&P500	Crude
Average	0.2251	0.2191
StDev	0.0576	0.0455
Skew	0.7801	0.5560
Ex. Kurt	0.9124	0.3933
AR(1)	0.9958	0.9932
#Observ.	44665	42425
CDF(0.1)	0.1578	0.1648
CDF(0.25)	0.1859	0.1858
CDF(0.5)	0.2178	0.2141
CDF(0.75)	0.2559	0.2492
CDF(0.9)	0.3023	0.2784

This table gives some basic statistics for the VPIN metric calculated for the E-mini S&P 500 future and the WTI (West Texas Intermediate) Crude Oil Future.

Table 3: The CDF of VPIN for different sample periods

Prob	CDF_1	CDF_2	CDF_Total
0.1	0.1711	0.1591	0.1648
0.2	0.1890	0.1792	0.1838
0.3	0.2030	0.1952	0.1989
0.4	0.2158	0.2101	0.2128
0.5	0.2284	0.2250	0.2267
0.6	0.2419	0.2409	0.2415
0.7	0.2571	0.2592	0.2583
0.8	0.2762	0.2824	0.2795
0.9	0.3050	0.3180	0.3119

This table provides critical values for the CDF of the E-mini S&P 500 VPIN for our entire sample and for the two subsamples: 1. Prior to the Flash Crash of May 6, 2010, and, 2. After the Flash Crash

Table 4: Conditional Probability Distributions of VPIN and Absolute Return

Absolute return between two consecutive buckets

VPIN percentiles	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	2.00%	>2.00%
0.05	96.56%	3.33%	0.11%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.10	96.33%	3.44%	0.23%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.15	91.54%	7.56%	0.73%	0.11%	0.06%	0.00%	0.00%	0.00%	0.00%
0.20	90.41%	8.24%	0.96%	0.28%	0.11%	0.00%	0.00%	0.00%	0.00%
0.25	90.47%	8.74%	0.79%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.30	89.45%	9.87%	0.62%	0.00%	0.06%	0.00%	0.00%	0.00%	0.00%
0.35	88.21%	10.55%	1.02%	0.06%	0.17%	0.00%	0.00%	0.00%	0.00%
0.40	84.72%	13.20%	1.69%	0.23%	0.17%	0.00%	0.00%	0.00%	0.00%
0.45	80.88%	17.26%	1.64%	0.17%	0.06%	0.00%	0.00%	0.00%	0.00%
0.50	81.90%	14.89%	2.65%	0.34%	0.06%	0.17%	0.00%	0.00%	0.00%
0.55	79.74%	17.55%	2.09%	0.56%	0.00%	0.06%	0.00%	0.00%	0.00%
0.60	79.30%	18.05%	2.09%	0.39%	0.11%	0.00%	0.06%	0.00%	0.00%
0.65	79.30%	16.92%	2.88%	0.39%	0.28%	0.11%	0.06%	0.00%	0.06%
0.70	75.06%	20.60%	3.22%	0.85%	0.11%	0.06%	0.06%	0.06%	0.00%
0.75	68.25%	24.99%	5.25%	1.18%	0.11%	0.11%	0.11%	0.00%	0.00%
0.80	62.32%	27.58%	6.49%	2.65%	0.51%	0.28%	0.17%	0.00%	0.00%
0.85	62.81%	26.52%	7.67%	1.86%	0.51%	0.23%	0.34%	0.06%	0.00%
0.90	56.38%	29.35%	9.20%	2.93%	1.24%	0.51%	0.06%	0.11%	0.23%
0.95	43.71%	30.51%	16.02%	5.75%	2.14%	0.79%	0.51%	0.17%	0.39%
1.00	39.56%	29.12%	16.42%	7.62%	3.27%	1.64%	0.90%	0.73%	0.73%

Table 4a – $Prob\left(\left|\frac{P_\tau}{P_{\tau-1}} - 1\right| \mid VPIN_{\tau-1}\right)$

Absolute return between two consecutive buckets

VPIN percentiles	0.25%	0.50%	0.75%	1.00%	1.25%	1.50%	1.75%	>=2%
0.05	6.28%	0.98%	0.14%	0.00%	0.00%	0.00%	0.00%	0.00%
0.10	6.27%	1.02%	0.28%	0.00%	0.00%	0.00%	0.00%	0.00%
0.15	5.96%	2.23%	0.90%	0.44%	0.63%	0.00%	0.00%	0.00%
0.20	5.88%	2.43%	1.17%	1.11%	1.26%	0.00%	0.00%	0.00%
0.25	5.89%	2.59%	0.97%	0.00%	0.00%	0.00%	0.00%	0.00%
0.30	5.82%	2.92%	0.76%	0.00%	0.63%	0.00%	0.00%	0.00%
0.35	5.74%	3.12%	1.24%	0.22%	1.89%	0.00%	0.00%	0.00%
0.40	5.51%	3.90%	2.07%	0.89%	1.89%	0.00%	0.00%	0.00%
0.45	5.26%	5.10%	2.00%	0.67%	0.63%	0.00%	0.00%	0.00%
0.50	5.33%	4.40%	3.24%	1.33%	0.63%	4.29%	0.00%	0.00%
0.55	5.19%	5.19%	2.55%	2.22%	0.00%	1.43%	0.00%	0.00%
0.60	5.16%	5.34%	2.55%	1.56%	1.26%	0.00%	2.50%	0.00%
0.65	5.16%	5.00%	3.52%	1.56%	3.14%	2.86%	2.50%	2.22%
0.70	4.88%	6.09%	3.93%	3.33%	1.26%	1.43%	2.50%	2.22%
0.75	4.44%	7.39%	6.42%	4.67%	1.26%	2.86%	5.00%	0.00%
0.80	4.06%	8.16%	7.94%	10.44%	5.66%	7.14%	7.50%	0.00%
0.85	4.09%	7.84%	9.39%	7.33%	5.66%	5.71%	15.00%	2.22%
0.90	3.67%	8.67%	11.25%	11.56%	13.84%	12.86%	2.50%	13.33%
0.95	2.84%	9.02%	19.60%	22.67%	23.90%	20.00%	22.50%	22.22%
1.00	2.57%	8.61%	20.08%	30.00%	36.48%	41.43%	40.00%	57.78%

Table 4b – $Prob\left(VPIN_{\tau-1} \left| \left|\frac{P_\tau}{P_{\tau-1}} - 1\right|\right.\right)$

This table gives the conditional probability distributions of the VPIN toxicity metric and absolute return. Table 4(a) shows distributions of returns at time τ given VPIN at time τ-1. Table 4 (b) shows distributions of VPIN at time τ-1 given returns at time τ.

Table 5: VPIN and Volatility Events

VPIN percentiles	Absolute return between any two intermediate buckets								
	0.25%	**0.50%**	**0.75%**	**1.00%**	**1.25%**	**1.50%**	**1.75%**	**2.00%**	**>2.00%**
0.05	25.00%	16.67%	25.00%	8.33%	0.00%	8.33%	8.33%	8.33%	0.00%
0.10	27.27%	27.27%	3.03%	9.09%	3.03%	0.00%	9.09%	0.00%	21.21%
0.15	39.58%	12.50%	10.42%	4.17%	6.25%	4.17%	2.08%	8.33%	12.50%
0.20	30.99%	25.35%	15.49%	8.45%	9.86%	0.00%	4.23%	1.41%	4.23%
0.25	21.69%	19.28%	22.89%	12.05%	6.02%	9.64%	2.41%	2.41%	3.61%
0.30	17.50%	25.00%	20.00%	13.75%	7.50%	3.75%	3.75%	2.50%	6.25%
0.35	19.05%	26.19%	16.67%	13.10%	5.95%	8.33%	5.95%	2.38%	2.38%
0.40	10.71%	16.67%	23.81%	13.10%	11.90%	7.14%	4.76%	2.38%	9.52%
0.45	15.91%	19.32%	13.64%	13.64%	14.77%	9.09%	6.82%	1.14%	5.68%
0.50	19.59%	16.49%	19.59%	16.49%	6.19%	4.12%	5.15%	4.12%	8.25%
0.55	18.39%	13.79%	18.39%	13.79%	11.49%	4.60%	6.90%	4.60%	8.05%
0.60	14.04%	10.53%	21.05%	17.54%	7.02%	7.02%	1.75%	10.53%	10.53%
0.65	12.00%	12.00%	8.00%	10.00%	14.00%	8.00%	2.00%	12.00%	22.00%
0.70	12.07%	10.34%	12.07%	8.62%	6.90%	8.62%	6.90%	6.90%	27.59%
0.75	14.04%	14.04%	10.53%	15.79%	5.26%	3.51%	10.53%	0.00%	26.32%
0.80	10.53%	8.77%	7.02%	8.77%	10.53%	3.51%	1.75%	7.02%	42.11%
0.85	8.33%	13.89%	5.56%	8.33%	11.11%	8.33%	5.56%	5.56%	33.33%
0.90	0.00%	0.00%	0.00%	10.00%	10.00%	20.00%	0.00%	10.00%	50.00%
0.95	0.00%	7.69%	7.69%	7.69%	0.00%	0.00%	0.00%	0.00%	76.92%
1.00	0.00%	0.00%	0.00%	0.00%	10.00%	0.00%	0.00%	0.00%	90.00%

This table shows the maximum exposure of the market maker to price movements following a transition of the VPIN metric from one level to the next. It is computed using10 buckets per day and a sample length of 350 (about 1.6 months).

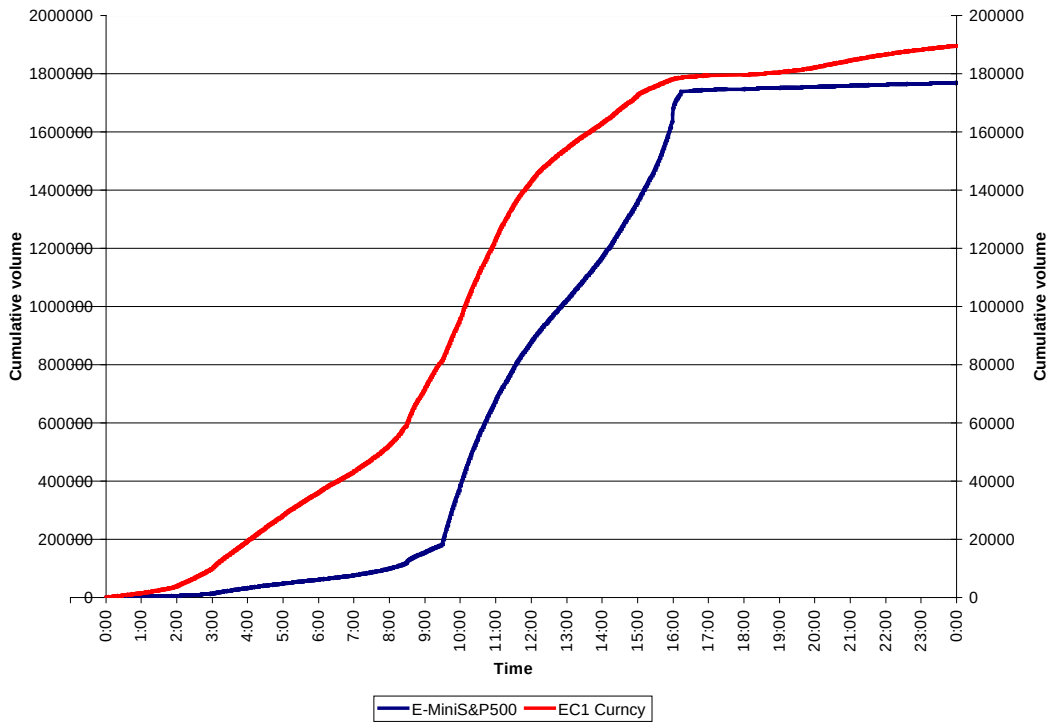

Figure 1: Average daily volume profile of E-Mini S&P Future and EC1 Future

This figure shows the intra-day volume pattern in the E-Mini S&P futures contract and the Eurodollar/USDollar futures contract (the EC1 contract). E-Mini volume is measured in contracts on the left axis while the EC1 volume is measured in contracts on the right axis.

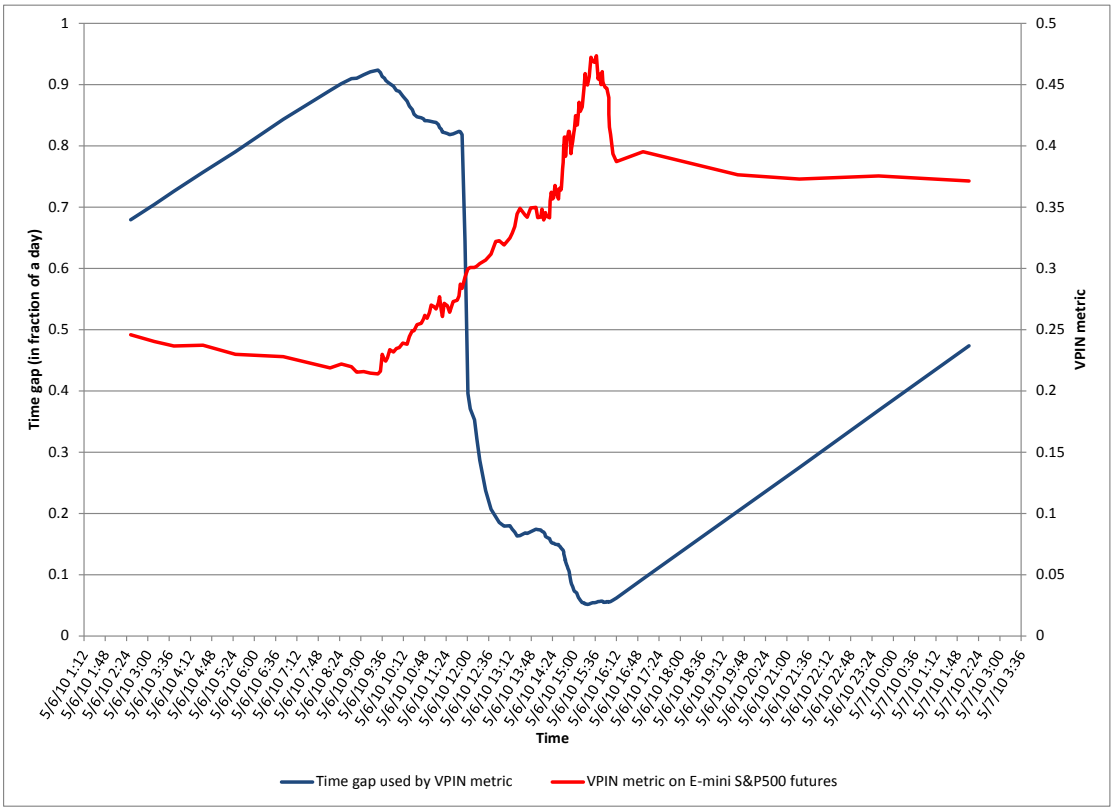

Figure 2: Time gap and VPIN on May 6, 2010

This Figure demonstrates how the time gap used to compute the E-Mini S&P futures contract changed over the course of May 6. The Figure also shows how the VPIN metric changed over the day. The time gap is measured on the left axis while the VPIN is measured over the right axis.

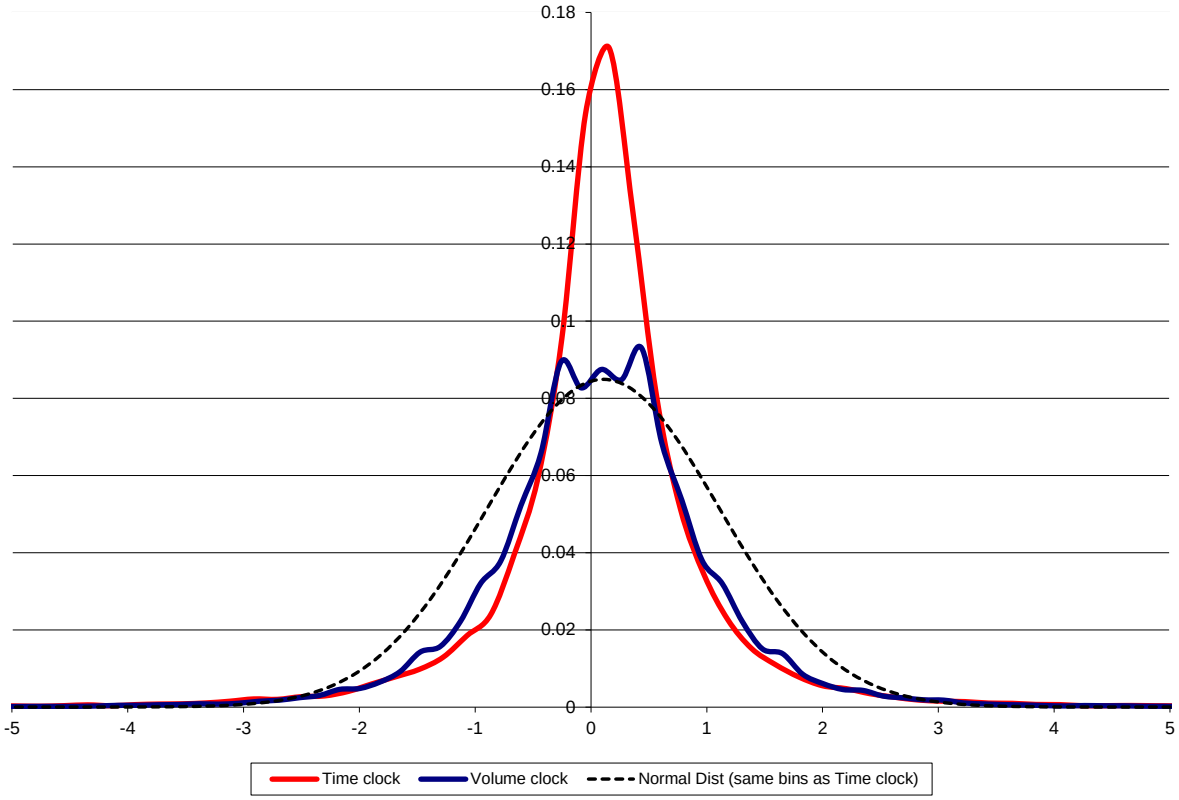

Figure 3: Graphs of Price Changes for E-mini S&P 500 futures, sampled by regular time intervals and regular volume intervals

This figure shows the distribution of normalized price changes for the E-Mini S&P futures contract. Our sample period is January 1, 2008 – August 1, 2011. We draw an average of 50 price observations per day equally spaced in time for the time clock and equally spaced by volume for the volume clock. We compute first order differences and standardize each sample. The black line gives the normal distribution.

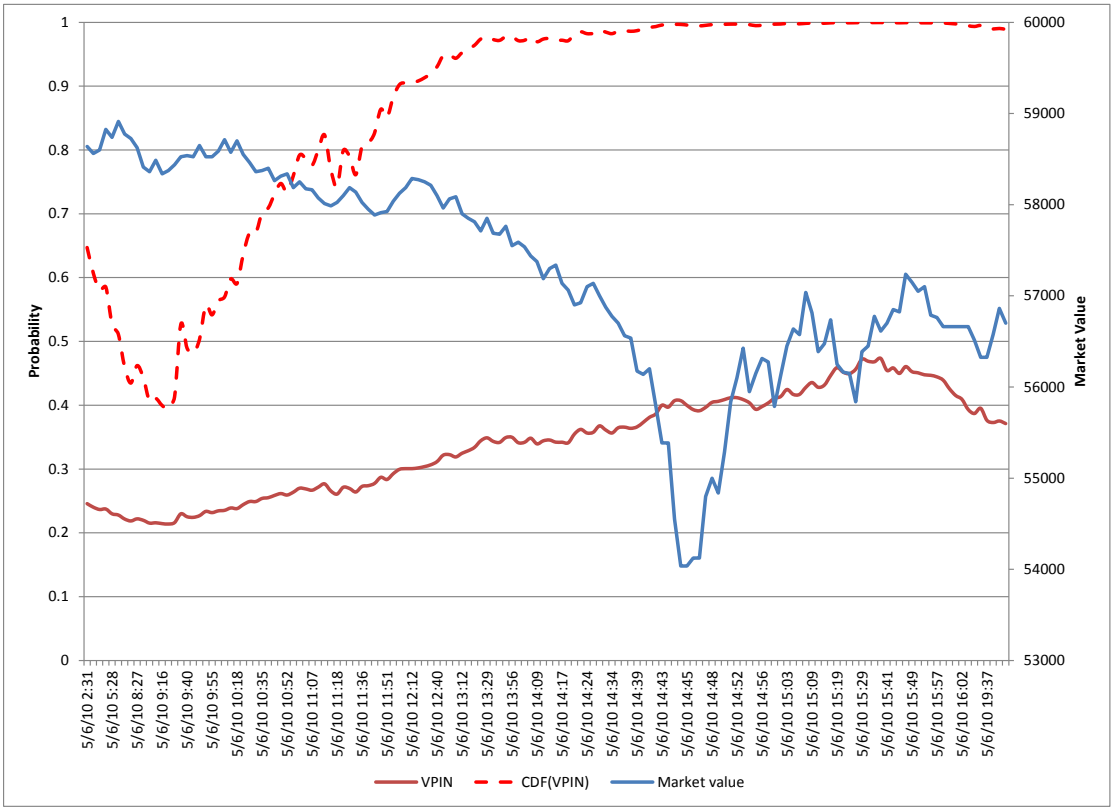

Figure 4: The VPIN Toxicity Metric and Trade Classification

(a) VPIN estimated on 1-minute time bars bulk classification

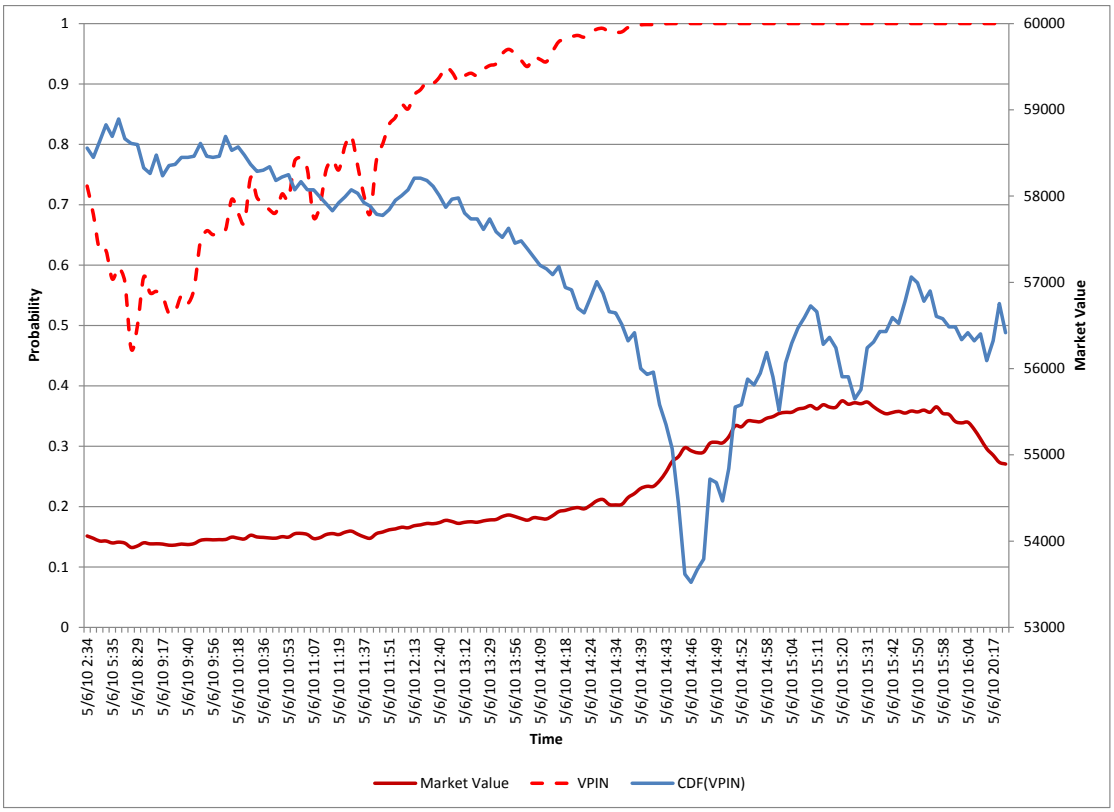

Figure 4: The VPIN Toxicity Metric and Trade Classification

(b) – VPIN estimated on 10-second time bars bulk classification

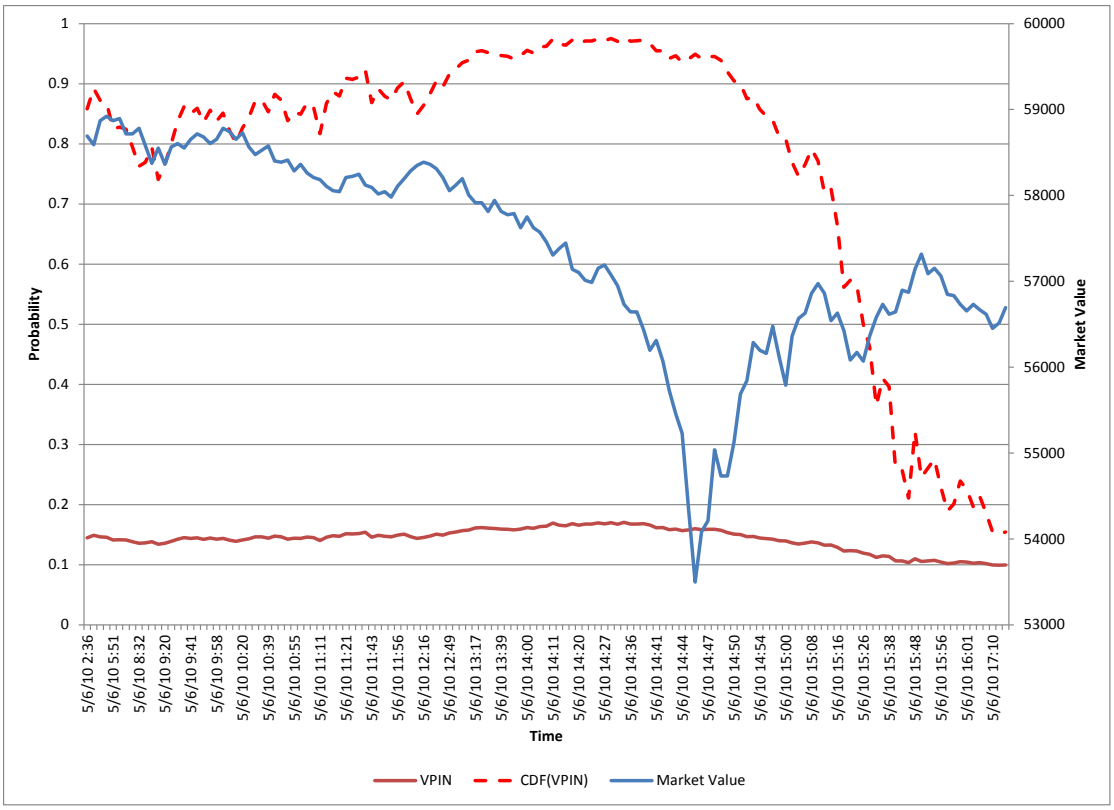

Figure 4: The VPIN Toxicity Metric and Trade Classification

(c) – VPIN estimated on trade-by-trade Lee-Ready classification

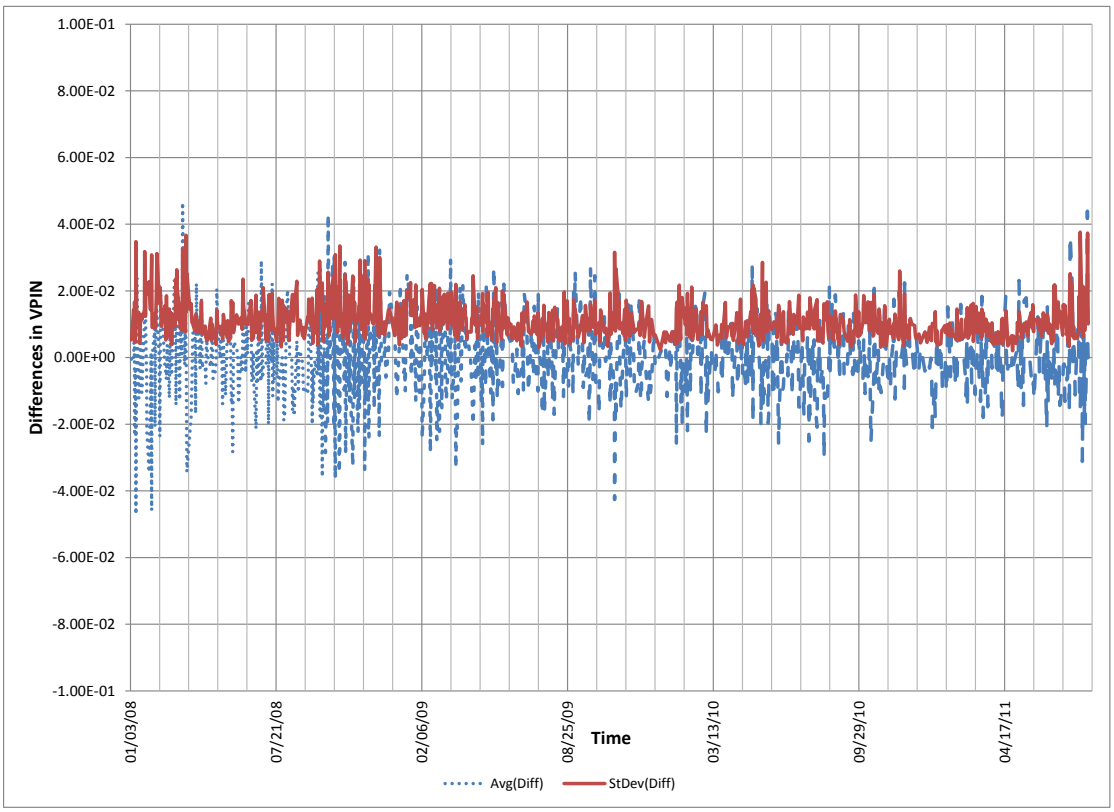

Figure 5: Stability of VPIN Toxicity Metric Estimates

This figure shows the differences in VPIN and their cross-sectional standard deviations from computing 1,000 alternative VPIN trajectories for the E-Mini S&P futures contract.

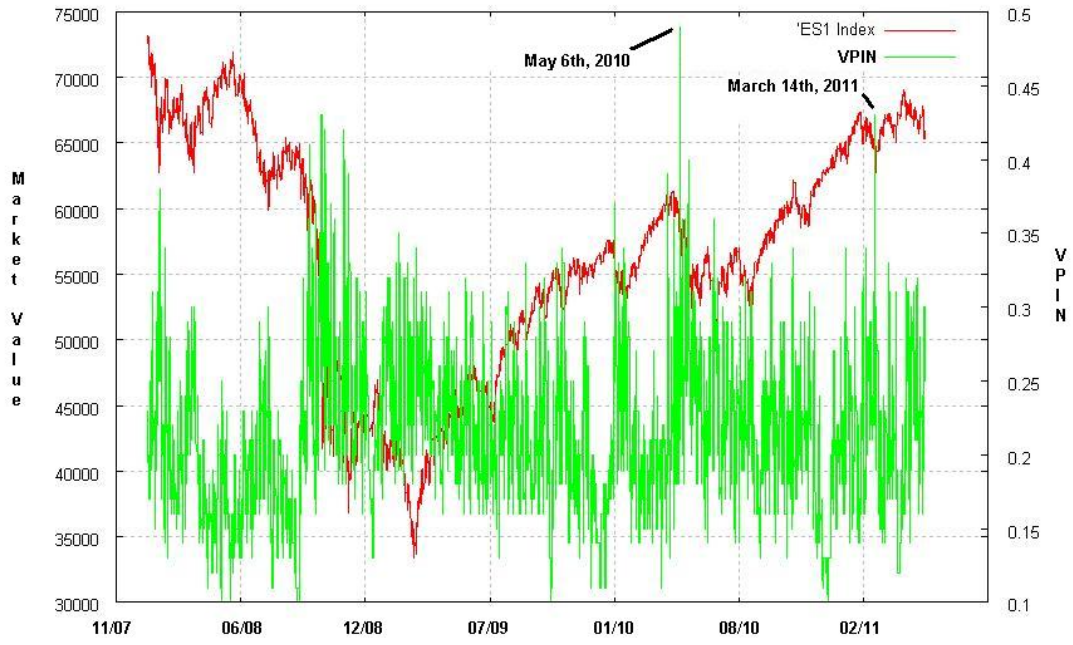

Figure 6: VPIN for E-mini S&P 500 futures

This figure shows the evolution of the E-Mini S&P futures contract (measured on the left axis) and its VPIN metric (measured on the right axis). The sample period is January 1, 2008 – August 1. 2011.

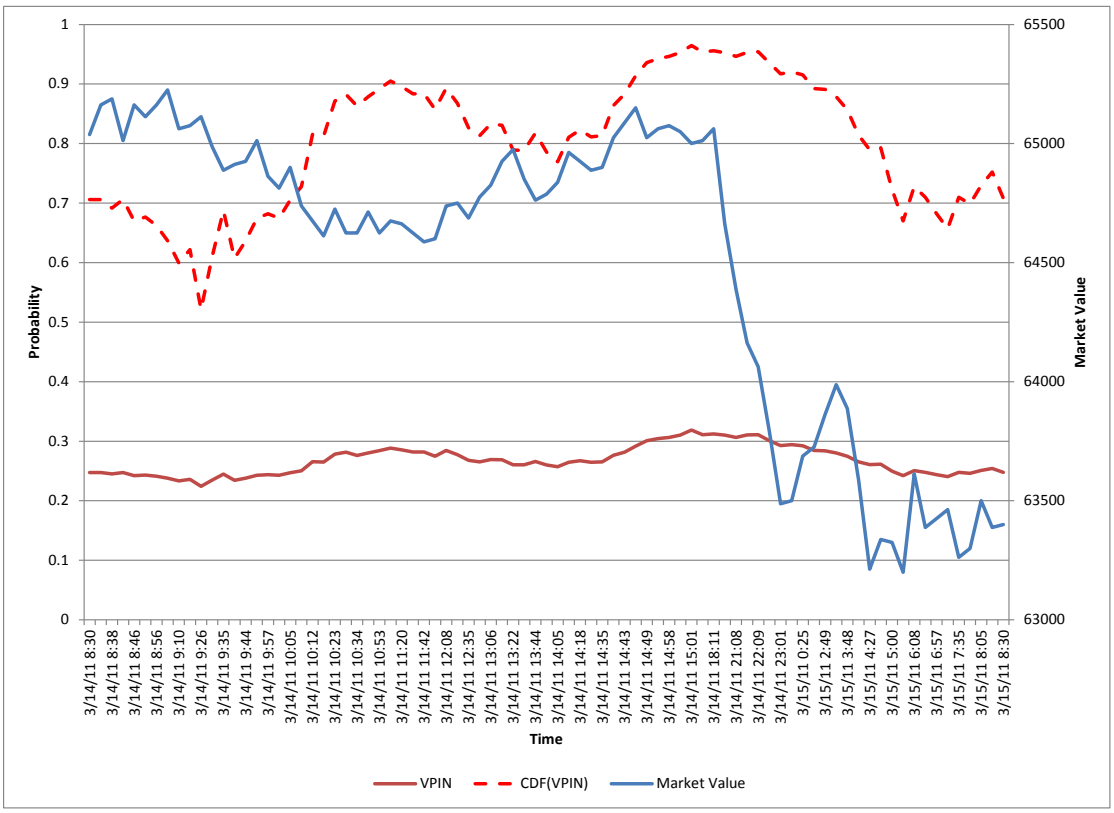

Figure 7: E-mini S&P 500 futures during the Fukushima Nuclear crisis

This figure shows the behavior of the E-Mini S&P futures contract and its VPIN on March 14, 2011.

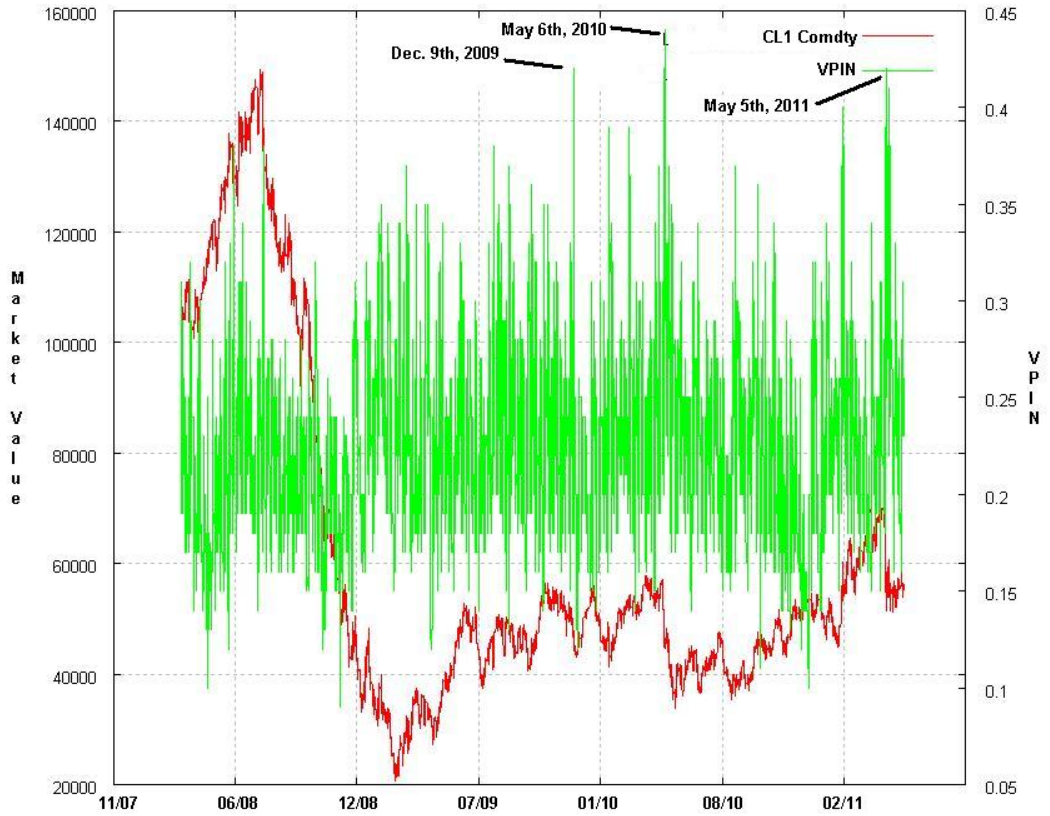

Figure 8: The VPIN Toxicity Metric for WTI Crude Oil Futures

This figure shows the evolution of the WTI futures contract (measured on the left axis) and its VPIN metric (measured on the right axis) for the sample period January 1, 2008 – August 1, 2011.

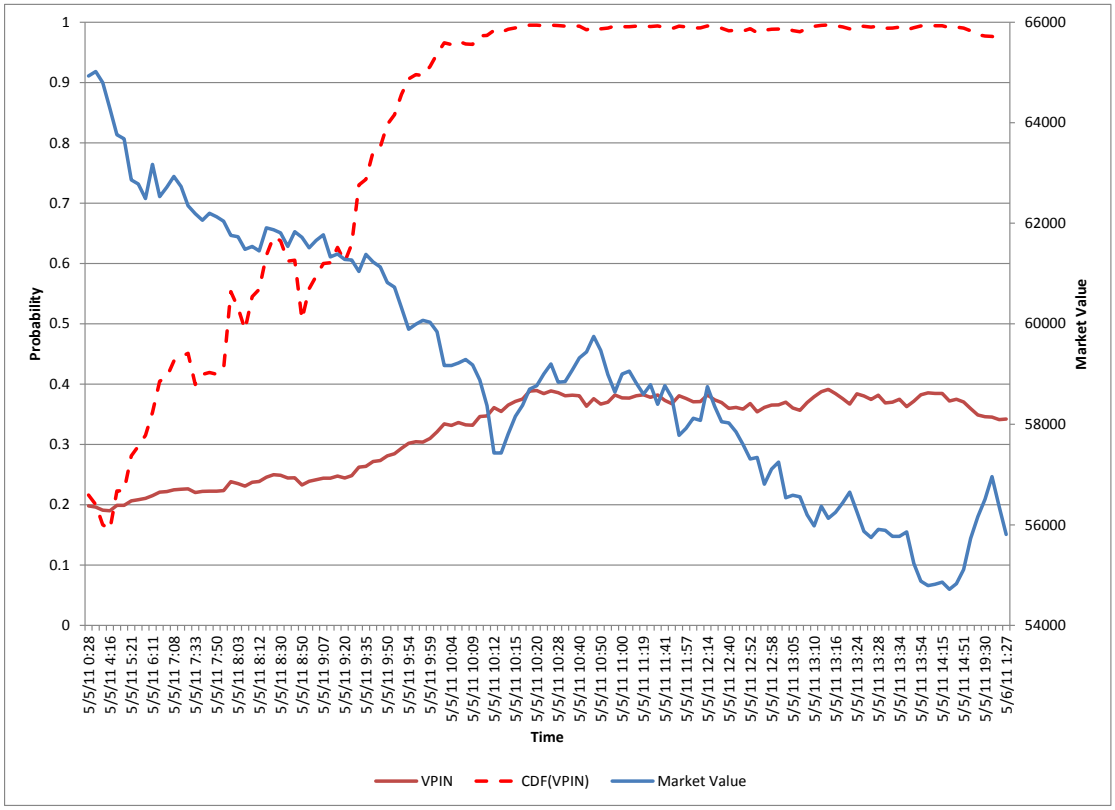

Figure 9: Crude on May 5th 2011

This figure shows the behavior of the WTI futures contract and its VPIN on March 14, 2011.

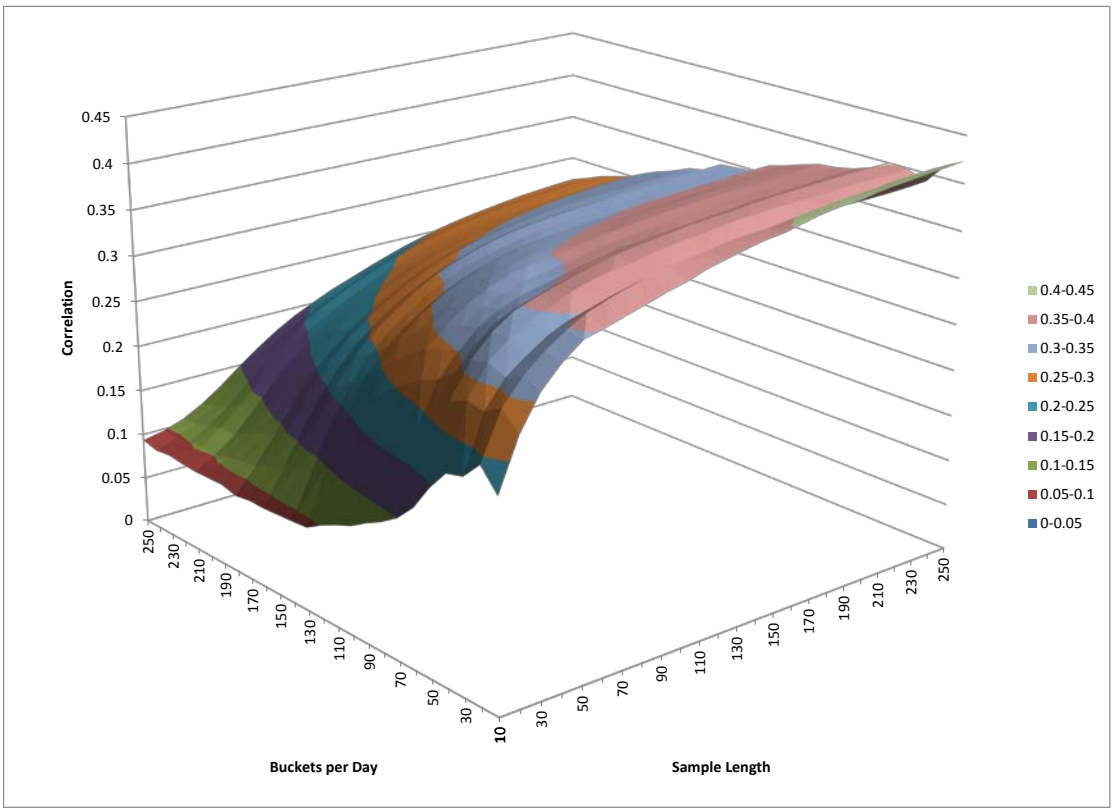

Figure 10: Correlation between VPIN and future volatility for E-mini S&P 500

This figure shows the correlation between VPIN and the following absolute returns for the E-Mini S&P futures contract for various combinations of buckets per day and the number buckets used to calculate VPIN (the sample length).

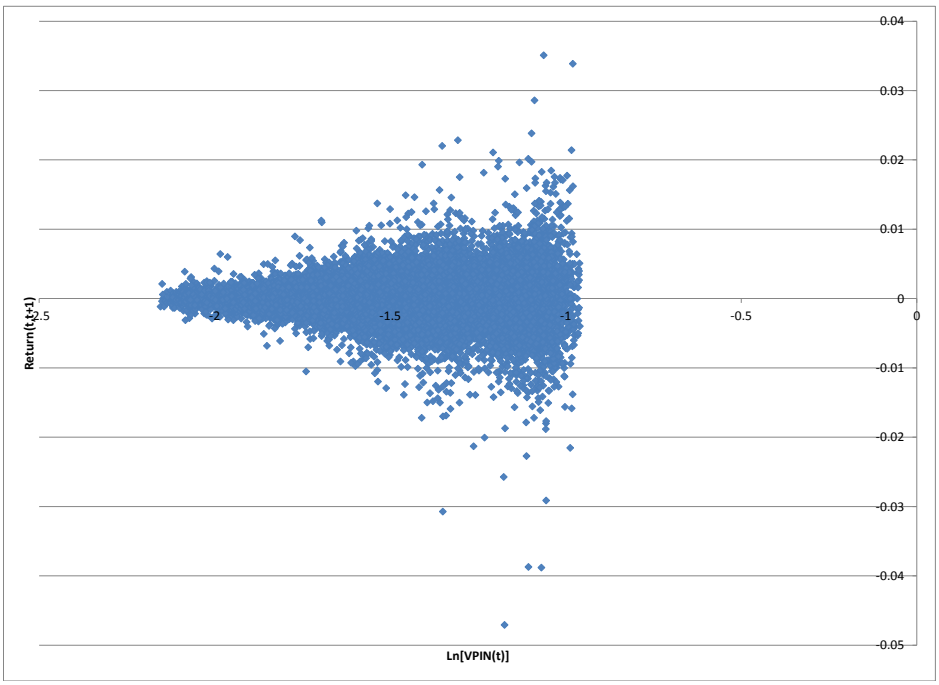

Figure 11: E-mini S&P 500's response to VPIN

This figure shows the return on the E-mini S&P 500 futures over the next volume bucket (1/50 of an average day's volume using the (*50,250)* combination) sorted by the log of the previous VPIN level.

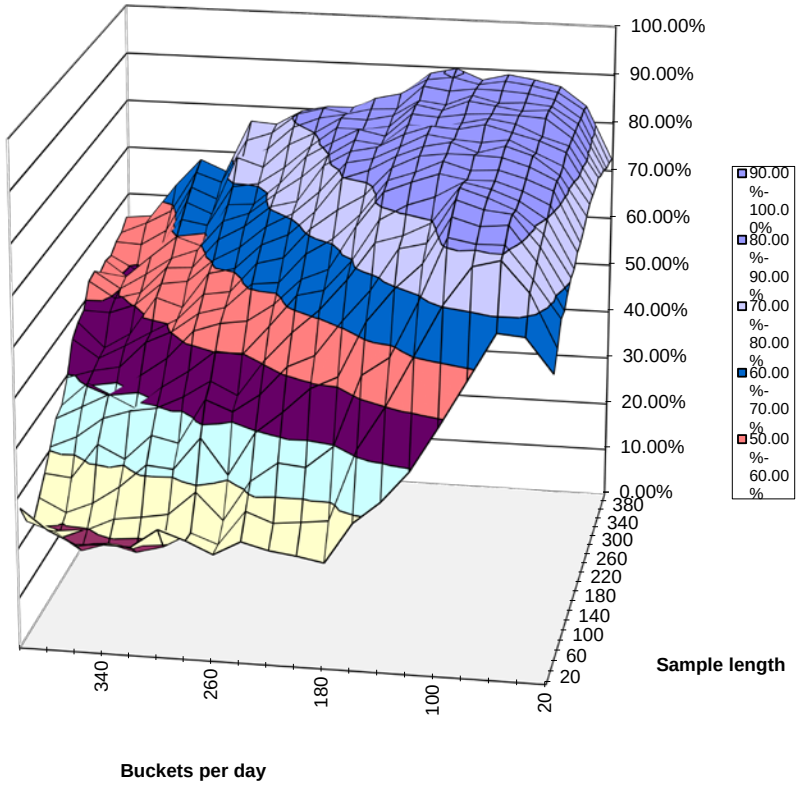

Figure 12: Heat Map of High VPIN and High Returns

This figure shows how parameter selection regarding bucket size and sample length influences the toxicity measure and subsequent volatility. The Figure shows $\text{Prob}\left(\text{CDF}(\text{VPIN}_{\tau-1}) > \frac{3}{4}\left|\left|\frac{P_\tau}{P_{\tau-1}} - 1\right| > 0.75\%\right)\right.$.

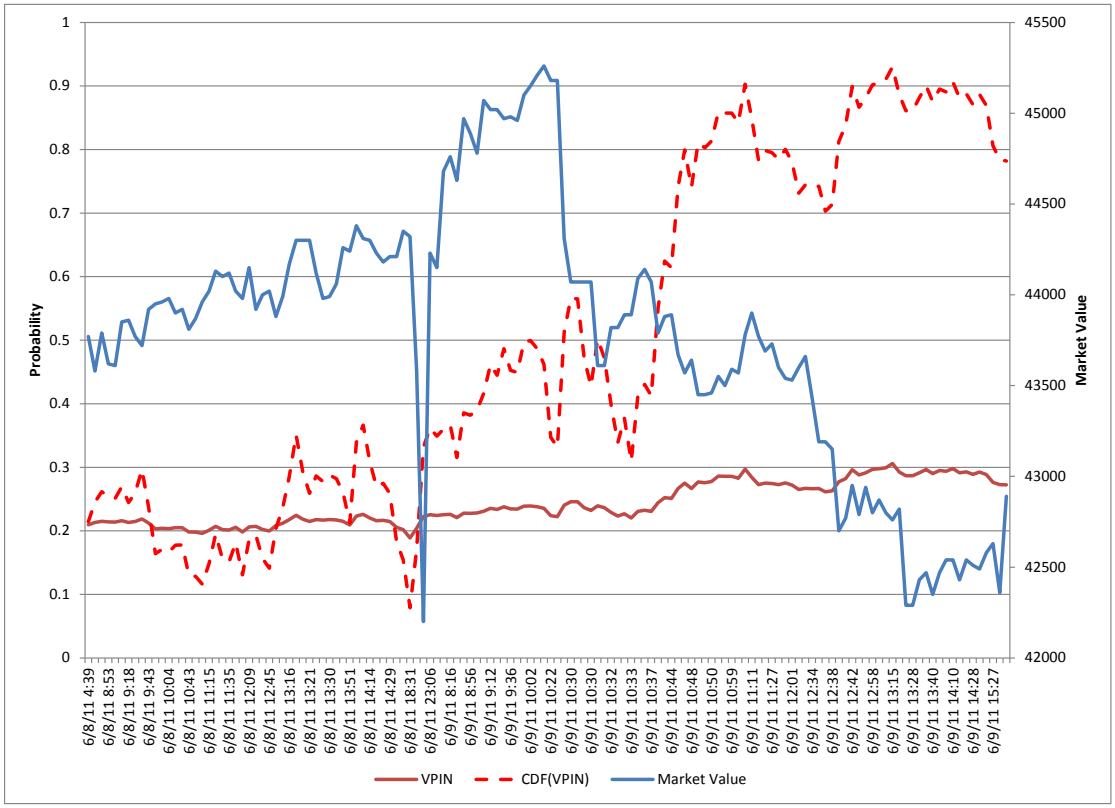

Figure 13: Natural Gas on June 8th 2011

This figure shows the behavior of the Natural Gas futures contract and its VPIN on June 8, 2011.

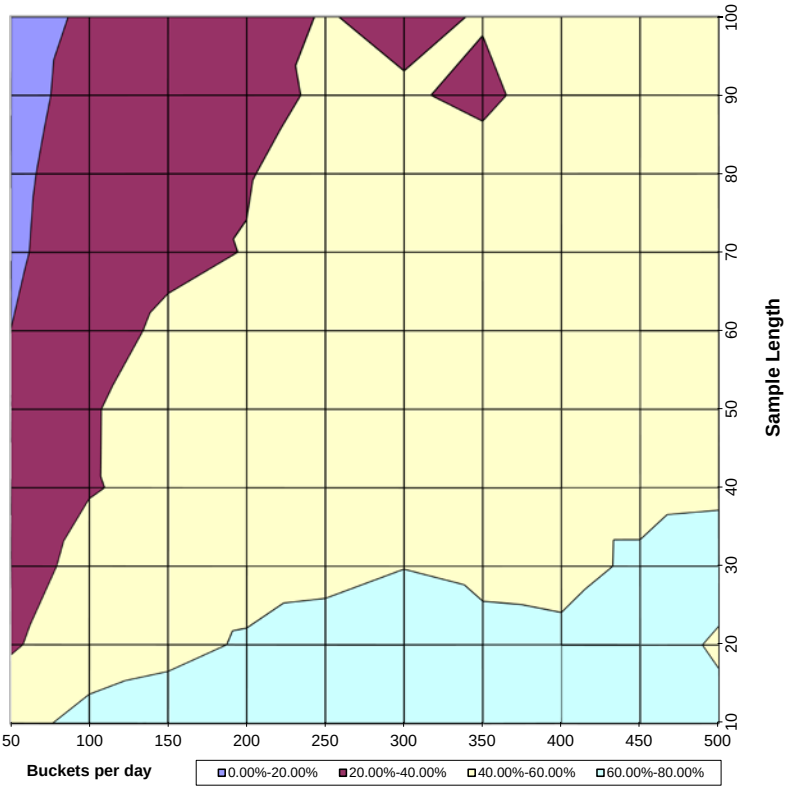

Figure 14: Probability of "True Positives"

This figure plots the probabilities of the largest absolute return being in excess of 0.75% while VPIN remains in any 5%-tile in the upper quartile of its distribution for various combinations of buckets per day and sample length.

REFERENCES

Admati, A. and P. Pfleiderer. 1988. A Theory of Intra-day Patterns: Volume and Price Variability, *Review of Financial Studies*, 1 (1): 3-40.

Ané, T. and H. Geman. 2000. Order flow, transaction clock and normality of asset returns, *The Journal of Finance*, 55: 2259–2284.

Bethel, W., D. Leinweber, O. Rübel and K. Wu. 2011. Federal Market Information Technology in the Post Flash Crash Era: Roles for Supercomputing, Working Paper, CIFT, Lawrence Berkeley National Laboratory.

Bloomfield, R., M. O'Hara, and G. Saar. 2005. The Make or Take Decision in an Electronic Market: Evidence on the Evolution of Liquidity, *Journal of Financial Economics*, 75 (1): 165-199.

Brogaard, J.. 2010. High Frequency Trading and Its impact on Market Quality, Working Paper, Northwestern University.

Brunnermeier, M. and L. Pedersen. 2009. Market Liquidity and Funding Liquidity, *Review of Financial Studies*, 22 (6): 2201-2238.

CFTC. 2010. Proposed Rules, *Federal Register*, 75 (112) (June 11): 33198-33202.

Chaboud, A., E. Hjalmarsson, C. Vega and B. Chiquoine. 2009. Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market, FRB International Finance Discussion Paper No. 980.

Clark, Peter K. 1973. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, *Econometrica*, 41: 135-155.

DeGennaro, R.P. and R.E. Shrieves. 1995. Public Information Releases, Private Information Arrival and Volatility in the FX Market, in *HFDF-1: First International Conference on High Frequency Data in Finance*, Volume 1. Olsen and Associates, Zurich.

Deuskar, P. and T. Johnson. 2011. Market Liquidity and Flow-driven Risk, *Review of Financial Studies*, 24 (3): 721-753.

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

Easley, D. and M. O'Hara. 1992. Time and the process of security price adjustment, *Journal of Finance*, 47: 576-605.

Easley, D., Kiefer, N., O'Hara, M. and J. Paperman. 1996. Liquidity, Information, and Infrequently Traded Stocks, *Journal of Finance*, September.

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

Easley, D., M. López de Prado and M. O'Hara. 2011a. The Microstructure of the Flash Crash, *The Journal of Portfolio Management*, Winter 2011

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

Engle, R. 1996. Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data, *Econometrica*, 66: 1127-1162.

Engle, R. and J. Lange. 2001. Predicting VNET: A Model of the Dynamics of Market Depth, *Journal of Financial Markets*, 4: 113-142.

Engle, R. and J. Russell. 2005. A Discrete-State Continuous-Time Model of Financial Transactions Prices and Times: The Autoregressive conditional Multinomial-Autoregressive Conditional Duration Model, *Journal of Business and Economic Statistics*, 23 (2): 166-180.

Fisher, R.A. 1915. Frequency Distribution of the Values of the Correlation Coefficient in Samples of an Indefinitely Large Population. *Biometrika*, 10: 507-521.

Foucault, T., Kadan, O., and E. Kandel. 2009. Liquidity Cycles and Make/Take Fees in Electronic Markets, http://ssrn.com/abstract=1342799.

Galant, R., A. Rossi, and G. Tauchen. 1992. Stock Prices and Volume, *Review of Financial Studies*, 5: 199-242.

Glosten, L. R. and P. Milgrom. 1985. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, *Journal of Financial Economics*, 14: 71-100.

Harris, L. 1986. Cross-Security Tests of the Mixture of Distribution Hypothesis, *Journal of Financial and Quantitative Analysis*, 21 (1): 39-46.

Hasbrouck, J. and G. Saar. 2010. Low Latency Trading, Working Paper, Cornell University.

Hendershott, T., Jones, C., and A. Menkveld. 2011. Does Algorithmic Trading Improve Liquidity?, *Journal of Finance*, 66 (1): 1-33.

Hendershott, T. and R. Riordan. 2009. Algorithmic Trading and Information, NET Institute Working Paper No. 09-08.

Huang, J. and J. Wang. 2009. Liquidity and Market Crashes, *Review of Financial Studies*, 22 (7): 2607-2643.

Iati, R. 2009. High-Frequency Trading Technology, TABB Group.

Jeria, D. and G. Sofianos. 2008. Passive Orders and Natural Adverse Selection, *Street Smart*, 33 (September 4).

Jones, C.M., G. Kaul and M.L. Lipton. 1994. Transactions, Volume and Volatility, *Review of Financial Studies* 7 (4): 631-651.

Kirilenko, A., A. P. Kyle, M. Samedi, and T. Tuzan. 2010. The Flash Crash: The Impact of High-Frequency Trading on an Electronic Market, Working paper.

Kyle, A. S. 1985. Continuous Auctions and Insider Trading, *Econometrica*, 53: 1315-1335.

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

Mandlebrot, B. and M. Taylor. 1967,. On the Distribution of Stock Price Differences, *Operations Research*, 15 (5): 1057-1062.

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

DISCLAIMER

The views expressed in this paper are those of the authors and not necessarily reflect those of Tudor Investment Corporation. No investment decision or particular course of action is recommended by this paper.