ASSESSING VPIN MEASUREMENT OF ORDER FLOW TOXICITY USING PERFECT TRADE CLASSIFICATION

Torben G. Andersen¹ Oleg Bondarenko²

¹Kellogg School, Northwestern University; NBER; CREATES
²University of Illinois at Chicago

Quantitative Finance Retrospective Workshop

The Fields Institute,
Toronto, Canada

October 28, 2013
VPIN = Volume-synchronized Probability of INformed trading, Variation and Extension to Concept of PIN

Developed in three recent papers by Easley, López de Prado and O’Hara (ELO):


ELO find

- VPIN is **Proxy for “Toxicity” of Order Flow**
- VPIN **Signaled Flash Crash** on May 6, 2010, as early as several Hours Ahead
- VPIN provides **Superior and more Timely Indicator of Future Market Volatility** than celebrated Option-Implied Volatility measure, VIX, the so-called “Fear Gauge”
- **VPIN leads VIX**, not otherwise
Recent ELO Developments include:

(1) “Flow Toxicity and Liquidity in a High-Frequency World,”

(2) “Bulk Classification of Trading Activity,”

(3) “The Volume Clock: Insights into the High Frequency Paradigm,”

(4) “Optimal Execution Horizon,”
SUGGESTED APPLICATIONS OF VPIN

- For Traders, Market Makers: Tool to Gauge Risk of Liquidity-induced Crash

- To Manage and Hedge this Risk: Traded Futures on VPIN?

- For Regulators and Exchanges: Tool to Monitor Market Conditions

- Circuit Breaker based on VPIN: May Help Prevent some Future Crashes?

- Joint CFTC-SEC Advisory Committee on Emerging Regulatory Issues to Address Challenges raised by the Flash Crash

- Patent for VPIN submitted to US Patent and Trademark Office
“‘Toxic’ Orders Can Predict Likelihood of Stock Market Crashes, Study Says"
(Bloomberg, October, 29, 2010):

“The measure would have been able to anticipate two hours in advance there was a high probability of a liquidity-induced event on May 6,” said Lopez de Prado, head of high-frequency futures at Tudor. “It measures order toxicity, or the probability the market is going to have persistent order imbalances that are going to damage market makers.”

“This would be much more effective than a circuit breaker,” he said. U.S. stock markets and the Securities and Exchange Commission instituted curbs in June that briefly halt trading in a security when its price moves rapidly. “That stops the infection after the infection is already widespread,” he said.
“‘Flow Toxicity’ Metric May Help Avoid Another Flash Crash: Study”
(Wall Street Journal, November 5, 2010):

“Complementary to circuit breakers based on price action, they could have circuit breakers based on our metric,” said Marcos Lopez de Prado.

As a measurement of risk, the VPIN metric could help market participants “anticipate a rise in volatility and estimate the risk of a liquidity-induced crash”.

“’Flash’ Crashes now Predictable, Thanks to Cornell-Developed Metric”
(Cornell University Chronicle, December, 1, 2010)
E-mini S&P 500 Futures Contract (CME Globex)

Best Bid & Offer Files, CME Group DataMine

February 10, 2006 – March 12, 2011 (Over 5 Years)

Transaction $i$ represented by Triplet $(t_i, p_i, s_i)$, where

- $t_i$ is Time (in Seconds),
- $p_i$ is Price, and
- $s_i$ is Trade Size.

Quote $i$ Contains

- Time Stamp in Second,
- Bid Price, Ask Price,
- Bid Depth, Ask Depth,

When many Events Occur in Same Second, we Know Correct Sequence
## Descriptive Trading Statistics

<table>
<thead>
<tr>
<th></th>
<th>Regular</th>
<th>Overnight</th>
<th>Holiday</th>
</tr>
</thead>
<tbody>
<tr>
<td># Days</td>
<td>1285</td>
<td>1285</td>
<td>30</td>
</tr>
<tr>
<td>Volume (1 min)</td>
<td>3973</td>
<td>208</td>
<td>69</td>
</tr>
<tr>
<td># Trades (1 min)</td>
<td>285</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td># Order Book Changes (1 min)</td>
<td>1730</td>
<td>175</td>
<td>60</td>
</tr>
<tr>
<td># BBO Changes (1 min)</td>
<td>26</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Notional Value, $Mln (1 min)</td>
<td>235</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Trade Size</td>
<td>13.9</td>
<td>8.9</td>
<td>7.3</td>
</tr>
<tr>
<td>Order Book Changes per Trade</td>
<td>6.1</td>
<td>7.5</td>
<td>6.4</td>
</tr>
<tr>
<td>Trades per BBO Changes</td>
<td>10.9</td>
<td>3.0</td>
<td>2.4</td>
</tr>
</tbody>
</table>

### Order Size: Average Daily Percentiles

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>10%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>99%</th>
<th>99.9%</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1.0</td>
<td>1.0</td>
<td>2.1</td>
<td>7.5</td>
<td>30.5</td>
<td>233.4</td>
<td>643.5</td>
<td>1683.5</td>
</tr>
</tbody>
</table>
Sequence Indicator Allows for “Perfect” Classification.

<table>
<thead>
<tr>
<th>Time</th>
<th>Sequence</th>
<th>BidPrice</th>
<th>BidSize</th>
<th>AskPrice</th>
<th>AskSize</th>
<th>TradePrice</th>
<th>TradeSize</th>
</tr>
</thead>
<tbody>
<tr>
<td>17:02:58</td>
<td>5770</td>
<td>1289.50</td>
<td>125</td>
<td>1289.75</td>
<td>98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:02:58</td>
<td>5780</td>
<td>1289.50</td>
<td>125</td>
<td>1289.75</td>
<td>99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:02:59</td>
<td>5790</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1289.75</td>
<td>5</td>
</tr>
<tr>
<td>17:02:59</td>
<td>5800</td>
<td>1289.50</td>
<td>125</td>
<td>1289.75</td>
<td>94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Split Volume at Opening Auction equally to Buy and Sell.

If Trade Price between Quotes, Allocate equally to Buy and Sell.

Positive Identification of Direction over 99.95% of Time.
**Figure**: This figure plots the price and cumulative signed order imbalance for May 6, 2010. The solid vertical lines indicate the official CTFC report timing of the “flash crash.”
1. Establishing the Transaction Sequence

- Fix Sample Period for the Analysis in Calendar Time, $[0, T]$.
- Each Individual Traded Contract represented by $(t_i, p_i)$.
- Trade of $v > 1$ Contracts $\leftrightarrow v$ trades of One Contract at Identical Price.
- $0 = t_0 \leq t_1 < \ldots \leq t_I \leq T$: Timing of Consecutive (One Unit) Trades.
- This Establishes Complete Transaction Record for Asset over $[0, T]$.
- In Total, $I$ contracts Traded over $[0, T]$, indexed $i = 1, \ldots, I$. 
2. Defining the Volume Bucket

- Define $V$ to indicate a fixed increment to cumulative volume.
- Leading choice by ELO is $\frac{1}{50}$th of daily average volume over sample.
- Days with high volume contain more buckets than days with lower volume.
- Measuring time in buckets represents Event or Business Time Transformation.
- Let $T_\ell$ indicate time when bucket $\ell$ has just been filled.
- Let $P_\ell$ be last transaction price within bucket $\ell$.
- Let $L$ be number of full volume buckets in $[0, T]$.
- $(T_\ell, P_\ell), \ell = 1, \ldots, L$, summarizes transaction record for buckets.
Constructing VPIN

3. Computing VPIN from Order Imbalances

- For each Bucket, \( \ell \), we need associated Order Imbalance Measure, \( OI_\ell \).
- VPIN is Moving Average over Last \( L \) OI Measures:

\[
VPIN_t = \frac{1}{L} \sum_{j=0}^{L-1} OI_{\ell-j}, \quad t = T_\ell.
\]

- VPIN is Computed when a Volume Bucket is Filled, so Random Timing.
- ELO Leading Choice is \( L = 50 \).
- VPIN, on average, Uses One Trading Day for VPIN Measurement.
- The MA(50) Filter implies fairly Aggressive Smoothing.
- Current Large OI Innovation has only \((1/50)\) Weight in VPIN.
- On the other hand, Things Move Quicker, when Activity is Elevated.
4. Defining Order Imbalances

- VPIN Requires Order Imbalance Measure for each Volume Bucket.

- Order Imbalance Determined via Trade Classification Scheme.

- Any given Scheme must Deliver \( b_\ell = \frac{V^B_\ell}{V} \).

- We have, \( V = V^B_\ell + V^S_\ell \), and \( 0 \leq b_\ell \leq 1 \).

- *Signed* Order Imbalance, SOI: \( \gamma_\ell = \frac{V^B_\ell - V^S_\ell}{V} = 2 b_\ell - 1 \).

- \( OI_\ell = \frac{|V^B_\ell - V^S_\ell|}{V} = |\gamma_\ell| = |2 b_\ell - 1| \).

- Trade Classification determines \( V^B_\ell \), and thus, in turn, \( b_\ell, \gamma_\ell, \) and \( OI_\ell \).

- Clearly, \( 0 \leq OI_\ell \leq 1 \).
5. Computing (Bucket) SOI from Smaller “Bars”

- Given Bucket (drop subscript $\ell$), Split into $Q$ units, or “Bars.”

- $V = V_1 + \ldots + V_Q$, and $\nu_q = V_q / V$.

- $b_q = V^B_q / V_q$, and $\gamma_q = 2b_q - 1$.

- $OI_\ell = \left| \sum_{q=1}^{Q_\ell} \left( V^B_{q,\ell} - V^S_{q,\ell} \right) \right| / V = \left| \sum_{q=1}^{Q_\ell} \gamma_{q,\ell} \cdot \nu_{q,\ell} \right|$.

- OI is Absolute Value of Volume-Weighted SOI over Bars in Bucket.

- Ultimately, Trade Classification performed over the Bars.
6. Defining Time and Volume Bars

- Bars defined via Fixed Increments in Calendar Time are labeled **Time Bars**.
  - Let $\delta$ be Fixed Length of Calendar Time Interval in Seconds.
  - Volume Buckets Contain Randomly Varying Number of Time Bars.
  - In Time Bars, $Q$ Reflects (Inversely) the Trading Intensity.

- Bars defined as Fixed Increment in Trading Volume are labeled **Volume Bars**.
  - Let $\nu$ be (Fixed) Proportion of Volume Bucket represented by Volume Bar.
  - Volume Bar contains $\nu \cdot V$ Traded Contracts.
  - $\frac{1}{V} \leq \nu \leq 1$, and $1 \leq Q \leq V$. 

7. Volume Aggregation Prior to Trade Classification

- Classification from More (Less) Aggregated Volume if Bar is Longer (Shorter).
- Large $\delta$ or $\nu$ Increases VPIN Measure.
- Diversification of SOI Measures within Bucket is Critical for OI Measure.
- Variation in Number of Bars Systematically Impact/Distort Properties.
- For Volume Bars, Direct Control over Diversification (Event Time Measure)
  - Contract-by-Contract Classification: $\nu = \frac{1}{V}$ and $Q = \frac{1}{\nu} = V$.
  - Bucket-Wide Classification: $\nu = 1$ and $Q = 1$.
  - Regular (Interim) Scenario: $\frac{1}{V} < \nu < 1$ and $1 < Q < V$.
- $OI = \frac{\gamma_1,\ell + \cdots + \gamma_Q,\ell}{Q}$.
8.B Tick Rule Classification (on Bulk Volume)

- Each Time Bar Classified as Buyer or Seller Initiated, using “Tick” Rule
- The $j^{th}$ Time Bar is Buy ($b_j = 1$), if either
  
  (I) $P_j > P_{j-1}$, or
  (II) $P_j = P_{j-1}$ and Previous Transaction is Buy

- Otherwise, Transaction is Sell ($b_j = 0$).

- Important: Entire Time Bar, rather than each Individual Transaction, Classified as Buy or Sell; Effectively, **Whole Time Bar Treated as Single Transaction**
8.C Bulk Volume Classification

Assign Buy Volume as Function of Price Change over Bar.

Letting $Z(\cdot)$ denote CDF of Standard Normal Variate,

$$\gamma_j = 2Z\left(\frac{\Delta P_j}{\bar{\sigma}}\right) - 1,$$

where $\bar{\sigma}$ is Sample (Unconditional) Standard Deviation of $\Delta P_j$.

BVC interprets Zero Price Change as Balanced Trading: $\gamma_j = 0$.

Large Positive (Negative) Price Change Translates into Large (Small) Proportion of Buy Volume, $\gamma_j \approx 1 (-1)$.

Notice: When Volume and Volatility High, $\gamma_j \approx -1$ or $1$. 
Main Features of VPIN:

- The Volume Bucket, $V$
- The Moving Average Parameter, $L$
- Transactions, Time Bars, Volume Bars
- Data Aggregation, $\delta$ or $\nu$
- The Trade Classification Rule, $b$
- Others

Implementation: $V = (1/50)^{th}$ of Average Daily Volume and $L = 50$

We Investigate:

- Tick Rule on Transactions,
- Time Bars, $\delta = 1, 10, 60, 300$ seconds,
- Volume Bars, $\nu = (0.02, 0.10) \cdot V$, (50 or 10 Volume Bars per Bucket)
ILLUSTRATING VPIN BUCKETING

1-minute Volume, 11:30–12:00

1-minute Volume, 13:30–14:00

TORENB. ANDERSEN, OLEG BONDARENKO

VPIN UNDER IDEAL TRADE CLASSIFICATION
**NOTATION**

**First Index Letter** denotes **Type of Data Aggregation**:

(R) Individual “Real” Trades based on tRansaction level data;

(T) “Time bars” based on fixed increments to calendar time;

(V) “Volume bars” based on fixed increments to trading volume;

**Second Index Letter** refers to **Trade Classification** Rule:

(A) “True” Trade Classification, based on Trade at Bid or Ask;

(B) Tick Rule Classification, as in ELO (2011a, 2011b, 2011c);

(C) Bulk-Volume Classification, as in ELO (2012a); **Unconditional** $\sigma_{\Delta P}$;

(D) Same as C, but **Conditional** $\sigma_{\Delta P}$ (1 week rolling window);

(E) Uninformative, using $L_2$ norm, U2-VPIN; using actual $\nu_{q,\ell}$.

**Third Index Letter** for **Aggregation Level** (for $\delta$ or $\nu$):

(1-4) For $\delta = 1, 10, 60, 300$ seconds;

(1-2) For $\nu = 0.02$ or $0.10 \cdot V$ contracts;
Tick Rule-VPIN and the Flash Crash

Price on 06-May-2010

VPIN, $\delta = 0, 10$ and 60 sec

VIX

Volume

Torben G. Andersen, Oleg Bondarenko

VPIN under Ideal Trade Classification
Suppose there are $Q$ Bars in Volume Bucket, with Number of Contracts Traded $V_1, \ldots, V_Q$, so $V_1 + \ldots + V_Q = V$.

Suppose these $Q$ Bars are Randomly Classified as Buys or Sells. Specifically, let $\gamma_1, \ldots, \gamma_Q$ be i.i.d. Trade Indicators, equal to $\pm 1$ with same Probability.

Corresponding Order Imbalance is

$$\text{OI} = \left| \frac{V_1}{V} \gamma_1 + \ldots + \frac{V_Q}{V} \gamma_Q \right| = \left| \nu_1 \gamma_1 + \ldots + \nu_Q \gamma_Q \right|. $$

Under Assumption of Equal Volume Sizes, $\nu_1 = \ldots = \nu_Q = \frac{1}{Q}$, Expected Order Imbalance $E[\text{OI}]$ can be derived Analytically,

$$E[\text{OI}] = F(Q) = \frac{(2q)!}{2^{2q} q! q!}, \quad \text{if} \quad Q = 2q, \quad \text{or} \quad Q = 2q + 1.$$
Expected Order Imbalance Function $F(Q)$

$$E[OI] \equiv F(Q) = \frac{(2q)!}{2^{2q} q! q!},$$

if $Q = 2q$, or $Q = 2q + 1$.

$$F(Q) \sim \sqrt{\frac{2}{\pi Q}},$$

for large $Q$. 
**Uninformed VPIN Measures: U1-VPIN and U2-VPIN**

- $F(Q)$ is Used to Construct the U1-VPIN Measure as:

$$U1-VPIN_t = \frac{1}{L} \sum_{\ell=1}^{L} F(Q_{\ell}),$$

where $Q_{\ell}$ is Number of Bars included in Volume Bucket Ending at $\tau_{\ell}$.

- U1-VPIN is “Uninformed” - independent of Price Information Variation Cannot be Attributed to Order Flow “Toxicity.”


**Key Question:** Does the Trade Classification provide Information beyond Observable Trading Pattern?
When the Weights $\nu_1, \ldots, \nu_Q$ are not equal, $E[OI]$ is Untractable. However, we can compute the $L^2$ Norm instead of the $L^1$ Norm:

$$\sqrt{E[OI^2]} = \sqrt{E[(\nu_1 b_1 + \cdots + \nu_Q b_Q)^2]} = \sqrt{\nu_1^2 + \cdots + \nu_Q^2} = |\nu|.$$ 

Therefore, we construct the **U2-VPIN Metric** as:

$$\text{U2-VPIN}_t = \frac{1}{L} \sum_{\ell=1}^{L} |\nu_\ell|,$$

where $\nu_\ell$ is Weight Vector in Volume Bucket ending at $T_\ell$. 
T-VPIN Highly Sensitive to Trading Activity:

As Trading Volume Rises, Number of Time Bars per Volume Bucket (Average $Q$) Decreases, mechanically Increasing T-OI and T-VPIN.

Approach Mixes Two Types of Clocks: Calendar Clock for Time Bars and Trade Clock for Volume Buckets.

To Break mechanical effect of Trading Activity, we Aggregate Transaction Data into Equal-Sized Volume Bars.

Volume Bar Measures provide Control for Volume Distortion.

Analogous to U-VPIN Control for Impact of Trade Classification.

Key Question: Does the Trade Classification provide Information beyond Trading Intensity – Does It Work in Event Time?
S&P 500, VIX AND TRADING VOLUME

Figure: This figure depicts daily values of the S&P 500 index, VIX, the daily volume, and 21-day moving average of the daily volume over our full sample, February 10, 2006 - March 22, 2011. The dashed line in the bottom panel shows the average daily volume over the whole sample.
Daily Maximum: TC3

Empirical CDF of Daily Maximum: TC3
Daily Maximum: TC3

Empirical CDF of Daily Maximum: TC3
Illustration: Actual Trade Sequence: BBBSSSSSB.

Actual Rule (A): \((1, 1, 1, 0, 0, 0, 0, 1)\)
Candidate Rule (K): \((1, 0, 1, 0, 1, 0, 1, 0)\)

For \(\nu = 1/8\), 50% Misclassification.
For \(\nu = 1/4\), 25% Misclassification.
For \(\nu = 1/2\), 25% Misclassification.
For \(\nu = 1\), 0% Misclassification.

Reflects Diversification (Aggregation) Result:

\[
\sum_{q=1}^{Q} |\hat{V}_q^B - V_q^B| \leq \left| \sum_{q=1}^{Q} (\hat{V}_q^B - V_q^B) \right|
\]

Always Compare Precision at Identical Aggregation Level!
Volume Buckets

\[ \text{MVB}_\ell = \frac{\sum_{q=1}^{Q} |\hat{v}_q^B - v_q^B|}{V} = \sum_{q=1}^{Q} |\hat{b}_q - b_q| \cdot v_q. \]

Simplification for Volume Bars: \( \text{MVB} = \frac{1}{Q} \sum_{q=1}^{Q} |\hat{b}_q - b_q| \).

Sample-Wide Measure:

\[ \text{MVB} = \frac{1}{L} \sum_{\ell=1}^{L} \text{MVB}_\ell. \]

Contract-by-Contract

\[ \text{MCC} = \frac{1}{I} \sum_{i=1}^{I} |\hat{b}_i - b_i|. \]
## Error Rates for Alternative Classification Schemes

### Panel A: MCC

<table>
<thead>
<tr>
<th>Rule</th>
<th>R</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>V1</th>
<th>V2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0.248</td>
<td>0.330</td>
<td>0.387</td>
<td>0.431</td>
<td>0.339</td>
<td>0.388</td>
</tr>
<tr>
<td>B</td>
<td>0.116</td>
<td>0.283</td>
<td>0.242</td>
<td>0.232</td>
<td>0.245</td>
<td>0.233</td>
<td>0.170</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>0.282</td>
<td>0.242</td>
<td>0.240</td>
<td>0.261</td>
<td>0.228</td>
<td>0.169</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>0.282</td>
<td>0.242</td>
<td>0.240</td>
<td>0.261</td>
<td>0.228</td>
<td>0.169</td>
</tr>
</tbody>
</table>

### Panel B: MVB

<table>
<thead>
<tr>
<th>Rule</th>
<th>R</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>V1</th>
<th>V2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0.038</td>
<td>0.072</td>
<td>0.141</td>
<td>0.264</td>
<td>0.065</td>
<td>0.121</td>
</tr>
<tr>
<td>B</td>
<td>0.023</td>
<td>0.040</td>
<td>0.045</td>
<td>0.086</td>
<td>0.172</td>
<td>0.041</td>
<td>0.043</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>0.040</td>
<td>0.045</td>
<td>0.086</td>
<td>0.172</td>
<td>0.041</td>
<td>0.043</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>0.040</td>
<td>0.045</td>
<td>0.086</td>
<td>0.172</td>
<td>0.041</td>
<td>0.043</td>
</tr>
</tbody>
</table>
ERROR RATES FOR ALTERNATIVE CLASSIFICATION SCHEMES

MCC: TB, TC, TD

MVB: TB, TC, TD

MCC: VB, VC, VD

MVB: VB, VC, VD
TIME VARIATION IN ERROR RATES (21-DAY MOVING AVERAGE)

MCC: RB, TC1, and TC3

MVB: RB, TC1, and TC3

MCC: RB, VC1, and VC2

MVB: RB, VC1, and VC2
TIME SERIES OF DAILY MAX VPIN VALUES

RA and RB

TC1 and TD1

TC3 and TD3

VC2 and VD2

TORBEN G. ANDERSEN, OLEG BONDARENKO
VPIN UNDER IDEAL TRADE CLASSIFICATION
## R-VPIN and V-VPIN Correlations

### Transaction-Based Classification (R-VPIN)

<table>
<thead>
<tr>
<th></th>
<th>RA</th>
<th>RB</th>
<th>Volume</th>
<th>VIX</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>1.00</td>
<td></td>
<td>-0.53</td>
<td>-0.72</td>
<td>-0.62</td>
</tr>
<tr>
<td>RB</td>
<td>0.96</td>
<td>1.00</td>
<td>-0.53</td>
<td>-0.73</td>
<td>-0.64</td>
</tr>
</tbody>
</table>

### Volume Bar-Based Classification (V-VPIN)

#### \( \nu = 0.02 \)

<table>
<thead>
<tr>
<th></th>
<th>VB</th>
<th>VC</th>
<th>VD</th>
<th>RA</th>
<th>RB</th>
<th>Volume</th>
<th>VIX</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.84</td>
<td>0.86</td>
<td>-0.45</td>
<td>-0.63</td>
<td>-0.53</td>
</tr>
<tr>
<td>VC</td>
<td>-0.59</td>
<td>1.00</td>
<td></td>
<td>-0.66</td>
<td>-0.71</td>
<td>0.69</td>
<td>0.76</td>
<td>0.80</td>
</tr>
<tr>
<td>VD</td>
<td>-0.17</td>
<td>0.72</td>
<td>1.00</td>
<td>-0.20</td>
<td>-0.25</td>
<td>0.51</td>
<td>0.32</td>
<td>0.45</td>
</tr>
</tbody>
</table>

#### \( \nu = 0.10 \)

<table>
<thead>
<tr>
<th></th>
<th>VB</th>
<th>VC</th>
<th>VD</th>
<th>RA</th>
<th>RB</th>
<th>Volume</th>
<th>VIX</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.65</td>
<td>0.66</td>
<td>-0.28</td>
<td>-0.44</td>
<td>-0.35</td>
</tr>
<tr>
<td>VC</td>
<td>-0.32</td>
<td>1.00</td>
<td></td>
<td>-0.65</td>
<td>-0.69</td>
<td>0.68</td>
<td>0.74</td>
<td>0.78</td>
</tr>
<tr>
<td>VD</td>
<td>0.36</td>
<td>0.42</td>
<td>1.00</td>
<td>0.22</td>
<td>0.17</td>
<td>0.25</td>
<td>0.01</td>
<td>0.19</td>
</tr>
</tbody>
</table>

*References: Torben G. Andersen, Oleg Bondarenko, *VPIN Under Ideal Trade Classification*
## T-VPIN Correlations

### Time Bar-Based Classification (T-VPIN)

#### $\delta = 1$

<table>
<thead>
<tr>
<th></th>
<th>TB</th>
<th>TC</th>
<th>TD</th>
<th>TE</th>
<th>RA</th>
<th>RB</th>
<th>Volume</th>
<th>VIX</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>TB</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC</td>
<td>0.71</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.63</td>
<td>0.62</td>
<td>-0.08</td>
<td>-0.39</td>
<td>-0.24</td>
</tr>
<tr>
<td>TD</td>
<td>0.74</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td>0.66</td>
<td>0.66</td>
<td>-0.13</td>
<td>-0.44</td>
<td>-0.31</td>
</tr>
<tr>
<td>TE</td>
<td>0.85</td>
<td>0.58</td>
<td>0.62</td>
<td>1.00</td>
<td>0.83</td>
<td>0.87</td>
<td>-0.48</td>
<td>-0.72</td>
<td>-0.64</td>
</tr>
</tbody>
</table>

#### $\delta = 10$

<table>
<thead>
<tr>
<th></th>
<th>TB</th>
<th>TC</th>
<th>TD</th>
<th>TE</th>
<th>RA</th>
<th>RB</th>
<th>Volume</th>
<th>VIX</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>TB</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC</td>
<td>0.71</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.35</td>
<td>0.30</td>
<td>0.30</td>
<td>-0.04</td>
<td>0.15</td>
</tr>
<tr>
<td>TD</td>
<td>0.85</td>
<td>0.94</td>
<td>1.00</td>
<td></td>
<td>0.60</td>
<td>0.56</td>
<td>0.03</td>
<td>-0.30</td>
<td>-0.11</td>
</tr>
<tr>
<td>TE</td>
<td>0.84</td>
<td>0.59</td>
<td>0.76</td>
<td>1.00</td>
<td>0.76</td>
<td>0.76</td>
<td>-0.16</td>
<td>-0.53</td>
<td>-0.38</td>
</tr>
</tbody>
</table>

#### $\delta = 60$

<table>
<thead>
<tr>
<th></th>
<th>TB</th>
<th>TC</th>
<th>TD</th>
<th>TE</th>
<th>RA</th>
<th>RB</th>
<th>Volume</th>
<th>VIX</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>TB</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TC</td>
<td>0.65</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.06</td>
<td>0.01</td>
<td>0.55</td>
<td>0.21</td>
<td>0.38</td>
</tr>
<tr>
<td>TD</td>
<td>0.87</td>
<td>0.87</td>
<td>1.00</td>
<td></td>
<td>0.47</td>
<td>0.43</td>
<td>0.22</td>
<td>-0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>TE</td>
<td>0.84</td>
<td>0.69</td>
<td>0.87</td>
<td>1.00</td>
<td>0.55</td>
<td>0.52</td>
<td>0.16</td>
<td>-0.28</td>
<td>-0.10</td>
</tr>
</tbody>
</table>

**Authors:** Torben G. Andersen, Oleg Bondarenko
Why such High VC1- and VC2-VPIN (TC3-VPIN) Correlation with RV?

Only Possible Reason is Trade Classification (otherwise Identical to other indices)

\[
\text{VC-VPIN} = \frac{1}{L} \sum_{\ell=1}^{L} \left| \frac{1}{Q} \sum_{q=1}^{Q} \gamma_{q,\ell} \right| = \frac{1}{L} \sum_{\ell=1}^{L} |\gamma_{\ell}|.
\]

Cumulates Absolute Values of HF Increments Linked to Price Changes!

Resembles (Distorted) Realized Volatility Measure
Why such High VC1- and VC2-VPIN (TC3-VPIN) Correlation with RV?

In Fact, “Bulk Return” \( r^b_\ell = \gamma_\ell \) Highly Correlated with Log-Returns.

Average Correlation over Volume Buckets: \( 0.86, 0.84 \) for VC1, VC2 Schemes.

VC-VPIN Measures Trivially Strongly Correlated with RV (and Volume).

TC-VPIN Measures will be Correlated with both RV and Volume.

VD-VPIN Controls for Recent RV – only RV Innovations Matter.
## Panel A: 5-Minute Forecast

<table>
<thead>
<tr>
<th>Reg</th>
<th>Const</th>
<th>RA</th>
<th>RB</th>
<th>TB3</th>
<th>TC3</th>
<th>TD3</th>
<th>TE3</th>
<th>Vol</th>
<th>VIX</th>
<th>RV</th>
<th>$\bar{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.10</td>
<td>-0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19.71</td>
</tr>
<tr>
<td></td>
<td>(25.50)</td>
<td>(-17.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.11</td>
<td>-0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.35</td>
</tr>
<tr>
<td></td>
<td>(25.43)</td>
<td>(-18.47)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>0.06</td>
<td>-0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>(14.90)</td>
<td>(-6.25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>-0.02</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9.91</td>
</tr>
<tr>
<td></td>
<td>(-4.03)</td>
<td>(13.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>0.03</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(9.16)</td>
<td>(2.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>0.05</td>
<td>-0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>(9.66)</td>
<td>(-2.71)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>-0.01</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33.42</td>
</tr>
<tr>
<td></td>
<td>(-4.27)</td>
<td>(19.66)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8)</td>
<td>-0.01</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48.42</td>
</tr>
<tr>
<td></td>
<td>(-8.31)</td>
<td>(30.66)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9)</td>
<td>0.00</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61.35</td>
</tr>
<tr>
<td></td>
<td>(3.17)</td>
<td>(98.96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10)</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>61.35</td>
</tr>
<tr>
<td></td>
<td>(1.81)</td>
<td>(54.54)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Panel B: 1-Day Forecast

<table>
<thead>
<tr>
<th>Reg</th>
<th>Const</th>
<th>RA</th>
<th>RB</th>
<th>TB3</th>
<th>TC3</th>
<th>TD3</th>
<th>TE3</th>
<th>Vol</th>
<th>VIX</th>
<th>RV</th>
<th>$\bar{R}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.10</td>
<td>-0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33.30</td>
</tr>
<tr>
<td></td>
<td>( 26.14)</td>
<td>(18.34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>0.11</td>
<td>-0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>37.40</td>
</tr>
<tr>
<td></td>
<td>( 25.98)</td>
<td>(18.84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>0.07</td>
<td>-0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td>( 16.27)</td>
<td>(-7.82)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>-0.01</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.87</td>
</tr>
<tr>
<td></td>
<td>(-1.94)</td>
<td>(12.14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>0.04</td>
<td>-0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>( 12.06)</td>
<td>(-0.20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>0.06</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>( 11.48)</td>
<td>(-4.49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>-0.01</td>
<td></td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>48.02</td>
</tr>
<tr>
<td></td>
<td>(-2.66)</td>
<td>(18.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8)</td>
<td>-0.01</td>
<td></td>
<td></td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>77.85</td>
</tr>
<tr>
<td></td>
<td>(-8.33)</td>
<td>(32.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9)</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>82.30</td>
</tr>
<tr>
<td></td>
<td>( 0.56)</td>
<td>(26.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10)</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>82.30</td>
</tr>
<tr>
<td></td>
<td>( 0.06)</td>
<td>(24.60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
THE FLASH CRASH

VPIN: RA and RB

VPIN: TC1, TC2, and TC3

VPIN: TB1, TB2, and TB3

VPIN: TE1, TE2, and TE3
ELO Metric for Evaluating VPIN

- VPIN Signals Impending Market Turbulence – NO, just reflects Volume, RV
- VPIN Predicts Future Short Run Volatility – NO, Subsumed by RV
- VPIN Leads VIX – NO, VIX Superior Volatility Predictor
- BVC Accuracy Superior – NO, Poor and Errors Correlate w/ Activity
**Features of VPIN Constructed from Ideal Classification (or RB):**

- Negative Correlation with Volatility and Volume.
- High Uncertainty → Thinning Order Book, Smaller Trade Sizes.
- Negative Correlation with TC3-VPIN.
- No Signal Effect for Crashes.
- Information in RA-SOI Destroyed by VPIN-Transformation.
- Is Flat across Trading Day up to Closing.

Classification Errors of Bar-Based VPIN Correlated with Volatility.
CONCLUSIONS

- Take In-Depth Look at Performance of TB-VPIN and TC-VPIN Measures.
- Contrast Performance to VPIN under Perfect Trade Classification.
- Also Contrasting to U-VPIN, RB-VPIN, xD-VPIN, and Vx-VPIN.
- ELO VPIN Positively Correlated with Volatility by Construction.
- Classification Errors Correlate with Volume/Volatility.
- No Auxiliary Forecast Power beyond known Activity Variables.
- No Formal Metric used for Evaluation.
- Association to Actual Trade Direction Broken.
- By What Formal Metric Does VPIN Help with Anything?
CONCLUSIONS

- Level of VPIN prior to Flash Crash Not Extraordinary.

- Level of VPIN (near) Record High following Crash.

- Increase in VPIN driven by Concurrent Volume/Volatility Innovations.

- Volume Bar Tick-Rule (VB-) VPIN Annihilates Results.

- We hope the Battery of Tests Undertaken here will be Helpful in Identifying more Suitable and Robust Market Stress Indicators.