

Lawrence Berkeley
National Laboratory

Testing VPIN on Big Data
-- Response to "Reflecting on the VPIN Dispute"

Kesheng Wu, E. Wes Bethel, Ming Gu, David Leinweber, and Oliver Ruebel

Lawrence Berkeley National Laboratory

One Cyclotron Road Berkeley, CA 94720

8/30/2013

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Testing VPIN on Big Data

– Response to "Reflecting on the VPIN Dispute"

Kesheng Wu, E. Wes Bethel, Ming Gu, David Leinweber, and Oliver Ruebel

The authors of this note are computer scientists and mathematicians from Lawrence Berkeley National Laboratory. Most of us are specialists in high-performance computing. In 2011, at the suggestion of a CFTC commissioner, we started examining the effectiveness of VPIN as a leading indicator of unusual market events such as the "Flash Crash" of 2010. The preliminary result of that work was published as a conference paper[1]. Our work was the first to show that VPIN could generate strong signals based on trading records of individual stocks [2]. On a number of stocks, VPIN generated strong signals more than an hour before the "Flash Crash." This lead-time could be valuable for market makers, regulators and traders in general. To further test the effectiveness of VPIN, we recently expanded our tests to a much larger set of test data, including the trading records of the 94 most liquid futures contracts from beginning of 2007 to middle of 2012. We set out to verify the following hypothesis: the volatility of trading is higher than the average in a time window (we call the event horizon) after VPIN value crosses a threshold. Our work carefully examined the parameters controlling the computation of VPIN as well as the event horizon and the VPIN threshold. In this largest-ever study of a leading indicator based on market microstructure, we examined 16,000 combinations of these parameters, and found that the average false positive rate can be reduced to 7% [3].

In their note titled "Reflecting on the VPIN Dispute," Andersen and Bondarenko [4] (AB) questioned the validity of using the false positive rate to measure the effectiveness of VPIN. However, their objection was based on an imperfect definition of the false positive rate, not taking into account the fact that we have addressed the key shortcoming of that definition. AB argued that VPIN was ineffective because it would produce too many events, but our report showed that the average number of events is quite modest on a large variety of data. They also raised a number of non-technical issues about our work. In the interest of maintaining the integrity of the research record, we would like to address their points.

False Positive Rate

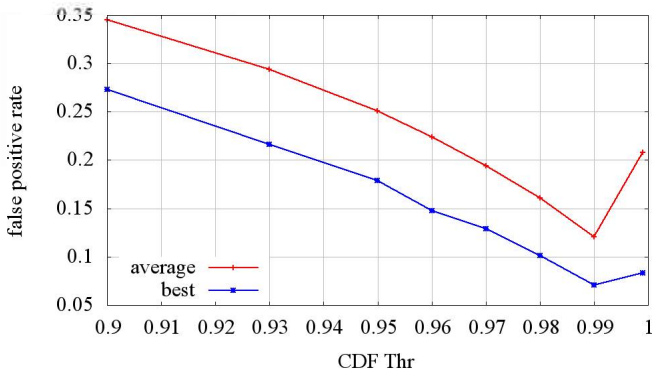
On page 8 of their note [4], AB provided the following table as evidence that the false positive rate should not be trusted.

Volatility percentile	70	80	90	95	99
False Positive rate	10.8	3.0	1.2	0.0	0.0

A false positive rate is the number of false positive events divided by the number of total events. In the above table, the 0 false positive rates are computed from cases where no events have volatility above the specified percentiles. Each of these test

cases has zero total events and zero false positive events, however 0 divided by 0 should not be recorded as 0. In our report, we have discussed how to avoid this division by 0 problem.

To avoid this division by 0 problem, whenever a test case produces no event, we would record the total number of events as 0.5, and at the same time record the number of false positive events as 0.5 as well. This set up is explained in the last paragraph on page 10 of our report. When no event is detected, we would report a false positive rate of 1, and not 0, as AB have done in the above table. Therefore, the false positive rates in our tests would increase when the threshold is set too high as illustrated in the figure on the right, which is also Figure 5a of our report [3]. In short, our false positive rate is mathematically sound.

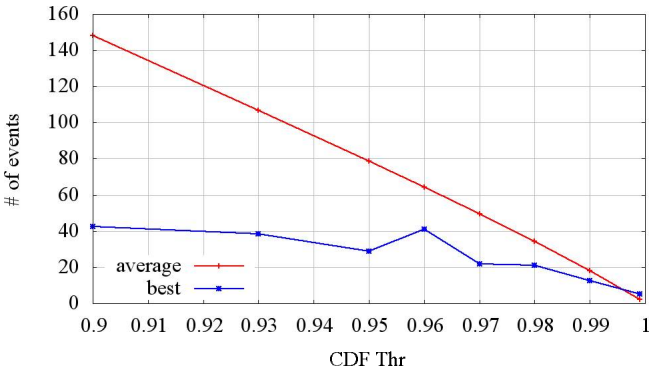


Furthermore, the false positive rates we reported were computed from a diverse set of trading instruments including equity, energy, precise metal, commodity, and interest rates among others. The 7% false positive rate was averaged over 94 futures contracts across 66 months. It is indeed a low false positive rate given the diversity and the duration of the data.

Number of VPIN Events

The first main point presented by AB in [4] could be boiled down to "using a CDF threshold leads to too many VPIN events." This assumption of "too many events" led AB to seek alternative criteria for declaring VPIN events, which in turn gave AB reasons to conclude that VPIN was ineffective. However, our tests on a variety of data and under a wide-range of conditions show that VPIN actually does not lead to "too many events."

In the first paragraph of Section 2 of their note [4], AB argued that there would be a VPIN event every two days using a CDF threshold of 0.99. However, this is not the case as shown in Figure 5b of our report [3] (also reproduced on the right). With 0.99 as the CDF threshold, the average number of VPIN events is less than 20 for the entire 66-month period from 2007 to 2012. Changing the CDF threshold to 0.9 does not significantly change the number of events either; the average number of events is about 40 with a good set of control parameters for VPIN (blue line) and about 140 with a random set of control parameters (red line).



By the definition of CDF, there is $(1-\tau)$ fraction of VPIN values with CDF greater than τ. This appears to be the basis of AB's expectation on the number of VPIN events. However, the VPIN values are computed on a time series, the values from nearby time buckets are highly correlated with each other. At the onset of a VPIN event, the CDF of VPIN crosses over the threshold. The next VPIN value will very likely be also above the threshold. It is common for VPIN values to keep on increasing for many subsequent time buckets. However, as long as these VPIN values are within the event horizon, we count them all as one event. Therefore, one VPIN event generally includes many time buckets with high VPIN values. Thus, the actual observed number of VPIN events is much less than AB had expected.

Comments on non-technical issues raised by AB.

AB first raised but then left aside the question on the independence of our work. In their note, AB began their last paragraph on page 2 with:

"A fourth issue is the findings of related independent research. ... ELO emphasize the studies by authors affiliated with the Lawrence Berkeley National Laboratories. Marcos Lopez de Prado is a Research Affiliate of this institution."

We emphatically refute any direct or indirect assertion that our work was biased due to Dr. Lopez de Prado's affiliation. Here are the relevant facts. Our work on VPIN [2] was suggested by a third party more than a year before Dr. Marcos Lopez de Prado became an affiliate with our institution. He has not been directly involved in the VPIN related publications. Furthermore, our work has been purely focused on evaluating the technique on various data sets, not on developing the method itself.

In footnote 9 of their article, AB stated:

"We have already contacted the authors behind the VPIN replication studies at the Lawrence Berkeley National Laboratories with an offer to exchange our series. We have not yet received a response."

This is factually inaccurate. We responded to AB's initial email within 24 hours and the discussion on the exchange of data series and comparison of technical details was ongoing in July 2013—about a month before their note was made public.

Acknowledgements

This work is supported in part by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 with University of California.

References

1. Bethel, E.W., et al., *Federal Market Information Technology in the Post Flash Crash Era: Roles for Supercomputing*, in *WHPCF*2011, ACM: New York, NY, USA. p. 23-30.

2. Bethel, E.W., et al., *Federal Market Information Technology in the Post-Flash Crash Era: Roles for Supercomputing.* The Journal of Trading, 2012. **7**(2): p. 9-25.
3. Wu, K., et al. *A Big Data Approach to Analyzing Market Volatility*. 2013.
4. Andersen, T.G. and O. Bondarenko, *Reflecting on the VPIN Dispute.* Journal of Financial Markets, (0).