# An Improved Version of the Volume-Synchronized Probability of Informed Trading (VPIN)*:* A Comment

David Easley, Cornell University

Marcos Lopez de Prado, Guggenheim Partners

Maureen O'Hara, Cornell University

Ke and Lin, hereafter referred to as KL, provide a valuable alternative approach to estimating the Volume-Synchronized Probability of Informed Trading (VPIN) measure that we introduced in a series of papers, Easley et al. (2011, 2012). VPIN is our modification of the PIN measure introduced in Easley et al. (1996, 1997) to make it applicable to high frequency markets. PIN is a measure of the fraction of trade that is information-based and it is based on a model in which information events are assumed to happen regularly in calendar time; in applications this is usually once per trading day. VPIN implicitly assumes that information events happen regularly in volume time rather than in calendar time; once per volume bucket which is some fixed number of shares traded. VPIN is related to PIN, in that they both are attempts to create measures of the importance of information-based trade, but they are of course different measures.

Direct estimation of PIN requires estimation of the parameters underlying the trading process as PIN is function of those parameters. Computing the likelihood function that is maximized over the parameters underlying PIN for high frequency data is difficult if not impossible. Computable approximations of the likelihood function are possible, but instead of approximating the likelihood function we chose to directly approximate PIN when it is viewed in volume time. That is, one of our purposes in introducing VPIN was to avoid having to estimate the underlying parameters of the trading process. Instead VPIN is defined to be the ratio of the absolute value of trade imbalance to total trade in a volume bucket. Given buys and sells (more on this later) VPIN is easy to compute.

One limitation of our approach to VPIN is that we ignore the information in how long it takes to fill a volume bucket. How quickly volume buckets fill clearly carries information about the likelihood of informed traders being present; something that we have previously investigated as the information content of time between trades in Easley and O'Hara (1992). KL show how to use the information contained in the amount of time that it takes volume buckets to fill and as a result they create an estimator of VPIN that improves on our approach. They show this new volume time measure makes their modified VPIN a better predictor order flow toxicity.

We applaud KL for figuring out how to use this extra information and for applying their extension of our approach to the predictability of toxicity issue. The one disadvantage of their approach is that, like the old PIN procedure, it requires estimation of the parameters underlying the trading process which are then used to create KL's modified VPIN. For markets in which these parameters can be reliably estimated this seems both a valuable and viable modification.

All of these measures (PIN, VPIN and KL's modification of VPIN) require assignments of buys and sells either in time periods (PIN) or in volume buckets (VPINs). This was reasonably straightforward in times past when many if not most trades were between market makers and order flow coming from traders. In this case it seemed reasonable to view the traders as the active side of the market and to assign trade to be a buy if the market maker sold, or equivalently if the trade occurred at or near the ask price set by the market maker, and similarly to be a sell if the trade price was at or near the bid price. In modern electronic markets without designated market makers this distinction no longer applies. Some data sets do have an aggressor flag and it's tempting to use that flag to assign trade direction. But we show in Easley et al. (forthcoming) that neither of these assignment procedures is effective. Instead we propose a bulk classification method (which KL use) to assign trade direction and we show that it produces a reasonable classification. Much like our VPIN work we view this bulk classification procedure as a first step in dealing with an important and difficult problem.

## References:


Easley, D., N. Kiefer, and M. O'Hara, 1997, One Day in the Life of a Common Stock, *Review of Financial Studies*, Vol. 10, No. 3.

Easley, D., N. Kiefer, M. O'Hara, and J. Paperman, 1996, Liquidity, Information and Infrequently Traded Stocks, *Journal of Finance*, Vol. 51, No. 4.

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

Easley, D., M. Lopez de Prado, and M. O'Hara, 2012, Flow Toxicity and Volatility in a High Frequency World, *Review of Financial Studies*, 25, 5, 1457-93.

Easley, D., M. Lopez de Prado, and M. O'Hara, forthcoming, Discerning Information from Trade Data, *Journal of Financial Economics*.

Easley, D. and M. O'Hara, 1992, Time and the Process of Security Price Adjustment, *Journal of Finance*, Vol. XLVII, No. 2.