VPIN with Limit Order Book Data to Predict Jumps Induced by Toxic Flow

Chulin Liang[1,2], Henry Ip[1], Hugh Christensen[1]

1. BMLL Technologies Ltd, 5 Fleet Place, London, UK, EC4M 7RD

2. Dept of Electrical and Electronic Engineering, Imperial College, London, UK, SW7 2AZ

{chulinliang, henrymanden, hughchristensen}@bmlltech.com

Introduction

VPIN is the ratio of average unbalanced volume to total volume in a period [1]. Heuristically, the VPIN metric "measures the fraction of volume-weighted trade that arises from informed traders as the informed tend to trade on one-side of the market and so their activity leads to unbalanced volume." In other words, there might be an increasing toxicity of orders entering the market. Toxic flow is a generic term which refers to the fact that some orders in the market have more information than you and so are deemed "toxic". Toxic flow can't, of itself, cause a market crash. However HFT activity that magnifies market movements could cause price volatility and jumps, and so those orders could also be described as toxic. In this poster we explore the use of VPIN to predict price jumps.

VPIN Algorithm

In this work, we use execution data for the 40 most liquid futures on CME from 2018-01-01 to 2018-06-31. A key feature of the data is that all the executions are labelled in the exchange feed if they are buyer or seller initiated. Thus unlike much VPIN work there is no requirement to use a trade classification algorithm. In its basic form, the VPIN algorithm does not use any order level data. To implement VPIN, executions need to be resampled into equal volume buckets with a size of V and the metric is updated in volume time instead of calendar time. Compared to standard VPIN, two modifications are made: (1) we use the signed rather than unsigned order imbalance [2]; (2) we applied a moving z-score standardization to VPIN using exponential weighted moving average. Detailed steps of the computation are listed in Algorithm 1.

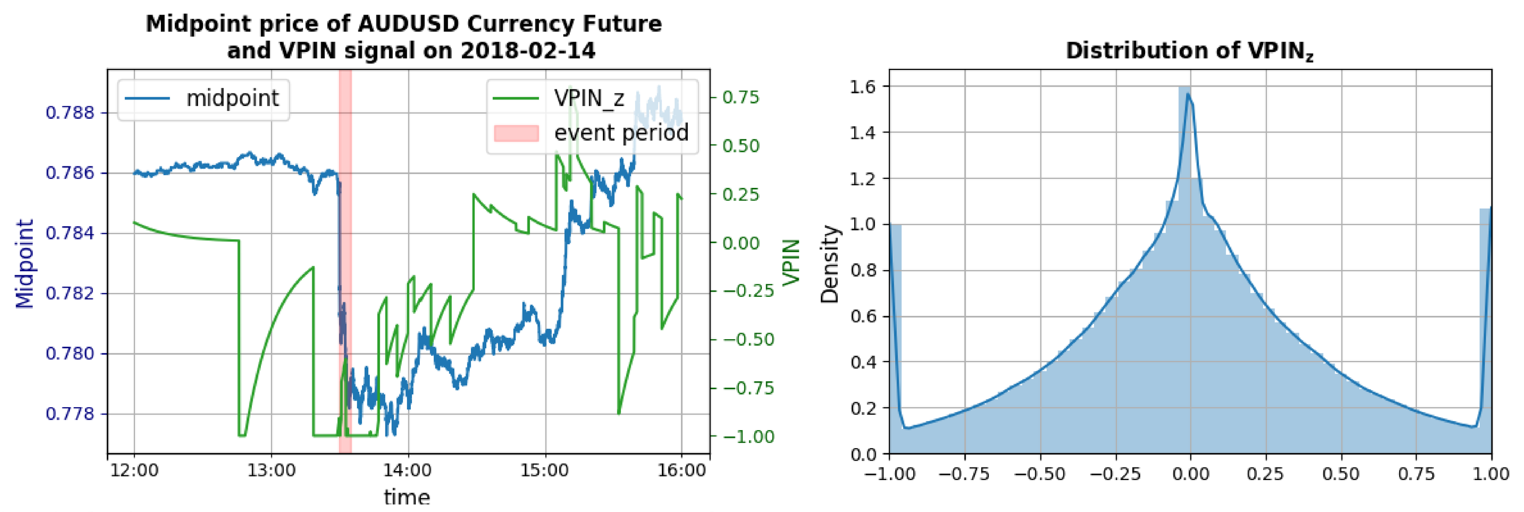
Alg. 1 Compute the VPIN metric

Data: Execution data containing time t, side (buy/sell) and size v.

Parameters: Number of bucket per day n, half-life of ewma mean t_μ, half-life of ewma volatility t_σ

1. Compute the averaged daily trades volume $\bar{V}$ of one security and set volume bucket size $V = \frac{1}{n}\bar{V}$.
2. For each observation, calculate the accumulated buy/sell volume at bucket τ, as well as the total accumulated volume $V_\tau = V_\tau^S + V_\tau^B$,
3. Once $V_\tau > V$, calculate the order imbalance at bucket τ: $OI_\tau = V_\tau^B - V_\tau^S$. The excess volume $V_\tau - V$ can be moved to the next bucket.
4. Repeat step 2 and 3 until $\tau >= n$, then compute the l_{th} VPIN estimate: $VPIN_l = \frac{\sum_{\tau=1}^{n} V_\tau^B - V_\tau^S}{nV}$
5. Discard the first bucket, update the index l and repeat step 2-4 to accumulate the next bucket.
6. For each security, first resample VPIN to 1 seconds per sample, then estimate the ewma mean μ_{VPIN} and ewma volatility μ_{VPIN} using t_μ and t_σ, finally apply the moving z-score with [-1,1] bound:

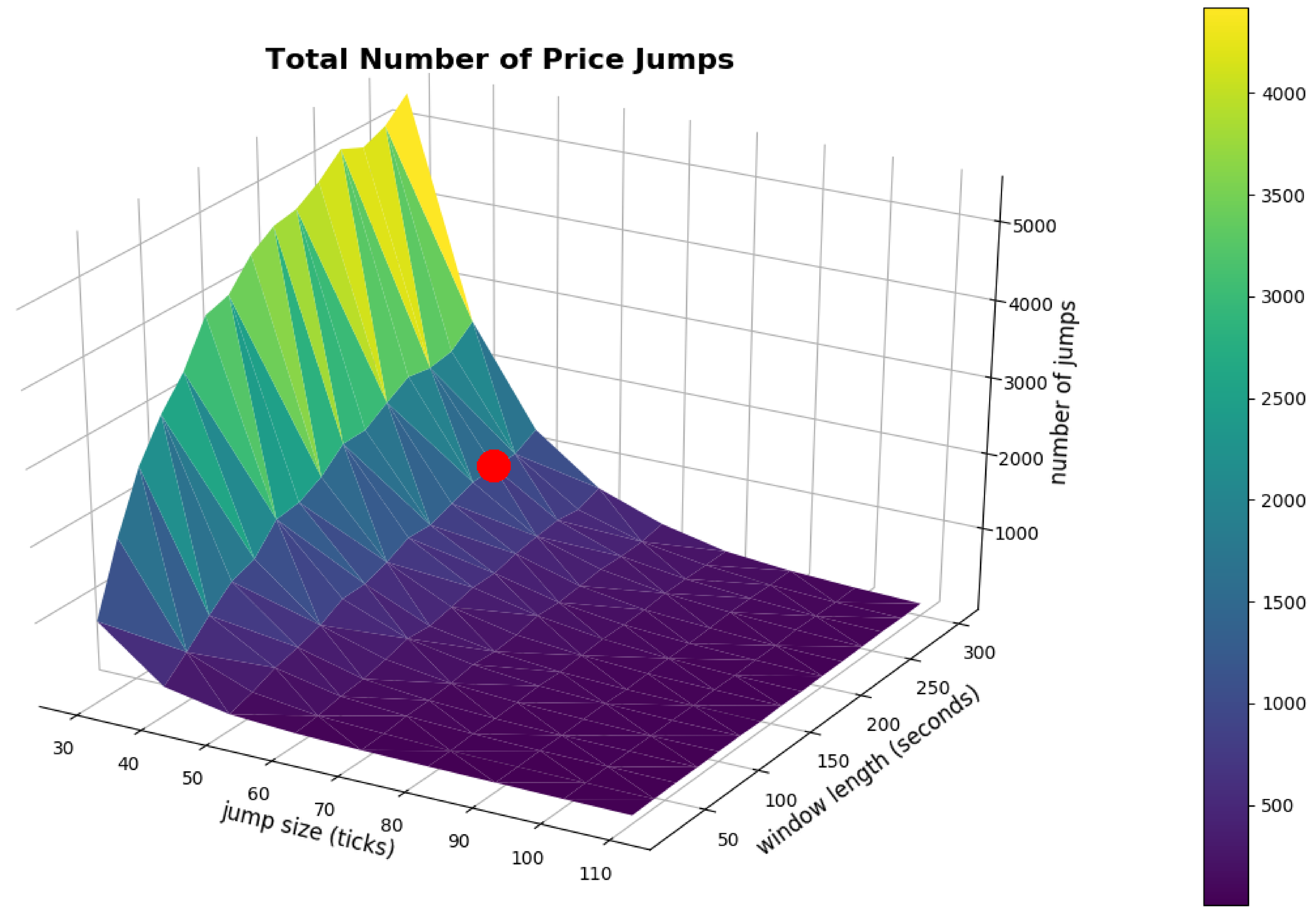
$$VPIN_z(n, t_\mu, t_\sigma) = max\Big(min\Big(\frac{VPIN - \mu_{VPIN}}{\sigma_{VPIN}}, -1\Big), 1\Big)$$



The left figure shows price and $VPIN_z$ of AUDUSD on 2018-02-14. A negative jump started at 13:30 and ended at 13:35, marked by the red shaded area. Note that the VPIN signal quickly drop to minus one before the event occur. The right figure shows the distribution of $VPIN_z$ for AUDUSD.

Price Jumps

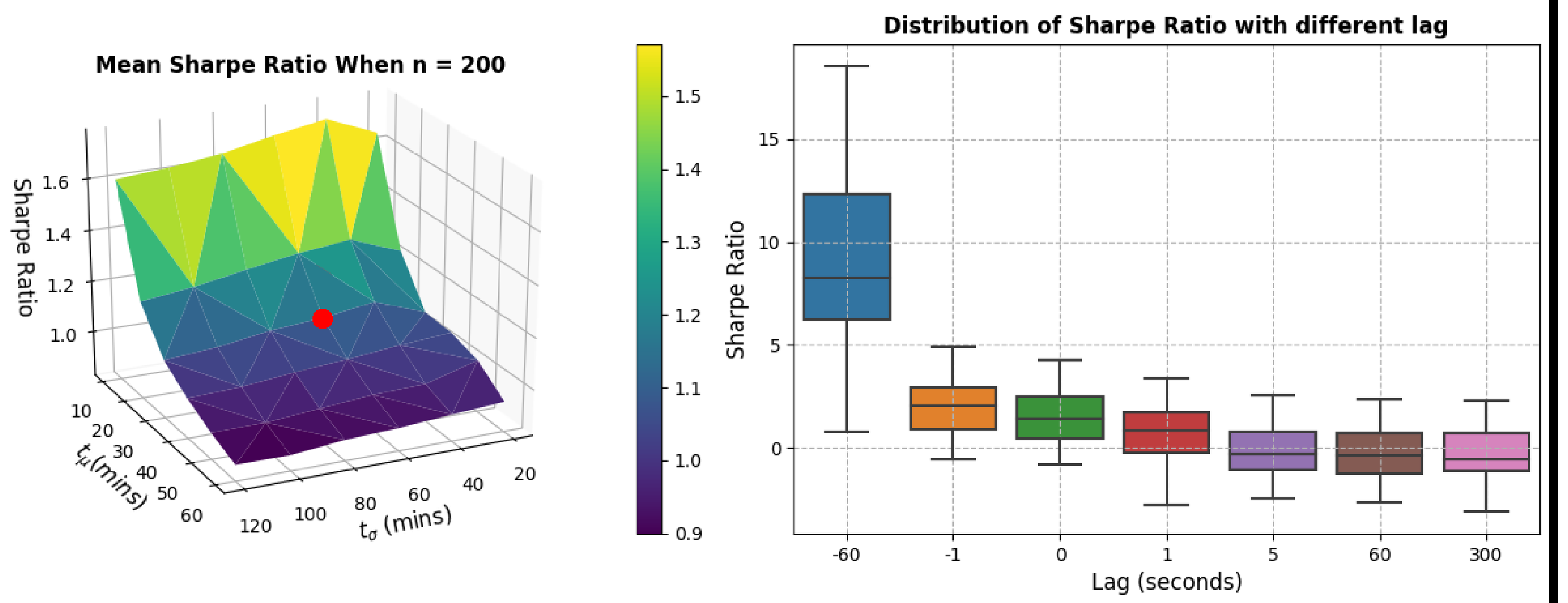
Jumps are defined as instantaneous and discrete moves in the price. In this work, we use approximate derivative of the midpoint price as an indicator. We first transform the price X_t into tick level price $\bar{X}_t$ by dividing its tick size and then compute the approximate derivative over a time window Δt: $f'(\bar{X}_t) = \frac{\bar{X}_t - \bar{X}_{t-\Delta t}}{t - \Delta t}$. We declare a jump event $J_{(i,t)}$ if $f'(\bar{X}_t) > M$, where M is the threshold we set.



The figure above shows the total number of jumps observed in the dataset with different M and Δ. For the rest of the analysis, we chose $M = 50$ and $\Delta t = 260s$ (marked by the red dot in the surface),which yield 1170 jumps (550 positive, 620 negative) in 6 months of data.

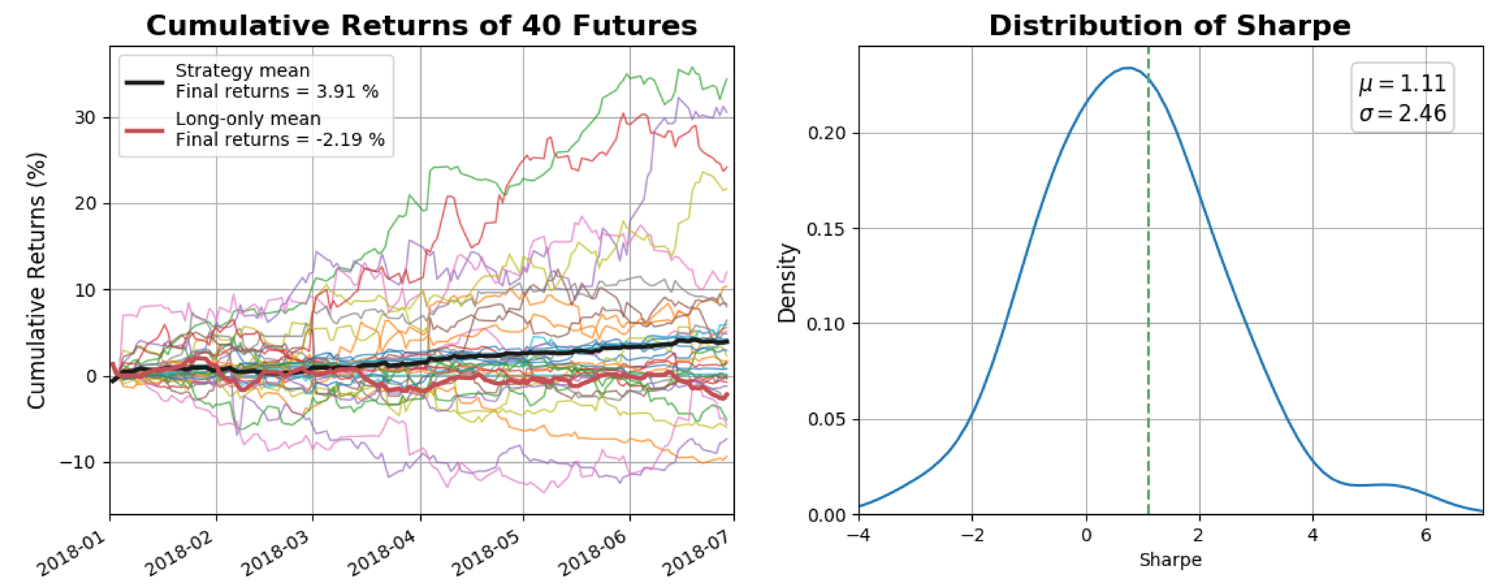
Parameter Optimization

We explored the relation between $VPIN_z$ and returns. Let $\mathscr{L}$ be the lag operator, then the strategy returns can be constructed by $SR = \mathscr{L}(VPIN_z, \tau = -1)\log(\frac{P_t}{P_{t-1}})$, where P is the midpoint price. We use annualized Sharpe $\sqrt{252}\left(\frac{\mu}{\sigma}\right)$ as the metric to optimise the parameters $\{n, t_\mu, t_\sigma\}$ described in Algorithm 1, and the results are shown in the left figure below. The right figure plots the Sharpe against different lag, where a negative lag means the price series is behind $VPIN_z$. The data we use has nanosecond time stamps and we chose to lag by one second going forwards.



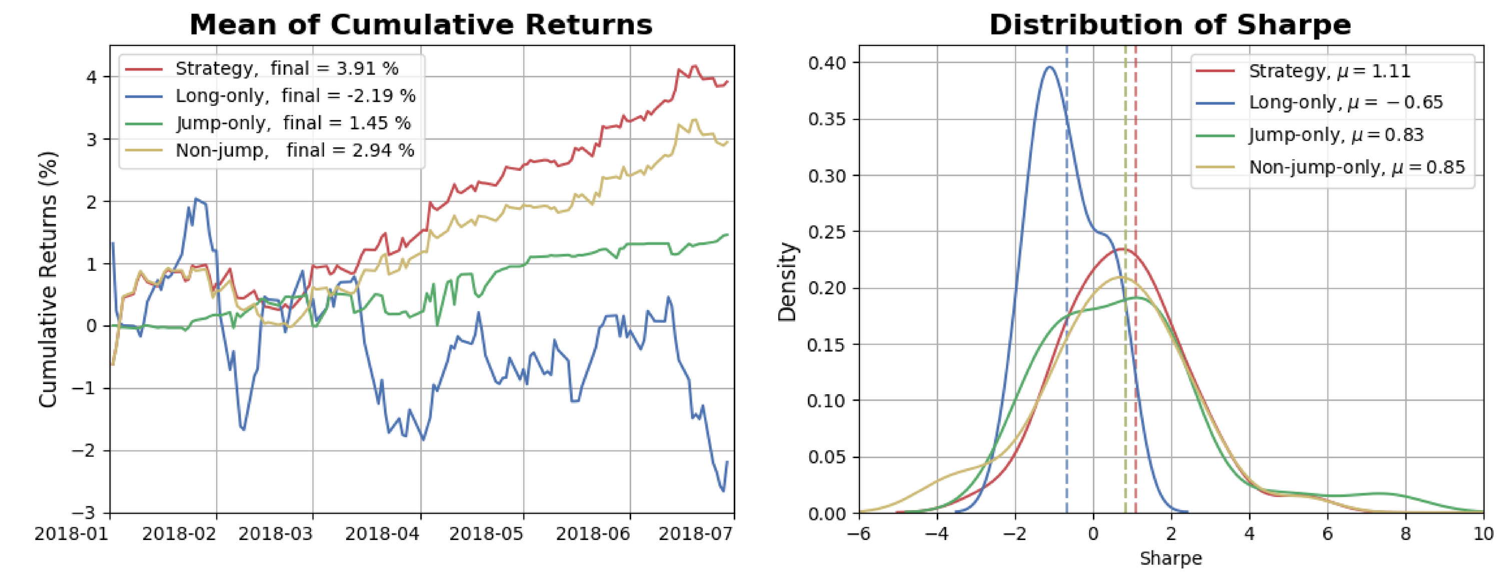
Simulation

We constructed $VPIN_z$ using some stable parameters $\{n = 200, t_\mu = 30, t_\sigma = 60\}$ and simulated the intraday strategy returns (excluding transaction costs and impact). For each contract for each trading day the daily return is calculated and the annualized Sharpe from there.



Predicting Jumps with VPIN

The simulation is repeated with an additional rule: using the time stamp of the price jump events, we allocate the corresponded return series into jump bucket, while the rest go to the non-jump bucket. The mean cumulative returns and the Sharpe of each bucket, as well as the mean strategy and long-only returns are displayed in figures below. What we can see is that the signal seems to be equally effective during both jump and non-jump periods. It is worth point out that the jump returns are nearly equal in size to the non-jump returns, despite them occurring over a significantly short time-scale.



Conclusions & Future Work & References

We have seen some evidence that VPIN does have predictive ability against future returns. We have seen that high-quality L3 data provides the additional fields removing the uncertainty for work with VPIN metrics. Future work could look at if this signals performance could be improved by using order book features [3].

[1] Easley, D., López de Prado, M., O'Hara, M. *Flow Toxicity and Liquidity in a High-frequency World. The Review of Financial Studies, Vol.25, pp.1457-1493, 2012*

[2] T.G. Andersen, O. Bondarenko *Assessing Measures of Order Flow Toxicity via Perfect Trade Classification. Review of Finance, Vol.19, pp.1-54, 2015*

[3] M. Paddrik et al. *Effects of Limit Order Book Information Level on Market Stability Metrics. White paper, Office of Financial Research. 2014*