Sensitivity Analysis of the VPIN Metric

Kesheng Wu, Jung Heon Song, Horst D. Simon
Lawrence Berkeley National Lab & UC Berkeley
http://crd.lbl.gov/cift/

Introduction

- Flash Crash of May 6, 2010 caused nearly one trillion dollars of equality to disappear from the market, but only to reappear minutes later.
- Challenge: can we detect or predict such an event before it happens?

Volume Synchronized Probability of Informed Trading (VPIN)

- Probability of Informed Trading (PIN) developed by Easley, Kiefer, O'Hara and Paperman [1996] intended to measure the information imbalance from prices of trades: it utilizes two levels of bins known as bars and buckets, generates bars and buckets based on time, e.g., all trades in a 10-second time window form a bar and all trades in a 5-minute time window form a bucket
- Volume Synchronized Probability of Informed Trading (VPIN) introduced volume bars to overcome the vastly varying arrival rates of trading activities (Easley, Lopez de Prado, O'Hara 2011)
- Bulk Volume Classification further reduces computational cost and avoids discrepancies with time stamps from different trading venues (Easley, Lopez de Prado, O'Hara 2012)
- There are strong anecdotal evidence that VPIN is an effective leading indicator based on available data during the Flash Crash of May 6 2010 (Figure 2)

Quantifying Effectiveness of VPIN

- Intuition
  - When VPIN is high, trouble ahead
  - Within a time window (called the event horizon) following the signal, the volatility is higher than usual
- Quantification
  - Use a threshold on a normalized version of VPIN so that the threshold can be keep the same for different trading instruments
  - Assume all time horizon are of the same length, measured with the number of buckets (expressed as a fraction of buckets per day)
  - Measure volatility with Maximum Intermediate Return (MIR), an instantaneous volatility measure
  - Compute MIR in an event against a random time interval of the same duration: if the MIR is high in the event, the trigger is considered predicting a true event, otherwise the prediction is false
  - **The prediction is effective if the false positive rate α is small**

Free Parameters for VPIN

(1) Nominal price of a bar π, (2) Buckets per day (BPD) β, (3) Support window σ, (4) Event horizon η, (5) Bulk Volume Classification (BVC) parameter ν, (6) Threshold for declaring VPIN event τ

Nonlinear Optimization with Mesh Adaptive Direct (NOMAD) Search

- Finds the optimal parameter combination that minimizes the false positive rate
- C++ implementation of NOMAD by Audet, Béchardand, Le Digabel [2008]

Uncertainty Quantification Toolkit (UQTK)

- Compute Sobol indices to measure the sensitivity of parameters using polynomial chaos expansion
- C++ implementation by Debusschere, Najm, Pébay, Knio, Ghanem, and Le Maître [2004]

Results

- Tested on 66-month trading activities of 94 most active futures contracts
- NOMAD was able to systematically examine the parameter space to produce much lower false positive rates than previous achieved: 20% →2%
- The number of bucket per day β has the strongest influence on the false positive rate, other than VPIN threshold τ

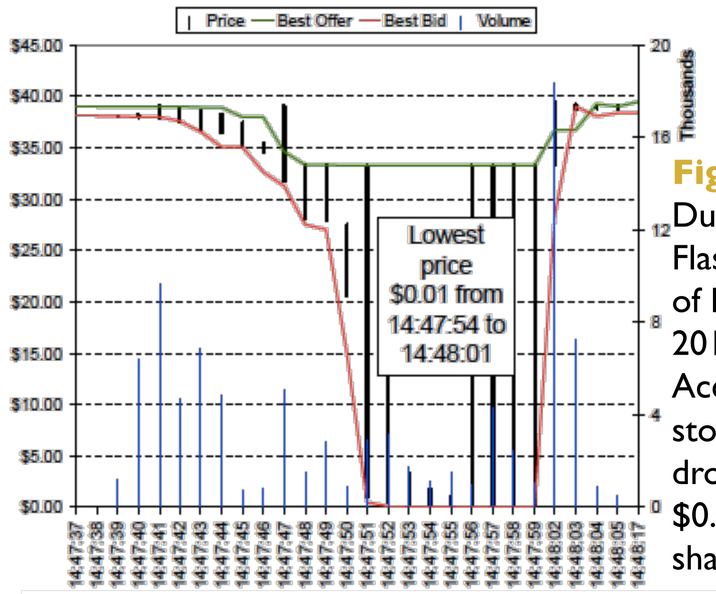


Figure 1: During the Flash Crash of May 6 2010, Accenture stock dropped to $0.01 per share.

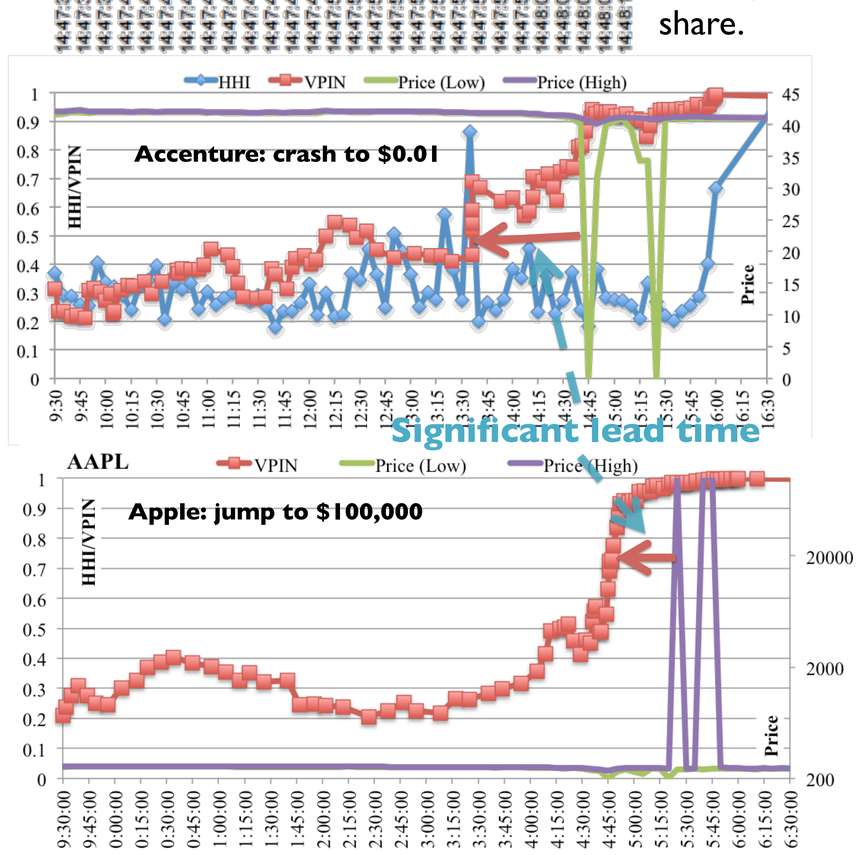
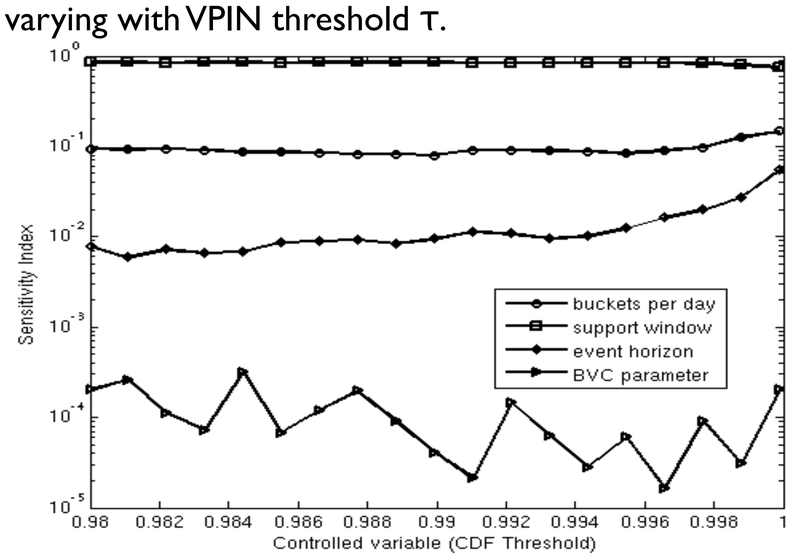


Figure 2: VPIN is among the early warning indicators that produced strong signals ahead of the flash crash.

Figure 3: Best results (lowest false positive rates α) produced by NOMAD

π	β	σ	η	υ	τ	α
Mean	1836	0.0478	0.0089	0.9578	0.9952	0.0258
Median	1528	0.1636	0.033	0.4611	0.9949	0.0340
Closing	1888	0.1578	0.0480	45.221	0.9942	0.0412
Closing	1600	0.3586	0.0482	10.371	0.9847	0.0458

Figure 4: Sobol Sensitivity index of parameters varying with VPIN threshold τ.



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