

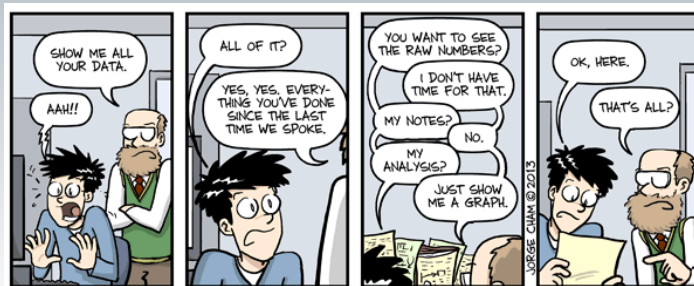
Twenty Years of VPIN

Stephen Rush

Department of Finance
The University of Connecticut School of Business

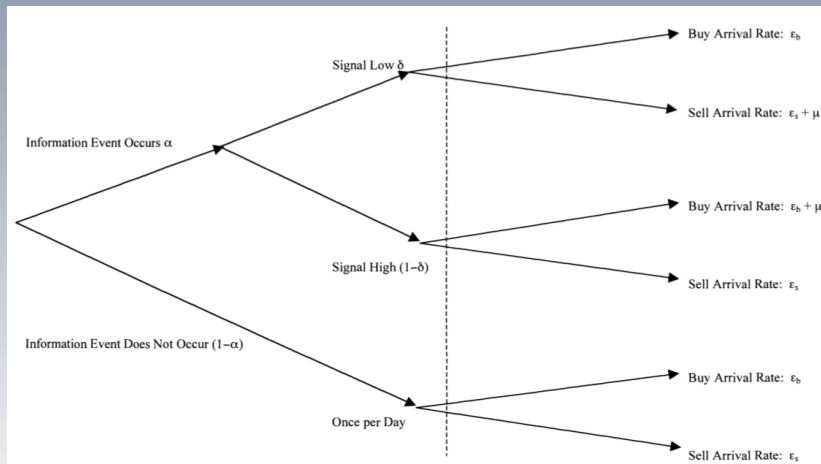
stephen.rush@business.uconn.edu

May 17, 2014



WWW.PHDCOMICS.COM

Probability of Informed Trading (PIN)



Probability of Informed Trading (PIN)

Lee-Reedy Algorithm (1991)

- Latency

Easley, O'Hara, and Paperman (1996)

- Original model

Easley, Hvidkjaer, and O'Hara (2002)

- Avg correlation between volume and PIN was -0.58
- Positive correlation between PIN and excess returns
- PIN matters less for large cap stocks

Volume Synchronized PIN (VPIN)

Bulk-Volume Classification

- Quotes are irrelevant
- Robust to HF trading

Easley,

Lopez de Prado, and O'Hara (2012)

- Changing information flow
- Volatility sampling
- CDF of VPIN



Liquidity or Informed Trading?

Liquidity

- Discount necessary to convert an asset to cash
- Likelihood of executing a trade
- Shown to be a state variable in all major asset classes

Informed Trading

- Need to execute with few trades \rightarrow greater expected value from trading
- No common component on time horizons > 1 day



TAQ Data from 1993 through 2013

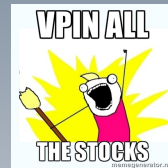
It's not big...it's large data.

Exchanges

- NYSE
- ARCA (continued tracking after merger)
- NASDAQ

Correlations

- Volume
- Returns



Model

“The best material model of a cat is another, or preferably the same, cat.” - Norbert Wiener



$$\text{VPIN}_{i,t} = \beta_1 \text{Market VPIN}_{i,t} + \beta_0 + \epsilon_{i,t} \quad (1)$$

Variables:

- MRP, SMB, and HML: Fama-French Factors
- UMD: Jegadeesh-Titman traded momentum factor
- PS: Pastor-Stambaugh traded liquidity factor
- RF: Risk-free Rate
- VPIN: Volume Synchronized Probability of Informed Trading

Econometrics

Single Stock Regressions

- Linear model with serially correlated errors

Panel Regressions

- Fixed Effects in both security and year
- First difference estimators for time dependence

Quantile Regressions

- Check 10%, 25%, 50%, 75%, 90%
- Idiosyncratic CDF factor remains significant



Predictive Value for Mergers and Acquisitions

Controls

- Fama-French Stock and Bond Factors
- Daily Momentum Factor
- Turnover and Spread

Methodologies

- Logit Model - significant but no foresight
- Multinomial Logit - significant but extremely low hit rate
- Event Study - not significant within a $T \pm 2$ day event period

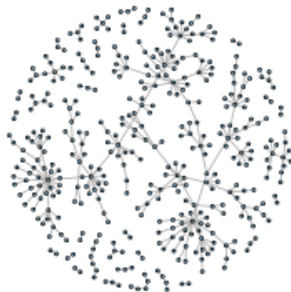
Network

VPIN Network in 1993



Network

VPIN Network in 2003



Network

VPIN Network in 2013



Tools

Hardware AKA 'MOAR CORES!':

- Assorted desktops running Windows 7, OS X, and Linux
- Discarded university servers (HP Proliant, IBM System X) running RStudio Server on Ubuntu Server
- Data Transfer: BitTorrent Sync (free unlimited space)
- Data Backup: Amazon Glacier - \$0.01/GB per month
- Total cost: < \$200

Not R Superstar:

- GAWK

Data Manipulation:

- foreach, doParallel, data.table, plyr, reshape2, fasttime, xts

Analysis:

- plm, bnlearn, nnet, quantreg

Visualization:

- ggplot2, texreg

Questions?

Code for the VPIN calculation in R is available on my website at:
<http://stephenrush.wordpress.com>

Email: stephen.rush@business.uconn.edu
Twitter: @ProbablePattern

