NYU STERN
NEW YORK UNIVERSITY · LEONARD N. STERN SCHOOL OF BUSINESS
CENTER FOR BUSINESS ANALYTICS

The Rise of Machines and Liquidity Patterns in Markets

Vasant Dhar

Professor and Co-Director, Center for Business Analytics
Stern School of Business
Editor-in-Chief, Big Data Journal
April 25, 2014

Key Points

- Patterns emerge before reasons for them become apparent, especially in complex systems such as financial markets
- Big Data has assumed a big role in emergence of machine intelligence: computers are not only running markets but playing a major role in data driven hypothesis **generation**
- What does liquidity "feel like" with machines running markets and why?
- What do regulators need in terms of technology capability to keep up with the complexity of financial markets?

Big Data and Search Algorithms: The Grist for Machine Intelligence

Data Science is the study of the generalizable extraction of knowledge from data*

The fire hose of data – quotes, cancellations, trades, etc is massive and getting larger – complexity has been on the rise: the analytically astute have an advantage

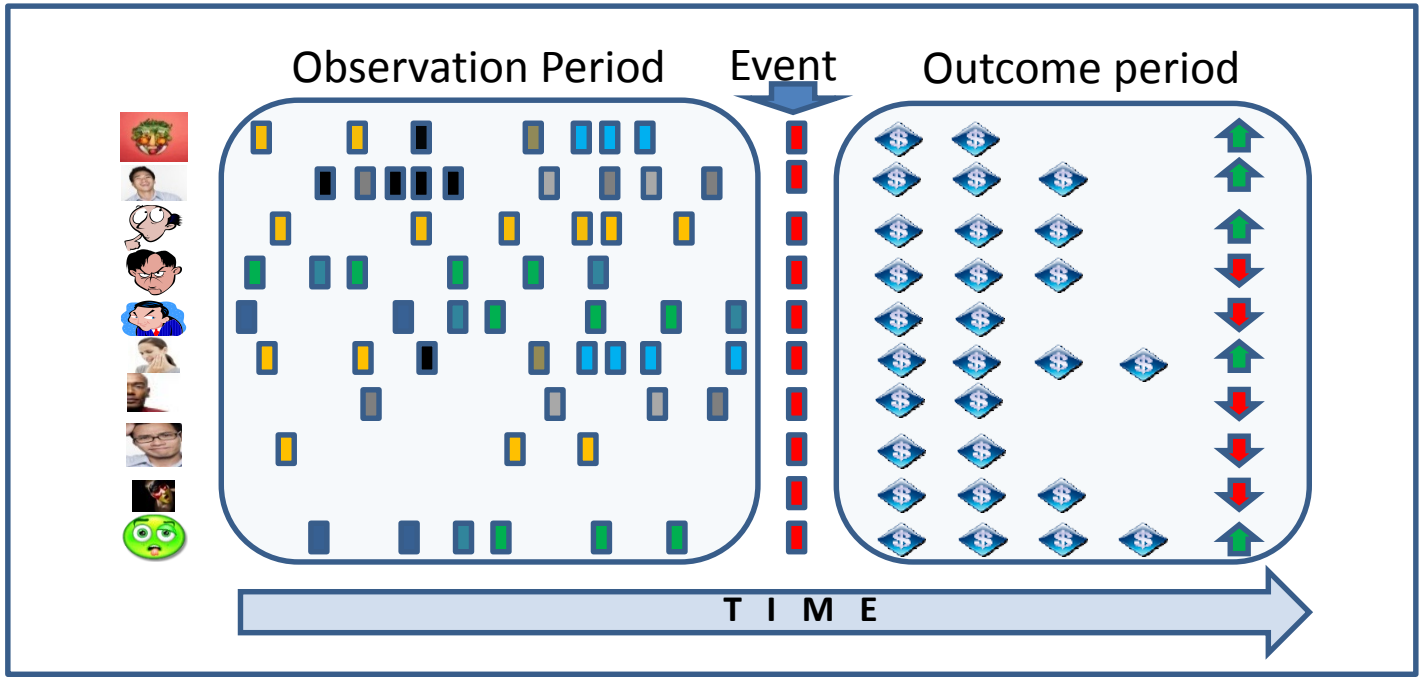
***Dhar, V., Data Science and Prediction, Communications of the ACM, Vol. 56 No. 12, December 2013**
http://cacm.acm.org/magazines/2013/12/169933-data-science-and-prediction/fulltext

Asking the Right Question

"Patterns Emerge Before Reasons for Them Become Apparent"

Asking the right question is therefore critical: "If only you knew what question to ask me, I'd give you very interesting answers from the data."

What is the Right Question Here?



Are bad outcomes associated with the blues?
Or with the grays?
Or the yellows in the absence of the blues?
Or is it more than three yellows or three blues?
Or is it the greens in "quick succession?"
Or … do we need to run additional experiments and gather more data?

Observational Data and Acquiring New Data

- Observational data may answer questions without explicitly asking anyone anything!
- It may also require an understanding of how the data are being generated
  - Are there "natural experiments" that are reflected in the data or are the data somehow biased through self selection?
  - Is it possible to **run** experiments to **get** additional data to amend the observed data?

Pattern/Anomaly Detection Mindset…

…assume that most data are “noise”

…challenge is to extract the “signal” that occurs very infrequently

…design algorithms accordingly

Internalization Patterns

Dark Pool / Internalization Rates by Volume and Stock Price – September 2012

Stock Price \ Consolidated Average Daily Volume	Under 100K CADV	100K-250K	250K-500K	500K-1MN	1MN-2MN	2MN-4MN	4MN-8MN	Above 8MN
Under $1	36.6%	36.5%	40.0%	40.5%	38.1%	45.4%	42.7%	41.9%
$1-$5	40.9%	44.2%	43.8%	43.4%	43.4%	40.7%	40.6%	46.9%
$5-$10	39.7%	41.5%	39.6%	37.5%	38.6%	35.9%	31.3%	37.4%
$10-$15	41.2%	40.3%	38.0%	36.8%	35.6%	31.0%	31.5%	35.6%
$15-$20	43.4%	42.7%	38.8%	33.1%	35.7%	30.7%	29.7%	28.5%
$20-$30	44.2%	42.8%	38.1%	34.5%	39.8%	32.3%	30.0%	34.7%
$30-$40	40.9%	37.3%	37.3%	31.6%	30.9%	29.9%	32.0%	31.0%
$40-$50	44.7%	34.2%	32.7%	33.5%	31.6%	30.8%	31.5%	27.8%
$50-$75	46.0%	37.7%	33.2%	31.9%	31.0%	29.2%	29.3%	25.9%
$75-$100	44.1%	38.5%	35.9%	31.5%	31.6%	30.5%	33.8%	28.4%
$100-$150	41.3%	34.7%	34.2%	35.0%	33.9%	26.4%	26.7%	27.8%
$150-$250	32.5%	29.2%	32.8%	28.9%	31.5%	27.8%	24.8%	33.8%
$250-$400	24.3%	NA	NA	32.4%	NA	34.7%	NA	NA
Above $400	31.4%	NA	31.5%	32.2%	NA	35.4%	NA	37.0%

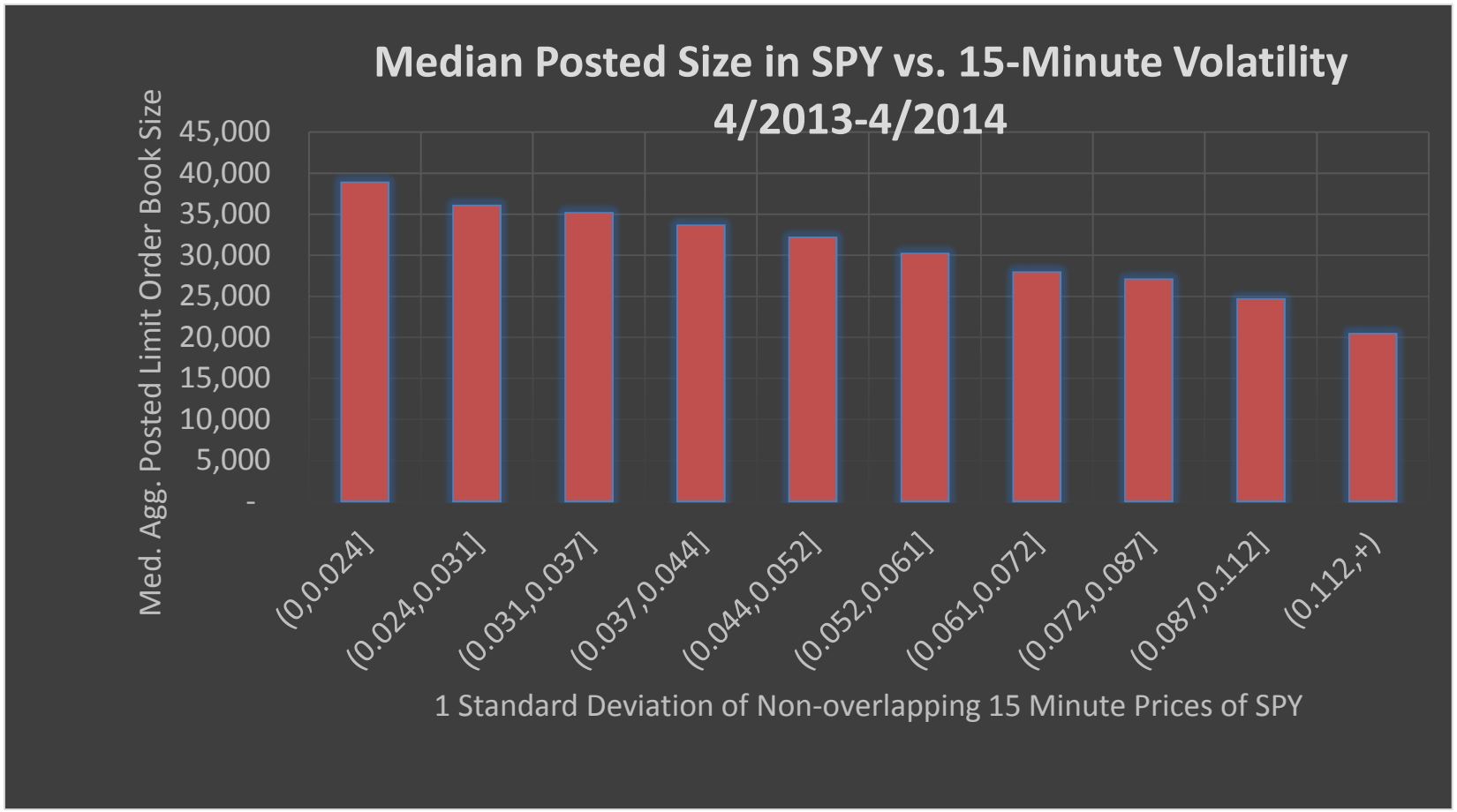
1. Is the penny spread artificially wide in these liquid stocks?

2. Do high-priced stocks require too much capital?

3. Is there insufficient liquidity to risk posting an order?

Source: TABB Group

Volatility Vs. Liquidity

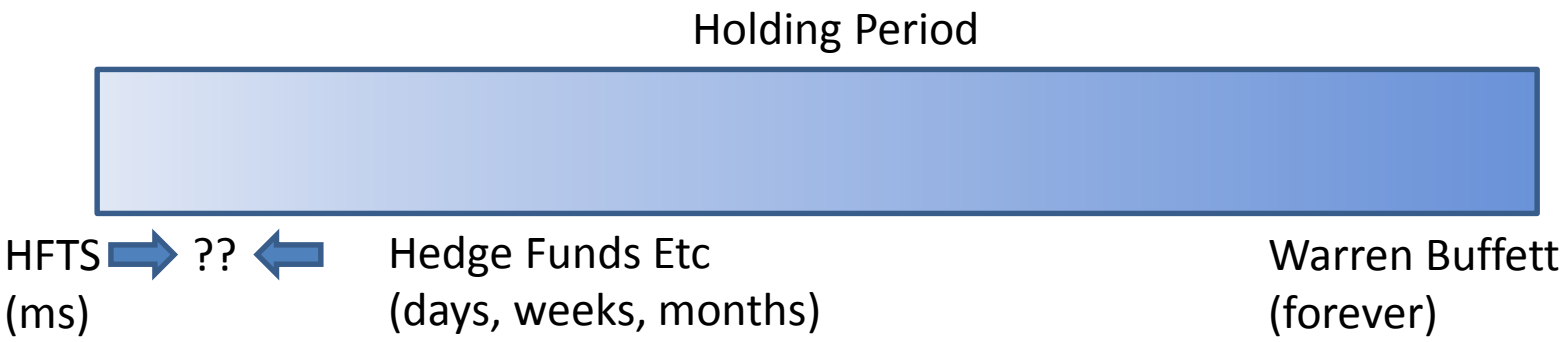


Source: T2AM (data from Tradeworx)

Is the "Displayed Liquidity" Illusory?

- It can be when you most need it, but the data suggest otherwise
  - But its always been that way to some extent even before machines
  - Perhaps the "cost of liquidation" is a better measure of liquidity: takes specific risk into account
- Making markets with penny spreads (as opposed to 6.25 cents in the old days) requires sophisticated real-time risk management
- We *wanted* a fast market and *encouraged fragmentation*, so it isn't surprising that machines are doing more of the work to minimize risk and use every bit of information available to do so rapidly

What if Speed Isn't an Advantage?

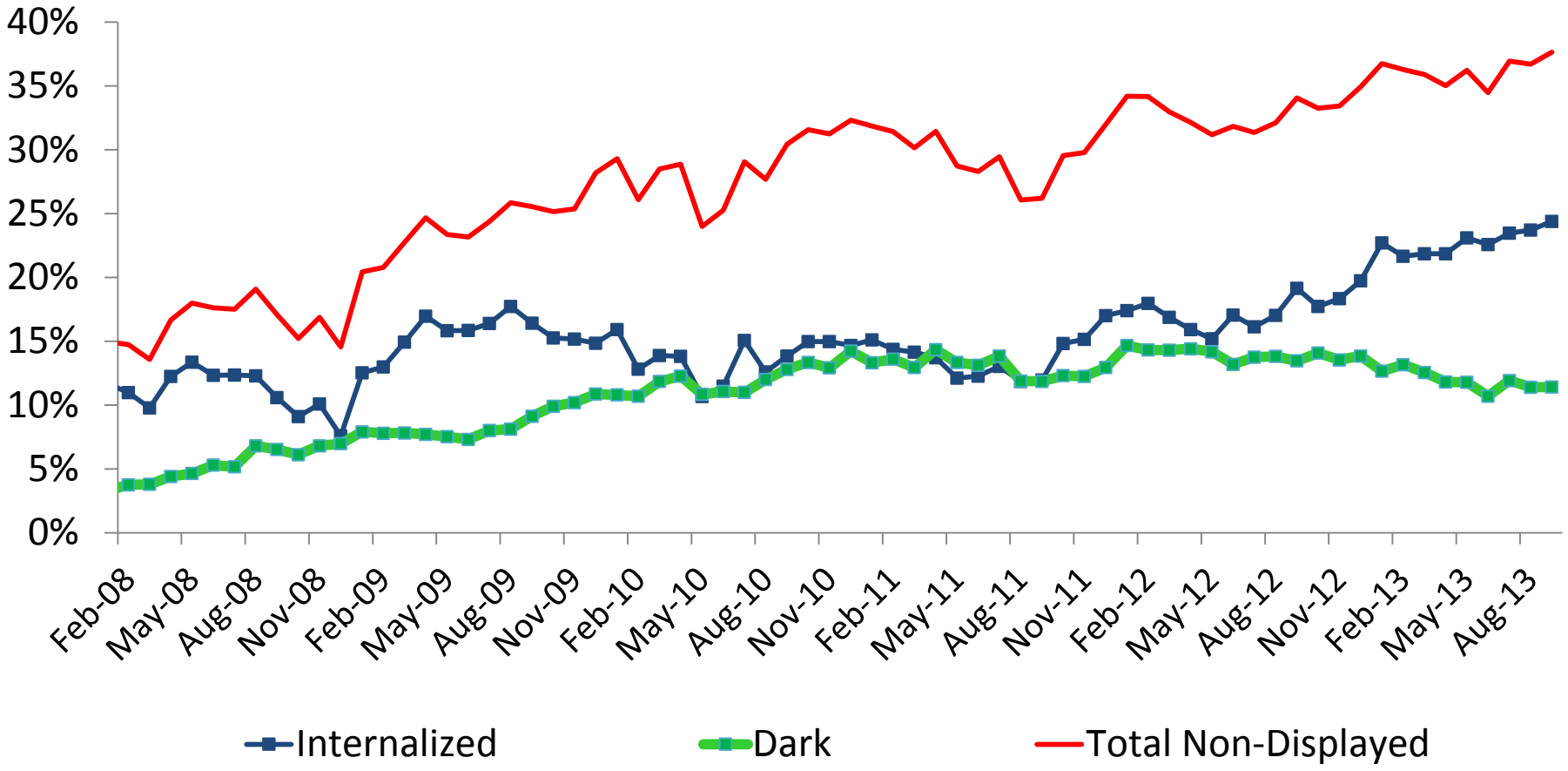


Is this one reason why HFT profitability is declining?

Summary

- People “closer” to the market tend to have an advantage
- People who understand the implications of new regulation ask the right questions for discovering patterns in “big data” through better predictive analytics
- HFTs can not only been faster but smarter because they need to control risk tightly in current-day markets where spreads are tight
- If markets are “rigged” or “unfair” this should be evident in the data, as long as regulators have the analytical sophistication to detect patterns that suggest as such irregularities
- NMS had unintended consequences, but you can’t punish people for being smarter!

Recent Data: Off-Exchange Activity Has Increased



Source: TABB Group