Market Microstructure and Algorithmic Trading

PIMS Summer School 2016
University of Alberta, Edmonton
Lecture 1: July 4, 2016

Robert Almgren

Robert Almgren background

PhD Applied Mathematics (gas dynamics)
Academic career (Univ Toronto)
  fluid dynamics & free boundary problems
  some papers on trade execution and costs
Bank of America Electronic Trading Services
  algorithmic execution in US equity markets
Quantitative Brokers
  algorithmic execution in fixed income and futures
  microstructure and trading expertise

Market microstructure: the study of detailed functioning of markets. Usually financial markets because large interest and good data.

Optimal trading: How to achieve your objectives in a given market structure.

Key concepts

Industrial

    details of market design
    high-frequency volatility
    intraday volume and volatility curves
    event response
    price forecast signals

Academic

    universal statistics of price motion
    impact of trading on markets
    optimal trading in various models

Course outline

1. Markets and Data
   What is market microstructure?
   What kinds of data do we get?

2. Tick Size, Reversion, and Volatility
   How do we compare different products?
   How can we measure volatility

3. Market Impact: effects of trading on mkts
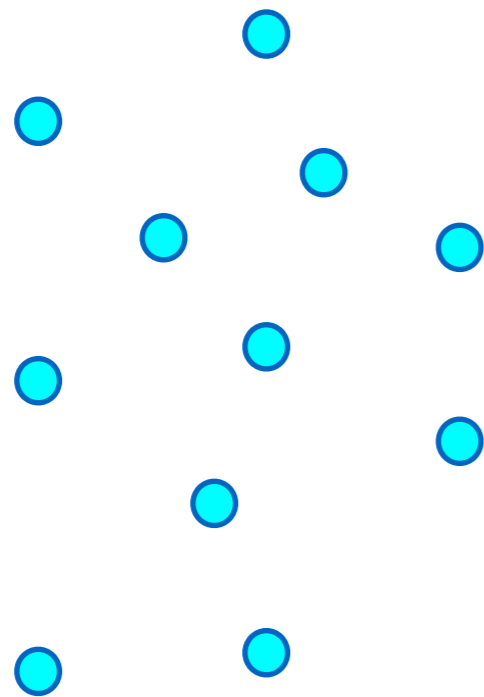   individual orders

   parent orders (meta orders)

4. Optimal Trading
   balancing risk and reward

   calculus of variations vs dynamic programming
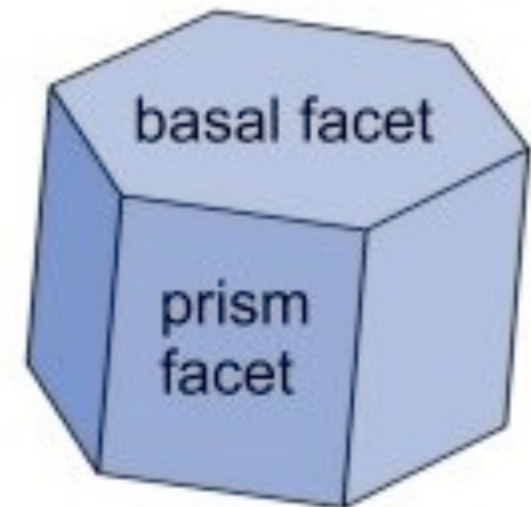
Metaphor for market microstructure
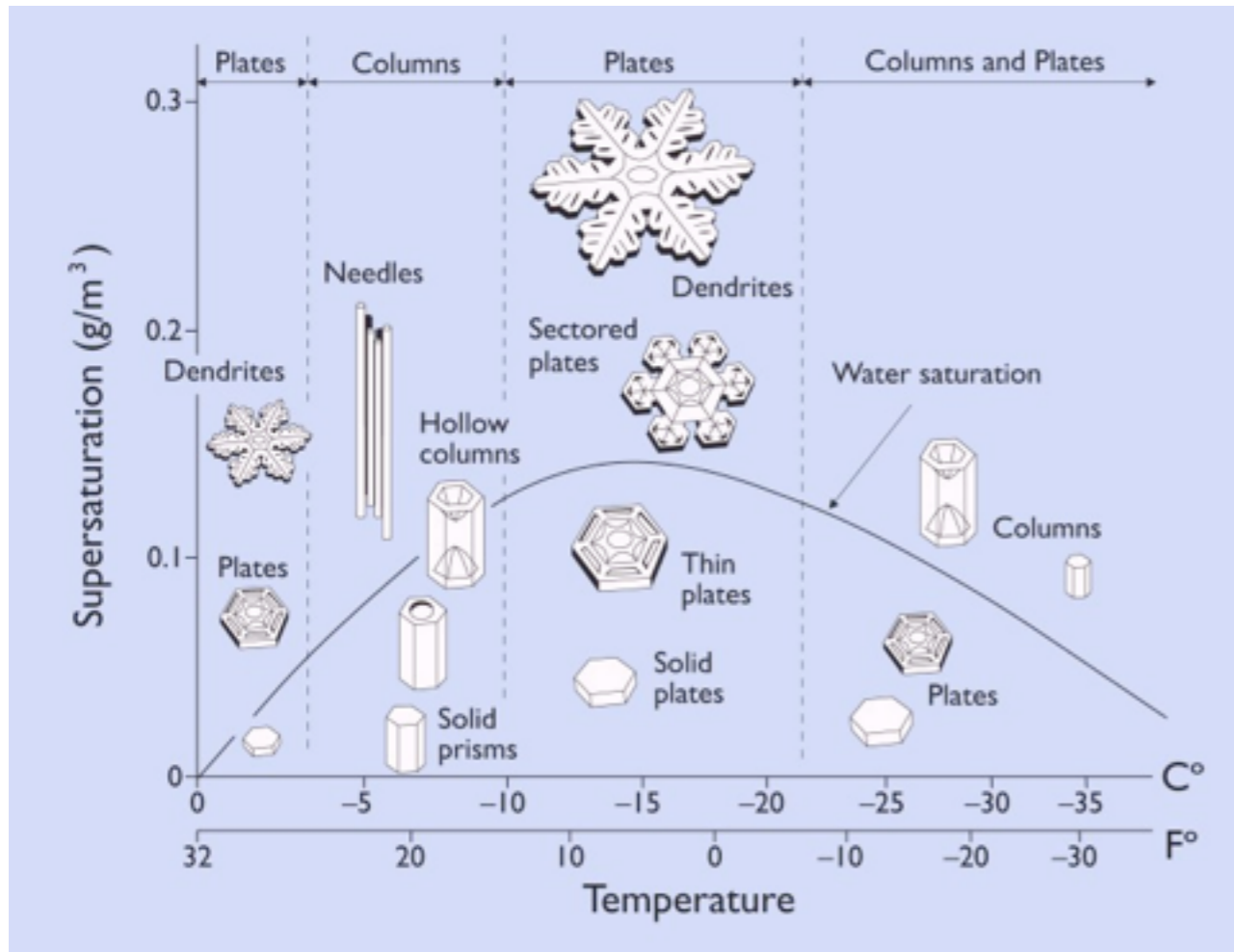(don't take this too seriously)

Water
vapor
in air

Cool
below
freezing

http://www.snowcrystals.com
Kenneth G. Libbrecht, Caltech

basal facet

prism
facet

Energy-
minimizing
ice crystal

What happens in reality

http://www.snowcrystals.com
Kenneth G. Libbrecht, Caltech

Process is as important as equilibrium state

Two kinds of finance

- Equilibrium finance
  optimal portfolio construction
  pricing of complex products
  fixed income yield curves

- Transaction finance
  trading costs (slippage)
  market structure and microstructure
  execution strategies

Equilibrium finance

The Pricing of Options and Corporate Liabilities

Fischer Black
University of Chicago

Myron Scholes
Massachusetts Institute of Technology

Journal of Political Economy 81 (1973) 637–654

In deriving our formula for the value of an option in terms of the price of the stock, we will assume "ideal conditions" in the market for the stock and for the option:

e) There are no transaction costs in buying or selling the stock or the option.

## Theory of rational option pricing

Robert C. Merton *J. Econom. Theory*, 3:373–413, 1971.

(1) *"Frictionless"* markets: there are no transactions costs or differential taxes. Trading takes place continuously and borrowing and short-selling are allowed without restriction.[41]

It is clearly possible to use a computer as a communications device between all the participants in the present system of trading. But we want to know instead whether it is possible to have an automated exchange providing markets as good as, or better than, the markets provided by current methods, and acting as the only intermediary between the ultimate buyers and sellers. Can a system be designed that will allow the user to place his order and then leave it to be executed by the system either immediately or over a period of time?
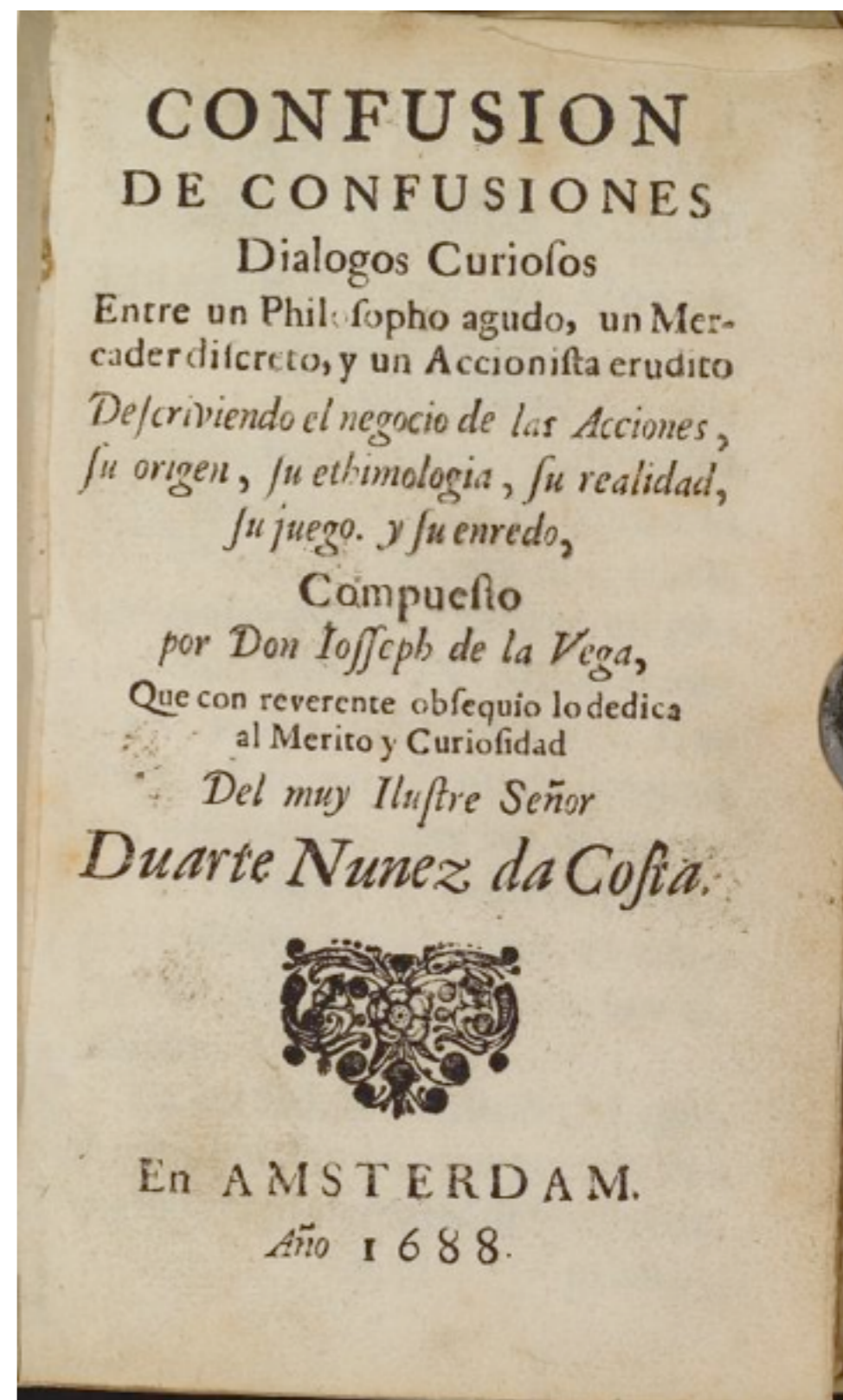
Buyers and sellers as a group will no longer have to pay a market maker for standing ready to buy for his inventory or sell from his inventory, or for taking the risk that his inventory will change in value over time. The function will be taken over by those buyers and sellers who do not need to have immediate execution of their orders.

A computer can even handle complicated trading strategies automatically. It can follow the sequence of transactions in one or more stocks automatically, and can activate orders whenever a "buy signal" or a "sell signal" (as defined by the designer of the strategy) appears. If stock prices move randomly, however, then it should not be profitable to use such strategies.

TOWARD A FULLY AUTOMATED STOCK EXCHANGE

by Fischer Black
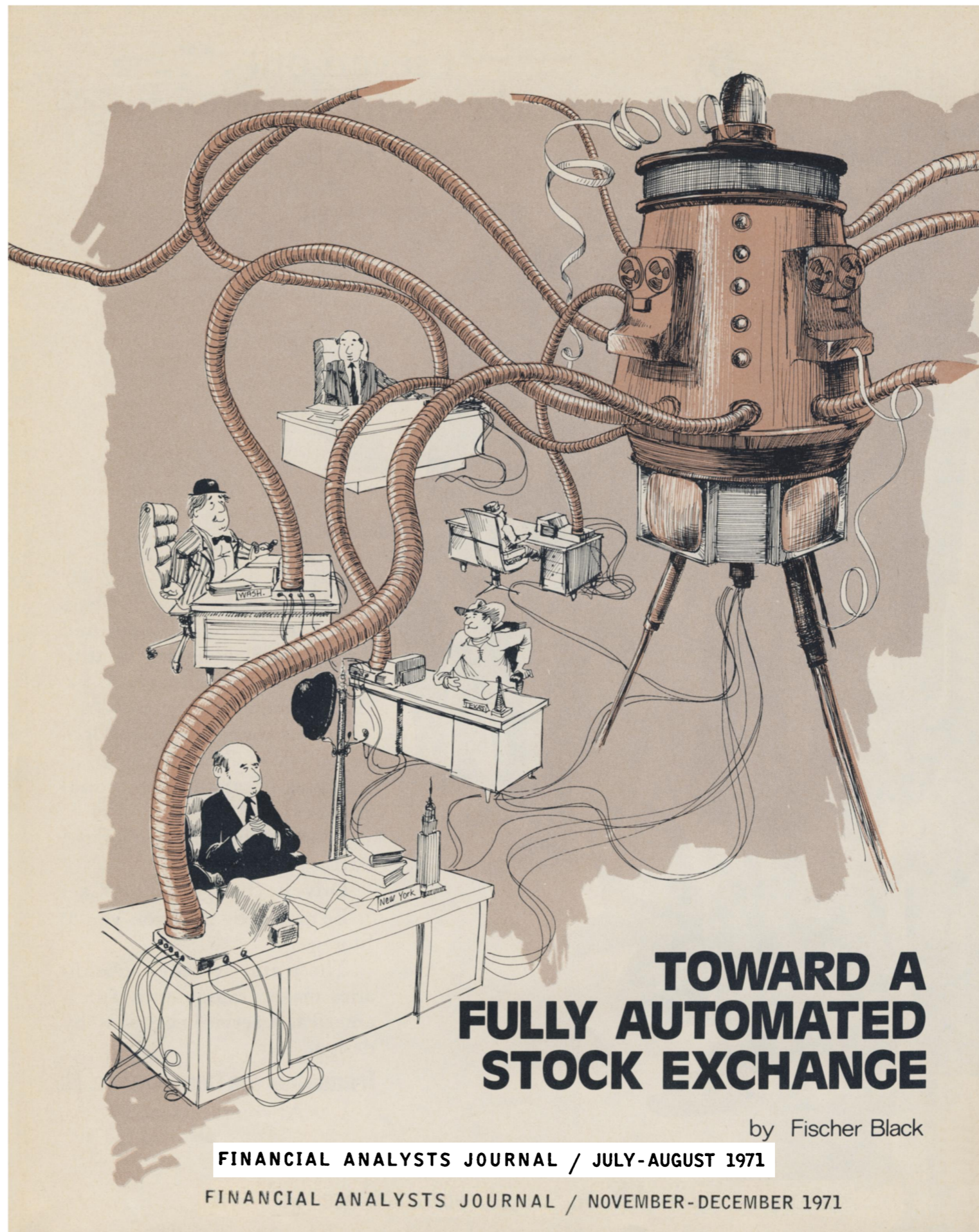
... this enigmatic business [trading] which is at once the fairest and most deceitful in Europe, the noblest and the most infamous in the world, the finest and the most vulgar on earth. It is a quintessence of academic learning and a paragon of fraudulence; it is a touchstone for the intelligent and a tombstone for the audacious, a treasury of usefulness and a source of disaster.

--Josef de la Vega
Confusion de Confusiones (1688)

CONFUSION
DE CONFUSIONES
Dialogos Curiosos
Entre un Philosopho agudo, un Mer-
cader discreto, y un Accionista erudito
Descriviendo el negocio de las Acciones,
su origen, su ethimologia, su realidad,
su juego. y su enredo,
Compuesto
por Don Ioseph de la Vega,
Que con reverente obsequio lo dedica
al Merito y Curiosidad
Del muy Iluftre Señor
Duarte Nunez da Costa.

En AMSTERDAM.
Año 1688.

Markets really still look like this:

not mathematics and statistics

Different financial product areas

Equity trading

   large culture of agency trading

   well developed trading technology

   intense focus on execution costs

Futures trading

   markets largely electronic, good market data

   little algorithmic execution

   wide variety of products

I will talk mostly about futures because of personal expertise, interest, and data

Fixed income trading

   executed largely through dealers

   quant emphasis on pricing

   little attention to execution costs

Interest of market microstructure

Exchange design
exchanges are for-profit institutions

futures exchanges monolithic: freedom to design

Regulation and societal benefit
risks and benefits of high-frequency trading

Trade execution
big business

key determinant of profits in trading

Scientific research
lots of interesting problems

applied area: need practical understanding

Exchange design

Exchanges must choose
   product design (futures)
   minimum price increment (tick size)
   trade hours
   matching algorithm (FIFO, pro rata, auction, ...)

Exchanges are profit-making institutions with their own goals

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
WASHINGTON, D.C. 20549

———————————————————————

FORM 10-K

———————————————————————

ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934

For the Fiscal Year Ended December 31, 2015

CME GROUP INC.

We offer the widest range of global benchmark products across all major asset classes, based on interest rates, equity indexes, foreign exchange, energy, agricultural commodities and metals. Our products include both exchange-traded and privately negotiated futures and options contracts and swaps. We bring buyers and sellers together through our CME Globex electronic trading platform across the globe and our open outcry trading facilities in Chicago and New York City, and provide hosting, connectivity and customer support for electronic trading through our co-location services. Our CME Direct technology offers side-by-side trading of exchange-listed and privately negotiated markets. We provide clearing and settlement services for exchange-traded contracts, as well as for cleared swaps, and provide regulatory reporting solutions for market participants through our global repository services in the United States, United Kingdom, Canada and Australia. Finally, we offer a wide range of market data services — including live quotes, delayed quotes, market reports and a comprehensive historical data service — and continue to expand into the index services business.

Our Competitive Strengths

We provide innovative ways to manage risk and offer a number of key differentiating elements that set us apart from others in our industry, including:

Highly Liquid Markets — Our listed futures and options markets provide an effective forum for our customers to manage their risk and meet their investment needs relating to our markets. We believe that our customers choose to trade on our centralized market due to its liquidity and price transparency. Market liquidity — or the ability of a market to absorb the execution of large purchases or sales quickly and efficiently, whereby the market recovers quickly following the execution of large orders — is key to attracting customers and contributing to a market's success.

Most Diverse Product Line — Our products provide a means for hedging, speculation and asset allocation relating to the risks associated with, among other things, interest rate sensitive instruments, equity ownership, changes in the value of foreign currency, credit risk and changes in the prices of agricultural, energy and metal commodities. The estimated percentage of clearing and transaction fees revenue contributed by each product line is as follows:

Superior Trading Technology and Distribution — We strive to provide the most flexible architecture in terms of bringing new technology, innovations and solutions to the marketplace. Our CME Globex electronic platform is the trading engine for our central limit order book markets, and is available on a global basis nearly 24 hours a day throughout the trading week. The CME Globex platform is accessible through a wide variety of vendor provided and custom built trading systems that benefit from our open application programming interface approach. For the privately negotiated markets, we offer brokers and customers the CME Direct system for arranging, executing, recording and risk-managing trades. CME Direct is a trading and analysis system that also includes CME Messenger for instant-message capabilities and CME Straight-Through Processing for connecting trade information directly with customer order management and risk management systems and is designed to reduce errors and improve efficiency. In 2015, 87% of our contract volume was conducted electronically.

Regulation and society

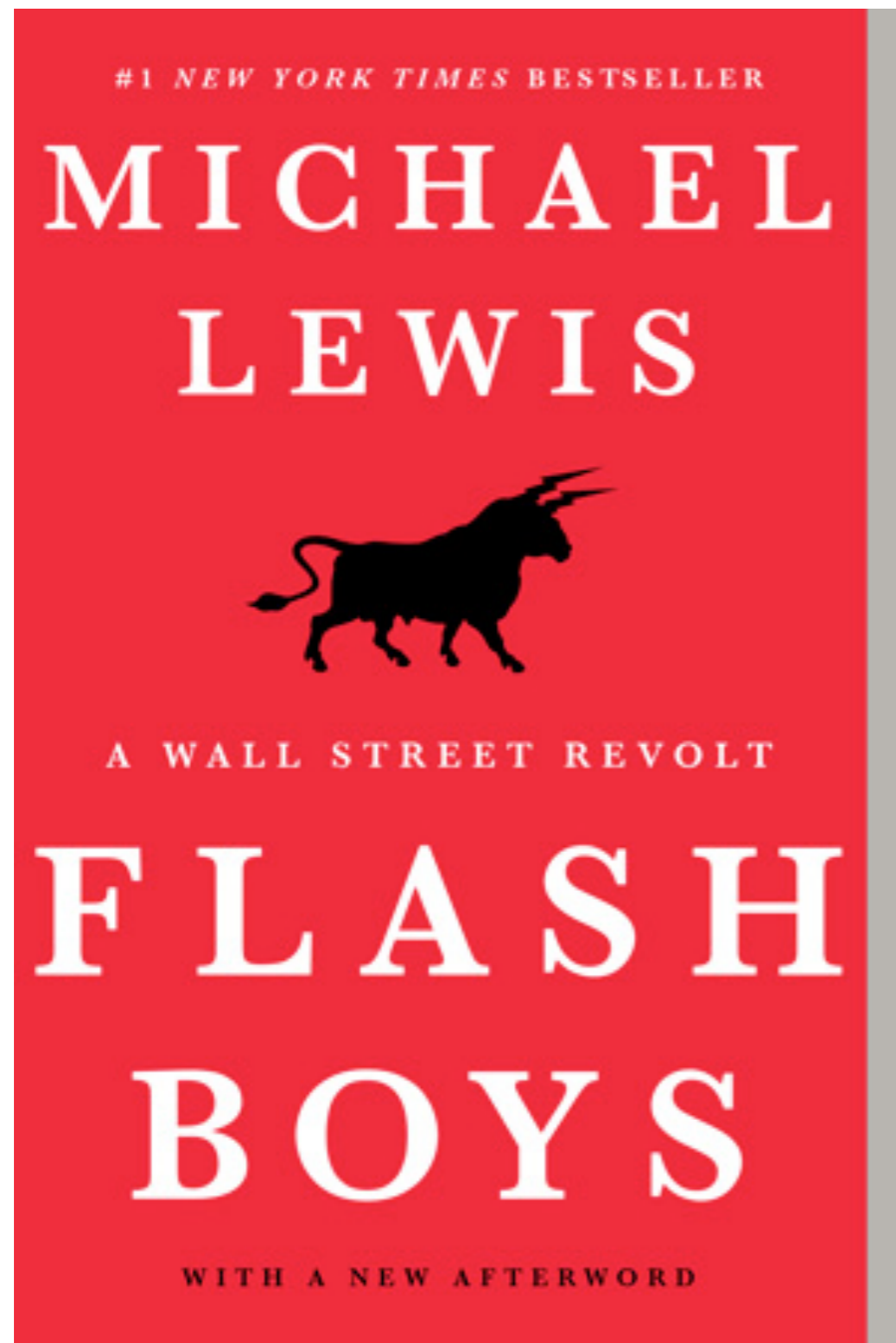Financial exchanges are important

investor value

price signals

liquid secondary market for auctions

Public impressions:

#1 *NEW YORK TIMES* BESTSELLER

MICHAEL LEWIS

A WALL STREET REVOLT

FLASH BOYS

WITH A NEW AFTERWORD

It used to be that when his trading screens showed 10,000 shares of Intel offered at $22 a share, it meant that he could buy 10,000 shares of Intel for $22 a share. He had only to push a button. By the spring of 2007, however, when he pushed the button to complete a trade, the offers would vanish. In his seven years as a trader, he had always been able to look at the screens on his desk and see the stock market. Now the market as it appeared on his screens was an illusion.

This made it impossible for Katsuyama to do his job properly. His main role as a trader was to play the middleman between investors who wanted to buy and sell big amounts of stock and the public markets, where the volumes were smaller. Say some investor wanted to sell a block of three million Intel shares, but the markets showed demand for only one million shares: Katsuyama would buy the entire block from the investor, sell off a million shares instantly and then work artfully over the next few hours to unload the other two million. If he didn't know the actual demand in the markets, he couldn't price the larger block. He had been supplying liquidity to the market; now whatever was happening on his screens was reducing his willingness to do that.

You may or may not buy Mr. Lewis's depiction of the high-frequency types as villains and those trying to thwart them as heroes. (If you ask me, there are no good guys in this story.) But either way, spending hundreds of millions of dollars to save three milliseconds looks like a huge waste. And that's part of a much broader picture, in which society is devoting an ever-growing share of its resources to financial wheeling and dealing, while getting little or nothing in return.

Paul Krugman 2014

SEC Approves IEX Proposal to Launch National Exchange, Issues Interpretation on Automated Securities Prices

Staff Issues Guidance on Speed Bumps

FOR IMMEDIATE RELEASE
2016-123

Washington D.C., June 17, 2016 — The Securities and Exchange Commission today approved Investors' Exchange LLC's (IEX) application to register as a national securities exchange. At the same time, the Commission issued an updated interpretation that will require trading centers to honor automated securities prices that are subject to a small delay or "speed" bump when being accessed.

"Today's actions promote competition and innovation, which our equity markets depend on to continue to deliver robust, efficient service to both retail and institutional investors," said SEC Chair Mary Jo White. "A critical role of the Commission's regulatory framework is to facilitate the ability of market participants to craft appropriate market-based initiatives, consistent with our mission to protect investors, maintain market integrity, and promote capital formation."

Equity Flash Crash

FINDINGS REGARDING THE MARKET EVENTS OF MAY 6, 2010

REPORT OF THE STAFFS OF THE CFTC AND SEC TO THE JOINT ADVISORY COMMITTEE ON EMERGING REGULATORY ISSUES

U.S. Commodity Futures Trading Commission
Three Lafayette Centre, 1155 21ˢᵗ Street, NW
Washington, D.C. 20581
(202) 418-5000
www.cftc.gov

U.S. Securities & Exchange Commission
100 F Street, NE
Washington, D.C. 20549
(202) 551-5500
www.sec.gov

SEPTEMBER 30, 2010

LESSONS LEARNED

The events summarized above and discussed in greater detail below highlight a number of key lessons to be learned from the extreme price movements observed on May 6.

One key lesson is that under stressed market conditions, the automated execution of a large sell order can trigger extreme price movements, especially if the automated execution algorithm does not take prices into account. Moreover, the interaction between automated execution programs and algorithmic trading strategies can quickly erode liquidity and result in disorderly markets. As the events of May 6 demonstrate, especially in times of significant volatility, high trading volume is not necessarily a reliable indicator of market liquidity.

May 6 was also an important reminder of the inter-connectedness of our derivatives and securities markets, particularly with respect to index products. The nature of the cross-market trading activity described above was confirmed by extensive interviews with market participants (discussed more fully herein), many of whom are active in both the futures and cash markets in the ordinary course, particularly with respect to "price discovery" products such as the E-Mini and SPY. Indeed, the Committee was formed prior to May 6 in recognition of the continuing convergence between the securities and derivatives markets, and the need for a harmonized regulatory approach that takes into account cross-market issues. Among other potential areas to address in this regard, the staffs of the CFTC and SEC are working together with the markets to consider recalibrating the existing market-wide circuit breakers – none of which were triggered on May 6 – that apply across all equity trading venues and the futures markets.

Another key lesson from May 6 is that many market participants employ their own versions of a trading pause – either generally or in particular products – based on different combinations of market signals. While the withdrawal of a single participant may not significantly impact the entire market, a liquidity crisis can develop if many market participants withdraw at the same time. This, in turn, can lead to the breakdown of a fair and orderly price-discovery process, and in the extreme case trades can be executed at stub-quotes used by market makers to fulfill their continuous two-sided quoting obligations.

As demonstrated by the CME's Stop Logic Functionality that triggered a halt in E-Mini trading, pausing a market can be an effective way of providing time for market participants to reassess their strategies, for algorithms to reset their parameters, and for an orderly market to be re-established.

November 24, 2015

CFTC Unanimously Approves Proposed Rule on Automated Trading

Washington, DC — The U.S. Commodity Futures Trading Commission (Commission) today unanimously approved proposed rules that mark a comprehensive regulatory response to the evolution of automated trading on U.S. designated contract markets (DCMs). The proposed rules, known collectively as Regulation Automated Trading or Regulation AT, represent a series of risk controls, transparency measures, and other safeguards to enhance the U.S. regulatory regime for automated trading. The notice of proposal will be open for a 90-day public comment period.

Regulation AT takes a multilevel approach by proposing risk control and other requirements for (a) market participants using algorithmic trading systems (ATSs), who are defined as "AT Persons" in the rulemaking, (b) clearing member futures commission merchants (FCMs) with respect to their AT Person customers, and (c) DCMs executing AT Person orders. The proposed rules are intended to reduce potential risks arising from algorithmic trading activity, by requiring the implementation of risk controls such as maximum order message and maximum order size parameters, and the establishment of standards for the development, testing, and monitoring of ATSs, among other requirements. AT Persons and clearing member FCMs would also be required to submit reports on their risk controls to DCMs, and maintain books and records regarding their risk controls and other algorithmic trading procedures for review by DCMs.
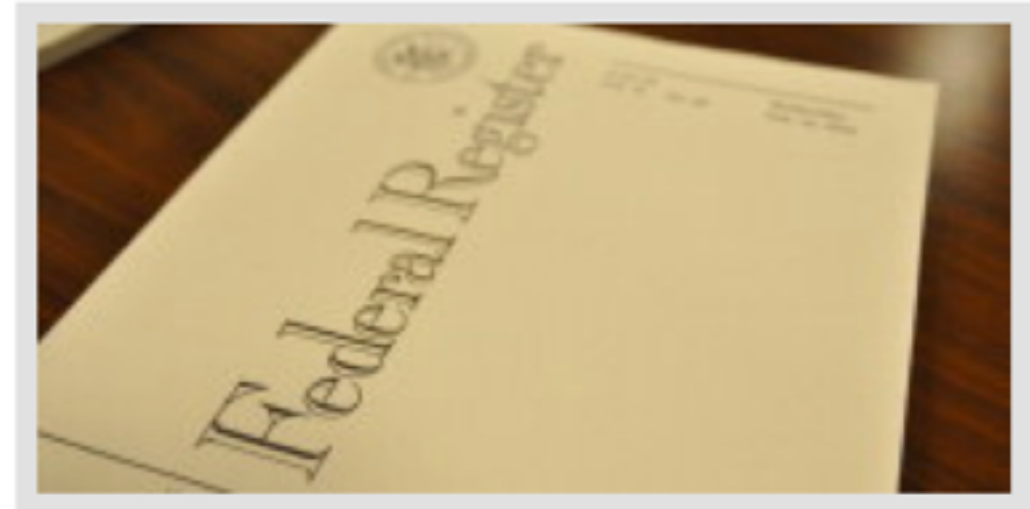
Interest rates "Flash Rally"

Joint Staff Report:
THE U.S. TREASURY MARKET ON OCTOBER 15, 2014

July 13, 2015

On October 15, 2014 ("October 15"), the market for U.S. Treasury securities, futures, and other closely related financial markets experienced an unusually high level of volatility and a very rapid round-trip in prices. Although trading volumes were high and the market continued to function, liquidity conditions became significantly strained. The yield on the benchmark 10-year Treasury security, a useful gauge for the price moves in other, related instruments that day, experienced a 37-basis-point trading range, only to close 6 basis points below its opening level. Intraday changes of greater magnitude have been seen on only three occasions since 1998 and, unlike October 15, all were driven by significant policy announcements. Moreover, in the narrow window between 9:33 and 9:45 a.m. ET, yields exhibited a significant round-trip without a clear cause, with the 10-year Treasury yield experiencing a 16-basis-point drop and then rebound. For such significant volatility and a large round-trip in prices to occur in so short a time with no obvious catalyst is unprecedented in the recent history of the Treasury market.

Figure 2.5: Historical Intraday Yield Ranges for 10-Year Treasury (Cash)
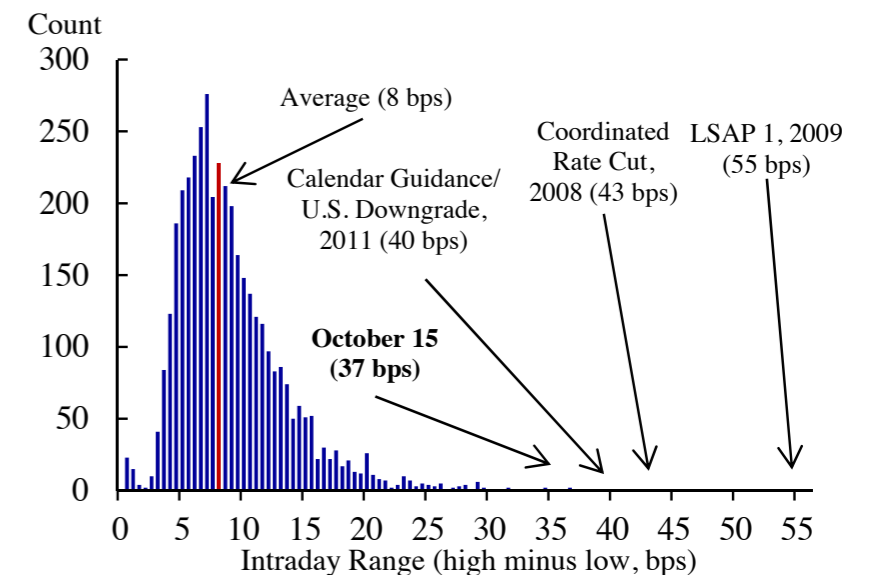
Note: Daily observations; 10/1998-10/2014
Source: Staff calculations, based on data from Bloomberg.

DEPARTMENT OF THE TREASURY

[Docket No. TREAS–DO–2015–0013]

Notice Seeking Public Comment on the Evolution of the Treasury Market Structure

**AGENCY:** Office of the Under Secretary for Domestic Finance, Department of the Treasury.

ACTION: Notice and Request for Information.

SUMMARY: The Department of the Treasury ("Treasury") is seeking public comment on structural changes in the U.S. Treasury market and their implications for market functioning; trading and risk management practices across the U.S. Treasury market; considerations with respect to more comprehensive official sector access to Treasury market data; and benefits and risks of increased public disclosure of Treasury market activity.

SUPPLEMENTARY INFORMATION: The U.S. Treasury market is the deepest and most liquid market in the world.[1] It plays a critical and unique role in the global economy, serving as the primary means of financing the U.S. federal government, a significant investment instrument and hedging vehicle for global investors, a risk-free benchmark for other financial instruments, and an important market for the implementation of monetary policy by the Federal Reserve System.

Treasury is seeking public comment on several specific questions that will inform the ongoing work related to the next steps identified in the JSR. This RFI is intended, in part, to seek information and viewpoints from a diverse group of stakeholders, including the general public, buy and sell-side market participants, academics, and industry groups regarding these and other structural changes in the Treasury market, and their implications for the depth, liquidity, and functioning of the market. This RFI is also intended to develop a holistic view of trading and risk management practices across U.S. Treasury futures and cash markets— including the various trading venues and modes of execution present in the cash market—and it seeks input on potential improvements in Treasury market policies, practices, and conduct.

10-year note auction, $21B

$200MM

10-Yr Note Auction
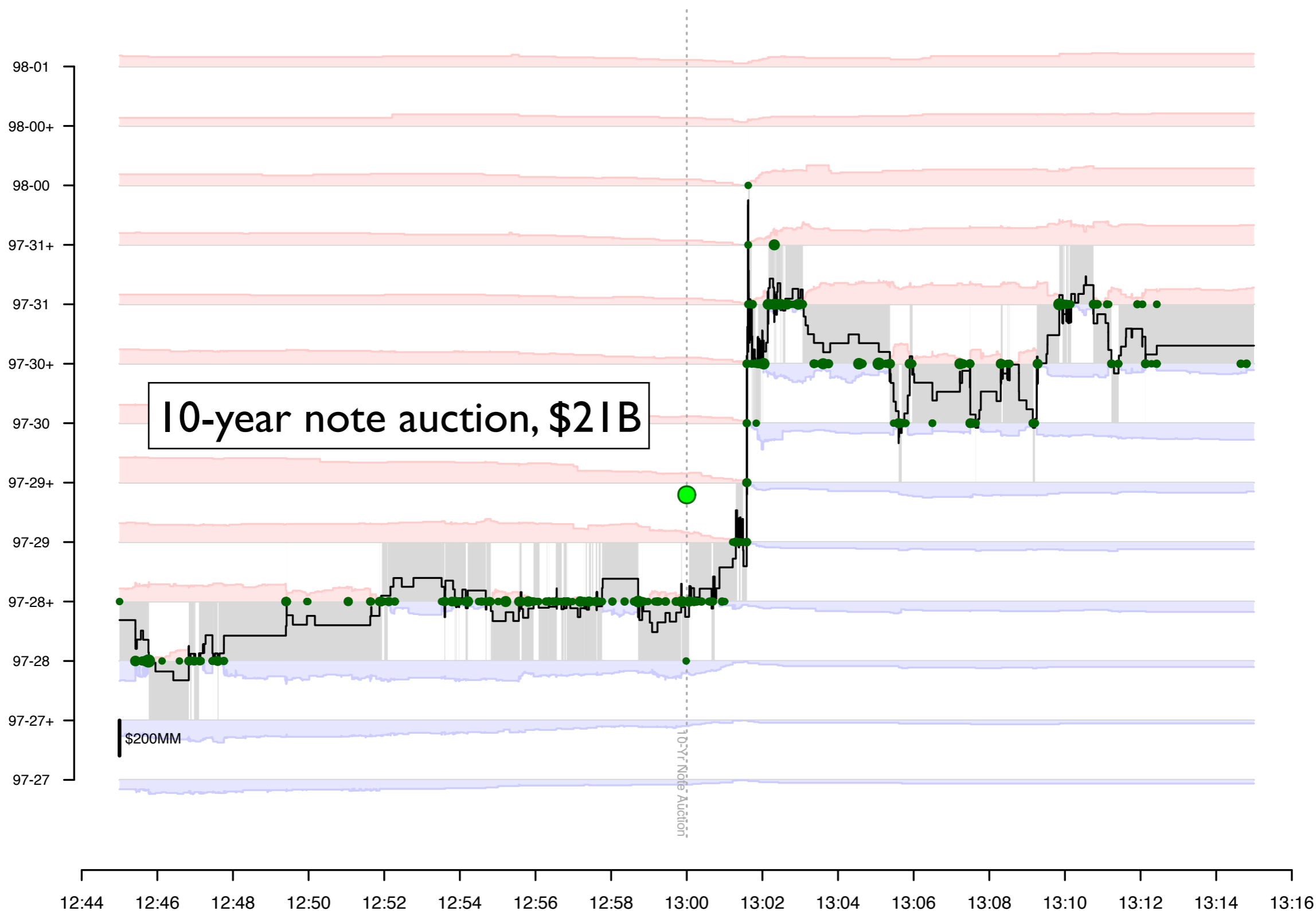
NY time, Sep 9, 2015

FT Trading Room

Investor lawsuits pile up claiming US Treasury market is rigged

Joe Rennison in New York

Investors have filed a flurry of court cases claiming banks and brokers have been rigging the Treasury bond market and increasing the cost of selling debt for the US government.

Twenty-three related cases have been filed, alleging the primary dealers that underwrite the US government's debt colluded to manipulate the price of US Treasuries to their benefit. US Treasury securities are sold through an auction process in which banks and brokers listed as "primary dealers" place bids for the number of bonds they wish to buy and at what price. Investors can use primary dealers to buy at the auction or purchase them directly.

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

CLEVELAND BAKERS AND TEAMSTERS PENSION FUND, CLEVELAND BAKERS AND TEAMSTERS HEALTH AND WELFARE FUND, AND MASTERINVEST KAPITALANLAGE GMBH, on behalf of themselves and all others similarly situated,

Case No.

Plaintiffs,

vs.

BANK OF NOVA SCOTIA, NEW YORK AGENCY; BMO CAPITAL MARKETS CORP.; BNP PARIBAS SECURITIES CORP.; BARCLAYS CAPITAL INC.;

CLASS ACTION COMPLAINT

JURY TRIAL DEMANDED

Regression Results for Model: Post-Auction Yield Change of Newly Auctioned Security = α + β x Yield Change in Secondary Market 2007-2014					
Maturity	Number of Observations	α Coefficient	β Coefficient	RMSE	Adjusted R-Squared
1M	142	-0.000129	0.661***	0.00100	0.13
3M			***	0.00047	0.39
6M	125	-0.000139***	0.084	0.00039	0.01
1Y	36	-0.000146***	0.497***	0.00028	0.31
2Y		-0.000035	0.698***	0.00018	0.71
	52	0.000001	0.906***	0.00004	0.98
5Y	72	-0.000122***	0.671***	0.00025	0.57
		***	0.555***	0.00022	0.46
10Y	83	-0.000068***	0.742***	0.00020	0.65
	71	-0.000093***	0.548***	0.00024	0.55

Note:

*** Designates statistical significance at the 99% confidence level. ** Designates statistical significance at the 95% confidence level. * Designates statistical significance at the 90% confidence level.

9. The data shows that the yields for these identical securities indeed repeatedly diverged between the auction and secondary markets, almost always in the direction of a higher yield (lower price) in the auction relative to the lower yield (higher price) in the secondary markets. Across all tenors (*i.e.*, lengths of time to maturity) of Treasuries, the yields of reissued Treasuries in the primary market were inflated in 69% of the auctions, by 0.91 basis points, a statistically significant disparity. This repeated bias cannot be explained as a result of random chance; instead, the only plausible explanation is that Defendants coordinated artificially to influence the results of the auctions in the primary market.

How exchanges work

Almost all have "continuous-time double-auction limit order book"

no "batching"
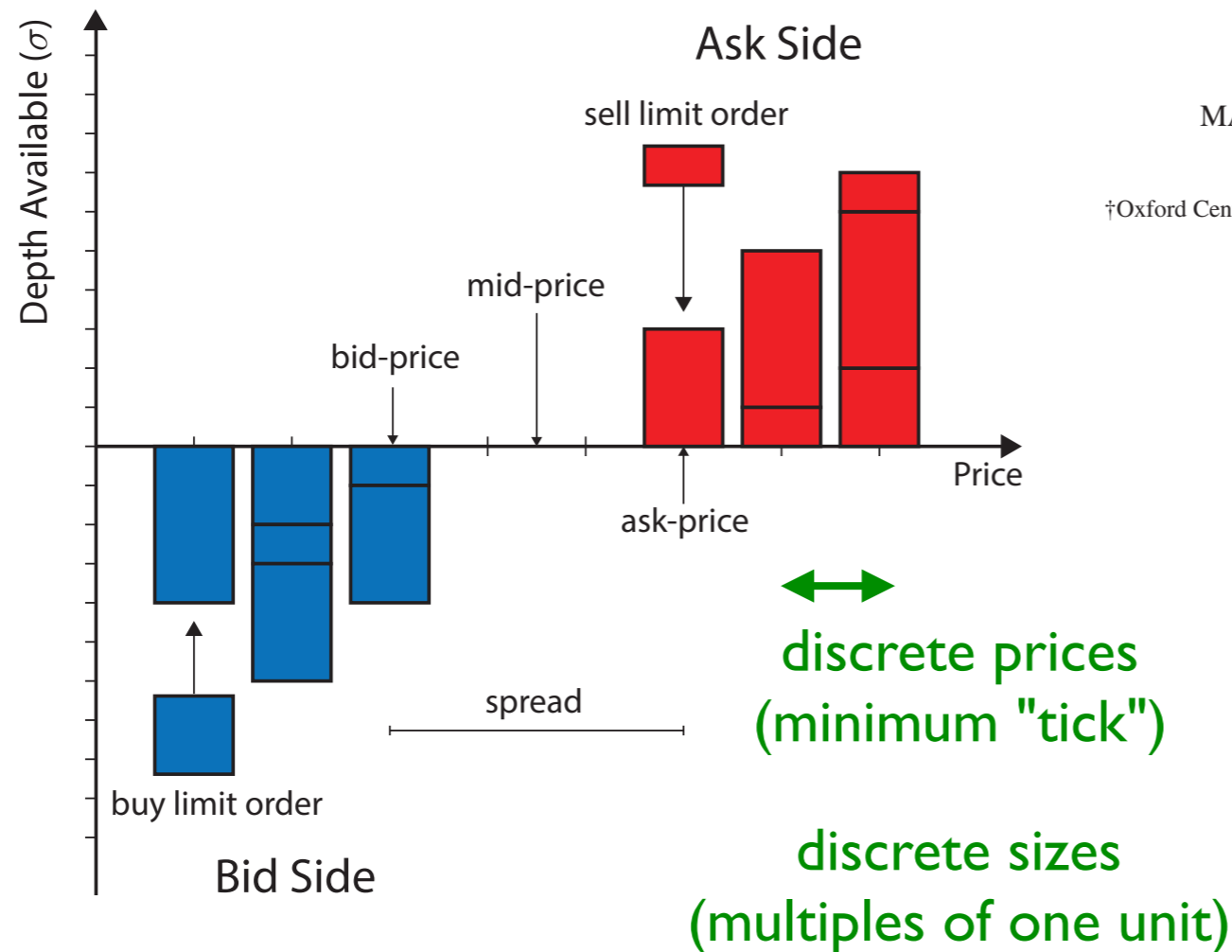
buys/sells symmetric

unfilled orders remain in book

FIG. 1 Schematic of an LOB.

Limit order books

MARTIN D. GOULD*†‡§, MASON A. PORTER†§, STACY WILLIAMS¶, MARK
MCDONALD¶, DANIEL J. FENN¶, and SAM D. HOWISON†‡

†Oxford Centre for Industrial and Applied Mathematics, Mathematical Institute, University of Oxford, Oxford OX1 3LB, UK
‡Oxford-Man Institute of Quantitative Finance, University of Oxford, Oxford, OX2 6ED, UK
§CABDyN Complexity Centre, University of Oxford, Oxford OX1 1HP, UK
¶FX Quantitative Research, HSBC Bank, London E14 5HQ, UK

Limit order books (LOBs) match buyers and sellers in more than half of the world's financial markets. This survey highlights the insights that have emerged from the wealth of empirical and theoretical studies of LOBs. We examine the findings reported by statistical analyses of historical LOB data and discuss how several LOB models provide insight into certain aspects of the mechanism. We also illustrate that many such models poorly resemble real LOBs and that several well-established empirical facts have yet to be reproduced satisfactorily. Finally, we identify several key unresolved questions about LOBs.

Alternative models

Specialist (former NYSE)
only one participant can post limit orders

Request for quote (RFQ)
send message to dealer indicating interest

Discrete time match (auction)
alleviate competition for speed

Time delay (IEX)

Complications

Complicated order types (stop, FOK, peg, etc)

Restrictions on counterparty

Interacting order books (implied quotes)

Fragmented markets "phantom liquidity"

Closely coupled markets (futures/cash/swaps)

Swaps arbitrage from clearing

Use market data to analyze order book

Quotes = changes in order book
  "unaggregated": each separate order
  "aggregated": quantity and number at each price
  "level 1": only inside quotes on each side
  "level 2": full depth (quotes at each price)
Trades = transactions
  triggered by marketable orders
  usually at least price, size, aggressor side
  may also include
    condition codes
    breakdown of order against multiple limit orders

Market data

Market data differs from real state in two ways:

1. Decision by exchange what to release

   Dark pool: release nothing

   NASDAQ ITCH: release full set of orders

   CME: size and number of orders at each level

   BBO: size and number of orders at top of book

2. Technological constraints (latency)

   trades and quotes through different paths

Exchange structure
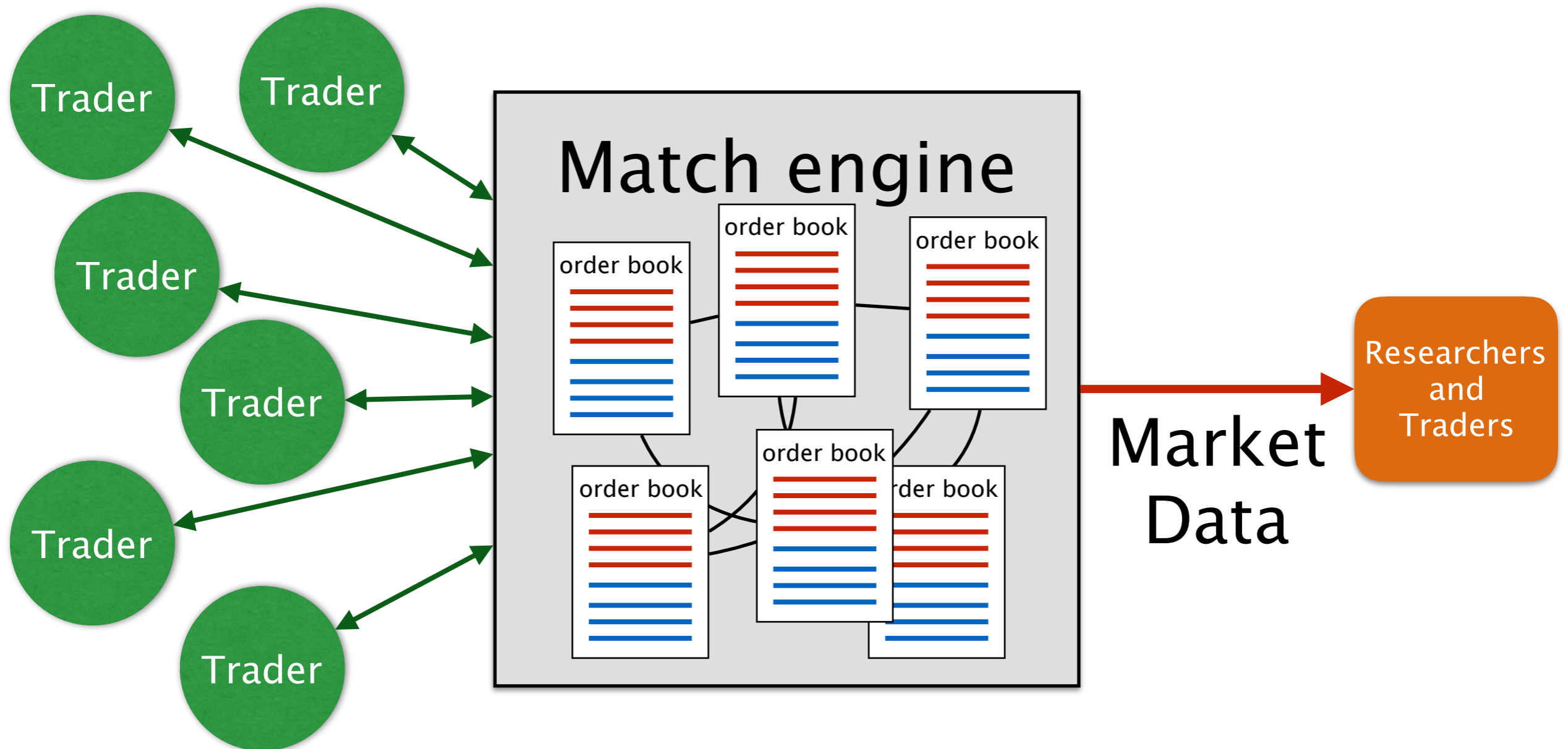
Data content

Trade connections

Trader → Orders, modifications, cancellations → **Match engine**

← Acknowledgments and fill reports
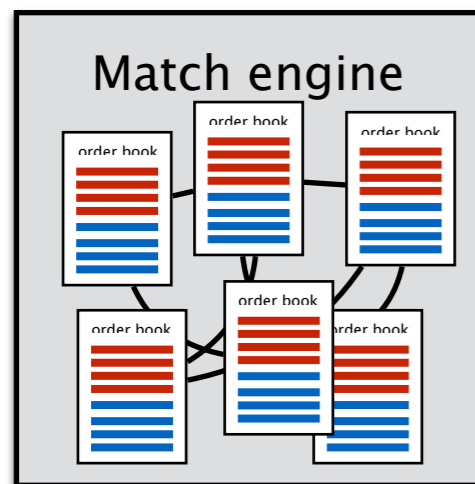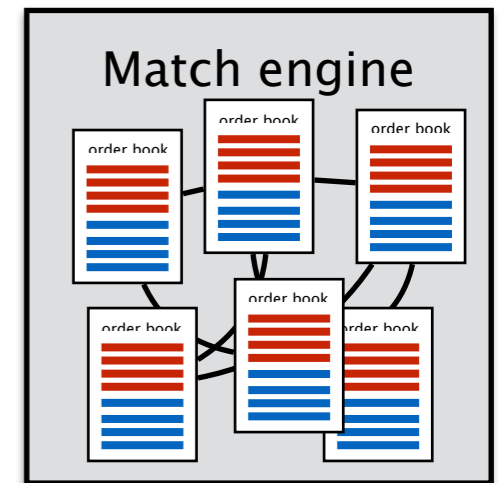
Match engine → **Market Data** →

Subset of full state:
- limited depth
- limited detail
- no client info
- possibly delayed

Challenges in using market data

Understand what process is happening
- matching algorithm
- implied quoting
- self-trade protections

Understand imperfections in data pipeline
- where it is recorded
- latencies and mis-sequencing

Data recording
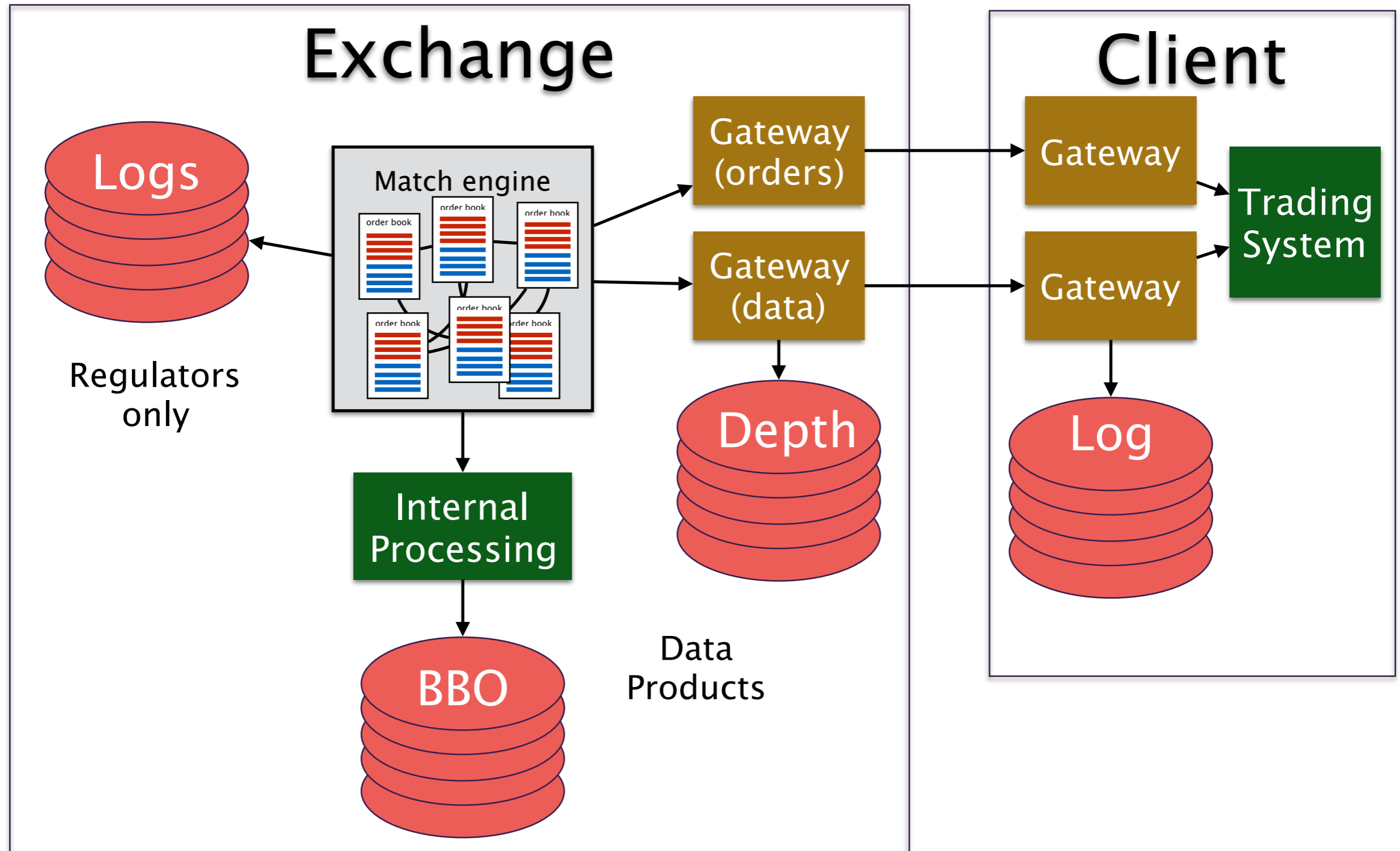
The Flash Crash: The Impact of High Frequency Trading on an Electronic Market*

Andrei Kirilenko—MIT Sloan School of Management & CFTC Chief Economist
Albert S. Kyle—University of Maryland
Mehrdad Samadi—University of North Carolina
Tugkan Tuzun—Board of Governors of the Federal Reserve System

Original Version: October 1, 2010
This version: May 5, 2014

B. The Data

For the day of the Flash Crash and three days prior to that, May 3-6, 2010, we examine transaction-level, "audit-trail" data for all regular transactions in the front-month June 2010 E-mini S&P 500 futures contract. These data come from the Trade Capture Report (TCR) dataset, which the CME provides to the Commodity Futures Trading Commission (CFTC) - the U.S. federal regulator of futures, options, and swaps markets.
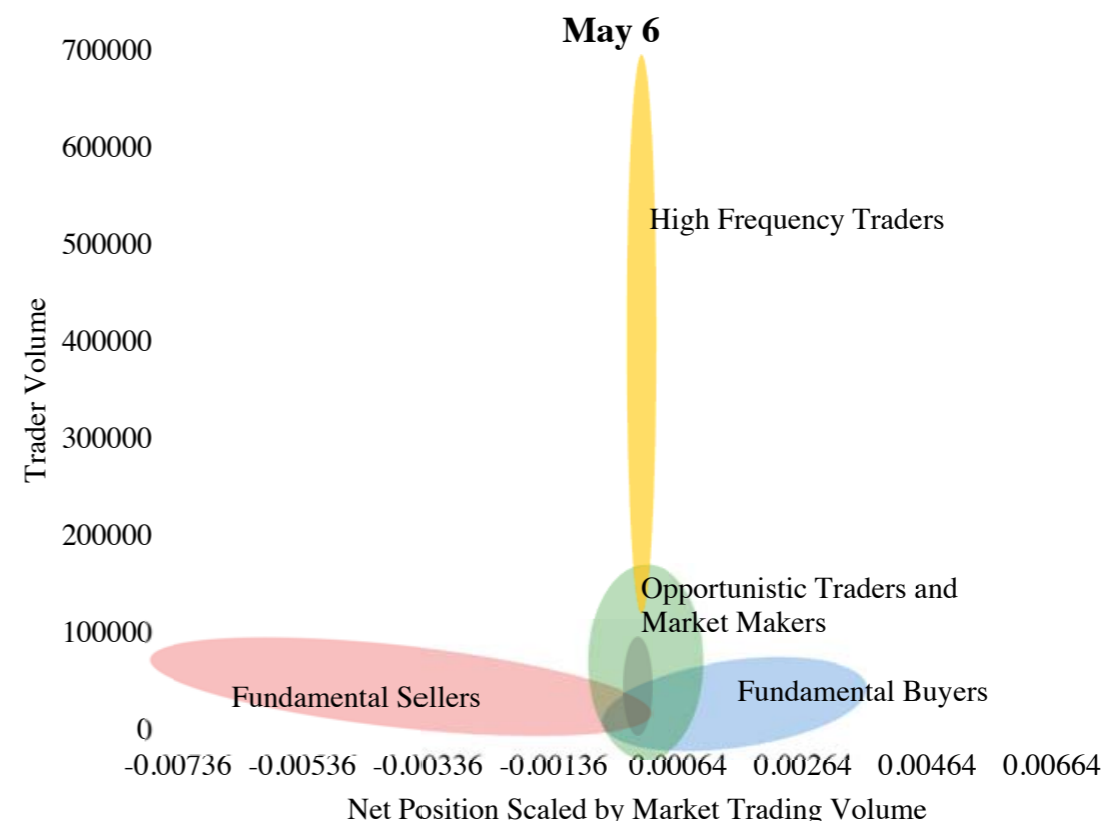
For each transaction, we utilize fields with the account numbers for the buyer and the seller, the price and quantity transacted, the date and time (to the nearest second), a sequence ID number which sorts trades into chronological order even within one second, order type (market order or limit order), and an "aggressiveness" indicator stamped by the CME Globex matching engine - "N" for the resting order and "Y" for the order that executed against a resting order.

The source data is confidential. This means that the results we present often provide a deliberately obscured illustration of what we have actually rigorously established and validated. Moreover, even though we have checked and re-checked our results, they are unlikely to be ever independently validated by other researchers. Even with these limitations though, we still believe that we owe to the public to provide the most informative analysis of the extraordinary stressful events that unfolded in the E-mini on May 6, 2010 and the lessons for market design that we can learn from these events.

May 6

High Frequency Traders

Opportunistic Traders and Market Makers

Fundamental Sellers

Fundamental Buyers

Trader Volume (vertical axis): 0, 100000, 200000, 300000, 400000, 500000, 600000, 700000

Net Position Scaled by Market Trading Volume (horizontal axis): -0.00736, -0.00536, -0.00336, -0.00136, 0.00064, 0.00264, 0.00464, 0.00664

THE WALL STREET JOURNAL.

'Flash Crash' Investigators Likely Missed Clues

Strategies of Navinder Sarao, trader accused of helping cause selloff, could have been spotted sooner, oversight committee members say

By **BRADLEY HOPE** and **ANDREW ACKERMAN**

April 26, 2015 7:03 p.m. ET

While investigators had access to the full set of data from that day, they focused on a subset related to actual trades, the committee members said. Had investigators delved deeper into the bigger set that included all the bids and offers entered, they said, they likely would have noticed that Mr. Sarao single-handedly put enormous pressure on a key futures contract tied to the S&P 500 stock index by making bids and quickly canceling them in a bluffing tactic known as "spoofing."

Historical Data

CME DataMine is the official source of the most comprehensive price information available for CME Group markets. With a broad array of data types including Market Depth, End-of-Day, Block Trades, and more, CME DataMine provides the information customers need to discover insights and capture market opportunity.

Available Data Types

Market Depth	Top-of-Book (BBO)	Time & Sales	End-of-Day	Block Trades

Market Depth

Market Depth FIX files provide all market data messages required to recreate the order book: five to ten orders deep in futures markets and three orders deep in options markets, as well as trade data for all CME Globex-traded products.

Full market data

Market Depth	Top-of-Book (BBO)	Time & Sales	End-of-Day	Block Trades

Top-of-Book (BBO) Data

Designed for those looking to better understand CME Globex liquidity, Top-of-Book contains all top bid, bid size, top ask, ask size, last trade, trade volume, and time-stamp data for CME Globex-traded products. This dataset includes best bids and best offers, with corresponding volume, for all CME Globex products.

Inside quotes + trades

Market Depth	Top-of-Book (BBO)	Time & Sales	End-of-Day	Block Trades

Time and Sales Data

Time and Sales files provide information on trades, as well as bids or offers that better the traded prices. This dataset contains the official record of trade times and prices, in addition to quantities on electronic trades only.

Trades only

FAQ

The CME DataMine Top of Book files (Best Bid/Offer) provide the top bid, bid volume, ask, ask volume, last trade price, and last trade volume of the order book for all CME Globex traded products.

How far back do you maintain Top of Book (BBO) records?

The start date of BBO records is dependent on the product. Product start dates can be found here.

What is the granularity of the timestamp for orders and trades?

Trades and orders are time stamped to the second.

Do the Top of Book files contain the same trade information as the Time & Sales files?

Yes, the Time& Sales information is a subset of the Top of Book data. Top of Book has the added benefit of providing the best bids and offers in addition to trade information.

Does Top of Book come from the same source as Market Depth and Time & Sales data?

Top of Book and Time & Sales files are produced from internal post-trade processing system. Market Depth files come directly from our FIX/FAST market data platform. The granularity of timestamps for the data offerings are different due to the sources.

What is the format of the Top of Book files?

CME DataMine Top of Book files are currently available exclusively in ASCII format. A file layout guide and examples are available here.

How are trades aggregated in the Top of Book files?

Trades that happen in the same second are each given a specific sequence number and are ordered in the same sequence that they were sent out over the market data feed.

Are spreads included in the Top of Book files?

Exchange Defined option spreads are displayed in outright options files.

Sep 2015 NASDAQ 100 futures:
10:00:05 to 10:00:06 CDT, Aug 7 2015

BBO data (research product)

```
q)h"select from trade where
date=2015.08.07,sym=`NQU5,time=10:00:05"
date       sym  time     seq      inst expir   prc     siz cond
-----------------------------------------------------------------
2015.08.07 NQU5 10:00:05 1090101 NQ   2015.09 4502     1
2015.08.07 NQU5 10:00:05 1090102 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090103 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090116 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090157 NQ   2015.09 4502     1
2015.08.07 NQU5 10:00:05 1090180 NQ   2015.09 4501.5  1
2015.08.07 NQU5 10:00:05 1090181 NQ   2015.09 4501.5  1
2015.08.07 NQU5 10:00:05 1090208 NQ   2015.09 4501.5  1
2015.08.07 NQU5 10:00:05 1090259 NQ   2015.09 4501.5  1
2015.08.07 NQU5 10:00:05 1090262 NQ   2015.09 4501.5  1
2015.08.07 NQU5 10:00:05 1090265 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090268 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090269 NQ   2015.09 4501.75 (3)
2015.08.07 NQU5 10:00:05 1090280 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090281 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090282 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090283 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090308 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090313 NQ   2015.09 4502     1
2015.08.07 NQU5 10:00:05 1090314 NQ   2015.09 4502     1
2015.08.07 NQU5 10:00:05 1090315 NQ   2015.09 4502     1
2015.08.07 NQU5 10:00:05 1090316 NQ   2015.09 4501.75 (2)
2015.08.07 NQU5 10:00:05 1090317 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090318 NQ   2015.09 4501.75 1
2015.08.07 NQU5 10:00:05 1090335 NQ   2015.09 4501.75 (4)
2015.08.07 NQU5 10:00:05 1090336 NQ   2015.09 4501.75 (3)
```

```
q)h"select sum siz by prc from
trade where date=2015.08.07,
sym=`NQU5,time=10:00:05"
prc     | siz
--------|  ---
4501.5  | 5
4501.75 | 24
4502    | 5
```

QB data (trading feed)

```
q)h"`rseq xasc select rseq,tptime,prc,siz,aggress,cond,mtch from
trade where date=2015.08.07,sym=`NQU5,tptime within
(2015.08.07T15:00:05;2015.08.07T15:00:06)"
rseq      tptime                    prc    siz aggress cond mtch
----------------------------------------------------------------
12445389 2015.08.07T15:00:05.017 450200 1   S             +
12445392 2015.08.07T15:00:05.017 450175 1   S             +
12445393 2015.08.07T15:00:05.017 450175 1   S
12445418 2015.08.07T15:00:05.019 450200 1   B             +
12445424 2015.08.07T15:00:05.020 450200 1   B             +
12445425 2015.08.07T15:00:05.020 450200 1   B
12445426 2015.08.07T15:00:05.020 450200 1   B
12445429 2015.08.07T15:00:05.025 450175 1   S             +
12445430 2015.08.07T15:00:05.025 450175 1   S
12445431 2015.08.07T15:00:05.025 450175 1   S
12445432 2015.08.07T15:00:05.025 450175 1   S
12445433 2015.08.07T15:00:05.025 450175 (4) S
12445434 2015.08.07T15:00:05.025 450175 (3) S
12445435 2015.08.07T15:00:05.025 450175 (2) S
12445436 2015.08.07T15:00:05.025 450175 1   S
12445437 2015.08.07T15:00:05.025 450175 1   S
12445472 2015.08.07T15:00:05.027 450175 1   S             +
12445499 2015.08.07T15:00:05.030 450175 1   B             +
12445502 2015.08.07T15:00:05.030 450175 1   B             +
12445503 2015.08.07T15:00:05.030 450175 (3) B
12445516 2015.08.07T15:00:05.033 450175 1   B             +
12445569 2015.08.07T15:00:05.171 450150 1   S             +
12445570 2015.08.07T15:00:05.171 450150 1   S
12445572 2015.08.07T15:00:05.171 450150 1   S             +
12445596 2015.08.07T15:00:05.307 450150 1   S             +
12445597 2015.08.07T15:00:05.307 450150 1   S
```
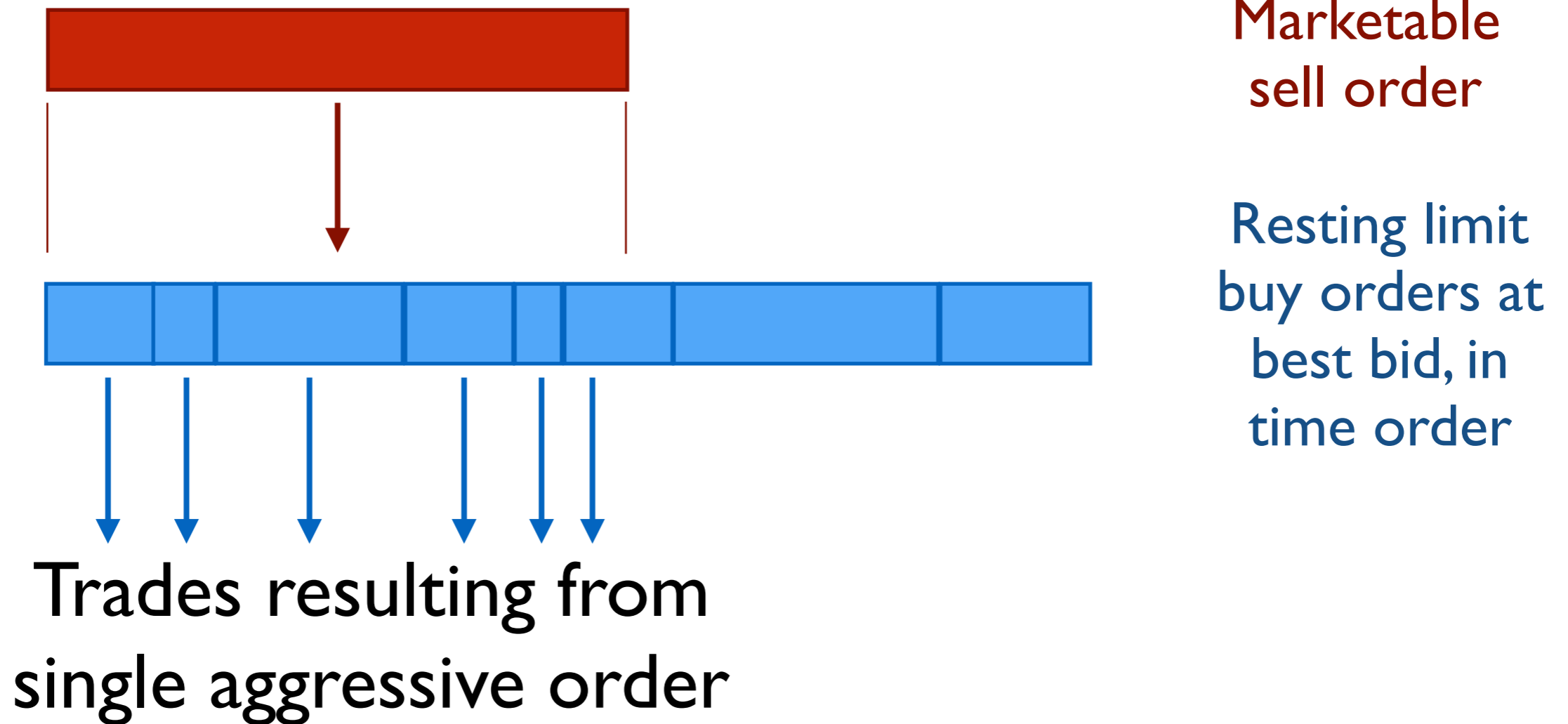
multiple fills from one aggressor

```
q)h"select sum siz by prc from trade where
date=2015.08.07,sym=`NQU5,tptime within
(2015.08.07T15:00:05;2015.08.07T15:00:06)"
prc    | siz
-------|  ---
450150 | 5
450175 | 24
450200 | 5
```

> **Sequence numbers (sequences are different!)**

What is "a trade"

Important for trade size modeling, and "transaction time"

Marketable sell order

Resting limit buy orders at best bid, in time order

Trades resulting from single aggressive order

Need to be able to identify, in market data, trades resulting from single incoming order

Example of careful use of data

Intraday Trading Invariance
in the E-mini S&P 500 Futures Market

Torben G. Andersen, Oleg Bondarenko
Albert S. Kyle and Anna A. Obizhaeva*

First Draft: July 2, 2014

Unfortunately, accurate identification of the number and size of trades by active investors is difficult given the tick record available from most exchanges. First, in electronic order book markets, it is typical to record large trades, which cross several standing limit orders, as separate trades. That is, a marketable order for, say, 12 units may be executed against four different limit orders on the book with sizes of 2, 1, 5, and 4 units, respectively. This will appear as four separate trades at the identical price, reflecting the execution of four (passive) limit orders, rather than as a single trade corresponding to one large active order.

The procedure behind the recording of trades and trade sizes is critical for our empirical tests. When an executable order arrives, it is often matched with more than one limit order resting at the top of the book at the time of execution. During our sample period, the exchange reported all contracts traded at the identical price against an incoming order as a single combined transaction quantity. Thus, the trade size is reported from the perspective of the active party placing the executable order.

Example of understanding market: match trades and quotes

Hypothesis: all trades at bid or ask

Friday, July 1, 2016

Front-month
SP500 futures

```
          tq|  n       v
          --|  ------------
prc<bid   -2|  331     13600
prc=bid   -1|  84696   825738
prc=ask    1|  77385   777455
prc>ask    2|  639     30223
```

Front-month
Crude Oil futures

```
             tq|  n       v
             --|  ----------
prc<bid      -2|  105     1134
prc=bid      -1|  7629    13587
bid<prc<ask   0|  1491    2204
prc=ask       1|  7375    14037
prc>ask       2|  94      623
```
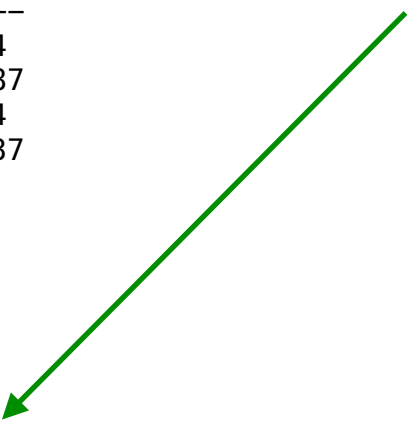
Dec 2017
Crude Oil futures

```
             tq|  n     v
             --|  --------
prc<bid      -2|  2     24
prc=bid      -1|  339   699
bid<prc<ask   0|  392   1165
prc=ask       1|  467   743
prc>ask       2|  3     81
```

Why so many trades within spread?
Implied quotes

Conclusions of lecture 1

Market microstructure is important
  to society
  scientifically
Applied and empirical subject
  need to understand how markets work
Data should be treated very skeptically
  Are trades within bid and ask? Or why not?
  How is data produced? What latencies?