Viewpoint: The Case for Frequent Batch Auctions

Eric Budish
University of Chicago, Booth School of Business

NBER Market Design @25 Conference
Session: Market Design in Financial Markets
May 12th, 2023

The Efficient Markets Hypothesis

- Fama (1970): "A market in which prices always 'fully reflect' available information is called 'efficient'"
- "Obviously an extreme null hypothesis ... we do not expect it to be literally true."
- Distinguishes 3 versions of the EMH, to "pinpoint the level of information at which the hypothesis breaks down"
  - Weak: past prices info
  - Semi-strong: all public info
  - Strong: all public and private info
- Fama concludes no evidence against EMH in weak or semi-strong forms, but evidence against strong form.
  - Translation: to beat the market you have to know something that the rest of the market doesn't know.

Modern Understanding of the EMH

- "We now know that asset prices are very hard to predict over short time horizons, but that they follow movements over longer horizons that, on average, can be forecasted" (2013 Nobel Committee).
- Debate: interpretation of the long-run predictability
  - Risk variation or behavioral inefficiency
  - Magnitudes, especially since non-trivial to exploit
  - (See Cochrane 2011 presidential address)
- Consensus: in short-run, EMH holds up pretty well
  - IGM Experts Panel: 100% agreement that "very few investors, if any, can consistently make accurate predictions about whether the price of an individual stock will rise or fall on a given day."
  - "If it is possible to predict with a high degree of certainty that one asset will increase more in value than another one, there is money to be made. *More importantly, such a situation would reflect a rather basic malfunctioning of the market mechanism.*" (2013 Nobel Committee)

The HFT Arms Race

- In 2010, Spread Networks invests $300mm to dig a high-speed fiber optic cable from NYC to Chicago
- Shaves round-trip data transmission time... from 16ms to 13ms
- Industry observers: 3ms is an "eternity"
- Joke at the time: next innovation will be to dig a tunnel, "avoiding the planet's pesky curvature"
- Joke isn't that funny... Spread's cable quickly obsolete!
- Arms race for speed continues — now commonly measured in microseconds (millionths) and even nanoseconds (billionths)
- As you'll see, on order of $10bn's per year
  - Hardware, software, communications links, and, perhaps most importantly, high-quality human capital.

The HFT Arms Race

- Question: *how could such tiny speed advantages be worth so much money?*

The HFT Arms Race

- Question: *how could such tiny speed advantages be worth so much money?*
- Fundamentals? No. 3 milliseconds too short to be about fundamentals.
  - Quarterly earnings released once per 8 billion ms ... and after market is closed!

The HFT Arms Race

- Question: *how could such tiny speed advantages be worth so much money?*
- Fundamentals? No. 3 milliseconds too short to be about fundamentals.
  - Quarterly earnings released once per 8 billion ms ... and after market is closed!
- Technical? Economists intrinsically skeptical.
  - "Technical strategies are usually amusing, often comforting, but of no real value." (Burton Malkiel, "A Random Walk Down Wall Street"
  - "A rather basic malfunctioning of the market mechanism"

Answer: Flawed Market Design

- The market design most widely used in financial markets around the world, called the "continuous limit order book":
  - Treats time as *continuous*
  - Processes requests to trade *serially*

Answer: Flawed Market Design

- The market design most widely used in financial markets around the world, called the "continuous limit order book":
  - Treats time as *continuous*
  - Processes requests to trade *serially*
- Continuous-time + serial processing $\rightarrow$ riskless arbitrage profits from symmetric public information

Answer: Flawed Market Design

- The market design most widely used in financial markets around the world, called the "continuous limit order book":
  - Treats time as *continuous*
  - Processes requests to trade *serially*
- Continuous-time + serial processing $\rightarrow$ riskless arbitrage profits from symmetric public information
- That is... *a violation of the weak-form and semi-strong form EMH, built directly into the market design.*

Answer: Flawed Market Design

- The market design most widely used in financial markets around the world, called the "continuous limit order book":
  - Treats time as *continuous*
  - Processes requests to trade *serially*
- Continuous-time + serial processing $\rightarrow$ riskless arbitrage profits from symmetric public information
- That is... *a violation of the weak-form and semi-strong form EMH, built directly into the market design.*
- These riskless arbitrage profits
  1. Are not supposed to exist in a well-functioning market
  2. Harm liquidity
  3. Induce a never-ending arms race for speed

Answer: Flawed Market Design

- The market design most widely used in financial markets around the world, called the "continuous limit order book":
  - Treats time as *continuous*
  - Processes requests to trade *serially*
- Continuous-time + serial processing $\rightarrow$ riskless arbitrage profits from symmetric public information
- That is... *a violation of the weak-form and semi-strong form EMH, built directly into the market design.*
- These riskless arbitrage profits
  1. Are not supposed to exist in a well-functioning market
  2. Harm liquidity
  3. Induce a never-ending arms race for speed
- Market design solution: put time into units ("discrete time") and process requests to trade in *batch*, using auctions.

Plan for Talk

- Part I: "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response" (Budish, Cramton and Shim, QJE, 2015)
- Part II: "Quantifying the High-Frequency Trading 'Arms Race'" (Aquilina, Budish and O'Neill, QJE, 2022)
- Part III: "A Theory of Stock Exchange Competition and Innovation: Will the Market Fix the Market?" (Budish, Lee and Shim, R&R JPE, 2023)
- Conclusion: Some Directions for Future Research

"The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response"

Eric Budish
Peter Cramton
John Shim

QJE, November 2015

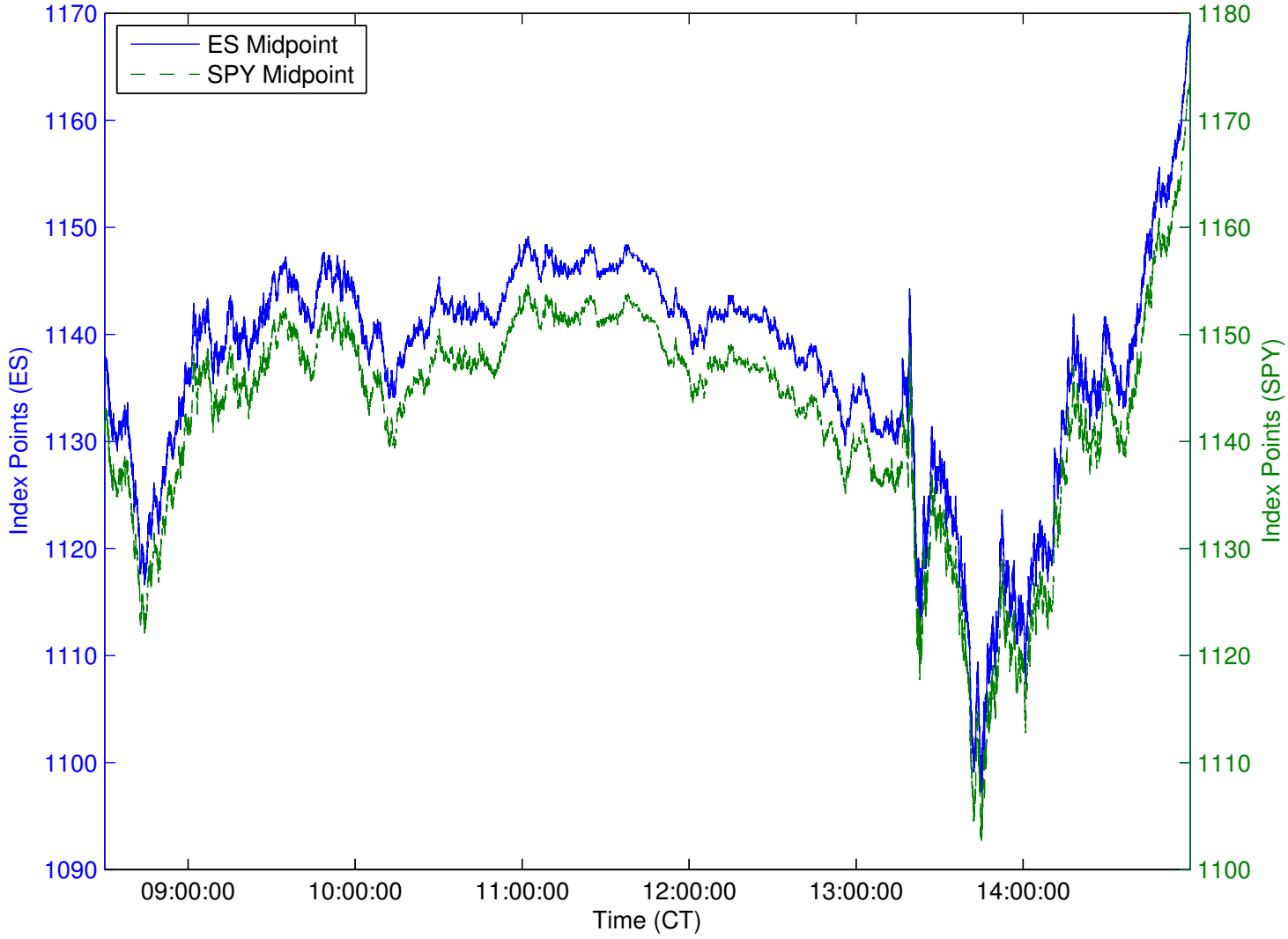
The Case for Frequent Batch Auctions

A simple idea: discrete-time trading.

1. **Empirical Facts: continuous market violates basic asset pricing principles at HFT time horizons.**
   - **Market correlations completely break down.**
   - **Frequent mechanical arbitrage opportunities.**
   - **Mechanical arbs $\rightarrow$ arms race. Arms race does not compete away the arbs, looks like a "constant".**
2. Theory: root flaw is continuous-time serial-process trading
   - Mechanical arbs are "built in" to market design. Sniping.
   - Harms liquidity.
   - Induces never-ending, wasteful, arms race for speed.
3. Solution: frequent batch auctions
   - Competition on speed $\rightarrow$ competition on price.
   - Enhances liquidity and stops the arms race.
   - Simplifies the market computationally.

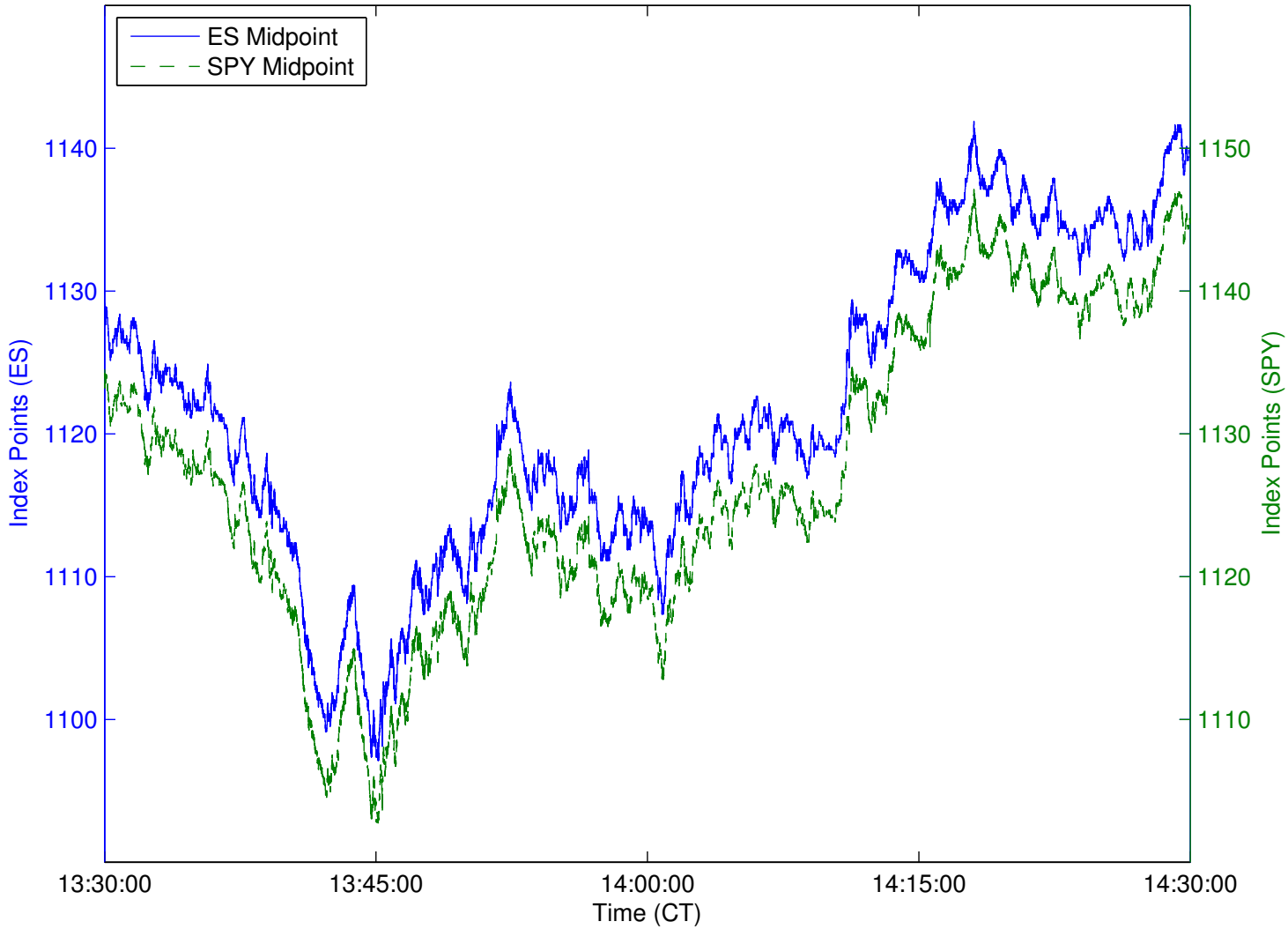
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 Day



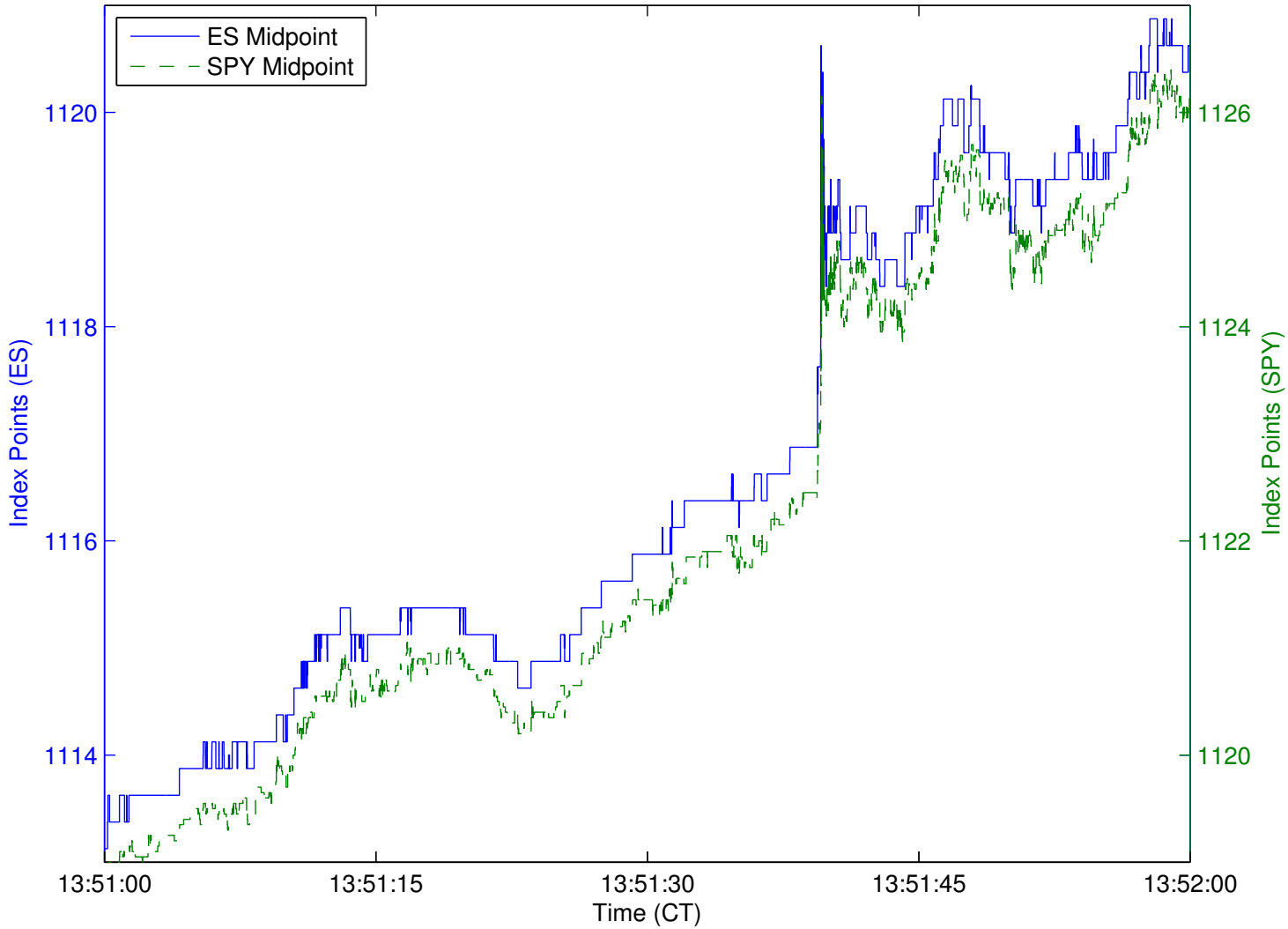
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 hour



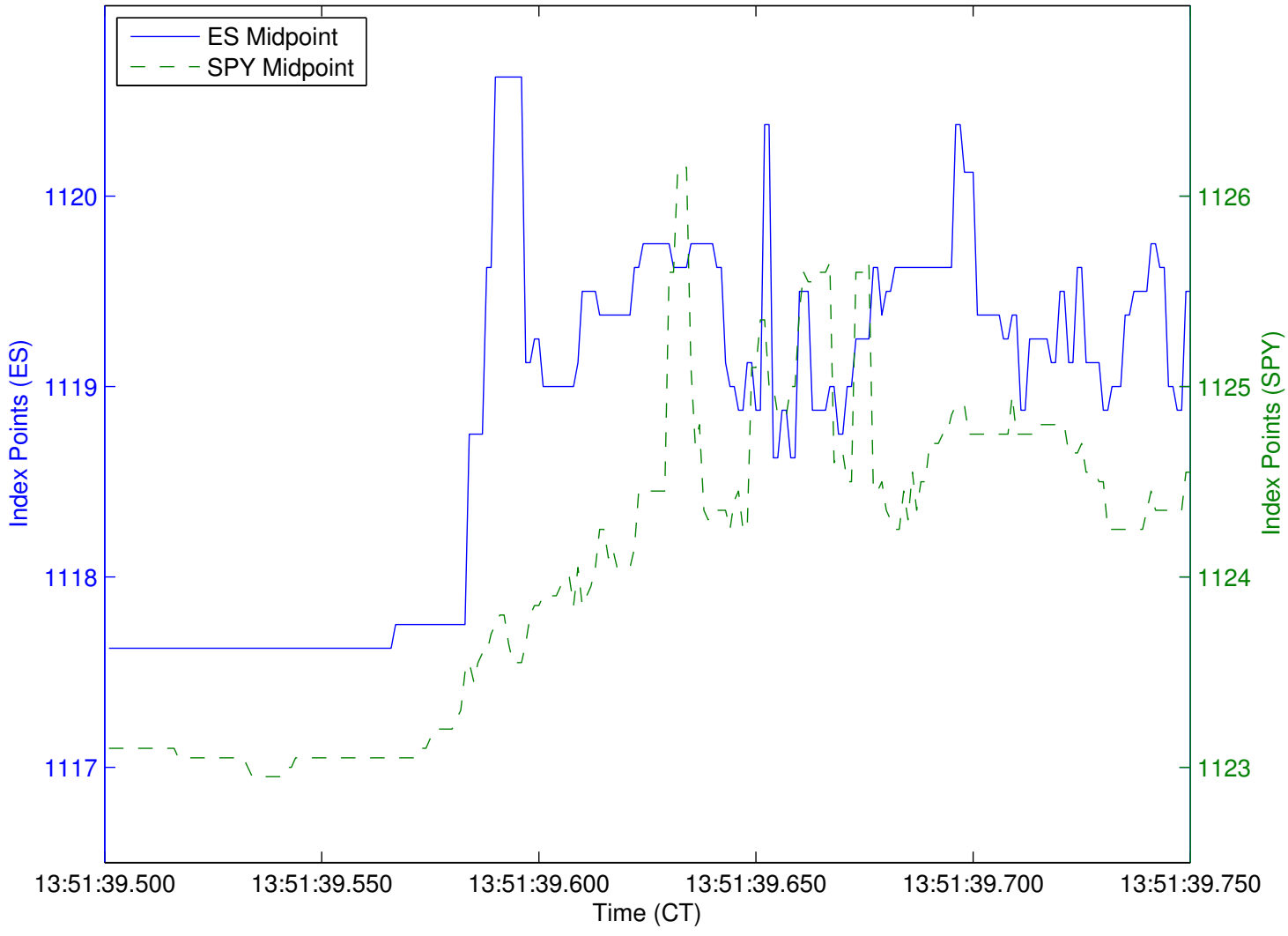
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 minute



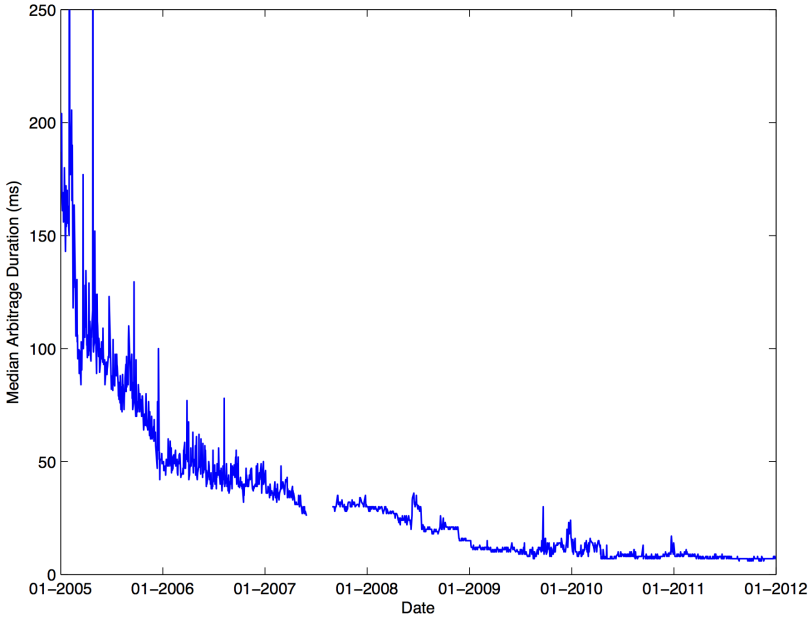
Market Correlations Break Down at High Frequency

ES vs. SPY: 250 milliseconds

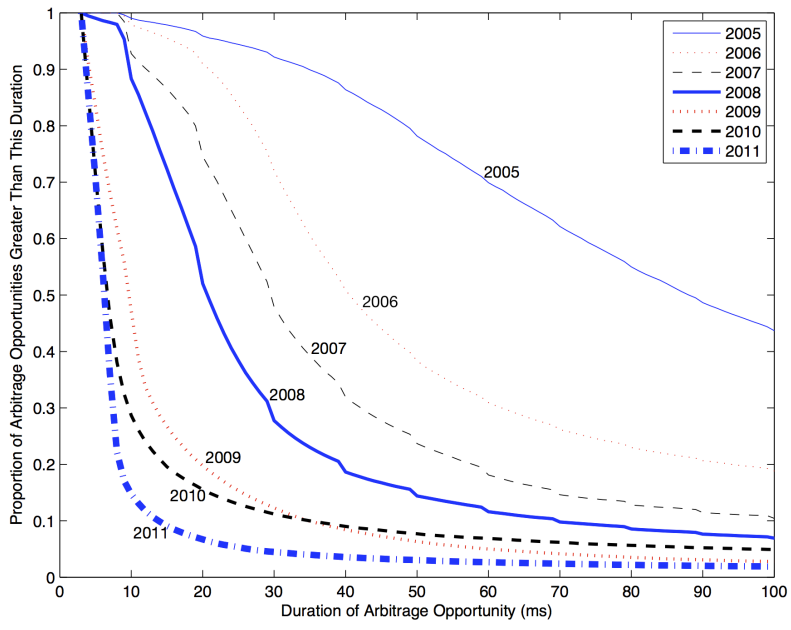


Arb Durations over Time: 2005-2011

Median over time

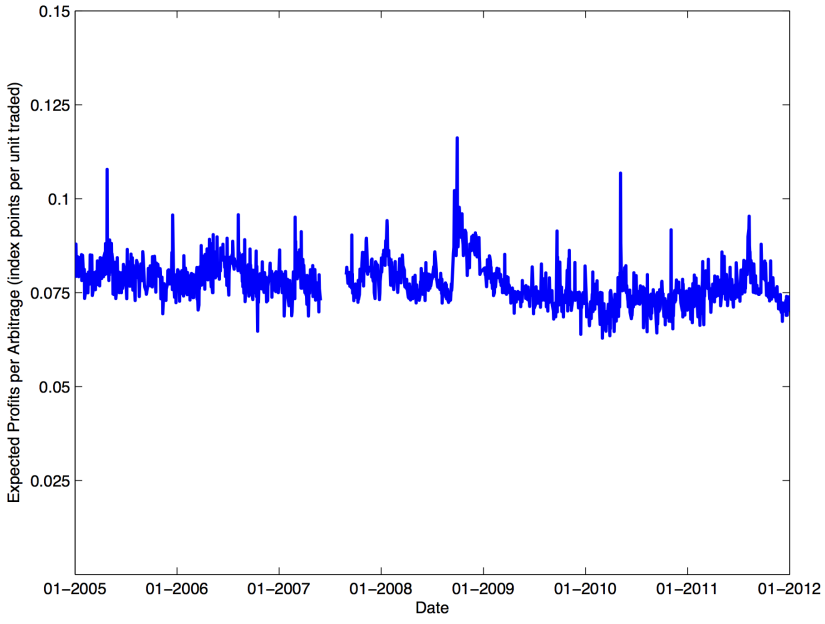


Distribution by year

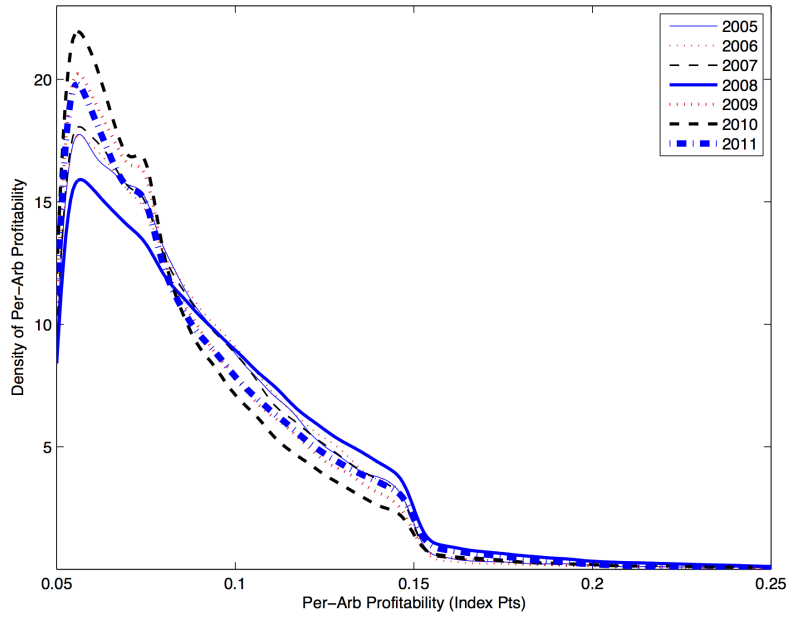


Arb Per-Unit Profits over Time: 2005-2011

Median over time

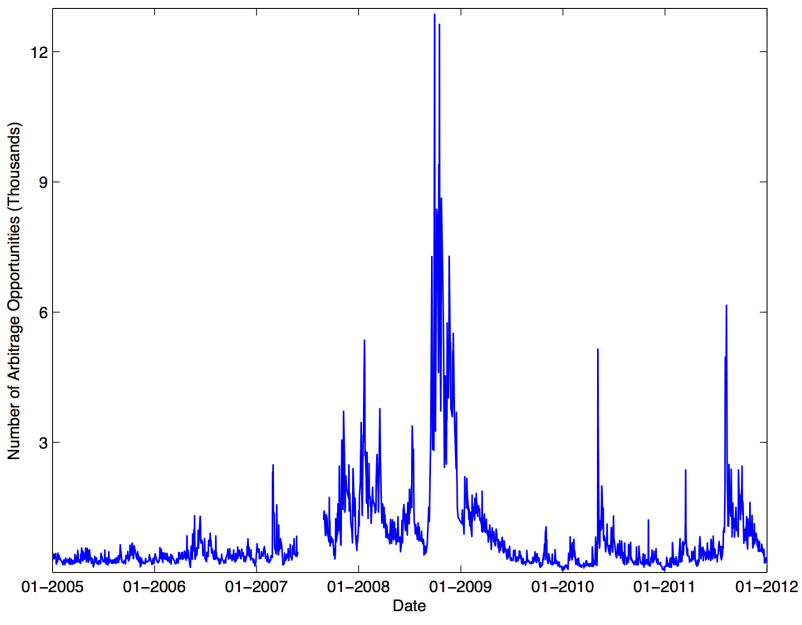


Distribution by year

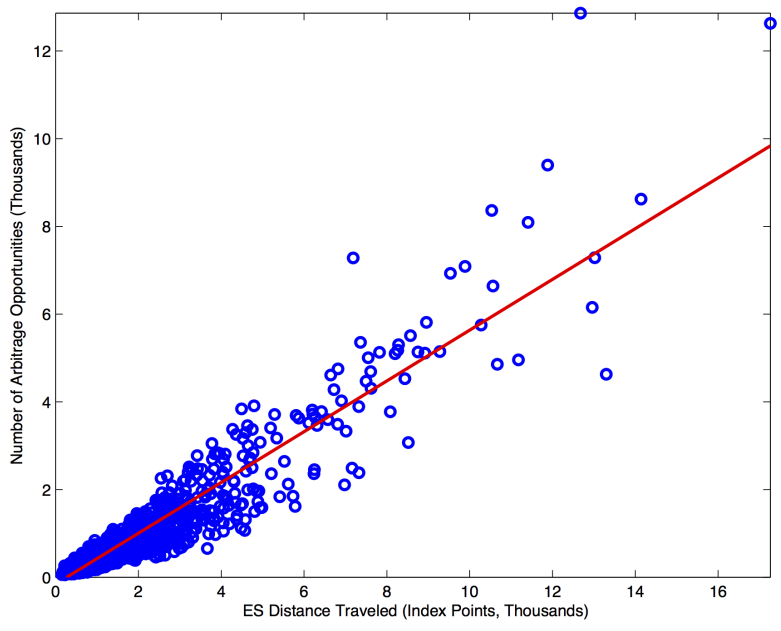


Arb Frequency over Time: 2005-2011

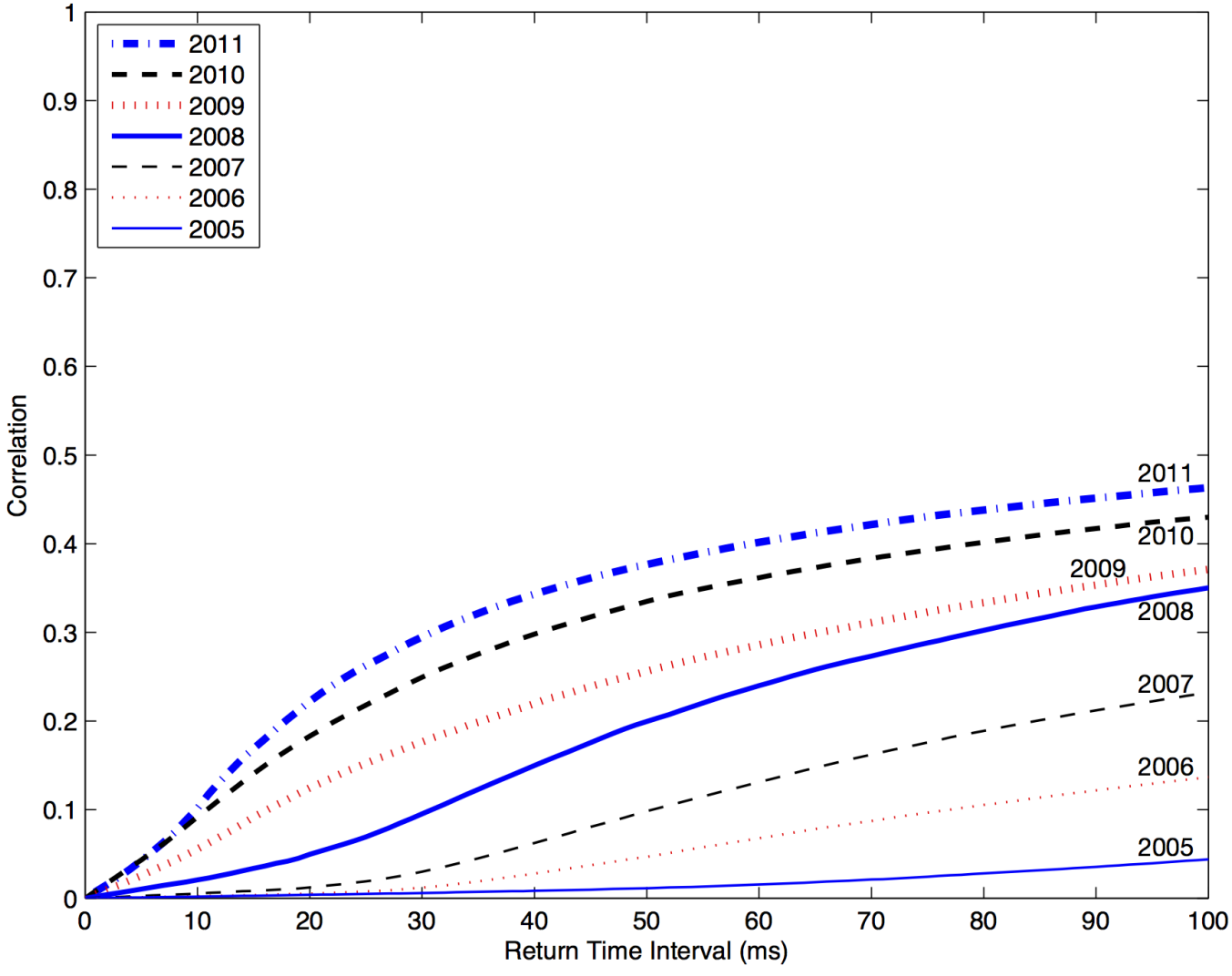
Frequency over time



Frequency vs. Volatility



Correlation Breakdown Over Time: 2005-2011



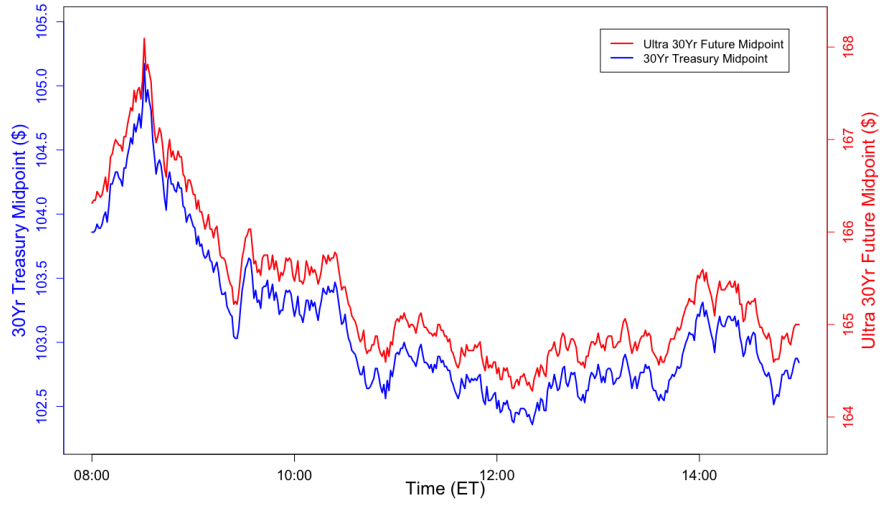
Races, Races, Races

- And ES-SPY is just the tip of the iceberg in the race for speed:
  1. Hundreds of trades very similar to ES-SPY: highly correlated, highly liquid

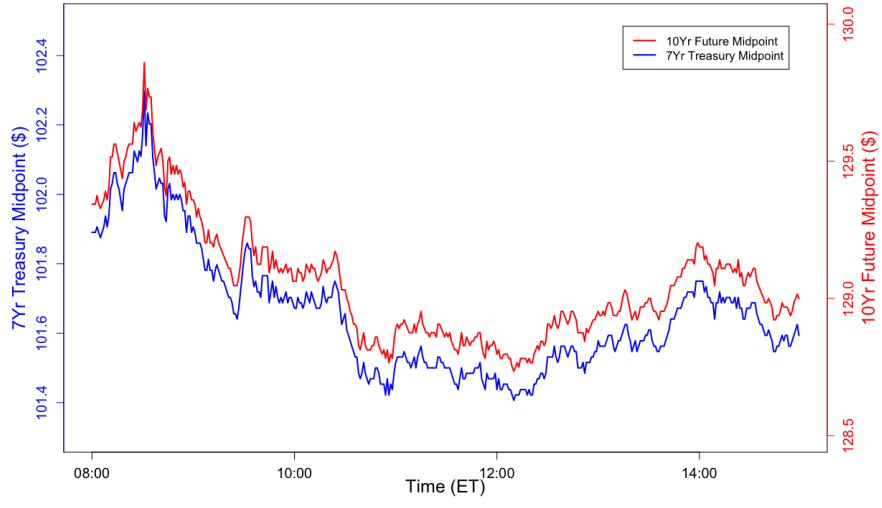
Highly Correlated Pairs

US Treasuries

30 Year Ultra Future vs. 30 Year Cash



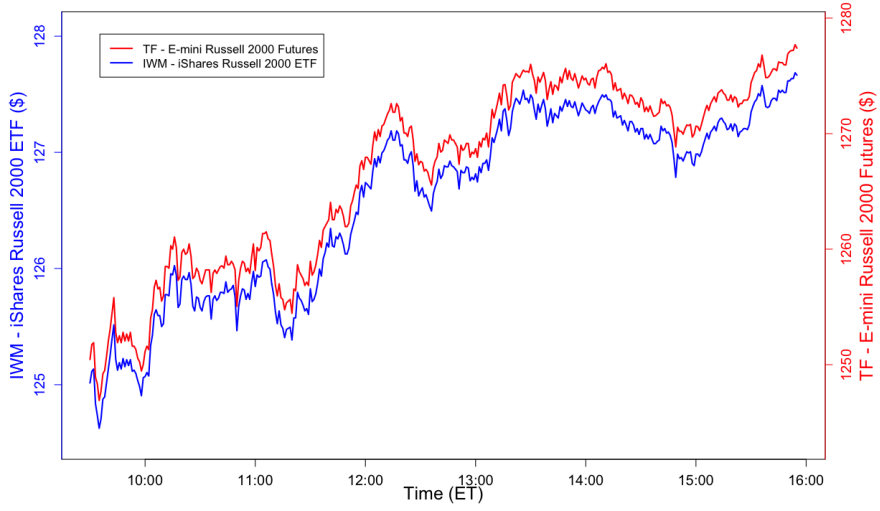
10 Year Future vs. 7 Year Cash



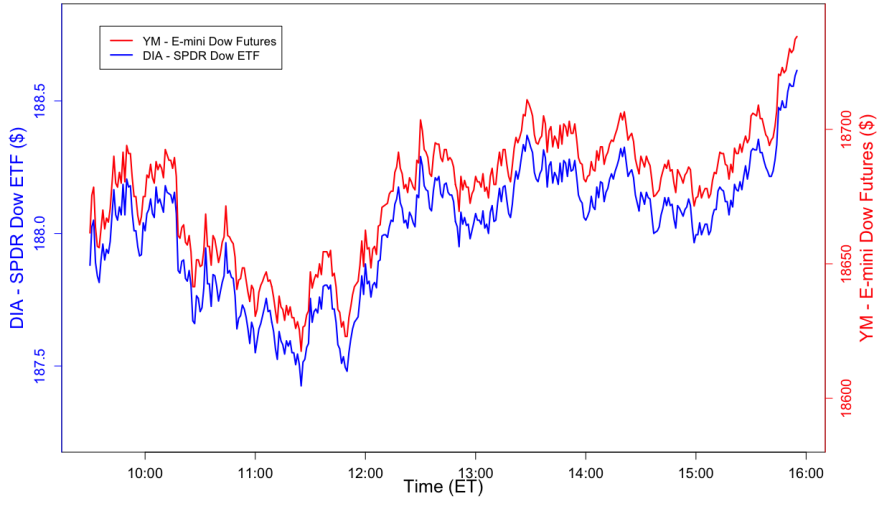
Highly Correlated Pairs

Equity Index

Russell 2000 Future vs. ETF



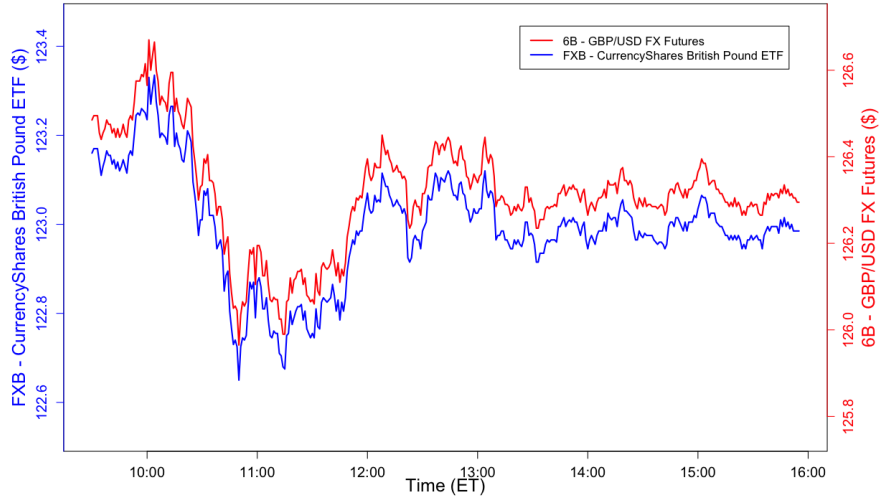
DOW Future vs. ETF



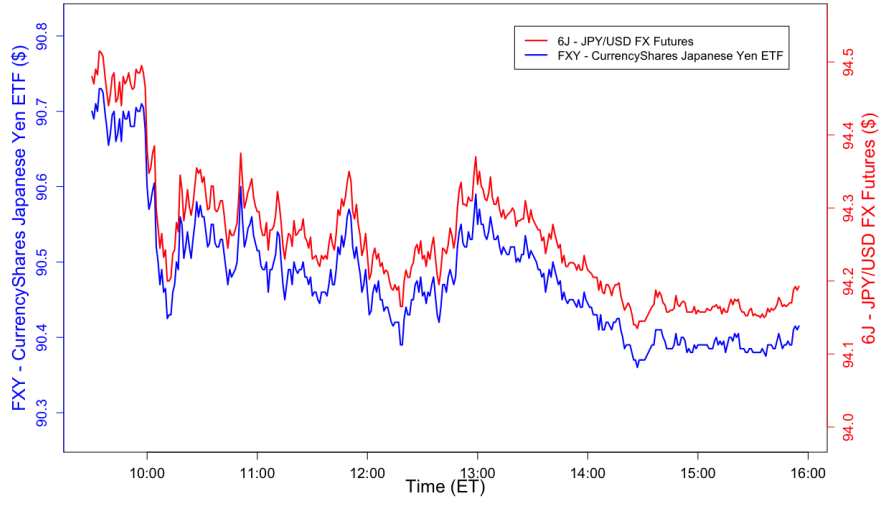
Highly Correlated Pairs

Foreign Exchange

GBP/USD Future vs. ETF



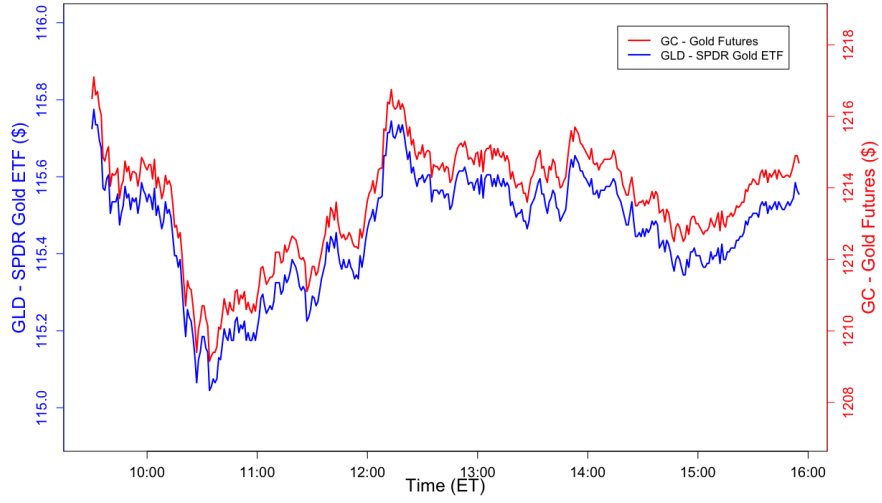
JPY/USD Future vs. ETF



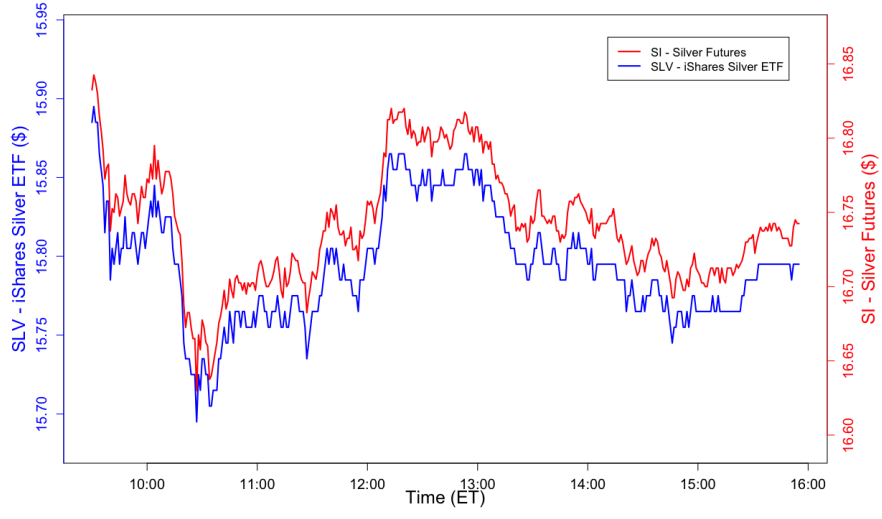
Highly Correlated Pairs

Commodities

Gold Future vs. ETF



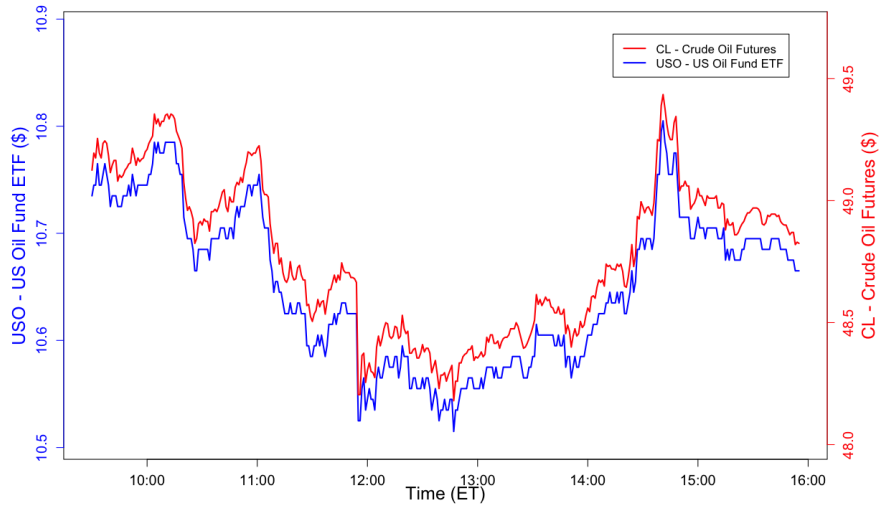
Silver Future vs. ETF



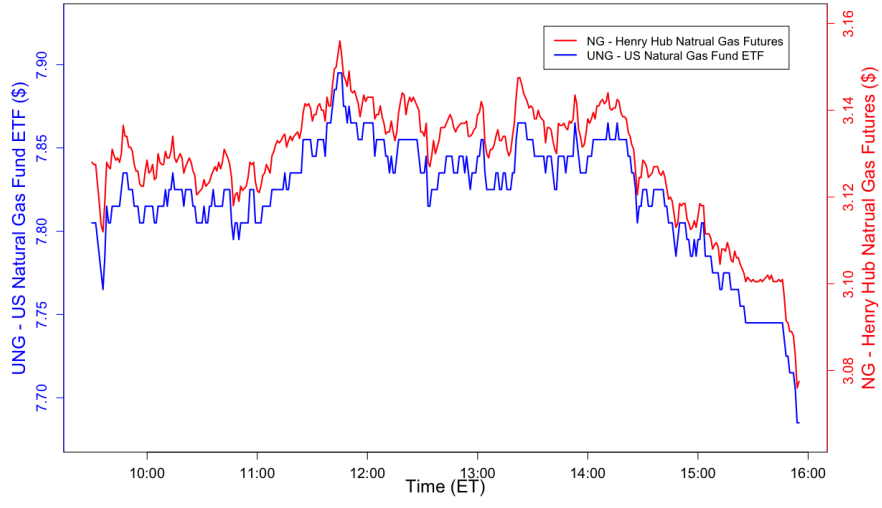
Highly Correlated Pairs

Commodities

Crude Oil Future vs. ETF

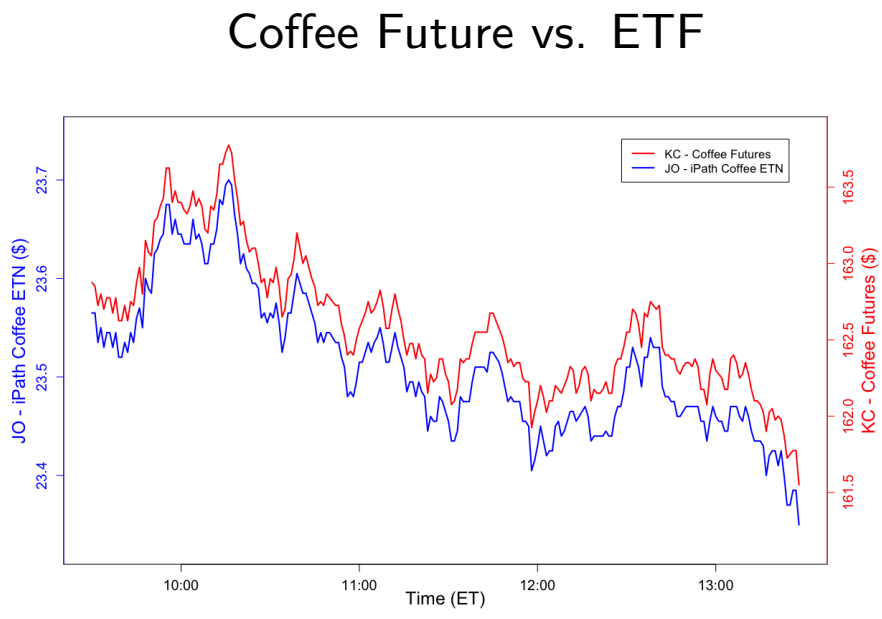


Natural Gas Future vs. ETF



Highly Correlated Pairs

Commodities



Other Highly Correlated Pairs

Partial List

E-mini S&P 500 Futures (ES) vs. SPDR S&P 500 ETF (SPY)
E-mini S&P 500 Futures (ES) vs. iShares S&P 500 ETF (IVV)
E-mini S&P 500 Futures (ES) vs. Vanguard S&P 500 ETF (VOO)
E-mini S&P 500 Futures (ES) vs. ProShares Ultra (2x) S&P 500 ETF (SSO)
E-mini S&P 500 Futures (ES) vs. ProShares UltraPro (3x) S&P 500 ETF (UPRO)
E-mini S&P 500 Futures (ES) vs. ProShares Short S&P 500 ETF (SH)
E-mini S&P 500 Futures (ES) vs. ProShares Ultra (2x) Short S&P 500 ETF (SDS)
E-mini S&P 500 Futures (ES) vs. ProShares UltraPro (3x) Short S&P 500 ETF (SPXU)
E-mini S&P 500 Futures (ES) vs. 500 Constituent Stocks
E-mini S&P 500 Futures (ES) vs. 9 Select Sector SPDR ETFs
E-mini S&P 500 Futures (ES) vs. E-mini Dow Futures (YM)
E-mini S&P 500 Futures (ES) vs. E-mini Nasdaq 100 Futures (NQ)
E-mini S&P 500 Futures (ES) vs. E-mini S&P MidCap 400 Futures (EMD)
E-mini S&P 500 Futures (ES) vs. Russell 2000 Index Mini Futures (TF)
E-mini Dow Futures (YM) vs. SPDR Dow Jones Industrial Average ETF (DIA)
E-mini Dow Futures (YM) vs. ProShares Ultra (2x) Dow 30 ETF (DDM)
E-mini Dow Futures (YM) vs. ProShares UltraPro (3x) Dow 30 ETF (UDOW)
E-mini Dow Futures (YM) vs. ProShares Short Dow 30 ETF (DOG)
E-mini Dow Futures (YM) vs. ProShares Ultra (2x) Short Dow 30 ETF (DXD)
E-mini Dow Futures (YM) vs. ProShares UltraPro (3x) Short Dow 30 ETF (SDOW)
E-mini Dow Futures (YM) vs. 30 Constituent Stocks
E-mini Nasdaq 100 Futures (NQ) vs. ProShares QQQ Trust ETF (QQQ)
E-mini Nasdaq 100 Futures (NQ) vs. Technology Select Sector SPDR (XLK)
E-mini Nasdaq 100 Futures (NQ) vs. 100 Constituent Stocks
Russell 2000 Index Mini Futures (TF) vs. iShares Russell 2000 ETF (IWM)
Euro Stoxx 50 Futures (FESX) vs. Xetra DAX Futures (FDAX)
Euro Stoxx 50 Futures (FESX) vs. CAC 40 Futures (FCE)
Euro Stoxx 50 Futures (FESX) vs. iShares MSCI EAFE Index Fund (EFA)
Nikkei 225 Futures (NIY) vs. MSCI Japan Index Fund (EWJ)
Financial Sector SPDR (XLF) vs. Constituents
Financial Sector SPDR (XLF) vs. Direxion Daily Financial Bull 3x (FAS)
Energy Sector SPDR (XLE) vs. Constituents
Industrial Sector SPDR (XLI) vs. Constituents
Cons. Staples Sector SPDR (XLP) vs. Constituents
Materials Sector SPDR (XLB) vs. Constituents
Utilities Sector SPDR (XLU) vs. Constituents
Technology Sector SPDR (XLK) vs. Constituents
Health Care Sector SPDR (XLV) vs. Constituents
Cons. Discretionary Sector SPDR (XLY) vs. Constituents
SPDR Homebuilders ETF (XHB) vs. Constituents
SPDR S&P 500 Retail ETF (XRT) vs. Constituents
Euro FX Futures (6E) vs. Spot EURUSD
Japanese Yen Futures (6J) vs. Spot USDJPY
British Pound Futures (6B) vs. Spot GBPUSD
Australian Dollar Futures (6B) vs. Spot AUDUSD
Swiss Franc Futures (6S) vs. Spot USDCHF
Canadian Dollar Futures (6C) vs. Spot USDCAD
Gold Futures (GC) vs. miNY Gold Futures (QO)
Gold Futures (GC) vs. Spot Gold (XAUUSD)
Gold Futures (GC) vs. E-micro Gold Futures (MGC)
Gold Futures (GC) vs. SPDR Gold Trust (GLD)
Gold Futures (GC) vs. iShares Gold Trust (IAU)
miNY Gold Futures (QO) vs. E-micro Gold Futures (MGC)
miNY Gold Futures (QO) vs. Spot Gold (XAUUSD)
miNY Gold Futures (QO) vs. SPDR Gold Trust (GLD)
miNY Gold Futures (QO) vs. iShares Gold Trust (IAU)
E-micro Gold Futures (MGC) vs. SPDR Gold Trust (GLD)
E-micro Gold Futures (MGC) vs. iShares Gold Trust (IAU)
E-micro Gold Futures (MGC) vs. Spot Gold (XAUUSD)
Market Vectors Gold Miners (GDX) vs. Direxion Daily Gold Miners Bull 3x (NUGT)
Silver Futures (SI) vs. miNY Silver Futures (QI)
Silver Futures (SI) vs. iShares Silver Trust (SLV)
Silver Futures (SI) vs. Spot Silver (XAGUSD)
miNY Silver Futures (QI) vs. iShares Silver Trust (SLV)
miNY Silver Futures (QI) vs. Spot Silver (XAGUSD)
Platinum Futures (PL) vs. Spot Platinum (XPTUSD)
Palladium Futures (PA) vs. Spot Palladium (XPDUSD)
Eurodollar Futures Front Month (ED) vs. (12 back month contracts)
10 Yr Treasury Note Futures (ZN) vs. 5 Yr Treasury Note Futures (ZF)
10 Yr Treasury Note Futures (ZN) vs. 30 Yr Treasury Bond Futures (ZB)
10 Yr Treasury Note Futures (ZN) vs. 7-10 Yr Treasury Note
2 Yr Treasury Note Futures (ZT) vs. 1-2 Yr Treasury Note
2 Yr Treasury Note Futures (ZT) vs. iShares Barclays 1-3 Yr Treasury Fund (SHY)
5 Yr Treasury Note Futures (ZF) vs. 4-5 Yr Treasury Note
30 Yr Treasury Bond Futures (ZB) vs. iShares Barclays 20 Yr Treasury Fund (TLT)
30 Yr Treasury Bond Futures (ZB) vs. ProShares UltraShort 20 Yr Treasury Fund (TBT)
30 Yr Treasury Bond Futures (ZB) vs. ProShares Short 20 Year Treasury Fund (TBF)
30 Yr Treasury Bond Futures (ZB) vs. 15+ Yr Treasury Bond
Crude Oil Futures Front Month (CL) vs. (6 back month contracts)
Crude Oil Futures (CL) vs. ICE Brent Crude (B)
Crude Oil Futures (CL) vs. United States Oil Fund (USO)
Crude Oil Futures (CL) vs. ProShares Ultra DJ-UBS Crude Oil (UCO)
Crude Oil Futures (CL) vs. iPath S&P Crude Oil Index (OIL)
ICE Brent Crude Front Month (B) vs. (6 back month contracts)
ICE Brent Crude (B) vs. United States Oil Fund (USO)
ICE Brent Crude (B) vs. ProShares Ultra DJ-UBS Crude Oil (UCO)
ICE Brent Crude (B) vs. iPath S&P Crude Oil Index (OIL)
Natural Gas (Henry Hub) Futures (NG) vs. United States Nat Gas Fund (UNG)

Races, Races, Races

- And ES-SPY is just the tip of the iceberg in the race for speed:
  1. Hundreds of trades very similar to ES-SPY: highly correlated, highly liquid
  2. Fragmented equity markets: can arbitrage SPY on NYSE against SPY on NASDAQ! Even simpler than ES-SPY.
  3. Race to respond to public news (eg Business Wire, Fed)
  4. Race to top of book (artifact of minimum price tick)

The Case for Frequent Batch Auctions

A simple idea: discrete-time trading.

1. Empirical Facts: continuous market violates basic asset pricing principles at HFT time horizons.
   - Market correlations completely break down.
   - Frequent mechanical arbitrage opportunities.
   - Mechanical arbs $\rightarrow$ arms race. Arms race does not compete away the arbs, looks like a "constant".
2. **Theory: root flaw is continuous-time serial-process trading**
   - **Mechanical arbs are "built in" to market design. Sniping.**
   - **Harms liquidity.**
   - **Induces never-ending, wasteful, arms race for speed.**
3. Solution: frequent batch auctions
   - Competition on speed $\rightarrow$ competition on price.
   - Enhances liquidity and stops the arms race.
   - Simplifies the market computationally.

Model: High-Level Idea

- Descendant of the famous Glosten Milgrom (1985) model
- Security x that trades on a continuous limit-order book market
- Publicly observable signal y of the value of security x. Jumps around Poisson.
- Purposefully strong assumption:
  - Fundamental value of x is *perfectly* correlated to the public signal y
  - x can always be costlessly liquidated at this fundamental value
  - Goal: "best case" scenario for price discovery and liquidity provision
- Players:
  - Investors: arrive stochastically, want to buy or sell one unit. No information.
  - Trading Firms: always present. Goal is to buy x at prices lower than y and sell at prices higher than y

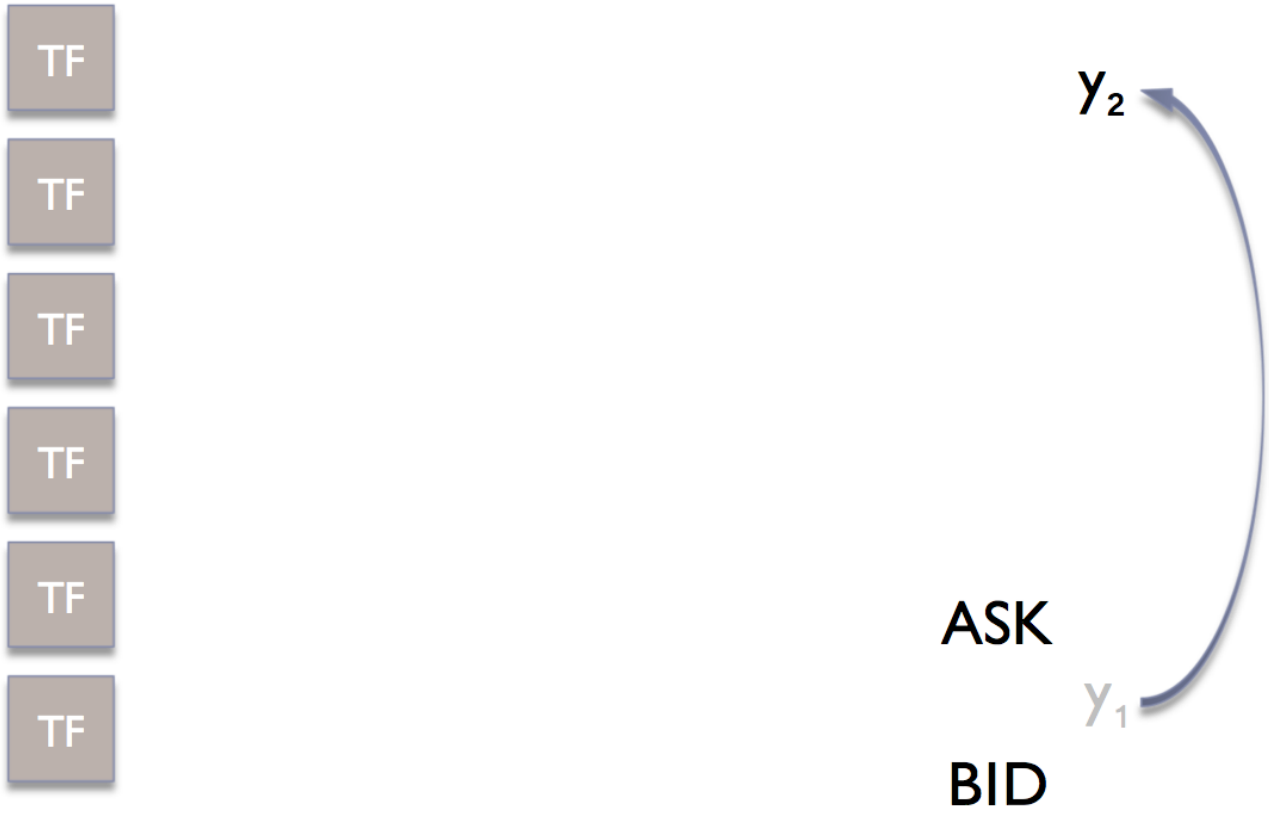
"Sniping"

- Given the model setup – no asymmetric information, no inventory costs, everyone risk neutral – one might conjecture that (Bertrand) competition among trading firms leads to effectively infinite liquidity for investors
  - That is, trading firms should offer to buy or sell x at price y in unlimited quantity at zero bid-ask spread
- But that is not what happens in the continuous limit order book market, due to a phenomenon we call "sniping" (or "latency arbitrage")

“Sniping”

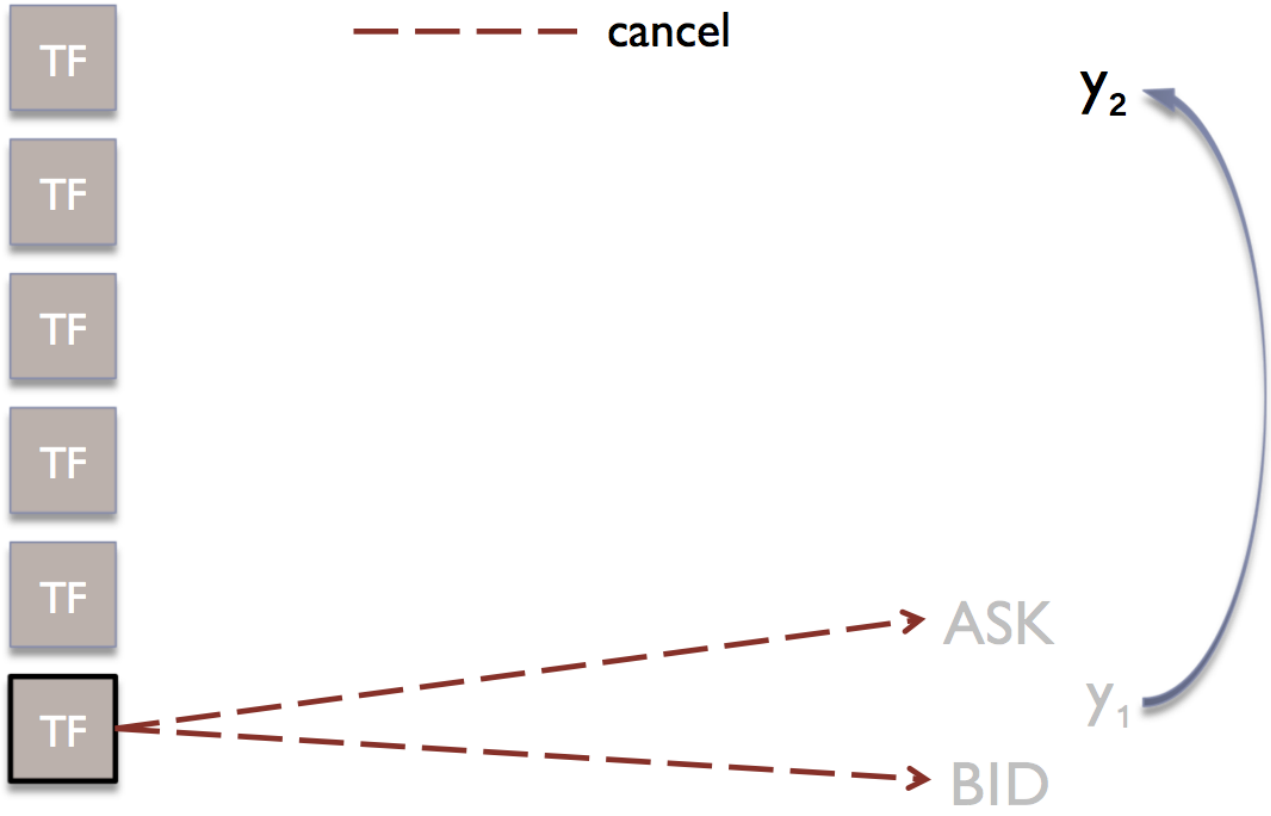
Fundamental value and bid-ask spread

“Sniping”



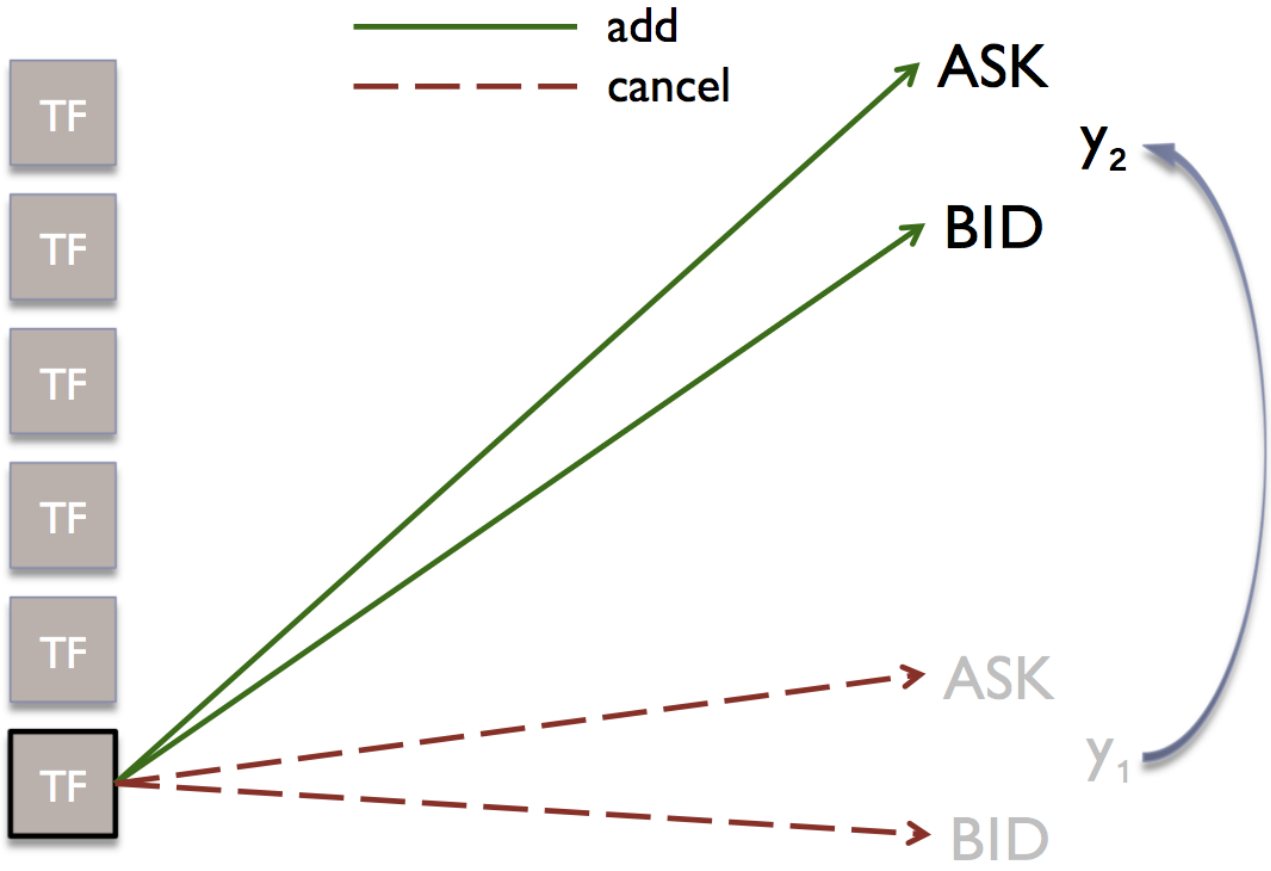
Fundamental value jumps

"Sniping"



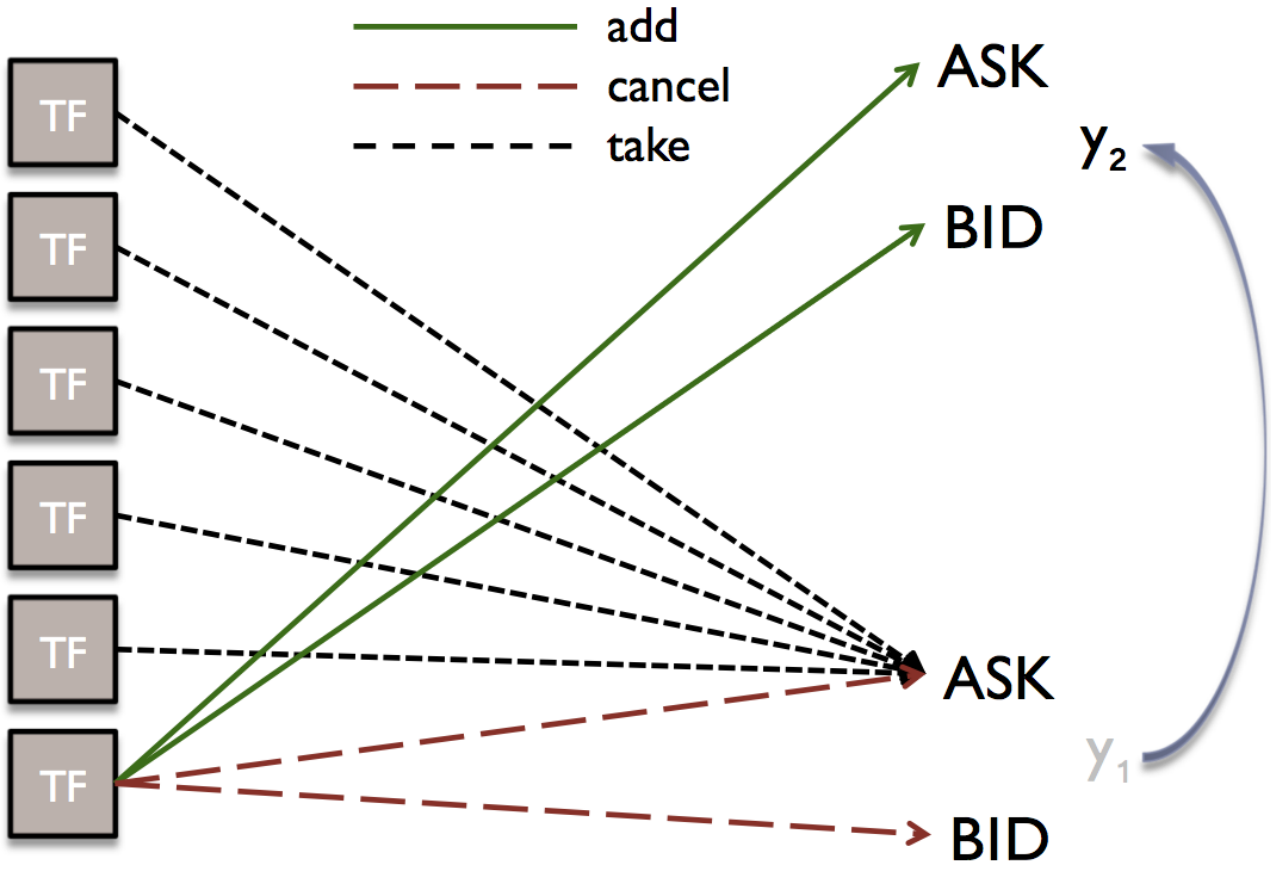
TFs providing liquidity send messages to cancel old quotes and add new quotes

"Sniping"



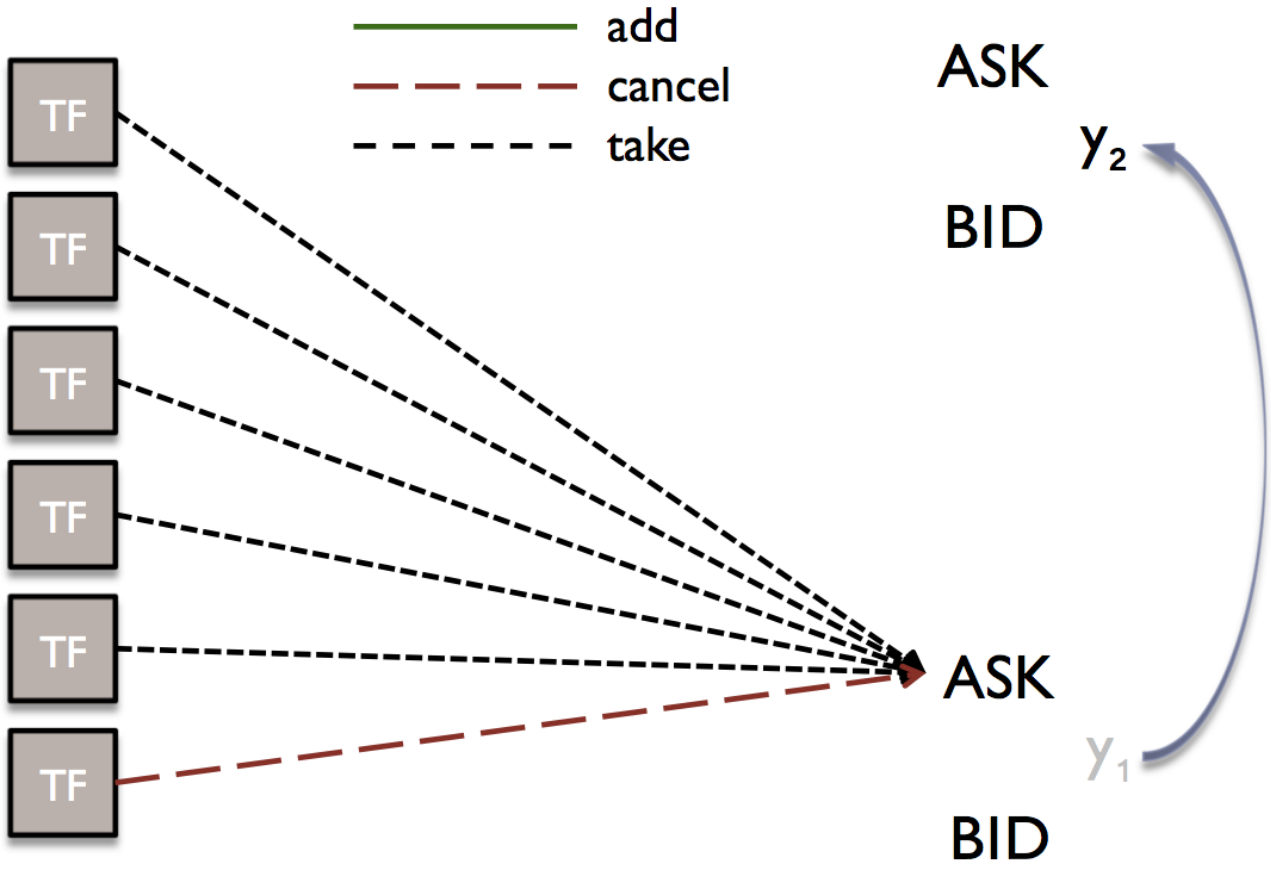
TFs providing liquidity send messages to cancel old quotes and add new quotes

"Sniping"



At same time, *other* TFs send messages to "snipe" the stale quotes

"Sniping"



Because the market design processes messages in *serial*, liquidity providers get sniped with probability $\frac{N-1}{N}$ ... even though the information was public and all TFs have the exact same technology

"Sniping"

- Hence, in a continuous limit order book, *symmetrically observed public information creates arbitrage rents*.
  - Mechanical arbs like ES-SPY are "built in" to the market design
- Not supposed to happen in an efficient market (Fama, 1970)
  - OK to make money from asymmetric information, but symmetric information is supposed to get into prices for free
- In equilibrium, these arbitrage rents are ultimately paid by investors
- 2013 Nobel citation: asset prices are predictable in the long run but "next to impossible to predict in the short run"
  - This is wrong: asset prices are extremely easy to predict in the *extremely* short run

Equilibrium Effects of Sniping

In equilibrium, the bid-ask spread has to be large enough to compensate liquidity providers for the cost of getting sniped.

- Equilibrium condition:

$$\lambda_{invest} \cdot \frac{s^*}{2} = \lambda_{jump} \cdot \Pr(J > \frac{s^*}{2}) \cdot \mathbb{E}(J - \frac{s^*}{2} | J > \frac{s^*}{2}) \quad (1)$$

- Uniquely pins down s. Interpretation:
  - LHS: revenue from investors due to non-zero bid-ask spread
  - RHS: rents to trading firms from mechanical arbitrages
- Endogenous entry yields an additional equation:

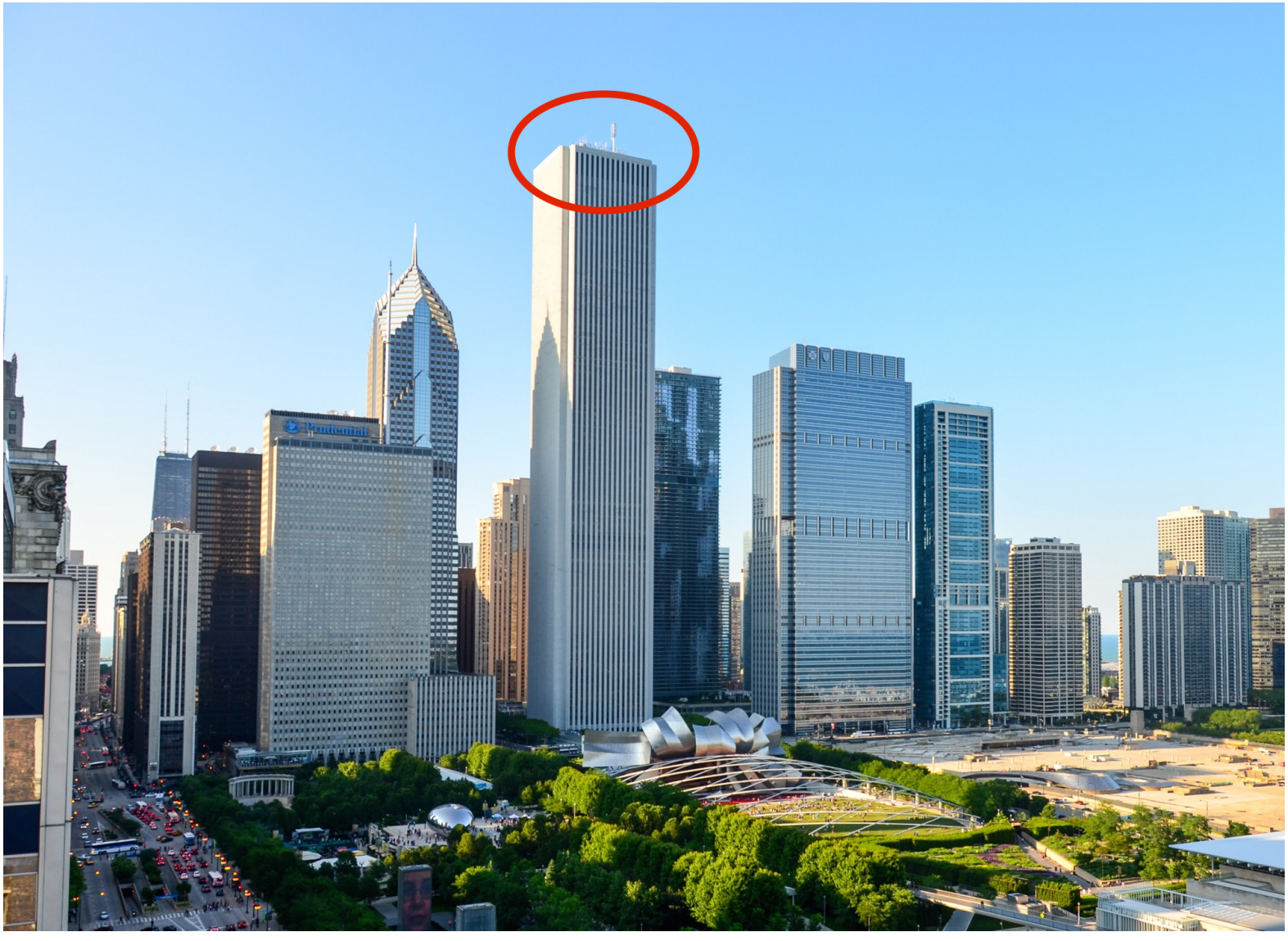
$$\lambda_{invest} \cdot \frac{s^*}{2} = N^* \cdot c_{speed}$$

- Economic interpretation: all of the expenditure by TFs on speed technology ultimately is borne by investors.
  - Arms-race prize = expenditures on speed = cost to investors
  - Remember: arms-race profits have to come from *somewhere*

The HFT Arms Race: Continued

The HFT Arms Race: Continued

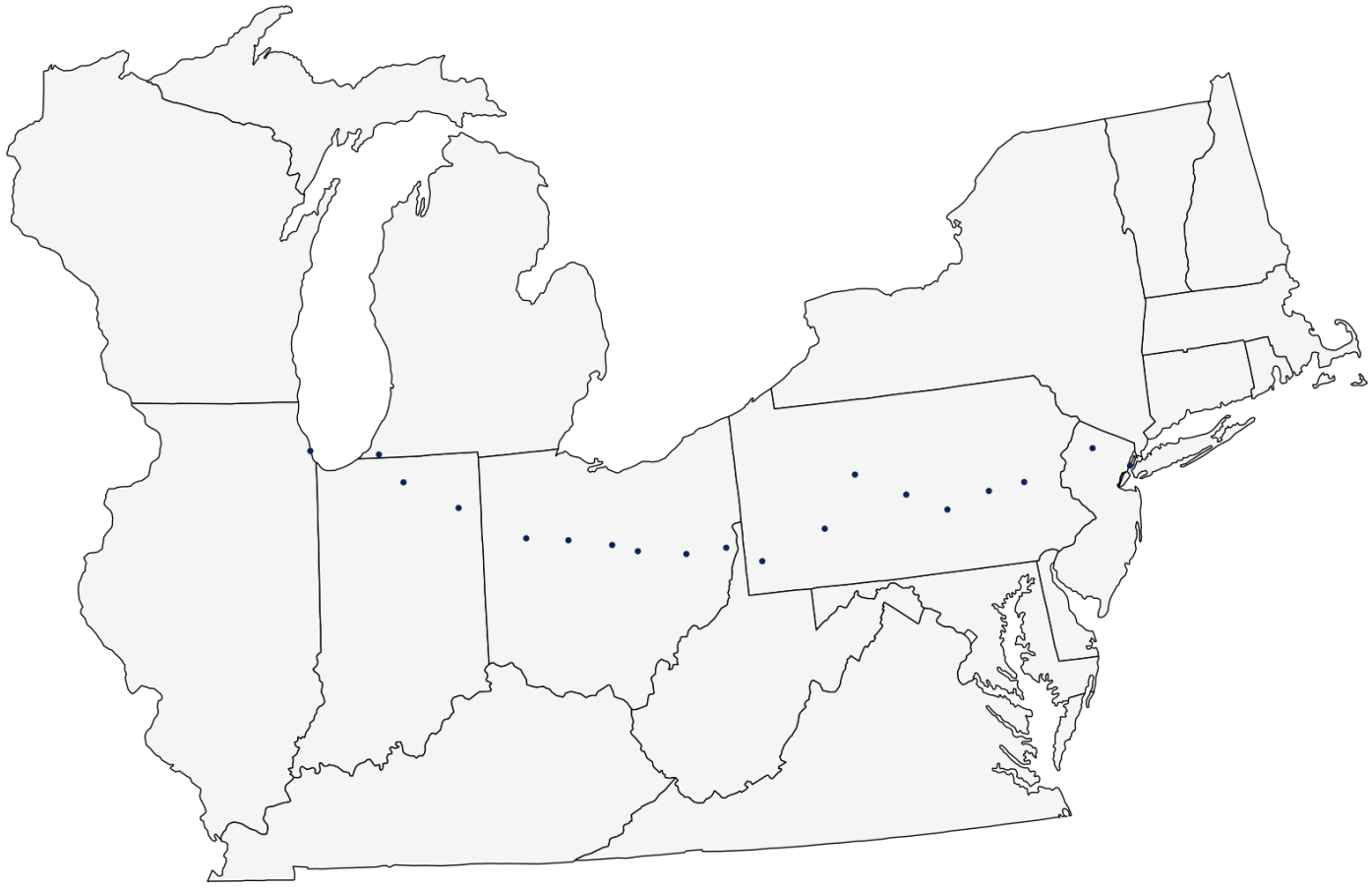
The HFT Arms Race: Continued



The HFT Arms Race: Continued

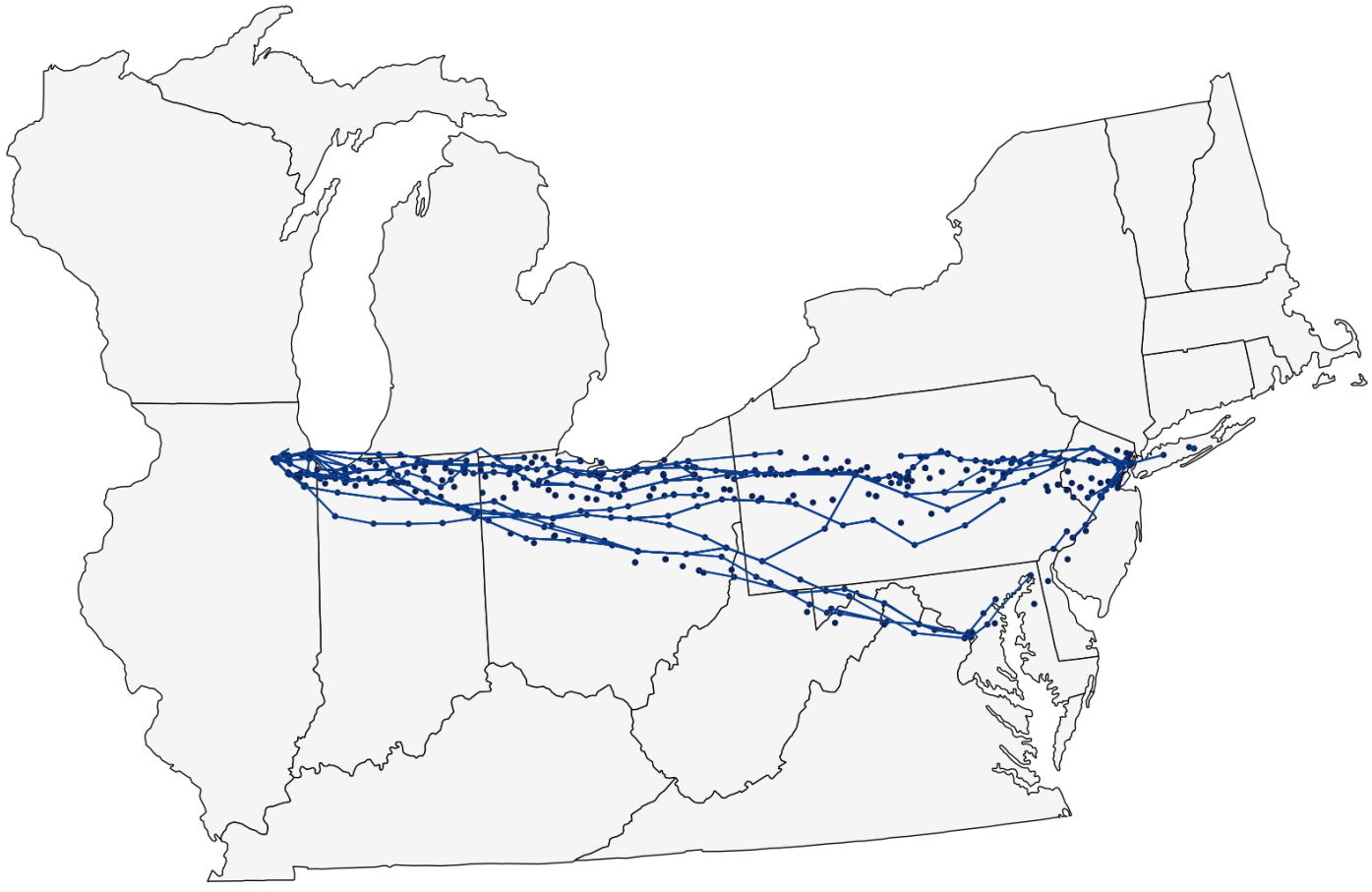
The HFT Arms Race: Continued

The HFT Arms Race: Continued

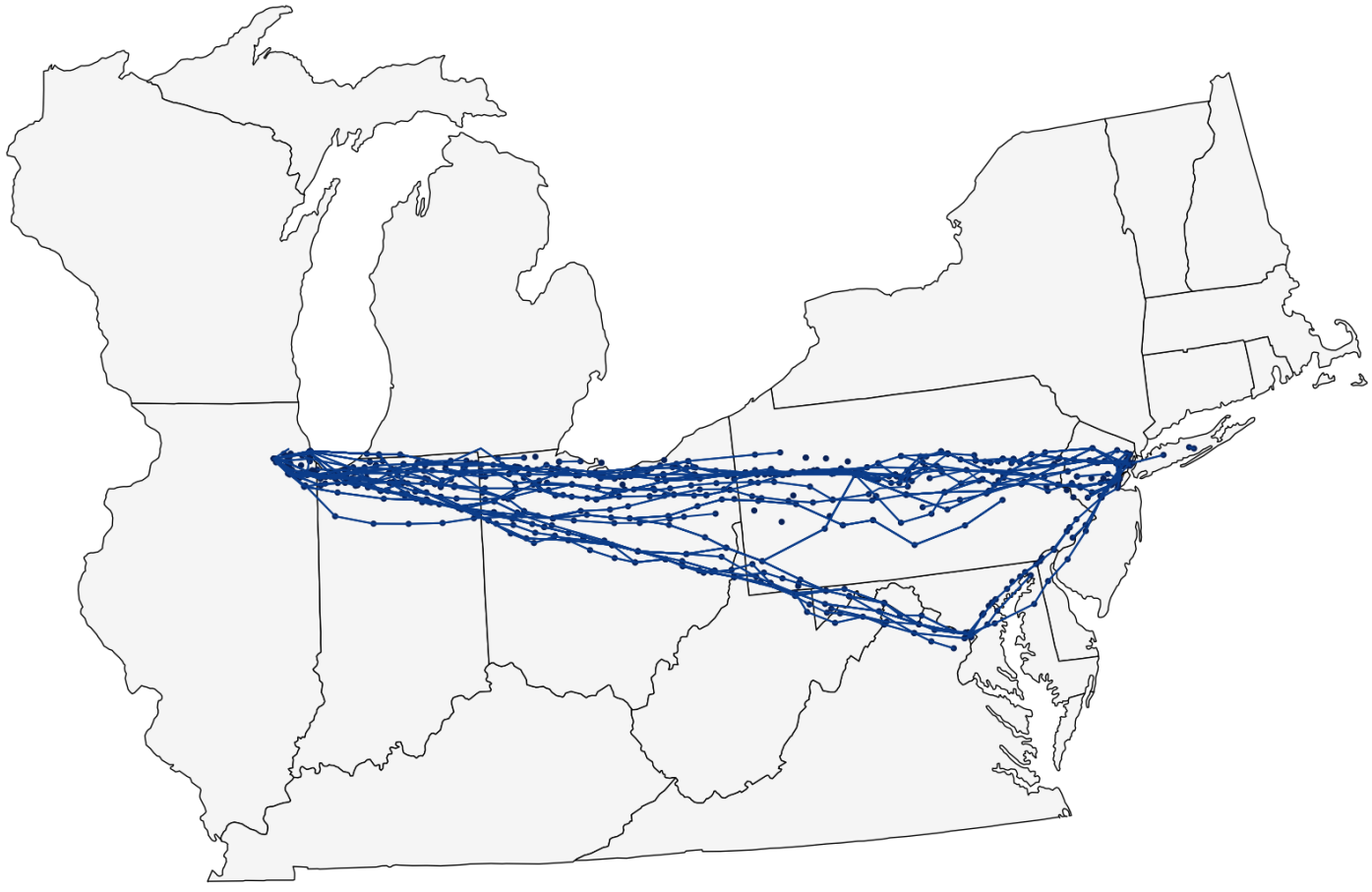


The HFT Arms Race: Continued

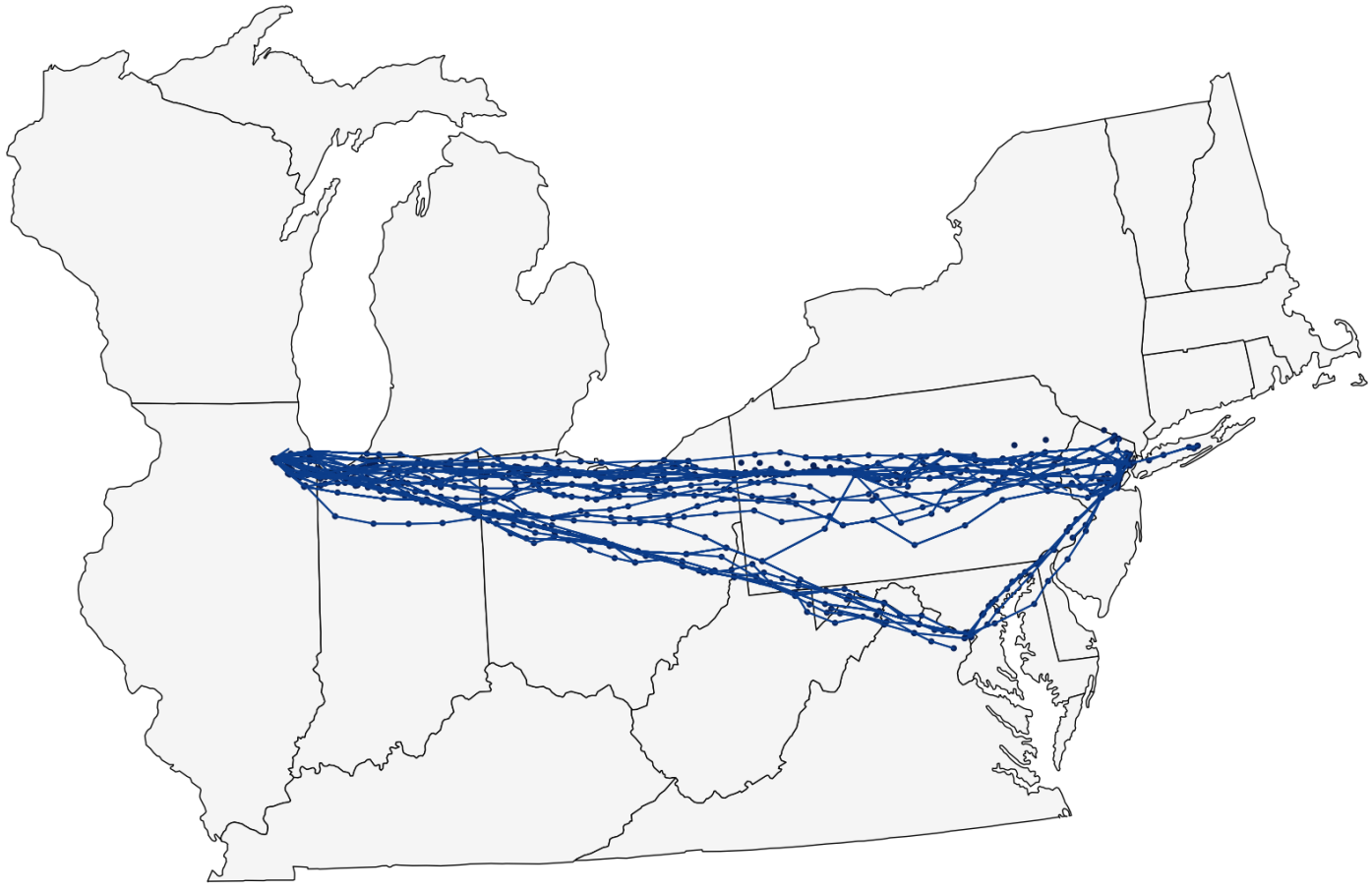
The HFT Arms Race: Continued



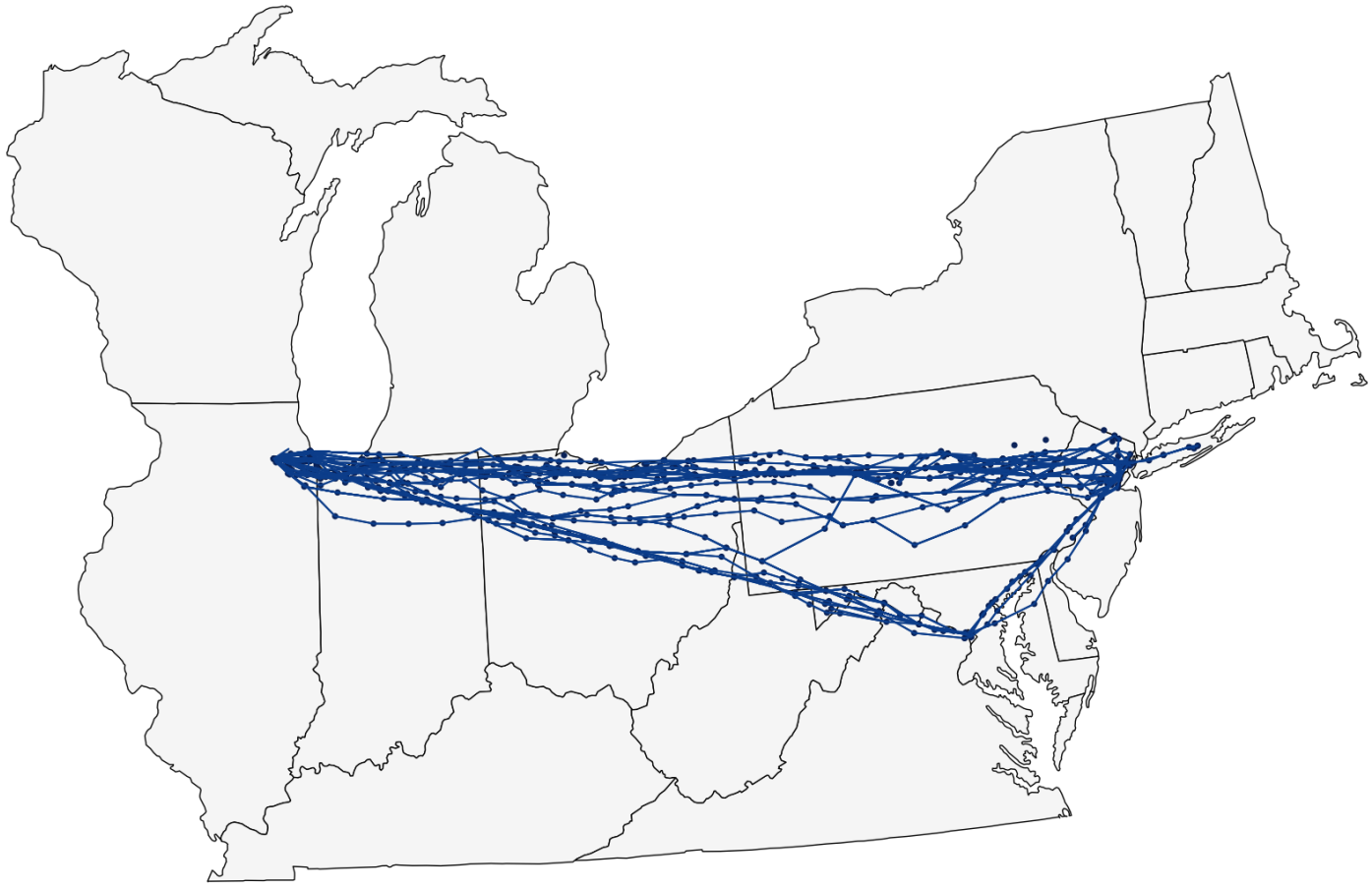
The HFT Arms Race: Continued



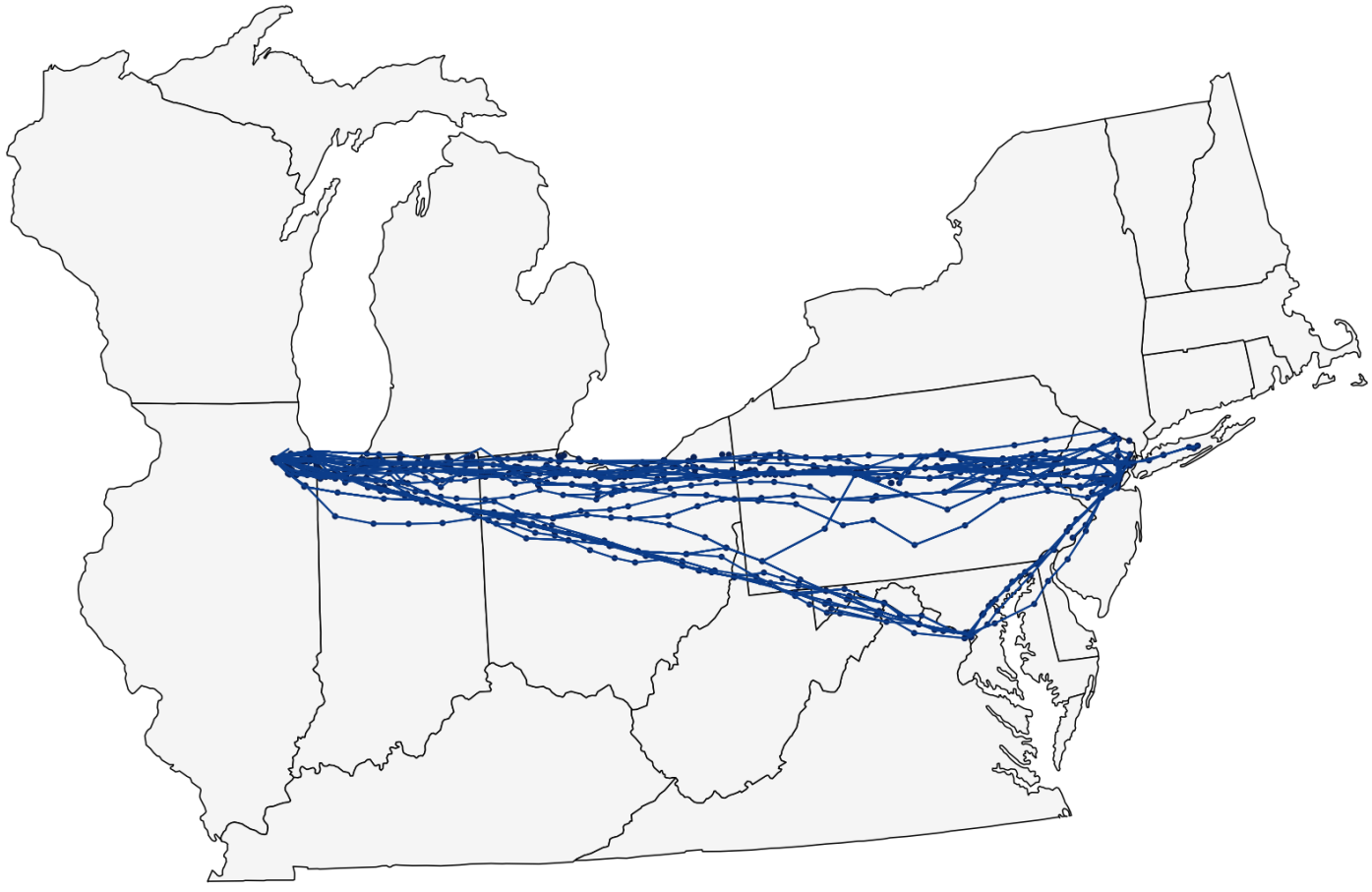
The HFT Arms Race: Continued



The HFT Arms Race: Continued



The HFT Arms Race: Continued



The Case for Frequent Batch Auctions

A simple idea: discrete-time trading.

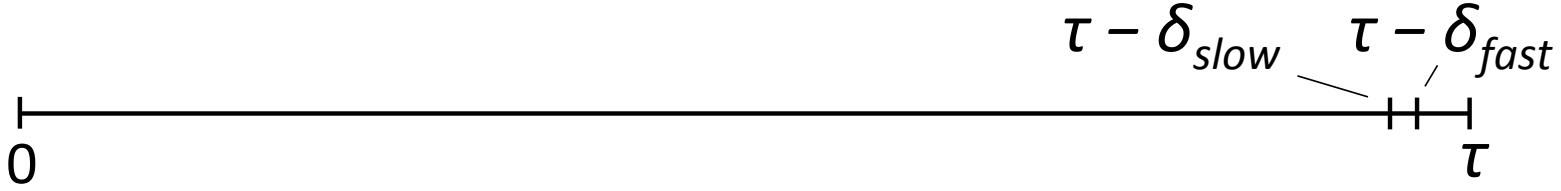
1. Empirical Facts: continuous market violates basic asset pricing principles at HFT time horizons.
   - Market correlations completely break down.
   - Frequent mechanical arbitrage opportunities.
   - Mechanical arbs $\rightarrow$ arms race. Arms race does not compete away the arbs, looks like a "constant".
2. Theory: root flaw is continuous-time serial-process trading
   - Mechanical arbs are "built in" to market design. Sniping.
   - Harms liquidity.
   - Induces never-ending, wasteful, arms race for speed.
3. **Solution: frequent batch auctions**
   - **Competition on speed $\rightarrow$ competition on price.**
   - **Enhances liquidity and stops the arms race.**
   - **Simplifies the market computationally.**

Frequent Batch Auctions: Overview

- High level: analogous to the current market design but for two key differences
  - Time is treated as discrete, not continuous
  - Orders are processed in batch, using an auction, not serially

- Some design details
  - Orders are just like traditional limit orders: price, quantity, direction. Remain outstanding until executed or canceled.
  - Auction is uniform price
  - Priority is price-time, but treating time as discrete
  - Information policy: same information as the continuous market, but disseminated in discrete time.

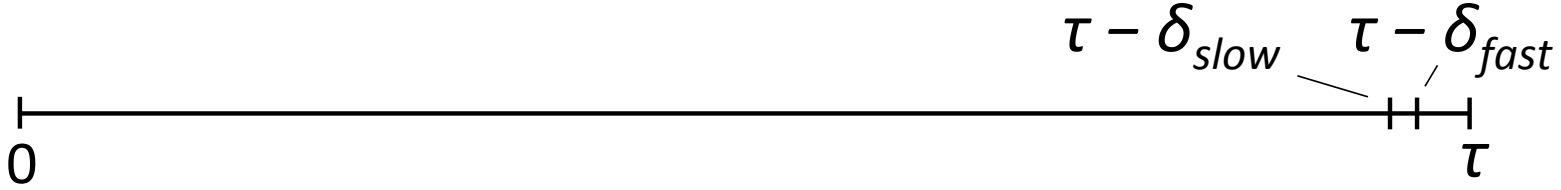
Why FBA Solves the Problem



Reason 1: Discrete time reduces the economic relevance of tiny speed advantages

- Most public information arrives at a time such that all market participants see it equally.
  - $0 \rightarrow \tau - \delta_{slow}$ everybody sees it
  - $\tau - \delta_{fast} \rightarrow \tau$ nobody sees it
  - $\tau - \delta_{slow} \rightarrow \tau - \delta_{fast}$ speed advantage relevent. Proportion $\frac{\delta}{\tau}$
- If the public information is information from past prices... proportion zero.
- Whereas: in the continuous market, the speed advantage is relevant for *ALL* public information.

Why FBA Solves the Problem



Reason 2: Auction changes the nature of competition. From competition on speed to competition on price

- Suppose:
  - Public information arrives in the critical window
  - There are some slow traders with stale quotes in the book
  - There are some fast traders who see the new information
- Continuous market: competition on speed, to snipe the stale quotes
- Batch auction market: competition on price!

Computational Benefits of Discrete Time

- Conceptual point
  - Continuous-time markets implicitly assume that computers and communications technology are infinitely fast.
  - Discrete time respects the limits of computers and communications.
- Examples
  - Regulatory paper trail has to be adjusted for relativity in continuous time.
  - Clock synchronization is a serious issue in continuous time.
  - Exchange matching engines occasionally become backlogged in continuous time (e.g., 5/6/2010 equities flash crash, 10/15/2014 treasuries flash rally).
  - Algos have to trade off error-checking for speed in continuous time (Donald MacKenzie, 2014).
- Advertistement: this is a good topic for research, at intersection of Econ + CS.

Quantifying the High-Frequency Trading"Arms Race"

Matteo Aquilina
Eric Budish
Peter O'Neill

QJE, February 2022

Measuring Latency Arbitrage

- This paper uses a simple new kind of data to measure latency arbitrage in a way that hasn't previously been possible: "Message data"

Measuring Latency Arbitrage

- This paper uses a simple new kind of data to measure latency arbitrage in a way that hasn't previously been possible: "Message data"
- Limit-order book data provide the complete "play-by-play" of the order book:
  - Every new limit order that posts to the book, every canceled order, every trade, etc.
  - Often with ultra-precise timestamps (or even firm IDs)

Measuring Latency Arbitrage

- This paper uses a simple new kind of data to measure latency arbitrage in a way that hasn't previously been possible: "Message data"
- Limit-order book data provide the complete "play-by-play" of the order book:
  - Every new limit order that posts to the book, every canceled order, every trade, etc.
  - Often with ultra-precise timestamps (or even firm IDs)
- But ... limit-order book data are missing the messages that *do not affect the state of the order book, because they fail*.
  - *Attempts* to snipe a stale quote that are too late
  - *Attempts* to cancel a stale quote that are too late

Measuring Latency Arbitrage

- This paper uses a simple new kind of data to measure latency arbitrage in a way that hasn't previously been possible: "Message data"
- Limit-order book data provide the complete "play-by-play" of the order book:
  - Every new limit order that posts to the book, every canceled order, every trade, etc.
  - Often with ultra-precise timestamps (or even firm IDs)
- But ... limit-order book data are missing the messages that *do not affect the state of the order book, because they fail*.
  - *Attempts* to snipe a stale quote that are too late
  - *Attempts* to cancel a stale quote that are too late
- Simple insight: these failure messages are a direct empirical signature of speed-sensitive trading
  - The essence of a race is that there are winners and losers ...
  - But limit order book data don't let you see the losers! Message data do!

Message Data, Simple Methodology

- We obtained message data from the London Stock Exchange (by a request under Section 165 of the Financial Service and Markets Act)

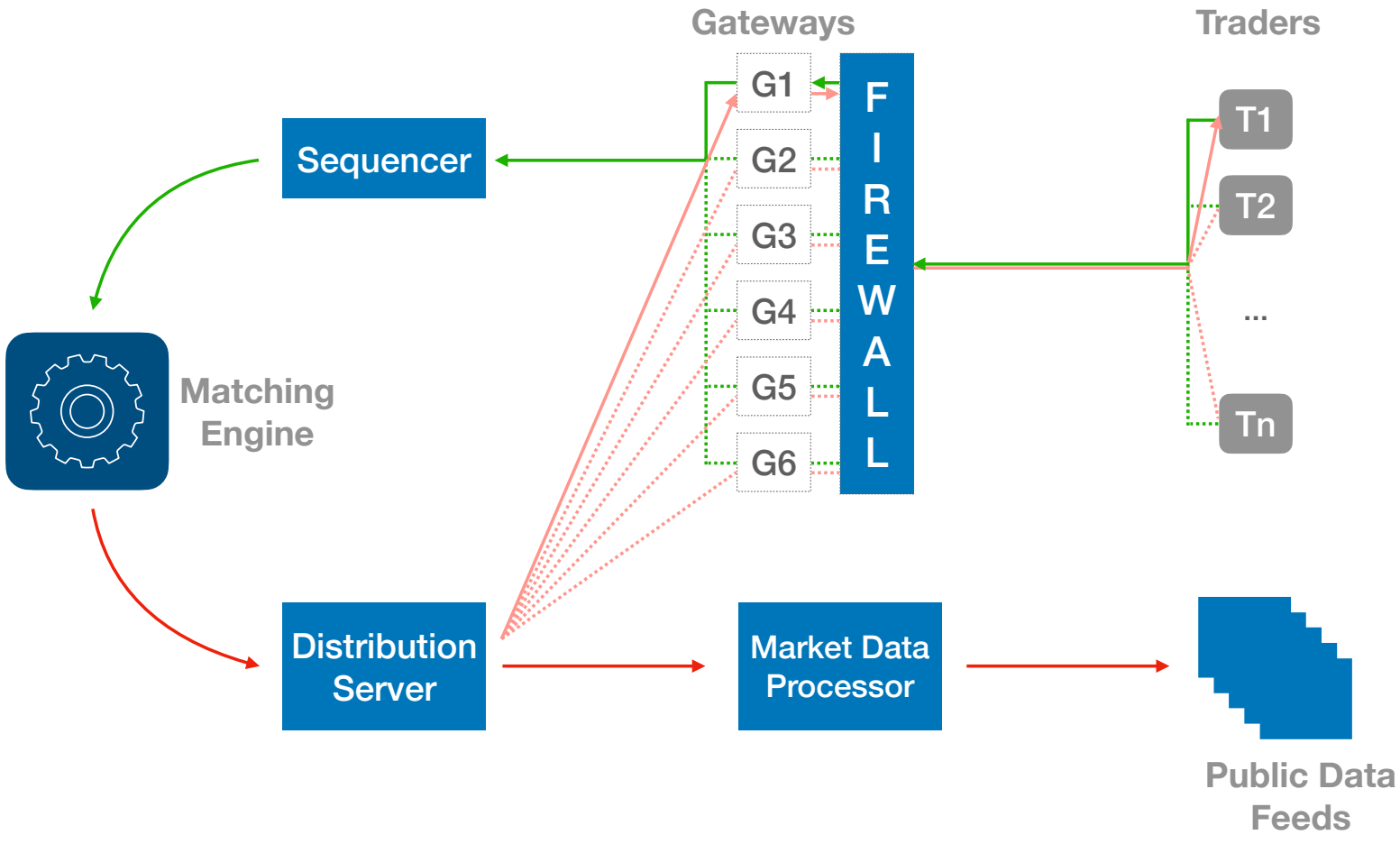
Message Data, Simple Methodology

- We obtained message data from the London Stock Exchange (by a request under Section 165 of the Financial Service and Markets Act)
  - All message data for all FTSE 350 stocks for a 9 week period in Fall 2015
  - Timestamps accurate to the microsecond (0.000001s)
  - Timestamps at the right location in the exchange architecture
  - Anonymized participant IDs

Message Data, Simple Methodology

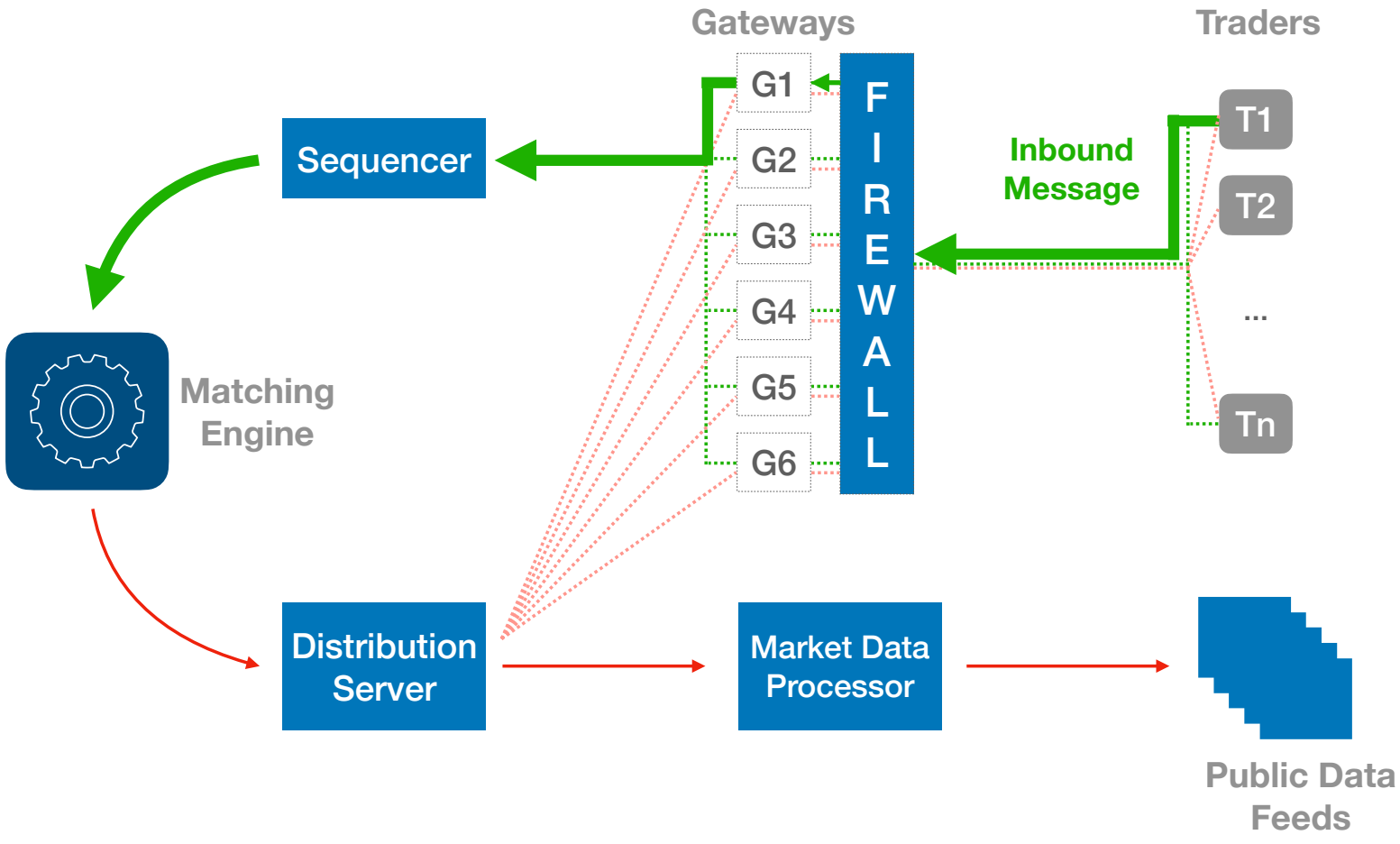
- We obtained message data from the London Stock Exchange (by a request under Section 165 of the Financial Service and Markets Act)
  - All message data for all FTSE 350 stocks for a 9 week period in Fall 2015
  - Timestamps accurate to the microsecond (0.000001s)
  - Timestamps at the right location in the exchange architecture
  - Anonymized participant IDs
- Using this data we can directly measure:
  - Quantity of races
  - How long they take
  - How many participants there are
  - The diversity / concentration of winners and losers
  - The economic stakes – per-race and overall

Exchange Schematic



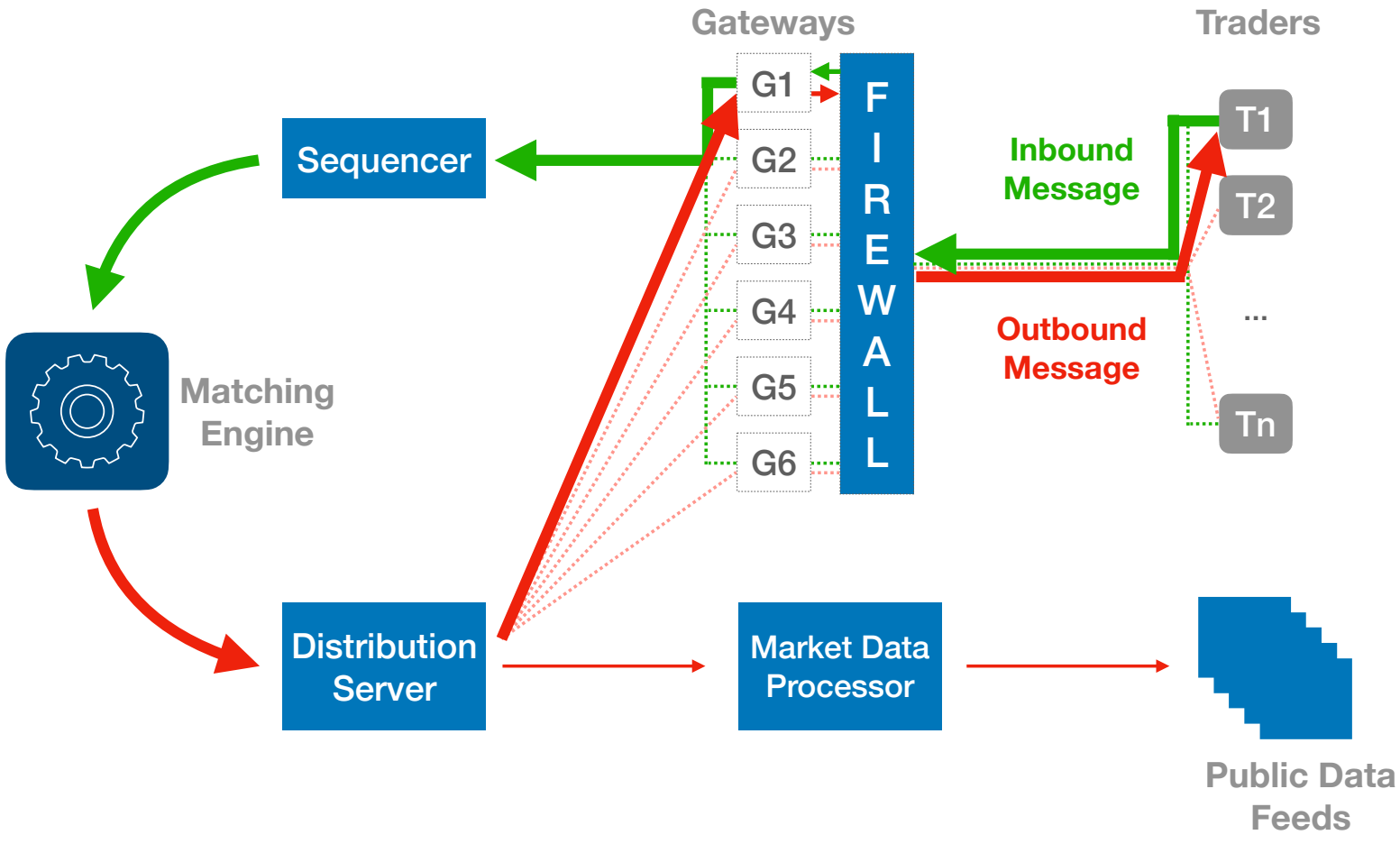
Notes: Please see the text of Section 2.1 for supporting details for this figure.

Exchange Schematic



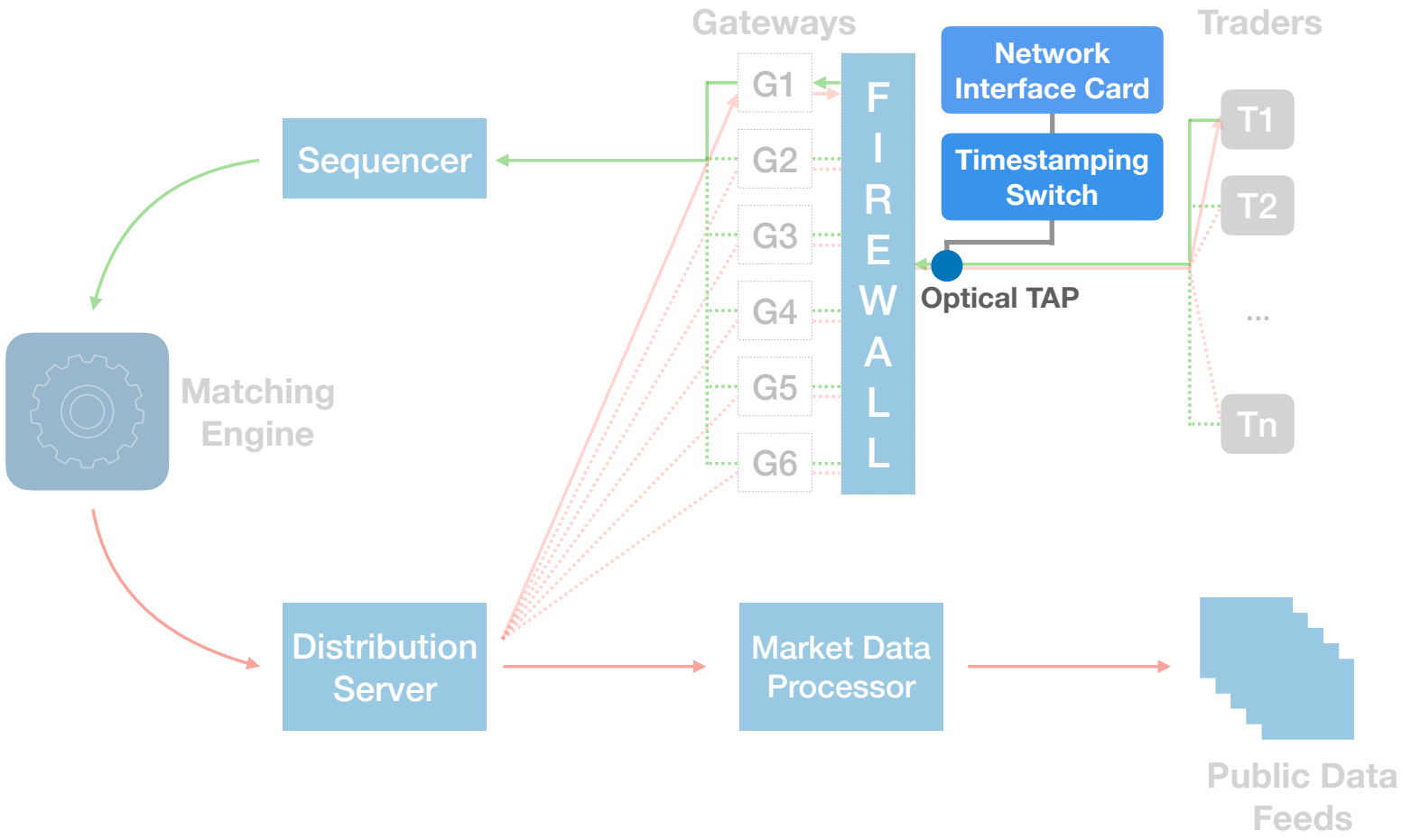
Notes: Please see the text of Section 2.1 for supporting details for this figure.

Exchange Schematic



Notes: Please see the text of Section 2.1 for supporting details for this figure.

Where the Message Data are Captured and Timestamped



Notes: Please see the text of Section 2.2 for supporting details for this figure.

Summary of Main Results

1. Races are frequent: one per minute per symbol for FTSE 100

Summary of Main Results

1. Races are frequent: one per minute per symbol for FTSE 100
2. Races are fast: mode is 5-10 microseconds

Summary of Main Results

1. Races are frequent: one per minute per symbol for FTSE 100
2. Races are fast: mode is 5-10 microseconds
3. Large volume in races: 22% of FTSE 100 volume

Summary of Main Results

1. Races are frequent: one per minute per symbol for FTSE 100
2. Races are fast: mode is 5-10 microseconds
3. Large volume in races: 22% of FTSE 100 volume
4. Race participation is concentrated: Top 6 win 82%, lose 87%. The top firms disproportionately snipe: Top 6 take 80%, provide 42%.

Summary of Main Results

1. Races are frequent: one per minute per symbol for FTSE 100
2. Races are fast: mode is 5-10 microseconds
3. Large volume in races: 22% of FTSE 100 volume
4. Race participation is concentrated: Top 6 win 82%, lose 87%. The top firms disproportionately snipe: Top 6 take 80%, provide 42%.
5. Races are small per race: average half a tick, 2GBP

Summary of Main Results

1. Races are frequent: one per minute per symbol for FTSE 100
2. Races are fast: mode is 5-10 microseconds
3. Large volume in races: 22% of FTSE 100 volume
4. Race participation is concentrated: Top 6 win 82%, lose 87%. The top firms disproportionately snipe: Top 6 take 80%, provide 42%.
5. Races are small per race: average half a tick, 2GBP
6. Adds up to meaningful proportion of price impact and effective spread: races are 31% of price impact, 33% of effective spread

Summary of Main Results

1. Races are frequent: one per minute per symbol for FTSE 100
2. Races are fast: mode is 5-10 microseconds
3. Large volume in races: 22% of FTSE 100 volume
4. Race participation is concentrated: Top 6 win 82%, lose 87%. The top firms disproportionately snipe: Top 6 take 80%, provide 42%.
5. Races are small per race: average half a tick, 2GBP
6. Adds up to meaningful proportion of price impact and effective spread: races are 31% of price impact, 33% of effective spread
7. Market design reform could meaningfully reduce the cost of liquidity: latency arbitrage tax is 0.42bps of volume. Eliminating latency arbitrage would reduce investors' cost of liquidity by 17%

Summary of Main Results

1. Races are frequent: one per minute per symbol for FTSE 100
2. Races are fast: mode is 5-10 microseconds
3. Large volume in races: 22% of FTSE 100 volume
4. Race participation is concentrated: Top 6 win 82%, lose 87%. The top firms disproportionately snipe: Top 6 take 80%, provide 42%.
5. Races are small per race: average half a tick, 2GBP
6. Adds up to meaningful proportion of price impact and effective spread: races are 31% of price impact, 33% of effective spread
7. Market design reform could meaningfully reduce the cost of liquidity: latency arbitrage tax is 0.42bps of volume. Eliminating latency arbitrage would reduce investors' cost of liquidity by 17%
8. Adds up to meaningful total "size of the prize": 0.42bps is about $5bn annually in global equities alone

Discussion of Magnitudes

- Whether magnitudes in our study seem large or small depends on the vantage point

Discussion of Magnitudes

- Whether magnitudes in our study seem large or small depends on the vantage point
- Cost per transaction: small.
  - Roughly half a tick per race.
  - Roughly 0.5 bps tax on trading.
  - Does not sound alarming.

Discussion of Magnitudes

- Whether magnitudes in our study seem large or small depends on the vantage point
- Cost per transaction: small.
  - Roughly half a tick per race.
  - Roughly 0.5 bps tax on trading.
  - Does not sound alarming.
- Overall sums: large.
  - 17%-33% reduction in cost of liquidity is huge
  - \$5bn per year in equities alone — not even counting futures, currencies, US Treasuries, etc.

Discussion of Magnitudes

- Whether magnitudes in our study seem large or small depends on the vantage point
- Cost per transaction: small.
  - Roughly half a tick per race.
  - Roughly 0.5 bps tax on trading.
  - Does not sound alarming.
- Overall sums: large.
  - 17%-33% reduction in cost of liquidity is huge
  - \$5bn per year in equities alone — not even counting futures, currencies, US Treasuries, etc.
- This creates a "Concentrated-Dispersed" problem in the sense of Mancur Olson, "The Logic of Collective Action"
  - Small enough that ordinary investors need not worry.
  - But: billions of dollars per year for a small number of parties in the speed race ...
  - ... who then have significant incentive to preserve status quo.

A Theory of Stock Exchange Competition and Innovation: Will the Market Fix the Market?

Eric Budish
Robin Lee
John Shim

R&R JPE May 2023

Incentives for Market Design Innovation

- Market design research usually focuses on designing the best possible market mechanism for a given problem
- This paper concerns a different, complementary question: suppose researchers have already designed an attractive mechanism — will it actually get adopted?
- What are the private incentives for stock exchanges to adopt frequent batch auctions?
  - Do exchanges' private innovation incentives align with what is socially efficient?
  - Will the market fix the market?

Will the Market Fix the Market? Summary of Main Results

- We study a model closely tailored to the institutional details of modern electronic financial exchanges
  - Players: exchanges, trading firms, informed traders, and uninformed investors.
  - Exchanges make a market design decision and set prices — for trading per se and for "speed technology"
  - TFs decide whether to buy speed technology, and then all market participants play a trading game
  - Regulatory details: stocks are fungible across exchanges ("Unlisted Trading Privileges") and market participants can frictionlessly search across exchanges ("Regulation National Market System")

Will the Market Fix the Market? Summary of Main Results

- Subgame in which all exchanges use status quo market design ("Continuous")
  - Trading fees are perfectly competitive ($f = 0$).
  - Exchanges capture economic rents from speed technology ($F > 0$).
  - Aligns with empirical facts we document
    - Trading fees are very competitive. \$0.0001 per share per side.
    - Speed technology fees are large and growing. \$1bn+ per year for US stock exchanges.

Will the Market Fix the Market? Summary of Main Results

- Subgames in which an exchange innovates (adopts "Discrete")
  - Result 1: if a single exchange adopts FBA's, it wins share and earns profits in *any* equilibrium. Not chicken-and-egg.
  - Result 2: if multiple exchanges adopt FBA's , then FBA "wins" ... but profits are zero. Trading fees are competitive, no more speed rents. (Regulatory mandate, imitation)
  - Result 3: there exists an equilibrium in which all incumbent exchanges maintain the status quo market design. Intuition: cooperation in the repeated prisoner's dilemma

Will the Market Fix the Market? Policy Implications

- Surprise: if there is an innovator, it would actually work
  - The difficulty is not that the new market design would not get off the ground (as in many other platform environments), but lack of economic incentive
  - Intuition: the same frictionless search that causes trading fees to be brutally competitive in the status quo, also helps the innovator get off the ground ... and also makes the innovator very vulnerable to imitation and with that perfect competition.
- Implication: a regulatory "push" might be enough
  - A "mandate" would certainly work
  - But a "push" that tips the balance of incentives, enough to get an initial adopter, might also be enough

Recent Policy Progress

SEC Proposes Rule to Enhance Competition for Individual Investor Order Execution

FOR IMMEDIATE RELEASE
2022-225

Washington D.C., Dec. 14, 2022 — The Securities and Exchange Commission today proposed a rule that would require certain orders of individual investors to be exposed to competition in fair and open auctions before such orders could be executed internally by any trading center that restricts order-by-order competition.

"Today's markets are not as fair and competitive as possible for individual investors — everyday retail investors. This is in part because there isn't a level playing field among different parts of the market: wholesalers, dark pools, and lit exchanges," said SEC Chair Gary Gensler. "Further, the markets have become increasingly hidden from view, especially for individual investors. These everyday individual investors don't have the full benefit of various market participants competing to execute their marketable orders at the best price possible. Thus, today's proposal is designed to bring greater competition in the marketplace for retail market orders. I think it makes sense for the market, and for everyday individual investors, to allow the broader market to compete for their orders."

Individual investors use marketable orders for stocks listed on U.S. securities exchanges (NMS stocks) when they seek to trade immediately at the best available prices in the market. Currently, retail brokers route more than 90 percent of these orders to a small group of off-exchange dealers, known as wholesalers. This routing practice is known as a type of segmentation and

SECURITIES AND EXCHANGE COMMISSION

17 CFR Parts 240 and 242

[Release No. 34-96495; File No. S7-31-22]

RIN 3235-AM57

Order Competition Rule

AGENCY: Securities and Exchange Commission.

ACTION: Proposed rule.

SUMMARY: The Securities and Exchange Commission ("Commission") is proposing to amend the regulation governing the national market system ("NMS") under the Securities Exchange Act of 1934 ("Exchange Act") to add a new rule designed to promote competition as a means to protect the interests of individual investors and to further the objectives of an NMS. The proposed rule would prohibit a restricted competition trading center from internally executing certain orders of individual investors at a price unless the orders are first exposed to competition at that price in a qualified auction operated by an open competition trading center. The proposed rule would also include limited exceptions to this general prohibition. In addition, the Commission is proposing to amend the regulation governing the NMS to add new defined terms included in the proposed rule.

DATES: Comments should be received on or before March 31, 2023.

Conclusion

Summary: the Case for FBAs

- My work looks at HFT from the perspective of market design
- Root problem isn't "evil HFTs", it's continuous-time / serial-process trading.
- Continuous + Serial $\rightarrow$ built-in violation of EMH
- Empirical evidence:
  - Sniping is a (shockingly) large percentage of financial market volume
  - Small per race, but it adds up. \$100bn+ NPV.
- Solution: discrete + batch. "Frequent batch auctions."
  - Eliminates sniping. No more arbitrage rents from symmetric public information.
  - Enhances liquidity
  - Stops the arms race
  - Simplifies the market computationally
- Private-sector innovation could solve the problem, but might not (hasn't yet). Might take a regulatory push.

Directions for Future Research, I

- Direction I: just keep doing the work! Of inventing and analyzing new, useful market designs.
- "Flow Trading", with Peter Cramton, Pete Kyle, Mina Lee, David Malec
  - Takes FBA idea to the trade of arbitrary portfolios of assets (arbitrary linear combinations with real-valued positive/negative weights)
  - Portfolios are at the heart of finance ... trade them directly!
  - Builds solution to "correlation breakdown" directly into market design: can trade a "Buy X, Sell Y" portfolio, preventing prices of X and Y from diverging in the first place.
  - Requires marriage of FBA design to the idea of "smooth trading" over time in Kyle and Lee (2017)
  - (Origin story: NBER Market Design, Fall 2017)
- Great topic for research: what is the optimal batch interval? What are the costs and benefits of a significantly longer batch interval? Important work on this topic by Du and Zhu (2017).
- Great topic for research: rents versus efficiency in finance (e.g., Philippon, 2015)

Directions for Future Research, II

- Direction II: measurement.
- Advertisement: the code for our message data study is publicly available (github.com/ericbudish/HFT-Races). Please use it!
  - U.S. equities market would be really interesting to study (fragmentation, complexity)
  - More asset classes: ETFs, futures, currencies, treasuries, options, corporate bonds
- Science of science point:
  - The "Quantifying the HFT Arms Race" paper was a "big data" style research project with a theory heart
  - Theory + Data as <u>complements</u>

Directions for Future Research, III

- Direction III: Political economy of market design adoption
  - Issue highlighted in the "Will the Market Fix the Market?" paper
  - Good ideas don't always implement themselves

Directions for Future Research, III

- Direction III: Political economy of market design adoption
  - Issue highlighted in the "Will the Market Fix the Market?" paper
  - Good ideas don't always implement themselves
- This is something I've been thinking about a lot in the past decade. Not just in the context of my research on HFT and the design of financial markets, but more generally:
  - What should we do as a profession when we have ideas where the social value is large, but private forces are opposed and the case for gov't addressing the problem isn't great either?
  - Ex: revenue-neutral carbon tax, congestion, many others

Friedman on Theory $\rightarrow$ Practice

There is enormous inertia—a tyranny of the status quo—in private and especially governmental arrangements. Only a crisis—actual or perceived—produces real change. When that crisis occurs, the actions that are taken depend on the ideas that are lying around. That, I believe, is our basic function [as economists]: to develop alternatives to existing policies, to keep them alive and available until the politically impossible becomes politically inevitable.

– *Milton Friedman, Capitalism and Freedom*

Roth and Zingales on Theory $\rightarrow$ Practice

- Al Roth: "We need to foster a still unfamiliar kind of design literature in economics ... if we nurture it to maturity, its relationship with current economics will be something like the relationship of engineering and physics, or of medicine and biology ("The Economist as Engineer", 2002)

- Luigi Zingales: "We should get more involved in policy (while not in politics). Policy work enjoys a lower status in our circles ... If profitable trading strategies are considered publishable research ... (AFA Presidential address, 2015) "

- The changes Roth and Zingales suggest seem especially important for ideas where
  - social value is large
  - concentrated interests are opposed
- When social and private align: natural economic forces help build the bridges
  - Index funds
  - Derivatives
  - Modern portfolio management
- When social and private diverge ...
- In the end I'm an optimist – wager that we'll see discrete-time trading *eventually* (and carbon taxes, etc.)
- But I wonder what we can do to speed up

Zooming Out: Private vs. Social Innovation Incentives

Private Incentives \ Social Incentives	+	-
+	Standard Case	Rent Seeking Business Stealing
-	Concentrated-Dispersed	-

Private vs. Social Incentives: Finance Innovations

Private Incentives \ Social Incentives	+	-
+	Option Pricing Index Funds Lifecycle Investing Portfolio Theory	Rent Seeking in Finance
-	Frequent Batch Auctions	-