

High-Frequency Trading and the Design of Financial Markets: The Case for Frequent Batch Auctions

Eric Budish
University of Chicago, Booth School of Business

a16z Seminar
July 24th, 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

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

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”

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”

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!

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 → 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 → 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 → 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 → 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, JPE, forthcoming)
- ▶ Part IV: Recent Research on “Flow Trading” (Budish, Cramton, Kyle, Lee, Malec)
- ▶ Conclusion: Directions for Future Research, Connections to Crypto Markets

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

Eric Budish
Peter Cramton
John Shim

QJE, November 2015

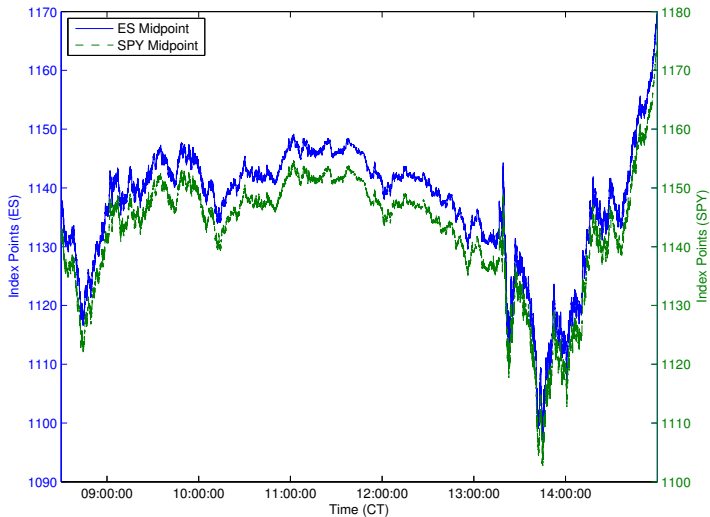
The Case for Frequent Batch Auctions

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 → 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 → competition on price.
 - ▶ Enhances liquidity and stops the arms race.
 - ▶ Simplifies the market computationally.

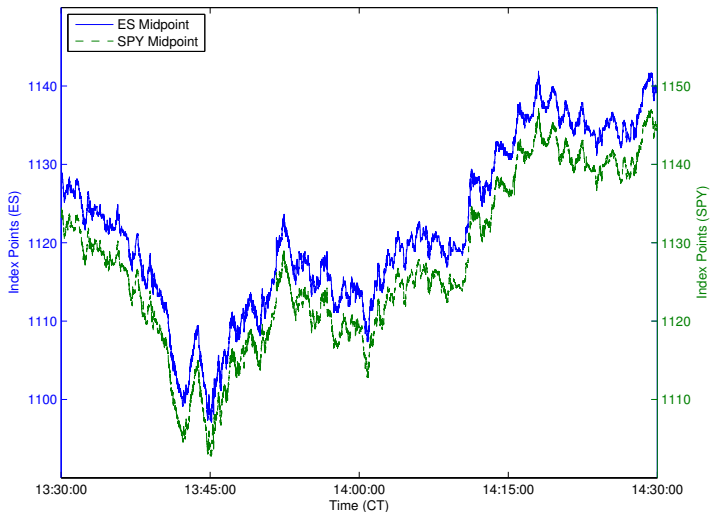
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 Day



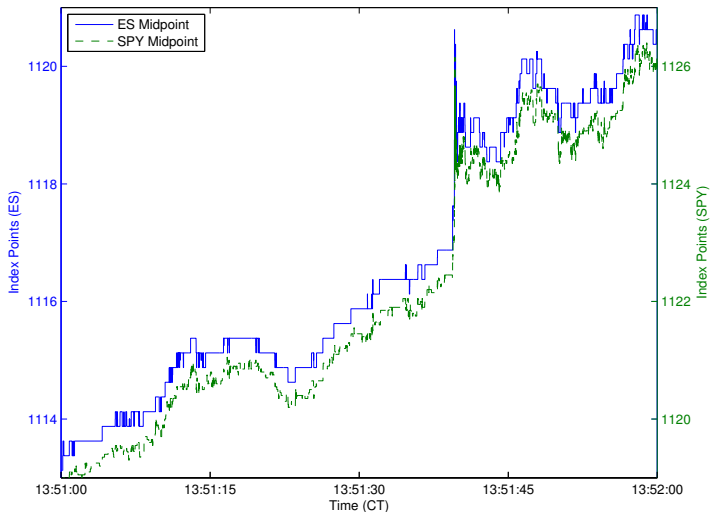
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 hour



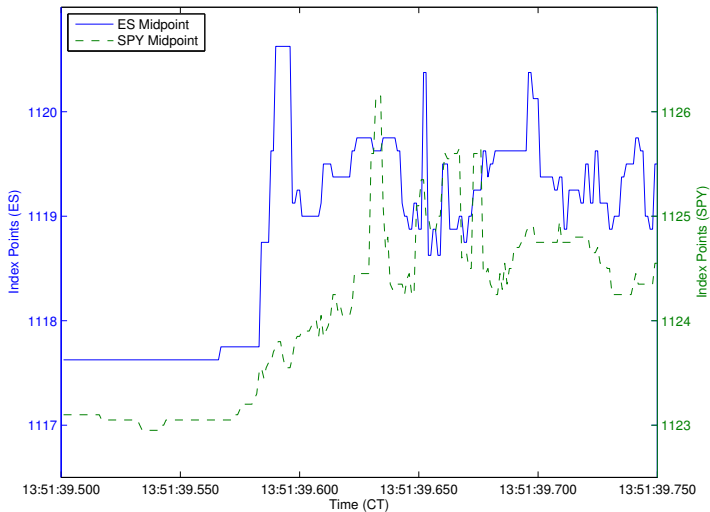
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 minute



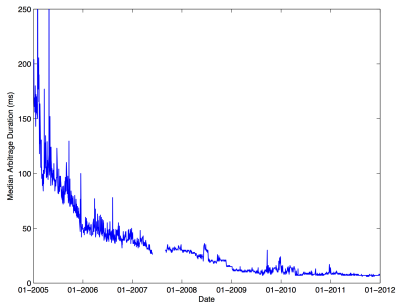
Market Correlations Break Down at High Frequency

ES vs. SPY: 250 milliseconds

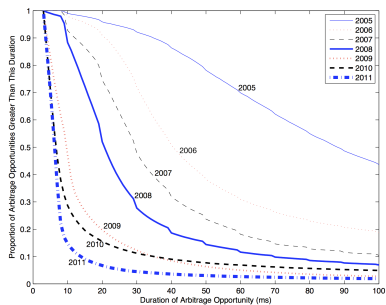


Arb Durations over Time: 2005-2011

Median over time

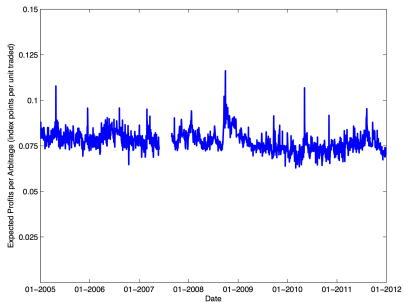


Distribution by year

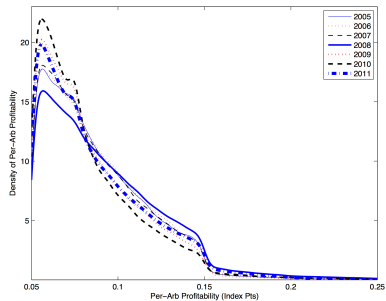


Arb Per-Unit Profits over Time: 2005-2011

Median over time

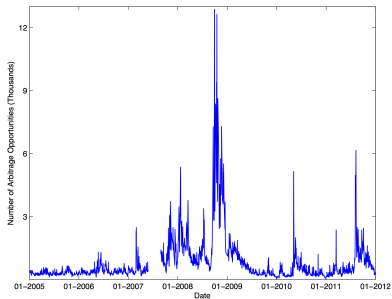


Distribution by year

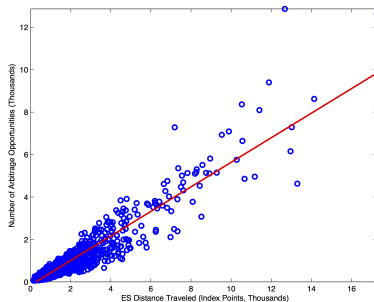


Arb Frequency over Time: 2005-2011

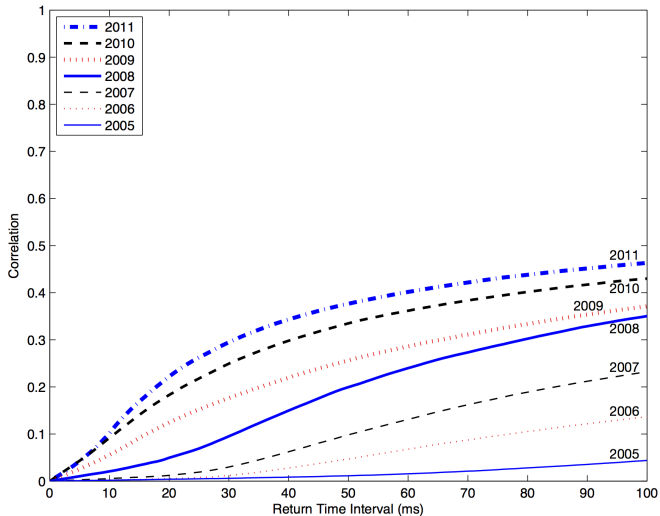
Frequency over time



Frequency vs. Volatility



Correlation Breakdown Over Time: 2005-2011



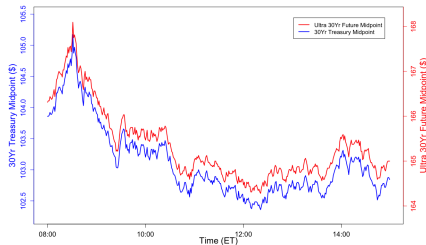
Races, Races, Races

- ▶ 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

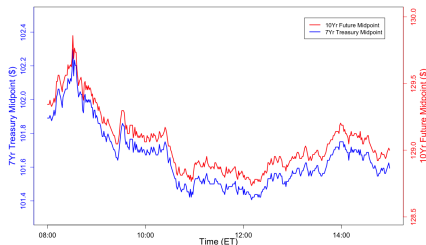
Highly Correlated Pairs

US Treasuries

30 Year Ultra Future vs. 30 Year Cash



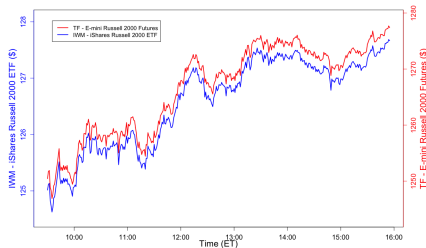
10 Year Future vs. 7 Year Cash



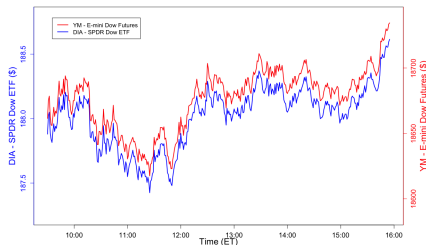
Highly Correlated Pairs

Equity Index

Russell 2000 Future vs. ETF



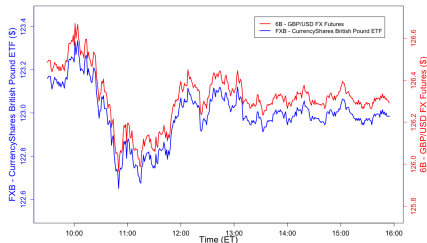
DOW Future vs. ETF



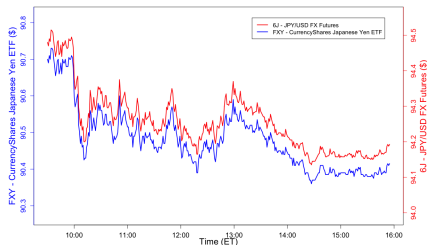
Highly Correlated Pairs

Foreign Exchange

GBP/USD Future vs. ETF



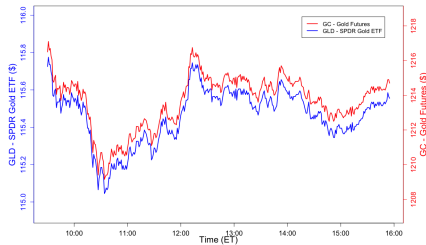
JPY/USD Future vs. ETF



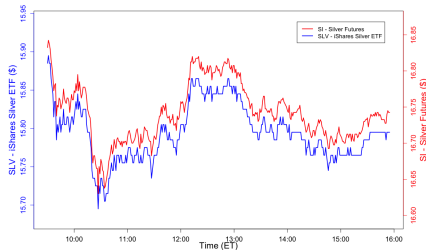
Highly Correlated Pairs

Commodities

Gold Future vs. ETF



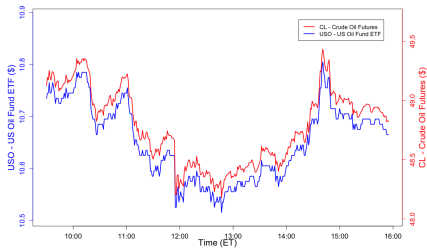
Silver Future vs. ETF



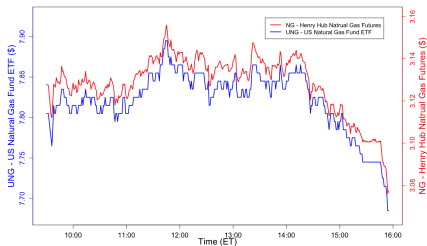
Highly Correlated Pairs

Commodities

Crude Oil Future vs. ETF



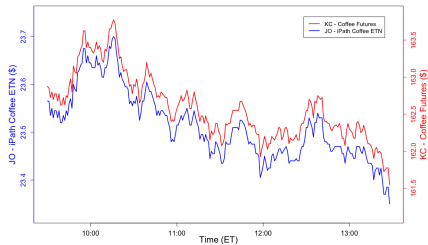
Natural Gas Future vs. ETF



Highly Correlated Pairs

Commodities

Coffee Future vs. ETF



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 (QQ)
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 (QQ) vs. E-micro Gold Futures (MGC)
miNY Gold Futures (QQ) vs. Spot Gold (XAUUSD)
miNY Gold Futures (QQ) vs. SPDR Gold Trust (GLD)
miNY Gold Futures (QQ) 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 (XPOUSD)
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 → 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 → 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”



ASK
 y_1
BID

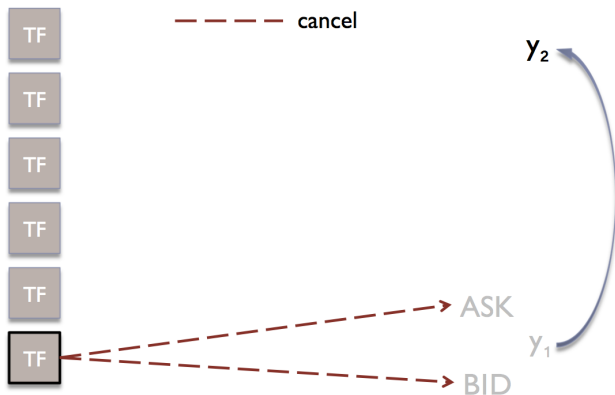
Fundamental value and bid-ask spread

“Sniping”



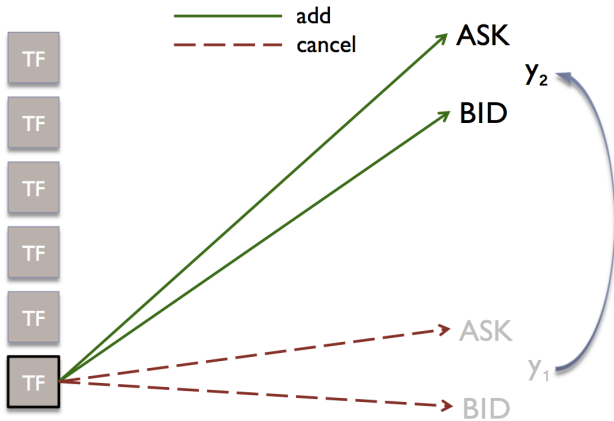
Fundamental value jumps

“Sniping”



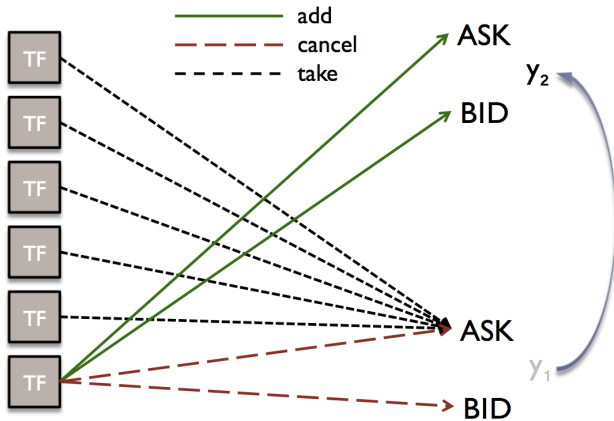
TFs providing liquidity send messages to cancel old quotes and add new quotes

“Sniping”



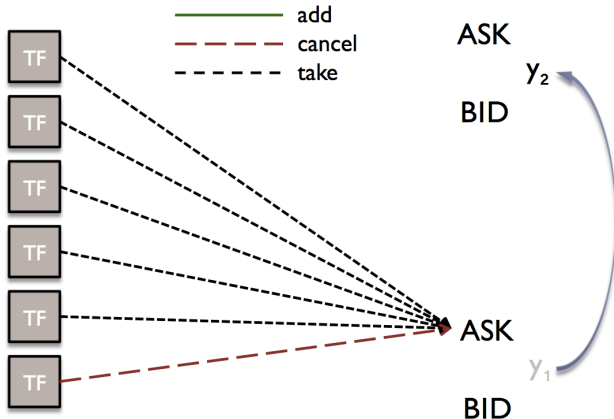
TFs providing liquidity send messages to cancel old quotes and add new quotes

“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

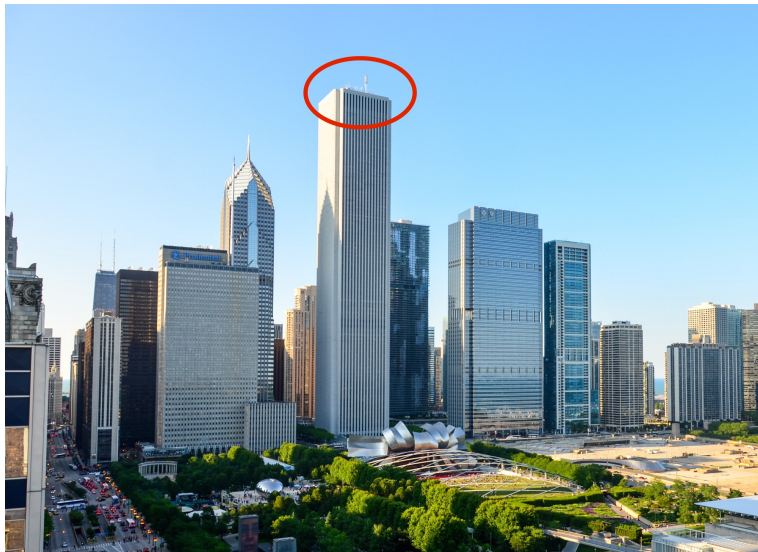
First Chicago-NYC Microwave Network



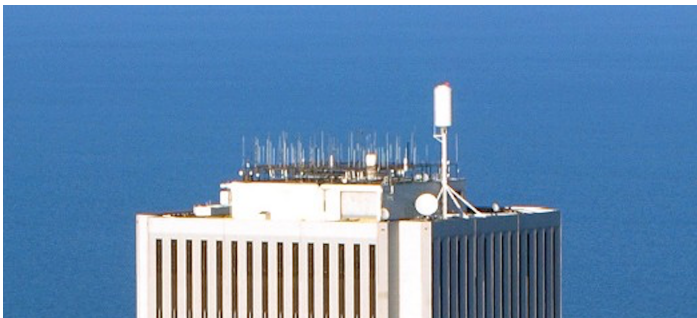
The HFT Arms Race: Continued



The HFT Arms Race: Continued



The HFT Arms Race: Continued



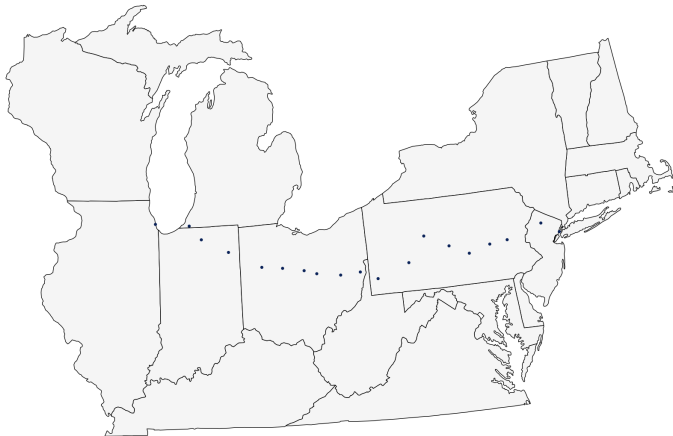
The HFT Arms Race: Continued

Active Microwave Networks in the Chicago-NYC-DC Region as of 2010-01-01



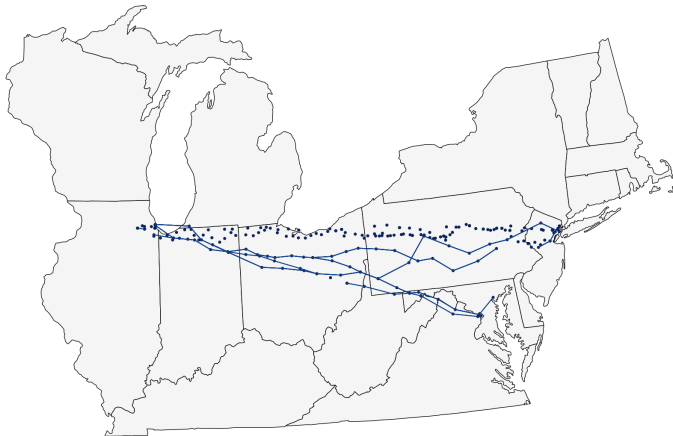
The HFT Arms Race: Continued

Active Microwave Networks in the Chicago-NYC-DC Region as of 2011-01-01



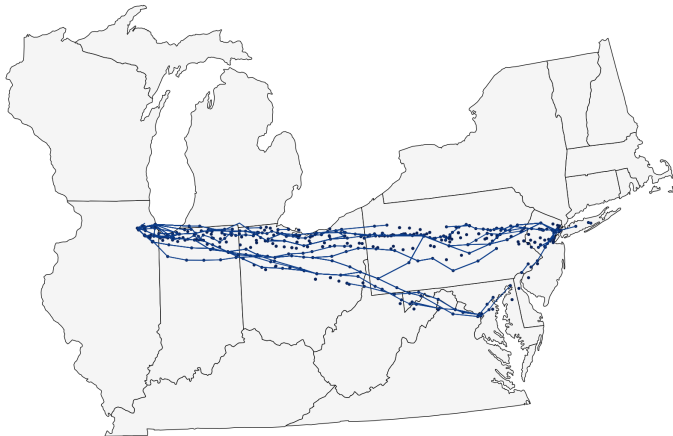
The HFT Arms Race: Continued

Active Microwave Networks in the Chicago-NYC-DC Region as of 2012-01-01



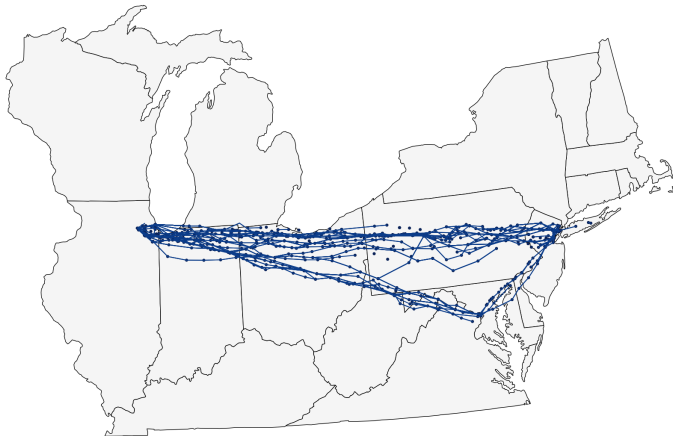
The HFT Arms Race: Continued

Active Microwave Networks in the Chicago-NYC-DC Region as of 2013-01-01



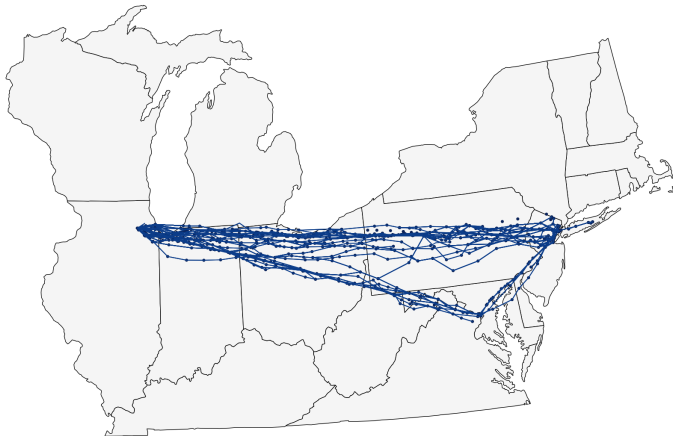
The HFT Arms Race: Continued

Active Microwave Networks in the Chicago-NYC-DC Region as of 2014-01-01



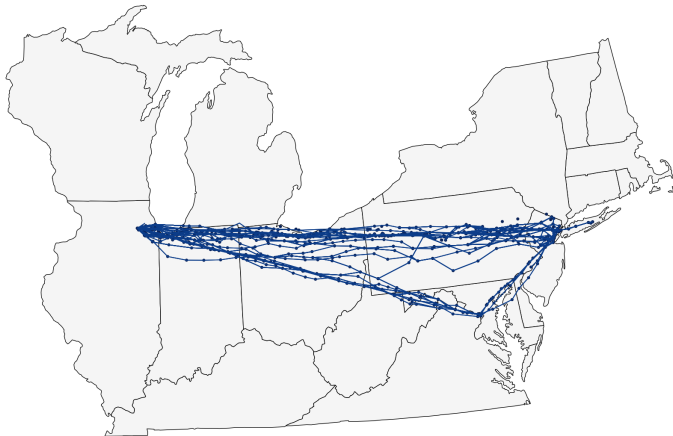
The HFT Arms Race: Continued

Active Microwave Networks in the Chicago-NYC-DC Region as of 2015-01-01



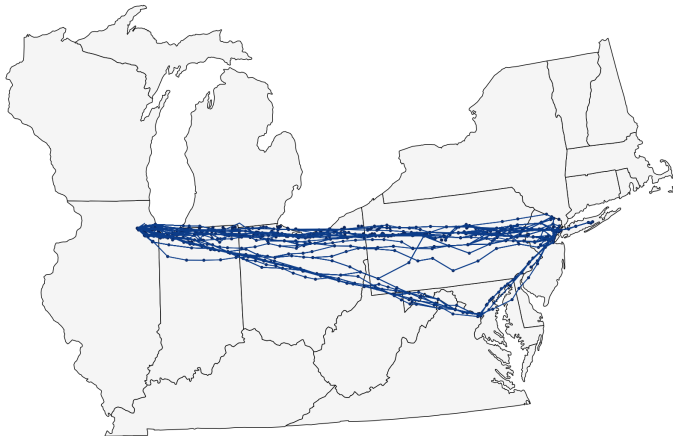
The HFT Arms Race: Continued

Active Microwave Networks in the Chicago-NYC-DC Region as of 2016-01-01



The HFT Arms Race: Continued

Active Microwave Networks in the Chicago-NYC-DC Region as of 2016-12-01



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 → 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 → 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

Frequent Batch Auctions: Definition

- ▶ The trading day is divided into equal-length discrete batch intervals, each of length $\tau > 0$.
- ▶ During each batch interval traders submit bids and asks
 - ▶ Can be freely modified, canceled, etc.
 - ▶ If an order is not executed in the current batch, it remains outstanding for the next batch, etc.
 - ▶ Just like standard limit orders

Frequent Batch Auctions: Definition

- ▶ The trading day is divided into equal-length discrete batch intervals, each of length $\tau > 0$.
- ▶ During each batch interval traders submit bids and asks
 - ▶ Can be freely modified, canceled, etc.
 - ▶ If an order is not executed in the current batch, it remains outstanding for the next batch, etc.
 - ▶ Just like standard limit orders
- ▶ At the end of each interval, the exchange aggregates all outstanding orders and computes supply and demand curves

Frequent Batch Auctions: Definition

- ▶ The trading day is divided into equal-length discrete batch intervals, each of length $\tau > 0$.
- ▶ During each batch interval traders submit bids and asks
 - ▶ Can be freely modified, canceled, etc.
 - ▶ If an order is not executed in the current batch, it remains outstanding for the next batch, etc.
 - ▶ Just like standard limit orders
- ▶ At the end of each interval, the exchange aggregates all outstanding orders and computes supply and demand curves
- ▶ If supply and demand intersect, then the market clears where supply equals demand, “uniform price”

Frequent Batch Auctions: Definition

- ▶ The trading day is divided into equal-length discrete batch intervals, each of length $\tau > 0$.
- ▶ During each batch interval traders submit bids and asks
 - ▶ Can be freely modified, canceled, etc.
 - ▶ If an order is not executed in the current batch, it remains outstanding for the next batch, etc.
 - ▶ Just like standard limit orders
- ▶ At the end of each interval, the exchange aggregates all outstanding orders and computes supply and demand curves
- ▶ If supply and demand intersect, then the market clears where supply equals demand, “uniform price”
- ▶ Priority: still price-time, but treat time as discrete.

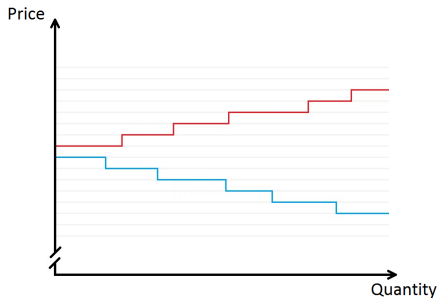
Frequent Batch Auctions: Definition

- ▶ The trading day is divided into equal-length discrete batch intervals, each of length $\tau > 0$.
- ▶ During each batch interval traders submit bids and asks
 - ▶ Can be freely modified, canceled, etc.
 - ▶ If an order is not executed in the current batch, it remains outstanding for the next batch, etc.
 - ▶ Just like standard limit orders
- ▶ At the end of each interval, the exchange aggregates all outstanding orders and computes supply and demand curves
- ▶ If supply and demand intersect, then the market clears where supply equals demand, “uniform price”
- ▶ Priority: still price-time, but treat time as discrete.
- ▶ Information policy: same info as in continuous, but disseminate info in discrete time
 - ▶ After each time interval, report all trades, and report all outstanding orders. (Discrete-time analog of reporting the state of the limit order book).

Frequent Batch Auctions: 3 Cases

Case 1: Nothing happens during the batch interval

- ▶ Very common case: most instruments, most 1ms periods, there is zero activity
- ▶ All outstanding orders carry forward to next interval
- ▶ Analogous to displayed liquidity in a LOB market

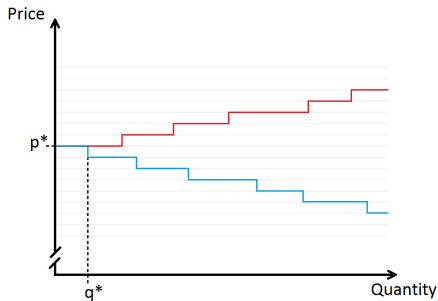


Case 1: No Trade

Frequent Batch Auctions: 3 Cases

Case 2: Small amount of trade

- ▶ Example: an investor arrives wanting to buy a small amount at market
- ▶ Demand will cross supply at the bottom of the supply curve
- ▶ Analogous to trading at the ask in a LOB market



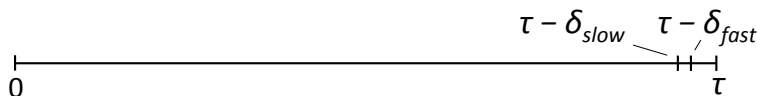
Case 2: Investor Buys q^*

Frequent Batch Auctions: 3 Cases

Case 3: Burst of activity in the interval

- ▶ Example: there is public news (jump in y) and many algos respond
- ▶ In this case, FBA and LOB are importantly different

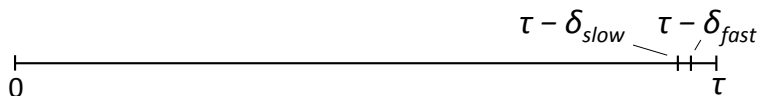
Why FBA Solves the Problem



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 relevant. 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)

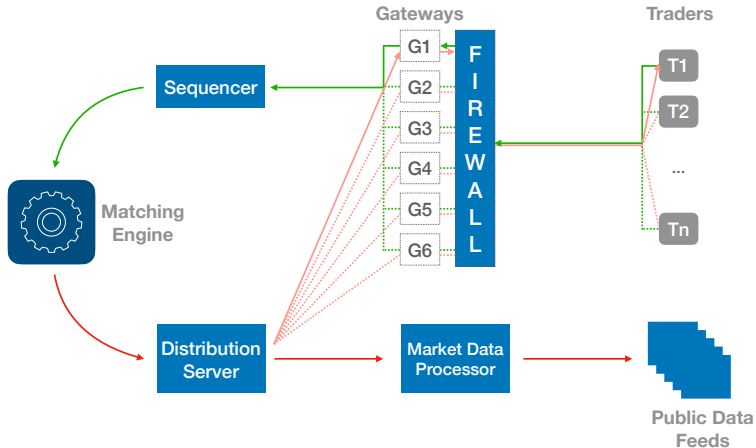
Message Data, Simple Methodology

- ▶ 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

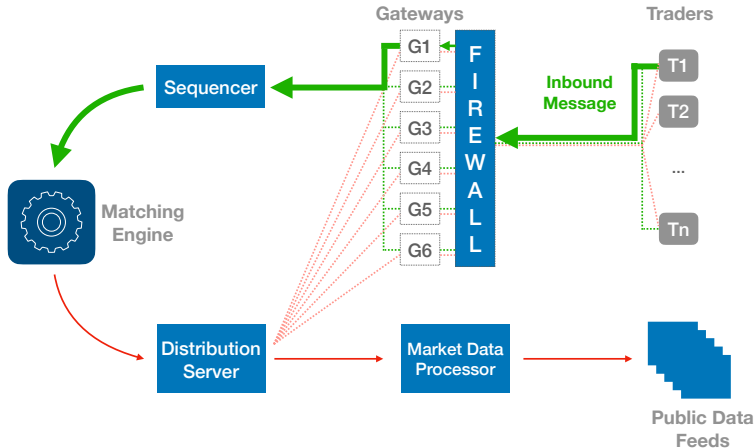
- ▶ We obtained message data from the London Stock Exchange (by a request under Section 165 of the Financial Service and Markets Act)
 - ▶ All message data for all FTSE 350 stocks for a 9 week period in Fall 2015
 - ▶ Timestamps accurate to the microsecond (0.000001s)
 - ▶ Timestamps at the right location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



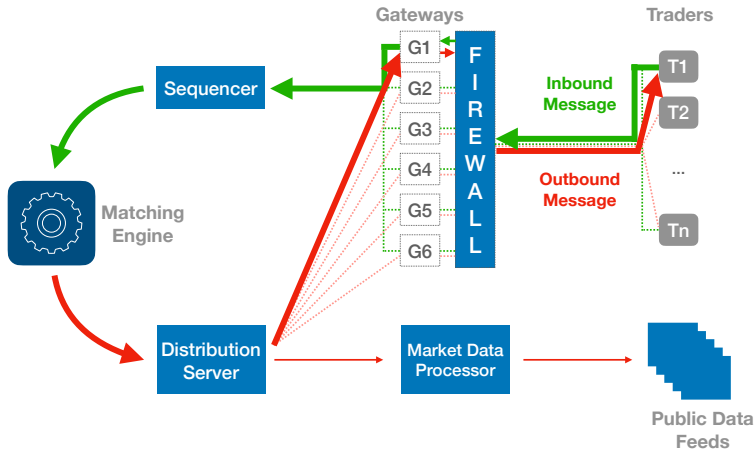
Notes: Please see the text of Section 2.1 for supporting details for this figure.

Exchange Schematic



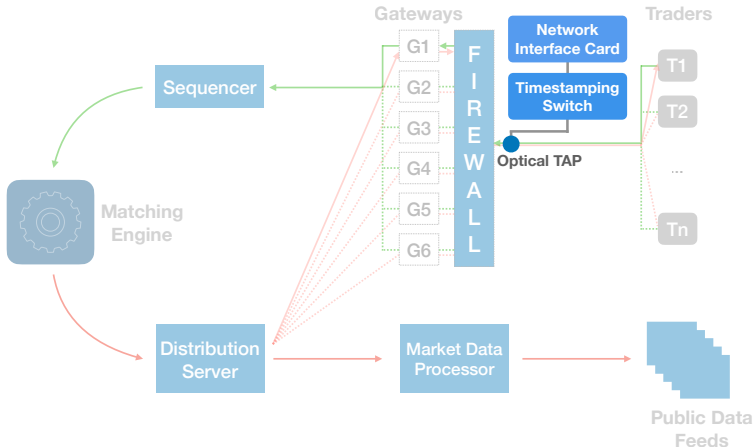
Notes: Please see the text of Section 2.1 for supporting details for this figure.

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.

Defining a Race

- ▶ The theory suggests that the empirical signature of a BCS-style latency-arbitrage race, as distinct from Glosten-Milgrom-style informed trading, is:
 1. Multiple market participants acting on the same symbol, price and side
 2. Either mix of takes + cancels (eqm emphasized in BCS) or all takes (if liquidity provider is slow)
 3. Some succeed, some fail
 4. All at the “same time”
- ▶ Items #1-#3 are relatively straightforward to implement
 - ▶ Please see paper for various sensitivities
- ▶ Item #4 is harder
 - ▶ In theory, such a thing as “same time”
 - ▶ But in data, no two things happen at *exactly* the same time
 - ▶ Main approach: “information horizon” (avg, 200 microseconds)
 - ▶ Alternative: wide range of sensitivities (from 50us to 3ms)

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

JPE, Forthcoming

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.

Flow Trading

Eric Budish
Peter Cramton
Pete Kyle
Mina Lee
David Malec

Working Paper, March 2023

Flow Trading: Key Ideas

- ▶ Takes FBA idea to the trade of arbitrary portfolios of assets
 - ▶ Arbitrary linear combinations with real-valued positive/negative weights
 - ▶ Complements, substitutes
- ▶ Portfolios and arbitrage are at the heart of finance ... trade portfolios directly! Engage in arbitrage 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)
 - ▶ Piecewise-linear, downward-sloping demand curves, continuous in price and quantity, with quantity expressed as “flows”
 - ▶ “Buy a maximum of 1 portfolio unit per second until 1000 units are bought.”

Flow Trading: How Orders Work

- ▶ There are N assets indexed n , I orders indexed i
- ▶ An order is specified by a tuple $(\mathbf{w}_i, p_i^L, p_i^H, q_i, Q_i^{max})$
- ▶ Description of portfolio: vector of portfolio weights $\mathbf{w}_i \in \mathbb{R}^N$
 - ▶ Individual asset: one nonzero weight to buy (+) or sell (-)
 - ▶ Substitutes: mix of positive and negative weights
 - ▶ Complements: 500 positive index weights to buy the S&P 500
 - ▶ Market making: pair of orders with weights \mathbf{w}_i and $-\mathbf{w}_i$
- ▶ Two limit prices for the portfolio ($p_i^L = \$50.30$, $p_i^H = \$50.40$)
 - ▶ Meaning: trade full amount at p_i^L or better, declining linearly in interval $[p_i^L, p_i^H]$
 - ▶ Negative prices if selling ($p_i^L = -\$50.40$ and $p_i^H = -\$50.30$).
Always $p_i^L < p_i^H$
- ▶ Maximum execution rate ($q_i = 2.00$ units per second)
- ▶ Cumulative quantity to be executed ($Q_i^{max} = 10000$ units)

Math: One Portfolio Order

- ▶ Order i is specified by the tuple $(\mathbf{w}_i, p_i^L, p_i^H, q_i, Q_i^{max})$
- ▶ Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)$ denote a vector of N asset prices. The price of the portfolio is the weighted sum:

$$p_i = \boldsymbol{\pi}^T \mathbf{w}_i$$

- ▶ Assume the order's cumulative purchased quantity is not within q_i of Q_i^{max} . The execution rate x_i of order i is given by:

$$x_i = D^i(p_i) = q_i \cdot \text{trunc} \left(\frac{p_i^H - p_i}{p_i^H - p_i^L} \right)$$

$$\text{where } \text{trunc}(x) := \begin{cases} 1, & \text{for } x \geq 1 \\ x, & \text{for } 0 < x < 1 \\ 0, & \text{for } x \leq 0 \end{cases}$$

Math: Market Clearing

- ▶ Market clears in N assets
- ▶ At price vector $\boldsymbol{\pi} \in \mathbb{R}^N$, the exchange converts each order i 's demand for portfolio units to demand for underlying assets by multiplying its portfolio weights \mathbf{w}_i . Summing over the I orders yields the excess demand vector:

$$\text{Excess Demand Vector} = D(\boldsymbol{\pi}) := \sum_{i=1}^I D^i(\boldsymbol{\pi}^T \mathbf{w}_i) \cdot \mathbf{w}_i$$

- ▶ The exchange seeks to find a market clearing price vector

$$D(\boldsymbol{\pi}) = 0, \quad (N \text{ equations in } N \text{ unknowns})$$

- ▶ In which case each order i executes at rate

$$x_i = D^i(\boldsymbol{\pi}^T \mathbf{w}_i), \quad (\text{scalar equation in portfolio units } x_i)$$

Existence Theorems

- ▶ Our problem of finding market-clearing prices is formulated as two optimization problems
 - ▶ A primal problem of finding quantities that maximize “as-bid” dollar value (quadratic program)
 - ▶ A dual problem of finding prices that minimize the cost of non-clearing prices
- ▶ **Theorem 1 (Existence and Uniqueness of Optimal Quantities).** *There exists a unique vector of trade rates x^* which solves the as-bid maximization problem.*
- ▶ **Theorem 2 (Existence of Market-Clearing Prices).** *There exists a vector of market-clearing prices π^* which solves the dual problem.*
- ▶ Proofs are relatively straightforward.

Why is Existence so Simple?

- ▶ The language allows users to define arbitrary portfolios, including complements and substitutes
- ▶ In general equilibrium theory and indivisible goods literatures, complements especially make existence hard (Arrow-Debreu-McKenzie, Starr 1969, Hatfield-Kominers-Westkamp 2021, Baldwin-Klemperer 2019)
- ▶ Yet here the existence proof is simple. Why?
 - ▶ Goods are infinitely divisible
 - ▶ Portfolio demand schedules are downward sloping
 - ▶ Utility for each order is defined only on the line segment associated with the portfolio weights (not defined off diagonal)
 - ▶ Quantities are bounded. Zero trade is feasible.
 - ▶ No in-order contingencies or linkages across multiple orders. (This limits the comps and subs).
- ▶ A sweet spot? Expressive enough to be useful, and existence is guaranteed.

Computation

- ▶ Question: prices and quantities exist. Can we compute them?
- ▶ Many economic settings where prices and quantities are known to exist but hard to find (Scarf and Hansen, 1973)
- ▶ Many economic settings where prices are trivial to compute—one asset version of our problem is an easy example!
- ▶ Our problem lies in between
- ▶ Plan
 - ▶ Show that gradient method works. “Easier than Scarf’s problem”
 - ▶ Result: gradient method convergence slow (confirmed in simulations)
 - ▶ Add “Exchange as market maker of last resort” which enables interior point methods
 - ▶ Result: faster in theory, and also in simulations
- ▶ Goal in mind: solve large problems in less than one second.

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 → 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
 - ▶ Could even add portfolios!

Topics for Discussion (1/2)

- ▶ Some research I would love to see in crypto markets:
- ▶ Quantification.
 - ▶ Someone should quantify sniping in crypto markets (follow cookbook in Aquilina, Budish and O'Neill 2022)
 - ▶ I bet it's massive - as % of volume and as % tax on trading
- ▶ Information and Exchanges
 - ▶ Reg NMS in U.S. stock market: exchanges must disseminate info updates about trades and quotes quickly, and are not allowed to (knowingly) execute trades against stale quotes
 - ▶ This limits magnitude of within-stock arbitrage (stock XYZ on exchange A vs. B)
 - ▶ No such limits in crypto markets — I'd bet LOTS of within-symbol arbitrage (ETH on exchange A vs. exchange B)
 - ▶ Would make crypto a good forum to study information policy

Topics for Discussion (2/2)

- ▶ Implementing FBAs
 - ▶ CowSwap approach for defi
 - ▶ Multi-asset, inspired by Flow Trading
 - ▶ A difficulty: no numeraire good
 - ▶ Solver competition approach
 - ▶ What are economics for a centralized exchange adopting? Do they face “will the market fix the market?” perverse incentives?
- ▶ MEV
 - ▶ There is of course a very close connection between the data/colo economics we study in Budish, Lee and Shim (2023) and the MEV issue in crypto markets.
 - ▶ Q1: quantification
 - ▶ Q2: exchange design to counter MEV (FBAs, are there other complementary ideas?)
 - ▶ Q3: *protocol* design to counter MEV. It would be very interesting to solve the problem at a level that is more abstract than financial exchanges

Friedman on Theory → 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

Zooming Out: Private vs. Social Innovation Incentives

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

Private vs. Social Incentives: Finance Innovations

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