Will the Market Fix the Market?

Eric Budish
University of Chicago, Booth School of Business

AEA / AFA Joint Luncheon
January 6th 2017
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!

- Question: how could such tiny speed advantages be worth so much money?
  - 3 milliseconds too short to be about fundamentals
  - Economists intrinsically skeptical about technical trading
The HFT Arms Race: Market Design Perspective

- My collaborators Cramton, Shim and I approached the HFT arms race from the perspective of market design.
  - We assume that HFT’s are optimizing with respect to market rules as presently given.
  - But, ask whether these are the right rules (avoids “is HFT good or evil?”).
  - Focus on the precise institutional details of the markets in question. Al Roth: “Economist as Engineer”
  - Milton Friedman: “rules of the game”.

Indeed, we find a subtle flaw in the design of modern financial exchanges.

Flaw: exchanges treat time as a continuous variable and process requests to trade serially.
The HFT Arms Race: Market Design Perspective

- My collaborators Cramton, Shim and I approached the HFT arms race from the perspective of market design.
  - We assume that HFT’s are optimizing with respect to market rules as presently given.
  - But, ask whether these are the right rules (avoids “is HFT good or evil?”).
  - Focus on the precise institutional details of the markets in question. Al Roth: “Economist as Engineer”
  - Milton Friedman: “rules of the game”.

- Indeed, we find a subtle flaw in the design of modern financial exchanges.

- Flaw: exchanges treat time as a *continuous variable* and process requests to trade *serially*
The HFT Arms Race: Market Design Perspective

- Continuous-time + serial processing $\rightarrow$ riskless arbitrage profits from symmetric public information
  - (info either technical or fundamental)

- 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.
  1. Transforms competition on speed into competition on price.
  2. Fixes the violation of EMH.
  3. Improves liquidity and stops the arms race.
Plan for Talk

- Part I: Budish, Cramton and Shim (QJE, 2015)
  - Empirical facts: continuous market violates basic asset pricing principles at HFT time horizons
  - Theory model: critique of the continuous limit order book
  - Market design solution: Frequent batch auctions

- Part II: Will the Market Fix the Market?
  - Main question re BCS: private vs. regulatory solution?
  - New research with Lee and Shim
  - Some concluding thoughts
“The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response”

Eric Budish
Peter Cramton
John Shim

QJE, November 2015
Brief Description of the Continuous Limit Order Book

- **Basic building block: limit order**
  - Specifies a price, quantity, and buy/sell (bid/ask)
  - “Buy 100 shares of XYZ at $100.00”
- **Traders may submit limit orders to the market any time during the trading day**
  - Also may cancel or modify outstanding limit orders at any time
  - Orders and cancelations are processed by the exchange one-at-a-time in order of receipt (serial process)

- **Set of outstanding orders is known as the limit order book**
- **Trade occurs whenever a new limit order is submitted that is either (i) bid $\geq$ lowest ask; (ii) ask $\leq$ highest bid**
  - New limit order is interpreted as accepting (fully or partially) one or more outstanding orders

<table>
<thead>
<tr>
<th>Price</th>
<th>Shares</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.40</td>
<td>7,176</td>
</tr>
<tr>
<td>62.39</td>
<td>6,050</td>
</tr>
<tr>
<td>62.38</td>
<td>3,900</td>
</tr>
<tr>
<td>62.37</td>
<td>2,600</td>
</tr>
<tr>
<td>62.36</td>
<td>1,000</td>
</tr>
<tr>
<td>62.35</td>
<td>800</td>
</tr>
<tr>
<td>62.34</td>
<td>2,050</td>
</tr>
<tr>
<td>62.33</td>
<td>3,200</td>
</tr>
<tr>
<td>62.32</td>
<td>4,300</td>
</tr>
<tr>
<td>62.31</td>
<td>5,600</td>
</tr>
</tbody>
</table>
Market Correlations Break Down at High Frequency
ES vs. SPY: 1 Day

The graph shows the index points for both ES and SPY from 09:00:00 to 14:00:00. The ES midpoint and SPY midpoint are plotted against time (CT) and index points (ES). The graph indicates a significant fluctuation in correlations between ES and SPY at high frequency.
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

![Graph showing Market Correlations Break Down at High Frequency between ES and SPY](image-url)
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
Latency Arb and Arms Race are “Constants”

To summarize:

- Competition does increase the speed requirements for capturing arbs (“raises the bar”)
- Competition does not reduce the size or frequency of arb opportunities
- Suggests we should think of latency arbitrage and the resulting arms race as a “constant” of the current market design
Total Size of the Arms Race Prize

- Estimate annual value of ES-SPY arbitrage is $75mm (we suspect underestimate, details in paper)

- 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

![GBP/USD Future vs. ETF Graph](image1)

JPY/USD Future vs. ETF

![JPY/USD Future vs. ETF Graph](image2)
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**

<table>
<thead>
<tr>
<th>Futures 1</th>
<th>Futures 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. SPDR S&amp;P 500 ETF (SPY)</td>
<td>Australian Dollar Futures (6B) vs. Spot AUDUSD</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. iShares S&amp;P 500 ETF (IVV)</td>
<td>Swiss Franc Futures (6S) vs. Spot USDCHF</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. Vanguard S&amp;P 500 ETF (VOO)</td>
<td>Canadian Dollar Futures (6C) vs. Spot USD CAD</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. ProShares UltraPro (3x) S&amp;P 500 ETF (UPRO)</td>
<td>Gold Futures (GC) vs. mNY Gold Futures (Q0)</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. ProShares Short S&amp;P 500 ETF (SH)</td>
<td>Gold Futures (GC) vs. Spot Gold (XAUUSD)</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. ProShares Ultra (2x) Short S&amp;P 500 ETF (SDS)</td>
<td>Gold Futures (GC) vs. E-micro Gold Futures (MGC)</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. ProShares UltraPro (3x) Short S&amp;P 500 ETF (SPXU)</td>
<td>Gold Futures (GC) vs. SPDR Gold Trust (GLD)</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. 500 Constituent Stocks</td>
<td>Gold Futures (GC) vs. iShares Gold Trust (IAU)</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. 9 Select Sector SPDR ETFs</td>
<td>mNY Gold Futures (Q0) vs. E-micro Gold Futures (MGC)</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. E-mini Dow Futures (YM)</td>
<td>mNY Gold Futures (Q0) vs. Spot Gold (XAUUSD)</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. E-mini Nasdaq 100 Futures (NQ)</td>
<td>mNY Gold Futures (Q0) vs. SPDR Gold Trust (GLD)</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. E-mini S&amp;P MidCap 400 Futures (EMD)</td>
<td>iShares Gold Trust (IAU)</td>
</tr>
<tr>
<td>E-mini S&amp;P 500 Futures (ES) vs. Russell 2000 Index Mini Futures (TF)</td>
<td>E-micro Gold Futures (MGC) vs. SPDR Gold Trust (GLD)</td>
</tr>
<tr>
<td>E-mini Dow Futures (YM) vs. SPDR Dow Jones Industrial Average ETF (DIA)</td>
<td>E-micro Gold Futures (MGC) vs. iShares Gold Trust (IAU)</td>
</tr>
<tr>
<td>E-mini Dow Futures (YM) vs. ProShares Ultra (2x) Dow 30 ETF (ODM)</td>
<td>E-micro Gold Futures (MGC) vs. Spot Gold (XAUUSD)</td>
</tr>
<tr>
<td>E-mini Dow Futures (YM) vs. ProShares UltraPro (3x) Dow 30 ETF (UDOW)</td>
<td>Market Vectors Gold Miners (GDX) vs. Direxion Daily Gold Miners Bull 3x (NUGT)</td>
</tr>
<tr>
<td>E-mini Dow Futures (YM) vs. ProShares Short Dow 30 ETF (DOG)</td>
<td>Silver Futures (SI) vs. mNY Silver Futures (QI)</td>
</tr>
<tr>
<td>E-mini Dow Futures (YM) vs. ProShares Ultra (2x) Short Dow 30 ETF (OXD)</td>
<td>Silver Futures (SI) vs. iShares Silver Trust (SVX)</td>
</tr>
<tr>
<td>E-mini Dow Futures (YM) vs. ProShares UltraPro (3x) Short Dow 30 ETF (S0DX)</td>
<td>mNY Silver Futures (QI) vs. iShares Silver Trust (SVX)</td>
</tr>
<tr>
<td>E-mini Dow Futures (YM) vs. 30 Constituent Stocks</td>
<td>mNY Silver Futures (QI) vs. Spot Silver (XAGUSD)</td>
</tr>
<tr>
<td>E-mini Nasdaq 100 Futures (NQ) vs. ProShares QQQ Trust ETF (QQQ)</td>
<td>Platinum Futures (PL) vs. Spot Platinum (XPTUSD)</td>
</tr>
<tr>
<td>E-mini Nasdaq 100 Futures (NQ) vs. Technology Select Sector SPDR (XLK)</td>
<td>Palladium Futures (PA) vs. Spot Palladium (XPDUSD)</td>
</tr>
<tr>
<td>E-mini Nasdaq 100 Futures (NQ) vs. 100 Constituent Stocks</td>
<td>Eurodollar Futures Front Month (ED) vs. (12 back month contracts)</td>
</tr>
<tr>
<td>Russell 2000 Index Mini Futures (TF) vs. iShares Russell 2000 ETF (IWM)</td>
<td>10 Yr Treasury Note Futures (ZN) vs. 5 Yr Treasury Note Futures (ZI)</td>
</tr>
<tr>
<td>Euro Stoxx 50 Futures (FESX) vs. Xetra DAX Futures (FDAX)</td>
<td>10 Yr Treasury Note Futures (ZN) vs. 30 Yr Treasury Bond Futures (ZB)</td>
</tr>
<tr>
<td>Euro Stoxx 50 Futures (FESX) vs. CAC 40 Futures (FCE)</td>
<td>10 Yr Treasury Note Futures (ZN) vs. 7-10 Yr Treasury Note</td>
</tr>
<tr>
<td>Euro Stoxx 50 Futures (FESX) vs. iShares MSCI EAFE Index Fund (EFA)</td>
<td>2 Yr Treasury Note Futures (ZT) vs. 1-2 Yr Treasury Note</td>
</tr>
<tr>
<td>Nikkei 225 Futures (NYF) vs. MSCI Japan Index Fund (EWJ)</td>
<td>2 Yr Treasury Note Futures (ZT) vs. iShares Barclays 1-3 Yr Treasury Fund (SHY)</td>
</tr>
<tr>
<td>Financial Sector SPDR (XLF) vs. Constituents</td>
<td>5 Yr Treasury Note Futures (ZF) vs. 4-5 Yr Treasury Note</td>
</tr>
<tr>
<td>Financial Sector SPDR (XLK) vs. Direxion Daily Financial Bull 3x (FAS)</td>
<td>30 Yr Treasury Bond Futures (ZB) vs. iShares Barclays 20 Yr Treasury Fund (TLT)</td>
</tr>
<tr>
<td>Energy Sector SPDR (XLE) vs. Constituents</td>
<td>30 Yr Treasury Bond Futures (ZB) vs. ProShares UltraShort 20 Yr Treasury Fund (TBT)</td>
</tr>
<tr>
<td>Industrial Sector SPDR (XLI) vs. Constituents</td>
<td>30 Yr Treasury Bond Futures (ZB) vs. ProShares Short 20 Year Treasury Fund (TFB)</td>
</tr>
<tr>
<td>Cons. Staples Sector SPDR (XLP) vs. Constituents</td>
<td>30 Yr Treasury Bond Futures (ZB) vs. 15+ Yr Treasury Bond</td>
</tr>
<tr>
<td>Materials Sector SPDR (XLB) vs. Constituents</td>
<td>Crude Oil Futures Front Month (CL) vs. (6 back month contracts)</td>
</tr>
<tr>
<td>Utilities Sector SPDR (XLU) vs. Constituents</td>
<td>Crude Oil Futures (CL) vs. ICE Brent Crude (B)</td>
</tr>
<tr>
<td>Technology Sector SPDR (XLK) vs. Constituents</td>
<td>Crude Oil Futures (CL) vs. United States Oil Fund (USO)</td>
</tr>
<tr>
<td>Health Care Sector SPDR (XLV) vs. Constituents</td>
<td>Crude Oil Futures (CL) vs. ProShares Ultra DJ-UBS Crude Oil (UCO)</td>
</tr>
<tr>
<td>Cons. Discretionary Sector SPDR (XLY) vs. Constituents</td>
<td>Crude Oil Futures (CL) vs. iPath S&amp;P Crude Oil Index (OIL)</td>
</tr>
<tr>
<td>SPDR Homebuilders ETF (XHB) vs. Constituents</td>
<td>ICE Brent Crude Front Month (B) vs. (6 back month contracts)</td>
</tr>
<tr>
<td>SPDR S&amp;P 500 Retail ETF (KRT) vs. Constituents</td>
<td>ICE Brent Crude (B) vs. United States Oil Fund (USO)</td>
</tr>
<tr>
<td>Euro FX Futures (6E) vs. Spot EURUSD</td>
<td>ICE Brent Crude (B) vs. ProShares Ultra DJ-UBS Crude Oil (UCO)</td>
</tr>
<tr>
<td>Japanese Yen Futures (6J) vs. Spot USDJPY</td>
<td>ICE Brent Crude (B) vs. iPath S&amp;P Crude Oil Index (OIL)</td>
</tr>
<tr>
<td>British Pound Futures (6B) vs. Spot GBPUSD</td>
<td>Natural Gas (Henry Hub) Futures (NG) vs. United States Nat Gas Fund (UNG)</td>
</tr>
</tbody>
</table>
Total Size of the Arms Race Prize

- Estimate annual value of ES-SPY arbitrage is $75mm (we suspect underestimate, details in paper)

- 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)

- Common sense extrapolation from our ES-SPY estimates suggest that the sums are substantial.
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: Preliminaries

- 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 $x$
- 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
- The public signal $y$ evolves as a compound Poisson jump process, symmetric with mean zero
  - Arrival rate $\lambda_{\text{jump}}$
  - “Jump size” distribution $J$ (dist. of absolute value of jumps)
Investors

- Represent end users of financial markets: mutual funds, pension funds, hedge funds, etc.
- Since there is no asymmetric information about fundamentals, could be called “liquidity traders” or “noise traders”
- Arrive stochastically to the market with an inelastic need to either buy or sell 1 unit of $x$
- Poisson arrival rate is $\lambda_{\text{invest}}$. Equal probability of need to buy vs. need to sell
- Mechanical strategy: trade at market immediately upon arrival
Players: Investors and Trading Firms

Trading Firms

- Equivalently: HFTs, market makers, algorithmic traders
- No intrinsic demand to buy or sell \( x \)
- Their goal is simply to buy \( x \) at prices lower than \( y \) and sell at prices higher than \( y \). Payoffs:
  - Buy \( x \) at price \( p \) at time \( t \): earn \( y_t - p \)
  - Sell \( x \) at price \( p \) at time \( t \): earn \( p - y_t \)

- Objective is to maximize profits per unit time
- Entry
  - Initially, \( N \geq 2 \) exogenously in the market.
  - Below we will endogenize entry via investment in speed technology.
Latency

Exogenous entry case

► No latency in observing $y$
  ► Trading firms observe innovations in the signal $y$ with zero time delay, for free.
► No latency in submitting orders to the exchange
  ► If multiple orders reach the market at the same time, the order in which they are processed is random (serial processing)
  ► Alternatively, orders are transmitted with small random latency, and processed in order of receipt (eg, colocation)
► Again, best case scenario for the continuous market

Endogenous entry case

► Will add latency in observing $y$
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.”
Suppose \( y \) jumps, e.g., from \( y_1 \) to \( y_2 \)

- This is the moment at which the correlation between \( y \) and \( x \) temporarily breaks down

Trading firms providing liquidity in the market for \( x \) send a message to the continuous limit order book

- Cancel old quotes based on \( y_1 \)
- Add new quotes based on \( y_2 \)
“Sniping”

- However, at the exact same time, other trading firms send a message to the continuous market attempting to “snipe” the stale quotes before they are adjusted
  - Buy at the old quotes based on $y_1$, before these quotes are withdrawn

- Since the continuous market processes messages in serial — that is, one at a time — it is possible that a message to snipe a stale quote will get processed before the message to cancel the stale quote

- In fact, not only possible but probable
  - For every 1 liquidity provider trying to get out of the way
  - $N-1$ other trading firms trying to snipe him
  - Hence, when there is a big jump, each liquidity provider gets sniped with probability $\frac{N-1}{N}$
“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.
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, \textit{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 \textit{extremely} short run
Equilibrium Effect on Liquidity

In equilibrium, the bid-ask spread is such that trading firms are indifferent between liquidity provision and sniping.

- Return to liquidity provision
  - Benefits: $\lambda_{\text{invest}} \cdot \frac{s}{2}$
  - Costs: $\lambda_{\text{jump}} \cdot \Pr(J > \frac{s}{2}) \cdot \mathbb{E}(J - \frac{s}{2} | J > \frac{s}{2}) \cdot \frac{N-1}{N}$

- Return to sniping
  - Benefits: $\lambda_{\text{jump}} \cdot \Pr(J > \frac{s}{2}) \cdot \mathbb{E}(J - \frac{s}{2} | J > \frac{s}{2}) \cdot \frac{1}{N}$

- Indifference condition:
  \[
  \lambda_{\text{invest}} \cdot \frac{s^*}{2} = \lambda_{\text{jump}} \cdot \Pr(J > \frac{s^*}{2}) \cdot \mathbb{E}(J - \frac{s^*}{2} | J > \frac{s^*}{2}) \tag{1}
  \]

- Uniquely pins down $s$. Interpretation:
  - LHS: revenue from investors due to non-zero bid-ask spread
  - RHS: rents to trading firms from mechanical arbitrages
Equilibrium, Endogenous Entry

- Now, endogenize entry.
  - Trading firms observe the signal $y$ with a small time delay, $\delta_{slow} > 0$, for free.
  - Can pay a cost $c_{speed}$ to reduce latency from $\delta_{slow}$ to $\delta_{fast}$. Let $\delta = \delta_{slow} - \delta_{fast}$.

- Equilibrium is very similar to above. Uniquely characterized by
  - Trading Firms’ indifference between liquidity provision and stale-quote sniping. Exact same $s^*$ as above.
  - Trading Firms’ free entry condition: TFs keep entering until the marginal TF earns zero profits. Characterizes $N^*$.

- Key new 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.
What's the Market Failure?

Chicago question: isn’t the arms race just healthy competition? what’s the market failure?
What’s the Market Failure?

Market Failure 1: Sniping
▶ Mechanical arb opportunities are “built in” to the market design
  ▶ These arb opportunities violate weak-form EMH (Fama, 1970)
  ▶ Rents from symmetrically observed public information
  ▶ “Such a situation would reflect a rather basic malfunctioning of the market mechanism.”

Market Failure 2: Arms Race
▶ The arb rents then induce an arms race for speed
▶ Mathematically, a prisoners’ dilemma
Remark I: Role of HFTs

- In our model HFTs endogenously perform two functions
  - Useful: liquidity provision / price discovery
  - Rent-seeking: sniping stale quotes

- The rent-seeking seems like zero-sum activity among HFTs
  - But this misses the economics: sniping is like a tax on liquidity provision, which in turn harms non-HFTs

- Clarification
  - Our results do not imply that on net HFT has been bad for liquidity or social welfare
  - Evidence is strong that humans → computers has on the whole been quite positive for markets (cf. Hendershott, Jones and Menkveld 2011), though gains appear to be mostly in late 90s / early 00s
  - Our results do say that sniping is bad for liquidity and the speed race is socially wasteful
Remark II: Arms Race is a “Constant”

- Arms race prize = expenditures on speed = cost to investors
  \[ = \lambda_{jump} \cdot \Pr(J > \frac{s^*}{2}) \cdot \mathbb{E}(J - \frac{s^*}{2} | J > \frac{s^*}{2}) \]

- Comparative static: the negative effects of the arms race on liquidity and welfare do not depend on either
  - the cost of speed (if speed is cheap, there will be more entry)
  - the magnitude of speed improvements (seconds, milliseconds, microseconds, nanoseconds, ...)

- The problem we identify is an equilibrium feature of continuous limit order books
  - not competed away as HFTs get faster and faster
  - ties in nicely with empirical results

- Implication: the race for speed will never end as long as we have continuous-time trading
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
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 by RSN Wireless as of 2016-12-01
The HFT Arms Race: Continued

How do you use Metamako?

“Metamako makes it easier to roll out and maintain complex network set ups. We are now able to more aggressively pursue latency goals.”

— Network Engineer, Financial Service Co.
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 → 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 submits 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: Definition

- The trading day is divided into equal-length discrete batch intervals, each of length $\tau > 0$.
- During each batch interval traders submits 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 submits 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 submits 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
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
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 → $\tau - \delta_{\text{slow}}$ everybody sees it
  - $\tau - \delta_{\text{fast}}$ → $\tau$ nobody sees it
  - $\tau - \delta_{\text{slow}}$ → $\tau - \delta_{\text{fast}}$ 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.
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!
Takeaways from Equilibrium Analysis

- If treat HFT entry as exogenous:
  - *Any* $\tau > 0$
  - Eliminates sniping.
  - Reduces the cost of liquidity by $\lambda_{\text{invest}} \frac{s^*}{2}$ (or multi-unit generalization of this).
  - I interpret “any” as $\tau$ long enough to enable genuine batch processing. (Likely range: 100 microseconds to 1 millisecond).

- If treat HFT entry as endogenous:
  - Sufficiently long $\tau > 0$.
  - Eliminates sniping (as above).
  - Reduces the cost of liquidity (as above).
  - Stops the arms race.
  - Rough calibration, given scale of modern speed race, suggests $\tau$ in range of 1ms to 100ms is sufficiently long.
Computational Benefits of Discrete Time

- It is also worth briefly mentioning the 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).
Alternative Responses to the HFT Arms Race

- Numerous alternative responses: mostly address symptoms, not root cause
- “Bans” on HFT
  - Message ratios, minimum resting times
  - Misunderstand cause and effect
- Taxes on HFT
  - Transaction tax directionally addresses sniping but is a blunt instrument (in our data: 10bps to reduce 90%).
  - Cancellation tax seems misguided, liquidity provision naturally requires cancellations as prices move
- IEX speed bump
  - Displayed (“lit”) part of market: no effect on sniping. Just adds 0.00035 seconds to all the race times.
  - Non-displayed (“dark”) part of market: eliminates sniping for orders “pegged” to prices elsewhere. New Zealand joke.
Summary of Budish Cramton and Shim

- We look 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

- 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
Response to BCS
FBA paper released publicly in July 2013, pretty quickly took on a life of its own

Attention from academia but also industry, policy, press

My approach (not sure if optimal, ex ante or ex post)

- Invest significant time beyond academia:
  - “Shoe leather”. Talks/meetings with numerous HFTs, Exchanges, Banks, Industry Groups, Regulatory Bodies. Essentially, any credible stakeholder who expressed sincere interest in the work.
  - Industry conferences. (“In the lion’s den”)

- Invest significant time learning institutional details
- No financial ties. I have tried to be an independent objective voice in a messy and charged debate (“good vs. evil”)
- Responsive to press, but didn’t do blog, twitter, op-ed, etc.
- Engage in formal policy process via comment letters
“We [should] carefully consider a proposal that I like very much. It was put forward by economists at the University of Chicago School of Business — not an enemy of free markets, the University of Chicago School of Business, by any means.”

“...fundamentally reorient the markets in a very simple way that would help restore confidence in them. Their proposals would reaffirm the basic concept that the best price – not the highest speed – should win.”

“Currently, on our exchanges, securities are traded continuously ... emphasizing speed over price.”

“The University of Chicago proposal – which I endorse – would, in effect, put a speed bump in place. Orders would be processed in batches after short intervals ... ensure that the price would be the deciding factor in who obtains a trade, not who has the fastest supercomputer and early access to market-moving information.”

“This structural reform – sometimes called “frequent batch auctions” – would help catch and cap the supercomputer arms race now underway. This is tremendously important ...”
“Today’s stock market is falling short. A wasteful arms race among high-frequency traders, the growth of dark pools (private trading venues) and assorted conflicts of interest have undermined its performance. If investors don’t trust the market, that hurts capital formation, not to mention retirement and college savings.”

“Fixing the problems will require more than a tweak here and there. One idea that’s winning converts would replace the 24-hour, continuous trading of stocks with frequent auctions at regular intervals. Why would that help? Because it would lessen the emphasis on speed and direct more attention to the price that investors are willing to pay for stocks, given the prospects of the companies concerned, their industries and the broader economy. The high-speed arms race would subside, because shaving another millisecond off the time it takes to trade would confer no benefit.”

“Goldman Sachs Group Inc., among others, is interested enough in frequent batch auctions that it’s working with Budish to find an exchange that will conduct a pilot program and a regulatory agency that will monitor the results. Mary Jo White, the Securities and Exchange Commission chair, indicated in a June 5 speech her interest in batch auctions. She should make it a priority to conduct a test program. It’s a promising idea.”
Happy to find something about which I agree with my colleague John Cochrane. One second is good enough liquidity.

johnhcochrane.blogspot.com/2014/02/budish
“I wasn’t expecting to say this, but you’re actually not a communist.” – Cliff Asness, Founder of AQR
Most Common Question: Private vs. Regulatory

- Can FBA get off the ground through private market forces alone, or is a regulatory intervention required?
- Will the market fix the market?
“Will the Market Fix the Market? A Theory of Stock Exchange Competition and Innovation”

Eric Budish
Robin Lee
John Shim
We must consider, for example, whether the increasingly expensive search for speed has passed the point of diminishing returns. I am personally wary of prescriptive regulation that attempts to identify an optimal trading speed, but I am receptive to more flexible, competitive solutions that could be adopted by trading venues. These could include frequent batch auctions or other mechanisms designed to minimize speed advantages.

A key question is whether trading venues have sufficient opportunity and flexibility to innovate successfully with initiatives that seek to deemphasize speed as a key to trading success in order to further serve the interests of investors. If not, we must reconsider the SEC rules and market practices that stand in the way.
SEC Chair Mary Jo White, “Enhancing Our Equity Market Structure”

We must consider, for example, whether the increasingly expensive search for speed has passed the point of diminishing returns. I am personally wary of prescriptive regulation that attempts to identify an optimal trading speed, but I am receptive to more flexible, competitive solutions that could be adopted by trading venues. These could include frequent batch auctions or other mechanisms designed to minimize speed advantages.

... 

A key question is whether trading venues have sufficient opportunity and flexibility to innovate successfully with initiatives that seek to deemphasize speed as a key to trading success in order to further serve the interests of investors. If not, we must reconsider the SEC rules and market practices that stand in the way.
SEC Chair Mary Jo White, “Enhancing Our Equity Market Structure”

We must consider, for example, whether the increasingly expensive search for speed has passed the point of diminishing returns. I am personally wary of prescriptive regulation that attempts to identify an optimal trading speed, but I am receptive to more flexible, competitive solutions that could be adopted by trading venues. These could include frequent batch auctions or other mechanisms designed to minimize speed advantages.

...

A key question is whether trading venues have sufficient opportunity and flexibility to innovate successfully with initiatives that seek to deemphasize speed as a key to trading success in order to further serve the interests of investors. If not, we must reconsider the SEC rules and market practices that stand in the way.
SEC Chair Mary Jo White, “Enhancing Our Equity Market Structure”

We must consider, for example, whether the increasingly expensive search for speed has passed the point of diminishing returns. I am personally wary of prescriptive regulation that attempts to identify an optimal trading speed, but I am receptive to more flexible, competitive solutions that could be adopted by trading venues. These could include frequent batch auctions or other mechanisms designed to minimize speed advantages.

... 

A key question is whether trading venues have sufficient opportunity and flexibility to innovate successfully with initiatives that seek to deemphasize speed as a key to trading success in order to further serve the interests of investors. If not, we must reconsider the SEC rules and market practices that stand in the way.
SEC Chair Mary Jo White, “Enhancing Our Equity Market Structure”

We must consider, for example, whether the increasingly expensive search for speed has passed the point of diminishing returns. I am personally wary of prescriptive regulation that attempts to identify an optimal trading speed, but I am receptive to more flexible, competitive solutions that could be adopted by trading venues. These could include frequent batch auctions or other mechanisms designed to minimize speed advantages.

... 

A key question is whether trading venues have sufficient opportunity and flexibility to innovate successfully with initiatives that seek to deemphasize speed as a key to trading success in order to further serve the interests of investors. If not, we must reconsider the SEC rules and market practices that stand in the way.
Private vs. Social Returns to Market Design Innovation

- Implicit presumption (?): *the market will fix the market.*
  - Natural instinct: if the current design is sufficiently inefficient, then surely there will be private incentive to fix the inefficiency.
  - Standard case when private and social returns to innovation and technology diffusion are aligned (e.g., Griliches, 1959).
  - But private incentives often ≠ social incentives
This paper: use theory and data to analyze whether the market will fix the market

1. We first build a simple new model of stock exchange competition that respects key institutional features. Goal is to understand the status quo.

2. We then use a variety of data to “validate” the model empirically — both trades-and-quotes data and financial filings

3. Last, we study the incentives for market design innovation through the lens of our model

For today’s talk, I’ll focus mostly on Part 3.
Exchange Competition Game

Model:

- Starting point: BCS model of continuous trading
- Generalize from a single passive exchange to $M \geq 2$ exchanges who are strategic players in the game
- Assume shares are perfectly divisible: this allows investors to buy “1” unit of $x$ across multiple exchanges
- Initially: exchanges all use status quo market design and strategically choose:
  1. Trading fees $f$
     - Fee is per share traded, paid by each side of the trade.
     - (Discuss more complex fee schedules in paper)
  2. Exchange-specific speed technology price $F$
     - Each exchange has the unique ability to sell co-location near its own servers, and proprietary fast data feed from its servers.
     - TFs who pay both $c_{\text{speed}}$ and $F$ are faster than TFs who don’t.
- Later: exchanges also strategically choose market design.
Key Institutional Details

Two key regulations shape stock exchange competition in the US

- **Unlisted Trading Privileges (UTP)**
  - Any stock can be bought or sold on any exchange.
  - Model: same asset x trades on all $M$ exchanges, fungible

- **Regulation National Market System (Reg NMS, 2005/07)**
  - Order Protection Rule: roughly, on an order-by-order basis, transaction must execute at the exchange(s) with the best quote.
  - Dissemination and Access Rules: exchanges must make quotes easily electronically accessible.
  - Model: frictionless search and multi-homing. On an order-by-order basis, it is zero-cost and mandatory to check quotes on all exchanges.

- For purpose of today’s talk: assume that FBA is allowed under Reg NMS. (June 2016 “De Minimis” ruling).
UTP + Reg NMS → Virtual Single Platform → Perfectly Competitive Trading Fees

![Diagram showing investor connections and exchange revenue with perfect competition on fees.]
Description of Equilibrium

Key economic points:

1. Virtual single platform.
   - Market shares coordinate behavior, via depth-volume relationship
   - Bid-ask spreads equivalent across all exchanges

2. Trading fees are perfectly competitive.

3. Exchanges have market power on Colo/Data fees.


- All borne out in the data.
Empirical Validation: Virtual Single Platform

Depth vs. Volume

Exchanges at best price
## Empirical Validation: Competitive Trading Fees

### Trading Fees by Exchange

<table>
<thead>
<tr>
<th>Exchange</th>
<th>Tape</th>
<th>Min User</th>
<th>Max User</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Take Fee</td>
<td>Make Fee</td>
<td>(\frac{\text{Take+Make}}{2})</td>
</tr>
<tr>
<td><strong>BATS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bats BZX</td>
<td>Tape A</td>
<td>0.0030</td>
<td>-0.0025</td>
</tr>
<tr>
<td>Bats BZX</td>
<td>Tape B</td>
<td>0.0030</td>
<td>-0.0025</td>
</tr>
<tr>
<td>Bats BZX</td>
<td>Tape C</td>
<td>0.0030</td>
<td>-0.0025</td>
</tr>
</tbody>
</table>

*Average fee on all BATS-operated exchanges (2015): \$0.000090*

| **NASDAQ** |            |          |          |            |          |          |
| Nasdaq     | Tape A     | 0.0030   | -0.0020  | 0.000500   | 0.0030   | -0.0032  | -0.000075 |
| Nasdaq     | Tape B     | 0.0030   | -0.0020  | 0.000500   | 0.0030   | -0.0032  | -0.000075 |
| Nasdaq     | Tape C     | 0.0030   | -0.0015  | 0.000750   | 0.0030   | -0.0031  | -0.000025 |

*Average fee on all Nasdaq-operated exchanges (2015): \$0.000112*

| **NYSE**   |            |          |          |            |          |          |
| NYSE       | Tape A     | 0.0030   | -0.0014  | 0.000800   | 0.0027   | -0.0022  | 0.000275  |
| NYSE Arca  | Tape A     | 0.0030   | -0.0020  | 0.000500   | 0.0030   | -0.0031  | -0.000050 |
| NYSE Arca  | Tape B     | 0.0030   | -0.0020  | 0.000500   | 0.0029   | -0.0023  | 0.000275  |
| NYSE Arca  | Tape C     | 0.0030   | -0.0020  | 0.000500   | 0.0030   | -0.0032  | -0.000100 |

*Average fee on all NYSE-operated exchanges (2015): \$0.000115*

**Note:** Average fee for standard, regular hours trading in US equities.
Empirical Validation: Profitability of Colo / Data

Exchange Revenue Breakdown

**BATS US equities**
- Port Fees: 24.8%
- Transaction & Clearing: 31.2%
- Market Data: 44.0%

**CME**
- Transaction & Clearing: 83.7%
- Market Data: 12.0%
- Access Other: 2.6%
Empirical Validation: Profitability of Colo / Data

- Exchanges make significant money from colo/data:
  - BATS exchange family, for which data is cleanest
    - 68.8% of US equities revenue from Colo and Data: $178.4M
    - 2015 US equities operating income, overall: $149.2M
    - 2015 US equities operating income, w/o Colo/Data: -$29.2M
  - NYSE/NASDAQ similar, but data less clean
  - Estimate NYSE + NASDAQ + BATS US equities colo and data revenue is around $1bn per year

- Colo/data significant proportion of HFT expenses
  - KCG/Getco: >60% of pre-communications trading profits, each year 2012-2015

- (Futures exchanges very different: CME earns 83.7% of revenue from trading and clearing, 14.6% from Colo and Data)
Empirical Validation: Profitability of Colo / Data

- Nasdaq connectivity offerings for co-located customers (included in IEX comment letter):

<table>
<thead>
<tr>
<th>CONNECTIVITY TYPE</th>
<th>NRC</th>
<th>MRC</th>
<th>LATENCY IMPROVEMENT*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1G ULTRA</td>
<td>$1,500</td>
<td>$2,500</td>
<td>5-12 μs vs. 1G</td>
</tr>
<tr>
<td>10G</td>
<td>$1,000</td>
<td>$10,000</td>
<td>8-15 μs vs. 1G</td>
</tr>
<tr>
<td>10G ULTRA</td>
<td>$1,500</td>
<td>$15,000</td>
<td>5-9 μs vs. 10G</td>
</tr>
<tr>
<td>40G</td>
<td>$1,500</td>
<td>$20,000</td>
<td>3-8 μs vs. 10G</td>
</tr>
</tbody>
</table>

*Median roundtrip latency measured from order to acknowledgement for the NASDAQ Stock Market™ using Corvil Net
Empirical Validation: Money Pump Constraint
Trading Fees by Exchange

<table>
<thead>
<tr>
<th>Exchange</th>
<th>Tape</th>
<th>Min User</th>
<th>Max User</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Take Fee</td>
<td>Make Fee</td>
</tr>
<tr>
<td>BATS</td>
<td></td>
<td>Take Fee</td>
<td>Make Fee</td>
</tr>
<tr>
<td>Bats BZX</td>
<td>Tape A</td>
<td>0.0030</td>
<td>-0.0025</td>
</tr>
<tr>
<td>Bats BZX</td>
<td>Tape B</td>
<td>0.0030</td>
<td>-0.0025</td>
</tr>
<tr>
<td>Bats BZX</td>
<td>Tape C</td>
<td>0.0030</td>
<td>-0.0025</td>
</tr>
</tbody>
</table>

Average fee on all BATS-operated exchanges (2015): $0.000090

<table>
<thead>
<tr>
<th>NASDAQ</th>
<th></th>
<th>Take Fee</th>
<th>Make Fee</th>
<th>Take+Make</th>
<th>Take Fee</th>
<th>Make Fee</th>
<th>Take+Make</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasdaq</td>
<td>Tape A</td>
<td>0.0030</td>
<td>-0.0020</td>
<td>0.000500</td>
<td>0.0030</td>
<td>-0.0032</td>
<td>-0.000075</td>
</tr>
<tr>
<td>Nasdaq</td>
<td>Tape B</td>
<td>0.0030</td>
<td>-0.0020</td>
<td>0.000500</td>
<td>0.0030</td>
<td>-0.0032</td>
<td>-0.000075</td>
</tr>
<tr>
<td>Nasdaq</td>
<td>Tape C</td>
<td>0.0030</td>
<td>-0.0015</td>
<td>0.000750</td>
<td>0.0030</td>
<td>-0.0031</td>
<td>-0.000025</td>
</tr>
</tbody>
</table>

Average fee on all Nasdaq-operated exchanges (2015): $0.000112

<table>
<thead>
<tr>
<th>NYSE</th>
<th></th>
<th>Take Fee</th>
<th>Make Fee</th>
<th>Take+Make</th>
<th>Take Fee</th>
<th>Make Fee</th>
<th>Take+Make</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYSE</td>
<td>Tape A</td>
<td>0.0030</td>
<td>-0.0014</td>
<td>0.000800</td>
<td>0.0027</td>
<td>-0.0022</td>
<td>0.000275</td>
</tr>
<tr>
<td>NYSE Arca Tape A</td>
<td>0.0030</td>
<td>-0.0020</td>
<td>0.000500</td>
<td>0.0030</td>
<td>-0.0031</td>
<td>-0.000050</td>
<td></td>
</tr>
<tr>
<td>NYSE Arca Tape B</td>
<td>0.0030</td>
<td>-0.0020</td>
<td>0.000500</td>
<td>0.0029</td>
<td>-0.0023</td>
<td>0.000275</td>
<td></td>
</tr>
<tr>
<td>NYSE Arca Tape C</td>
<td>0.0030</td>
<td>-0.0020</td>
<td>0.000500</td>
<td>0.0030</td>
<td>-0.0032</td>
<td>-0.000100</td>
<td></td>
</tr>
</tbody>
</table>

Average fee on all NYSE-operated exchanges (2015): $0.000115

Note: Average fee for standard, regular hours trading in US equities.
Incentives for Market Design Innovation

- Overall, the data suggests that the simple model is sensible.
- So now let’s study the incentives for market design innovation.
- First, the good news...
What Happens if FBA Enters

- Suppose an FBA exchange enters, de novo, and charges a trading fee of zero. All other exchanges are fixed at Continuous, also at $f = 0$. 

Reasonable prior: coordination problem, multiple equilibria.

But in fact, a unique equilibrium: Discrete gets 100% share.

Intuition:

- TFs strictly prefer liquidity on Discrete over Continuous, to avoid cost of getting sniped, provided that investors notice.

- Zero search and multi-homing costs, via Reg NMS, gets investors to notice.

- Essentially: two otherwise identical markets, one with a tax and one without, the one without the tax wins if there are zero frictions.

(N.B. 100% should not be taken literally)

Argument works if Discrete charges any fee smaller than the sniping savings. Discrete gets compensated for eliminating sniping.
What Happens if FBA Enters

- Suppose an FBA exchange enters, de novo, and charges a trading fee of zero. All other exchanges are fixed at Continuous, also at $f = 0$.
- Reasonable prior: coordination problem, multiple equilibria.
What Happens if FBA Enters

- Suppose an FBA exchange enters, de novo, and charges a trading fee of zero. All other exchanges are fixed at Continuous, also at $f = 0$.
- Reasonable prior: coordination problem, multiple equilibria.
- But in fact, a unique equilibrium: Discrete gets 100% share.

Intuition:

- TFs strictly prefer to offer liquidity on Discrete over Continuous, to avoid cost of getting sniped, provided that Investors notice
- Zero search and multi-homing costs, via Reg NMS, gets Investors to notice.
- Essentially: two otherwise identical markets, one with a tax and one without, the one without the tax wins if there are zero frictions.
- (N.B. 100% should not be taken literally)

Argument works if Discrete charges any fee smaller than the sniping savings. Discrete gets compensated for eliminating sniping.
But ... Incumbent Response → Bertrand Trap

- How will incumbents respond if Discrete enters?
- If initial Discrete enters at fee $0 < f_1 < \text{sniping} \ldots$ then an incumbent will switch to Discrete with fee $f_2 < f_1 \ldots$ fees get competed down to zero. Same perfect competition on trading fees as above.
- Entrant’s profits are zero
- Classic “Bertrand trap” prevents innovator from capturing the social value of innovation
- First source of tension between private and social incentives to innovate
Now suppose an Incumbent exchange adopts FBA

As before:
- Unique equilibrium is Discrete market design gets 100% share
- Bertrand trap

Difference: incumbent’s profits are zero under Discrete, but were positive under Continuous
- Incumbents lose rents from speed technology

Second source of tension between private and social incentives to innovate

Can model the game among incumbents as a repeated prisoners’ dilemma, where status quo is an equilibrium among incumbents.
Chief Economist of NASDAQ, at a public academic event in Nov 2013 when asked whether NASDAQ would consider trying frequent batch auctions:

Technologically, we could do it. **The big issue**, one of the big issues for us, when I talked about cost, the cost we would bear, **would be getting [the SEC] to approve it**, which would take a lot of time and effort, and if we got it approved, it would **immediately be copied by everybody else**. . . . So we would have essentially no first-mover advantage if we put it in there, **we would have no incentive to go through the lift of creating [the new market design]**.
Summary of Analysis

- Summary:
  - If FBA enters, the new market design wins significant share (100% in stylized model)
  - But Bertrand competition on fees implies de novo entrant profits = 0
  - Incumbent profits < 0 (because lose rents from speed race)
  - Doesn’t look so good for the market fixing the market!
  - Optimistic spin: 0 ... is pretty close to positive?
  - Prediction: if an FBA enters, it is most likely to come from
    - de novo entrant (e.g., IEX)
    - incumbent with low share (e.g., CHX)
    - futures exchange, where Bertrand trap not an issue because no fungibility (e.g., CME)
  - Least likely: incumbent stock exchange with large revenues from speed race
  - Robust conclusion: private incentives <<<< social incentives
Regulatory Response

- Analysis suggests a potential role for regulatory intervention
  1. Well-documented market failure in status quo of the market
  2. Model suggests that private-market incentives alone might not fix the problem

- Model suggests the following would facilitate the market fixing the market:
  1. Reduce adoption costs for first entrant (e.g., proactively clarify that FBA is allowed)
  2. Reduce tick-size constraints, so that if there is a first entrant, market can tip
  3. Provide modest exclusivity period, or some other “push” to provide incentive to first entrant

- A more direct response: put time into discrete units
  - Unit could be guided by physical limits of computers & communications
  - Orders received at the same discrete time would have the same time priority
Political Economy of Regulation

- Political economy of addressing sniping isn’t great:
  1. Concentrated harm / dispersed benefits
  3. Nuance. HFT has both positive and negative aspects.
  4. Regulatory risk-reward not great.
  5. Chicken-and-egg problem
    - Regulators would like empirical proof that FBA works, but this requires someone to try it first
    - Even a pilot test is a major undertaking (no real progress here)
    - One bit of progress: Financial Conduct Authority in UK gathering new kinds of data to permit evaluation

- So, in addition to friction in private incentives to adopt, there are potential frictions in the regulatory response
Conclusion
Role of Academics: Theory → Practice

- Well identified market failure. Strikes at the core of how we think efficient markets are supposed to work.
Role of Academics: Theory → Practice

- Well identified market failure. Strikes at the core of how we think efficient markets are supposed to work.
- Causes an industrial arms race, harms investors, and makes markets unnecessarily complex. Likely >$100bn NPV
- Solution reasonably simple.
Role of Academics: Theory → Practice

- Well identified market failure. Strikes at the core of how we think efficient markets are supposed to work.
- Causes an industrial arms race, harms investors, and makes markets unnecessarily complex. Likely >$100bn NPV
- Solution reasonably simple.
- The case for market forces fixing the market failure isn’t great.
Role of Academics: Theory → Practice

- Well identified market failure. Strikes at the core of how we think efficient markets are supposed to work.
- Causes an industrial arms race, harms investors, and makes markets unnecessarily complex. Likely >$100bn NPV
- Solution reasonably simple.
- The case for market forces fixing the market failure isn’t great.
- And the political economy for the regulator to fix the market also isn’t great.
Role of Academics: Theory → Practice

- Well identified market failure. Strikes at the core of how we think efficient markets are supposed to work.
- Causes an industrial arms race, harms investors, and makes markets unnecessarily complex. Likely >$100bn NPV
- Solution reasonably simple.
- The case for market forces fixing the market failure isn’t great.
- And the political economy for the regulator to fix the market also isn’t great.

- So a question I think about a lot is what to do about this
- 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
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 → Practice

- My second thought is more of a question, which is whether there is more that we can do as a profession to help bring our good ideas from theory to 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

- 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 ... ”
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
How do you use Metamako?

“Metamako makes it easier to roll out and maintain complex network set ups. We are now able to more aggressively pursue latency goals.”

— Network Engineer, Financial Service Co.
Metamako makes it easier to roll out and maintain complex network set ups. We are now able to more aggressively pursue latency goals.

— Network Engineer, Financial Service Co.
But they shouldn’t :)