Financial Markets Microstructure
Lecture 16

Algo-trading, High-Frequency Trading, and Blockchain

Egor Starkov

Københavns Unversitet

Spring 2025

Previously on FMM

Corporate governance has a lot of connection to company's financial market performance

- access to capital affected by liq-ty
- liq-ty and corporate control are somewhat antithetical
- firm can use stock price as market's feedback on its decisions or as benchmark of CEO performance
- firms have some ways in which they can improve the liquidity of their stocks

Today on FMM...

- Algo-trading!
- High-frequency trading!
- Cryptocurrencies!
- and more...

Digital Markets

"...It should come as no surprise then that the financial system exhibits a Moore's Law of its own – from 1929 to 2009 the total market capitalization of the US stock market has doubled every decade. The total trading volume of stocks in the Dow Jones Industrial Average doubled every 7.5 years during this period, but in the most recent decade, the pace has accelerated: now the doubling occurs every 2.9 years, growing almost as fast as the semiconductor industry."

Kirilenko and Lo [2013]

Digital Markets

- The digital revolution of the past few decades has reshaped financial markets as much as (if not more than) any other aspect of our lives
- The quote above mentions the "extensive margin" akin to the Moore's Law
- But the "intensive margin" is also at work
  - Index funds, automated arbitrage, automated execution & market-making only made possible by computers
- In addition to Moore's Law, Murphy's Law does not fail either
  - If something can go wrong it will, and the scope for failures is as big as ever these days. See Kirilenko and Lo [2013] (pp.60-67) for five stories.

More on algo-trading

- Algorithms allow for a lot of stuff:
  - HFT (later today)
  - better hedging though some automated hedges
  - but also for better execution via order-splitting.
- Beason and Wahal [2019] give some (actually a lot of) info on how algorithms work for large institutional investors (a typical counterpart to HFT nowadays)
  - I also put on absalon a 2020 SEC report on algo-trading that gives a bird's-eye overview
- "Parent" orders are split (by algorithms) into many "child" orders that are routed to markets

Institutional algo-trading

- The average parent order attempts to trade \$287,000 over 84 minutes, equivalent to 4.80 percent of volume over the duration of the order.
  - avg: 63.1 runs per parent (avg 10m total duration), 8.8 children per run
- Of the 300 million child orders, less than 0.40 percent are market orders.
  - By comparison, retail investors usage of market orders is over 50 percent
  - $\approx$ 80% are limit orders, $\approx$ 20% are PEG orders – dark limit orders that are dynamically "pegged" to the NBBO
  - Of the limit orders, 24% are marketable, 65% are passive, rest inside the spread
- Many orders are unfilled (even marketable)
  - Conditional on filled, median time-to-trade=5ms
  - Even unfilled orders have price impact

High-frequency trading: introduction

- **HFT**: Refers to computerized, algorithmic trading at high pace: fastest participants take advantage of opportunities before others
- **Speed is key**: for instance, in 2010, a USD 300 million cable was laid between Chicago and New Jersey (Nasdaq)
- **Ubiquitous**: estimated to account for more than 50% of volume in the US and more than 25% in Europe
- **Recent phenomenon(?)**: the effect on markets is still not well understood. Few empirical studies and fewer theoretical models
- **Today**: Look at two models of HFT

Biais, Foucault, and Moinas [2015]

- Simple model of fast trading and investment in speed
- Look at equilibrium behavior and welfare implications
- Endogenize the choice of whether to be fast or slow
  - optimal decision depends on size of trader

Model: basics

- **Institution**: continuum of profit-maximizing financial institutions indexed by i, zero endowment, trade one unit
- **Time**: $\tau \in \{0, 1, 2\}$
- **Asset value**: $u_i = v + y_i$, where v is the fundamental value and y the institution's private value
  - **Fundamental value**: $v \in \{\mu - \epsilon, \mu + \epsilon\}$, equal probability, realized at $\tau = 2$
  - **Private value**: $y_i \in \{\delta, -\delta\}$, equal probability and i.i.d. across investors, observed at $\tau = 1$
- **Trading**: Occurs at $\tau = 1$ after private values are learned

Model: high-frequency trading

- **HFT**: Fraction α of the institutions invest at $\tau = 0$ to become HFT
- **Information**: Let's call HFT *fast* institutions (viz. *slow* institutions)
  - Fast institutions have better information: learn v at $\tau = 1$ whereas slow institutions learn v at $\tau = 2$
  - Fast institutions find a trading opportunity with probability one, slow institutions with probability $\rho < 1$
- **Timing** (within period $\tau = 1$):
  1. Each institution i observes y_i, and if fast, observes v
  2. Each institution i finds a trading opportunity or not. If yes, chooses whether to buy/sell/abstain (one-unit trades only): $d_i \in \{-1, 0, 1\}$.
  3. Liquidity providers execute order d_i at price $\mathbb{E}(v|d_i)$ (implicit assumption of market maker competition + no aggregate order flow transparency)

Information

- **Fundamental value:** Good news/bad news:
  - Good news refer to high value: $v = \mu + \epsilon$.
  - Bad news refer to low value: $v = \mu - \epsilon$.
- **Fast institutions** (FI) have the following types
  - *GH*: Good news, high private valuation
  - *GL*: Good news, low private valuation
  - *BH*: Bad news, high private valuation
  - *BL*: Bad news, low private valuation
- **Slow institutions** (SI), on the other hand, are either
  - *H*: high private valuation
  - *L*: low private valuation

Equilibrium analysis

- **No fast trading**: If $\alpha = 0$ all orders execute at μ
- **Active fast trading**: Now suppose $\alpha > 0$. Let

$$\beta_j^F : \text{prob. that fast institution type } j \text{ buys}$$
$$\beta_j^S : \text{prob. that slow institution type } j \text{ buys (cond. on trading opp-ty)}$$

- **High value**: Fast GH types have highest possible valuation: $\beta_{GH}^F = 1$
- **Low value**: Fast BL types have lowest possible valuation: $\beta_{BL}^G = 0$
- **Buy side**: Let $a = \mathbb{E}[v|buy]$. Use above observation and Bayes' Rule to get

$$a = \mu + \frac{\alpha\frac{1+\beta_{GL}^F-\beta_{BH}^F}{4}}{(1-\alpha)\rho\frac{\beta_H^S+\beta_L^S}{2}+\alpha\frac{1+\beta_{GL}^F+\beta_{BH}^F}{4}}\epsilon.$$

Multiple equilibria

- **Multiple equilibria**: Often markets have several equilibria.
- **Self-fulfilling expectations**: This is caused by the endogenous price:
  - If you think buyers will have high valuations $\to$ set high a
  - If a is high, only traders with high valuations will buy
- **Assumption**: $\frac{\epsilon}{2} < \delta < \epsilon$: both v and y matter; v more so. Then

$$V_{GH}^F > V_H^S > V_{GL}^F > \mu > V_{BH}^F > V_L^S > V_{BL}^F,$$

  where V_j^i is the value of a type-i institution with type-j information.
- **Equilibrium types**: We focus here on the pure strategy equilibria. There will be three types of equilibria: P1, P2 and P3.

- **P1**: $\mu \leq a < \mu + \epsilon - \delta$. Fast institutions with good news always buy, slow institutions buy if private value high:

$$\beta^F_{GL} = \beta^S_H = 1 \text{ and } \beta^F_{BH} = \beta^S_L = 0$$

- **P2**: $\mu + \epsilon - \delta < a < \mu + \delta$. Fast institutions with good news and high value buy, but don't trade if information is conflicting. Slow institutions buy if private value high:

$$\beta^F_{GL} = \beta^S_L = \beta^F_{BH} = 0 \text{ and } \beta^S_H = 1$$

- **P3**: $a = \mu + \epsilon$. Fast institutions with good news and high value buy, other types of institutions don't trade (*crowding out*):

$$\beta^F_{GL} = \beta^S_H = \beta^F_{BH} = \beta^S_L = 0$$

- Note: as usual, we look on one side of mkt so "sell" same as "abstain" in betas above

When do we have multiple equilibria?

- P3 equilibrium: Always exists. If dealers believe only fast institutions trade $\rightarrow$ high a. But then only optimal for fast institutions to trade.
- **Proposition**: P1 equilibrium exists if $\alpha < \alpha_{P1} \equiv \frac{\rho(\epsilon-\delta)}{\rho(\epsilon-\delta)+\delta}$. Proof.
  1. Suppose institutions expect $a = \mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon$
  2. Notice $\mu - \epsilon + \delta < \mu < a$: FI never buy with bad news ($\beta^F_{BH} = 0$ optimal)
  3. $\alpha < \alpha_{P1} \Rightarrow a < \mu + \epsilon - \delta$: good news imply expected FI gains from buying, regardless of private valuation ($\beta^F_{GL} = 1$ is optimal).
  4. Notice $\mu + \delta > \mu + \epsilon/2 > \mu + \epsilon - \delta > a$: SI with high valuation always buys ($\beta^S_H = 1$ is optimal).
  5. Conditional on this, $\mathbb{E}[v|buy] = \mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon$ (a is optimal price). $\square$
- Similarly, can find values of α s.t. P2 equilibrium exists
- P1 is Pareto dominant for $\alpha < \alpha_{P1}$.

Institution gain (Fast Institutions)

- FI gain in P1 equilibrium is (focus on buy side):

$$\begin{aligned}
\mathbb{E}[u-a|buy,FI] &= \mathbb{E}[u|buy,FI] - \left[\mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon\right] \\
&= \mathbb{E}[u|v=\mu+\epsilon] - \left[\mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon\right] \\
&= \mu+\epsilon - \left[\mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon\right] \\
&= \frac{(1-\alpha)\rho}{\alpha+(1-\alpha)\rho}\epsilon \equiv \pi_F(\alpha)
\end{aligned}$$

- Notice $\pi'_F(\alpha) < 0$.

Institution gain (Slow Institutions)

- SI gain in $P1$ equilibrium is (focus on buy side):

$$\begin{aligned}
\rho\mathbb{E}[u-a|buy,SI] &= \rho\left(\mathbb{E}[u|buy,SI] - \left[\mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon\right]\right) \\
&= \rho\left(\mathbb{E}[u|y_i=\delta] - \left[\mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon\right]\right) \\
&= \rho\left(\mu+\delta - \left[\mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon\right]\right) \\
&= \rho\left(\delta - \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon\right) \equiv \pi_S(\alpha)
\end{aligned}$$

- Notice $\pi'_S(\alpha) < 0$.

Institution gain

- Both $\pi_F(\alpha)$ and $\pi_S(\alpha)$ are decreasing in α: all trader types lose when proportion of fast traders increases
- This result holds generally (when we focus on the Pareto-dominant equilibria)
  - Fast institutions lose because there is more price impact of trades (more adverse selection)
  - Higher price impact dissuades slow institutions from trading (crowding out)
- So more HFT is always 'bad' for existing traders, but beneficial for institutions that switch to become HFT
  - Note: $\pi_F(\alpha) - \pi_S(\alpha) > 0$ is independent of α
- Shkilko and Sokolov [2020]: periods when HFT is disrupted are characterized by less adverse selection, lower trading costs

Endogenous acquisition of technology

- **Cost**: At $\tau = 0$, a trader can become a fast institution at cost C
- **Markets**: There is a size-N continuum of markets (this will simplify the maths). An institution of type n can participate in $n \leq N$ markets
- **Participation**: Type is distributed according to "pdf" $h(n)$ on $[0, N]$ with

$$h(n) = \frac{N}{n}$$

- **Optimal investment**: Invest in becoming fast institution if

$$\pi_F(\alpha) \cdot n - C \geq \pi_S(\alpha) \cdot n$$
$$\Leftrightarrow n \geq \frac{C}{\pi_F(\alpha) - \pi_S(\alpha)} \equiv n(\alpha)$$

Endogenous acquisition of technology (2)

- Notice that $n \cdot h(n) = N$: the total number of investments made by type-n institutions is N for all n
- Thus: equal amount of n and n' investors *within each market*.
- In other words, n is uniformly distributed within each market: $n \sim U[0, N]$ such that we get the following fixed-point problem

$$\alpha = \mathbb{P}(n \geq n(\alpha)) = \frac{N - n(\alpha)}{N}$$

Endogenous acquisition of technology (3)

- Authors find equilibrium with $\partial\alpha/\partial C < 0$.
- Welfare result:

If $\rho > 1/2$ then welfare-maximizing value of α is 0.

- Hence: in 'well-functioning' markets, equilibrium has too much HFT
- Because HFT effects are:
  - *more trading opportunities* – personal and social benefit
  - *pvt info about v* – personal benefit, social cost (worse prices for everybody)

BFM Conclusion

- Fast trading exacerbates adverse selection, but is individually appealing
- If the markets are already reasonably good at matching traders with opportunities, fast trading may be strictly bad for welfare

Budish, Cramton, and Shim [2015]

- **Claim**: there is an *arms race* in HFT (perpetual wasteful investment in gaining advantage) and this is a result of bad market design
- **Solution**: must go to the root and construct better markets rather than imposing taxes etc.
- **Proposal**: Authors propose to replace the *continuous auction* with *frequent batch auctions*
  - frequent = every 0.1s
- **Paper**: Claims are backed up with a great deal of data and a (very!) simple model

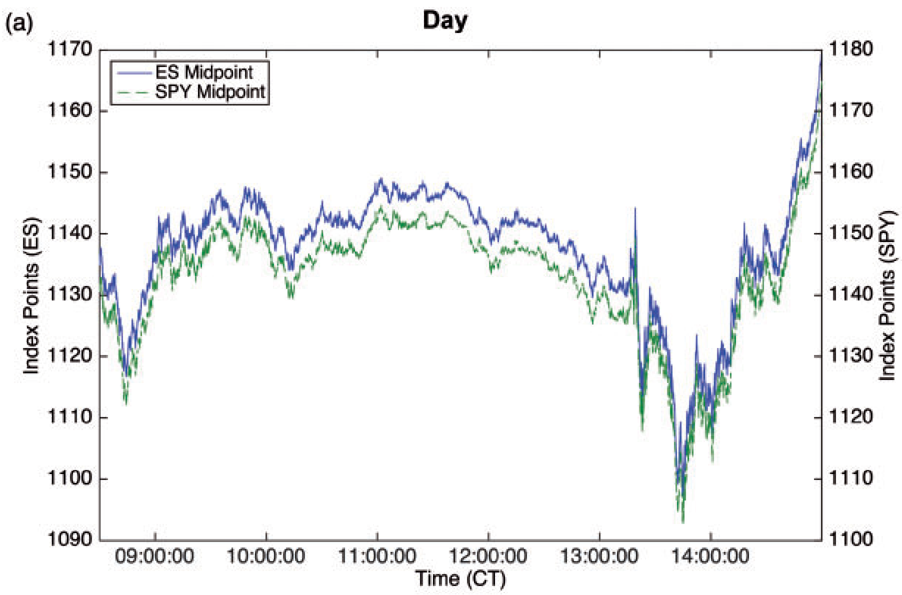
Correlations and arbitrage

The authors make three points

1. **Vanishing correlations**: For short enough latency (time intervals), correlations between almost identical assets break down
2. **Arbitrage**: This leads to arbitrage possibilities
3. **Perpetual situation**: These arbitrage possibilities do not vanish over time, suggesting that competition does not make them disappear

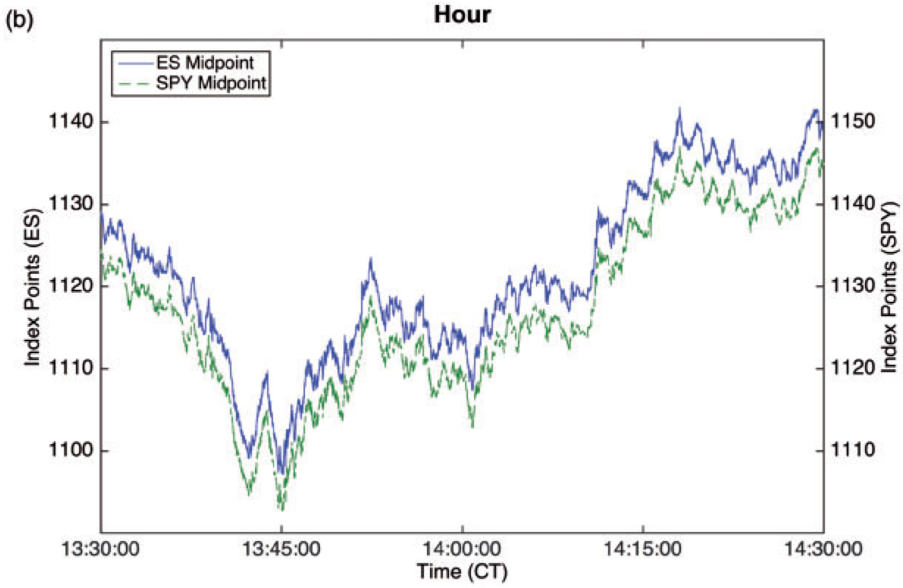
Budish, Cramton, and Shim [2015]

ES and SPY are the two largest instruments tracking S&P500. In theory perfectly correlated. Panel (a) shows a trading day.



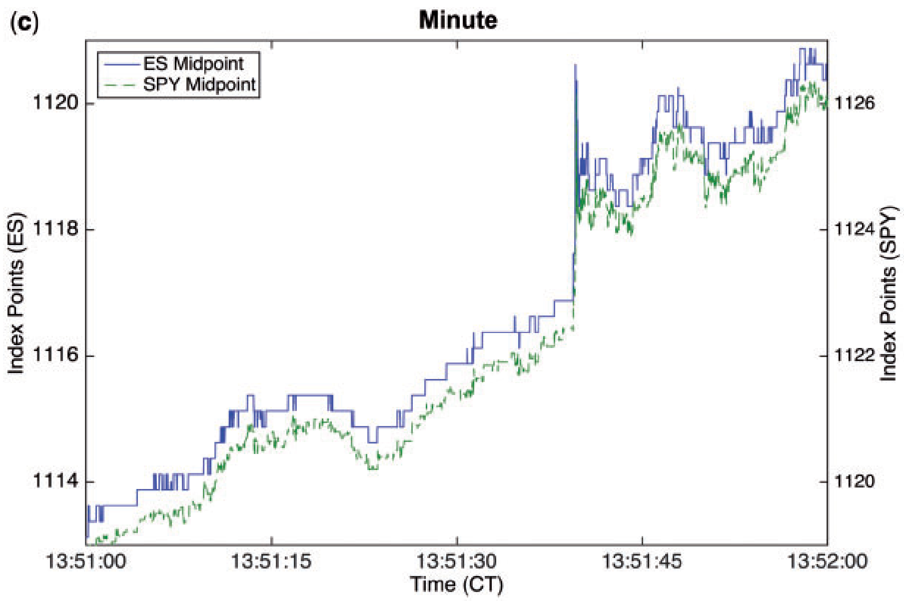
Budish, Cramton, and Shim [2015]

ES and SPY are the two largest instruments tracking S&P500. In theory perfectly correlated. Panel (b) shows a trading hour.



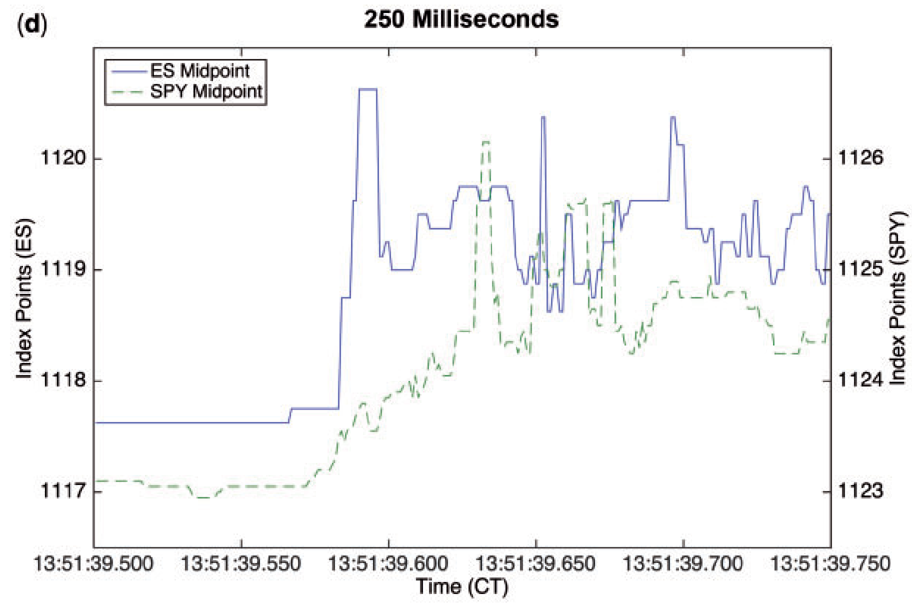
Budish, Cramton, and Shim [2015]

ES and SPY are the two largest instruments tracking S&P500. In theory perfectly correlated. Panel (c) shows a trading minute.



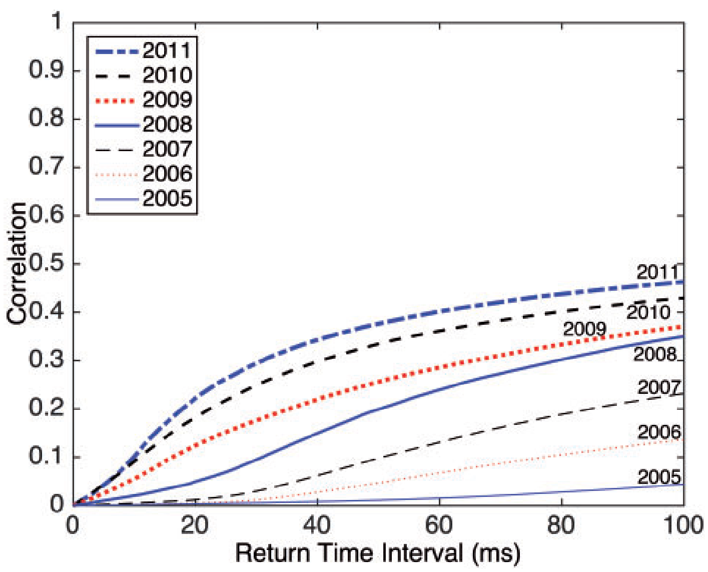
Budish, Cramton, and Shim [2015]

ES and SPY are the two largest instruments tracking S&P500. In theory perfectly correlated. Panel (d) shows a high-frequency breakdown.



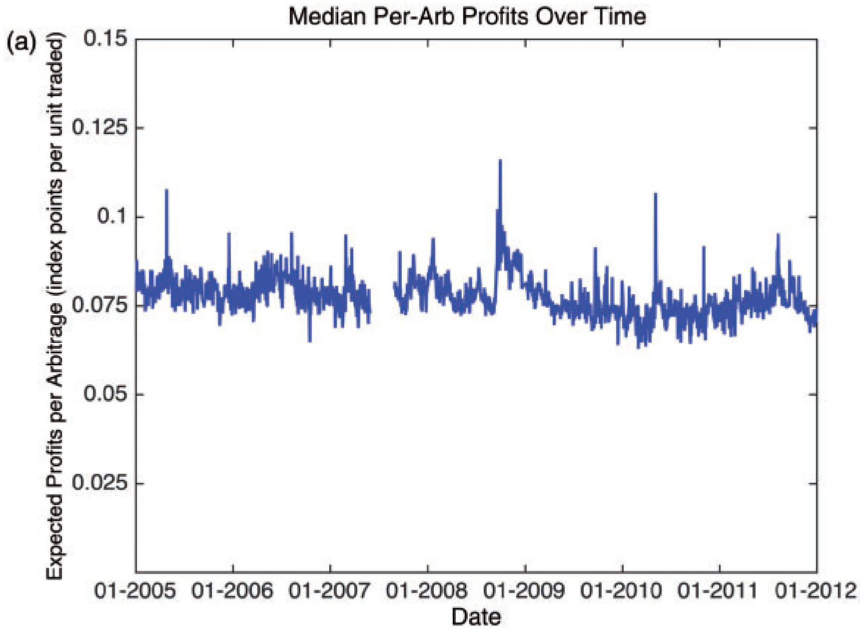
Budish, Cramton, and Shim [2015]

The figure below shows the correlation between ET and SPY by time interval for different years.



Budish, Cramton, and Shim [2015]

The figure below shows median arbitrage profits over time. Very stable, total $\sim$ \$75m/yr



Model of a continuous market

Security

- **Value**: There is a signal y, which is perfectly correlated with the value x. Signal y follows compound Poisson distribution.
- **Comp. Poisson**: Jumps arrive at rate λ_{jump} and have size $J \sim F_{jump}$.

Players

- **Noise traders**: arrive according to Poisson process (λ_{invest})
  - Want to buy/sell one unit
  - Incur cost of delay, so use marketable limit orders $\simeq$ market orders
- **HFTs**: There are N HFT firms who use market or limit order

Order processing

- If multiple orders/messages at same time, uniform random draw to determine first to be processed

Equilibrium

Equilibrium properties: Focus on equilibrium with the following properties

- **Endogenous market maker**: 1 HFT *endogenously* takes the role of liquidity provider. Refer to this as the market maker (MM).
- **Adverse selection**: $N-1$ HFTs act as *stale quote snipers*

Market maker

- **Quotes**: Suppose signal is y. Set $a = y + \frac{s}{2}$ and $b = y - \frac{s}{2}$ where s is the spread.
- **News**: If news arrive and the new signal is y', send message to cancel quotes a and b and post new ones: $a' = y' + \frac{s}{2}$ and $b' = y' - \frac{s}{2}$. Noise traders are slower at receiving news.

Snipers

- Trade if $|y' - y| > \frac{s}{2}$.

Equilibrium (2)

- **Market maker profits**: The MM flow profits (per dt period, normalized by dt) are

$$\lambda_{invest} \cdot \frac{s}{2} - \lambda_{jump} \cdot \mathbb{P}\left(J > \frac{s}{2}\right) \cdot \mathbb{E}\left[J - \frac{s}{2} \mid J > \frac{s}{2}\right] \cdot \frac{N-1}{N}.$$

- **Sniper profits**: The profits to stale-quote snipers are

$$\lambda_{jump} \cdot \mathbb{P}\left(J > \frac{s}{2}\right) \cdot \mathbb{E}\left[J - \frac{s}{2} \mid J > \frac{s}{2}\right] \cdot \frac{1}{N}.$$

- **Equilibrium condition**: Make HFT indifferent btw MM and sniper:

$$\lambda_{invest} \cdot \frac{s}{2} = \lambda_{jump} \cdot \mathbb{P}\left(J > \frac{s}{2}\right) \cdot \mathbb{E}\left[J - \frac{s}{2} \mid J > \frac{s}{2}\right].$$

- **Lack of competition**: Spread s does not depend on N.

Continuous auction market versus batch

- **Conclusion**: There will be a positive bid-ask spread in continuous market (even as $N \to \infty$), despite no asymmetric information (kind of)
- **Market failure**: Authors argue that this failure is built into the market via processing mechanism
- **Proposed solution**: Frequent batch auction
  - Auction every τ moments. Fast institutions have latency δ_{fast} and slow institutions latency δ_{slow}. Three intervals, depending on when public signal arrives:
    1. $[0, \tau - \delta_{slow}]$: all institutions trade, no AS
    2. $[\tau - \delta_{slow}, \tau - \delta_{fast}]$: only fast institutions trade, AS
    3. $[\tau - \delta_{fast}, \tau]$: no institution trade (inefficient)
- **Outcome**: Before, fast trader always has advantage; now only $\frac{\delta}{\tau}$ of the time, where $\delta = \delta_{slow} - \delta_{fast}$. If $\delta = 100$ microseconds and $\tau = 100$ milliseconds. Then $\frac{\delta}{\tau} = \frac{1}{1000}$. Large reduction in HFT importance.

HFT Conclusion

- Effects of speed are similar to those of informed trading
- By design, continuous trading generates arbitrage opportunities
- Firms overinvest in speed in attempts to reap these arbitrage profits
- Risk of being sniped contributes to the spread
- While HFTs can serve as liquidity providers, they do not actually contribute to narrowing the spread
- Use better market design (batch auctions) to improve this

Blockchain and cryptocurrencies

- Our discussion would be incomplete without mentioning blockchain and cryptocurrencies, the biggest trend of 2017
  - blockchain is a "distributed ledger" technology
  - crypto uses blockchain to record transactions in some tokens
- In addition to below, you can find some economic discussion of crypto in Nica, Piotrowska, and Schenk-Hoppé [2017] and Halaburda, Haeringer, Gans, and Gandal [2020]

How should it work?

- Cryptos (bitcoin, ethereum) are like distributed payment systems
- You can translate that to a financial market:
  - Say coins serve as shares of some company
  - Or there is a decentralized exchange that records stock ownership transactions in a blockchain

Why the hype?

- Decentralization: no exchange to profit from traders $\Rightarrow$ lower order costs
  - Even when the market is dominated by exchanges, you do not need to use them to trade (in principle)
- Transparency: transaction history is visible, order flow is visible, counterparty's trading history visible
  - note: there is very little anonymity, contrary to what some say!
- Smart contracts: algotrading by design

Why not?

- Limited processing capacity: block size and frequency is $\approx$fixed
  - Visa: 150m tx/day; Bitcoin: 300k tx/day (15.02.21)
- Order costs and execution risk: you have to bid for your transaction to be accepted into a block.
  - This is on top of execution risk from other sources (for limit orders)
  - Average order costs fluctuate over time graph
  - There are concerns that miners inflate fees (Lehar and Parlour [2020])
- Delay: blocks are only processed rarely (one per 10 min avg for bitcoin) graph
- Clearing and settlement: without a trusted mediator, counterparty and security risks instensify pic
- No transparency requirements: it is more difficult to enforce disclosure of financial info by firms/coin issuers

Revealed preference

- But in the end, the final users (traders) don't care about the fancy technology in the backend
- They choose whichever is (1) cheaper and (2) more convenient to use
- So web3 converged to the same centralized system we had before:
  - crypto is traded via a few centralized exchanges (Binance, Coinbase, ...)
  - NFTs barely exist(ed?) outside OpenSea and Rarible
- Signal founder has a nice post about it.