

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

Eric Budish, Peter Cramton and John Shim

Spring 2014 Q-Group Seminar

8 April 2014

The HFT Arms Race: Example



- ▶ 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 is already obsolete!
- ▶ Not tunnels, but microwaves (first 10ms, then 9ms, now 8.5ms).
- ▶ Analogous races occurring throughout the financial system, sometimes measured as finely as microseconds or nanoseconds
 - ▶ Last month alone: Lasers, BusinessWire

The HFT Arms Race: Market Design Perspective

- ▶ We examine the HFT arms race from the perspective of market design.
 - ▶ We assume that HFT's are optimizing with respect to market rules as they're presently given
 - ▶ But, ask whether these are the right rules
 - ▶ Avoids much of the “is HFT good or evil?” that seems to dominate the discussion of HFT
 - ▶ Instead, ask at a deeper level what is it about market design that incentivizes arms race behavior, and is this design optimal
- ▶ Central point: HFT arms race is a *symptom* of a basic flaw in modern financial market design: continuous-time trading.
- ▶ Proposal: make time discrete.
 - ▶ Replace continuous-time limit order books with *discrete-time frequent batch auctions*: uniform-price sealed-bid double auctions conducted at frequent but discrete time intervals, e.g., every 1 second or 100ms.
 - ▶ Minimal departure from CLOB that eliminates the problems caused by continuous-time trading

Frequent Batch Auctions

A simple idea: make time discrete.

1. Continuous limit-order books don't actually “work” in continuous time: market correlations completely break down; frequent technical arbitrage opportunities
2. Technical arbitrage opportunities → arms race. Arms race does reduce the duration of the market failures; arms race does not actually fix the market failures
3. Theory model: critique of the CLOB market design
 - ▶ Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)
 - ▶ Identifies the costs of the arms race
 - ▶ Harms liquidity (spreads, depth)
 - ▶ Socially wasteful
4. Frequent Batch Auctions as a market design response
 - ▶ Benefits: eliminates arms race, enhances liquidity, enhances market stability
 - ▶ Cost: investors must wait a small amount of time to trade

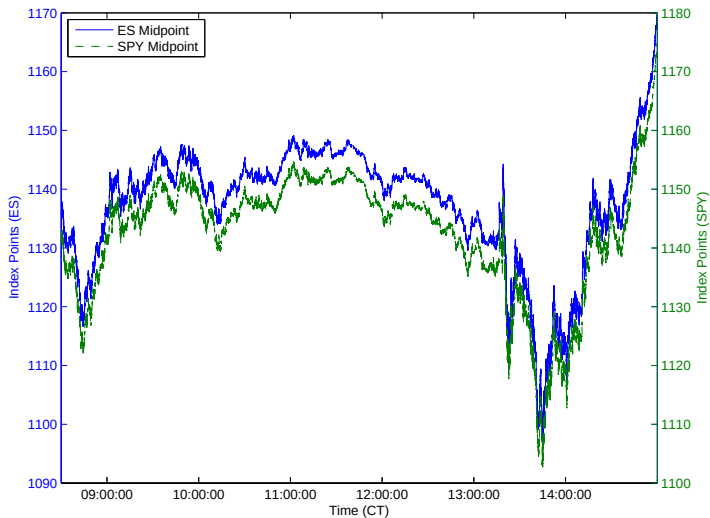
Frequent Batch Auctions

A simple idea: make time discrete.

1. **Continuous limit-order books don't actually “work” in continuous time: market correlations completely break down; frequent technical arbitrage opportunities**
2. Technical arbitrage opportunities → arms race. Arms race does reduce the duration of the market failures; arms race does not actually fix the market failures
3. Theory model: critique of the CLOB market design
 - ▶ Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)
 - ▶ Identifies the costs of the arms race
 - ▶ Harms liquidity (spreads, depth)
 - ▶ Socially wasteful
4. Frequent Batch Auctions as a market design response
 - ▶ Benefits: eliminates arms race, enhances liquidity, enhances market stability
 - ▶ Cost: investors must wait a small amount of time to trade

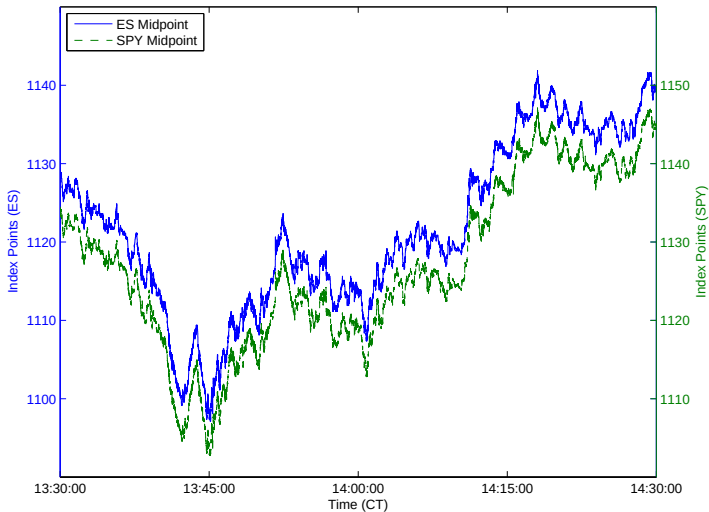
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 Day



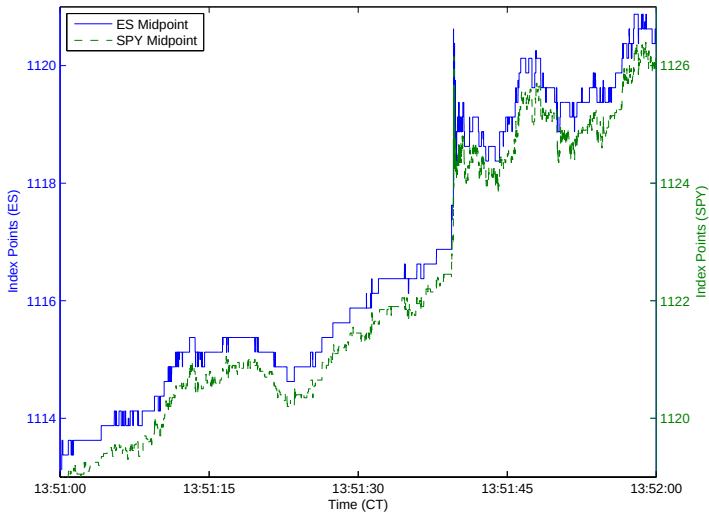
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 hour



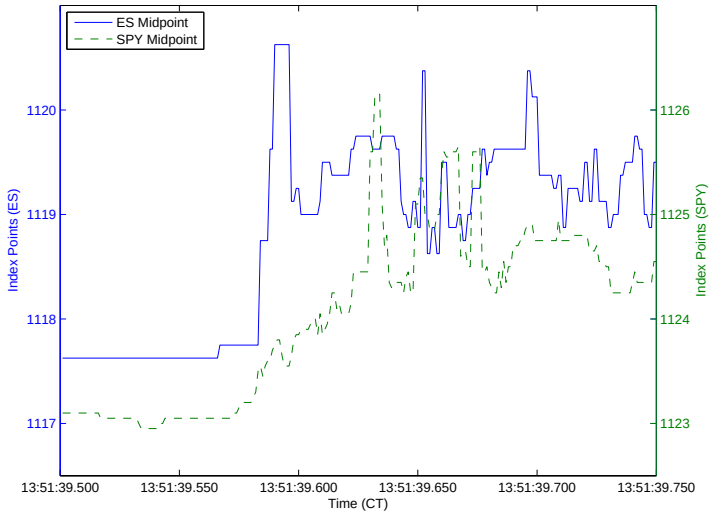
Market Correlations Break Down at High Frequency

ES vs. SPY: 1 minute



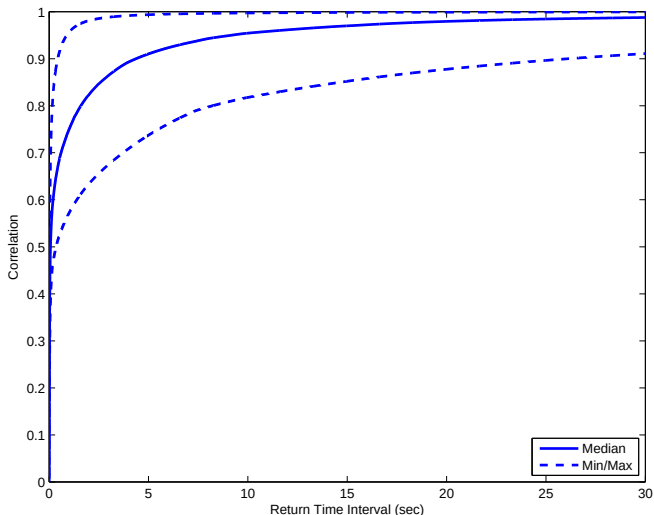
Market Correlations Break Down at High Frequency

ES vs. SPY: 250 milliseconds



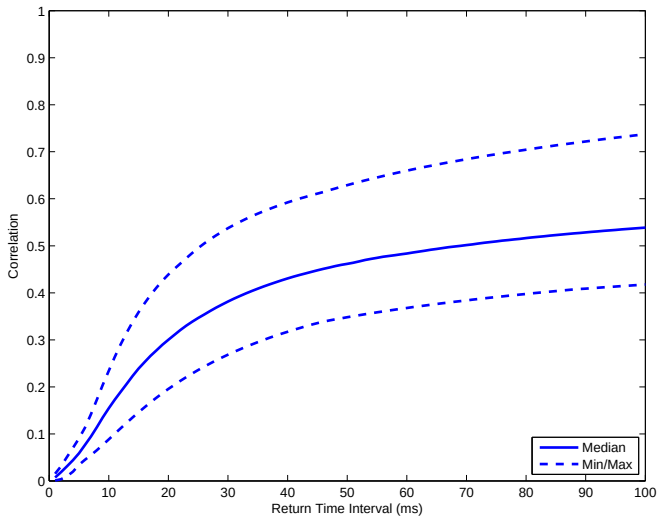
Market Correlations Break Down at High Frequency

ES vs. SPY: Correlations by Time Interval in 2011



Market Correlations Break Down at High Frequency

ES vs. SPY: Correlations by Time Interval in 2011



Correlation Breakdown: Equities

Equities Pairs: Correlations by Time Interval in 2011

	1 ms	100 ms	1 sec	10 sec	1min	10 min	30 min
HD-LOW	0.008	0.101	0.192	0.434	0.612	0.689	0.704
GS-MS	0.005	0.094	0.188	0.405	0.561	0.663	0.693
CVX-XOM	0.023	0.284	0.460	0.654	0.745	0.772	0.802
AAPL-GOOG	0.001	0.061	0.140	0.303	0.437	0.547	0.650

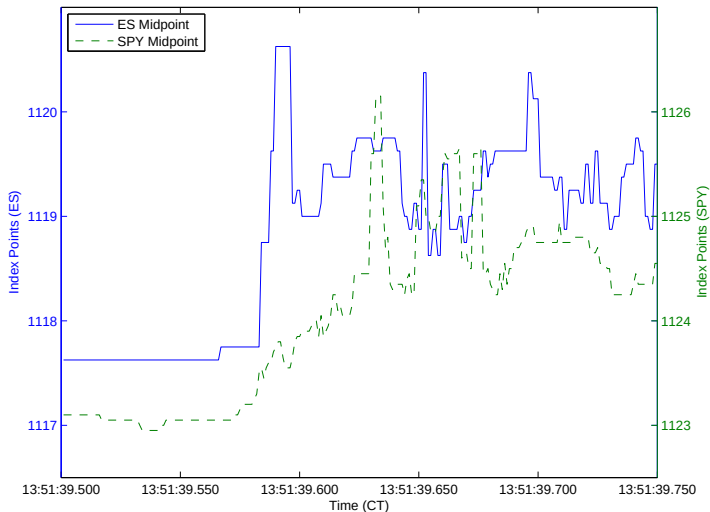
Correlation Breakdown: Equities

Equities Correlation Matrix: Correlations by Time Interval in 2011

	AAPL	XOM	GE	JNJ	IBM
	1 ms				
AAPL	1.000				
XOM	0.005	1.000			
GE	0.002	0.005	1.000		
JNJ	0.003	0.010	0.004	1.000	
IBM	0.002	0.005	0.002	0.004	1.000

	30 Min				
AAPL	1.000				
XOM	0.495	1.000			
GE	0.508	0.571	1.000		
JNJ	0.349	0.412	0.440	1.000	
IBM	0.554	0.512	0.535	0.464	1.000

Correlation Breakdown = Technical Arb Opportunities



Technical Arb Profits: ES and SPY

Summary Stats, 2005-2011

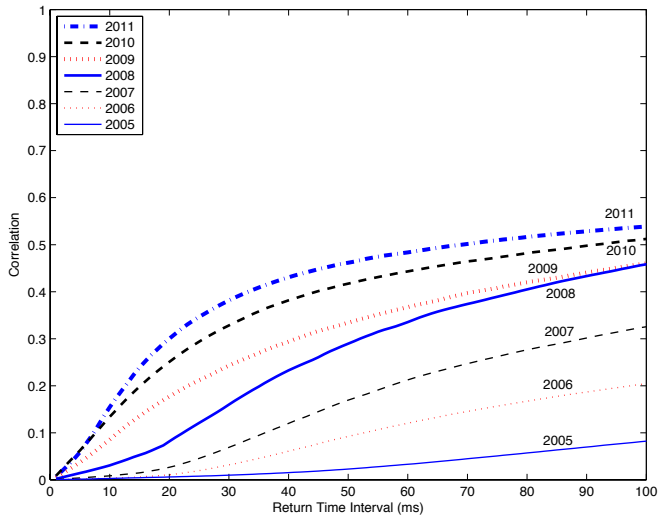
	Mean	Percentile						
		1	5	25	50	75	95	99
# of Arbs/Day	801	118	173	285	439	876	2498	5353
Qty (ES Lots)	13.83	0.20	0.20	1.25	4.20	11.99	52.00	145.00
Profits/Arb (Idx Pts)	0.09	0.05	0.05	0.06	0.08	0.11	0.15	0.22
Profits/Arb (\$)	98	1	1	5	17	50	258	927
Profits/Day-NYSE (\$)	79k	5k	9k	18k	33k	57k	204k	554k
Profits/Day-All (\$)	306k	27k	39k	75k	128k	218k	756k	2333k

Frequent Batch Auctions

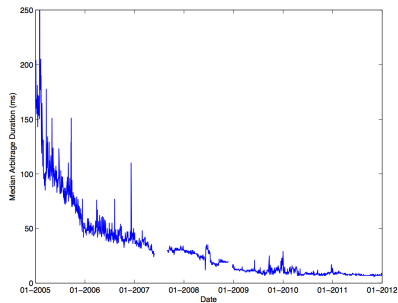
A simple idea: make time discrete.

1. Continuous limit-order books don't actually “work” in continuous time: market correlations completely break down; frequent technical arbitrage opportunities
2. **Technical arbitrage opportunities → arms race. Arms race does reduce the duration of the market failures; arms race does not actually fix the market failures**
3. Theory model: critique of the CLOB market design
 - ▶ Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)
 - ▶ Identifies the costs of the arms race
 - ▶ Harms liquidity (spreads, depth)
 - ▶ Socially wasteful
4. Frequent Batch Auctions as a market design response
 - ▶ Benefits: eliminates arms race, enhances liquidity, enhances market stability
 - ▶ Cost: investors must wait a small amount of time to trade

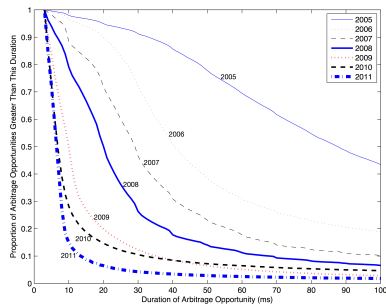
Correlation Breakdown Over Time



Arb Durations over Time: 2005-2011

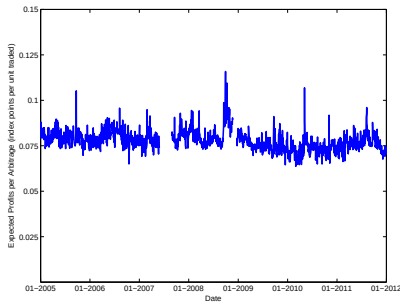


(a) Median over time

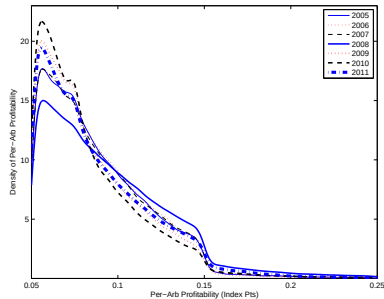


(b) Distribution by year

Arb Per-Unit Profits over Time: 2005-2011

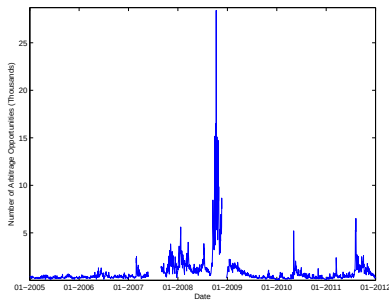


(c) Median over time

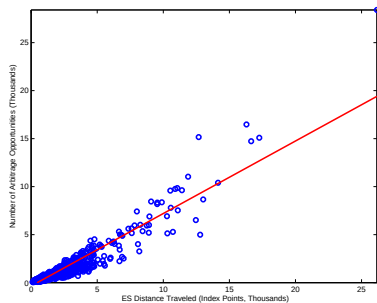


(d) Distribution by year

Arb Frequency over Time: 2005-2011



(e) Median over time



(f) Frequency vs. Volatility

Arms Race is a “Constant” of the Market Design

- ▶ Results suggest that the arms race is a mechanical “constant” of the continuous limit order book.
 - ▶ Rather than a profit opportunity that is competed away over time
- ▶ Correlation Breakdown
 - ▶ Competition does increase the speed with which information is incorporated from one security price into another security price
 - ▶ Competition does not eliminate correlation breakdown
- ▶ Technical Arbitrage
 - ▶ Competition does increase the speed requirements for capturing arbs (“raises the bar”)
 - ▶ Competition does not reduce the size or frequency of arb opportunities
- ▶ These facts both inform and are explained by our model

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. Correlations that are high but far from one can also be exploited in a statistical sense. Example: GS-MS
 4. Race to top of book (artifact of minimum tick increment and maker-taker)
 5. Race to respond to public news (eg Business Wire, Fed)

We don't attempt to put a precise estimate on the total prize at stake in the arms race, but common sense extrapolation from our ES-SPY estimates suggest that the sums are substantial.

Frequent Batch Auctions

A simple idea: make time discrete.

1. Continuous limit-order books don't actually "work" in continuous time: market correlations completely break down; frequent technical arbitrage opportunities
2. Technical arbitrage opportunities → arms race. Arms race does reduce the duration of the market failures; arms race does not actually fix the market failures
3. **Theory model: critique of the CLOB market design**
 - ▶ **Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)**
 - ▶ **Identifies the costs of the arms race**
 - ▶ **Harms liquidity (spreads, depth)**
 - ▶ **Socially wasteful**
4. Frequent Batch Auctions as a market design response
 - ▶ Benefits: eliminates arms race, enhances liquidity, enhances market stability
 - ▶ Cost: investors must wait a small amount of time to trade

Model: Key Idea

Key idea: in a CLOB, any time there is public information, there is a race. This race ultimately harms liquidity provision.

- ▶ Why? Consider the race from a liquidity provider's perspective
 - ▶ Suppose there is a publicly observable news event that causes his quotes to become “stale”
 - ▶ E.g., a change in the price of a highly correlated security (ES/SPY), Fed announcement
 - ▶ 1 of him, trying to adjust his stale quotes
 - ▶ Many others, trying to “pick off” his stale quotes
 - ▶ In a continuous limit order book, messages are processed one-at-a-time in serial ...
 - ▶ so the 1 usually loses the race against the Many ...
 - ▶ Even if he, too, is at the cutting edge of speed
- ▶ Incentive to be fast
 - ▶ Snipers: win race to pick off stale quotes
 - ▶ Liquidity providers: get out of the way of the snipers!

Model: Key Idea

- ▶ This technical cost of providing liquidity – liquidity-providing HFTs getting “picked off” by other HFTs in the race to respond to symmetrically observable public news – is incremental to the usual fundamental costs of providing liquidity
 - ▶ Asymmetric information, inventory costs, search costs
- ▶ In a competitive market, picking off costs get passed on to investors
 - ▶ Thinner markets, wider bid-ask spreads
- ▶ Ultimately, in equilibrium of our model, all of the \$ spent in the arms race come out of the pockets of investors
 - ▶ Arms-race prize = expenditures on speed = cost to investors
 - ▶ Remember: arms-race profits have to come from *somewhere*

Model: Additional Remarks

The Arms-Race is a “Constant”

- ▶ Comparative static: the negative effects of the arms race 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

Model: Additional Remarks

Role of HFTs

- ▶ In our model HFTs endogenously perform two functions
 - ▶ Useful: liquidity provision / price discovery
 - ▶ Rent-seeking: picking off stale quotes
- ▶ HFTs are indifferent between these two roles in equilibrium of our model
- ▶ The rent-seeking seems like zero-sum activity among HFTs (all good fun!)
 - ▶ but we show that it ultimately harms real investors
- ▶ Frequent batching preserves the useful function but eliminates the rent seeking function (or at least reduces)

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?

Our model yields two responses

1. Model shows that the arms race can be interpreted as a prisoners' dilemma
 - ▶ If all HFTs could commit not to invest in speed, they'd all be better off
 - ▶ But, each individual HFT has incentive to deviate and invest in speed
2. Model shows that a violation of the efficient market hypothesis is built in to the market design
 - ▶ Violations of the the weak-form EMH are intrinsic to the CLOB
 - ▶ You can make money from purely technical information (and HFTs do!)
 - ▶ Core issue: continuous markets process messages in serial (i.e., one-at-a-time)
 - ▶ Even for public / technical info (e.g., a jump in ES): *somebody is always first to react*

Frequent Batch Auctions

A simple idea: make time discrete.

1. Continuous limit-order books don't actually "work" in continuous time: market correlations completely break down; frequent technical arbitrage opportunities
2. Technical arbitrage opportunities → arms race. Arms race does reduce the duration of the market failures; arms race does not actually fix the market failures
3. Theory model: critique of the CLOB market design
 - ▶ Shows that the arms race is a never-ending, equilibrium, feature of the market design (helps explain empirical facts)
 - ▶ Identifies the costs of the arms race
 - ▶ Harms liquidity (spreads, depth)
 - ▶ Socially wasteful
4. **Frequent Batch Auctions as a market design response**
 - ▶ **Benefits:** eliminates arms race, enhances liquidity, enhances market stability
 - ▶ **Cost:** investors must wait a small amount of time to trade

Frequent Batch Auctions: Overview

- ▶ High level: analogous to a CLOB, except time is discrete
- ▶ Discrete time then necessitates batch processing, using an auction

Frequent Batch Auctions: Details

Order Submission

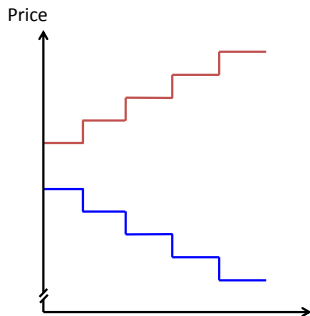
- ▶ During the batch interval (eg 1 second), traders submit bids and asks as price-quantity pairs
 - ▶ Can be freely modified, withdrawn at any time
 - ▶ If an order is not executed in the batch at time t , it automatically carries over for $t + 1, t + 2$, etc., until it is either executed or withdrawn
 - ▶ Analogous to standard limit orders

Frequent Batch Auctions: Details

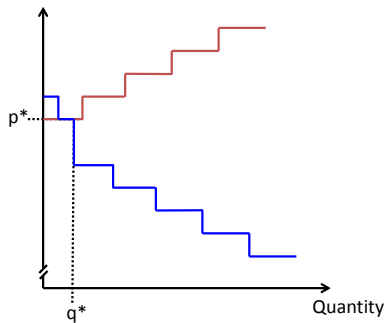
Auction

- ▶ At the conclusion of each batch interval, the exchange “batches” all of the outstanding orders, and computes market-level supply and demand curves
- ▶ Case 1: supply and demand don't cross
 - ▶ No trade
 - ▶ All orders remain outstanding for the next batch auction
 - ▶ We expect this case to be quite common
 - ▶ Analogous to liquidity provision in a CLOB
- ▶ Case 2: supply and demand cross (at price p^*)
 - ▶ All bids strictly greater than p^* and all asks strictly lower than p^* transact their full quantity at p^* (“uniform price”)
 - ▶ Bids and asks of exactly p^* may get rationed
 - ▶ Suggested rationing rule: pro-rata with time priority across batch intervals but not within batch intervals

Frequent Batch Auctions: Illustrated



(a) Case 1: No Trade



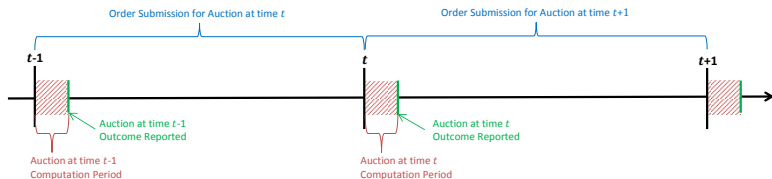
(b) Case 2: Trade

Frequent Batch Auctions: Details

Reporting

- ▶ After the auction is computed, the following information is announced publicly
 - ▶ The market-clearing price (or the outcome “no trade”)
 - ▶ The quantity cleared (possibly zero)
 - ▶ The supply and demand curves
 - ▶ Optional: individual bids comprising the supply and demand curves
- ▶ Key point: orders for the time t batch auction are *not* visible during the time t batch interval. Instead, announced *after* the time t auction is conducted. This is to prevent gaming.
- ▶ Analogous to current practice under the CLOB
 - ▶ (i) submit order; (ii) exchange processes order; (iii) public announcement
 - ▶ For auction at time t , traders see bids and asks at time $t - 1, t - 2, t - 3, \dots$ (see market as it was a latency ago)

Frequent Batch Auctions: Illustrated



Why and How Batching Eliminates the Arms Race

There are two reasons why batching eliminates the arms race:

1. Batching reduces the value of a tiny speed advantage
 - ▶ If the batch interval is 1 second, a 1 millisecond speed advantage is only $\frac{1}{1000}$ th as useful
2. Batching transforms competition on speed into competition on price
 - ▶ Ex: the Fed announces policy change at 2:00:00.000pm ...
 - ▶ Continuous market: competition manifests in a race to react. *Someone is always first.*
 - ▶ Batched market: competition simply drives the price to its new correct level for 2:00:01.000. Lots of orders reach the exchange by the end of the batch interval.

Computational Benefits of Frequent Batching

- ▶ Overall
 - ▶ Continuous-time markets implicitly assume that computers and communications technology are infinitely fast.
 - ▶ Discrete time respects the limits of computers and communications. Computers are fast but not infinitely so.
- ▶ Algorithmic traders
 - ▶ Continuous: Always uncertain about current state; temptation to trade off robustness for speed
 - ▶ Discrete: Everyone knows state at time t before decision at time $t + 1$
- ▶ Exchanges
 - ▶ Continuous: Computational task is mathematically impossible; latencies and backlog unavoidable
 - ▶ Discrete: Computation is easy
- ▶ Regulator
 - ▶ Continuous: Audit trail is difficult to parse; who knew what when? in what order did events occur across markets?
 - ▶ Discrete: Simple audit trail; state at $t, t + 1, \dots$

Summary: Costs and Benefits of Frequent Batching

► Benefits

- Enhanced liquidity
 - Narrower spreads
 - Increased depth
- Eliminate socially wasteful arms race
- Computational / market stability benefits of batching

► Costs

- Investors must wait until the end of the batch interval to transact
- Transition costs
- Unintended consequences

Open Questions

- ▶ Another Chicago question: if this is such a good idea, why hasn't an exchange already tried it? Potential reasons:
 - ▶ Coordination challenge
 - ▶ Regulatory ambiguities
 - ▶ Interests in the current market structure
- ▶ Issues due to fragmentation of US equities markets
 - ▶ What is equilibrium if there are both batch and continuous exchanges operating in parallel?
 - ▶ Mechanics if multiple exchanges each run batch (how to ensure law of one price)
 - ▶ Interaction with Reg NMS
- ▶ Implementation details
 - ▶ Optimal τ , and how this varies by security
 - ▶ Tick sizes?
 - ▶ Circuit breakers?
 - ▶ Role for small amount of randomization?
 - ▶ Further details of information policy

Summary

- ▶ We take a market design perspective to the HFT arms race.
 - ▶ Propose a simple idea: make time discrete. Replace (continuous-time) limit-order books with (discrete-time) frequent batch auctions.
1. Continuous-time markets are a fiction: correlations break down; frequent technical arbs
 2. Technical Arbs → Arms Race. Arms Race is a “constant” of the market design.
 3. Theory: root cause is the CLOB market design
 - ▶ Arms race is a never-ending, equilibrium feature of the CLOB
 - ▶ Arms race harms liquidity and is socially wasteful
 4. Frequent Batch Auctions as a market design response
 - ▶ Benefits: eliminates arms race, enhances liquidity, enhances market stability
 - ▶ Costs: investors must wait a small amount of time to trade, law of unintended consequences