Empirical Limitations on High Frequency Trading Profitability

Michael Kearns
Computer and Information Science, University of Pennsylvania
SAC Capital, New York

Market Microstructure: Confronting Many Viewpoints
Paris, 9 December 2010

Joint work with:
Yuriy Nevmyvaka (SAC Capital)
Alex Kulesza (University of Pennsylvania)

Background and Motivation

- HFT media/furor/controversy/witch hunt/investigations
- Many activities we might regard as fundamentally distinct are being conflated:
  - co-location, low-latency access
  - algorithmic trading (optimized execution)
  - dark pools and flash trading
  - market-making, liquidity provision
  - front-running
- Lots of "guesstimates" and back-of-the envelope calculations of HFT profits
- Almost no careful, data-centric empirical studies (Brogaard 2010)
- This talk: a large-scale, systematic, data-centric and "model-free" (almost) calculation of the *maximum profitability* (overestimate) of certain types of HFT
- An extensive accounting exercise

Premise and Assumptions

- To conduct precise experiments, must commit to some definition of HFT
- Equate HFT with *short holding periods*
- Rationale: if your alpha is consistently realized over minutes or hours, you don't need picosecond latency
- Divide HFT strategies into two (very) broad categories:
    - aggressive: cross the spread to enter every trade
    - passive: exclusively employ (non-marketable) limit orders every trade
- Here: focus exclusively on *aggressive HFT*
- Rationale: passive HFT is liquidity provision and therefore "benign" (Brogaard 2010)
- This work: empirically (over)estimate *total market size* (profitability) of aggressive HFT in 2008 for all ~6K U.S. equities
- Note: Sharpe ratios generally a misleading/inappropriate measure for HFT
    - returns vs. capacity
- Fundamental tension: trading costs (spreads) vs. short-term price changes

Methodology: Overview

• Using internal QAT (ITCH) message data from NASDAQ, perform full reconstruction of order books throughout all of 2008 (9:45AM – 3:45PM).

• Divide time into discrete "instances" at 10ms intervals, conditioned on there being any change to the top of the books since the last instance.

• Permit trading at every instance. Use the order books *only* to compute the prices of entering and exiting trades crossing the spread and walking the books. Books are reset to their historical states at every new instance; thus there is *no long-term market impact* in our simulations.

• Simulate the *Omniscient Trader* at every instance, and compute its total profitability for a given name in 2008.

• Apply the above methodology to 19 higher-liquidity NASDAQ names; use TAQ data and regression methodology to scale to larger universe and composite exchange (details later).

The Omniscient Trader (OT)

- Has a single parameter: *holding period h* (seconds)

- At each time t, the OT may either buy or sell v shares, for any integer v >= 0. The purchase or sale of the v shares occurs at *market prices;* thus the OT must cross the spread and (potentially) walk the book for the v shares.

- If at time t the OT bought/sold v shares, at time t+h it *must liquidate* this position and sell/buy the shares back, again by crossing the spread and paying market prices on the opposing book.

- At each time t, the OT makes only that trade (buying or selling, and the choice of v) that *optimizes (absolute) profitability.* This may often mean doing nothing.

- Holding periods examined: 10ms, 100ms, 500ms, 1s, 2s, 3s, 4s, 5s, 10s

- Also permitted *variable* holding period

Sources of Optimism/Overestimation

- Omniscience! In reality must *predict* profitable direction and size
- No fees or commissions paid by OT
- Zero latency for OT
- No market impact for OT
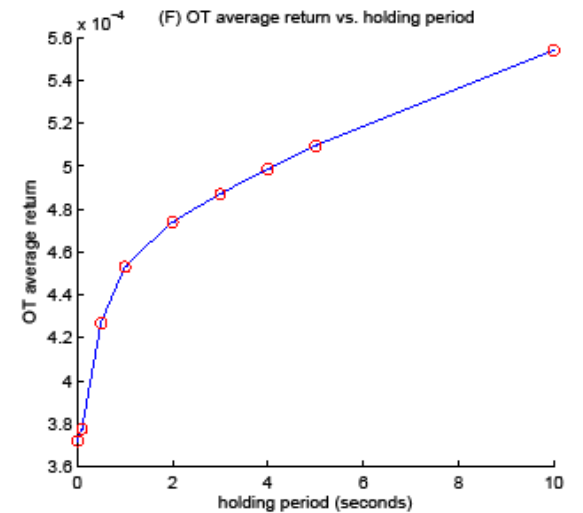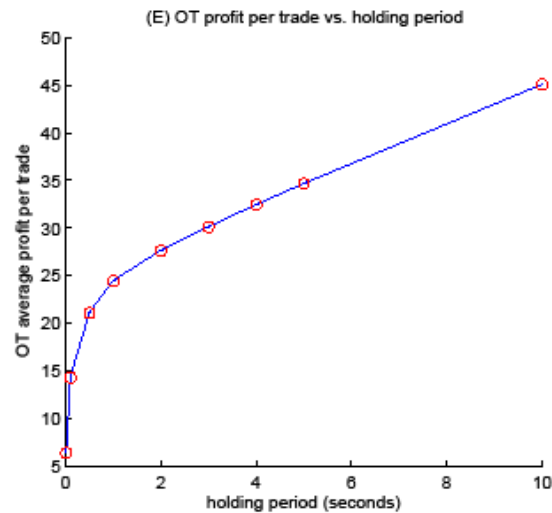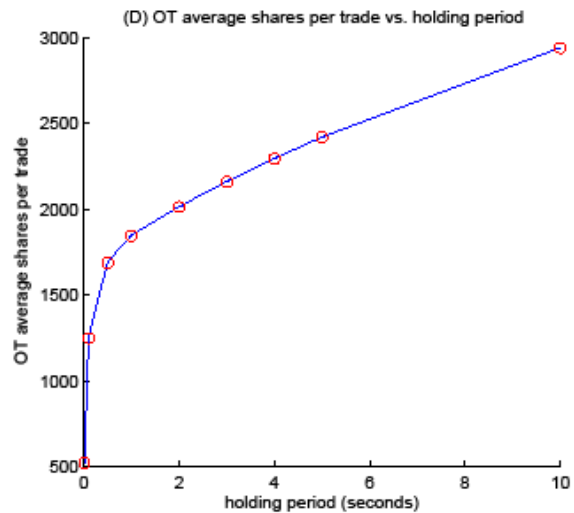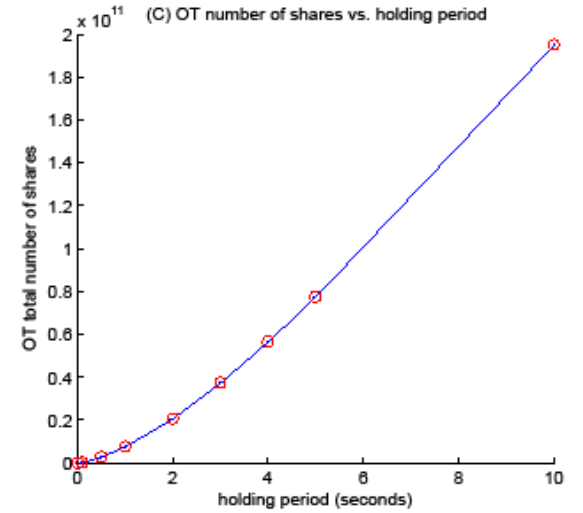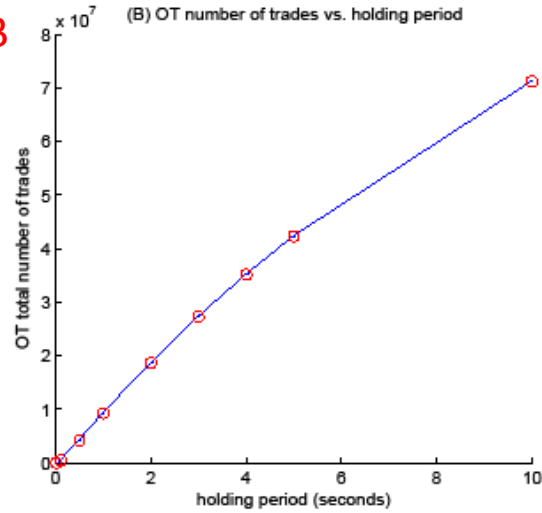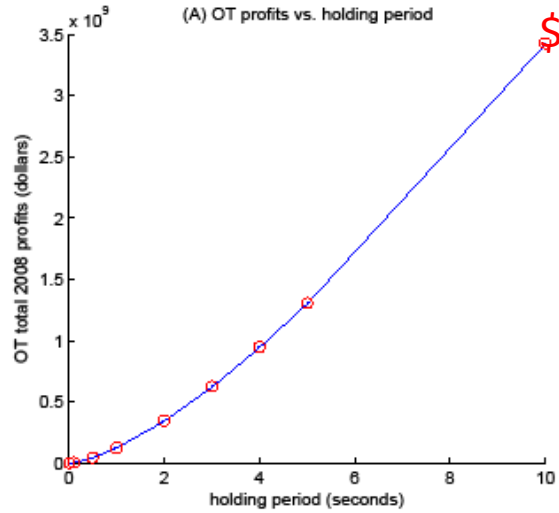- Overcounting of instances

What We're Not Accounting For

- Passive HFT: market-making/liquidity provision *(Are these "benign"?)*
- Conditional holding periods *(What does "high frequency" mean?)*
- Positive market impact: e.g. inducing momentum
- Cross-exchange plays: dark pools, flash trading
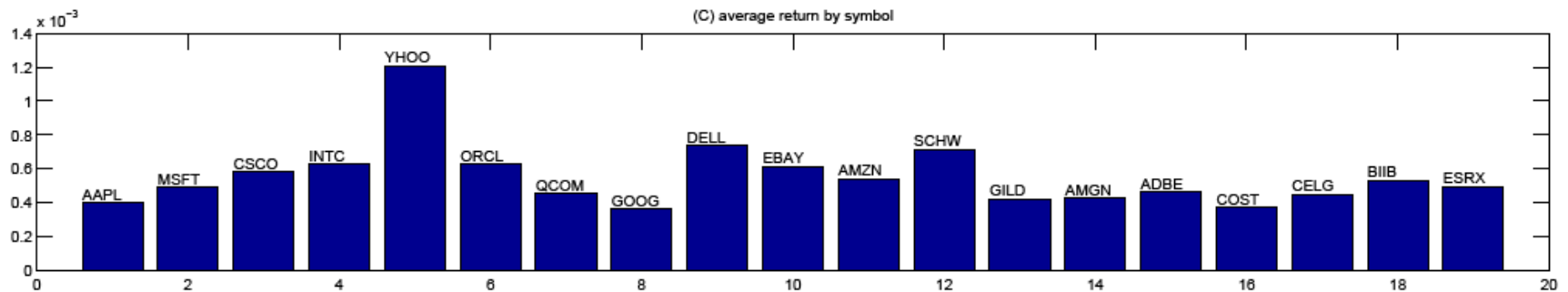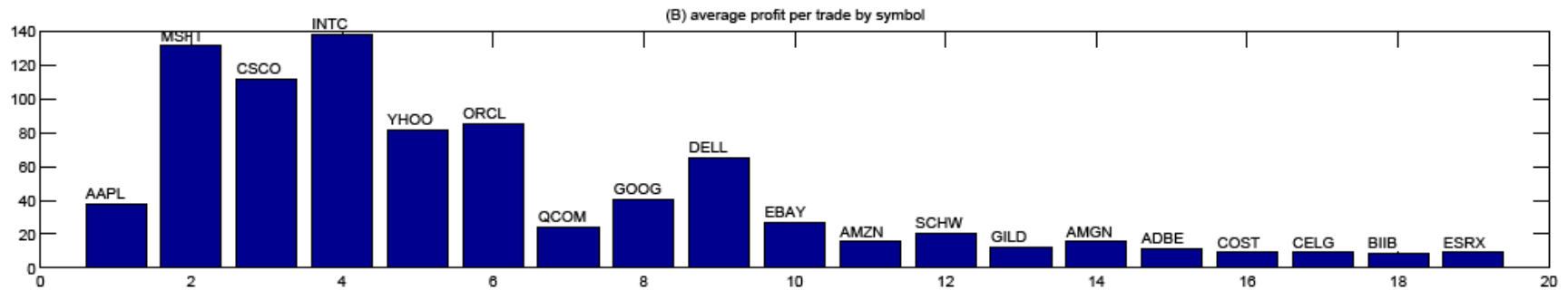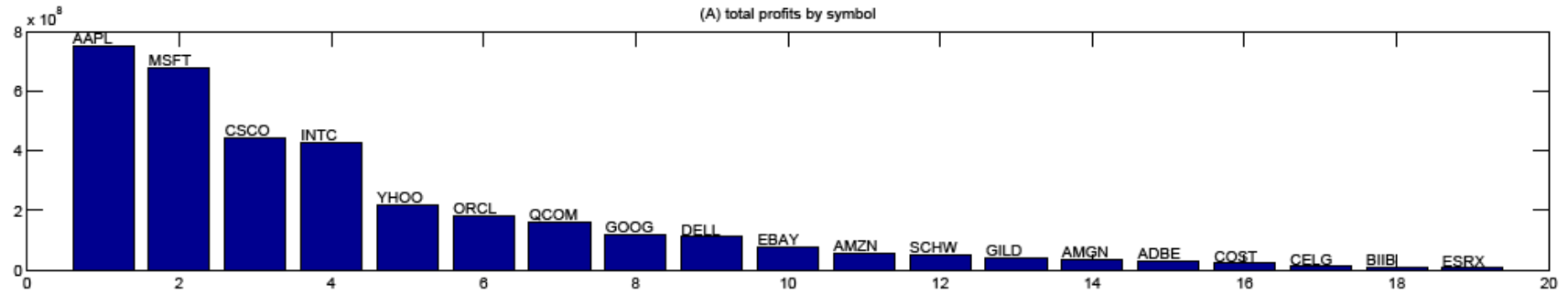- Non-equity instruments: futures, FX, ETFs, etc.
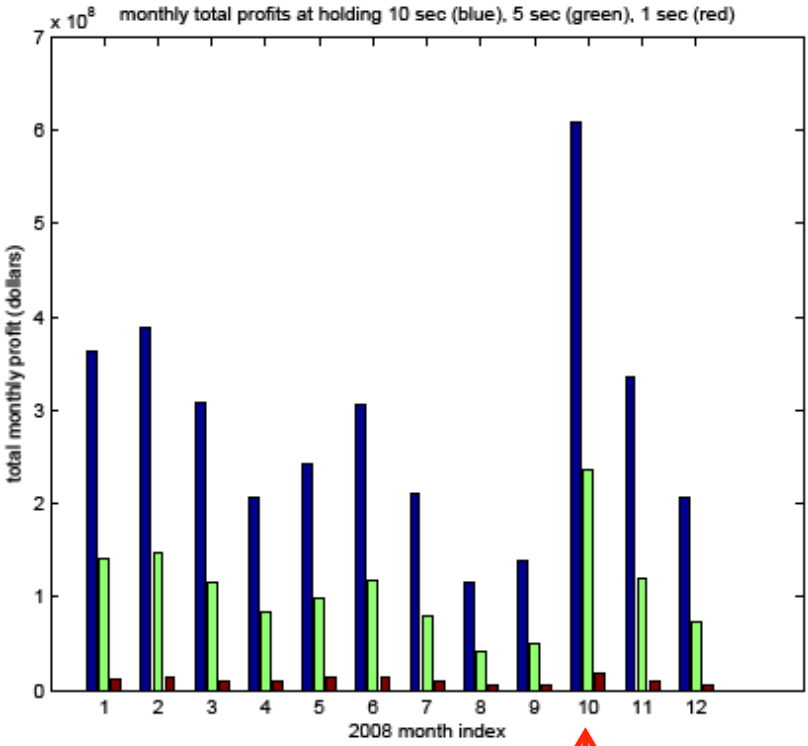- Non-U.S. markets

Results on the 19 Names

Aggregated Statistics vs. Holding Period

Per-Name Breakdown, h = 10s

(A) total profits by symbol

(B) average profit per trade by symbol

(C) average return by symbol

2008 Aggregate Monthly Profits

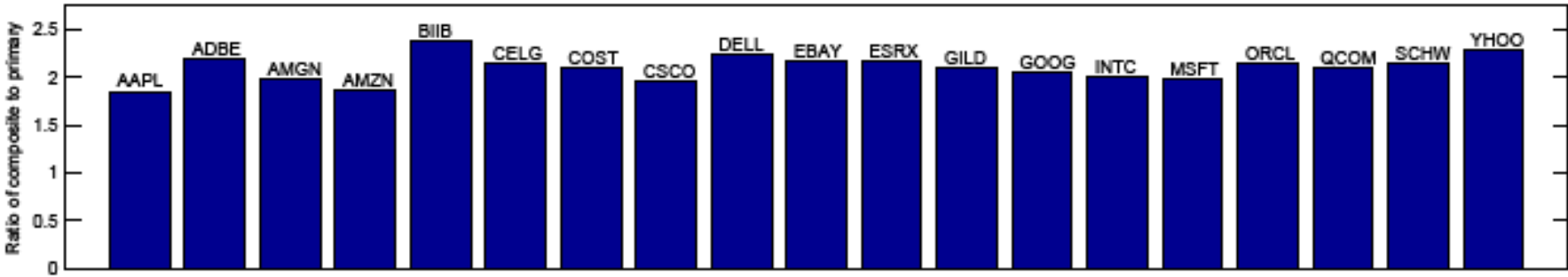monthly total profits at holding 10 sec (blue), 5 sec (green), 1 sec (red)

crisis peak
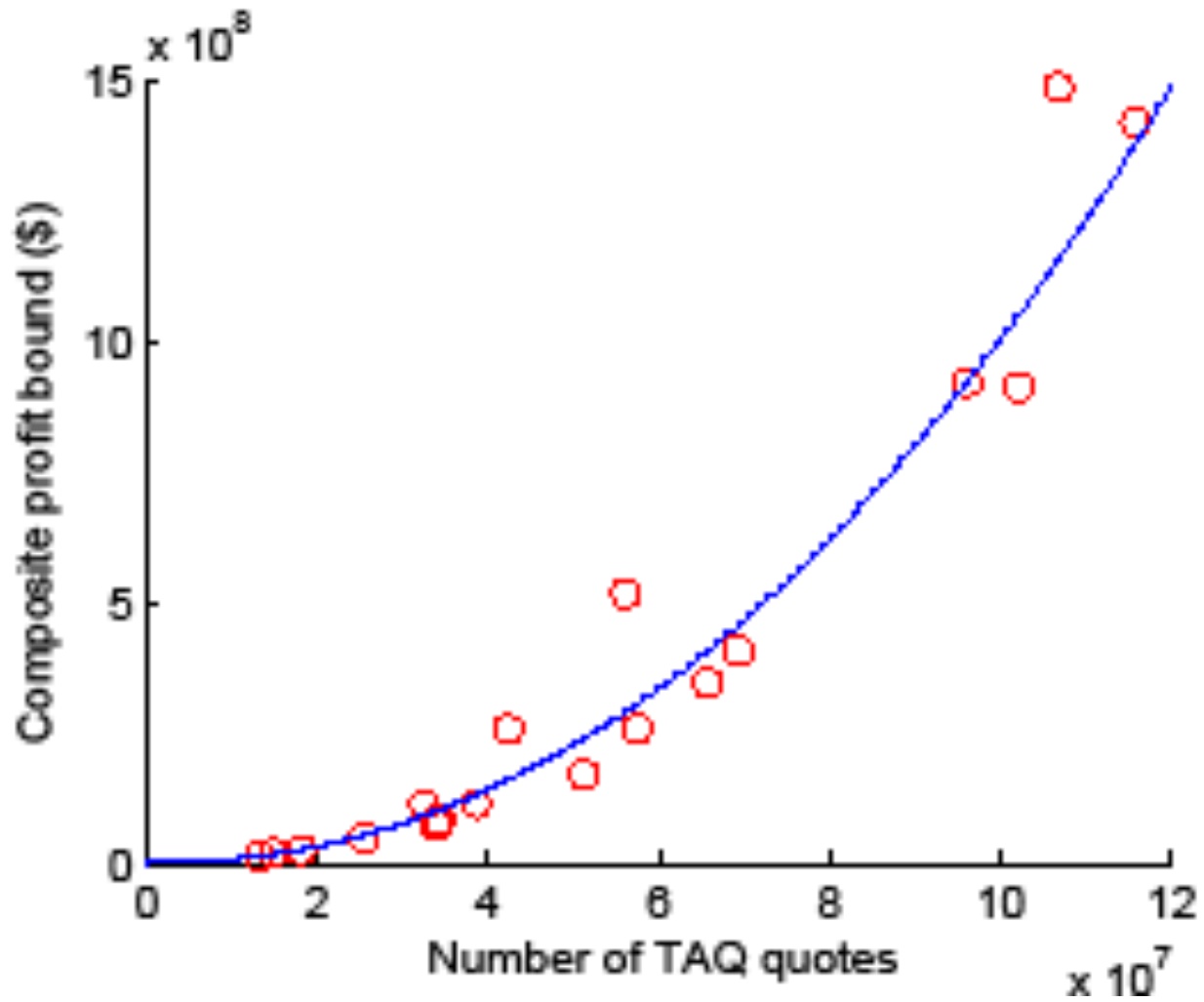
Scaling to the Full Universe

Scaling Methodology: Overview

• Computation (can't even uncompress statically) and data (only have NASDAQ OBs) preclude running methodology on all names and exchanges

• TAQ data includes bid/ask prices and volumes for primary and secondary

• Can thus run *modified OT* on TAQ data: can only consume bid/ask volumes

• On original 19 names, check correspondence between OT profits on full NASDAQ OBs and TAQ primary data (sanity check)

• On original 19 names and TAQ composite data, estimate OT profitability

• Now have 19 <name, composite profitability> pairs

• Fit two-parameter, non-linear regression model mapping number of TAQ quotes to profitability

• Use TAQ quotes to (over)estimate profitability on full universe of 6,279 US stocks
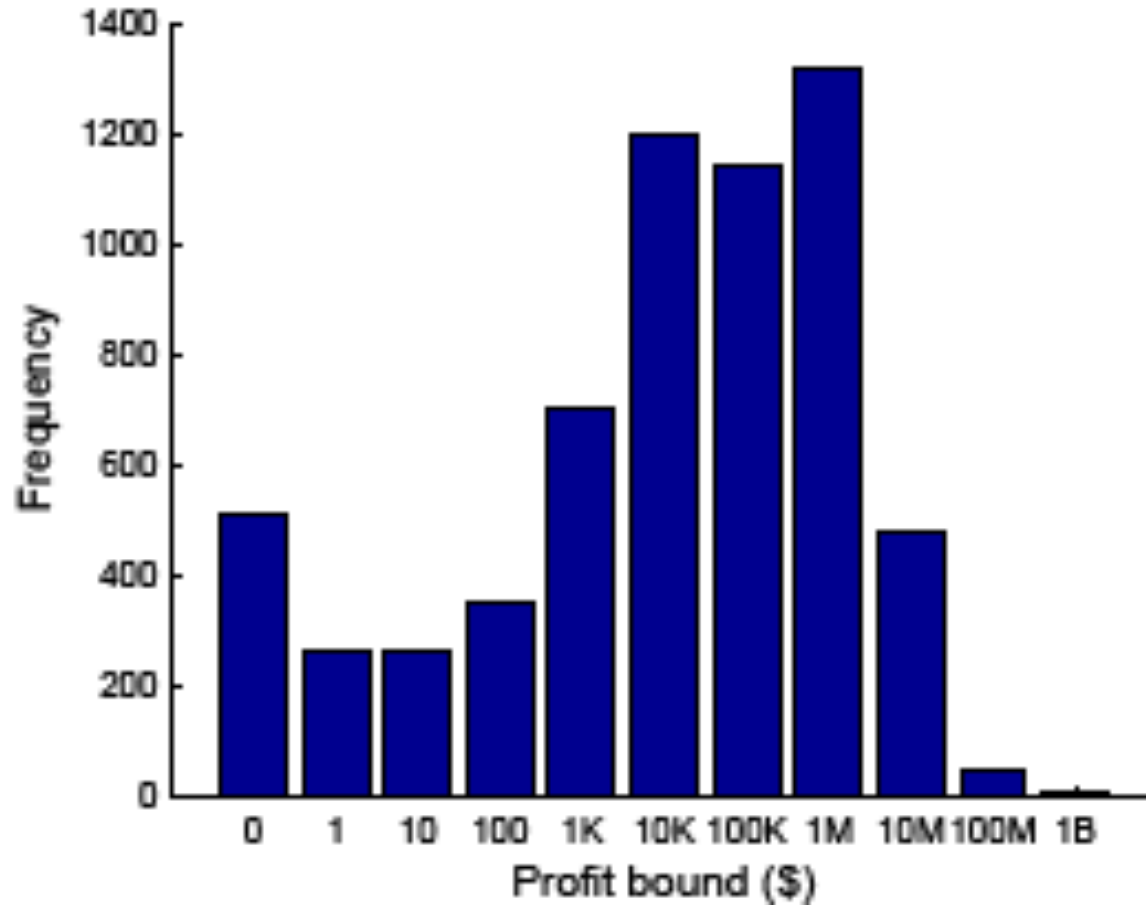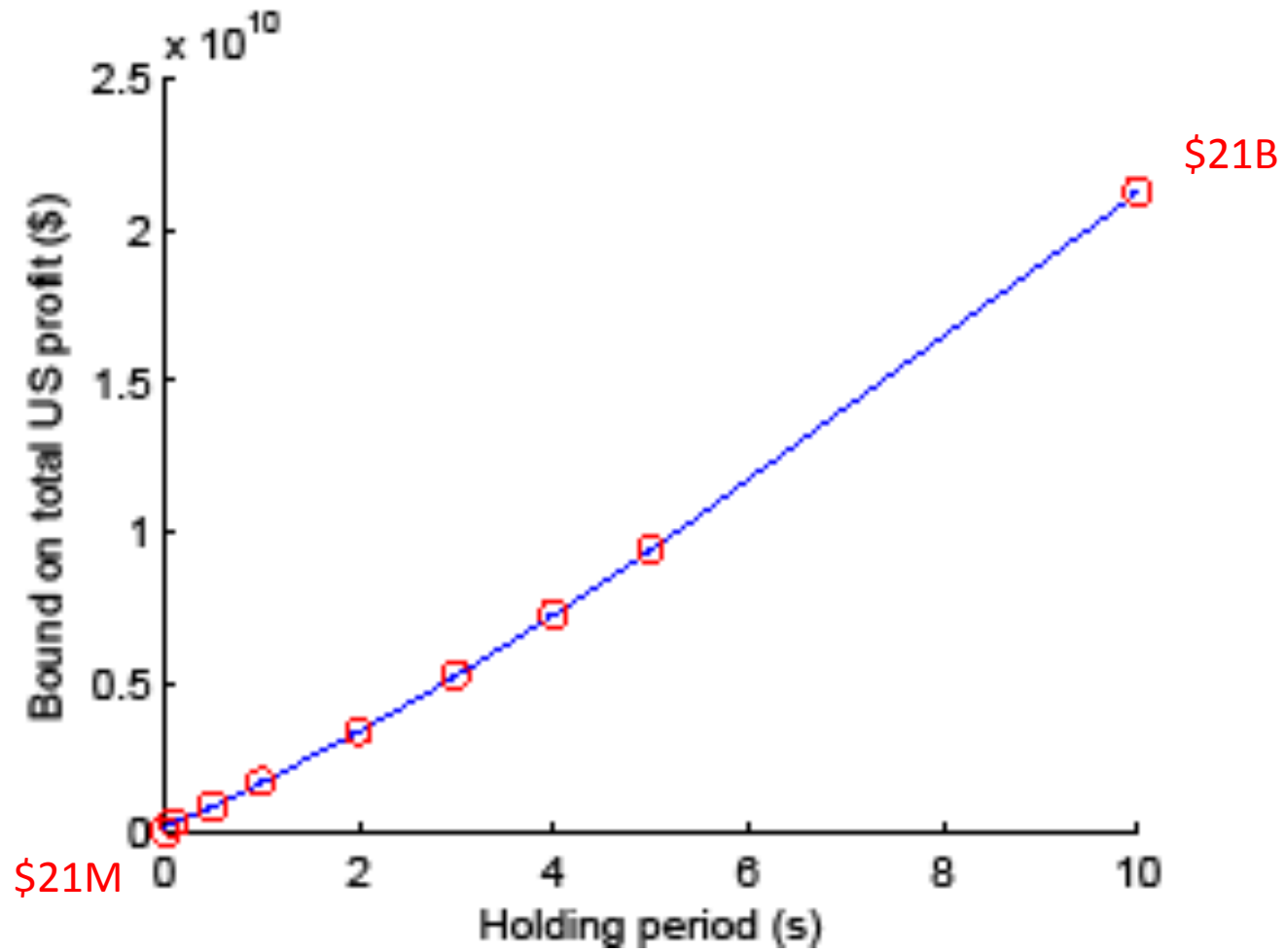
Composite/Primary Ratios (10s)

Histogram of 2008 Composite Profit Overestimates (10s)

Profit Bound vs. Holding Period, Full Universe

Closing Remarks

- $21B vs *~ $52 trillion* (TAQ) trading volume in same universe annually (<0.05%)
- 10% omniscience & no losses → $2.1B (consistent with Brogaard 2010)
- Allow optimal exit point within a bounded window → ~50% increase
- Already simulating zero latency; no market impact or fees for taking liquidity
- Figures for 2009/10 likely much lower due to 2008 volatility

- Some parties are getting rich from HFT. Should society be concerned?

Scaling Methodology: Overview

• Computation (can't even uncompress statically) and data (only have NASDAQ OBs) preclude running methodology on all names and exchanges

• TAQ data includes bid/ask prices and volumes for primary and secondary

• Can thus run *modified OT* on TAQ data: can only consume bid/ask volumes

• On original 19 names, check correspondence between OT profits on full NASDAQ OBs and TAQ primary data (sanity check)

• On original 19 names and TAQ composite data, estimate OT profitability

• Now have 19 <name, composite profitability> pairs

• Fit two-parameter, non-linear regression model mapping number of TAQ quotes to profitability

• Use TAQ quotes to (over)estimate profitability on full universe of 6,279 US stocks

Closing Remarks

$21B vs ~ $52 trillion trading volume (TAQ) annually in same universe (<%0.05)