A BACKGROUND AND CHRONOLOGICAL TAKE ON HIGH FREQUENCY TRADING

by

Quinn Murray

A thesis submitted to The University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
April 2016

Approved by

Advisor: Professor Bonnie Van Ness

Reader: Professor Robert Van Ness

Reader: Professor Dwight Frink
ABSTRACT

The purpose of this thesis is to analyze, in context of high frequency trading, potential market manipulation techniques (i.e. market making arbitrage, statistical arbitrage, market structure arbitrage, and directional strategies), and to review the subject from a chronological perspective from the 1960s onward, covering topics not limited to Regulation National Market System, the flash crash of May 6, 2010, and the August 24, 2015 market crash.

To date, high frequency trading’s effects on the United States market place have been well documented. This thesis will speculate about the true nature—whether adverse or beneficial-- of this fascinating, evolutionary, highly scrutinized topic.
# TABLE OF CONTENTS

INTRODUCTION.........................................................................................................................5
PREDATORY TRADING TECHNIQUES..........................................................................................8
THE ORIGINS AND EVOLUTION OF HFT: A CHRONOLOGICAL TAKE.................................31
CONCLUSION.............................................................................................................................46
REFERENCES.............................................................................................................................48
ACKNOWLEDGEMENTS.............................................................................................................54
INTRODUCTION

“Wall Street, like a clever pervert, is often suspected but seldom understood and never convicted”

-Michael Lewis

The familiar cliché of gaudily dressed men waving arms and placing orders on a stock market floor is no longer: trading now occurs on the basis of algorithms formulated not by financiers or programmers, but instead by quantum physicists, climate scientists, and theoretical mathematicians—some of the most formidable minds in the world.¹ In the face of the market now dominated by computers running complex algorithms that have sanctified every last microsecond (one millionth of a second), regulators have been left entirely perplexed, and perhaps, impotent.

Wielding the weapons of speed and complexity, high frequency traders (here forth, HFT) spoof investors via numerous predatory trading techniques that the author will later cover in detail.

The thinking behind HFT isn’t unique in any sense of the word—it’s all relative. Although some closely associate HFT with high-speed computers, HFT really describes how market participants use technology to gain information and act upon it in advance of the rest of the market. For example, near the advent of the

telescope, market merchants would use telescopes and look out to the sea to
determine the cargo hold of incoming merchant ships. If the merchant could
determine which goods were soon to arrive on these ships, they could sell off their
excess supply in the market before the incoming goods could introduce price
competition. Another example, British banker and speculator Nathan Rothschild
had an agent at the Battle of Waterloo in 1815. Rothschild’s agent saw that
Napoleon was losing, rushed back to the coast and hired a boat to sail through a
storm to England. On getting the news, Rothschild made his way to the London
Stock Exchange and acted as though he wanted to sell British shares, giving the
impression that British commander, Wellington, had lost the battle. Everybody
pitched in to sell and all the while, Rothschild quietly bought up available British
shares before the news arrived of British victory. Rothschild and the market
merchant’s cases obviously suggest speedier information allow for a speculative
advantage—a point that regulators don’t seem to understand.

Speed-centric trading began in 1971 with NASDAQ, the first computer-driven
exchange. While some form of HFT has been around for a long time, its true
potential was first uncovered on "Black Monday"—October 19, 1987—with the first
whole market flash crash, which resulted from an exponential propagation of
program trading. Even though some thought that “Black Monday” would have

---

taught traders and regulators a lesson, it merely accelerated the incursion of computerized and algorithmic trading into regular markets. In fact, HFT now accounts for nearly three quarters of all exchange-based trading volume. And although today’s HFT-made markets may impress with their steadily decreasing execution times, their apparent efficiency, and narrow spreads, reality will soon obliterate any naïve notions of transparency in our markets and reveal their true opaque composition. Put plainly, "high frequency trading," in a sense, is a misnomer-- it instead should be called high frequency front running, or high frequency market manipulation. And immoral as these practices may seem, they are just as legal as the packaging of sub-prime mortgages into tradable securities prior to the financial crisis. The question remains as to how this issue should be addressed.

---

PREDATORY TRADING TECHNIQUES

The U.S. Commodity Futures Trading Commission acknowledged in 2011 it does not “purport to have a perfect definition” for HFT, but proposed a “seven-part test for what constitutes HFT”\(^7\):

1. The use of extraordinarily high-speed order submission/cancellation/modification systems with speeds in excess of five milliseconds or generally very close to minimal latency of a trade.

2. The use of computer programs or algorithms for automated decision making where order initiation, generating, routing, and execution are determined by the system without human direction for each individual trade or order.

3. The use of colocation services, direct market access, or individual data feeds offered by exchanges and others to minimize network and other types of latencies.


5. High daily portfolio turnover and/or a high order-to-trade ratio intraday.

6. The submission of numerous orders that are cancelled immediately or within milliseconds after submission.

7. Ending the trading day in as close to a flat position as possible (not carrying significant, un-hedged positions overnight).

In summation, high frequency traders need high-computing power and ultra-low latency in order to go about their business. They get it by renting server space from the stock exchanges (a process called colocation). HFT also need to access vast amounts of data-- obtained primarily from stock exchanges-- which they analyze in order to detect patterns in the markets. They then trade, capitalizing on those patterns. And, in many cases, the exchanges pay them to trade.8

One of the primary trading tactics HFT firms make use of is market making arbitrage—a strategy especially catered to by all of the for-profit exchanges.

As exchanges have become publicly traded entities, they've realigned their priorities to maximizing firm valuation and thus, revenue growth. On any given exchange, the largest traders by volume are HFT market makers, and are thus the most willing to pay for an exchange's expensive services such as enriched data feeds and colocation.9 HFT market makers and are the modern electronic replacements of the human specialists and market makers of the past whose jobs were to match customer orders as well as maintain “a fair and orderly market.”10

HFT market makers make money not only from buying stocks and selling them at a higher price, but also from the exchanges, which reward them with

---

rebates for “adding liquidity.”

Because of these exchange rebates, designated market makers can make money even when they buy and sell a stock at the same price. A critical factor here is speed. And a key for the designated market maker’s success, therefore, is being as close as possible to the exchange (utilizing colocation). In the second quarter of 2015 alone, NASDAQ paid out $216 million in transaction rebates.

Designated market makers have several things tipping the scales in their favor versus the traditional investor and retail trader. Their collocated speed and technology usually places them at the top of the queue at any price. Designated market makers also run on virtual autopilot, flipping stocks with the aim of maximizing their rebates. Whether intraday markets are directionless or trending, their technological prowess and speed gets them to the best prices for their own account faster than any other participant. In market infrastructural experts Joe Saluzzi’s and Sal Arnuk’s paper, “What Ails Us About High Frequency Trading,” the Themis Trading partners made the following analogy:

Imagine you are at the grocery store. You take your cart to one of the five apparently empty checkout lines. Suddenly, nine carts instantaneously

---

appear ahead of you. You scratch your head and move to lane two. The same thing happens. You soon find that whenever you move into a new lane, a multitude of carts appear ahead of you in line. Why? Because the supermarket has sold the right for those carts to do so. Thus, you can never be at the head of the line, no matter what you do, short of paying the exchanges a large fee to have the same right...

By pure intuition, it may make sense to assume that designated market makers provide liquidity to the rest of the market by always trading. But in reality, they provide liquidity only when they want and turn it off when things get too risky—or better put, when markets are stressed. During times of uncertainty, the HFT designated market makers exit their positions, spreads widen, and sometimes, such as the May 6, 2010 Flash Crash, there is a liquidity vacuum.\textsuperscript{16} To quote Andy Haldane, the Executive Director of Financial Stability at the Bank of England, “HFT designated market makers add liquidity in a monsoon and absorb it in a drought.”\textsuperscript{17}

Take, for instance, the April 22, 2013 Google mini-flash crash. In a flurry of trading at 9:37 a.m., Google fell from $796 to $775 in about \( \frac{3}{4} \) of a second, likely as a result of real orders meeting phantom liquidity,\textsuperscript{18} and then rebounded to $793 a second later. The drop involved 307 trades and 57,255 shares from 10 exchanges and dark pools (it’s noteworthy, too, that there were five orders placed for every

\begin{footnotes}
\item Jones, Charles M. "What Do We Know About High-Frequency Trading?" Columbia Business School (2013). Web.
\end{footnotes}
The scenario resembled a giant game of hot potato, but with huge financial consequences.

According to a study conducted by Woodbine Associates, a capital markets research and advisory firm, suboptimal order routing decisions cost institutional investors as much as $4.5 billion per year.\(^\text{20}\)

HFT firms also engage in statistical arbitrage- a tactic in operation for decades. Today, statistical arbitrage HFT trade “rich” versus “cheap” in 11 different public exchanges and roughly 44 alternative trading systems (dark pools).\(^\text{21}\) They also take advantage of Wall Street’s greatest success/growth story of recent years:

---


---
the proliferation of thousands of exchange traded funds (ETFs) to give them even more baskets to statistically compare and trade. In fact, NASDAQ alone offers over 40,000 indexes—more indexes than index components—which suggests more ETF arbitrage opportunities, more volume, and therefore, more HFT revenues. With more indexes than index components, one can’t help but wonder whether many ETFs have been developed for the express purpose of statistical arbitrage trading.

Another predatory strategy, market structure and latency arbitrage is designed to exploit built-in weaknesses in the market’s structure. As a foreword, it’s worth discussing Regulation National Market System (here forth, Reg NMS). Reg NMS created the concept of a National Best Bid and Offer (here forth, NBBO)—an aggregation of the best-priced orders on all exchanges. Meaning, if NYSE had a participant with an order to buy a stock at $10.00 and did not have a matching order at $10.00, while another exchange, say NASDAQ, did, then the NYSE would have to route that buy order to NASDAQ, which would then match the buy and sell order.²³ As a result, Reg NMS effectively commoditized trading destinations, making speed of execution paramount (so much so, that only 27% of stocks listed on NYSE are actually traded on NYSE, whereas in 2007 80% of stocks traded on their respective listed exchanges²⁴). The slower, behemoth, specialist-oriented NYSE was forced to become a fast, lean, electronic market. These changes brought by Reg NMS have turned the market from an investor-focused mechanism-- which welcomed traders

---

and investors of all types and speeds-- to a sub-second, trader-focused mechanism, where the concerns and confidence of investors have become an afterthought. Even the exchanges realize a review of market structure is essential to market stability and soundness, and important to investor confidence in the functioning of the world’s largest equity markets-- that there is certainly room to improve with respect to market stability and efficiency.\(^{25}\) In sum, via Reg NMS, the SEC has inadvertently allowed markets to become based on business models rife with conflicts of interest that cater to hyper-short-term traders.

While Reg NMS was being proposed, debated, and enacted between 2004 and 2007, the stock exchanges were busy readying new demutualized business models to better allow for revenue growth. The exchanges knew full well that future primary revenue sources would stem from technology services as opposed to old-fashioned trade matching.

Post-Reg NMS, the exchanges have become responsible for collectively maintaining the publicly available stock quotes, commonly referred to as the Security Information Processor (here forth, SIP)—basically the aggregation of quotes across all public exchanges to create a universal NBBO for every security. Data transfers from all the data centers of the exchanges into one central processor, which then spits out a universal bid/ask spread. Regulators created the SIP in order to tackle what they saw as ill effects of increased fragmentation of public exchanges (a phenomena they began to notice prior to the Securities Acts Amendment of

1975), but the SIP came with many unintended consequences as well. For one, it
directly enabled predatory latency arbitrage strategies.

The process of all the exchange feeds delivering their market data to the SIP
happens in well under a second, but given the fact that signals travel various speeds
based on the servers’ proximity to each other, there is a slight lag in relation to
“direct feeds.” Every exchange has a proprietary “members-only” direct feed that
provides the fastest quote data to those who subscribe to the feed. The closer you
are to the exchange servers, the faster you will be able to receive and transmit stock
pricing data. Note too, that when it comes to latency, today’s world makes sacred
every last microsecond (as mentioned previously).

HFT firms pay large sums of money to collocate servers directly next to
exchange servers where they can get faster access to data than the marginal
investor receives from the slower SIP (and effectively see into the future). Because
of this advantage, they know the quote of any given stock microseconds before
those looking at the public SIP. If the actual price for a stock is higher than what is
available on the slower markets, HFTs swoop in, buy it, and sell it at the higher
price. It’s like Biff Tannen in Back to the Future Part II, who makes a fortune betting
on sports lines he already knows the outcome to. It’s no wonder HFT titan Virtu
Financial, Inc. lost money on just one trading day in six years.

---

The University of Chicago, 3 Nov. 2015. Web.
28 Mamudi, Sam, and Leslie Picker. “Virtu Celebrates Another Year Without a Single
Exchanges also offer customers access to private data feeds (note the difference between private data feeds and “members-only” direct feeds), which supply traders with information on the life of any given order. These compiled feeds provide information on what time an order was placed and whether or not it was revised, whether an order had a partial fill and was then cancelled, or even whether an order revised its price and how often. HFT firms find this information immensely valuable because it allows them to model the behavior of institutional and retail investor orders to predict the near-future price movements of stocks.

To put the importance of data feeds to HFT firms in retail/consumer terms, companies like Amazon leverage their marketing databases to mine data and find patterns in customer behavior. For instance, if the consumer buys a certain good every 45 days from Amazon.com, a pattern is forged—a pattern that the retailer can model. Thus, if the customer’s behavior changes, smart retailers with intricate databases react and respond. A similar analogy can be made in the consumer banking industry. For example: if a bank’s customer direct deposits a check every two weeks, and that pattern suddenly stops, the bank can monitor that behavior (or lack thereof) and act accordingly to prompt a message. In sum, behaviors can be monitored and modeled to the customer’s benefit and to the benefit of the brand. It’s worth noting, though, that in HFT, said behavior monitoring tends to be to the advantage of HFT firms, not to the entity who is making the trade.

---

Naturally, HFT proponents argue colocation, direct access feeds, and private data feeds are available to anyone, but in truth, they are only worth much to well-capitalized companies that can build and run HFT strategies on a large scale.

Colocation cannot be overemphasized—it’s a huge business for exchanges, which, again, are for-profit enterprises. HFT firms pay millions of dollars to house servers in exchange facilities, and the demand is unimaginably high for space, connectivity, speed and bandwidth.

In some ways, colocation is a new version of an old phenomenon. Think about why, even before the advent of the telephone, so many brokerage firms originally located their headquarters in New York City: they were coloating with the NYSE so that their brokers could bring in trades to the exchange quickly and efficiently.

Because HFT firms have access to faster data from the various exchanges’ direct feeds while the rest of the market gets quotes from the SIP, latency arbitrage HFT strategies are able to slip in ahead of orders that hit the tape and skim money off of trades. The strategies function as taxes on investors—the money garnished comes directly from investors, who pay more for the stocks they buy and receive less for the stocks they sell. Technically, this tactic takes advantage of loopholes created by conflicts of interest and misaligned incentive structures. But “front running” may be a better way to describe latency arbitrage.

The issue is further compounded by geography. To illustrate, consider an investor that splits a block trade and sends two buy orders to exchanges A and B,

---

where B is geographically further away from the trader. The order reaches exchange A in, say, one millisecond and executes, while the second order takes two milliseconds to reach exchange B, located in the next town. However, the collocated HFT servers at exchange A receive notification of the first trade 0.5 milliseconds after the trade happens, leaving them with half a millisecond to act on this information before the second trade reaches exchange B.\textsuperscript{31} Remember too that a millisecond in today’s market might as well be an eternity.

Illustration courtesy of Nanex LLC

In an open letter addressing criticism and opposition during their application to become a regulated stock exchange, IEX Group, Inc. touched on market structure and latency arbitrage strategies:\textsuperscript{32}

Colocation practices of existing exchanges, coupled with exponentially increasing market data fees, deserve heightened attention... Colocation practices, coupled with the sale of low-latency market data, are intentionally


\textsuperscript{32} Lee, Sophia (General Counsel, IEX Group, Inc.), letter dated 13 Nov. 2015 to Mr. Brent J. Fields (Secretary, Securities and Exchange Commission), Securities and Exchange Commission website, http://www.sec.gov/comments/10-222/10222-20.pdf
designed to cause market participants to pay for preferred access, and result in delayed access for those who will not or cannot pay these prices. An important related issue is the latency difference between the SIPs and proprietary data feeds. The Commission has taken important steps to address that issue by requiring exchanges to provide time stamps on their messages sent to the SIPs, and by urging greater investment in the SIPs, but more work remains.

The last of the predatory tactics used by HFT firms that the author will cover in detail, Directional strategies, look to profit from the anticipation of directional movement in securities prices (which is significant because HFT firms require a hedged risk). For instance, momentum ignition strategies involve a firm submitting a series of orders in an attempt to “ignite” a rapid directional price movement. This concept, called “spoofing,” tricks algorithms and manual traders into buying or selling aggressively. Momentum ignition strategies may also look to trigger existing stop-loss orders to ignite further price movement. Unlike most other HFT strategies, spoofing is actually illegal.

Quote stuffing also falls under the category of directional strategies, and is incredibly encompassing and pervasive, affecting over 74% of US listed equities. With quote stuffing, HFT firms slam vast numbers of orders into the system, cancelling them before anyone can react, with the aim of slowing the transit of information to competitors, or of creating confusion from which they can profit. Orders are entered at higher rates, cancelled at higher rates, are for shorter

durations, are executed at lower rates, and are smaller in size during quote stuffing events. Take the following into consideration: after dividing the trading day into 390 one-minute segments, and calculating the intraday variation in quoting activity by computing the average standard deviation of the number of quotes submitted in the one-minute segments for rolling twenty-day windows, Egginton, Van Ness and Van Ness were able to identify 58,737 unique quote stuffing events (where “quote stuffing” is defined as an intense quoting episode where the level of quoting exceeds the previous twenty-day mean number of quotes-per-minute by at least 20 standard deviations) with durations ranging from one to ten minutes.34


HFTs aren’t just “managing risk” in their market making activities when they enter and cancel 95% of their orders without a trade. In reality, HFT firms are trying to create momentum. HFT firms try to mislead institutional algorithmic orders into following along these price changes, which are created with the intention to manipulate a stock price higher after the HFT firm has already bought the stock at a cheaper price.\textsuperscript{35} That, and to leave an incremental price impact on outstanding orders.

A digression: Bloomberg Executive Editor, Tracy Alloway, noted on November 11, 2015 that the number of six standard deviation broad-market moves have been increasing at an alarming rate—a number that could quite possibly “go parabolic” (to put this into perspective: a six standard deviation move should theoretically happen once every 62 years or so).\textsuperscript{36} HFT predatory trading strategies are the root of concern here.

After reviewing said predatory HFT tactics, it’s worth noting many industry participants have taken an unwavering stance that there is nothing wrong with current market structure and status quo—that predatory HFT tactics are merely an illusion. It’s easy to tell the comment below is from a HFT firm:\textsuperscript{37}

\begin{quote}


\textsuperscript{37} Greg O’Connor (Compliance Manager, Wolverine Trading), letter dated April 21, 2010 to Elizabeth M. Murphy (Secretary, Securities and Exchange Commission), Securities and Exchange Commission Website, www.sec.gov/comments/s7-02-10/s70210-143.pdf
\end{quote}
Our general view is the market structure that has evolved is effective and beneficial, especially with a view toward the public investor. The securities marketplace has witnessed a growth of technology in concert with regulatory reforms that have proven to equalize access to all investors, both professional and retail, reduce latency, narrow spreads and lower costs. Rijper, Sprenkeler, and Kip reinforced the aforementioned stance, making a case that HFT improves price discovery, supplies market liquidity, narrows spreads, reduces volatility, and improves overall market quality.\(^{38}\)

Financial journalist Matt Levine has also taken a very public position in support of HFT:\(^{39}\)

HFTs are reacting instantly to demand, avoiding being picked off by informed investors and making the stock market more efficient. HFT... undercut big banks’ gut-instinct-driven market making with tighter spreads and cheaper trading costs. Big HFTs like Knight/Getco and Virtu trade vast volumes of stock while still taking in much less money than the traditional market makers: $688 million and $623 million in 2013 market-making revenue, respectively, for Knight and Virtu, versus $2.6 billion in equities revenue for Goldman Sachs and $4.8 billion for JPMorgan. Even RBC made 594 million Canadian dollars trading equities last year. The HFT make money more consistently than the old-school traders, but they also make less of it.

---


Columbia University professors Merritt Fox and Lawrence Glosten also argue that HFT tactics aren’t really unfair—that they just favor informed investors over uninformed ones, and eliminating HFT would do the opposite.40

In Michael Lewis’ book Flash Boys, market structure expert Don Bollerman made an interesting analogy pertinent to HFT:41

On the savannah, are the hyenas and vultures the bad guys? We have a boom in carcasses on the savannah. So what? It’s not their fault. The opportunity is there.

But, HFT proponents were quieted on May 6, 2010, when markets failed and were exposed as a conflicted and rigged game in which only the connected insiders stood a chance. Every investor and market participant in the United States had been sold a lie: HFT liquidity was a blessing that lowered costs and helped investors, and it would be there in stressful markets like the market makers and specialists that it replaced.42

During the market events of May 6, 2010 (commonly referred to as the “flash crash”), the Dow Jones Industrial Average and Standard & Poor’s 500 indexes lost approximately a trillion dollars of market value— and reabsorbed the majority of that value—all within a 36-minute span. To put matters into context, before the flash crash, U.S. markets had opened and trended downward throughout the day in

response to worries regarding the concurrent debt crisis in Greece. At 2:42:46 p.m., with the DJIA down more than 300 points for the day, the equity markets began to fall rapidly, dropping an additional 600 points in five minutes for a loss of 998.5 points for the day by 2:47 p.m. 20 minutes later, by 3:07 p.m., the market had miraculously pared most of the 600 point drop.\(^{43}\) It was the third-largest intraday point swing in market history, at 1,010.14 points (eclipsed only by the intraday swings of August 25, 2015 and October 14, 2008—all after the passing of Reg NMS).\(^ {44}\)

The Securities and Exchange Commission and Commodities Futures Trading Commission released a joint report on the flash crash on October 1, 2010, in which they thoroughly described the event’s causation. According to the report,\(^ {45}\) in the hours before the nosedive, volatility was unusually high and liquidity thin, thanks to a barrage of unsettling (but not by any means catastrophic) political and economic news. The main trigger for the sudden decline, the report suggests, was a $4 billion sell order in E-Mini futures (75,000 contracts) on the S&P 500 index by Waddell & Reed, a mutual fund group. Because this automated algorithmic trade was programmed to take account of trading volume, not price or time, it was executed


unusually rapidly: in 20 minutes, instead of the several hours that would be typical for such an order. HFT firms initially helped to absorb the sell pressure, buying E-Mini contracts. Ten minutes later, however, HFT firms responded to the amount of orders flooding the market and began forcefully selling to reduce their long positions. The sell algorithm used by the mutual fund group responded to this increased volume by increasing the rate at which it fed orders into the market, creating a feedback loop with obviously negative consequences.

The feedback loop created two liquidity crises, the report says: one at the broad index level in the E-Mini, the other in individual shares. HFTs began quickly buying and reselling to each other E-Mini contracts. This “hot potato” trading generated lots of volume, but little net buying. Traditional buyers were unable (or unwilling) to step in, and the depth of the buying market for E-Minis and S&P 500-tracking ETFs fell to a mere 1% of its previous high earlier that morning.46

The second liquidity crunch, in individual stocks, began when automated trading systems used by market makers and other large liquidity providers paused, as they were designed to do when prices move beyond certain thresholds. Traders were then left to assess the risks of restarting trading. A number of participants reported that because prices had fallen precipitously across many types of securities, they “feared the occurrence of a cataclysmic event of which they were not yet aware, and that their strategies were not designed to handle,” says the report.47

Some market makers reacted to this increased risk by widening the spreads between the levels at which they would buy or sell. Others withdrew completely. Some resorted to manual trading but could not keep up with the explosion in volume. It did not help that market makers in over-the-counter markets began routing their orders to the exchanges, where they competed with other orders for immediately available but dwindling liquidity.\(^{48}\)

Nanex’s Eric Hunsader described the flash crash a little more bluntly than the SEC’s and CFTC’s report:\(^{49}\)

> At 14:42:44-- and again four seconds later, HFT market maker software detected that its inventory limit was exceeded and needed to reduce its position. It accomplished this, not by passively selling, but by immediately, and aggressively (on a scale of 1-10 of aggressive behavior, it was an 11) selling its entire inventory onto the market. This caused a sudden price drop in not only the E-Mini, but also SPY, its components, and the options for those symbols, all of which led to massive blasts of order cancellation and replacement messages: a tsunami of message traffic that reverberated through Wall Street networks at the speed of light and swamped all networks and computers in its path. This was the beginning of widespread system delays and is the reason why many other market participants either

\[^{48}\text{One Big, Bad Trade.} The Economist Newspaper, 01 Oct. 2010. Web.\]
pulled out, or sharply curtailed their activity. That, is what caused the flash crash.

The Dow Jones Industrial Average had fallen further during the flash crash than it did on news of Lehman Brothers’ collapse in 2008. But that crash elapsed over the course of a day: the flash crash spanned minutes. Circuit breakers designed to halt trading after unnatural price swings work only until 2:30 p.m.; there were no regulatory mechanisms in place to protect investors during the 2:42 p.m. crash.50

Stock exchanges agreed in retrospect to cancel trades that were 60 percent or more away from their prices at 2:40 p.m., before the selloff intensified. Transactions in 326 securities, 70 percent of which were ETFs, were broken that day. About 20,800 separate trades were voided, including many when orders hit stub quotes (orders placed well off a stock’s market price).51

But the flash crash of May 6, 2010 hasn’t been the last. According to Hunsader, by the end of 2013 there were, on average, 24 mini-flash crashes in individual stocks every day.52 See the U.S. Treasury market flash crash of October 15, 2014,53 the August 24, 2015 flash crash,54 or the aforementioned small-scale Google mini-flash crash for example.

54 Durden, Tyler. "The Stunning Comparisons Between The "Flash Crash" Of August 24, 2015 And May 6, 2010." The Stunning Comparisons Between The "Flash Crash" Of
Flash crashes and front running occur on a regular basis, and said problems stem from exchanges’ blatant abdication of their self-regulatory organization (here forth, SRO) obligations. SRO’s have the power to create and enforce industry regulations and standards. Their priority should be to protect investors through the establishment of rules that promote ethics and equality.\textsuperscript{55} And as such, exchanges shouldn't be able to sell consumer trading data, shouldn’t be able to sell access to direct feeds, and shouldn’t allow HFT and other exchange customers to colocate, and so on. SEC Commissioner Luis A. Aguilar admitted in a May 8, 2013 letter that there’s a “need for robust Commission oversight of SRO activities to enhance investor protection, maintain fair, orderly, and efficient markets, and facilitate capital formation.”\textsuperscript{56} To expand on the point of investor protection, the Mutual Fund Investor Confidence Act of 2003 illustrates that there were times in the past when regulators acknowledged the detriment to investors of practices that took advantage of slower market participants, who traditionally have been investors as opposed to [HFT].\textsuperscript{57}
In June 2001, the International Organization of Securities Commissions issued a report relevant to the conflicts of interest faced by exchanges when contemplating enforcement action against customers.\(^58\)

Due to increased pressure to generate investment returns for shareholders, a for-profit exchange may be less likely to take enforcement action against customers or users who are a direct source of [significant] income for the exchange.

The fundamental purpose for government regulation of the financial services industry is to make sure the critical personal interests of consumers are not lost in the arena of government supervision. Once regulators’ consumer protection responsibilities are satisfied, it is fair to ask how the system of regulation can be made most compatible with the demands of commercial competition without sacrificing the needs of consumers. As such, why is the marginal consumer not being protected to the full extent of the law? Why is the overall efficiency of the national market system deferring to the needs of professional traders, many of whom rarely intend to hold a position overnight?\(^59\) Why hasn’t effective legislation been passed—or for that matter, been enforced?

---


In an interview with CNBC Monday Morning, Berkshire Hathaway Vice Chairman and buy-and-hold investing figurehead Charlie Munger criticized HFT:60

[HFT is] the functional equivalent of letting rats into a granary. It does the rest of the civilization no good at all.

Regulators are getting ahead of themselves and need to assess their motives. Civil money penalties are not teaching exchanges their lesson. The SEC needs to prescribe cease and desist orders to stop predatory HFT techniques, or place exchanges under conservatorship altogether.

---

THE ORIGINS AND EVOLUTION OF HFT: A CHRONOLOGICAL TAKE

Although HFT itself didn’t go primetime until after Reg NMS was revised in 2005, so many events led to the proliferation of HFT.

In the 1960s, daily trading volume was soaring to new heights—markets saw 10 million daily share volumes for the first time (a digression: 10 million is a paltry number relative to today’s billion-plus daily trading volume). But in this era before automated systems and processing, Wall Street soon became overwhelmed. The sheer mass of stock certificates and other paperwork led to the NYSE “paperwork crisis,” which forced the NYSE to restrict trading to four days a week. In an exchange report released in February of 1971, NYSE president Robert Haack wrote, “it was a time of painful lessons.”61

Haack went on to spearhead the automation of trades through computers, which provided much-needed relief from the paperwork logjam and strongly incentivized a further increase in transactions62 (computers cost the same whether busy or idle, so the more transactions a firm processed, the lower its cost per transaction).

---

The NYSE’s failures brought light to blatant inefficiencies and needed change, spurring NASDAQ’s conception on February 4, 1971—the world’s first electronic stock market. With NASDAQ came market fragmentation; the same stock sometimes traded at different prices at different trading venues, and the NYSE ticker tape did not report transactions of NYSE-listed stocks that took place on other exchanges/over-the-counter markets. This fragmentation led to the SEC’s pursuit of a national market system, which they were able to facilitate after congress passed the Securities Acts Amendments of 1975.

Physicists began to arrive on Wall Street in the late 1970s, in the midst of a post-securities acts amendment “financial revolution.” Finance had become more complicated and risky—it required increasingly sophisticated mathematical expertise to sift through the suddenly computer-run market that came with NASDAQ. As technology became more available, there was an increased demand for code-fluent “quants” that could write computer programs to calculate bond and stock option prices.

The programs written by “quants” were largely to blame for the Black Monday crash of October 19, 1987. The Dow fell 22.61% in a single day—the largest decline in U.S. stock market history.

Paolo Pasquariello, assistant professor of finance at the University of Michigan, described program trading’s role in the Black Monday crash as such:

---

At the time, it was largely believed program trading accounted for approximately 10% of all trades... Computers in 1987 hurt the market from being too primitive and unable to predict human behavior. [Computers wouldn’t] stop and ask the question, ‘Should I sell?’ They’d simply make the trade that they were programmed to execute... Technology is only as good as the people behind it.

The scare brought by Black Monday sparked intense political and media scrutiny of markets that culminated with William Christie and Paul Schultz’s 1994 report, “Why Do NASDAQ Market Makers Avoid Odd-Eighth Quotes?” which inexplicitly revealed acts of collusion by NASDAQ. In response to the report, an antitrust lawsuit was filed against NASDAQ and the SEC took action, embarking upon a series of regulations that teed up today’s fragmented equity markets. Whether the SEC knew it at the time or not, its actions in the mid-90s led to thousands of lost jobs, the replacement of human trading by computer systems, and the rise of algorithmic trading and HFT.

In January 1997, the SEC implemented its first regulations in response to NASDAQ’s antitrust suit, the Order Handling Rules, in an attempt to add more transparency to markets. These Order Handling Rules integrated electronic

communication networks (here forth, ECNs) into the NASDAQ system, and between 1997 and 1998, the ECN business model took off.69

Initially, the SEC saw ECNs as a means to “further tap competition to augment market integrity and quality.”70 However, the Commission soon realized it needed to put a lid on the relatively lightly regulated ECNs, so it passed Regulation ATS (here forth, Reg ATS) in 1998. Reg ATS primarily made it so ECNs had to submit to either becoming Self-Regulatory Organizations (here forth, SROs) or exchanges, one or the other. Thus, ECNs were no longer able to hide orders from the public eye—a win for HFT and the inception of the “algorithmic trading revolution.”71 With more quotes being displayed, algorithmic traders were able to model order books and predict prices with greater certainty than ever before.

They NYSE did not have to worry much about Reg ATS, because of a rule (Rule 390) on the NYSE that stated members must receive permission from the exchange’s management before conducting trades on listed securities anywhere other than the trading floor.72 So, while NASDAQ was under assault by the ECNs, the NYSE was shielded underneath its own umbrella—but only for a while.

---

Acknowledging mounting pressure from the SEC ("remove all anti-competitive barriers!"), the NYSE’s board voted to rescind Rule 390 in May 2000. This move allowed for a comfortable, swift entryway to the market as a whole for algorithmic traders and HFT alike.

The next domino to fall in favor of then-embryonic HFT was the decimalization of U.S. equity markets on August 9, 2001. Meaning, instead of eight or 16 price points per dollar, stocks suddenly had 100. This move came with unintended consequences, in particular, by allowing pinging and sniffing for order flow to become much easier. Limit orders that were displayed by retail and institutional investors could easily be stepped in front of because it only took a penny to price improve, rather than an eighth of a dollar.

Saluzzi and Arnuk described the effects of decimalization on market maker economics in their book, Broken Markets:

The effect on small and mid-cap market maker economics was devastating. Margin compression drove many market makers out of business. The void would later be filled by high frequency traders who posed as market makers, but with none of the affirmative and negative obligations that specialist and NASDAQ market makers had. These obligations would require a NYSE specialist, for example, to step into the market when there was no liquidity.

---


but step back when it was abundant. No such obligations exist for today’s market makers.

Put another way, decimalization allowed HFT more price points to arbitrage, and further instigated the rise of HFT.

The technological modernization set in motion by Reg ATS, the repeal of the NYSE Rule 390, and the decimalization of US equity markets eventually antiquated the concept of floor trading in favor of virtual trading. Because the physical presence of the trader was no longer important, the costs of inducting members on exchanges naturally fell, which in turn lessened overall trading costs. Thus, membership fees were no longer of significance to exchanges, reducing the importance of mutual dependence and co-operation, which is why exchanges then took their own interests at heart and pursued routes of demutualization.76

NASDAQ was the first to demutualize, issuing shares to the public in 2000. NASDAQ's decision to demutualize drew a lot of attention to competition among individual markets. In particular, to NASDAQ's technological prowess far exceeding that the NYSE’s at the time, despite the fact that the NYSE continued to maintain 80% of all trade execution market share.77 The SEC acknowledged changes in technology over the previous decade, and decided to take extensive action (pointing a gun squarely at the head of the NYSE), publishing the first draft of Reg NMS in February 2004.

Reg NMS was billed as a way to “modernize and strengthen” the existing national market system.\textsuperscript{78} Reg NMS mandated that any order to buy or sell a stock had to go to the trading venue that offered the best price. After all, such a system had become possible with the growth of electronic trading (yet was not feasible in the days of the NYSE “paperwork crisis”).

\textit{Dark Pools: The Rise of Machine Traders and the Rigging of the U.S. Stock Market} author Scott Thompson explained Reg NMS’s effect on the NYSE and overall U.S. equity market as such:\textsuperscript{79}

The NYSE’s floor often took ten or twenty seconds to execute a trade, light-years compared with split-second trading on electronic venues…. It was a problem that had stifled [HFT] and forced them to stick with their bread-and-butter NASDAQ stocks. Reg NMS would put an end to this regime by allowing firms to “trade through” human-controlled manual markets. If [HFT] wanted to buy IBM, a NYSE-listed stock, they could go straight… to [any given ECN], because the NYSE’s human dealers were too slow. [HFT] could do this even if IBM shares were cheaper at the NYSE. Reg NMS, in essence, decreed speed was more important than price. As such, Reg NMS represented a potential deathblow to the Big Board, where price—along with human trading—was paramount.

In hindsight, while Reg NMS made some intuitive sense, it spawned a vast tangle of complications that drove markets to their current conflicted state. The

NYSE scrambled at the news of Reg NMS, and acquired notable ECNs Archipelago and Euronext as a means to ramp up its technological capabilities. To finance said acquisitions, the NYSE went public on March 8, 2006.

Reg NMS exposed every crevice of the NYSE’s 80% market share—an untapped gold mine—to HFT. And in the years following Reg NMS’s 2007 implementation, HFT took off. A new equity market had arrived: a market completely subservient to HFT. A market designed to maximize the number of collisions between ordinary investors and HFT—for the benefit of HFT, exchanges, Wall Street banks, and online brokerage firms.\textsuperscript{80} Around that market, an entire ecosystem had arisen. The amount of displayed liquidity had fallen, average trade sizes plummeted (in 2004, the average trade size of a NYSE-listed stock was 724 shares. By 2009, the average size had dropped to 268\textsuperscript{81}), and “passive investing” had become comical jargon. In fact, in the 1960s an average share of a company was held for four years, a figure that had fallen to eight months by 2000, en route to two months in 2008. Today, the average share is held for 10 seconds—a number that’s trending even further downward.\textsuperscript{82}

By early 2008, the number of new HFT entrants in the U.S. equity markets was growing steady, according to Ted Myerson, president of FTEN, a HFT service-providing company. Excess profits had attracted increased competition. In fact, at that point, so many HFT firms had launched (only $10 million to $100 million is

---


needed in startup capital), that some estimated HFT market share at the time to be as high as two-thirds of all trades.\textsuperscript{83} The amount of new entries to the line of business suggests the sheer profitability of HFT, especially amidst the financial crisis of 2007-2008.

Even with its successes, HFT as a line of business wasn't exposed to mass criticism and the public eye until July 2009, when Sergey Aleynikov, a Goldman Sachs computer programmer, faced prosecution for stealing proprietary code with the potential to “manipulate markets in unfair ways.”\textsuperscript{84}

Within days Aleynikov's arrest made national news. Reporters quickly learned from Aleynikov’s LinkedIn profile that he was a programmer at Goldman Sachs. Finally, a light had been cast on HFT—how could one man have access to code with the potential to manipulate markets?

The Google trend search results above show the term \textit{High Frequency Trading} first making waves in July 2009—around the time of Aleynikov's arrest.


HFT immediately became a topic of concern among market critics. *Forbes* magazine ran an exposé in September 2009 that called HFT, “the most wrenching, and controversial, transition in the history of U.S. securities markets.”

Still, HFT backers did whatever possible to quell public outrage and protect their profits. A representative of HFT firm GETCO wrote to the SEC on April 27, 2010:

We believe that the current national market system is performing extremely well. For instance, the performance during the 2008 financial crisis suggests that our equity markets are resilient and robust even during times of stress and dislocation.

On May 6, 2010, SEC chairman Mary Schapiro gave a speech to fellow regulators saying, “In the wake of the financial crisis, it has become a cliché that regulators cannot keep up with innovators in a market as dynamic as finance. I don’t think that’s true.”

Coincidentally, just hours later, markets were devastated by the flash crash—a crash unlike any other in financial market history. The flash crash sparked a crisis of faith in the market. Retail cash started flowing out of stock funds at an alarming pace.

---

86 McCarthy, John A. (General Counsel, Getco, LLC), letter dated 27 April, 2010 to Elizabeth M. Murphy (Secretary, Securities and Exchange Commission), Securities and Exchange Commission website, https://www.sec.gov/comments/s7-02-10/s70210-158.pdf
rate after May 6, 2010, falling every single month for the rest of the year even as the stock market rebounded and vaulted higher.⁸⁸

In the years after the flash crash, fear of another potential crash lingered—a crash where markets perhaps wouldn’t pare the majority of losses.

Ratings agency Standard & Poor’s responded to the crisis of market confidence among investors in September 2013, warning exchanges collectively of a credit downgrade. In their report, they wrote:

Stock exchanges... have become more prone to operational risk. We believe these [operational] factors could put downward pressure on [exchange credit] ratings over the next few years.⁹⁹

Standard and Poor’s fears were affirmed months later in November 2013, when the NASDAQ Options Market shut down for almost a day following a spike in volume that hampered the exchange’s ability to accept orders and distribute quotes. And although the NASDAQ Options Market isn’t a dominant player in the options market, “[it’s] a big deal in the sense that there’s another glitch,” said Themis Trading Partner Joe Saluzzi. “These are small tremors that are building and building, and eventually it will lead to a larger system outage.”⁹⁰

With that said, regulators and exchanges alike had still yet to institute real structural change to the current financial system. Flash Boys: A Wall Street Revolt author Michael Lewis made his stance on the subject clear: ”The bad incentives that

---

are the heart of the problem, which in fact were the heart of the financial crisis, like poorly incentivized financial intermediaries, are still there.”

Lewis and Saluzzi’s candid fears were affirmed on August 24, 2015, when the Dow Jones Industrial Average plunged almost 1100 points by 9:35 a.m., shortly after markets opened. The media characterized the crash as a freak occurrence, largely attributable to universal panic-selling by investors spooked at news of China’s devaluation of their currency. But in reality, the entire occurrence harked back to the May 6, 2010 flash crash. In fact, it may have been worse.

In early trading, the entire market was sent in a freefall by news of China’s devaluation of their currency. Roughly 15 minutes before the market open, at 9:15 a.m., the E-Mini S&P 500 contract reached its 5% extended-hours limit down price, and remained there until the regular trading session began at 9:30 a.m. HFTs walked away from the impending increased market unpredictability and volume, and inadvertently created a feedback loop after the commencement of regular trading that led to more volatility (the volatility index soared to levels not seen since March 2009\(^2\)), and less liquidity. In fact, liquidity levels were so unbelievably low that trough liquidity levels on May 6, 2010 were matched throughout the entire August 24, 2015 trading day.\(^3\)

---


Market volatility was so extraordinary that from 9:30-9:31 a.m., NASDAQ’s ETF QQQ had its largest one-minute price swing in history.94

One-minute range (percentage of price) of NASDAQ’s ETF QQQ since 2007. Courtesy of Nanex, LLC.

---

The market's only saving grace was the various major market indices (S&P 500, NASDAQ, DJIA) hitting their 7% limit down levels for the first time in history.95

Once the selling wave had exhausted itself, HFTs were able to front run buy order after buy order, leading to a 5% market move in less than ten minutes. For reference, the average total trading range of the E-Mini S&P 500 over an entire day for the past five years is 1.3%, and a 2% move in two minutes is considered to be a “black swan” event (a ten-plus standard deviation move under normal statistical assumptions).96 It was here that prominent HFT firm Virtu Financial Inc. had one of its most profitable days in history at the expense of most other market participants.97


---

By 12:00 p.m., the E-Mini S&P 500 had recovered from -7% to -1.5%—a prompt rebound that perhaps discouraged the SEC and its capable hands of instituting any sort of real reform—at least until the time of this paper’s publication.
CONCLUSION

In February 2015, the FCC voted to curb broadband Internet service providers’ attempts to divide the Internet into pay-to-play fast lanes for Internet and media companies that can afford it (like Walmart and Amazon) and slow lanes for everyone else. The FCC decided the Internet should be regulated as a public utility—that pay-for-play fast lanes would hamper small business growth potential and create an unfair online business environment.\(^98\) Obviously, the FCC viewed Internet access as too important to let broadband providers be the ones making the rules, and the SEC should take an analogous stance on market structure. Fast Internet service isn’t considered a discretionary good, and broadband Internet providers do not have wide latitude to price their offerings at whatever the market will bear. So why should exchanges? The government has embraced the notion that Internet is like electricity, and should therefore take a similar stance with exchanges. Under the current regulatory climate, the exchanges have realized that they must offer the fastest speed at the lowest price to their HFT clients because HFTs represented the majority of the exchanges’ business. The exchanges also realized that the HFT community was willing to spend billions of dollars per year to

get an edge on their competitors. The problem with arms races, however, is that they are zero-sum games. And like any arms race, the result is a cycle of spending that leaves everyone in the same relative position, only poorer. What’s happening with exchanges and HFT is a net drain on social welfare. Although exchanges are a hub for capitalism and free enterprise, the exchanges’ drive for profits have counterbalanced their own long-standing goals to protect the investor and help companies raise capital.

It’s time for a change: after all, what kind of a culture develops at a place that is for-profit, on Wall Street, and given carte blanche legal immunity? Exchanges need to be put into conservatorship because there are no relevant self-policing mechanisms intact, and exchanges have been abdicating their SRO obligations time after time again.

The suspicion has grown that something about how the HFT market operates—something it’s doing or not doing—may be what’s causing the trends of American decline. Although trading costs have never been cheaper, something is wrong; it’s time to take a closer look at why regulators thought replacing Wall Street with computers was a good idea. It’s time to restore the idea that exchanges are quasi-government utility-type organizations, and not bottom-line driven, publicly traded, shareholder-focused companies. It’s time to acknowledge that shareholder interests and investor interests are not always the same.

———

REFERENCES


Greg O'Connor (Compliance Manager, Wolverine Trading), letter dated April 21, 2010 to Elizabeth M. Murphy (Secretary, Securities and Exchange Commission), Securities and Exchange Commission Website, www.sec.gov/comments/s7-02-10/s70210-143.pdf


Lee, Sophia (General Counsel, IEX Group, Inc.), letter dated 13 Nov. 2015 to Mr. Brent J. Fields (Secretary, Securities and Exchange Commission), Securities and Exchange Commission website, http://www.sec.gov/comments/10-222/10222-20.pdf


McCarthy, John A. (General Counsel, Getco, LLC), letter dated 27 April, 2010 to Elizabeth M. Murphy (Secretary, Securities and Exchange Commission), Securities and Exchange Commission website, https://www.sec.gov/comments/s7-02-10/s70210-158.pdf


ACKNOWLEDGEMENTS

I would like to thank first the Sally McDonnell Honors College for providing me with the platform to write this undergraduate thesis. Also for providing financial support necessary to conduct ample research on HFT-- to travel to New York City and meet and interview the executive board members of IEX Group, among other endeavors. It was the opportunity of a lifetime.

I would also like to thank my thesis advisor, Dr. Bonnie Van Ness, for her countless hours of contribution to my thesis research and publication. To my second and third thesis readers Dr. Dwight Frink and Dr. Robert Van Ness: thank you as well for your time and support.

This paper would not have been possible without Mr. Christopher Powers and Mr. Dennis Stanek-- thank you for opening my eyes to HFT. And finally, I would like to acknowledge Mr. Eric Hunsader, Mr. Joe Saluzzi and Mr. Sal Arnuk for their enduring anti-HFT efforts, and for their willingness to respond to me on Twitter (the best research portal).