HIGH FREQUENCY TRADING AND ITS IMPACT ON MARKET QUALITY

Jonathan A. Brogaard *
Northwestern University
Kellogg School of Management
Northwestern University School of Law
JD-PhD Candidate
j-brogaard@kellogg.northwestern.edu

First Draft: July 16, 2010

November 22, 2010

*I would like to thank my advisors, Thomas Brennan, Robert Korajczyk, Robert McDonald, and Annette Vissing-Jorgensen, for the considerable amount of time they have spent discussing this topic with me; the Zell Center for Risk Research for its financial support; and the many faculty members and PhD students at the Kellogg School of Management, Northwestern University and at the Northwestern University School of Law for assistance on this paper. Please contact the author before citing this preliminary work.

Abstract

In this paper I examine the impact of high frequency trading (HFT) on the U.S. equities market. I analyze a unique dataset to study the strategies utilized by high frequency traders (HFTs), their profitability, and their relationship with characteristics of the overall market, including liquidity, price discovery, and volatility. The 26 HFT firms in the dataset participate in 68.5% of the dollar-volume traded. I find the following key results: (1) HFTs tend to follow a price reversal strategy driven by order imbalances, (2) HFTs earn gross trading profits of approximately $2.8 billion annually, (3) HFTs do not seem to systematically engage in a non-HFTr anticipatory trading strategy, (4) HFTs' strategies are more correlated with each other than are non-HFTs', (5) HFTs' trading levels change only moderately as volatility increases, (6) HFTs add substantially to the price discovery process, (7) HFTs provide the best bid and offer quotes for a significant portion of the trading day and do so strategically so as to avoid informed traders, but provide only one-fourth as much book depth as non-HFTs, and (8) HFTs may dampen intraday volatility. These findings suggest that HFTs' activities are not detrimental to non-HFTs and that HFT tends to improve market quality.

1 Introduction

This paper examines the role of high frequency trading (HFT; HFTs refers to multiple high frequency traders and HFTr for a single trader) in the U.S. equities market.[1] HFT is a type of investment strategy whereby stocks are rapidly bought and sold by a computer algorithm and held for a very short period, usually seconds or milliseconds.[2] The advancement of technology over the last two decades has altered how markets operate. No longer are equity markets dominated by humans on an exchange floor conducting trades. Instead, many firms employ computer algorithms that receive electronic data, analyze it, and publish quotes and initiate trades. Firms that use computers to automate the trading process are referred to as algorithmic traders; HFTs are the subset of algorithmic traders that most rapidly turn over their stock positions. Today HFT makes up a significant portion of U.S. equities market activity, yet the academic analysis of its activity in the financial markets has been limited. This paper aims to start filling the gap.

The rise of HFT introduces several natural questions. The most fundamental is how much market activity is attributable to HFTs. In my sample 68.5% of the dollar-volume traded involves HFT.[3] A second question that arises is what HFTs are doing. Within this question lie many concerns regarding HFT, including whether HFTs systematically anticipate and trade in front of non-HFTs, flee in volatile times, and earn exorbitant profits.[4] My findings do not validate these concerns. The third integral question is how it is impacting asset pricing characteristics. This paper is an initial attempt to answer this difficult question. The key characteristics I analyze are price discovery, liquidity, and volatility. I find that HFTs add substantially to the price discovery process, frequently provide inside quotes while providing only some additional liquidity depth, and may dampen intraday volatility.

Specifically this paper addresses the following eight questions:

High Frequency Traders' Profitability and Determinants:

1. What determinants influence HFTs' market activity?

[1]While this paper examines several important questions regarding HFT, it does not attempt to analyze flash quotes, latency arbitrage, quote stuffing, or the order book dynamics of HFTs. These are important issues but beyond the scope of this paper.

[2]A more detailed description of HFT is provided in Appendix B.

[3]I estimate that HFTs participate in 77% of the dollar-volume traded in U.S. equities.

[4]"Systematically anticipate and trade in front of non-HFTs" refers to anticipatory trading, whereby HFTs predict when a non-HFTr is about to buy (sell) a stock and a HFTr takes the same position prior to the non-HFTr. The HFTr then buys (sells) at a lower (higher) price than the non-HFTr and can turn around and sell (buy) the stock to (from) the non-HFTr at a small profit. I avoid the term "front running" as it has illegal connotations and is typically used when there is a fiduciary duty between the involved parties to the above-described activity. As the HFTs in my dataset are identified only as propriety trading firms they almost certainly have no fiduciary obligation to non-HFTs.

2. How profitable is HFT?
3. Do HFTs systematically engage in anticipatory trading?
4. Are HFTs' strategies more correlated than non-HFTs'?
5. How does HFTs' fraction of market activity change with volatility?

High Frequency Traders' Impact on Asset Prices:

6. Do HFTs add to the price discovery process?
7. Do they provide liquidity?
8. How does HFT impact volatility?

To address (1), I examine the determinants of HFTs' fraction of market activity at the stock level. Performing an OLS regression analysis, I find that HFTs trade more in large value stocks. I next analyze the factors influencing HFTs' 10-second buy and sell decisions using an ordered logit regression. I find past returns are important and so perform a logit regression analysis on past returns for different HFTs' buying/selling and liquidity providing/demanding activities. The results suggest HFTs engage in a price reversal strategy. In addition, the results are strongest for past returns that are associated with a buyer-seller order imbalance. To analyze (2) I sum HFTs' stock purchases and sales over the trading day, and at the end of each day I net to zero any outstanding shares held by HFTs, closing the positions at the average price for that day. I estimate HFTs generate around \$2.8 billion in gross annual trading profits and on a per \$100 traded earn three-fourths of a penny. The per-dollar traded profit is about one-seventh that of traditional market makers. Assuming HFTs hold capital to meet their largest one-hour inventory build-up, their investment strategies have an annualized Sharpe ratio of 4.5.

For (3) I compare the probability of seeing different trading patterns if trading was independent of trader-type history with the actual frequency of seeing such a pattern. The results do not support the claim that HFTs engage in anticipatory trading. I take a similar approach for analyzing (4). I compare the probability of seeing different trade-partners' trades, assuming trade partners are independent, with the actual frequency of seeing different trade-partners. I find that HFTs trade with each other less than expected, suggesting their strategies are more correlated with each other than are non-HFTs'. I use two techniques to address (5). I analyze HFTs' fraction of market activity across different intraday volatility levels and different 15-minute price change magnitudes. I find that HFTs reduce their liquidity supply as day-level volatility increases, increase it as 15-minute price changes increase, increase their liquidity demand as day-level volatile increases and decrease it as 15-minute price changes increase. Second, I use information shocks to test how volatility impacts HFTs' fraction of market activity. The information shocks I consider are days surrounding stocks' quarterly earnings announcements and the week of the Lehman Brothers failure. I find that higher volatility induces HFTs to participate in a larger fraction of shares traded.

I analyze (6) by implementing three Hasbrouck measures of price discovery, the permanent price impact, aggregate information variance decomposition, and information share (1991a, 1991b, 1995). Each measure indicates that HFTs are an important part of the price discovery process. For (7) I consider the price impact trades would have in a partial equilibrium setting if HFTs' limit orders were unavailable to fill marketable orders. The results suggest HFTs provide some book depth, but only a fraction of that provided by non-HFTs. I also look at whether HFTs provide liquidity to informed traders by implementing an adjusted permanent price impact measure and find that HFTs avoid trading with informed traders. Finally, I address (8) using two techniques. First, I use the September 2008 short-sale ban as a natural experiment that exogenously removed a varying portion of HFTs from affected stocks. Second, I construct hypothetical price paths in a partial equilibrium setting assuming the absence of different portions of HFTs' liquidity providing and liquidity taking activities and compare the intraday volatility between the different price paths. Both results suggest HFTs may dampen volatility.

The rest of the paper proceeds as follows: Section 2 describes the related literature. Section 3 discusses the data. Section 4 provides descriptive statistics. Section 5 analyzes HFTs' profitability and activity. Section 6 analyzes HFTs' impact on asset prices. Section 7 presents my conclusions.

2 Literature Review

A small but growing group of academic papers address questions regarding HFT. The theoretical work relating to HFT shows that, depending on the model, HFTs may improve or degrade market characteristics. Cvitanic and Kirilenko (2010) build the first theoretical model to address how HFTs impact market conditions. Their main findings are that when HFTs are present transaction prices will differ from their HFTr-free price; when a HFTr is present, the distribution of transaction prices will have thinner tails and more mass near the mean; and as humans increase their order submissions, liquidity proportionally increases. While Cvitanic and Kirilenko (2010) build a theoretical framework that directly addresses HFT, other work has been conducted to understand how market quality will be impacted when investors have different investment time horizons. Froot, Scharfstein, and Stein (1992) find that short-term speculators may put too much emphasis on short term information and not enough on stock fundamental information. The result is a decrease in the informational quality of asset prices. Vives (1995) finds that the market impact of short term investors depends on how information arrives. The informativeness of asset prices is impacted differently based on the arrival of information: "with concentrated arrival of information, short horizons reduce final price informativeness; with diffuse arrival of information, short horizons enhance it" (Vives, 1995). The theoretical work on short horizon investors suggests that HFT may either benefit or harm the informational quality of asset prices.

The empirical work relating to HFT either uses indirect proxies for HFT, studies algorithmic trading (AT), or examines the May 6, 2010 "flash crash." Regarding work on indirect proxies for HFT, Kearns, Kulesza, and Nevmyvaka (2010) show that HFTs' profits in the U.S. equities market have an upper bound of $21.3 billion from demand-taking trading. They come to this conclusion by analyzing NYSE Trade and Quote (TAQ) data under the assumption that HFTs initiate trades only when they will be profitable if held for a pre-determined time period. Hasbrouck and Saar (2010) study millisecond strategies by observing trading activity around order book adjustments. They find that such low-level activity reduces volatility and spreads, and increases book depth.

The empirical AT literature finds that AT either has no impact or reduces volatility, increases liquidity, and adds to the price discovery process.[5] Hendershott and Riordan (2009) use data from the Deutsche Boerse DAX index stocks to examine the information content in AT. In their dataset, ATs supply 50% of liquidity. They find that AT increases the efficiency of the price process and contributes more to price discovery than does human trading. Also, they find a positive relationship between ATs providing the best quotes for stocks and the size of the spread, suggesting that ATs supply liquidity when the payoff is high and take liquidity when doing so is inexpensive. The study finds little evidence of any relationship between volatility and AT. Hendershott, Jones, and Menkveld (2010) utilize a dataset of NYSE electronic message traffic to proxy for algorithmic liquidity supply and study how AT impacts liquidity. The time period of their analysis surrounds the staggered introduction of autoquoting on NYSE, so they use this event as an exogenous instrument for AT.[6] The study finds that AT increases liquidity and lowers bid-ask

[5] I am unaware of a study that provides an estimate of the fraction of U.S. equities market activity involving ATs and so I am unable to determine what fraction of ATs are HFTs.

[6] "Autoquoting" is a technology put in place in 2003 by the NYSE to assist specialists in their role of displaying the best bid

spreads. Chaboud, Hjalmarsson, Vega, and Chiquoine (2009) look at AT in the foreign exchange market. Like Hendershott and Riordan (2009), they find no causal relationship between AT and volatility. They find that human order flow is responsible for a larger portion of the return variance, which suggests that humans contribute more to the price discovery process than do algorithms in currency markets. Gsell (2008) takes a simple algorithmic trading strategy and simulates the impact it would have on markets if implemented. He finds that the low latency of algorithmic traders reduces market volatility and that the large volume of trading increases AT's impact on market prices.[7]

Studies on the May 6, 2010 flash crash find that HFTs did not ignite the downfall, but they disagree as to whether HFTs enhanced the magnitude of the decline. The joint SEC and CFTC (September 30, 2010) official report describes in detail the events of that day. It finds that HFTs initially provided liquidity to the large sell order that was identified as the cause of the crash, but that after fundamental buyers withdrew from the market, HFTs, and all liquidity providers, also stopped trading and providing competitive quotes. Kirilenko, Kyle, Samadi, and Tuzun (2010) provide additional insight to the activities of different traders on May 6, 2010 in the E-mini S&P 500 stock index futures market. They conclude that while HFTs did not ignite the flash crash, their activities enhanced market volatility. Easley, de Prado, and O'Hara (2010) argue that order flow toxicity, a term referring to a higher likelihood of a trade resulting in a loss, was the cause of the flash crash. The order flow toxicity signalled days in advance the increased likelihood of a liquidity-induced crash.

This paper adds to the literature in a variety of ways. It is the first to document the prominence of HFT in the U.S. equities market. It estimates HFTs' profits, documents their trading strategy, and tests for anticipatory trading by HFTs. It also examines HFTs' order book depth, analyzes which traders supply liquidity to informed traders, and uses the 2008 short sale ban as a natural experiment to study HFTs' impact on volatility. Finally, it applies methodologies previously used in the AT literature to study HFTs' trading strategy correlation and their contribution to price discovery.

---

and offer. It was implemented under NYSE Rule 60(e) and provides an automatic electronic update, as opposed to a manual update by a specialist, of customers' best bid and offer limit orders.

[7]"Latency" in HFT nomenclature refers to the time it takes to receive, process, and respond to market information. Appendix B provides more detail.

3 Data

3.1 Standard Data

The data in this paper come from a variety of sources. I use CRSP data when considering daily data not included in the HFT dataset. Compustat data are used to incorporate stock characteristics in to the analysis. Trade and Quote (TAQ) data are used to incorporate additional intraday information. Finally, I use the CBOE S&P 500 Volatility Index (VIX) to capture market-wide volatility.

3.2 High Frequency Trading Data

The unique dataset (the HFT dataset) used in this study contains trades, inside quotes, and the order book for 120 U.S. stocks, whose symbols, company names, and market capitalization groups are listed in Table A-1.[8] The Trade data contain all trades that occurred on the Nasdaq exchange during regular trading hours in 2008, 2009, and 02/22/2010 - 02/26/2010, excluding trades that occurred at the opening, closing, and during intraday crosses.[9] The trades are millisecond timestamped and identify what type of trader (HFTr or non-HFTr) is supplying liquidity and demanding liquidity. By supplying (or providing) liquidity I mean the limit order standing on the order book that was hit by a marketable order (i.e., a market order or a more recent limit order taking the opposite side of the transaction and that crossed prices). The liquidity demander (or taker or initiator) is the market participant who enters the marketable order.[10] The Quote data are from 02/22/2010 - 02/26/2010 and include the best bid and ask that is being offered at all times by HFTs and non-HFTs. The Book data are from the first full week of the first month of each quarter in

[8]The stocks were selected by Terrence Hendershott and Ryan Riordan.

[9]Nasdaq offers opening, closing, and intraday crosses. A cross is a two-step batch order whereby in the first step Nasdaq accumulates all outstanding orders entered into the cross system and sets a preliminary transaction price. If there is an imbalance in orders it displays the price to dealers and they can submit orders. Given the final number of orders, the transaction price is set.

[10]As there are "flash trades" in the data set, let me briefly discuss what they are and how they show up in the data. Flash quoting is a technology that Nasdaq, BATS, and DirectEdge implemented to facilitate trading on their exchanges. Nasdaq ran the program from April 2009 to July 2009. A market participant who was going to enter a market order had the option of flashing his quote. For instance, if person A put in a market buy order on Nasdaq and selected for the order to "flash" if not fillable on Nasdaq and Nasdaq did not have the national best offer, then before Regulation NMS required Nasdaq to send the order to the exchange with the best offer price, the following events would occur. Person B, likely a HFTr, would be shown the market order for 20-30 milliseconds and in that time could place an offer matching or bettering the national best offer. If person B did not provide the offer, the trade would route to the other exchange. If person B did respond to the flashed quote, then the trade would execute on Nasdaq between persons A and B. In my data this would show up as person A being the liquidity provider (think of the flashable market order as a 30 millisecond limit order that converts to a market order) and person B would be the liquidity taker, however, the price the transaction occurred at would be at the offer, even though the liquidity taker was selling.

2008 and 2009, 09/15/2008 - 09/19/2008, and 02/22/2010 - 02/26/2010. They include the ten best bids and offers available on Nasdaq's order book, the type of trader that supplied the order, and whether the order was displayed or hidden.

The HFT dataset distinguishes messages from 26 firms that were identified by Nasdaq as engaging primarily in high frequency trading. This was determined by Nasdaq based on known information regarding the different firms' trading styles and also on the firms' website descriptions. The characteristics of firms identified as being HFTs are the following: They engage in proprietary trading; that is, they do not have customers but instead trade their own capital. They use sponsored access providers whereby they have access to the co-location services and can obtain large-volume discounts and reduce latency. They tend to switch between long and short net positions several times throughout the day, whereas non-HFT firms rarely do so. Orders by HFT firms are of a shorter time duration than those placed by non-HFT firms. Also, HFT firms normally have a lower ratio of trades per orders placed than non-HFT firms.

Some firms that others may define as HFTs are not labeled as such in the dataset. Potential HFT firms are excluded if they fall into one of the following categories: brokerage firms that provide direct market access and other powerful trading tools to their customers; proprietary trading firms that are a desk of a larger, integrated firm, like a large Wall Street bank; independent firms that are engaged in HFT activities, but route their trades through a Market Participant ID (MPID) of a non-HFT firm;[11] small firms that engage in HFT activities.

4 Descriptive Statistics

I compare the HFT dataset to Compustat and show it is representative on many dimensions, but on average contains stocks with larger market capitalizations. I also compare it to TAQ and find it to be similar on most measures. Next, I provide summary statistics showing the fraction of market activity involving HFTs. Finally, I provide summary statistics for HFTs' fraction of time at the best bid or offer and frequency of quote changes.

[11] MPIDs are necessary for those firms that directly interact with Nasdaq's computer systems and for those required to have them by the Financial Industry Regulatory Agency (FINRA).

4.1 The High Frequency Trading Dataset Sample Characteristics

Table A-2 Panel A reports the 120 stocks in the HFT dataset compared to the Compustat database. It includes the market capitalization, market-to-book ratio, industry, and listing exchange summary statistics and provides t-statistics for the differences in means. The Compustat stocks consist of all stocks in the Compustat database that were listed on either Nasdaq or the NYSE and that had the requisite data available. The data for both the Compustat and HFT stocks are for fiscal year end on December 31, 2009. If a stock's year-end is on a different date, I use the fiscal year-end that is most recent, but prior to December 31, 2009.

Whereas the average Compustat stock has a market capitalization of $3.37 billion, the average HFT dataset stock is $17.6 billion and the difference is statistically significant. The HFT dataset includes stocks with market capitalizations ranging from $80 million to $175.9 billion. The average market-to-book ratio for the HFT dataset is 2.65 and 13.81 for the Compustat database. This difference is not statistically significant. On Industry the HFT dataset matches Compustat in seven industries, but it overweights Manufacturing and underweights Energy and Other. The industries are determined based on the Fama-French 10 industry designation from SIC identifiers. Finally, half the HFT dataset stocks are listed on the NYSE and half on the Nasdaq exchange. This is statistically different from Compustat.[12] The HFT dataset provides a robust variety of industries, market capitalizations, and market-to-book values.

Table A-2 Panel B describes the market characteristics of the HFT dataset stocks and the NYSE and Nasdaq stocks with the requisite data available in the TAQ database for 02/22/2010 - 02/26/2010. The reported statistics include the quoted half-spread, stock price, bid size, offer size, daily volume traded, number of trades, and average trade size. The average half-spread in the HFT database is $.07 while in TAQ it is $.13 and the difference is statistically significant. The average HFT dataset number of trades is 3,090 while TAQ's is 1,371 trades and the difference is statistically significant. The HFT database average bid size is 23,880 shares and the average offer size is 24,240 shares. These values are larger in the HFT dataset but are not statistically significantly different from the TAQ database. Finally, the average trade size in the HFT dataset is 208 shares while in TAQ it is 243 shares and the difference is statistically significant.

[12] To clarify, the HFT dataset comes from the Nasdaq exchange. Of the 120 stocks in it, 50% are listed on Nasdaq and 50% are listed on the NYSE. The listing exchange does not determine where trading occurs. Different firms can route their orders to different exchanges, and under Regulation NMS that exchange can execute the order if it is displaying the national best bid and offer (NBBO); otherwise it is required to route the order to the exchange that is offering the NBBO.

4.2 High Frequency Trading's Prominence in Trading Activity

Table 1 looks at the prevalence of HFT in the U.S. equities market. The reported results are for the aggregate day level fraction of trading involving HFTs. Panel A measures the fraction of dollar-volume activity involving HFTs, Panel B measures the fraction of trades involving HFTs, and Panel C measures the fraction of shares traded involving HFTs. Within each panel are three categories. The first, HFT-All, reports the fraction of activity where HFTs either demand liquidity, supply liquidity, or do both. The second, HFT-Demand, reports the fraction of activity where HFTs demand liquidity. The third, HFT-Supply, shows the fraction of activity where HFTs supply liquidity. Within each category I report the findings by stock size, with each group having 40 stocks and the row Overall reporting the unconditional results. The reported summary statistics include the mean, standard deviation, minimum, 5th percentile, 25th percentile, median, 75th percentile, 95th percentile, and maximum fraction of market activity involving HFTs.

Panel A shows that HFTs are involved in 68.5% of all dollar-volume activity in the sample and that their involvement varies from 60.4% to 75.9%. They demand liquidity in 42.8% of dollar-volume activity and supply it in 41.1%. Panel B shows that HFTs' involvement in the fraction of trades is even larger at 73.8%, which implies they are involved in smaller dollar-volume trades. They demand liquidity in 43.6% of all trades and supply it in 48.7%. Panel C reports that HFTs' involvement in the fraction of shares traded is 71.6%, and they demand liquidity in 38.4% of all shares traded and supply it in 47.3%. In each panel and category HFTs' fraction of activity increases with stock size.

Many HFTs act as market makers even if they are not registered as such. Of the 26 HFT firms, some mainly provide liquidity while others mainly take it. Although I cannot observe it directly in the data, conversations with market participants indicate that even registered market makers will take liquidity at times. In traditional microstructure models, the observation that HFTs take as much liquidity as they provide would preclude them from being market makers. However, empirical papers such as Chae and Wang (2003) and Van der Wel, Menkveld, and Sarkar (2008) find that market makers frequently take liquidity, make informational-based trades, and earn a significant portion of their profits from non-liquidity providing activities.

4.2.1 High Frequency Trading's Fraction of the Market - Time Series

Figure 1 presents time series graphs of HFTs' fraction of market activity in addition to Table 1's summary statistics, as the graphical analysis shows whether HFTs' market participation is abnormally high or low during key market events, such as the Lehman Brothers collapse. The three graphs relate to Table 1 Panels A - C's analysis of HFTs' activity in terms of dollar-volume, trades, and shares. In each graph are three lines that correspond to the three categories in each of Table 1's Panels, HFT-All, HFT-Demand, and HFT-Supply.

HFTs' involvement level fluctuates minimally and there are no apparent abnormal withdrawals or increases during key market events. The correlation coefficient between the VIX, the standard measure of market-wide expected volatility, and the different measures of HFTs' fraction of market activity is strongly positive for the different measures of HFT market participation. The dollar-volume correlation with VIX is: All 0.71, Supply 0.35, Demand 0.72. Along all measures, HFTs' fraction of market activity fluctuates +/- 8% on a day-to-day basis. Especially of note, there is no abnormally large drop, or increase, in HFT in September 2008, when the U.S. equities market was especially volatile.

4.3 High Frequency Trading's Prominence at the Inside Quotes

In this section I provide summary statistics on the amount of time HFTs supply the inside bid or offer. I report the average fraction of the day HFTs provide the inside bid or offer for a stock. When HFTs and non-HFTs both offer the best bid or offer, I include the time HFTs are at the inside quotes. Table 2 shows the results.

Table 2 reports statistics on HFTs' fraction of calendar time at the best quotes, fraction of tick time at the best quotes, and fraction of quote revisions. Panel A contains statistics regarding HFTs' fraction of calendar time at the best bid and offer. Panel B and C divide up the observations so that Panel B contains summary statistics for days when a stock's spread is lower than its average spread, and Panel C reports summary statistics when its spreads are higher than its average spread. Panels D - F report summary statistics based on tick time, weighting each observation equally regardless of the calendar time of the quote. Panel D relates to Panel A, and Panels E and F relate to Panels B and C, respectively. Panel G reports summary statistics on the fraction of total quote changes originating from HFTs. In each panel I report the summary statistics based on stock market capitalization with each group containing 40 stocks, and also the total row that reports the unconditional summary statistics.

The results show how frequently HFTs provide inside quotes. In all of the panels, HFTs' fraction of time at the inside quotes increases with stock size, suggesting that HFTs are more competitive in their quotes for larger stocks. Panel A reports that HFTs provide the best bid and offer quotes 65.3% of the calendar time. The results when conditioning on spread size show how the spread impacts HFTs' quoting activity for different size stocks. While there is little change in HFTs' fraction of time at the inside quotes for small and medium size stocks, for large stocks HFTs increase their inside quote time from 80.5% to 85.7%. This is consistent with HFTs attempting to profit when the price of liquidity is high as is found in Foucault and Menkveld (2008) and Hendershott and Riordan (2009).

The unconditional tick time results reported in Table 2 Panel D are similar to those found in Panel A, but the spread-conditioned results differ in an important way. For small and medium size stocks, on days when a stock's spread is lower than average, HFTs spend more time at the best bid and offer than during high spread days. This suggests that when the liquidity premium is low, HFTs increasingly change their quotes. For large stocks the spread appears not to have an impact. Table 2 Panel G provides summary statistics on HFTs' quote revision activity. It reports the fraction of total revisions that were due to HFTs.[13] Quote cancelations and revisions have been found to have net economically significant benefits by reducing the non-execution cost that would otherwise occur (Fong and Liu, 2010). Panel G shows that HFTs are less active than non-HFTs in their quote revisions for small stocks, but their revisions increase with stock size so that they initiate 68.4% of all quote revisions for large stocks.

5 High Frequency Traders' Profitability and Determinants

This section analyzes HFTs' trading activity and their profitability in terms of six questions: (1) What determinants influence HFTs' fraction of trading in a stock? (2) What factors drive HFTs' decision to buy or sell? (3) How profitable is HFT? (4) Do HFTs systematically engage in anticipatory trading? (5) Are their strategies more correlated than non-HFTs'? (6) How active are HFTs in volatile markets?

The answers clarify the role of HFTs in the U.S. equity markets. I find that HFTs trade in large market capitalization stocks, with lower market-to-book ratios, larger spreads, and more depth. Their buy and sell decisions depend heavily on past returns interacted with order imbalances and suggest HFTs engage in a

[13]Looking at changes in the inside quotes is only a proxy for quote revisions and cancelations as a quote change will also occur when a trade is executed that takes the inside quote's standing order.

price reversal strategy. I estimate that HFTs earn gross annual profits of approximately $2.8 billion. There is no evidence that HFTs engage in anticipatory trading. HFTs' strategies are more correlated than non-HFTs'. They tend to decrease their supply of liquidity moderately and increase their demand for liquidity as volatility increases at the day level, but during extreme 15-minute price movements they increase their supply of liquidity and decrease the amount they demand. Finally, I find evidence indicating that higher volatility drives HFTs to increase their trading participation. The results in this section suggest that HFTs' activities are not detrimental to non-HFTs.

5.1 Day-Level High Frequency Trading Determinants

In this section I analyze the factors that influence HFTs' fraction of trading across stocks and time. The time series graphs in Figure 1 show that HFTs' fraction of market activity varies over time, and the summary statistics in Table 1 show that HFTs' fraction of market activity varies across stocks as well. To study the factors influencing HFTs' fraction of trading across stocks and time, I perform an OLS regression analysis using their activity in each stock on each day as an observation. I find that HFTs are involved in a larger fraction of trading in stocks with larger market capitalizations and lower market-to-book ratios.

I run the following regression:

$$\begin{aligned}\mathrm{H}_{i,t} =& \alpha + \mathrm{MC}_i * \beta_1 + \mathrm{MB}_t * \beta_2 + \mathrm{VIX}_i * \beta_3 + \sigma_{i,t} * \beta_4 + \mathrm{SP}_{i,t} * \beta_5 + \\ & \mathrm{DEP}_{i,t} * \beta_6 + \mathrm{TS}_{i,t} * \beta_7 + \mathrm{NV}_{i,t} * \beta_8 + \mathrm{AC}_{i,t} * \beta_9 + \epsilon_{i,t},\end{aligned}$$

where $\mathrm{H}_{i,t}$ is the fraction of shares traded involving a HFTr in stock i on day t, MC is the log market capitalization as of December 31, 2009, MB is the market to book ratio as of December 31, 2009, which is Winsorized at the 99th percentile, VIX is the S&P 500 Chicago Board of Exchange Volatility Index (scaled by 10^{-3}), σ is the ten-second realized volatility summed up over the day (scaled by 10^{-5}), SP is the average time-weighted dollar spread between the bid and offer (scaled by 10^{-1}), DEP is the average time-weighted depth available at the inside bid and ask in dollars (scaled by 10^{-3}), TS is the average dollar-volume size of a non-HFTr-only trade (trades where non-HFTs both supplied liquidity and demanded it) (scaled by 10^{-5}), NV is the dollar-volume of non-HFTr-only transactions, normalized by market capitalization (and scaled by 10^{-6}), and AC is the absolute value of a one-period autoregressive process (AR(1)) analyzed at ten-second intervals (scaled by 10^{-2}). Standard errors are clustered by stock.

These variables should capture different stock and time characteristics that may influence HFTs' behavior. Market capitalization and market-to-book are used to incorporate established asset pricing information of importance; VIX to capture overall market volatility; stock-level volatility to measure stock-specific price fluctuations; spread, depth, non-HFTr trade size, and non-HFTr dollar-volume traded as measures of liquidity; autocorrelation to detect whether HFTs are more active in stocks that are more predictable. Some variables in this regression analysis may be endogenously determined, which can bias the coefficients. Therefore, I perform the same regression analysis with only the most plausibly exogenous variables: market capitalization, VIX, and the market-to-book ratio. The analysis is done for all trading days in the HFT dataset.

Table A-3 reports the results. Columns (1) - (6) show the standardized coefficients, and columns (7) - (12) show the regular coefficients.[14] I perform the regression with the dependent variable capturing the fraction of shares involving HFT in any capacity (All), as the liquidity taker (Dem.), and as the liquidity supplier (Sup.).

The results of the full regression analysis show how important and positive is the relationship between HFTs' fraction of market activity and market capitalization. The market-to-book variable has a negative coefficient, suggesting HFTs prefer value stocks. The VIX coefficient implies HFTs increase their demand for liquidity in volatile markets and decrease their supply. The stock volatility coefficient is only statistically significant for HFT-demand and it is negative. The spread coefficient is only statistically significant for the HFT-all regression and is positive. The depth coefficient is negative for HFT-demand and positive for HFT-supply. The average non-HFTr trade size coefficient is positive for HFT-demand and negative for HFT-supply. The Non HFT dollar-volume traded is not statistically significant. Finally, the autocorrelation coefficient is only statistically significant for HFT-supply and is positive. The restricted regression does not change the sign or statistical significance of any of the included regressors and has minimal impact on the magnitude of their coefficients.

[14] Standardized coefficients are calculated by running an OLS regression analysis after all the variables have been demeaned and have been scaled by their standard deviations. The standardized coefficients' interpretation is that a one standard deviation change in an independent variable is expected to change the dependent variable by β standard deviations. The regressors underlying scale of units are irrelevant due to the pre-regression scaling. Thus, the larger the standardized coefficient, the larger the impact that variable has on the dependent variable.

5.2 10-Second High Frequency Trading Determinants

This section examines the factors that influence HFTs' buy and sell decisions. I begin by testing a variety of potentially important variables in an ordered logistic regression analysis. The results show the importance of past returns. I carry out a logistic regression analysis distinguishing the dependent variables based on whether the HFTr is buying or selling and whether the HFTr is providing liquidity or taking liquidity. Finally, I include order imbalance in the logit analysis and find that the interaction between past order imbalance and past returns drives HFTr activity and is consistent with HFTs engaging in a short term price reversal strategy.

5.2.1 Analysis of Potential Determinants

I analyze the decisions a HFTr must make at every moment: Does it buy, sell, or do nothing. I model this setting by using a three-level ordered logistic regression analysis and consider the activities by the 26 HFT firms together as a representative HFT agent. The ordered logit analysis is such that the lowest decision is to sell, the middle option is to do nothing, and the highest option is to buy. The approach is similar to that used in Hausman, Lo, and MacKinlay (1992) except that in this case the dependent variable is a ten-second buy/do nothing/sell decision and not a transaction-by-transaction price process. I group HFT activity into ten-second bins throughout the trading day.[15]

Each ten-second interval for each stock is an observation in the ordered logit regression:

$$\begin{aligned}\text{HFT}_{i,t} = \alpha \quad & +\beta_{1-11} * \text{Ret}_{i,t,0-10} + \beta_{12-22} * \text{SP}_{i,t,0-10} + \beta_{23-33} * \text{DEPB}_{i,t,0-10} + \\ & \beta_{34-44} * \text{DEPA}_{i,t,0-10} + \beta_{45-55} * \text{NT}_{i,t,0-10} + \beta_{56-66} * \text{NV}_{i,t,0-10} + \epsilon_{i,t},\end{aligned}$$

where HFT is -1 during the ten-second period t if HFTs were, on net, selling shares of stock i, 0 if HFTs performed no transactions or bought as many shares as they sold, and 1 if, on net, HFTs purchased shares. SP is the average time weighted spread, where spread is the best offer price minus the best bid price, DEPB is the average time-weighted best bid depth in dollars. DEPA is the average time-weighted best offer depth in dollars. NT is the number of non-HFTr trades that occurred, and NV is the non-HFTr dollar-volume of shares exchanged. Stock fixed effects are implemented and standard errors are clustered by stock.

[15] I carried out the same analysis using other time intervals including 250-milliseconds, 1-second and 100-second periods. The results from these alternative time horizons are economically similar.

I include the contemporaneous and lagged values for each of the explanatory variables to capture how the evolution of the stock characteristics impact HFTs' trading decisions. Each explanatory variable has a subscript $0-10$. This represents the number of lagged time periods away from the event occurring in the time t dependent variable. Subscript 0 represents the contemporaneous value for that variable. Thus, the betas represent row vectors of 1x11 and the explanatory variables column vectors of 11x1. I select the explanatory variables to test which trading strategy HFTs are implementing. It could be that they engage in a momentum strategy, price reversal strategy, spread-premium strategy, or intertemporal volume smoothing strategy.

Table A-4 shows the results for the marginal effects at the means for the probability of a HFTr buying stock i at time t. There is sporadic statistical significance in the DEPA, TS, NT and NV variables, and no statistical significance in SP. The coefficients of importance are the returns, which have strong statistical significance throughout the lagged time periods. The negative coefficient is interpreted as when prices have been falling in the past there is an increased probability that HFTs will buy the stock now. The results suggest that past and contemporaneous spreads, depth, and volume are not primary factors in HFTs' trading decisions. Alternatively, these variables may be primary factors but, due to endogeneity, are not captured by the logit regression analysis. Nonetheless, the results strongly support a price reversal strategy.

5.2.2 Lagged Returns' Importance Based on Trade and Liquidity Type

I further analyze the main finding from the previous section, that HFTs engage in a price reversal strategy, by distinguishing HFTs' buy and sell activities based on whether they are supplying liquidity or taking it. In this analysis I run six different logit models with the dependent variable being one of the following: HFTs buying, HFTs buying and supplying liquidity, HFTs buying and demanding liquidity, HFTs selling, HFTs selling and supplying liquidity, and HFTs selling and demanding liquidity.

Table A-5 shows the results for the marginal effects at the means for the probability of a HFTr meeting the dependent variable's criteria for stock i at time t. The logit models incorporate stock fixed effects and standard errors are clustered by stock. The first three columns report the logistic regression analyses with HFTs, on net, buying, with varying liquidity requirements. The results maintain the price reversal strategy finding and the buy-supply results show the price reversal strategy being stronger when HFTs are deciding to supply liquidity. The last three columns report the logistic regression analyses with HFTs, on net, selling. Again, the results support the price reversal strategy findings, but for the sell results, it is stronger when HFTs demand liquidity. While the magnitude and statistical significance vary across the different HFTs' decisions, they are each consistent with a price reversal strategy.[16]

5.2.3 Order Imbalance and Lagged Returns

I further refine the price reversal strategy analysis to incorporate the influence of short-term order imbalances. I find that using past returns and past order imbalances more precisely identify the factors influencing HFTs' buying and selling activity. In the recent volatility-volume literature Chan and Fong (2000) and Chordia and Subrahmanyam (2004) find that taking into account order imbalances significantly reduces the remaining volatility-volume relationship. This suggests order imbalance plays an important role in price fluctuations. Given HFTs' short investment horizon I test whether short-term order imbalances are an important factor in their trading strategy.

I rerun the logistic models from the previous section but include three independent variables: past returns, a past order-imbalance dummy, and their interaction. As in the previous regression analysis, $\text{HFT}_{i,s}$ takes on one of six definitions: HFTs buying, HFTs buying and supplying liquidity, HFTs buying and demanding liquidity, HFTs selling, HFTs selling and supplying liquidity, and HFTs selling and demanding liquidity.

$$\text{HFT}_{i,t} = \alpha + \text{Ret}_{i,1-10} * \beta_{1-10} + \text{OIB}_{i,1-10} * \beta_{11-20} + \text{OIB}_{i,1-10} * \text{Ret}_{i,1-10} * \beta_{21-30} + \epsilon_{i,t}, \qquad (1)$$

where $\text{Ret}_{i,1-10}$ is the return for stock i in period s, s is the number of 10-second time periods prior to the time t, $\text{OIB}_{i,1-10}$ is a dummy variable derived from the order imbalance that equals 1 if $\text{OIB}_{i,s} = \frac{\text{Buy Initiated Shares}_{i,s} - \text{Sell Initiated Shares}_{i,s}}{\text{Shares Outstanding}_{i,s}}$ is ≤ 0 for the buy regressions, and ≥ 0 for the sell regressions.

Table 3 reports the results for the marginal effects at the means for the probability of a HFTr meeting the dependent variable's criteria for stock i at time t. The logit models incorporate stock fixed effects and standard errors are clustered by stock. The first three columns look at HFTs' decision to buy. In all three columns the interaction term between order imbalance and returns are strongly significant, whereas

[16] I carried out the same exercise for non-HFTs buying and selling decisions and found the opposite results for this group. Their aggregate short-horizon activity suggests they engage in a momentum trading strategy. Interpreting the findings in this section as an investment strategy for HFTs is reasonable, as it is from only 26 firms and by definition HFTs make their purchase and sale decisions based on short-term information. The same is not true for non-HFTs because many of the firms and individuals in this group are not basing their buy and sell decisions on high frequency price fluctuations.

the return-only coefficients are of smaller magnitude and less statistically significant than before. The interpretation of the negative interaction term is that, when there is an order imbalance in favor of selling, HFTs are more likely to buy after prices have declined. In addition, HFTs' probability of buying only increases slightly when prices have declined, but there is no selling order imbalance.

The last three columns report the results for the sell logit regressions. Like the buy logit results, in all three regressions the interaction term between order imbalance and returns are strongly significant, whereas the return-only coefficients are of a smaller magnitude and less statistically significant than before. Recall that OIB is redefined for the Sell logit models so as to make the interpretation easier. In the Sell regressions OIB $= 1$ if the order imbalance is ≥ 0. The interpretation of the positive interaction term is that, when there is an order imbalance in favor of buying, HFTs are more likely to sell after prices have increased. In addition, the reduced statistical significance and magnitude of the return-only coefficients suggest that HFTs' probability of selling does not increase as much when prices have increased, but there is no buying order imbalance.[17]

5.3 The Profitability of High Frequency Trading

In this section I estimate the profits HFTs earn from their U.S. equities trading activity. I find they earn gross profits of \$2.8 billion annually. This equates to three-fourths of a penny for every \$100 traded, a seventh as much as traditional market makers. After estimating the capital required to carry out their trading activities, I calculate that HFTs obtain a pre-expense annualized Sharpe ratio of 4.5.

The HFT dataset allows for an estimate of the profitability, but with limitations. First, the HFT dataset contains only 120 stocks out of the several thousand listed on NYSE and Nasdaq.[18] Second, I can only observe trades occurring on Nasdaq.[19] This impacts my ability to determine precisely the level of HFT activity and also the inventory held by HFTs. Finally, I can only observe HFT firms' activities in the aggregate and so cannot calculate the profitability between the firms.

I first calculate the profitability of the HFT firms in the HFT dataset and subsequently I extrapolate

[17] I carry out the same exercise for non-HFTs' buying and selling decisions. I find a similar price reversal strategy in the interactive term. However, the Ret variables are also statistically significant and positive (negative) for the buy (sell) decision, suggesting that a group of non-HFTs are more likely to buy when the there are more buy than sell orders and returns have been positive.

[18] The 120 stocks have a combined market capitalization of \$2.1 trillion at the end of 2009, a fraction of Compustat stocks' combined market capitalization of \$17.2 trillion.

[19] On average 20-30% of trading activity occurs on Nasdaq.

from this to determine HFTs' overall U.S. equities trading profits. I assume all HFT actions come from one representative HFTr. To determine the daily profitability of the HFTs, I sum the amount spent on purchasing shares and the amount received from selling shares for the trades in the HFT dataset. I assume that HFTs close the day with a net zero position in each stock.[20] In the HFT dataset, on most days HFTs do not end the day with an exact net zero position in each stock. I correct for this by assuming any excess shares were traded at the mean stock price for that day.

The daily profitability for each stock is calculated as follows:

$$\begin{aligned}\text{Profit} &= \sum_{t=1}^{T} \left[\mathbf{1}_{\text{Sell,t}} * \text{Price}_t * \text{Shares}_t - \mathbf{1}_{\text{Buy,t}} * \text{Price}_t * \text{Shares}_t\right] + \\ &\quad E(\text{Price}) * \sum_{t=1}^{T} \left[\mathbf{1}_{\text{Buy,t}} * \text{Shares}_t - \mathbf{1}_{\text{Sell,t}} * \text{Shares}_t\right],\end{aligned}$$

where $1_{\textbf{Sell}}$ is a dummy indicator equal to one if HFTs sold a stock in transaction t and zero otherwise, $1_{\textbf{Buy}}$ is similarly defined for HFTs buying, Price_t is the price at which transaction t occurred, Shares_t is the number of shares exchanged in transaction t. The second term in the equation corrects for the end-of-day inventory balance. Summing up the Profit for each stock on a given day results in the total HFT profitability for that day. The result is that on average HFTs make $298,000 per day from their Nasdaq trades in these 120 stocks. Figure 2 displays the time series of HFT profitability per day. The graph is a five-day moving average of HFTs' daily profitability for the 120 stocks in the HFT dataset. Profitability varies substantially over time, even after smoothing out the day-to-day fluctuations.[21] HFTs' profitability per dollar traded is $.000072.[22] I estimate HFTs Sharpe ratio assuming HFTs require capital to be able to fund the maximum inventory imbalance that occurs in any one hour period. In the HFT dataset this means HFTs require $117 million in capital ($4.68 billion when considering the entire U.S. equities market). This implies HFTs have an annualized Sharpe ratio of 4.5 before expenses.

I extrapolate from the within-sample estimate of profitability to the entire U.S. equities market. To do so, I estimate the fraction of shares involving HFTs using the exogenous regression coefficient estimates

[20]This assumption is consistent with conversations I have had with HFTs who say they avoid holding positions overnight. In addition, it is consistent with HFTs balancing their inventory throughout the day so as never to accumulate a large long or short position, which can be observed in the HFT dataset from the fact that HFTs typically switch between being long and short in a stock several times a day.

[21]The most profitable day in the figure is September 29, 2008. On this day the House of Representatives failed to pass the initial TARP bill. The least profitable day is January 23, 2009. I am unaware of any significant event occurring on this date.

[22]I calculate this by taking the average daily profit of HFTs from trades in the HFT dataset and divide it by the average daily HFT - non-HFT dollar-volume traded ($0.298 million/$4,145.5 million).

in the HFTs' fraction of the market regression found in Table A-3, column 10.[23] I multiply the estimated HFTs' fraction of market involvement by the fraction of dollar-volume where HFTs are not trading with each other.[24]

$$\widehat{\text{HFT}}_{i,t} = 0.78[-0.127 + \text{Market Cap}_{i,t} * 0.080 - \text{Market/Book}_{i,t} * 0.015 + \text{VIX}_t * 0.582],$$

where Market Cap is the log of the daily shares outstanding of stock i multiplied by the closing price of stock i, Market/Book is the ratio of stock i's market value divided by the Compustat book value based on the most recent preceding quarterly report, Winsorized at the 99th percentile, and VIX is the S&P 500 Chicago Board of Exchange Volatility Index. $\widehat{\text{HFT}}$ is calculated for each stock on each day. I multiply $\widehat{\text{HFT}}$ by the dollar-volume traded for each stock on each day in 2008 and 2009, and I multiply this value by the profit per dollar traded of \$.000072 estimated in the in-sample analysis. I am assuming that the profit per dollar traded is the same across stocks. I divide the total sum by two as the data covers a two-year period:

$$\text{HFT }\widehat{\text{Annual}}\text{ Profit} = \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\widehat{\text{HFT}}_{i,t} * \text{DVolume}_{i,t} * 0.000072\right] \tag{2}$$

The result of this calculation is that HFTs gross profit is approximately \$ 2.8 billion annually on trading of \$39.3 trillion (\$50.4 trillion when including HFT-to-HFT trades, which implies they are involved in 77% of all U.S. equities dollar-volume traded).[25]

No adjustment is made for transaction costs yet. Such costs will be relatively small as HFTs pay to trade only when they take liquidity and they receive a rebate when they provide liquidity. For example, Nasdaq offers \$.20 per 100 shares to high-volume liquidity providers. On the other hand, Nasdaq charges \$.25 per 100 shares to traders who take liquidity. I use the fraction of shares where HFTs take liquidity and

[23]As the regression analysis conducted in Table A-3 is not bound between 0 and 1 I set a lower limit of 0% and an upper limit of 95% of HFTs' fraction of market participation.

[24]The fraction of dollar-volume HFTs are not trading with each other is 0.780, which I obtain by calculating the dollar-volume where HFTs are on only one side of a trade divided by the total dollar-volume in which HFTs are involved: $(\text{DVolume}_{HN} + \text{DVolume}_{NH})/(\text{DVolume}_{HH} + \text{DVolume}_{HN} + \text{DVolume}_{NH})$. As I am trying to determine the profitability of the HFT industry, when I multiply the average profit per dollar traded I only consider the dollars traded with non-HFTs, as a trade between two HFTs has a net zero profit for the industry.

[25]This number is less than what others have estimated. An article by the Tabb Group claimed HFTs made around \$21 billion annually. However, the \$2.8 billion annually from U.S. equities is in line with other claims. For instance, a *Wall Street Journal* article states that Getco made around \$400 million in 2008 across all of its divisions (it trades on fifty exchanges around the world and in equities, commodities, fixed income, and foreign exchange. Even if Getco, one of the largest HFT firms, earned \$150 million of that profit from U.S. equities it would still be in line with my findings.

provide liquidity from Table 1 Panel A to estimate the explicit Nasdaq trading costs. Using the average stock price in 2008 and 2009 of $37.10 and the finding that HFTs take liquidity in 50.98% of the dollar-volume they trade and provide it in 49.02%, I estimate HFTs annual net exchange transaction costs to be $399.6 million.

Besides knowing how profitable HFTs are, it is also informative to know how profitable they are compared to traditional market makers. Hasbrouck and Sofianos (1993) and Coughenour and Harris (2004) study the trading activity and profitability of the NYSE specialists. In the HFT dataset HFTs earn less than 1/100th of a penny ($0.000072) per dollar traded. Using the reported summary data in Coughenour and Harris (2004), specialists in 2000 (after decimalization, before decimalization reported in parentheses) made $0.00052 ($0.000894) per dollar traded in small stocks, $0.00036 ($0.00292) per dollar traded in medium stocks, and $.00059 ($.0025) per dollar traded in large stocks.[26] From this perspective, HFTs are almost one-seventh as expensive as traditional market makers.[27] Part of this earning discrepancy could be that HFTs and traditional market makers are being compensated for different kinds of risks. Whereas HFTs tend to close the day in a neutral position, traditional market markets often hold inventory for days or weeks.

5.4 Testing Whether High Frequency Traders Systematically Engage in Anticipatory Trading

In this section I test whether HFTs systematically anticipate and trade in front of non-HFTs ("anticipatory trading") (SEC, January 14, 2010). It may be that HFTs predict and buy (sell) a stock just prior to when a non-HFTr buys (sells) the stock. If this is the case, HFTs are profiting at the expense of non-HFTs.[28]

To determine whether HFTs are implementing an anticipatory trading strategy, I analyze the frequency

[26] I calculate the gross profits in Coughenour and Harris (2004) by applying the following formula: $\text{Profit Per Dollar Traded}_{\text{Stock Size}} = \text{Specialists Gross Profits}_{\text{Table 5}} / (\text{Price}_{\text{Table 1 Panel A}} * \text{Total Shares Traded}_{\text{Table 1 Panel C}} * \text{Specialist Share Participation Rates}_{\text{Table 2 Panel A}})$.

[27] I do not make adjustments for inflation.

[28] Anticipatory trading is not itself an illegal activity. It is illegal when a firm has a fiduciary obligation to its client and uses the client's information to front run its orders. In my analysis, as HFTs are propriety trading firms, they do not have clients and so the anticipatory trading they may be conducting would likely not be illegal. Where HFT and anticipatory trading may be problematic is if market manipulation is occurring that is used to detect orders. It may be the case that "detecting" orders would fall in to the same category of behavior as that resulted in a $2.3 million fine to Trillium Brokerage Services for "layering". Trillium was fined for the following layering strategy: Suppose Trillium wanted to buy stock X at $20.10 but the current offer price was $20.13, Trillium would put in a hidden buy order at $20.10 and then place several limit orders to sell where the limit orders were sufficiently below the bid price to be executed. Market participants would see this new influx of sell orders, update their priors, and lower their bid and offer prices. Once the offer price went to $20.10, Trillium's hidden order would execute and Trillium would then withdraw its sell limit orders. FINRA found this violated NASD Rules 2110, 2120, 3310, and IM-3310 (now FINRA 2010, FINRA 2020, FINRA 5210, and also part of FINRA 5210).

of observing different marketable order sequences. In the data, anticipatory trading would show up as a HFTr-initiated buy (sell) order just prior to a non-HFTr-initiated buy (sell) order. If the trading sequence is independent of the trader type, it should be equally likely to observe a HFTr-initiated trade prior to a non-HFTr-initiated trade as to observe the reverse, a non-HFTr-initiated trade prior to a HFTr-initiated trade.

To formalize, let $T_{t-1}T_t$ represent the sequence of trades where T is the type of trader, H a HFTr and N a non-HFTr, and t the transaction time sequence of events. If systematic anticipatory trading by HFTs is occurring, then I would see: Prob(HN) $>$ Prob(NH) and if it were not occurring, I would observe Prob(HN) $\leq$ Prob(NH). Prob() is defined as Prob(x) $= \frac{n_x}{n_t}$ where x represents the trade sequence of interest, n_x represents the total number of times such a sequence is observed, and n_t represents the total number of sequences observed. It is important to compare the sequence to its reflection so as to nullify differences in the probabilities of observing a HFTr- or non-HFTr-initiated trade.

For each stock and each day I analyze the probability of seeing different trading patterns. Besides the two-period sequence, I consider the three-, four-, five- and six-period sequences. The five different calculations are:

$$
\begin{aligned}
AT_1 &= \text{Prob(HN) - Prob(NH)} \\
AT_2 &= \text{Prob(HHN) - Prob(NHH)} \\
AT_3 &= \text{Prob(HHHN) - Prob(NHHH)} \\
AT_4 &= \text{Prob(HHHHN) - Prob(NHHHH)} \\
AT_5 &= \text{Prob(HHHHHN) - Prob(NHHHHH).}
\end{aligned}
$$

An $AT > 0$ is consistent with anticipatory trading, while an $AT \leq$ is not. I calculate the statistical significance incorporating Newey-West standard errors to correct for the time-series correlation in observations. I summarize the results by stock market capitalization.

Table 4 Panel A shows the results. Column (1) shows the results for AT_1, column (2) AT_2, column (3) AT_3, column (4) AT_4, and column (5) AT_5. For all the sequences across all of the stock sizes, the average AT is negative, which is inconsistent with HFTs systematically engaging in an anticipatory trading strategy. The number of stocks in which the AT is statistically significantly less than zero increases with stock size and decreases with the sequence length.[29]

If HFTs engage in anticipatory trading, it is most likely to occur when future trading activity is the most

[29] I do the same analysis for Sell orders and find economically similar results.

predictable. Thus, I perform the same analysis as above except that I condition the probabilities on the non-HFT trade being for more than 500 shares.[30] That is, $AT_1 = \text{Prob(HN}|\text{N}_{\text{shares}} > 500) - \text{Prob(NH}|\text{N}_{\text{shares}} > 500)$, where N_{shares} represents the number of shares traded in the non-HFTr-initiated trade. Table 4 Panel B shows the results. Like Panel A, regardless of stock size or sequence length, AT is negative or zero. For no stock is AT of any sequence length statistically significantly greater than zero. This additional test finds no evidence suggesting HFTs systematically engage in anticipatory trading.

These findings suggest HFTs as a whole are not engaging in non-HFTr anticipatory trading. However, I cannot conclude there is no anticipatory trading. It could be that the multiple strategies HFTs use cancel out the informativeness of this approach to detect anticipatory trading. It could also be that when one non-HFTr initiated order executes it is a signal that other non-HFTr initiated orders are coming into the market and so HFTs quickly initiate their own orders. The sequence may then look like NHHHN, which would show up in the results as there being one of each of the following: NH, HN, NHH, HHN, NHHH, HHHN, and from this sequence, AT_1, AT_2, and AT_3 would equal zero.

5.5 Testing Whether High Frequency Traders' Strategies Are More Correlated than Non High Frequency Traders'

In this section I test whether HFTs' strategies are more correlated than non-HFTs'. A concern is that, if HFTs use similar trading strategies, they may exacerbate market movements. To determine whether HFTs' strategies are more correlated than non-HFTs', I examine the frequency at which HFTs trade with each other and compare it to a benchmark model used in Chaboud, Hjalmarsson, Vega, and Chiquoine (2009) that produces theoretical probabilities of different types of trading partners (demander - supplier) under the assumption that traders' activities are independent of their trader partners. I compare the actual occurrence of different trades to the predicted amount. I find HFTs trade with each other less than predicted, suggesting their strategies are more correlated than non-HFTs.

There are four trade partner combinations, HH, HN, NH, NN, where the first letter represents the liquidity demander and the second the liquidity supplier and N represents a non-HFTr and H a HFTr. Let H_s be the number of HFT liquidity suppliers, H_d be the number of HFT liquidity demanders, N_s be the number of non-HFT liquidity suppliers, and N_d be the number of non-HFT liquidity demanders. Then

[30] I do the same analysis for 300 and 1000 shares with economically similar results.

there are the following probabilities:

$$\text{Prob(HFT - supply)} = \frac{H_s}{N_s + H_s} = \alpha_s$$
$$\text{Prob(non-HFT - supply)} = \frac{N_s}{N_s + H_s} = 1 - \alpha_s$$
$$\text{Prob(HFT - demand)} = \frac{H_d}{N_d + H_d} = \alpha_d$$
$$\text{Prob(non-HFT - demand)} = \frac{N_d}{N_d + H_d} = 1 - \alpha_d.$$

And so,

$$\text{Prob(HH)} = (\alpha_d)(\alpha_s)$$
$$\text{Prob(NH)} = (1 - \alpha_d)(\alpha_s)$$
$$\text{Prob(HN)} = (\alpha_d)(1 - \alpha_s)$$
$$\text{Prob(NN)} = (1 - \alpha_d)(1 - \alpha_s).$$

As a result, the following fraction holds: $\frac{\text{Prob(NN)}}{\text{Prob(NH)}} \equiv \frac{\text{Prob(HN)}}{\text{Prob(HH)}}$. Let RN = $\frac{\text{Prob(NN)}}{\text{Prob(NH)}}$ be the non-HFTr demanding liquidity ratio and RH = $\frac{\text{Prob(HN)}}{\text{Prob(HH)}}$ be the HFTr demanding liquidity ratio. When there is more non-HFTr stock volume than HFTr stock volume, then Prob(NN) $>$ Prob(NH) and Prob(HN) $>$ Prob(HH). However, regardless of the volume of transactions by non-HFTs and HFTs, the difference of ratios R = RH - RN will equal zero as non-HFTs will take liquidity from other non-HFTs in the same proportion as HFTs take liquidity from other HFTs. Therefore, if R = 0, it is the case that HFTs and non-HFTs trade with each other as much as expected when their trading strategies are equally correlated. If R $>$ 0, then it is the case that HFTs trade with each other less than expected or that HFTs trade with non-HFTs more than expected. In the data Prob() is calculated as Prob(x) = $\frac{n_x}{n_t}$ where x represents the desired trade liquidity supplier and demander, n_x represents the total number of times such a transaction is observed, and n_t represents the total number of transactions observed.

Table A-6 Panel A shows the results. The stocks are sorted into three market capitalization groups, as well as an Overall category. The column Mean R shows the average result of RH - RN. The column Std. Dev. R is the standard deviation of R across the stocks. The column Mean % Days R $>$ 0 is average percent of days when R is greater than zero. The column Mean % Days R $<$ 0 is the average percent of days when R is less than zero. The column Stat. Sign. $<$ 0 is the number of stocks whose average R is statistically significantly less than zero. The column Stat. Sign. $>$ 0 is the number of stocks whose average R is statistically significantly greater than zero. I report the results in three groups based on stock size, as well as the overall results. As the stock size increases, R decreases. For small stocks the mean R is 8.4, for medium it is 3.3, and for large it is 0.8. For no stocks is R statistically significantly less than zero, while for 91 it is statistically significantly greater than zero. This suggests that HFTs trade with each other less than expected or that HFTs trade with non-HFTs more than expected. The interpretation of this result is that HFTs' trading strategies are more correlated than non-HFTs'.

If HFTs tend not to carry positions overnight, then as the trading day nears an end HFTs should increase their trading with each other to offset each others' inventory imbalances. To check this, I perform the same analysis as above, but consider only trading in the last 15-minutes of the day. The results are in Table A-6 Panel B and are consistent with the market maker story. For small and medium stocks, the number of stocks whose R is statistically significantly greater than zero declines. In addition, for all stocks, the average percent of trades where R is greater than 0 decreases.

5.6 Analysis of the High Frequency Trading - Volatility Relationship

5.6.1 Examination of How High Frequency Trading Changes as Day-Level Volatility Varies

The fact that HFTs are around during normal times, but during extreme market conditions may reduce their trading activity is a serious concern. In this section I study how HFTs behave in different levels of volatility. I am interested in understanding the relationship between HFT and volatility as volatility levels change. To do so, I build a graphical representation of the HFT-Volatility relationship. Of interest is how HFTs either pull back or increase their trading activity as volatility changes. The results are in Figure 3. The X-axis for each graph is 100 bins grouped together based on the V-Level value:

$$\text{V-Level}_{i,t} = \frac{V_{i,t} - E(V_i)}{E(V_i)} * \frac{1}{\sigma_i},$$

where $V_{i,t}$ is the 15-minute realized volatility for stock i on day t, and σ_i is the standard deviation of stock i's V. The V-Level variable is the scaled deviation from the mean, where it is scaled by the standard deviation of a stock's volatility, σ_i. Without the scaling by σ_i I would essentially be plotting HFTs' fluctuation across stocks, with more volatile stocks, which tend to be smaller stocks, being further to the right on the X-axis.

For the observations in each bin, I calculate the abnormal HFTr activity, HFT-Level:

$$\text{HFT-Level}_j = \sum_{\text{V-Level}_{i,t} \epsilon j} \frac{1}{N_j} \left[\frac{\text{HFT}_{i,t} - E(\text{HFT}_i)}{E(\text{HFT}_i)} \right] \tag{3}$$

where HFT is the fraction of shares in which HFTs are involved and j is the V-Level bin for which stock i at time t has been grouped. N is the number of observations in bin j. By evaluating HFT-Level at the stock-by-stock level, I am in effect controlling for stock-specific effects. Figure 3 contains three graphs, HFT - All, which looks at all HFTs' activity, HFT - Supply, which considers only HFTs' liquidity supply activity, and HFT - Demand, which considers only HFTs' liquidity demand activity. In each graph there are four lines. The bin-by-bin HFT-Level, a nine-bin centered moving average of the HFT-Level, and the upper and lower 95 % confidence intervals.

The first graph, HFT - All, is almost flat across volatility levels. Even on the most volatile days, HFTs' overall activity does not seem to increase or decrease substantially. However, when volatility is low, HFTs' activity is slightly lower than average. The second figure shows HFTs' supply of liquidity. HFTs provide 10% more liquidity than usual on very low volatility days. The level of HFTs' liquidity declines steadily as volatility picks up. At the highest levels of volatility HFTs provide 10% less liquidity than on average. The third figure displays HFTs taking liquidity. On the least volatile days HFTs take about 7% less liquidity than normal, and on the most volatile days they take 6% more liquidity than on average.

These results show that as volatility increases, HFTs provide liquidity less often and take it more often, but the change is neither precipitous or large. In particular, on the most volatile days, HFTs do not stop providing liquidity. On these days, there does seem to be a transfer of HFT activity from supplying liquidity to demanding liquidity. This could be consistent with the market maker story. On volatile days HFTs' inventory will need to be rebalanced and they will have to demand liquidity to unload positions. The change in liquidity taking is also consistent with a statistical arbitrage story. When prices are steady there are fewer arbitrage opportunities and so HFTs make fewer marketable trades. When prices are volatile, there are more arbitrage opportunities for which HFTs will step in and demand liquidity.

5.6.2 Examination of How High Frequency Trading Changes During Intra-day Price Movements

While the above analysis and Figure 3 look at day-level volatility, higher-frequency price fluctuations are also of interest to more precisely understand HFTs' trading activity during short-term price movements. I use 15 minute intervals in this analysis and instead of looking at volatility I examine returns during the 15-minute period. I can separate the analysis along three dimensions: whether returns are positive or negative, whether HFTs are buying or selling, and whether HFTs are supplying or demanding liquidity.

The variables are similarly defined as above, but adjusted to look at returns, not volatility:

$$\text{Ret-Level}_{i,t,m} = \frac{\text{Ret}_{i,t,m}}{\sigma_i},$$

where $\text{Ret}_{i,t,m}$ is the maximum return for stock i on day t during 15-minute period m. That is, Ret is calculated based on the maximum and minimum prices during a 15-minute period. If the minimum price occurred prior to the maximum price, it is considered a positive return period. The opposite scenario defines a negative return period. I re-define HFT-Level as:

$$\text{HFT-Level}_j = \sum_{\text{Ret-Level}_{i,t,m} \epsilon j} \frac{1}{N_j} \left[\frac{\text{HFT}_{i,t,m} - E(\text{HFT}_i)}{E(\text{HFT}_i)} \right] \tag{4}$$

HFT-Level takes on one of five definitions: All Activity, Buy-Demand, Buy-Supply, Sell-Demand, or Sell-Supply where each defines HFT-Level as the percent of all trades that occur in the market that satisfy the criteria implied in the name, where the Buy/Sell refers to HFTs' activity, and Supply/Demand refers to HFTs' role in the transaction. I remove observations where the return was zero for that period or where fewer than 30 trades occurred.

The results for price increases are in Figure A-2 and the results for price declines are in Figure A-3. In each figure there are four lines: the bin-by-bin HFT-Level, a five-bin centered moving average of the HFT-Level, and the upper and lower 95% confidence intervals. The figures suggest that HFTs do not drive price fluctuations. During the greatest price increases, HFTs buy and demand liquidity less than normal. During the large price increases, HFTs provide more liquidity than normal. The same relationship is true during price declines: during the largest price declines, HFTs decrease their liquidity demand and increase the liquidity they provide. The 15-minute analysis shows that as prices fluctuate more than normal, HFTs supply more liquidity, and demand less liquidity, than on average, which is contrary to HFTs fleeing markets in volatile times.

5.6.3 Analysis of How Volatility Impacts High Frequency Trading

The above results do not address the endogeneity between HFT and volatility. To evaluate whether HFTs increase their trading because volatility is higher or vice versa, I analyze situations in which there are exogenous shocks to volatility. Exogenous shocks to volatility can come from new information entering the public domain. Thus, a natural time to expect exogenous shocks to volatility is during quarterly stock earnings announcements. In the HFT sample dataset, days on which stocks announce their quarterly earnings have higher volatility than the average non-announcement day for that stock. The difference is small but statistically significant. Using OLS regression analysis, I perform the following regression at the daily stock level:

$$HFT_{i,t} = \alpha + \mathbf{1}_{QEA,i,t} * \beta_1 + \mathrm{VIX}_t * \beta_2 + \epsilon_{i,t}, \tag{5}$$

where $HFT_{i,t}$ takes on different definitions: it is either (1) the percent of shares in stock i in which HFTs were involved, (2) the percent of shares in stock i in which HFTs were involved and were demanding liquidity, or (3) the percent of shares in stock i in which HFTs were involved and were supplying liquidity. The Quarterly Earnings Announcement variable, $\mathbf{1}_{QEA}$, is a dummy variable that equals one for stock i if the observation is on the day of or the day after stock i reports its quarterly earnings, and zero otherwise. Stock fixed effects are implemented and standard errors are clustered by stock. The results are in Table A-7 Panel A; the coefficient on the quarterly earnings announcement dummy is insignificant for HFT-All, statistically significant and negative for HFT-Demand, and statistically significant and positive for HFT-Supply. These results suggest that when volatility rises for exogenous reasons, HFTs take less liquidity and increase the liquidity they supply.

Another time in which there was an identifiable exogenous shock to volatility was the week of September 15 - September 19, 2008, the week in which Lehman Brothers collapsed, during that time significant amounts of information were being discovered by market participants. I conduct an analysis similar to that done for the Quarterly Earnings announcements but now with a dummy $\mathbf{1}_{LF}$, which equals one for observations during September 15, 2008 - September 19, 2008 and zero otherwise. I run the regression at the aggregate day level for trading days in 2008.

$$HFT_t = \alpha + \mathbf{1}_{LF,t} * \beta_1 + \mathrm{VIX}_t * \beta_2 + \epsilon_{i,t}, \tag{6}$$

where the variables are defined as in the previous equation. Table A-7 Panel B reports the results. The Lehman Week dummy coefficient has a positive and statistically significant value for the HFT - Supply regression, implying that during the Lehman Week HFTs supplied more liquidity than normal. The Lehman Week coefficient in the other two regressions, HFT - All and HFT - Demand, is not statistically significantly different from zero. The results from the quarterly earnings announcement and the Lehman failure week are consistent in that the HFT - Supply regression analyses both produce statistically significantly greater than zero coefficients on the variable capturing exogenous volatility. The coefficients of interest in the other two regressions are not statistically significantly different from zero. The results in this section suggest that an increase in volatility due to new information causes HFTs to increase their supply of liquidity.

6 High Frequency Trading and Asset Pricing

In this section I analyze HFTs' impact on asset pricing. I focus on the market quality measures price discovery, liquidity, and volatility by addressing three questions: (1) Do HFTs contributes to the price discovery process? (2) How much liquidity do HFTs provide? (3) How does HFT impact volatility? My findings suggest that HFTs improve market quality. I find that HFTs contribute significantly to the price discovery process. Also, they frequently offer the best bid and offer prices. However, they provide less depth than non-HFTs and they tend to avoid supplying liquidity to informed traders. Finally, I find evidence that HFTs may dampen market volatility.

6.1 Analysis of High Frequency Traders' Role in the Price Discovery Process

In this section I utilize three Hasbrouck price discovery methodologies to see whether HFTs provide new information to the market. The Hasbrouck methodologies are similar to those found in Hendershott and Riordan (2009) and other papers. First, I implement the permanent price impact measure that utilizes an impulse response function. The results show the amount of permanent price adjustment from trades by the two types of traders. This value is interpreted as the private information impounded in to the stock price by different traders (Hasbrouck, 1991a,b). Second, I use the aggregate information variance decomposition, a technique that takes the results of the impulse response function calculated for the permanent price impact measure and relates the different types of traders' trades to the price discovery process. Finally, I implement the information share approach, which takes the innovations in HFTs' and non-HFTs' quotes and decomposes the variance of the common component of the price to attribute contribution to the efficient price path between the two types of traders. Across all three measures, I find that HFTs make significant contributions to the price discovery process.

6.1.1 The Permanent Price Impact of Trades Initiated by High Frequency Traders and Non High Frequency Traders

To measure the information content of trades by HFTs and non-HFTs, I calculate the permanent price impact of HFTs' and non-HFTs' trades. Hendershott and Riordan (2009) perform a similar calculation for algorithmic trading, while others, such as Barclay, Hendershott, and McCormick (2003), use the technique to compare information from different markets.[31] I calculate the impulse response function on a trade-by-trade basis using 10 lags for trades by HFTs and non-HFTs. I estimate the impulse response function for each stock for each day. As in Barclay, Hendershott, and McCormick (2003) and Hendershott and Riordan (2009), I estimate three equations, a midpoint quote return equation, a HFTr trade equation, and a non-HFTr trade equation. The 10-lag vector auto regression (VAR) is:

$$r_t = \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=0}^{10} \beta_i q_{t-i}^H + \sum_{i=0}^{10} \gamma_i q_{t-i}^N + \epsilon_{1,t}, \tag{7}$$

$$q_t^H = \sum_{i=1}^{10} \delta_i r_{t-i} + \sum_{i=0}^{10} \rho_i q_{t-i}^H + \sum_{i=0}^{10} \zeta_i q_{t-i}^N + \epsilon_{2,t}, \tag{8}$$

$$q_t^N = \sum_{i=1}^{10} \pi_i r_{t-i} + \sum_{i=0}^{10} \nu_i q_{t-i}^H + \sum_{i=0}^{10} \psi_i q_{t-i}^N + \epsilon_{3,t}. \tag{9}$$

where t is a trade-event based time indicator, i indicates the number of lagged events, q^H is a signed indicator for HFTr trades taken a value of +1 for a HFTr initiated buy, -1 for HFTr initiated sell, and 0 otherwise, q^N is a similarly designated indicator for non-HFTr trades, r_t is the quote midpoint to quote midpoint return between trade changes, and the Greek letters are the coefficients for the different regressors.

After estimating the VAR, I invert it to get the vector moving average (VMA) model to obtain:

$$\begin{bmatrix} r_t \\ q_t^H \\ q_t^N \end{bmatrix} = \begin{bmatrix} a(L) & b(L) & c(L) \\ d(L) & e(L) & f(L) \\ g(L) & h(L) & i(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}, \tag{10}$$

where the vectors $a(L)$ - $i(L)$ are lag operators. Hasbrouck (1991a) interprets the impulse response function for HFT, $\sum_{t=0}^{10} b(L)$, as the private information content of an innovation in HFT. The non-HFT impulse response function is $\sum_{t=0}^{10} c(L)$ and is the private information content of an innovation in non-HFT.[32] The expected portion of a trade should not impact prices and so should not show up in the impulse response function; however, the unexpected portion, the innovation, of a trade should influence the price of future trades. The impulse response function estimates this impact on future trades. By examining a

[31] The HFT dataset is especially well suited for this as it is in milliseconds and thus minimizes the problem of multiple trades occurring within one time period, as occurs with data denoted in seconds.

[32] The impulse response function is a technology first used in the macro-economic literature to determine the impact of an exogenous shock to the economy as it worked its way through the economy. Hasbrouck (1991a) and Hasbrouck (1991b) took this methodology and applied it to the microstructure literature.

unit shock to $\epsilon_{1,t}$ from either a HFTr trade (non-HFTr trade), the coefficients in $b(L)$ ($c(L)$) provide the expected permanent price impact that will result.

Table 5 Panel A shows the results of the HFT and non-HFT impulse response function for 10 events into the future.[33] The results are in terms of basis points. I divide the stocks into three groups based on stock market capitalization. All three size groups show that a HFTr-initiated trade has a larger permanent price impact than a non-HFTr-initiated trade.

For each stock I estimate the statistical significance of the difference of the impulse response functions for the HFT and non-HFT over five trading days using a t-test. The t-test is adjusted using Newey-West standard errors to account for the time-series correlation in observations. As stock size increases, the number of stocks with the HFT impulse response function being statistically significant and larger than the non-HFT's decreases from 3 to 0, and the reverse, with HFT's impulse response function being larger, increases from 0 to 21 stocks. This suggests that HFTs' trades provide more private information than do non-HFTs' trades. This is similar to the findings for algorithmic trades in Hendershott and Riordan (2009).

I now test whether the price impact is immediate or gradual over the ten future time periods. If there is an immediate overreaction to a HFTr's trade, this would support the claim that HFTs increase market volatility. I use the same VAR and VMA results from above but analyze the difference between the long-run (LR; 10-event forecast horizon) and short-run (SR; immediate) impulse response functions. The results are in Table 5 Panel B.

The results do not support the overreaction hypothesis to HFTs' trades. A significant portion of the price impact from HFTs' trades comes immediately: 1/3 for small and medium stocks and 1/2 for large stocks. The remainder of the price impact is imputed over the next several trades. For each stock I estimate the statistical significance of the difference of the LR - SR impulse response functions for the HFT and non-HFT over five trading days using a t-test. The t-test is adjusted using Newey-West standard errors to account for the time-series correlation in observations. Only one stock has a HFT LR-SR impulse response function difference that is statistically significant and less than the non-HFT's, while 52 stocks have differences that are statistically significant in the other direction. Table 5 Panel A suggests that HFTs' trades have more private information than non-HFTs' trades and Panel B supports that the difference is persistent and increases beyond the immediate impact of the trade.

[33] I do a similar calculation for 20, 50, and 100 events and obtain similar results.

6.1.2 The Aggregate Amount of Information in Trades Initiated by High Frequency Traders and Non-High Frequency Traders

In this section I examine HFTs' and non-HFTs' roles in the price formation process. The permanent price impact section above shows that HFTr-initiated trades add important information to the market, but the methodology does not directly estimate the importance of HFT and non-HFT in the overall price formation process. To examine this, I follow Hasbrouck (1991b), which decomposes the variance of the efficient price into the portion of total price discovery that is correlated with HFTs' trades and non-HFTs' trades.[34] The results indicate that HFTs contribute more to price discovery.

To perform the variance decomposition, the return series r_t(using midpoint returns to avoid the bid-ask bounce) is separated into its random walk component m_t and stationary component s_t:

$$r_t = m_t + s_t, \tag{11}$$

where m_t represents the efficient price, $m_t = m_{t-1} + w_t$, and w_t is a random walk with $Ew_t = 0$, and s_t is the transitory price component. Letting $\sigma^2_{\epsilon_1} = E\epsilon^2_1$, $\sigma^2_{\epsilon_2} = E\epsilon^2_2$, and $\sigma^2_{\epsilon_3} = E\epsilon^2_3$, I decompose the variance of the efficient price, m_t, into trade-correlated and trade-uncorrelated changes:

$$\sigma^2_w = (\sum_{i=0}^{10} a_i)^2\sigma^2_{\epsilon_1} + (\sum_{i=0}^{10} b_i)^2\sigma^2_{\epsilon_2} + (\sum_{i=0}^{10} c_i)^2\sigma^2_{\epsilon_3}, \tag{12}$$

where the a, b, c are as defined in the previous section the lag coefficients found in the VMA matrix. The $(\sum_{i=0}^{10} b_i)^2\sigma^2_{\epsilon_2}$ term represents the portion of the efficient price variance attributable to HFT and the $(\sum_{i=0}^{10} c_i)^2\sigma^2_{\epsilon_3}$ term represents the non-HFT proportion of the efficient price variance. The $(\sum_{i=0}^{10} a_i)^2\sigma^2_{\epsilon_1}$ term is the already public information portion of price discovery.

The results are in Table A-8 Panel A. I report the average contribution by HFTs and non-HFTs to each stock over the five days. I summarize the findings by combining observations based on stock market capitalization. The t-statistic for the difference between the HFT and non-HFT contribution is adjusted for its time-series correlation with Newey-West standard errors. The contribution to the Returns component (the public information) is the public information related to price discovery, the difference between 1 and the sum of the HFT and non-HFT components. As the stock size increases, more of the overall price discovery contribution comes from trades, as opposed to returns. For the small stocks, HFTs' average contribution

[34] This analysis was used in Hendershott and Riordan (2009) to determine that algorithmic traders contribute more to price discovery than do human traders.

is less than non-HFTs' and the difference is statistically significant for 10 stocks. For medium size firms, the contribution by HFTs and non-HFTs is on average equal. Six stocks in this size category have non-HFTs providing a larger contribution to price discovery than HFTs and the difference being statistically significant, and 11 stocks have HFTs' contribution being larger and the difference being statistically significant. Finally, in the large stocks, HFTs on average contribute more to the price discovery process, and this difference is statistically significant for 28 stocks.

6.1.3 The Information Share in Quotes Supplied by High Frequency Traders and Non-High Frequency Traders

This section examines the role of quotes placed by HFTs and non-HFTs in the price discovery process, whereas the previous two sections analyzed the role of trades. I use the Information Share (IS) methodology introduced by Hasbrouck (1995) and used in, among others, Chaboud, Hjalmarsson, Vega, and Chiquoine (2009) and Hendershott and Riordan (2009).[35]

The IS methodology is similar to the Variance Decomposition technique used above, but focuses on the evolution of traders' quotes. First, I calculate the HFT and non-HFT midpoint quote price paths. Next, if prices follow a random walk then I can represent the change in price as a vector moving average (VMA). I can decompose the VMA variance into the lag operator coefficients and the variance of the different market participants' price paths. The market participants' variance is considered the contribution of that participant to the information in the price discovery process. The VMA distinguishes the variance of the random walk and the coefficients of the VMA innovations.

More precisely, I utilize the following framework. The HFTr price process is calculated from the HFTr quote midpoint, MP_t^{HFT} = $(\text{InsideBid}_t^{HFT} + \text{InsideAsk}_t^{HFT})/2$ for HFT, and the same is done for the non-HFTr quote midpoint. The price process for HFTs and non-HFTs is $p_t^{HFT} = m_t + \epsilon_t^{HFT}$ and $p_t^{nHFT} = m_t + \epsilon_t^{nHFT}$ respectively, and the common efficient price path is the random walk process, $m_t = m_{t-1} + u_t$.

The price change vector of the HFTr and non-HFTr price process can be modeled as a VMA:

$$\Delta p_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} \ldots, \tag{13}$$

where $\epsilon_t = [\epsilon_t^{HFT}, \epsilon_t^{nHFT}]'$ and is the information coming from HFTr quotes and non-HFTr quotes. The variance σ_u^2 can be decomposed as:

[35] The IS methodology has been used to determine which of several markets contributes more to price discovery, and, as will be done here, to determine which type of market participant contributes more to the price discovery process.

$$\sigma_u^2 = \left[\begin{array}{cc} \Psi_{HFT} & \Psi_{nHFT} \end{array} \right] \left[\begin{array}{cc} \sigma^2_{HFT} & \sigma^2_{HFT,nHFT} \\ \sigma^2_{HFT,nHFT} & \sigma^2_{nHFT} \end{array} \right] \left[\begin{array}{c} \Psi_{HFT} \\ \Psi_{nHFT} \end{array} \right], \tag{14}$$

where Ψ represent the lag operator vector from Equation 13 and the sigmas represent the $Var(\epsilon_t)$ from Equation 13.[36]

The results are reported in Table A-8. I report the average contribution by HFTs and non-HFTs for each stock over the five days. The IS between HFTs and non-HFTs sums to 1 for each stock-day estimation. I summarize the findings by combining observations based on stock market capitalization. The t-statistic for the difference between the HFT and non-HFT contribution is adjusted for its time-series correlation with Newey-West standard errors. The IS of a trader type is measured as that participant's contribution to the total variance of the common component of the price. Across stock sizes, HFTs' average IS is larger than non-HFTs', and the difference increases with stock size. Also, HFTs' IS is larger than non-HFTs' and the difference is statistically significant for 38 stocks, while non-HFTs' Information is larger than HFTs' and the difference is statistically significant only 13 times. These results suggest that in quotes, as in trades, HFTs add considerably to the price discovery process.

6.2 High Frequency Traders' Supply of Liquidity

This section analyzes HFTs' supply of liquidity. I examine the stock and day level determinants that influence HFTs' time at the best bid or offer. I find that HFTs are more likely to provide the inside quotes for large stocks that do not exhibit autocorrelation and that have thinner quote depth. Next, I examine the depth of liquidity provided by HFTs and non-HFTs. I find that HFTs provide less book depth than non-HFTs. Finally, I analyze which trader type provides more liquidity to informed traders and find that HFTs avoid trading with informed traders.

6.2.1 Day-Level High Frequency Trading Determinants of Time at the Inside Quotes

In this section I analyze the factors that influence HFTs' fraction of time providing the best bid and offer across stocks and days. The summary statistics in Table 2 show that HFTs' fraction of time providing the best bid and offer varies across stocks and across time. I perform an OLS regression analysis similar to that found in Table A-3 for each stock on each day. I run the following regression:

[36] As the quote data I use are updated every time a new inside bid or ask is posted by a HFT or a non-HFT, the diagonal values of the covariance matrix should be nearly perfectly identified. That is, as the limit order book is updated every millisecond for which an order arrives, there should be no contemporaneous correlation between HFT and non-HFT quote changes.

$$\begin{aligned}\mathrm{H}_{i,t} =& \alpha + \mathrm{MC}_i * \beta_1 + \mathrm{MB}_t * \beta_2 + \mathrm{VIX}_i * \beta_3 + \sigma_{i,t} * \beta_4 + \mathrm{SP}_{i,t} * \beta_5 + \\ & \mathrm{DEP}_{i,t} * \beta_6 + \mathrm{TS}_{i,t} * \beta_7 + \mathrm{NV}_{i,t} * \beta_8 + \mathrm{AC}_{i,t} * \beta_9 + \epsilon_{i,t},\end{aligned}$$

where $\mathrm{H}_{i,t}$ is the fraction of calendar time a HFTr was providing the best bid or offer in stock i on day t. The rest of the variables are defined the same as in Section 5.1. I also perform a restricted regression analysis that includes only market capitalization, market-to-book, and the VIX.

Table A-9 reports the results. Columns (1) and (2) show the standardized coefficients, and columns (3) and (4) show the regular coefficients. The results show that market capitalization is the most important determinant of the time HFTs provide the best bid or offer and it has a positive coefficient. The other coefficients that are statistically significant are negative, suggesting that HFTs prefer to provide the inside quotes when there is less quote depth, non-HFTr dollar-volume, and when the return autocorrelation is closer to zero. In the restricted regression the Market Capitalization coefficient remains statistically significant and positive. The Market / Book coefficient is negative and becomes statistical significant. The VIX coefficient, which is not statistically significant, changes signs. The magnitudes of the market capitalization and market-to-book variables increase.

6.2.2 High Frequency Traders' and Non-High Frequency Traders' Book Depth

In this section I examine the depth of liquidity on the order book arising from HFTs and non-HFTs. I use one-minute snapshots of the order book to analyze how much limit orders provided by HFTs reduce the price impact from a trade in a partial equilibrium setting. I find that, while HFTs provide a sizeable amount of order book depth, it is strictly less than that provided by non-HFTs.

I consider the price impact of different size trades hitting the book with and without different types of traders. First, I report in Table 6 Panel A the average price impact different size trades would have if they were to hit Nasdaq's order book and be able to access all standing limit orders. I separate the results into three categories based on stock size. I report the basis point impact and the dollar value impact. As the number of simulated shares hitting the book increases, the price impact increases. Also, the price impact for a given size trade decreases as stock size increases.

Next I consider how much limit orders provided by HFTs reduce the price impact from a trade in a partial equilibrium setting. I do so by comparing the price that would result after a given trade size with access to all available limit orders with the price that would result from the same trade size having access only to non-HFTs' limit orders. In Table 6 Panel B I report the price impact difference in basis points and dollars. The results suggest HFTs provide more liquidity in large stocks than in small. The price impact is monotonically increasing and almost linear within each stock size category.

In Table 6 Panel C the results are reported for the same analysis but with consideration to an elimination of non-HFTs' limit orders. The price impact is substantially more than in Panel B across all stock size categories and for all trade sizes. These results suggest that, while HFTs provide some liquidity depth, it is only a fraction of that provided by non-HFTs.

A concern with this analysis is the endogeneity of limit orders (Rosu, 2009) and the information they may contain (Harris and Panchapagesan, 2005; Cao, Hansch, and Wang, 2009). That is, a market participant who sees a limit order at a given price or in a certain quantity (or absence thereof) may alter his behavior as a result. The analysis I do is only for a partial equilibrium setting and does not attempt to incorporate general equilibrium dynamics into the results. Also, the most important part of this table is the comparison between HFTr and non-HFTr book depth. It is not clear whether, once a market participant observed a given limit order, he would place his own limit order, place a marketable order, or withdraw from the market. In a general equilibrium setting it is not *a priori* clear whether the price impact would be larger or smaller than the results in Table 6.[37]

Next, I show how order book depth by different traders varies over time. While Table 6 presents the average price impact in different settings, Figure 4 depicts the time series price impact a 1000-share trade would have if it were to hit the order book. There are three graphs: the total price impact the trade would have with all available liquidity accessible, the price impact difference from removing HFTs' limit orders, and the price impact difference from removing non-HFTs' limit orders. The order book data is available only during 10 5-day windows. The X-axis identifies the first day in the 5-day window.[38] There is considerable variation over time in the overall depth of the order book and the amount provided by HFTs and non-HFTs. Even so, HFTs on average provide strictly less liquidity than do non-HFTs on each day.[39]

[37]In addition, this concern should be further dampened as market participants can always choose not to display their limit orders.

[38]That is, the observation 01-07-08 is followed by observations on January 8th, 9th, 10th, and 11th of 2008. The next observation is for April 7, 2008 and is followed by the next four consecutive trading days.

[39]The correlation coefficient between the VIX and the non-HFTs / HFTs book ratio is -.38, thus when expected volatility is high, the difference between HFTs and non-HFTs book depth narrows.

6.2.3 Examining Who Supplies Liquidity to Informed Traders

Here I test whether HFTs are able to avoid providing liquidity to informed traders. To test this I implement the Permanent Price Impact measure described and utilized in Section 6.1.1, but with adjustments. Whereas in Section 6.1.1 the Permanent Price Impact measure was used to determine which trader-type initiated trades that permanently moved prices, here I use the measure to determine which trader-type *supplies* liquidity to trades that permanently move prices.

I apply the same technique as the Price Impact measure but consider who is *supplying* liquidity to informed traders. Specifically, I change the q^H and q^N variables from Equation 7 to take their definitions from what type of trader is supplying liquidity and whether the liquidity supplier is a buyer or a seller. q^H equals +1 when a HFTr supplies liquidity and is selling, -1 when a HFTr supplies liquidity and is buying, and 0 when a HFTr is not supplying the liquidity, The q^N value is similarly defined for non-HFTs. I run the same VAR and VMA as in Section 6.1.1 to obtain $\sum_{t=0}^{10} b(L)$, which is interpreted as the private information content from a trade with liquidity supplied by a HFTr. The non-HFT impulse response function is $\sum_{t=0}^{10} c(L)$ and is the private information content of a trade supplied by a non-HFTr.

The results are in Table 5 Panel C. The column HFT is the private information from HFT- supplied trades and the nHFT column is the private information from non-HFT-supplied trades. If HFTs are better than non-HFTs at avoiding informed traders, then HFT should be less than nHFT. The results are in terms of basis points. I divide the stocks into three groups based on stock market capitalization. For each stock I estimate the statistical significance in the difference of the impulse response function for the HFT and non-HFT five trading days using a t-test. The t-test is adjusted using Newey-West standard errors to account for the time-series correlation in observations.

The results show that HFTs are better able to avoid informed traders for large market capitalization stocks. For the small stocks HFTs on average are more likely than non-HFTs to provide liquidity to informed traders. No small stock has HFT being larger than nHFT and the difference being statistically significant, but two have nHFT being larger than HFT and the difference being statistically significant. For the medium stocks, HFTs and non-HFTs are equally likely to supply liquidity to informed traders. Two stocks have HFT being larger than nHFT and the difference being statistically significant, and two have the difference being statistically significant in the other direction. For the large stocks, non-HFTs on average provide more liquidity to informed traders. Twenty stocks have nHFT being greater than HFT and the difference being statistically significant, while none have a statistically significant difference in the other direction. The results suggest that HFTs are more selective than non-HFTs in how they supply liquidity, especially in large stocks, and this culminates in their not matching the best bid or offer when they suspect they will be trading with informed traders.

6.3 High Frequency Trading's Impact on Volatility

The final market quality measure I analyze is the causal relationship between HFT and volatility. I have already considered volatility in previous areas, both the general relationship and also the impact of an exogenous shock to volatility on HFTs' market participation. In this section I examine whether HFT impacts volatility. I implement two methodologies to examine the question. First, I do an event study around the September 2008 short-sale ban. Second, I compare the price paths of stocks with and without HFTs being part of the data generation process in a partial equilibrium setting. Both methodologies suggest HFT tends to dampen volatility.

6.3.1 A Natural Experiment Around the Short-Sale Ban

In this section I utilize an exogenous shock to HFTs' activity to study the impact HFT has on volatility. The exogenous shock I study is the September 19, 2008 ban on short selling on 799 financial stocks.[40] Of the 120 stocks in the HFT dataset, 13 were on the ban list.

The ban indirectly stopped some HFTs from trading in the banned stocks. While the ban did not explicitly require HFT firms to stop trading the affected stocks, it did undermine their trading strategy. The strategies used by HFTs require them to be able to freely switch between being long or short a stock.[41] To verify that the short sale was a *de facto* ban on HFT, I graph the time series activity of HFTs' fraction of market involvement in Figure A-1.[42] The figure shows that HFTs' activity dropped precipitously for the 13 affected stocks during the ban. One reason HFT did not drop to zero for the affected stocks is that a portion of HFT firms were designated market makers and not subject to the short-sale ban restrictions.[43]

[40] The ban was in place until October 9, 2008.

[41] I cannot observe this in the data, but have been told by HFT firms that this is the case.

[42] The fraction of the market is normalized so that HFTs' fraction of trading in the affected and unaffected stocks are equal on September 1, 2010.

[43] The initial announcement of the short-sale ban provided a limited exemption for option market makers, but on September 22, 2008 the SEC extended the option market maker exemption to sell short the 799 affected stocks.

I use the variation in HFTs' fraction of market activity for the 13 affected stocks before and after the implementation of the short-sale ban to study how HFT impacts volatility. My findings are not picking up the impact of the short-sale ban on volatility as I analyze only the 13 affected stocks. I control for time-series variation in HFT and volatility by matching each affected stock with a stock in the unaffected group based on the two weeks prior to the short-sale ban HFT fraction of the market.

More specifically, I implement the following steps. First, I match each affected stock with one unaffected stock based on the average fraction of shares traded by HFTs in the two weeks prior to the short-sale ban. Second, I calculate F, the difference between the percent change in HFTs' fraction of a stock's activity before and after the short-sale ban implementation for the affected stock and its matched stock:

$$\mathrm{F}_i = \frac{\text{HFT-Lev}_{i,\text{affected, post}} - \text{HFT-Lev}_{i,\text{affected, pre}}}{\text{HFT-Lev}_{i,\text{affected, pre}}} - \frac{\text{HFT-Lev}_{i,\text{unaffected, post}} - \text{HFT-Lev}_{i,\text{unaffected, pre}}}{\text{HFT-Lev}_{i,\text{unaffected, pre}}}, \quad (15)$$

where HFT-Lev is the fraction of shares involving HFTs for the affected or unaffected stock in pair i either before (pre) the short-sale ban or after (post) it. I consider HFT-Lev for all HFTs' involvement, HFTs' demanding liquidity, and HFTs' supplying liquidity. The pre- and post-time periods refer to the day prior to (09/17/2008) and the day the short-sale ban went into effect (09/19/2008), the average value for the week prior to (09/11/2008 - 09/17/2008) and the week after (09/19/2008 - 09/25/2008) the start of the short-sale ban, or the average value for the full 11 days before (09/03/2008 - 09/17/2008) and during the ban (09/19/2008 - 10/05/2008). Third, I calculate σ, the difference between the volatility change before and after the short-sale ban implementation for the affected stock and its matched stock:

$$\sigma_i = \frac{\sigma\text{-Lev}_{i,\text{affected, post}} - \sigma\text{-Lev}_{i,\text{affected, pre}}}{\sigma\text{-Lev}_{i,\text{affected, pre}}} - \frac{\sigma\text{-Lev}_{i,\text{unaffected, post}} - \sigma\text{-Lev}_{i,\text{unaffected, pre}}}{\sigma\text{-Lev}_{i,\text{unaffected, pre}}}, \quad (16)$$

where σ-Lev is the daily summed 15-minute realized volatility for the affected or unaffected stock in pair i either before (pre) the short-sale ban or after (post) it. The pre- and post- time periods refer to the day before and after the short-sale ban, or the average value for the days before and after the short-sale ban. Finally, I implement an OLS regression to analyze the impact of HFT on volatility:

$$\sigma_i = \alpha + \mathrm{F}_i * \beta_1 + \epsilon_i, \quad (17)$$

The results are in Table 7. The first column, HFT - All, shows the results for all HFTs' involvement in the HFT-level definition, the second column shows the results for HFTs' demanding liquidity, and the third column shows the results for HFTs' supplying liquidity. Panel A shows the results from the 1-day analysis, Panel B shows the 1-week results, and Panel C shows the results for the full ban period. The 1-day analysis shows F being statistically significant at the 10% level and negative for the HFT - demand regression. The interpretation is that, as HFTs' ability to demand liquidity decreases, volatility increases. The results of the 1-week analysis show no statistical significance. The full period (11-day) results shows F being statistically significant at the 10% level and negative for the HFT - supply regression. The interpretation of this result is that as HFTs' ability to supply liquidity decreases, volatility increases. While the limited number of observations restrict the statistical significance of the results, the findings from the natural experiment suggest that HFTs' trading may dampen intraday volatility.

6.3.2 Comparing Actual and Alternative Price Paths

In this section I examine the impact HFTs have on volatility by comparing the observed volatility and the volatility that would have occurred had HFTs not participated in trading and had others kept their quoting and trading activity the same. While this methodology sets aside the well-established theoretical importance of a stock's price path on traders' behavior, it provides a partial equilibrium understanding of HFTs' impact on volatility.

I compare the actual price path with different hypothetical price paths assuming HFTs were not involved in the market in different capacities. I consider three hypothetical price paths: HFTs engage in no trading, HFTs demand liquidity but do not supply liquidity, and HFTs supply liquidity but do not demand any liquidity. To simulate HFTs not supplying any liquidity, I assume that the price impact for trades in which HFTs were providing liquidity was that which would occur if HFTs were not part of the order book. As the book depth is available only in one-minute snapshots, I assume the order book's depth remains constant around the stock price for trades occurring during the previous sixty seconds. To simulate HFTs not demanding liquidity, I assume that trades initiated by HFTs were not part of the price path. For the next hypothetical trade, I use the next trade to occur after time t that was initiated by a non-HFTr. I assume that after HFTs' activity was removed, the price path reverted and the remaining trades occurred at the same prices as actually occurred in the real price path. I calculate the daily 1-minute realized volatility for each stock on each day from February 22, 2010 to February 26, 2010.

Table A-10 shows the results. The t-statistics for the individual stocks use Newey-West standard errors to account for the time-series correlation. Panels A - C report the results. Panel A reports the actual and hypothetical price paths' volatility when removing all HFT activity. The volatility increases substantially across all stock sizes under this hypothetical price path. Panel B reports the hypothetical volatility when HFTs demand liquidity but do not supply any. The volatility increases across stock sizes here as well. Panel C reports the hypothetical volatility when HFTs supply liquidity but do not initiate trades. Here the volatility is unchanged. The results suggest HFT has no impact or reduces intraday volatility. The findings here, along with those in the short-sale ban volatility analysis, suggest that HFT may dampen volatility.

7 Conclusion

In this paper I examine HFT and its role in U.S. equity markets. I aim to provide a better understanding of the behavior of HFTs and their impact on market quality. I document that HFTs are involved in a large portion of U.S. equities activity. In the HFT dataset they are involved in 68.5% of the dollar-volume traded. Regarding HFTs' trading activity, I find that HFTs tend to follow a price reversal strategy driven by order imbalances and that their trading activity changes only moderately during the most volatile times. I also find that HFTs' strategies are more correlated than non-HFTs'. Finally, I find no evidence supporting HFTs engaging in an anticipatory trading. HFT is a profitable endeavor; I estimate HFTs earn gross trading profits of approximately $2.8 billion annually, and obtain an annualized pre-expense Sharpe ratio of 4.5.

I also analyze HFTs' impact on market quality, focusing on price discovery, liquidity, and volatility. I find HFTs add substantially to the price discovery process and that HFTs' supply of liquidity is mixed. They are frequently at the inside bid and offer, yet the depth of liquidity they provide on the order book is much less than that provided by non-HFTs. In addition, HFTs are strategic with their liquidity provisions and tend to avoid trading with informed traders. Finally, I find evidence that HFT dampens intraday volatility. Overall, these results suggest that HFTs' activities are not detrimental to non-HFTs and that HFT tends to improve market quality.

References

M.J. Barclay, T. Hendershott, and D.T. McCormick. Competition among trading venues: Information and trading on electronic communications networks. *Journal of Finance*, 58(6):2637–2665, 2003.

C. Cao, O. Hansch, and X. Wang. The information content of an open limit-order book. *Journal of Futures Markets*, 29(1):16–41, 2009.

A. Chaboud, E. Hjalmarsson, C. Vega, and B. Chiquoine. Rise of the machines: Algorithmic trading in the foreign exchange market. *International Finance Discussion Papers*, 2009.

J. Chae and A. Wang. Who makes markets? Do dealers provide or take liquidity. *MIT Working Paper*, 2003.

K. Chan and W.M. Fong. Trade size, order imbalance, and the volatility-volume relation. *Journal of Financial Economics*, 57(2):247–273, 2000.

T. Chordia and A. Subrahmanyam. Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics*, 72(3):485–518, 2004.

J.F. Coughenour and L. Harris. Specialist profits and the minimum price increment. *Univ. of Delaware Working Paper*, 2004.

J. Cvitanic and A.A. Kirilenko. High frequency traders and asset prices. *Cal. Tech. Working Paper*, 2010.

C Duhigg. Stock traders find speed pays, in milliseconds. *New York Times*, July 23, 2009.

D. Easley, M.M.L. de Prado, and M. O'Hara. The microstructure of the 'flash crash': Flow toxicity, liquidity crashes and the probability of informed trading. *Cornell Univ. Working Paper*, 2010.

K.Y.L. Fong and W.M. Liu. Limit order revisions. *Journal of Banking & Finance*, 34(8):1873–1885, 2010.

T. Foucault and A.J. Menkveld. Competition for order flow and smart order routing systems. *The Journal of Finance*, 63(1):119–158, 2008.

K.A. Froot, D.S. Scharfstein, and J.C. Stein. Herd on the street: Informational inefficiencies in a market with short-term speculation. *Journal of Finance*, 47(4):1461–1484, 1992.

M. Gsell. Assessing the impact of algorithmic trading on markets: A simulation approach. *Proceedings of the 16th European Conference on Information Systems*, 2008.

L.E. Harris and V. Panchapagesan. The information content of the limit order book: Evidence from NYSE specialist trading decisions. *Journal of Financial Markets*, 8(1):25–67, 2005.

J. Hasbrouck. Measuring the information content of stock trades. *Journal of Finance*, 46(1):179–207, 1991a.

J. Hasbrouck. The summary informativeness of stock trades: An econometric analysis. *Review of Financial Studies*, 4(3):571–595, 1991b.

J. Hasbrouck. One security, many markets: Determining the contributions to price discovery. *Journal of Finance*, pages 1175–1199, 1995.

J. Hasbrouck and G. Saar. Low-latency trading. *NYU Working Paper*, 2010.

J. Hasbrouck and G. Sofianos. The trades of market makers: An empirical analysis of NYSE specialists. *Journal of Finance*, 48(5):1565–1593, 1993.

J.A. Hausman, A.W. Lo, and A.C. MacKinlay. An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, 31(3):319–379, 1992.

T. Hendershott and P.C. Moulton. Speed and stock market quality: The NYSE's Hybrid. *UC Berkeley Working Paper*, 2010.

T. Hendershott and R. Riordan. Algorithmic trading and information. *UC Berkeley Working Paper*, 2009.

T. Hendershott, C.M. Jones, and A.J. Menkveld. Does algorithmic trading improve liquidity? *Journal of Finance, Forthcoming*, 2010.

M. Kearns, A. Kulesza, and Y. Nevmyvaka. Empirical limitations on high frequency trading profitability. *Univ. of Pennsylvania Working Paper*, 2010.

A.A. Kirilenko, A.S. Kyle, M. Samadi, and T. Tuzun. The flash crash: The impact of high frequency trading on an electronic market. *CFTC Working Paper*, 2010.

I. Rosu. A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641, 2009.

SEC. Concept release on equity market structure. January 14, 2010.

SEC and CFTC. Findings regarding the market events of may 6, 2010. September 30, 2010.

M. Van der Wel, A. Menkveld, and A. Sarkar. Are market makers liquidity suppliers? *VU Univ. Working paper*, 2008.

X. Vives. Short-term investment and the informational efficiency of the market. *Review of Financial Studies*, 8(1):125–160, 1995.

Figure 1: Time Series of High Frequency Trading's Fraction of Trade Activity. The figure shows the aggregate daily time series fraction of trades involving HFT in 2008 and 2009. The three graphs measure involvement differently. The first graph, HFT - Fraction of Dollar-Volume, is calculated using dollar-volume activity; the second graph, HFT - Fraction of Trades, considers the number of trades; the third graph, HFT - Fraction of Shares, uses the number of shares. Three lines appear in each graph, one for HFTs' percent of market activity in any capacity, another for the percent of market activity when a HFTr provides liquidity, and another for when a HFTr takes liquidity.

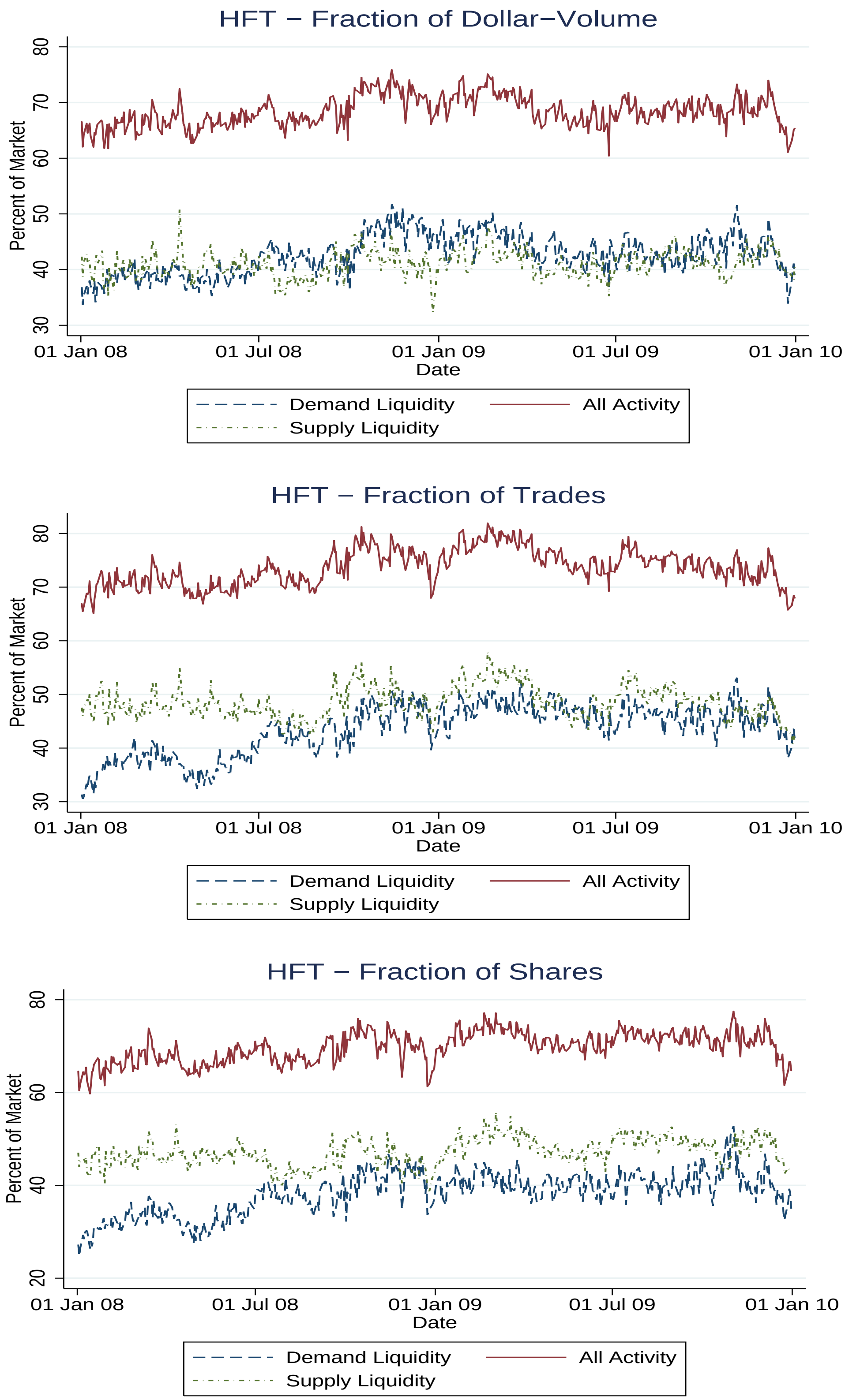
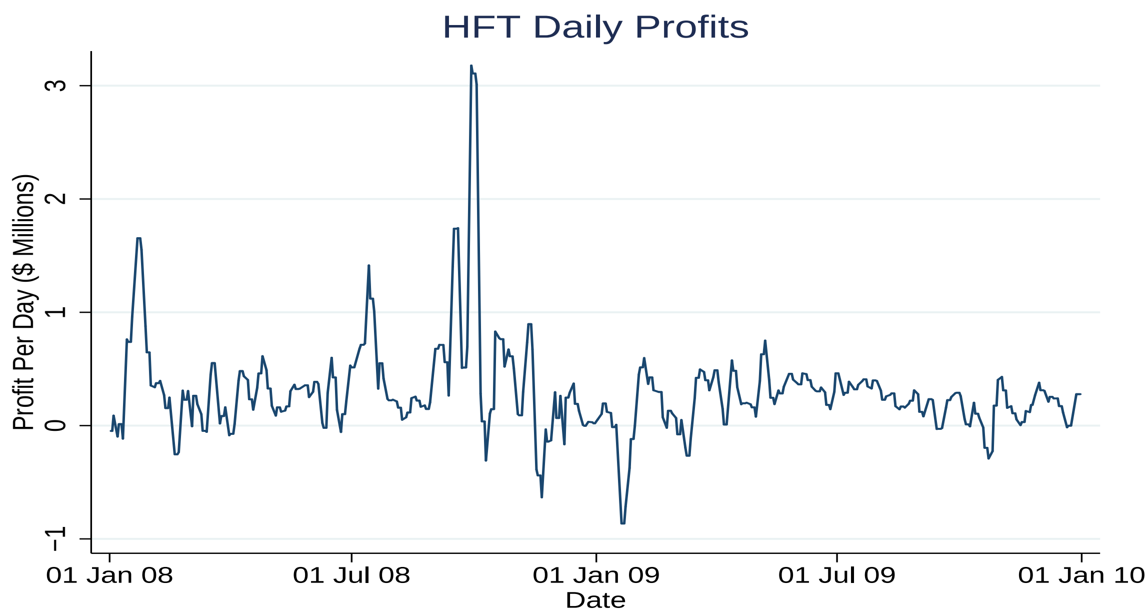


Figure 2: Time Series of the Daily Profitability of High Frequency Trading. The figure contains two graphs that show how profitability varies over time. The first graph, HFT Daily Profits, displays the centered 5-day moving average profits for HFTs. The daily profit value is calculated by summing HFTs' purchases and sales on that day for each stock. For any end-of-day inventory imbalance the required number of shares are assumed traded at the average share price for the day in order to end the day with a net zero position in each stock. The time series spans 2008 and 2009. The second graph, HFT Daily Profits - Intraday Volatility, includes both the centered 5-day moving average profits for HFTs but also a measure of intraday volatility. The intraday volatility measure is calculated for each day $\frac{SP_{High} - SP_{Low}}{SP_{Close}}$, where SP_{High} is the S&P 500's intraday high, SP_{Low} is the S&P 500's intraday low, and SP_{Close} is the S&P 500's closing value.



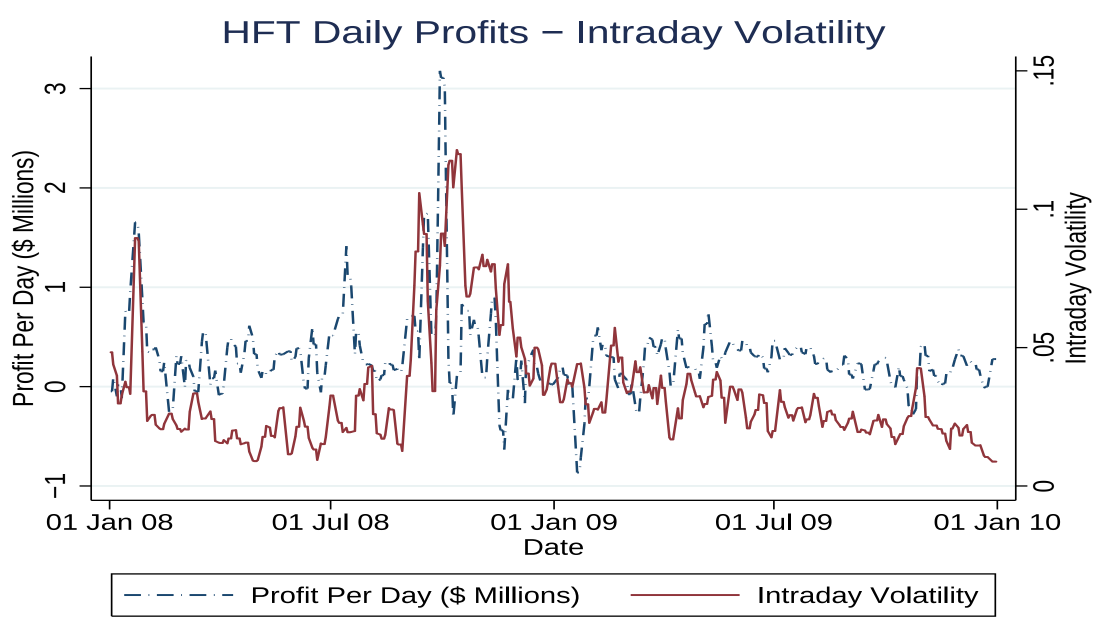


Figure 3: High Frequency Traders' Participation as Day-Level Volatility Changes. This figure depicts how HFTs' participation varies as day-level volatility changes. The X-axis represents 100 bins grouped together based on the V-Level value: V-Level$_{i,t} = \frac{V_{i,t} - E(V_i)}{E(V_i)} * \frac{1}{\sigma_i}$, where $V_{i,t}$ is the 15-minute realized volatility for stock i on day t, and σ_i is the standard deviation of stock i's V. The V-Level variable is the scaled deviation from the mean, where it is scaled by the standard deviation of a stock's volatility, σ_i. The Y-axis is the percent change from HFTs' average trading activity, HFT-Level$_j = \sum_{\text{V-Level}_{i,t} \epsilon j} \frac{1}{N_j} \left[\frac{\text{HFT}_{i,t} - E(\text{HFT}_i)}{E(\text{HFT}_i)} \right]$ where HFT is the fraction of shares in which HFTs are involved and j is the V-Level bin stock i at time t is falls into. N is the number of observations in bin j. There are three graphs, HFT and Volatiltiy - All, which looks at all HFTs activity, HFT and Volatility - Liquidity Supply, which only considers HFTs' supplying liquidity activity, and HFT and Volatility - Liquidity Demand, which only considers HFTs' demanding liquidity activity. In each graph there are four lines. The bin-by-bin HFT-Level, a nine-bin centered moving average of the HFT-Level, and the upper and lower 95 % confidence intervals.

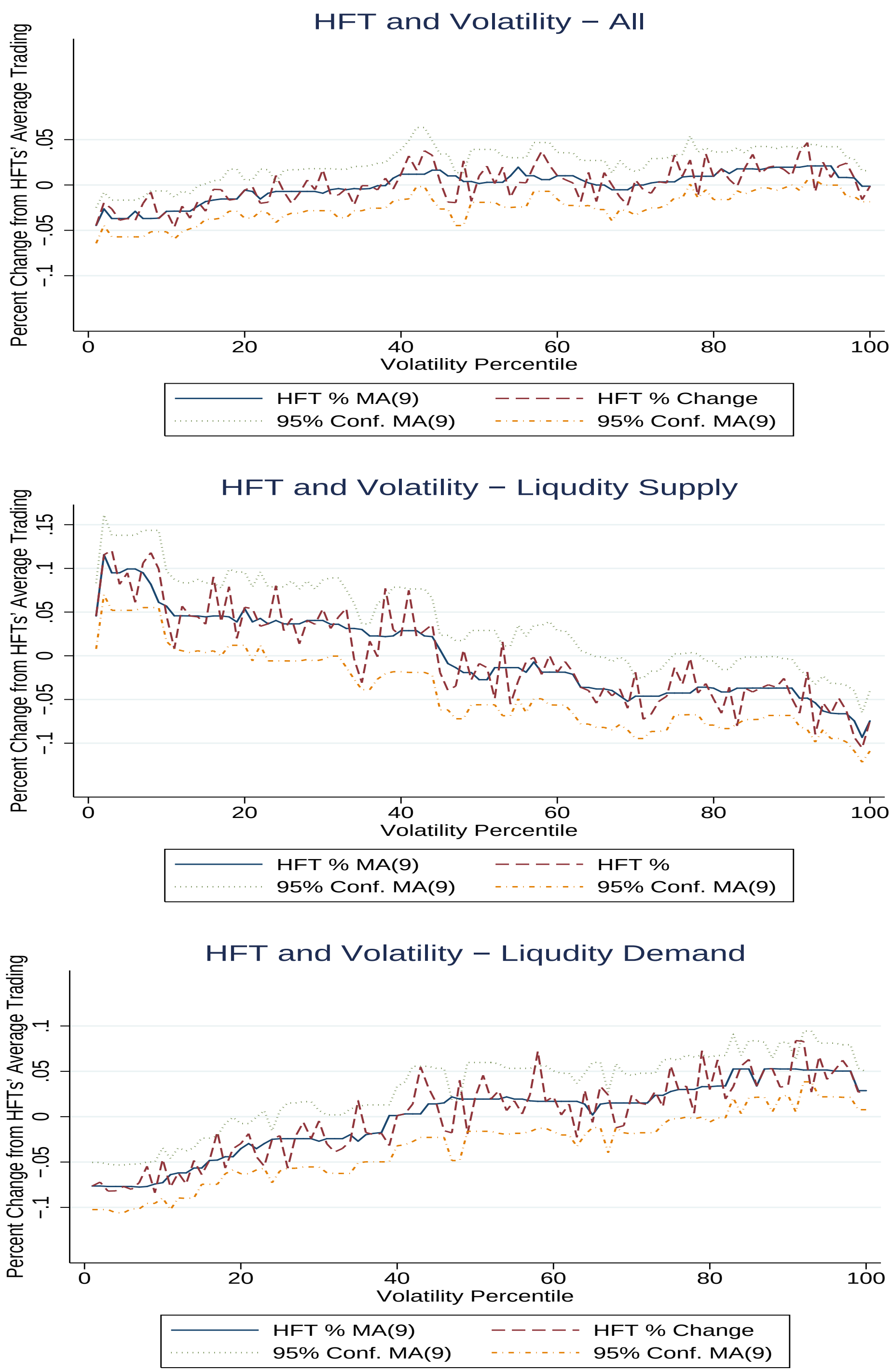


Figure 4: Time Series of High Frequency Traders' and Non High Frequency Traders' Book Depth. This Figure analyzes the depth of the order book and how much depth different types of traders provide by analyzing the price impact of a 1000 share trade hitting the order book with and without different types of traders. There are three graphs. The first, Price Impact of a 1000 Share Trade, examines the total price impact a 1000 share trade would have with all available liquidity accessible. The second graph, Additional Price Impact without HFTs on the Book, depicts the additional price impact that would occur from removing HFTs' limit orders The third graph, Additional Price Impact without non-HFTs on the Book, graphs the additional price impact from removing non-HFTs' limit orders. The daily dollar price impact value is calculated giving equal weight to each stock. The order book data is available during 10 5-day windows. The X-axis identifies the first day in the 5-day window. That is, The observation 01-07-08 is followed by observations on January 8th, 9th, 10th, and 11th of 2008. The next observation is for April 7, 2008 and is followed by the next four consecutive trading days. To separate the 5-day windows I enter a zero-impact trade, creating the evenly spaced troughs.

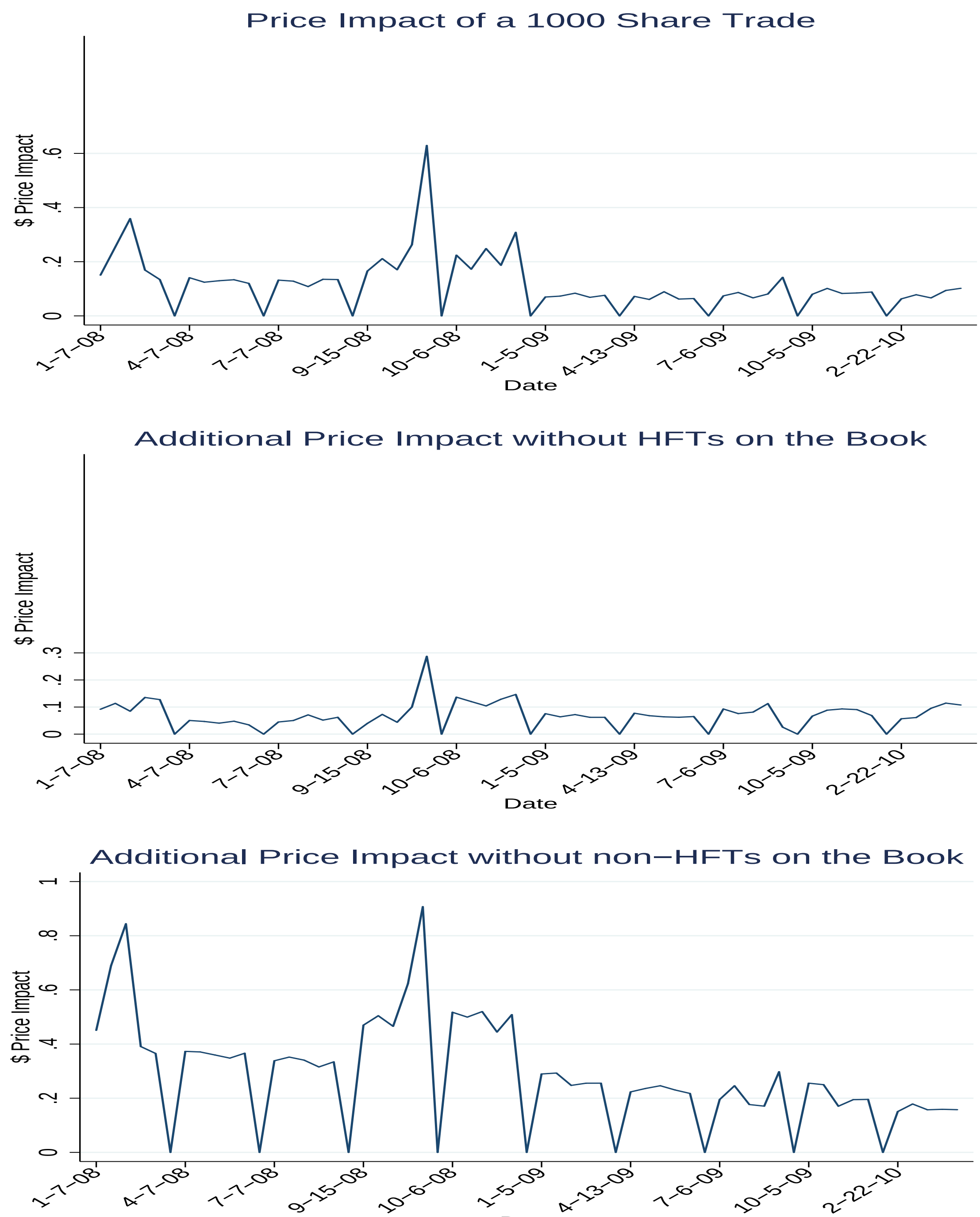


Table 1: High Frequency Trading's Fraction of Trade Activity. This table reports the aggregate day-level fraction of trading involving HFTs for all trades in the HFT dataset. Panel A measures the fraction of dollar-volume in which HFTs are involved , Panel B measures the fraction of trades involving HFTs, and Panel C measures the fraction of shares traded involving HFTs. Within each panel are three categories. The first category, HFT-All, reports the fraction of activity where HFTs are either demanding liquidity, supplying liquidity, or doing both. The second category, HFT-Demand, reports the fraction of activity where HFTs are demanding liquidity. The third category, HFT-Supply, shows the fraction of activity where HFTs are supplying liquidity. Within each category I report the findings by stock size, with each bin having 40 stocks and the row Overall reporting the unconditional results. The reported summary statistics include the mean, standard deviation, minimum, 5th percentile, 25th percentile, median, 75th percentile, 95th percentile, and maximum fraction of activity involving HFTs.

Panel A: HFT Dollar-Volume Market-wide Participation

Stock Size	Mean	Std. Dev.	Min.	5%	25%	50%	75%	95%	Max.
HFT-All									
Small	35.07	4.496	23.93	28	32.16	34.83	37.98	42.45	52.26
Medium	50.02	4.955	37.91	41.97	46.42	49.83	53.62	58.28	62.27
Large	69.36	2.73	61.31	65.01	67.38	69.19	71.33	74.11	76.51
Overall	68.49	2.762	60.44	64.23	66.49	68.27	70.48	73.3	75.85
HFT-Demand									
Small	24.4	5	11.44	16.73	20.92	24.16	27.55	32.9	46.27
Medium	36.43	5.099	21.06	27.94	32.77	36.76	40.29	44.38	49.35
Large	43.09	3.674	33.86	37.16	40.43	43.08	45.94	49.22	52.04
Overall	42.75	3.64	33.66	37.01	40.14	42.69	45.42	48.93	51.72
HFT-Supply									
Small	13.53	3.836	6.84	8.994	11.07	12.45	14.96	21.15	31.7
Medium	19.83	4.375	11.74	14.62	16.3	18.68	23	28.13	32.14
Large	42.06	2.398	33.29	38.08	40.41	41.99	43.83	45.97	51.7
Overall	41.13	2.393	32.37	37.15	39.49	41.07	42.86	45.05	50.75

Panel B: HFT Trades Market-wide Participation

HFT-All									
Small	37.03	4.218	26.7	30.72	34	37.01	39.56	43.87	54.16
Medium	57.94	5.495	44.68	48.83	53.69	58.07	62.33	66.68	71.36
Large	76.35	3.119	67.83	71.44	74.02	76.28	78.63	81.47	83.72
Overall	73.77	3.344	65.11	68.48	71.33	73.72	76.11	79.5	81.91
HFT-Demand									
Small	24.49	3.94	12.66	18.61	21.5	24.15	27.38	31.2	37.43
Medium	39.8	4.926	27.72	31.34	36.11	39.9	43.61	47.45	55.52
Large	44.58	4.939	30.94	35.39	40.95	45.73	48.41	51.03	54.67
Overall	43.64	4.759	30.52	34.74	40.19	44.78	47.31	49.93	53.12
HFT-Supply									
Small	15.93	4.38	7.093	10.51	12.92	15.14	18.19	23.79	36.38
Medium	27.99	5.437	17.08	20.45	23.18	27.25	32.53	36.82	41.58
Large	51.56	2.931	43.37	46.6	49.68	51.37	53.43	56.56	60.43
Overall	48.65	2.99	40.75	43.92	46.62	48.33	50.48	53.95	57.89

Panel C: HFT Shares Market-wide Participation

HFT-All									
Small	33.55	4.423	21.54	26.6	30.56	33.16	36.3	41.01	50.84
Medium	52.32	5.123	38.31	43.57	48.61	52.39	55.98	60.85	64.39
Large	71.61	3.224	61.51	66.09	69.23	71.87	74.12	76.15	79.32
Overall	69.88	3.274	59.76	64.43	67.33	70.2	72.37	74.73	77.47
HFT-Demand									
Small	22.1	4.144	10.38	16.11	19.15	21.73	24.91	29.34	40.05
Medium	34.51	4.608	22.5	26.77	31.14	34.88	37.85	41.88	47.3
Large	38.98	4.869	24.87	30.36	36.08	39.64	42.4	45.54	54.38
Overall	38.44	4.665	24.81	30.19	35.69	38.99	41.78	44.98	52.77
HFT-Supply									
Small	14.3	4.411	6.411	9	11.19	13.27	16.29	22.63	32.75
Medium	25.39	5.251	15.15	18.16	20.66	24.59	29.66	34.52	38.15
Large	49.2	2.911	40.73	44.25	47.27	49.33	51.38	53.48	57.19
Overall	47.27	2.989	38.73	42.31	45.34	47.32	49.46	51.83	55.77

Table 2: High Frequency Trading's Fraction of the Inside Quote Activity. This table reports the summary statistics based on HFTs best bid and offer activity in each stock on each day. Panels A - F report the percent of the time HFTs offer the same or better bid or offer as non-HFTs. Panel G reports the fraction of quote changing activity arising from HFTs. Panels A - C measure time based on calendar time; Panels D - F measure time based on tick time, weighting each observation equally regardless of the calendar time of the quote. Panels A and D report the unconditional percent of time HFTs are at the best bid or offer. I divide the observations into days on which the bid - ask spread is lower and higher than average for each stock. I report the conditional percent of time for days where the spread is lower than average in Panels B and E, and higher than average in Panels C and F. Within each category I report the findings by stock size, with each bin having 40 stocks and the row Total reporting the unconditional results. The reported summary statistics include the mean, standard deviation, minimum, 5th percentile, 25th percentile, median, 75th percentile, 95th percentile, and maximum. For Panels A - F the column "% of Inside" reports the average percent of the total shares HFTs provide at the inside quotes, conditional on HFTs being at the inside quotes. The data are from 02/22/2010 - 02/26/2010.

Panel A: All

Stock Size	Mean	Std. Dev.	% Of Inside	Min.	5%	25%	50%	75%	95%	Max.
Small	50.04	26.88	73.44	.0116	2.035	37.23	54.32	67.56	92.97	97.4
Medium	61.84	23.37	68.12	3.547	15.42	50.77	59.95	77.06	99.33	99.95
Large	84.47	20.58	59.85	21.57	32.44	79.41	94.26	99.2	99.89	99.95
Total	65.29	27.68	67.14	.0116	8.192	50.19	64.87	92.25	99.8	99.95

Panel B: Low Spread

Stock Size	Mean	Std. Dev.	% Of Inside	Min.	5%	25%	50%	75%	95%	Max.
Small	50.99	28.71	72.70	.0888	2.198	29.35	54.32	70.64	95.56	97.4
Medium	62.65	24.33	67.40	3.547	13.93	50.72	60.3	81.12	99.57	99.95
Large	80.53	22.69	60.48	21.57	32.34	67.73	91.61	98.94	99.89	99.95
Total	63.39	27.96	67.39	.0888	8.032	47.92	62.98	88.66	99.73	99.95

Panel C: High Spread

Stock Size	Mean	Std. Dev.	% Of Inside	Min.	5%	25%	50%	75%	95%	Max.
Small	48.78	24.35	74.40	.0116	1.355	43.86	54.25	62.28	83.77	96.02
Medium	60.51	21.77	69.32	6.606	22	50.82	58.68	68.11	98.33	99.82
Large	85.68	20.1	59.36	24.68	38.96	82.19	95.9	99.22	99.88	99.94
Total	67.14	27.19	66.83	.0116	11.38	51.53	65.13	95.42	99.82	99.94

Panel D: Tick Time - All

Stock Size	Mean	Std. Dev.	% Of Inside	Min.	5%	25%	50%	75%	95%	Max.
Small	51.58	27.52	76.99	.0905	2.366	38.64	59.85	69.46	89.32	96.42
Medium	62.37	22.47	72.22	1.653	17.74	50.37	60.76	76.37	98.2	99.74
Large	84.87	18.82	63.02	31.88	40.95	80.96	94.2	98.24	99.27	99.6
Total	66.28	27.01	70.74	.0905	7.731	51.49	65.51	92.35	99.07	99.74

Panel E: Tick Time - Low Spread

Stock Size	Mean	Std. Dev.	% Of Inside	Min.	5%	25%	50%	75%	95%	Max.
Small	53.98	28.93	75.88	.0905	2.25	36.71	61.6	72.28	94.56	96.42
Medium	64	23.14	71.39	1.653	11.38	53.72	63.05	79.13	98.57	99.74
Large	83.58	19.88	62.97	31.88	40.62	77.36	93.78	98.19	99.34	99.42
Total	65.97	27.14	70.61	.0905	8.096	53.17	66.41	89.91	99.06	99.74

Panel F: Tick Time - High Spread

Stock Size	Mean	Std. Dev.	% Of Inside	Min.	5%	25%	50%	75%	95%	Max.
Small	48.34	25.29	78.48	.2433	2.482	40.56	56.95	64.77	78.67	87.18
Medium	59.76	21.24	73.55	7.629	25.35	48.74	57.42	65.14	97.58	99.68
Large	85.97	17.89	63.07	34.92	43.19	82.17	94.78	98.24	99.22	99.6
Total	66.65	26.91	70.91	.2433	7.227	50.4	64.83	93.63	99.07	99.68

Panel G: HFT Percent of Quote Revisions / Changes

Stock Size	Mean	Std. Dev.	Min.	5%	25%	50%	75%	95%	Max.
Small	44.34	15.75	.9324	9.678	37.88	45.11	53.89	67.56	76.53
Medium	55.05	13.37	.4702	34.02	50.03	55.08	60.93	80.81	85.14
Large	68.42	7.75	44.72	50.58	64.31	69.46	73	79.53	83.99
Total	55.94	16.09	.4702	25.2	46.66	57.18	68.32	78.55	85.14

Table 3: 10-Second High Frequency Trading Determinants: Order Imbalance and Lagged Returns. This table reports the results from performing the logit regression: $\text{HFT}_{i,t} = \alpha + \text{Ret}_{i,1-10} * \beta_{1-10} + \text{OIB}_{i,1-10} * \beta_{11-20} + \text{OIB}_{i,1-10} * \text{Ret}_{i,1-10} * \beta_{21-30} + \epsilon_{i,t}$, $\text{HFT}_{i,s}$ takes on one of six definitions: (1) HFTs buying, (2) HFTs buying and supplying liquidity, (3) HFTs buying and demanding liquidity, (4) HFTs selling, (5) HFTs selling and supplying liquidity, and (6) HFTs selling and demanding liquidity. The dependent variable equals one if, on net, HFTs are engaging in that activity and zero otherwise. The explanatory variables include $\text{Ret}_{i,1-10}$, the return for stock i in period s, where s is the number of 10-second time periods prior to the time t, $\text{OIB}_{i,1-10}$, a dummy variable derived from the order imbalance that equals 1 if $\text{OIB}_{i,s} = \frac{\text{Buy Initiated Shares}_{i,s} - \text{Sell Initiated Shares}_{i,s}}{\text{Shares Outstanding}_{i,s}}$ is ≤ 0 for the Buy regressions, and ≥ 0 for the Sell regressions. Each explanatory variable is followed by a subscript between 1 and 10 that represents the number of lagged time periods away from the event occurring in the time t dependent variable. I perform the regression on all stocks from 02/22/2010 - 02/26/2010. Stock fixed effects are implemented and standard errors are clustered by stock. The table reports the results for the marginal effects at the means for the probability of the dependent variable equaling one for stock i at time t.

	(1) Buy - ALL	(2) Buy - Supply	(3) Buy - Demand	(4) Sell - ALL	(5) Sell - Supply	(6) Sell - Demand
Ret_1	2.414	-0.291	3.433**	-4.150***	0.920	-3.639***
	(1.622)	(0.992)	(1.141)	(1.193)	(0.994)	(0.852)
Ret_2	0.479	-0.872	1.724	-2.196	0.479	-1.717
	(1.503)	(1.077)	(0.966)	(1.527)	(0.967)	(1.095)
Ret_3	-3.872*	-2.692	-0.964	0.419	2.124*	-0.735
	(1.621)	(1.400)	(0.698)	(1.554)	(1.029)	(1.004)
Ret_4	-2.794	-3.535***	0.964	-2.024	1.701	-2.417*
	(1.458)	(0.999)	(0.980)	(1.429)	(1.085)	(0.953)
Ret_5	-2.104	-2.874**	0.949	-0.935	2.167*	-1.892
	(1.384)	(0.946)	(0.984)	(1.560)	(0.933)	(1.005)
Ret_6	-2.236	-3.421**	1.250	0.487	1.999*	-0.586
	(1.527)	(1.120)	(0.929)	(1.141)	(0.855)	(0.822)
Ret_7	-0.459	-2.432**	1.998*	-1.085	1.878	-1.857*
	(1.149)	(0.910)	(0.923)	(1.400)	(1.053)	(0.903)
Ret_8	-1.415	-2.344*	1.097	-1.004	0.385	-0.711
	(1.343)	(0.987)	(0.784)	(1.434)	(1.058)	(1.114)
Ret_9	0.818	-0.477	1.375	-1.648	1.679	-2.298**
	(1.481)	(0.930)	(1.079)	(1.327)	(0.957)	(0.870)
Ret_{10}	-0.262	-0.348	0.306	-2.694	0.651	-2.728**
	(1.390)	(1.044)	(0.888)	(1.413)	(0.903)	(1.033)
$\text{OIB}_1 * \text{Ret}_1$	-27.51***	-12.44***	-16.09***	24.94***	14.80***	11.79***
	(1.861)	(1.240)	(1.368)	(1.822)	(1.427)	(1.359)
$\text{OIB}_2 * \text{Ret}_2$	-17.82***	-8.165***	-10.40***	18.48***	10.52***	9.319***
	(1.583)	(1.311)	(1.164)	(1.799)	(1.264)	(1.489)
$\text{OIB}_3 * \text{Ret}_3$	-12.77***	-8.107***	-5.454***	14.64***	7.393***	8.507***
	(1.752)	(1.415)	(1.044)	(1.749)	(1.332)	(1.272)
$\text{OIB}_4 * \text{Ret}_4$	-12.08***	-6.067***	-6.776***	15.58***	7.262***	9.170***
	(1.789)	(1.261)	(1.244)	(1.827)	(1.342)	(1.377)
$\text{OIB}_5 * \text{Ret}_5$	-13.69***	-7.129***	-7.285***	13.15***	5.008***	8.788***
	(1.662)	(1.286)	(1.086)	(1.817)	(1.245)	(1.203)
$\text{OIB}_6 * \text{Ret}_6$	-12.50***	-4.875***	-7.963***	12.40***	5.824***	7.523***
	(1.876)	(1.348)	(1.214)	(1.595)	(1.321)	(1.146)
$\text{OIB}_7 * \text{Ret}_7$	-10.73***	-3.235*	-7.770***	11.26***	6.077***	6.078***
	(1.582)	(1.394)	(1.198)	(2.060)	(1.648)	(1.251)
$\text{OIB}_8 * \text{Ret}_8$	-8.208***	-3.056*	-5.590***	9.174***	6.316***	4.030***
	(1.695)	(1.251)	(1.106)	(1.845)	(1.497)	(1.177)
$\text{OIB}_9 * \text{Ret}_9$	-11.89***	-5.715***	-6.559***	12.32***	5.023***	7.991***
	(2.015)	(1.445)	(1.214)	(1.529)	(1.389)	(1.123)
$\text{OIB}_{10} * \text{Ret}_{10}$	-7.928***	-2.888	-5.320***	10.22***	5.209***	6.022***
	(1.966)	(1.476)	(1.220)	(1.697)	(1.096)	(1.292)
N	1389393	1389393	1377873	1389393	1354772	1389393

Marginal effects; Standard errors in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Analysis of whether High Frequency Traders Systematically Engage in Anticipatory Trading. The table shows the results from an analysis to detect whether HFTs engages in anticipatory trading. I analyze the frequency of observing different marketable order sequences. If trading is independent of the trader type it should be equally likely to observe a HFTr-initiated trade prior to a non-HFTr-initiated trade as it is to observe the reverse, a non-HFTr-initiated trade prior to a HFTr-initiated trade. Let $T_{t-1}T_t$ represent the sequence of trades where T is the type of trader, H for a HFTr and N for a non-HFTr, and t is the transaction time sequence of events. If systematic anticipatory trading by HFTs is occurring then I would see: Prob$(HN) >$ Prob(NH) and if it were not occurring I would observe Prob$(HN) \leq$ Prob(NH). Prob() is defined as Prob$(x) = \frac{n_x}{n_t}$ where x represents the desired sequence of trades, n_x represents the total number of times such a sequence is observed and n_t represents the total number of sequences observed. For each stock and each day I analyze the probability of seeing different trading patterns. I consider the two-, three-, four-, five-, and six-period sequence. The five different calculations are: $AT_1 =$ Prob(HN) - Prob (NH), $AT_2 =$ Prob(HHN) - Prob (NHH), $AT_3 =$ Prob(HHHN) - Prob (NHHH), $AT_4 =$ Prob(HHHHN) - Prob (NHHHH), and $AT_5 =$ Prob(HHHHHN) - Prob (NHHHHH), where the trade sequences represent times $(T_{t-5}, T_{t-4}, T_{t-3}, T_{t-2}, T_{t-1}, T_t)$. An $AT > 0$ is consistent with anticipatory trading, while an $AT \leq$ is not. I calculate the statistical significance incorporating Newey-West standard errors to correct for the time-series correlation in observations. I summarize the results based on stock market capitalization. Panel A shows the results for seeing the sequence under analysis regardless of trade size. Panel B shows the results for seeing the sequence under analysis conditional on the N trade being for more than 500 shares. That is, $AT_1 =$ Prob$(HN|N_{\text{shares}} > 500) -$ Prob$(NH|N_{\text{shares}} > 500)$, where N_{shares} represents the number of shares traded in the non-HFTr initiated trade.

Panel A - Unconditional Sequence Probabilities

Sequence	AT_1			AT_2			AT_3			AT_4			AT_5		
		Stat. Sign.			**Stat. Sign.**			**Stat. Sign.**			**Stat. Sign.**			**Stat. Sign.**	
Stock Size	**Mean**	**> 0**	**< 0**	**Mean**	**> 0**	**< 0**	**Mean**	**> 0**	**< 0**	**Mean**	**> 0**	**< 0**	**Mean**	**> 0**	**< 0**
Small	-0.0012	0	25	-0.0007	0	18	-0.0005	0	18	-0.0003	1	18	-0.0002	1	14
Medium	-0.0022	0	38	-0.0016	0	32	-0.0011	1	30	-0.0008	1	25	-0.0006	3	23
Large	-0.0028	0	40	-0.0026	0	40	-0.0021	0	40	-0.0017	0	40	-0.0013	0	40
Overall	-0.0018	0	103	-0.0014	0	90	-0.0010	1	88	-0.0008	2	83	-0.0006	4	77

Panel B - Conditional Sequence Probabilities

Sequence	AT_1			AT_2			AT_3			AT_4			AT_5		
		Stat. Sign.			**Stat. Sign.**			**Stat. Sign.**			**Stat. Sign.**			**Stat. Sign.**	
Stock Size	**Mean**	**> 0**	**< 0**	**Mean**	**> 0**	**< 0**	**Mean**	**> 0**	**< 0**	**Mean**	**> 0**	**< 0**	**Mean**	**> 0**	**< 0**
Small	-0.0001	0	15	0.0	0	12	0.0	0	7	0.0	0	6	0.0	0	6
Medium	-0.0001	0	30	-0.0001	0	23	0.0	0	16	0.0	0	16	0.0	0	12
Large	-0.0003	0	40	-0.0002	0	40	-0.0002	0	40	-0.0002	0	40	-0.0001	0	35
Overall	-0.0002	0	85	-0.0001	0	75	-0.0001	0	62	-0.0001	0	62	-0.0001	0	53

Table 5: The Permanent Price Impact of Trades by High Frequency Traders and Non High Frequency Traders. This table reports results from different permanent price impact measures calculated for each stock on each day. Panel A reports the average long-run (10 events in the future) impulse response function for HFTr- and non-HFTr-initiated trades. The interpretation is that the larger a trader's impulse response function the more private information from that trader's trades. Panel B reports the average long-run - short-run (10 events in the future - the contemporaneous event) impulse response function for HFTs and non-HFTs. A larger long run - short run impulse response difference suggests the information in the initial trade takes time to be imputed, whereas a negative difference suggests there is an immediate overreaction to the trade's information. Panel C reports the average long-run impulse response function for HFTr- and non-HFTr-supplied trades. A larger value suggests that trader-type provides more liquidity to informed traders. In Panel A and B I define q^H and q^N based on who is demanding liquidity. In Panel C I define the q based on the liquidity *supplier's* activity and trader type. I define q^H to be a +1 when a HFTr liquidity supplier sells and -1 when a HFTr liquidity supplier buys, The q^N value is similarly defined for non-HFTr supplied trades. Each panel reports the average percent of days when HFTs' price impact is greater than non-HFTs'. The table also reports the number of stocks for which the difference between HFT and non-HFT is statistically significant. In each panel I group the results into three groups based on stock market capitalization, and also report the overall results. I perform this analysis for 02/22/2010 - 02/26/2010. I calculate statistical significance incorporating Newey-West standard errors to correct for the time-series correlation in observations.

Panel A: Long-Run Impulse Response Functions

Stock Size	HFT	Std. Dev. HFT	non-HFT	Std. Dev. HFT	Mean % Days HFT > non-HFT	Stat. Sign HFT < non-HFT	Stat. Sign HFT > non-HFT
Small	3.60	18.06	1.08	9.31	64	3	0
Medium	2.31	2.81	1.20	1.91	74.21	1	12
Large	1.07	0.41	0.81	0.41	78.46	0	21
Overall	2.15	9.14	1.02	4.75	73.33	4	33

Panel B: Long-Run - Short-Run Impulse Response Functions

Stock Size	HFT	Std. Dev. HFT	non-HFT	Std. Dev. HFT	Mean % Days HFT > non-HFT	Stat. Sign HFT < non-HFT	Stat. Sign HFT > non-HFT
Small	1.25	9.98	-0.77	8.69	0.61	0	1
Medium	0.81	2.49	0.33	1.75	0.64	1	14
Large	0.54	0.36	0.36	0.32	0.81	0	37
Overall	0.81	5.17	0.07	4.45	0.70	1	52

The Supply of Liquidity to Informed Traders: The Permanent Price Impact of Trades based on who Supplies Liquidity.

Panel C: Liquidity Providing Impulse Response Functions

Stock Size	HFT	Std. Dev. HFT	non-HFT	Std. Dev. HFT	Mean % Days HFT > non-HFT	Stat. Sign HFT < non-HFT	Stat. Sign HFT > non-HFT
Small	2.76	10.99	1.45	5.24	58.40	0	2
Medium	1.20	2.07	1.17	1.90	45.26	2	2
Large	0.55	0.45	0.91	0.60	12.82	20	0
Overall	1.33	5.64	1.14	2.87	36.08	22	4

Table 6: High Frequency Traders and Non High Frequency Traders Book Depth. This table reports the partial equilibrium price impact resulting from different sized market buy orders hitting the book with and without different liquidity providers. Panel A reports the average price impact different size trades would have given the displayed and hidden liquidity on the HFT dataset book snapshots. Panel B and C look at the *additional* price impact in partial equilibrium that an average trade of varying sizes would have if certain liquidity providers did not have orders on the order book. Panel B considers the additional price impact that would occur if there were no HFTs on the order book. Panel C considers the impact if no non-HFTs were on the book. I provide the analysis for stock sizes from 100 - 1000 in increasing trade size increments. I Winsorize the upper tail of the price impact values at the 99.5% level. I report both the basis point price impact and also the dollar size price impact. I divide the results into three even groups based on stock size and also report the overall results. I find similar results for market sell orders that are not reported.

Panel A: Price Impact of Trade - All Orders Available

Stock Size	Small		Medium		Large		All	
Trade Size	Basis	Dollars	Basis	Dollars	Basis	Dollars	Basis	Dollars
100	28.186	0.047	7.754	0.023	0.968	0.006	12.503	0.026
200	44.564	0.082	12.934	0.039	1.785	0.012	20.077	0.045
300	57.099	0.108	17.071	0.051	2.476	0.017	25.952	0.060
400	66.522	0.130	20.979	0.063	3.074	0.021	30.661	0.072
500	77.965	0.150	24.769	0.075	3.657	0.026	36.013	0.085
600	88.908	0.169	28.082	0.085	4.224	0.030	41.032	0.096
700	96.987	0.185	30.901	0.094	4.736	0.033	44.890	0.105
800	104.228	0.200	33.767	0.103	5.244	0.037	48.478	0.115
900	110.736	0.215	36.730	0.113	5.748	0.041	51.848	0.124
1000	122.937	0.241	40.541	0.125	6.283	0.045	57.450	0.139

Panel B: Additional Price Impact without HFTs on the Book

Stock Size	Small		Medium		Large		All	
Trade Size	Basis	Dollars	Basis	Dollars	Basis	Dollars	Basis	Dollars
100	25.468	0.035	8.099	0.024	0.941	0.006	11.684	0.022
200	32.082	0.047	9.091	0.027	1.194	0.008	14.350	0.028
300	36.670	0.057	9.885	0.030	1.428	0.011	16.254	0.033
400	41.509	0.068	11.392	0.035	1.708	0.013	18.496	0.039
500	47.953	0.080	13.195	0.041	1.978	0.016	21.381	0.046
600	54.177	0.092	14.670	0.046	2.276	0.019	24.090	0.053
700	60.055	0.104	16.129	0.051	2.582	0.022	26.679	0.060
800	66.104	0.118	17.791	0.057	2.899	0.026	29.397	0.068
900	73.009	0.136	19.604	0.064	3.233	0.029	32.462	0.077
1000	77.594	0.142	21.463	0.071	3.589	0.033	34.761	0.083

Panel C: Additional Price Impact without non-HFTs on the Book

Stock Size	Small		Medium		Large		All	
Trade Size	Basis	Dollars	Basis	Dollars	Basis	Dollars	Basis	Dollars
100	324.482	0.412	72.360	0.209	9.171	0.056	137.654	0.228
200	364.129	0.469	89.941	0.257	10.749	0.066	157.540	0.267
300	386.254	0.495	102.242	0.289	12.046	0.074	169.605	0.289
400	401.202	0.511	111.877	0.313	13.327	0.082	178.330	0.305
500	404.739	0.517	118.162	0.330	14.448	0.089	181.997	0.315
600	420.002	0.530	125.549	0.349	15.745	0.097	190.084	0.329
700	423.937	0.531	129.333	0.359	16.849	0.104	193.047	0.334
800	435.158	0.539	133.601	0.369	17.972	0.111	198.659	0.343
900	437.783	0.540	136.075	0.374	18.955	0.117	200.699	0.344
1000	438.884	0.541	137.231	0.377	19.948	0.122	201.265	0.344

Table 7: High Frequency Trading's Impact on Volatility: The Short-Sale Ban Natural Experiment. This table shows the results of a natural experiment conducted around the short-sale ban of September, 2008 to analyze how HFT impacts intraday volatility. Thirteen stocks in the HFT dataset were part of the ban. I match each affected stock with one unaffected stock based on the average fraction of shares traded by HFTs in the two weeks prior to the short-sale ban. Next, I calculate F, the difference between the percent change in HFTs' fraction of a stock's activity before and after the short-sale ban implementation for the affected stock and its matched stock: $\mathrm{F}_i = \frac{\text{HFT-Lev}_{i,\text{affected, post}} - \text{HFT-Lev}_{i,\text{affected, pre}}}{\text{HFT-Lev}_{i,\text{affected, pre}}} - \frac{\text{HFT-Lev}_{i,\text{unaffected, post}} - \text{HFT-Lev}_{i,\text{unaffected, pre}}}{\text{HFT-Lev}_{i,\text{unaffected, pre}}}$, where HFT-Lev is the fraction of shares involving HFTs for the affected or unaffected stock in pair i either before (pre) the short-sale ban or after (post) it. I consider HFT-Level for all HFTs' involvement, HFTs' demanding liquidity, and HFTs' supplying liquidity. The pre and post time periods refer to the day prior to (09/17/2008) and the day the short-sale ban went into effect (09/19/2008) for Panel A, the average value for the week prior to (09/11/2008 - 09/17/2008) and the week after (09/19/2008 - 09/25/2008) the start of the short-sale ban for Panel B, or the average value for the full 11 days before (09/03/2008 - 09/17/2008) and during (09/19/2008 - 10/05/2008) the ban in Panel C. I then calculate σ, the difference between the volatility change before and after the short-sale ban implementation for the affected stock and its matched stock: $\sigma_i = \frac{\sigma\text{-Lev}_{i,\text{affected, post}} - \sigma\text{-Lev}_{i,\text{affected, pre}}}{\sigma\text{-Lev}_{i,\text{affected, pre}}} - \frac{\sigma\text{-Lev}_{i,\text{unaffected, post}} - \sigma\text{-Lev}_{i,\text{unaffected, pre}}}{\sigma\text{-Lev}_{i,\text{unaffected, pre}}}$, where σ-Lev is the daily summed 15 minute realized volatility for the affected or unaffected stock in pair i either before (pre) the short-sale ban or after (post) it. Finally, I implement an OLS regression analysis to analyze the impact of HFT on volatility: $\sigma_i = \alpha + \mathrm{F}_i * \beta_1 + \epsilon_i$.

Panel A: 1-day Before and After Short-Sale Ban

	HFT - All	HFT - Demand	HFT - Supply
F	-0.703	-1.482^{+}	0.0225
	(0.790)	(0.716)	(0.220)
Constant	0.889**	0.676*	1.063***
	(0.283)	(0.260)	(0.216)
Observations	13	13	13
Adjusted R^2	-0.018	0.215	-0.090

Panel B: 1-week Before and After Short-Sale Ban

	HFT - All	HFT - Demand	HFT - Supply
F	-0.415	-0.732	-0.0726
	(0.552)	(0.558)	(0.249)
Constant	0.314	0.236	0.401*
	(0.200)	(0.189)	(0.169)
Observations	13	13	13
Adjusted R^2	-0.037	0.056	-0.083

Panel C: Full Period Before and After Short-Sale Ban

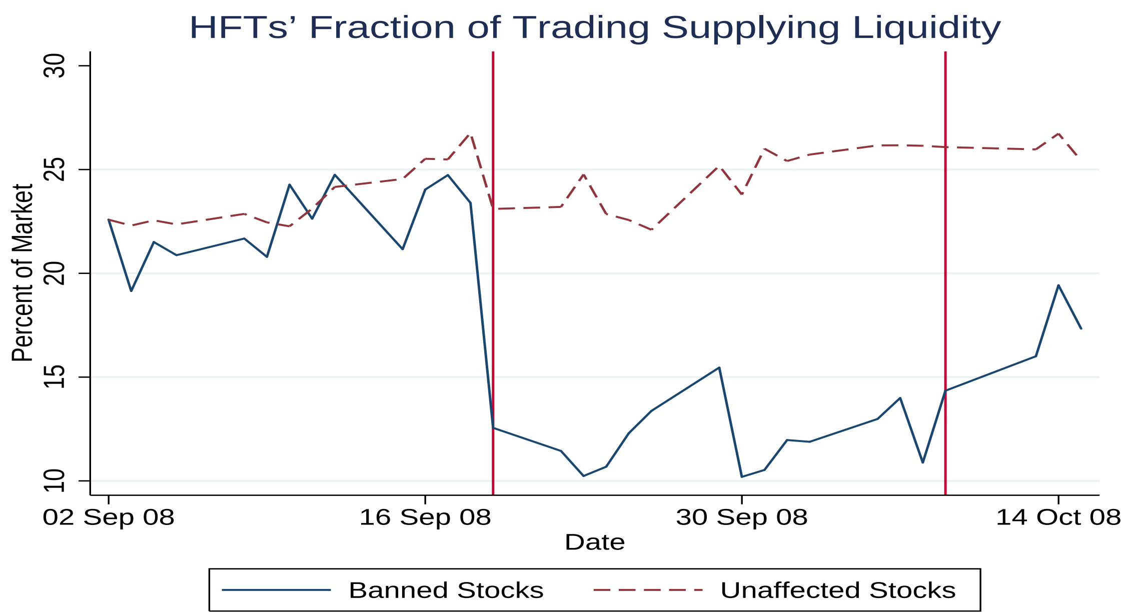
	HFT - All	HFT - Demand	HFT - Supply
F	-0.694	-0.484	-0.276^{+}
	(0.389)	(0.391)	(0.149)
Constant	0.0494	0.123	0.126
	(0.125)	(0.117)	(0.0907)
Observations	13	13	13
Adjusted R^2	0.154	0.042	0.168

Marginal effects; Standard errors in parentheses

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix A: Figures and Tables

Figure A-1: High Frequency Trading's Fraction of the Market around the Short-Sale Ban. The figure shows how HFTs' fraction of dollar-volume trading varied surrounding the September 19, 2008 SEC imposed short-sale ban on many financial stocks. In the HFT dataset, 13 stocks are in the ban. The two graphs plot HFTs' fraction of dollar-volume traded for the banned stocks and for the unaffected stocks. The first graph reports the fraction of dollar-volume where HFTs supplied liquidity. The second reports the fraction where HFTs demanded liquidity. I normalize the banned stocks' percent of the market in both graphs so that the affected and unaffected stocks have the same percent of the market on September 2, 2008. The two vertical lines represent the first and last day of the short-sale ban.



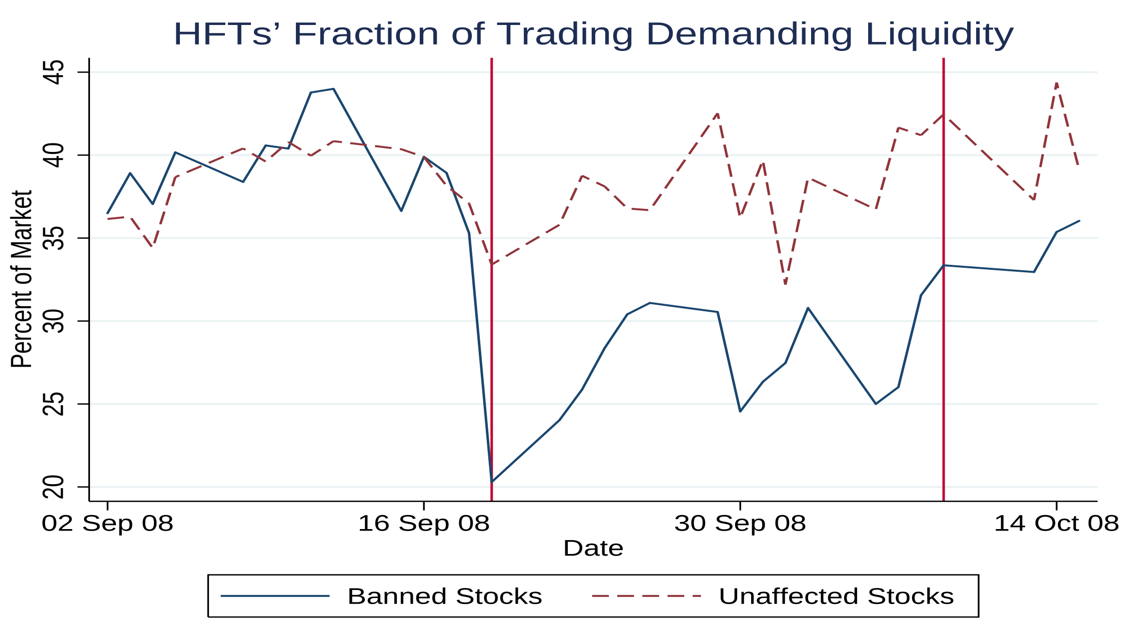


Figure A-2: High Frequency Traders' Participation during Intra-day Price Increases. The figure shows how HFTs' participation varies during different magnitude price increases. The X-axis represents 100 bins based on the Ret-Level value: $\text{Ret-Level}_{i,t,m} = \frac{\text{Ret}_{i,t,m}}{\sigma_i}$, where $\text{Ret}_{i,t,m}$ is the maximum return for stock i on day t during period m. That is, Ret is calculated based on the maximum and minimum price during a 15-minute period. If the minimum price occurred prior to the maximum price it is considered a positive return period. The higher the percentile, the larger the price increase. The Y-axis is the percent change from HFTs' average trading activity, $\text{HFT-Level}_j = \sum_{\text{Ret-Level}_{i,t,m} \epsilon j} \frac{1}{N_j} \left[\frac{\text{HFT}_{i,t,m} - E(\text{HFT}_i)}{E(\text{HFT}_i)} \right]$, where HFT takes on five definitions, which are the graphs' labels: All , Liquidity Demand - Buy, Liquidity Supply - Buy, Liquidity Demand - Sell, and Liquidity Supply - Sell. The definition of HFT refers to the fraction of dollar-volume traded that satisfies the specified trading criteria: where the Buy/Sell refers to HFTs' action, and Liquidity Supply/ Liquidity Demand refers to their liquidity role in the transaction. I remove observations where the return was 0 for that period or where less than 30 trades occurred. In each graph there are four lines. The bin-by-bin HFT-Level, a five-bin centered moving average of the HFT-Level, and the upper and lower 95 % confidence intervals.

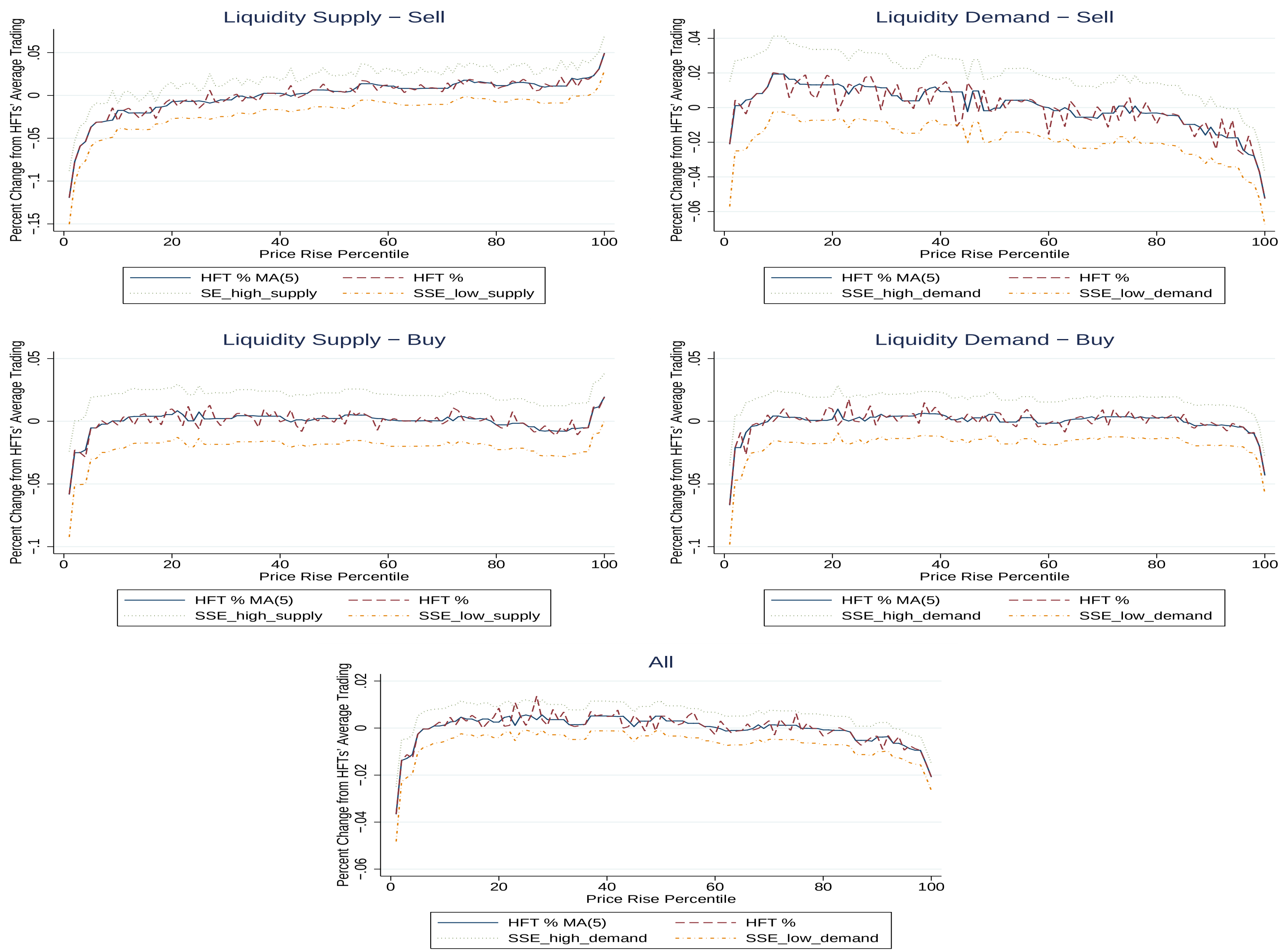


Figure A-3: High Frequency Traders' Participation during Intra-day Price Decreases. The figure shows how HFTs' participation varies during different magnitude price decreases. The X-axis represents 100 bins based on the Ret-Level value: $\text{Ret-Level}_{i,t,m} = \frac{\text{Ret}_{i,t,m}}{\sigma_i}$, where $\text{Ret}_{i,t,m}$ is the maximum return for stock i on day t during period m. That is, Ret is calculated based on the maximum and minimum price during a 15-minute period. If the minimum price occurred after the maximum price it is considered a negative return period. The higher the percentile, the larger the price decrease. The Y-axis is the percent change from HFTs' average trading activity, $\text{HFT-Level}_j = \sum_{\text{Ret-Level}_{i,t,m} \epsilon j} \frac{1}{N_j} \left[\frac{\text{HFT}_{i,t,m} - E(\text{HFT}_i)}{E(\text{HFT}_i)}\right]$, where HFT takes on five definitions, which are the graphs' labels: All , Liquidity Demand - Buy, Liquidity Supply - Buy, Liquidity Demand - Sell, and Liquidity Supply - Sell. The definition of HFT refers to the fraction of dollar-volume traded that satisfies the specified trading criteria: where the Buy/Sell refers to HFTs' action, and Liquidity Supply/ Liquidity Demand refers to their liquidity role in the transaction. I remove observations where the return was 0 for that period or where less than 30 trades occurred. In each graph there are four lines. The bin-by-bin HFT-Level, a five-bin centered moving average of the HFT-Level, and the upper and lower 95 % confidence intervals.

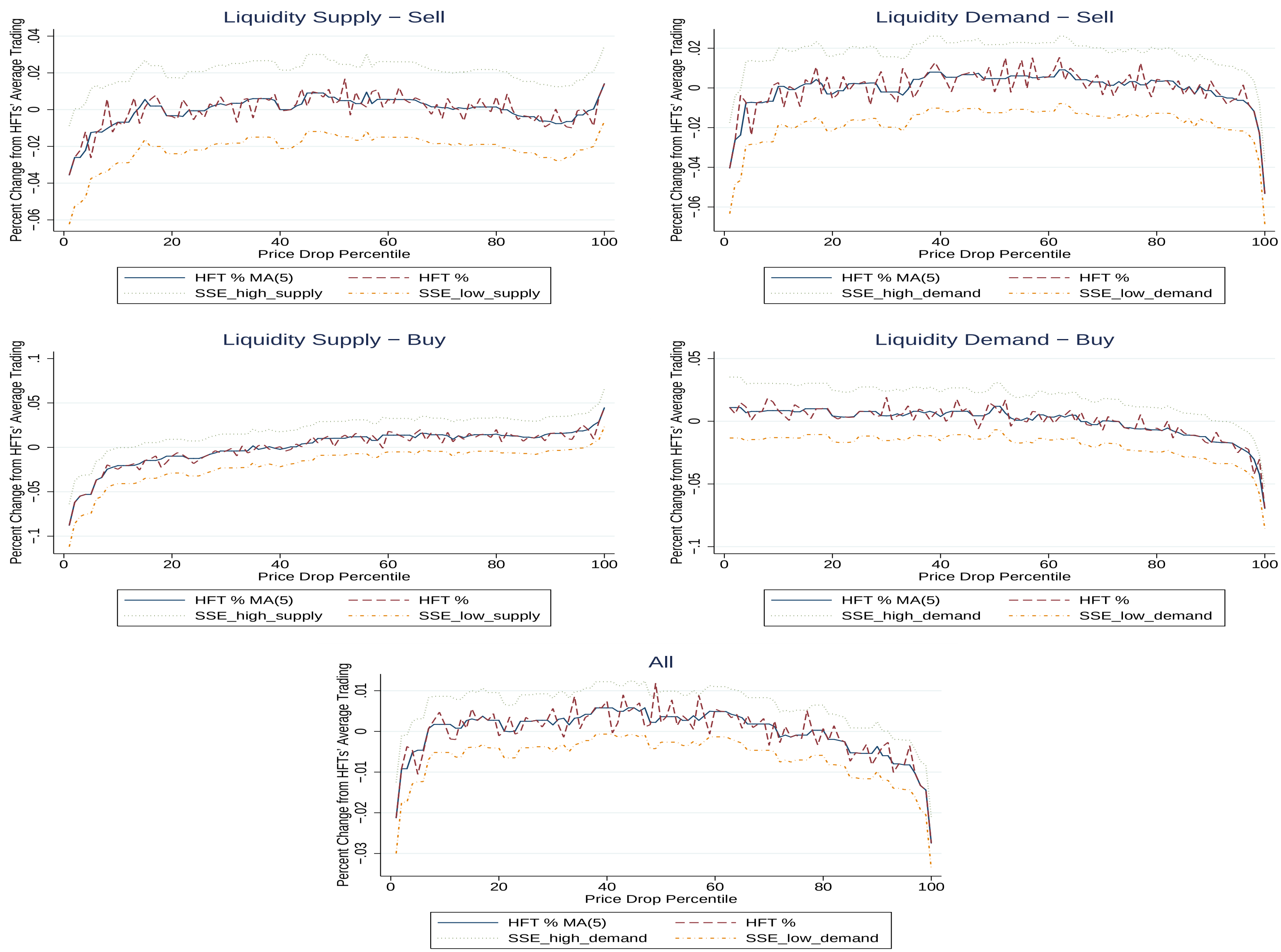


Table A-1: The 120 Stocks in the High Frequency Trading Dataset. This table reports the tickers and company names for the stocks in the HFT dataset. The columns Small, Medium, and Large are based on stock market capitalization as of the end-of-year 2009 Compustat valuation. The largest market capitalization in Small is $.5 billion and the smallest in Medium is $1.1 billion. The largest market capitalization in Medium is $3.7 billion and the smallest in Large is $11.7 billion.

Stock Size	**Small**		**Medium**		**Large**
Ticker	**Stock Name**	**Ticker**	**Stock Name**	**Ticker**	**Stock Name**
ABD	ACCO BRANDS CORP	AMED	AMEDISYS INC	AA	ALCOA INC
AINV	APOLLO INVESTMENT CORP	ARCC	ARES CAPITAL CORP	AAPL	APPLE INC
ANGO	ANGIODYNAMICS INC	AYI	ACUITY BRANDS INC	ADBE	ADOBE SYSTEMS INC
APOG	APOGEE ENTERPRISES INC	BARE	BARE ESCENTUALS INC	AGN	ALLERGAN INC
AZZ	AZZ INC	BRE	BRE PROPERTIES	AMAT	APPLIED MATERIALS INC
BAS	BASIC ENERGY SERVICES INC	BXS	BANCORPSOUTH INC	AMGN	AMGEN INC
BW	BRUSH ENGINEERED MATERIALS	CBT	CABOT CORP	AMZN	AMAZON.COM INC
BZ	BOISE INC	CCO	CLEAR CHANNEL OUTDOOR	AXP	AMERICAN EXPRESS CO
CBEY	CBEYOND INC	CETV	CENTRAL EUROPEAN MEDIA	BHI	BAKER HUGHES INC
CBZ	CBIZ INC	CHTT	CHATTEM INC	BIIB	BIOGEN IDEC INC
CDR	CEDAR SHOPPING CENTERS INC	CKH	SEACOR HOLDINGS INC	BRCM	BROADCOM CORP
CPSI	COMPUTER PROGRAMS & SYSTEMS	CNQR	CONCUR TECHNOLOGIES INC	CB	CHUBB CORP
CRVL	CORVEL CORP	COO	COOPER COMPANIES INC	CELG	CELGENE CORP
CTRN	CITI TRENDS INC	CPWR	COMPUWARE CORP	CMCSA	COMCAST CORP
DCOM	DIME COMMUNITY BANCSHARES	CR	CRANE CO	COST	COSTCO WHOLESALE CORP
DK	DELEK US HOLDINGS INC	CRI	CARTER'S INC	CSCO	CISCO SYSTEMS INC
EBF	ENNIS INC	CSE	CAPITALSOURCE INC	CTSH	COGNIZANT TECH SOLUTIONS
FFIC	FLUSHING FINANCIAL CORP	CSL	CARLISLE COS INC	DELL	DELL INC
FPO	FIRST POTOMAC REALTY TRUST	ERIE	ERIE INDEMNITY CO	DIS	DISNEY (WALT) CO
FRED	FREDS INC	EWBC	EAST WEST BANCORP INC	DOW	DOW CHEMICAL
IMGN	IMMUNOGEN INC	FCN	FTI CONSULTING INC	EBAY	EBAY INC
IPAR	INTER PARFUMS INC	FL	FOOT LOCKER INC	ESRX	EXPRESS SCRIPTS INC
KNOL	KNOLOGY INC	FMER	FIRSTMERIT CORP	GE	GENERAL ELECTRIC CO
KTII	K-TRON INTERNATIONAL INC	FULT	FULTON FINANCIAL CORP	GENZ	GENZYME CORP
MAKO	MAKO SURGICAL CORP	GAS	NICOR INC	GILD	GILEAD SCIENCES INC
MDCO	MEDICINES CO	ISIL	INTERSIL CORP	GLW	CORNING INC
MFB	MAIDENFORM BRANDS INC	JKHY	HENRY (JACK) & ASSOCIATES	GOOG	GOOGLE INC
MIG	MEADOWBROOK INS GROUP INC	LANC	LANCASTER COLONY CORP	GPS	GAP INC
MOD	MODINE MANUFACTURING CO	LECO	LINCOLN ELECTRIC HLDGS INC	HON	HONEYWELL INTERNAT. INC
MRTN	MARTEN TRANSPORT LTD	LPNT	LIFEPOINT HOSPITALS INC	HPQ	HEWLETT-PACKARD CO
MXWL	MAXWELL TECHNOLOGIES INC	LSTR	LANDSTAR SYSTEM INC	INTC	INTEL CORP
NC	NACCO INDUSTRIES	MANT	MANTECH INTL CORP	ISRG	INTUITIVE SURGICAL INC
NXTM	NXSTAGE MEDICAL INC	MELI	MERCADOLIBRE INC	KMB	KIMBERLY-CLARK CORP
PBH	PRESTIGE BRANDS HOLDINGS	NSR	NEUSTAR INC	KR	KROGER CO
PPD	PREPAID LEGAL SERVICES INC	NUS	NU SKIN ENTERPRISES	MMM	3M CO
RIGL	RIGEL PHARMACEUTICALS INC	PNY	PIEDMONT NATURAL GAS CO	MOS	MOSAIC CO
ROCK	GIBRALTAR INDUSTRIES INC	PTP	PLATINUM UNDERWRITERS	PFE	PFIZER INC
ROG	ROGERS CORP	ROC	ROCKWOOD HOLDINGS INC	PG	PROCTER & GAMBLE CO
RVI	RETAIL VENTURES INC	SF	STIFEL FINANCIAL CORP	PNC	PNC FIN. SVCS GROUP INC
SJW	SJW CORP	SFG	STANCORP FINANCIAL GROUP	SWN	SOUTHWESTERN ENERGY

Table A-2: Comparing the High Frequency Trading Dataset to the Compustat and TAQ Datasets. Panel A in this table compares the HFT and Compustat datasets. The Compustat stocks included in the comparison consist of all stocks with data available and that are listed on either Nasdaq or NYSE, which includes 4,238 stocks. The table reports market capitalization, market-to-book ratio, industry, and listing exchange summary statistics and provides the t-statistic for the differences in means. The Compustat and the HFT dataset data are for stocks 2009 end-of-year report. If a stock's year-end was not 12/31/2009, I use the most recent preceding annual figures. The industries are categorized based on the Fama-French ten industry groups. Panel B compares the market characteristics of the HFT dataset to stocks in the TAQ database that are listed on NYSE or Nasdaq and have the requisite data, which results in 4,045 stocks. I look at each stock's market characteristics for the five trading days 02/22/2010 - 02/26/2010. For each dataset I report the half spread, stock price, bid size, offer size, daily volume traded, number of trades, and size of a trade. For each variable I report the t-statistic for the differences in means between the two datasets.

Panel A: HFT Database v. Compustat Database

	HFT Database					Compustat Database					
	Mean	**Std. Dev.**	**Min.**	**Median**	**Max.**	**Mean**	**Std. Dev.**	**Min.**	**Median**	**Max.**	**T-Test**
Market Cap. (millions)	17588.24	37852.38	80.6025	1743	197012.3	3367	13881	.0229	403.3	322334	10.2063
Market-to-Book	2.650	3.134	-11.7799	1.837	20.0406	13.81	690.8	-688.5	1.452	44844	-0.1770
Industry - Non-Durables	.0333	.1802				.04389	.2049				-0.5583
Industry - Durables	.025	.1567				.01982	.1394				0.3999
Industry - Manufacturing	.1666	.3742				.09297	.2904				2.7169
Industry - Energy	.0083	.0912				.04578	.209				-1.9568
Industry - High Tech	.1583	.3665				.1824	.3862				-0.6740
Industry - Telecom.	.05	.2188				.02737	.1632				1.4819
Industry - Wholesale	.0916	.2897				.08377	.2771				0.3076
Industry - Health Care	.15	.3585				.1081	.3105				1.4522
Industry - Utilities	.0333	.1802				.02855	.1666				0.3094
Industry - Other	.2833	.4525				.3702	.4829				-1.9469
Exchange - NYSE	.5	.5020				.3905	.4879				2.4221
Exchange - Nasdaq	.5	.5020				.6095	.4879				-2.4221
Observations	120					4238					

Panel B: HFT Database v. TAQ Database

	HFT Database					TAQ Database					
Variable	**Mean**	**Std. Dev.**	**Min.**	**Median**	**Max.**	**Mean**	**Std. Dev.**	**Min.**	**Median**	**Max.**	**T-Test**
Quoted Half Spread (Dollars)	.0711	.0863	.0057	.0479	.7675	.1349	.2016	.005	.07142	2.786	-7.699
Stock Price (Dollars)	35.42	40.02	4.623	26.46	347.4	26.11	30.96	5.01	18.90	705.95	11.72
Bid Size (Hundreds of Shares)	23.88	68.69	1.216	3.615	989.3	15.55	231.1	1	3.47	18770	0.877
Offer Size (Hundreds of Shares)	24.24	70.69	1.115	3.857	882.6	14.31	186.8	1	3.642	10441	1.294
Daily Volume Traded (Millions of Dollars)	31.46	76.1	.1053	5.77	1174	8.721	25.93	.0053	1.206	1174	19.105
Number of Trades	3090	3326	66	1420	15472	1371	2005	1	552.5	20996	20.113
Size of a Trade (Shares)	207.5	172.3	104	160.3	3263	243.3	457.9	1	171.3	33333	-1.902
Observations	600					20225					

Table A-3: Day-Level Determinants of High Frequency Trading's Fraction of Trade Activity. This table shows the result of an OLS regression with observations at the stock and day level. I implement the regression: $\mathrm{H}_{i,t} = \alpha + \mathrm{MC}_i * \beta_1 + \mathrm{MB}_t * \beta_2 + \mathrm{VIX}_i * \beta_3 + \sigma_{i,t} * \beta_4 + \mathrm{SP}_{i,t} * \beta_5 + \mathrm{DEP}_{i,t} * \beta_6 + \mathrm{TS}_{i,t} * \beta_7 + \mathrm{NV}_{i,t} * \beta_8 + \mathrm{AC}_{i,t} * \beta_9 + \epsilon_{i,t}$, where $\mathrm{H}_{i,t}$ is the fraction of shares traded involving a HFTr in stock i on day t, MC is the log market capitalization as of December 31, 2009, MB is the market to book ratio as of December 31, 2009, which is Winsorized at the 99th percentile, VIX is the S&P 500 Chicago Board of Exchange Volatility Index (scaled by 10^{-3}), σ is the ten-second realized volatility summed up over the day (scaled by 10^{-5}), SP is the average time-weighted dollar spread between the bid and offer (scaled by 10^{-1}), DEP is the average time-weighted depth available at the inside bid and ask in dollars (scaled by 10^{-3}), TS is the average dollar-volume size of a non-HFTr-only trade (trades where non-HFTs both supplied liquidity and demanded it) (scaled by 10^{-5}), NV is the dollar-volume of non-HFTr-only transactions, normalized by market capitalization (and scaled by 10^{-6}), and AC is the absolute value of a one-period autoregressive process (AR(1)) analyzed at ten-second intervals (scaled by 10^{-2}). Standard errors are clustered by stock. Columns 1 - 6 show the standardized beta coefficients and columns 7 - 12 display the regular coefficients. The dependent variables for the column are: the percent of shares HFTs are involved in for 1, 4, 7, and 10; the percent of shares in which HFTs demand liquidity for 2, 5, 8, and 11 ; the percent of shares in which HFTs supply liquidity for 3, 6, 9, and 12. Columns 4 - 6 and 10 - 12 show the results for the restricted regression analysis.

	Standardized Beta Coefficients						OLS Coefficients					
	All (1)	**Dem. (2)**	**Sup. (3)**	**All (4)**	**Dem. (5)**	**Sup. (6)**	**All (7)**	**Dem. (8)**	**Sup. (9)**	**All (10)**	**Dem. (11)**	**Sup. (12)**
Market Cap. (MC)	0.791***	0.558***	0.773***	0.768***	0.529***	0.772***	0.082***	0.043***	0.066***	0.080***	0.041***	0.066***
							(0.0050)	(0.0050)	(0.0041)	(0.0038)	(0.0043)	(0.0040)
Market / Book (MB)	-0.129**	-0.088	-0.124***	-0.139**	-0.043	-0.193***	-0.013**	-0.007	-0.011***	-0.015**	-0.003	-0.017***
							(0.0044)	(0.0044)	(0.0029)	(0.0045)	(0.0046)	(0.0038)
VIX	0.030***	0.100***	-0.050***	0.037***	0.099***	-0.036***	0.462**	1.173***	-0.638***	0.582***	1.163***	-0.455***
							(0.1638)	(0.1709)	(0.1443)	(0.1546)	(0.1467)	(0.1111)
Volatility (σ)	-0.002	-0.010***	0.008				-0.239	-0.761***	0.645			
							(0.1944)	(0.1691)	(0.3367)			
Average Spread (SP)	0.047*	0.013	0.053				0.440*	0.091	0.410			
							(0.1829)	(0.3513)	(0.3316)			
Average Depth (DEP)	0.049	-0.172***	0.264***				0.314	-0.833***	1.402***			
							(0.2005)	(0.1875)	(0.2392)			
Non HFT Trade Size (TS)	-0.058*	0.097*	-0.203***				-0.213*	0.267*	-0.610***			
							(0.1019)	(0.1077)	(0.1442)			
Non HFT $-Volume (NV)	0.006	-0.007	-0.001				0.457	-0.386	-0.030			
							(1.2807)	(1.1864)	(0.8853)			
Autocorrelation (AC)	0.001	-0.013	0.016***				0.052	-0.349	0.466***			
							(0.3829)	(0.3707)	(0.0912)			
Constant							-0.145***	-0.040	-0.236***	-0.127***	-0.029	-0.231***
							(0.0375)	(0.0371)	(0.0269)	(0.0306)	(0.0321)	(0.0276)
Observations	60800	60800	60800	60995	60995	60995	60800	60800	60800	60995	60995	60995
Adjusted R^2	0.549	0.306	0.620	0.545	0.278	0.544	0.549	0.306	0.620	0.545	0.278	0.544

Standardized beta coefficients for (1) - (6); Standard errors statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-4: 10-Second Determinants of High Frequency Trading: Testing Potential Determinants. This table shows the results from performing an ordered logit analysis to determine what factors influence HFTs' buying and selling decision. I run the following ordered logit regression: $\text{HFT}_{i,t} = \alpha + \beta_{1-11} * \text{Ret}_{i,t,0-10} + \beta_{12-22} * \text{SP}_{i,t,0-10} + \beta_{23-33} * \text{DEPB}_{i,t,0-10} + \beta_{34-44} * \text{DEPA}_{i,t,0-10} + \beta_{45-55} * \text{NT}_{i,t,0-10} + \beta_{56-66} * \text{NV}_{i,t,0-10} + \epsilon_{i,t}$, where HFT is -1 during the ten-second period t HFTs were, on net, selling shares of stock i, it is 0 if HFTs performed no transactions or they bought as many shares as they sold, and it is 1 if, on net, HFTs purchased shares. DEPB is the average time-weighted best bid depth in dollars. DEPA is the average time-weighted best ask depth in dollars. SP is the average time-weighted spread, where spread is the best offer price minus the best bid price. NT is the number of non-HFTr trades that occurred, and NV is the non-HFTr dollar-volume of shares exchanged. Stock fixed effects are implemented and standard errors are clustered by stock. I include the contemporaneous and lagged values for each of the explanatory variables. Each explanatory variable has a subscript $0 - 10$. This represents the number of lagged time periods away from the event occurring in the time t dependent variable. Subscript 0 represents the contemporaneous value for that variable. Thus, the betas represent row vectors of 1x11 and the explanatory variables column vectors of 11x1. I perform the regression on all stocks from 02/22/2010 - 02/26/2010. Stock fixed effects are implemented and standard errors are clustered by stock. The table reports the results for the marginal effects at the means for the probability of a HFTr buying stock i at time t.

Variable	**Coefficient**	**Std. Error**	**Variable**	**Coefficient**	**Std. Error**
Ret_0	-5.555	(15.81)	DEPA_0	-1.74e-7	(1.39e-7)
Ret_1	-5.238***	(1.572)	DEPA_1	7.03e-8	(8.33e-8)
Ret_2	-5.012***	(1.189)	DEPA_2	1.00e-7**	(3.88e-8)
Ret_3	-6.445***	(1.134)	DEPA_3	-3.40e-8	(5.91e-8)
Ret_4	-4.565***	(1.002)	DEPA_4	8.99e-9	(4.71e-8)
Ret_5	-4.533***	(0.875)	DEPA_5	5.12e-8	(5.58e-8)
Ret_6	-4.471***	(0.815)	DEPA_6	7.37e-8	(4.75e-8)
Ret_7	-3.131***	(0.733)	DEPA_7	-7.24e-8	(5.02e-8)
Ret_8	-1.677*	(0.693)	DEPA_8	2.13e-8	(3.36e-8)
Ret_9	-1.928*	(0.752)	DEPA_9	2.96e-8	(4.63e-8)
Ret_{10}	-1.038	(0.589)	DEPA_{10}	-2.51e-8	(4.29e-8)
SP_0	0.000851	(0.00646)	NT_0	-0.00158**	(0.000585)
SP_1	-0.000905	(0.00741)	NT_1	0.00137**	(0.000463)
SP_2	-0.00415	(0.00432)	NT_2	-0.0000161	(0.000306)
SP_3	0.00738	(0.00513)	NT_3	-0.000632*	(0.000310)
SP_4	-0.000591	(0.00758)	NT_4	0.000618*	(0.000278)
SP_5	0.00301	(0.00517)	NT_5	-0.000489*	(0.000240)
SP_6	-0.00592	(0.00736)	NT_6	0.000396	(0.000320)
SP_7	-0.00563	(0.00951)	NT_7	-0.000183	(0.000327)
SP_8	0.0108	(0.00727)	NT_8	0.000164	(0.000272)
SP_9	-0.00247	(0.00490)	NT_9	0.000331	(0.000275)
SP_{10}	-0.00884	(0.00717)	NT_{10}	-0.000176	(0.000279)
DEPB_0	2.09e-7	(1.47e-7)	NV_0	1.53e-7	(1.42e-7)
DEPB_1	-7.25e-8	(8.33e-8)	NV_1	-1.40e-7	(1.35e-7)
DEPB_2	-1.20e-7**	(4.59e-8)	NV_2	-9.23e-8	(2.88e-7)
DEPB_3	-4.86e-8	(5.13e-8)	NV_3	2.30e-7	(2.25e-7)
DEPB_4	-6.04e-8	(4.00e-8)	NV_4	-1.97e-7	(1.41e-7)
DEPB_5	-1.20e-7	(1.13e-7)	NV_5	2.87e-7**	(1.09e-7)
DEPB_6	8.71e-8	(6.75e-8)	NV_6	-8.86e-8	(1.77e-7)
DEPB_7	-1.12e-8	(5.94e-8)	NV_7	3.55e-8	(1.28e-7)
DEPB_8	-5.30e-8	(3.21e-8)	NV_8	1.72e-7	(1.34e-7)
DEPB_9	-8.53e-9	(6.09e-8)	NV_9	-3.63e-7	(1.97e-7)
DEPB_{10}	1.79e-7*	(7.80e-8)	NV_{10}	3.54e-7	(2.58e-7)
N	1356143				

Marginal effects; Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-5: 10-Second Determinants of High Frequency Trading: Lagged Returns Importance Based on Trade and Liquidity Type. This table reports the results from performing a logit regression with dependent variable equal to 1 if during the ten-second interval (1) HFTs, on net, sell in a given ten-second period, (2) HFTs, on net, sell and supply liquidity, and (3) HFTs, on net, sell and demand liquidity, (4) HFTs, on net, buy in a given ten-second period, (5) HFTs, on net, buy and supply liquidity, and (6) HFTs, on net, buy and demand liquidity, and 0 otherwise. Each explanatory variable is followed by a subscript between 1 and 10. This represents the number of lagged time periods away from the event occurring in the time t dependent variable. I perform the regression on all stocks from 02/22/2010 - 02/26/2010. Stock fixed effects are implemented and standard errors are clustered by stock. The table reports the results for the marginal effects at the means for the probability of the dependent variable equaling one for stock i at time t.

	(1)	(2)	(3)	(4)	(5)	(6)
	Buy - All	**Buy - Supply**	**Buy - Demand**	**Sell - All**	**Sell - Supply**	**Sell - Demand**
Ret_1	-6.747***	-6.128***	-0.432	5.234**	-0.911	7.567***
	(1.678)	(1.035)	(1.223)	(1.652)	(0.945)	(0.775)
Ret_2	-5.675***	-4.484***	-1.169	5.186***	1.228	4.709***
	(1.308)	(1.000)	(0.814)	(1.352)	(0.842)	(0.853)
Ret_3	-7.347***	-4.152***	-3.356**	6.429***	2.565***	4.476***
	(1.206)	(0.591)	(1.049)	(1.109)	(0.663)	(0.722)
Ret_4	-6.252***	-2.916***	-3.522***	4.880***	1.299	4.272***
	(1.193)	(0.761)	(0.895)	(1.177)	(0.776)	(0.790)
Ret_5	-6.333***	-2.508**	-4.038***	4.239***	1.903*	2.688***
	(1.018)	(0.819)	(0.719)	(1.112)	(0.853)	(0.737)
Ret_6	-6.171***	-2.882***	-3.442***	5.014***	2.190***	3.250***
	(0.997)	(0.652)	(0.793)	(1.039)	(0.616)	(0.736)
Ret_7	-3.294**	-1.809*	-1.542*	3.256***	0.390	3.507***
	(1.043)	(0.765)	(0.691)	(0.882)	(0.601)	(0.809)
Ret_8	-2.316*	-1.496*	-0.849	0.943	-0.131	1.338*
	(0.946)	(0.633)	(0.804)	(0.933)	(0.765)	(0.647)
Ret_9	-2.668**	-1.278	-1.478*	2.520*	0.278	2.747**
	(1.004)	(0.801)	(0.647)	(0.981)	(0.662)	(0.849)
Ret_{10}	-2.068*	-2.114**	0.0880	0.285	-0.958	1.639*
	(0.908)	(0.648)	(0.764)	(0.926)	(0.685)	(0.747)
N	1377798	1366278	1377798	1377798	1377798	1343177

Marginal effects; Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-6: Testing whether High Frequency Traders Strategies are More Correlated than Non High Frequency Traders. This table reports the difference in frequencies of different types of trade pairs (demander - supplier) to analyze how correlated HFTs' strategies are with each other compared to how correlated non-HFTs' strategies are with each other. There are four trade partner combinations, HH, HN, NH, NN, where the first letter represents the liquidity demander and the second the liquidity supplier and N represents a non-HFTr and H a HFTr. The table reports information regarding R. R = $\frac{\text{Prob(NN)}}{\text{Prob(NH)}} - \frac{\text{Prob(HN)}}{\text{Prob(HH)}}$. Let RN = $\frac{\text{Prob(NN)}}{\text{Prob(NH)}}$ be the non-HFTr demanding liquidity ratio and RH = $\frac{\text{Prob(HN)}}{\text{Prob(HH)}}$ be the HFTr demanding liquidity ratio. R will equal zero when non-HFTs will take liquidity from other non-HFTs in the same proportion as HFTs take liquidity from other HFTs. Therefore, if R $= 0$, HFTs and non-HFTs trade with each other as much as expected when their trading strategies are equally correlated. If R > 0 then either HFTs trade with each other less than expected or HFTs trade with non-HFTs more than expected. In the data Prob() is calculated as Prob(x) = $\frac{n_x}{n_t}$ where x represents the desired trade liquidity supplier and demander, n_x represents the total number of times such a transaction is observed and n_t represents the total number of transactions observed. Panel A shows the results when considering the entire trading day. Panel B shows the results when analyzing the last 15-minutes of the trading day. I use all trade data in the HFT dataset. I calculate statistical significance incorporating Newey-West standard errors to correct for the time-series correlation in observations.

Panel A - All Day Trading

Stock Size	**Mean R**	**Std. Dev. R**	**Mean % Days R $>$ 0**	**Stat. Sign $<$ 0**	**Stat. Sign $>$ 0**
Small	8.37	191.38	53.70	0	17
Medium	3.29	78.16	76.65	0	34
Large	0.80	0.93	96.45	0	40
Overall	4.15	119.29	75.62	0	91

Panel B - Last 15-Minutes Trading

Stock Size	**Mean R**	**Std. Dev. R**	**Mean % Days R $>$ 0**	**Stat. Sign $<$ 0**	**Stat. Sign $>$ 0**
Small	2.15	79.10	24.57	1	5
Medium	7.58	117.72	58.54	0	19
Large	0.62	1.27	79.95	0	40
Overall	3.45	81.93	54.44	1	64

Table A-7: Analysis of how Volatility Impacts High Frequency Trading. This table shows the results from two different regressions used to test the impact volatility has on HFT activity. Panel A shows the results for the daily stock regression: $HFT_{i,t} = \alpha + \mathbf{1}_{QEA,i,t} * \beta_1 + \text{VIX}_t * \beta_2 + \epsilon_{i,t}$, where $HFT_{i,t}$ takes on different definitions: In column (1) it is the percent of shares in stock i in which HFTs were involved, in column (2) it is the percent of shares in stock i in which HFTs were involved and were demanding liquidity, in column (3) it is the percent of shares in stock i where HFTs were involved and were supplying liquidity. The Quarterly Earnings Announcement variable, $\mathbf{1}_{QEA}$, is a dummy variable that equals one for stock i if the observation is on the day of or the day after stock i reports its quarterly earnings, and zero otherwise. VIX is the daily S&P 500 Chicago Board of Exchange Volatility Index. The OLS regression uses data over the entire HFT dataset time horizon. Stock fixed effects are implemented and standard errors are clustered by stock. Panel B shows the results for a similar regression performed at the aggregate day level, but includes a dummy variable for the week of the Lehman Brothers failure, $\mathbf{1}_{LF}$ in place of $\mathbf{1}_{QEA}$: $HFT_t = \alpha + \mathbf{1}_{LF,t} * \beta_1 + \text{VIX}_t * \beta_2 + \epsilon_{i,t}$. $\mathbf{1}_{LF}$ equals one for observations during September 15, 2008 - September 19, 2008 and zero otherwise. I run the Lehman regression for all trading days in 2008.

Panel A - HFT - Exogenous Volatility, Quarterly Earnings

	(1) **HFT - ALL**	**(2)** **HFT - Demand**	**(3)** **HFT - Supply**
Quarterly EA Dummy ($\mathbf{1}_{QEA}$)	-0.00484	-0.0106***	0.00788***
	(0.00251)	(0.00249)	(0.00153)
VIX	0.000581***	0.00116***	-0.000453***
	(0.000153)	(0.000146)	(0.000111)
Constant	0.731***	0.371***	0.545***
	(0.00490)	(0.00468)	(0.00354)
Observations	61014	61014	61014
Adjusted R^2	0.740	0.597	0.792

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel B - HFT - Exogenous Volatility, Lehman Failure

	(1) **HFT - ALL**	**(2)** **HFT - Demand**	**(3)** **HFT - Supply**
Lehman Week Dummy ($\mathbf{1}_{LF}$)	0.012	-0.0025	0.022*
	(0.008)	(0.011)	(0.010)
VIX	0.00137***	0.00188***	0.000587***
	(0.000071)	(0.000097)	(0.000088)
Constant	0.634***	0.354***	0.386***
	(0.0025)	(0.0035)	(0.0032)
Observations	252	252	252
Adjusted R^2	0.596	0.595	0.157

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-8: Analysis of High Frequency Traders Role in the Price Discovery Process: Variance Decomposition and Information Share. Panel A reports the percent of the variance of the efficient price correlated with HFT and non-HFT trades. The remainder is in the Return column (unreported), the price discovery from publicly available information. The column labeled HFT is the percent of the variance of the efficient price correlated with HFTr-initiated trades, the column labeled non-HFT is the percent correlated with non-HFTr-initiated trades. The HFT and non-HFT values are averages of the HFT and non-HFT calculated for each stock on each day. Higher values imply that type of trader's trades add more to price discovery. The table reports the average percent of days when HFTs' contribution to price discovery is greater than non-HFTs'. The table also reports the number of stocks for which the difference between HFT and non-HFT is statistically significant. Panel B reports the Hasbrouck (1995) information share for HFT and non-HFT. The information share attributes different types of traders' quotes to the price discovery process. I report the HFTs' information share in column HFT and the non-HFTs' in column non-HFT. The higher the information share the great that trader's quotes attribution to price discovery. In each panel I group the results into three categories based on stock market capitalization, and also report the overall results. I perform this analysis for 02/22/2010 - 02/26/2010. I calculate statistical significance incorporating Newey-West standard errors to correct for the time-series correlation in observations.

Panel A: Variance Decomposition

Stock Size	HFT	Std. Dev. HFT	non-HFT	Std. Dev. HFT	Mean % Days HFT > non-HFT	Stat. Sign HFT < non-HFT	Stat. Sign HFT > non-HFT
Small	0.005	0.015	0.011	0.047	28.00	10	2
Medium	0.029	0.030	0.044	0.120	57.50	6	11
Large	0.137	0.107	0.075	0.081	88.72	0	28
Overall	0.056	0.086	0.043	0.092	57.82	16	41

Panel B: Information Share

Stock Size	HFT	Std. Dev. HFT	non-HFT	Std. Dev. HFT	Mean % Days HFT > non-HFT	Stat. Sign HFT < non-HFT	Stat. Sign HFT > non-HFT
Small	0.527	0.330	0.473	0.330	50.00	5	9
Medium	0.632	0.312	0.368	0.312	64.74	2	15
Large	0.585	0.300	0.415	0.300	57.44	6	14
Overall	0.581	0.316	0.419	0.316	57.39	13	38

Table A-9: Day-Level Determinants of High Frequency Traders' Time at Inside Quotes This table shows the result of an OLS regression with observations at the stock and day level. I implement the regression: $\mathrm{H}_{i,t} = \alpha + \mathrm{MC}_i * \beta_1 + \mathrm{MB}_t * \beta_2 + \mathrm{VIX}_i * \beta_3 + \sigma_{i,t} * \beta_4 + \mathrm{SP}_{i,t} * \beta_5 + \mathrm{DEP}_{i,t} * \beta_6 + \mathrm{TS}_{i,t} * \beta_7 + \mathrm{NV}_{i,t} * \beta_8 + \mathrm{AC}_{i,t} * \beta_9 + \epsilon_{i,t}$, where $\mathrm{H}_{i,t}$ is the fraction of calendar time a HFTr was providing the best bid or offer in stock i on day t, MC is the log market capitalization as of December 31, 2009, MB is the market to book ratio as of December 31, 2009, which is Winsorized at the 99th percentile, VIX is the S&P 500 Chicago Board of Exchange Volatility Index (scaled by 10^{-3}), σ is the ten-second realized volatility summed up over the day (scaled by 10^{-5}), SP is the average time-weighted dollar spread between the bid and offer (scaled by 10^{-1}), DEP is the average time-weighted depth available at the inside bid and ask in dollars (scaled by 10^{-3}), TS is the average dollar-volume size of a non-HFTr-only trade (trades where non-HFTs both supplied liquidity and demanded it) (scaled by 10^{-5}), NV is the dollar-volume of non-HFTr-only transactions, normalized by market capitalization (and scaled by 10^{-6}), and AC is the absolute value of a one-period autoregressive process (AR(1)) analyzed at ten-second intervals (scaled by 10^{-2}). Standard errors are clustered by stock. Columns 1 and 2 show the standardized beta coefficients and columns 3 and 4 display the regular coefficients. Columns 2 and 4 include only exogenous variables. The regression analysis uses all the HFT dataset quote data, which spans 02/22/2010 - 02/26/2010.

	Standardized Beta Coefficients		**OLS Coefficients**	
	(1)	**(2)**	**(3)**	**(4)**
Market Cap. (MC)	0.608***	0.760***	0.05764***	0.07250***
			(0.006065)	(0.005136)
Market / Book (MB)	-0.072	-0.130*	-0.00704	-0.01269*
			(0.005220)	(0.006285)
VIX	-0.012	0.011	-3.55201	3.37655
			(4.656054)	(4.206044)
Volatility (σ)	0.025		158.45870	
			(243.251073)	
Average Spread (SP)	-0.045		-0.00018	
			(0.000117)	
Average Depth (DEP)	-0.163***		-3.10e-08***	
			(6.07e-09)	
Non HFT Trade Size (TS)	-0.072		-0.20808	
			(0.195249)	
Non HFT $-Volume (NV)	-0.155***		-26.86858***	
			(5.398572)	
Autocorrelation (AC)	-0.249***		-38.69878***	
			(8.593787)	
Constant			0.30230*	-0.03573
			(0.116133)	(0.091196)
Observations	597	600	597	600
Adjusted R^2	0.621	0.532	0.621	0.532

Standardized beta coefficients in columns (1) and (2) ; Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A-10: High Frequency Trading's Impact on Volatility: Comparing Real and Alternative Price Paths. This table reports the intraday realized volatility from the actual price path and alterative hypothetical price paths that remove certain HFTr activities. To calculate the volatility I sum the the one-minute realized volatilities for each stock on each day and compare its actual value with what it would be if certain HFT activity had not occurred. I calculate three alternative price paths and their volatility levels. In Panel A I remove both HFTs' liquidity supply and demand activity. In Panel B I remove just HFTs' supply of liquidity. In Panel C I remove only HFTs' liquidity demand activity. To remove HFTs' liquidity demand activity I take out all HFTr-initiated trades. I leave all non-HFTr activity the same and assume that they traded at the same price as in the actual price path. To remove HFTs' supply of liquidity I use the order book snapshots to determine the additional price impact a trade would have if there were no HFTs supplying liquidity on the order book. I assume that even though trades where HFTs previously were providing liquidity will have occurred at different prices, that the price path returns to its actual level thereafter. If HFTs increase volatility then by "trimming" the price path I should see volatility decrease by removing their trades. If they are reducing volatility or not impacting it I should see volatility increase or remain unchanged. R-V refers to the real realized volatility, A-V refers to the alternative price path's realized volatility. I group the results into three categories based on stock market capitalization, and also report the overall results. I perform this analysis for 02/22/2010 - 02/26/2010. I calculate statistical significance incorporating Newey-West standard errors to correct for the time-series correlation in observations.

Panel A - Impact on Volatility from Removing HFT Supply and Demand

Stock Size	Mean R-V	Std. Dev. R-V	Mean A-V	Std. Dev. A-V	Mean % Days A-V > R-V	Stat. Sign R-V > A-V	Stat. Sign R-V < A-V
Small	0.138	0.090	0.191	0.135	98.50	0	33
Medium	0.139	0.083	0.165	0.098	94.50	0	28
Large	0.167	0.048	0.178	0.051	90.00	0	22
Overall	0.148	0.077	0.178	0.101	94.33	0	83

Panel B - Impact on Volatility from Removing HFT Supply

Stock Size	Mean R-V	Std. Dev. R-V	Mean A-V	Std. Dev. A-V	Mean % Days A-V > R-V	Stat. Sign R-V > A-V	Stat. Sign R-V < A-V
Small	0.138	0.090	0.192	0.135	98.00	0	39
Medium	0.139	0.083	0.169	0.097	94.50	0	30
Large	0.167	0.048	0.184	0.053	98.50	0	29
Overall	0.148	0.077	0.182	0.101	97.00	0	98

Panel C - Impact on Volatility from Removing HFT Demand

Stock Size	Mean R-V	Std. Dev. R-V	Mean A-V	Std. Dev. A-V	Mean % Days A-V > R-V	Stat. Sign R-V > A-V	Stat. Sign R-V < A-V
Small	0.138	0.090	0.138	0.090	37.00	0	0
Medium	0.139	0.083	0.139	0.085	42.50	0	0
Large	0.167	0.048	0.167	0.049	52.50	0	0
Overall	0.148	0.077	0.148	0.078	44.00	0	0

Appendix B: Background on High Frequency Trading

HFT is a recent phenomenon. One of its earliest references in main stream media was a New York Times article published on July 23, 2009 (Duhigg, July 23, 2009). Not until March 2010 did it have an entry on Wikipedia. Even the prominent firms that engage in HFT are young. For example, Tradebot, a large HFT firm that frequently makes up over 5% of all trading activity and was one of the earliest HFT firms, has only existed since 1999. Whereas only recently an average trade on the NYSE took ten-seconds to execute, (Hendershott and Moulton, 2010), now some firms' entire trading strategy is to buy and sell stocks multiple times within a second.

The rise of HFT is a result of two important changes that have increased the ability and desirability of trading fast and frequently. First, the decimalization of quotes, the change from having bid and offer prices being quoted in eighths to having them quoted in pennies, has allowed for more minute price variation. The smaller price increments makes trading during short horizons less risky as a tick in the wrong direction now can cause a penny per-stock loss whereas previously it would cost an eight of a dollar. Second, there have been technological advances in the ability to quickly analyze information and to rapidly transport data between locations. From these changes HFTs evolved. The enhancements in speed allow them to more quickly examine data and respond to market changes, while the decimalization of quotes allows them to profit from penny-sized price fluctuations.

HFT is a subset of algorithmic trading (AT). AT is defined as "the use of computer algorithms to automatically make trading decisions, submit orders, and manage those orders after submission" (Hendershott and Riordan, 2009). AT and HFT are similar in that they both use automatic computer generated decision making technology. However, they differ in that ATs may have holding periods that are minutes, days, weeks, or longer, whereas HFTs hold their position for a very short time and try to close the trading day in a neutral position. Thus, HFT is a subset of AT, but not all AT is HFT.

A new nomenclature has developed with the rise of HFT. Terms such as pinging, flashing, latency, co-location, and quote stuffing are regularly associated with HFT. These terms refer to different strategies and technical activities. Pinging is a type of market activity a HFT firm may use to seek out hidden liquidity and to try and better understand the full supply and demand for a particular stock beyond what is displayed on the order book. HFTs will issue immediate-or-cancel orders to try and detect additional liquidity in dark pools, ECNs, or hidden exchange orders. If there is hidden liquidity on one of these venues the HFTr's order will execute, otherwise it will be canceled.

Another concept in HFT is flash trading. Flash trading was a short lived and controversial option provided to liquidity demanders. Three exchanges, Nasdaq, BATS, and DirectEdge offered the program. A flash quote was an order type whereby when an exchange that was not offering the national best bid or offer (NBBO) received a market order it "flashed" the order to market participants who had registered to receive such flashes, giving them the option to take the other side of the trade before routing the order to the exchange with the NBBO.

Latency, to HFTs, refers to to the processing times required in different steps of the investment process. There are three main stages in which latency can be considered: the time between data generation and the time that data can be transferred to a firm, the time between the firm receiving the data and the firm finishing processing the data, and the time between the firm entering an order and the exchange receiving the order. There are natural limitations to how fast certain processes can occur. If a firm located in Chicago places an order to NYSE with its servers stationed in New York, even traveling at the speed of light there will be at least a four millisecond time delay between the order being sent by the firm and it being received by the exchange. One reason to minimize latency is to take advantage of latency arbitrage, a strategy whereby a firm profits from data flow inefficiencies between markets and/or exchanges. A less benign type of latency arbitrage is Quote stuffing. Quote stuffing refers to the alleged use of excessive message traffic from one firm for the specific purpose of creating latency arbitrage opportunities. The opportunities would arise from two potential areas. First, the high volume of orders would not have to be processed by the quote-generating firm but would have to be processed by other HFT firms, giving the quote-generating firm an advantage. Alternatively, if the volume is high enough it may cause the data generating process from the exchange receiving the high quote volume to lag other exchanges, creating a latency arbitrage.

A key way to reduce latency is to use co-location. Co-location refers to the practice of market participants renting space from a computer server center located next to an exchange to minimize the time it takes for a market message to arrive at the exchange. So, for example, if the previously mentioned Chicago firm had its HFT program running on a co-located server, the time between the program entering an order and the exchange receiving it would fall to microseconds.