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

September 20, 2010

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

Abstract

This paper examines the impact of high frequency traders (HFTs) on the U.S. equity market. I analyze a unique data set to study the strategies utilized by HFTs, their profitability, and their relationship with characteristics of the overall market, including liquidity, price discovery, and volatility. The 26 high frequency trading (HFT) firms in my dataset participate in 74% of all trades and make up a larger percent of large market capitalization firms. I find the following key results: (1) HFTs tend to follow a price reversal strategy driven by order imbalances, (2) HFTs make approximately $3 billion annually, (3) HFTs do not seem to systematically front run non-HFTs, (4) HFTs rely on a less diverse set of strategies than do non-HFTs, (5) HFTs trading level changes 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, but only around one-fourth of the book depth as do non-HFTs, and (8) HFTs do not seem to increase volatility and may in fact reduce it.

1 Introduction

"The evolution of financial markets has raised innumerable policy issues relating to market structure and stability."[1]

1.1 Motivation

Financial markets continuously evolve. Whenever a change in the market composition occurs, it is important to study the impact of the new development. In the 1980's the pertinent issues were program trading (Harris et al., 1994) and the expansion of option markets (Skinner, 1989). In the 1990's it was allowing the public to place limit orders (Barclay et al., 1999). In the early 2000's it was algorithmic trading (Hendershott et al., 2008), the decimalization of prices (Chung et al., 2004), and the introduction of electronic communication networks (Huang, 2002). Today it is high frequency trading (HFT; and I use HFTs to refer to high frequency traders). In this paper HFT is defined as a type of investment strategy whereby profits are attempted to be made by rapidly buying and selling stocks, with a typical holding period in terms of seconds or milliseconds. HFT has changed the composition of the market and has brought concerns with it. The fact that HFT is a new breed of trading with no trade-by-trade human interaction that can execute dozens of transactions faster than a blink of an eye is disconcerting and makes it important to understand the impact it is having on the market. HFT now makes up a large portion of the U.S. equity market activity, yet the academic analysis of its role in the financial markets is limited. This paper aims to start filling the gap.

Widespread interest exists in understanding the impact of HFT on market quality: HFTs argue they improve liquidity, enhance price discovery, and reduce volatility, while others express concern that HFT may exacerbate volatility, consume liquidity, and profit at the expense of more traditional investors. In the press HFT has received an increasing amount of attention with most of it emphasizing concerns with the practice. For example, on May 6, 2010 the Dow Jones Industrial Average dropped over 1,000 points in intraday trading in what has come to be known as the "flash crash". Afterward, some claimed HFTs drove down the market (Krudy, June 10, 2010). Others suggested a temporary withdrawal of HFTs from

[1](O'Hara, 1995) Pg. 2.

the market as exacerbating the fall (Lee, August 10, 2010).[2]

Congress and regulators have begun to take notice and vocalize concern with HFT. The Securities and Exchange Commission (SEC) issued a Concept Release regarding the topic on January 14, 2010 requesting feedback on how HFTs operate and what benefits and costs they bring with them (SEC, January 14, 2010). The Dodd Frank Wall Street Reform and Consumer Protection Act calls for an in depth study on HFT (Section 967(2)(D)). The Commodity Futures Trading Commission (CFTC) has created a technology advisory committee to address the development of high frequency trading. Talk of regulation on HFT has already begun. Given the lack of empirical foundation for such regulation, the framework for regulation is best summarized by Senator Ted Kaufman, ”Whenever you have a lot of money, a lot of change, and no regulation, bad things happen” (Kardos and Patterson, January 18, 2010). There has been a proposal (House Resolution 1068) to impose a per-trade tax of .25%. Some have suggested implementing fees when the number of canceled orders by a market participant exceeds a certain level, or limit the number of canceled orders. While others have recommended requiring quotes to have a minimum life before they can be canceled or revised. Before discussing regulation to restrict HFT it is useful to better understand what HFTs are doing and whether HFT is harming or benefiting markets.

In this paper I examine the empirical consequences of HFT on market functionality. I utilize a unique dataset that distinguishes HFT from non-HFT quotes and trades. This paper provides an analysis of HFTs behavior and their impact on financial markets. Such an analysis is necessary since to ensure properly functioning financial markets the SEC, CFTC, Congress and exchanges must set appropriate rules for traders. These rules should be based on the actual behavior and implications of market participants. It is equally important that investors understand whether or not new market developments, like the rise of HFT, benefit or harm them.

HFT is a recent phenomenon. It was brought to the general public's attention on July 23, 2009 in a New York Times article (Duhigg, July 23, 2009). Not until March 2010 did Wikipedia have an entry for HFT. Tradebot, a large player in the field who frequently makes up over 5% of all trading activity and was one of the earliest HFTs, has only been around since 1999. Whereas only recently an average trade on the NYSE took ten seconds to execute, (Hendershott and Moulton, 2007), now some firms' entire trading

[2]To date, the cause of the flash crash has not been determined, but the SEC says it will produce a report by the end of September. A preliminary report on May 6 was put out jointly by the SEC and CFTC on May 18 and provides a list of issues that may have contributed to the crash.

strategy is to buy and sell stocks multiple times within a mere second. The acceleration in speed has arisen for two main reasons: First, the change from stock prices trading in eighths to decimalization has allowed for more minute price variation. This smaller price variation makes trading with short horizons less risky as price movements are in pennies not eighths of a dollar. Second, there have been technological advances in the ability and speed to analyze information and to transport data between locations. As a result, a new type of trader has evolved to take advantage of these advances: the high frequency trader. Because the trading process is the basis by which information and risk become embedded into stock prices it is important to understand how HFT is being utilized and its place in the price formation process.

A type of trading that is similar to HFT, but fundamentally different is 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 AT may have holding periods that are minutes, days, weeks, or longer, whereas HFT by definition hold their position for a very short horizon and try to close the trading day in a neutral position. Thus, HFT must be a type of AT, but AT need not be HFT.

This paper studies HFT from a variety of viewpoints and hopes to answer two fundamental questions. First, what are the activities of HFTs? Specifically, what drives HFTs decision to buy or sell? How profitable is HFT? Are HFTs systematically front running non-HFTs? How diverse are HFTs strategies? What do HFTs do in volatile markets? Second, how does HFT impact market quality? Using research design techniques that try to overcome data limitations I ask whether HFT contributes to the price discovery process, affects liquidity, and generates or dampens volatility. Although this paper aims to address many issues raised regarding HFTs there are certain topics it does not attempt to address, these include but are not limited to flash quotes, latency arbitrage, quote stuffing, and HFTs order book dynamics.

I answer these questions by posing the following null hypotheses: (1) HFTs trade in a random fashion, (2) HFT is not profitable, (3) HFTs engage in systematic front running, (4) HFTs rely on similar strategies, (5) HFTs flee in volatile times, (6) HFTs do not add to the price discovery process, (7) HFTs do not provide liquidity, and (8) HFTs increase volatility.

To test whether HFTs trade in a random fashion I perform an ordered logit on HFTs decision to sell, not trade, or buy and find that lagged returns drives HFTs decisions. I further analyze the non-randomness of

HFTs behavior by looking specifically at HFTs buy and sell decisions based on whether they are supplying liquidity or demanding liquidity. The results suggest HFTs tend to perform a price reversal strategy. Finally I include order imbalance to the regression and find that it is lagged order imbalances more so than other types of returns that drive HFT behavior. The results are consistent with a price reversal strategy except for in the Buy-Demand column.

To test the profitability of HFTs I sum up the purchases and sales of HFT over the trading day, and at the end of each day I net out the outstanding shares held by HFT at the average price for that day. I find HFTs generate around $3 billion in gross annual trading profits.

To test whether HFTs engage in systematic front running I compare the probability of seeing different trading patterns if trading were random and compare it to the actual probability of seeing such a pattern. I observe that the probability of patterns consistent with front running do not appear more often than if trading were random. This is consistent with HFTs not systematically engaging in front running.

To test whether HFTs rely on similar trading strategies I compare the frequency of different types of trade exchanges if HFTs and non-HFTs had a similar number of strategies to the actual frequency of observing different trading types and I find evidence that supports HFTs relying on a less diverse set of strategies than non-HFTs.

To test whether HFTs flee in volatile times I take two approaches: I analyze their activity as day-level volatility increases and during varying degrees of 15-minute period price changes. I find that HFTs do reduce their liquidity-providing trades and increase their liquidity-taking trades during more volatile times in the day-level analysis, by about 10% and 5% respectively. the higher frequency 15-minute data fluctuates substantially but no clear pattern emerges. An extension of this question is whether HFTs decrease their trading as a result of volatility. I look at two types of shocks to volatility to try and capture an exogenous change in volatility to study the relationship: Days surrounding firms' quarterly earnings announcements and the week of the Lehman Brothers failure. Both approaches show HFTs tend to increase their trading during times of exogenous volatility.

To test whether HFTs do not add to the price discovery process I implement three Hasbrouck measures. I find the Price Impact, Aggregate Information Variance Decomposition, and Information Share approach all support HFTs having an important role in the price discovery process.

To test whether HFTs do not provide liquidity I examine what percent of the time HFTs provide better inside quotes than non-HFTs. In addition I consider what the price impact would be of trades of different sizes if HFTs suddenly stopped making limit orders. Both suggest HFTs play an important role in the provision of liquidity. A comparison of the book depth provided by HFT and by non-HFT reveal that HFTs tend to provide less depth to the market than may be expected given their participation level.

Finally, to test whether HFTs increase volatility I first analyze how an exogenous removal of varying amounts HFT during the short sale ban relates to volatility. Second I consider what volatility would have been had HFTs not participated in the market in varying capacities, with the assumption that other traders do not change their behavior. The results suggest HFTs do not impact volatility or may even decrease it.

The rest of the paper is as follows: Section 2 describes the related literature. Section 3 discusses the data. Section 4 provides descriptive statistics. Sections 5 analyzes hypotheses one to five that relate to HFTs market behavior, Section 6 analyzes hypotheses six to eight that relate to HFTs impact on market quality. Section 7 concludes.

2 Literature Review

HFT has received little attention to date in the academic literature. This is because until recently the concept of HFT did not exist. In addition, data to conduct research in this area has not been available. I am aware of at least two academic papers addressing HFT directly. Kearns, Kulesza, and Nevmyvaka (2010) show that the maximum amount of profitability that HFT can make based on TAQ data under the assumption that HFT enter every transaction that is profitable. The findings suggest that an upper bound on the profits HFT can earn per year is $21.3 billion.

Cvitanic and Kirilenko 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 transaction prices' distribution will have thinner tails and more mass near the mean, and as human increase their order submissions liquidity increases proportional.

HFT touches on a variety of related fields of research, the most relevant being algorithmic trading (AT). In principle AT is similar to HFT except that the holding period can vary. It is also similar to HFT in that data to study the phenomena are difficult to obtain. Recently though a growing literature on AT has developed.

Hendershott and Riordan (2009) use data from the firms listed in the Deutsche Boerse DAX index.

They find that AT supply 50% of the liquidity in that market. They find that AT increase the efficiency of the price process and that AT contribute more to price discovery than do human traders. Also, they find a positive relationship between AT providing the best quotes for stocks and the size of the spread. Regarding volatility, the study finds little evidence of any relationship between it and AT.

Hendershott, Jones, and Menkveld (2008) utilize a dataset of NYSE electronic message traffic, and use this as a proxy for algorithmic liquidity supply. The time period of their data surrounds the start of autoquoting on NYSE for different stocks and so they use this event as an exogenous instrument for AT.[3] The study finds that AT increases liquidity and lowers bid-ask spreads.

Chaboud, Hjalmarsson, Vega, and Chiquoine (2009) look at AT in the foreign exchange market. Like Hendershott and Riordan (2009), they find no evidence of there being a causal relationship between AT and price volatility of exchange rates. Their results suggest human order flow is responsible for a larger portion of the return variance.

Gsell (2008) takes a simple algorithmic trading strategy and simulates the impact it would have on markets. He finds that the low latency of algorithmic traders reduces market volatility, but that the large volume of trades increases the impact on market prices.[4]

Together these papers suggest that algorithmic trading as a whole improves market liquidity and does not impact, or may even decrease, price volatility. This paper fits in to this literature by focusing on a sub-sample of AT, only those with the shortest horizon, and studying its trading behavior and its impact on market quality. None of the above mentioned papers work with directly identified AT data and so I am unable to determine what fraction of AT is HFT.

While Cvitanic and Kirilenko build a theoretical framework that directly addresses HFT, other work has been conducted to understand what the impact on market quality will be of having investors with 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 fundamentals. The result is a decrease in the informational

[3]"Autoquote" is a technology put in place in 2003 by the NYSE to assist specialists in their role of displaying the best bid and offer. It was implemented under NYSE Rule 60(e) and provides an automatic electronic update, as opposed to manual update by a specialist, of customers' best bid and offer limit order.

[4]"Latency" refers to the speed at which a market participate can decide on making a market message and the time the exchange receives the message. To reduce latency "co-location" is typically used whereby market participants will rent space in a computer server center next to an exchange so as to minimize the time a market message takes to arrive at the exchange.

quality of asset prices.

Vives (1995) obtains the result 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 thus suggests that HFT may be benefit or may harm the informational quality of asset prices.

3 Data

3.1 Standard Data

The data in this paper comes from a variety of sources. It uses CRSP data when considering daily data not included in the Nasdaq dataset. Compustat data is used to incorporate firm characteristics in the analysis. TAQ data is used when intraday data for firms outside of the Nasdaq dataset are used. CBOE Index data is used to incorporate the CBOE S&P 500 Volatility Index (VIX) in certain instances.

3.2 Nasdaq High Frequency Data

The unique data set used in this study has data on trades and quotes on a group of 120 stocks. The Trade data consists of all trades that occur on the Nasdaq exchange during regular trading hours, excluding trades that occurred at the opening, closing, and during intraday crosses.[5] The Trade data used in this study includes those from all of 2008, 2009 and from February 22, 2010 to February 26, 2010. The trades include a millisecond timestamp at which the trade occurred and an indicator of what type of trader (HFTs or not) is providing or taking liquidity. By providing (or supplying) liquidity I mean that for a given trade the market participant had a limit order outstanding that was hit by a marketable order (or new limit order taking the opposite side of the transaction and that crossed prices). The liquidity taker (or demander) is the market participant who entered the marketable order. [6] The Quote data is from February 22, 2010 to

[5]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.

[6]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 quotes is a technology that Nasdaq, BATS, and DirectEdge had implemented to facilitate trading on their exchanges. Nasdaq ran the program for a few months between April, 2009 to July 2009. A market participant who was going to enter a marketable order had the option to flash his quote. So, for instance, if person A puts in a marketable buy order on Nasdaq and selects for the order to "flash" if not fillable on Nasdaq, and it turns out Nasdaq does not have the national best offer, then before Regulation NMS requires Nasdaq to send the order the exchange with the best offer price, the SEC approved the following

February 26, 2010. It includes the best bid and ask that is being offered by HFTs and by non-HFTs at all times throughout the day. The Book data is from the first full week of the first month of each quarter in 2008 and 2009, September 15 - 19, 2008, and February 22 - 26, 2010. It provides the 10 best price levels on each side of the market that are available on the Nasdaq book. Along with the standard variables for limit order data, the data show whether the liquidity was provided by HFTs or non-HFTs, and whether the liquidity was displayed or hidden.

The Nasdaq dataset consists of 26 traders that have been identified by Nasdaq as engaging primarily in high frequency trading. This was determined based on known information regarding the different firms' trading styles and also on the firms' website descriptions. The characteristics of firms that have been identified as being HFTs are the following: They engage in proprietary trading; that is, the firms do not have customers but instead trade their own capital. The HFT firms use sophisticated trading tools such as high-powered analytics and computing co-location services located near exchanges to reduce latency. The HFT firms engage in sponsored access providers whereby they have access to the co-location services and can obtain large-volume discounts. The HFT firms tend to switch between long and short net positions several times throughout the day, whereas non-HFT firms rarely switch from long to short net positions on any given day. 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 do non-HFT firms.

Firms that others may define as HFTs are not labeled as HFT firms here if they satisfy one of the following: brokerage firms who provide direct market access and other powerful trading tools to its customers; proprietary trading firms that are a desk of a larger, integrated firm, like a large Wall Street bank with multiple trading desks; an independent firm that is engaged in HFT activities, but who routes its trades through a Market Participant ID (MPID) of a non-HFT type firm;[7] firms that engage in HFT activities but are small.

The data is for a sample of 120 Nasdaq stocks whose ticker symbols are listed in table 1. These sample

events to happen; person B, likely a HFTr, would be shown the marketable 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 provider 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 person 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 marketable 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.

[7]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).

stocks were selected by Terrence Hendershott and Ryan Riordan. The stocks consist of a varying degree of market capitalization, market-to-book ratios, industries, and listing venues.

[Table 1 about here.]

4 Descriptive Statistics

Before entering the analysis section of the paper, as HFT data has not been identified before, I first provide the basic descriptive statistics of interest. I look at liquidity and trading statistics of the HFT sample and compare them to all stocks in the TAQ database. I then compare the firm characteristics of the HFT sample to Nasdaq and NYSE listed firms with market capitalizations greater than $10 million in the Compustat database. Finally, I provide summary statistics on the percent of the market trades in which HFT is involved, considering all types of trades, supplying liquidity trades, and demanding liquidity trades, as defined in the previous section.

4.1 Sample Characteristics

Panel A in table 2 describes the 120 stocks in the HFT database compared to the Compustat database. The table looks at the market capitalization, market-to-book ratio, industry, and listing exchange summary statistics and provides the t-statistic for the differences in means. The Compustat firms consist of all firms in the Compustat database with data available, that have a market capitalization greater than $10 million in 2009, and where listed on either Nasdaq or NYSE, which ammounts to 5,050 firms. The data for both the Compustat and the HFT firms are for fiscal year end on December 31, 2009. If a firm's year-end is on a different date, the fiscal year-end that is most recent, but prior to December 31, 2009, is used.

Whereas the average Compustat firm has a market capitalization of $3.5 billion, the average HFT database firm is $17.6 billion and the difference is statistically significant. The HFT database includes very small firms with a market capitalization of only $80 million, to the very large with a market capitalization of $175.9 billion. The average Market-to-Book ratio for the HFT database is 2.66 and is 14.18 for the Compustat database. This difference is not statistically significant.

Based on industry, the HFT database matches the Compustat database among many dimensions. Four of the ten industries though do vary by a statistically significant amount. The HFT database overweights Manufacturing and Health Care, and underweights Energy and Other. The industries are determined based on the Fama-French 10 industry designation from SIC identifiers. Finally, half the HFT database firms are listed on the NYSE and the other half on the Nasdaq exchange. This is not statistically different than the Compustat database.[8] The HFT database provides a robust variety of industries, market capitalization, and market-to-book values.

Panel B in table 2 describes the market characteristics of the 120 stocks in the HFT sample database and compares them to the full TAQ database which includes 7,537 firms. Each firm has 5 observations. These statistics are taken for the five trading days from February 22 to February 26, 2010. The statistics considered include half spreads, stock price, bid size, offer size, daily volume traded, number of trades, and size of a trade. The average half spread in the HFT database is \$.07, while in the TAQ database it is \$.17, this difference is statistically significant. The HFT database average bid size is 2380 shares and the average offer size is 2420 shares. These values are larger in the HFT dataset but are not statistically significantly different from the TAQ database. The average HFT dataset number of trades is 3,090 and is statistically significantly more than the 910 trades in the TAQ dataset. Finally, the average trade size in the HFT dataset is 208 shares while in TAQ it is 340 shares but the difference is not statistically significant. In conclusion the HFT database has smaller spreads and more trades than the average TAQ database but otherwise is statistically similar.

[Table 2 about here.]

4.2 HFT Trading Activity

Table 3 looks at the prevalence of HFT in the stock market. It captures this in a variety of ways. Panel A and B look at HFT activity at a day level, ignoring firm-by-firm variations. Panel C and D look at HFT activity at a firm-day level. That is, whereas Panel A and B each have 509 (252*2+5) observations, Panel C and D have 61080 ((252*2+5)*120) observations. Panel A (C) measures HFT activity based on the percent of dollar-volume involving HFTs for each day (day-firm). Panel B (D) measures HFT activity based on the percent of trades involving HFTs for each day (day-firm). Within each Panel are three rows. The row HFT - All shows the fraction of activity where HFTs are either demanding liquidity, supplying liquidity, or

[8] To clarify, the unique dataset utilized in this study comes from the Nasdaq exchange, 50% of the stocks in the sample are listed on the 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), or it is required to route the order to the exchange where the NBBO is being generated.

both. The row HFT - Demand shows the fraction of activity where HFTs are demanding liquidity. The row HFT - Supply shows the fraction of activity where HFTs are supplying liquidity.The summary statistics include the mean, standard deviation, minimum, 5th percentile, 25th percentile, median, 75th percentile, 95th percentile, and maximum fraction of activity in which HFTs are involved.

The results in Panel A show that HFTs are involved in 68.5% of all dollar-volume traded in the sample. Their level of daily involvement varies from 60.4% to 75.9%. They demand liquidity in 42.7% of all dollar-volume traded and supply it in 41.1%. Panel B shows that HFTs participate in 73.8% of all trades and that this varies at the day level from 65.1% to 81.9%. They demand liquidity in 43.6% of all trades and supply it in 48.7%.

These statistics are an aggregate for 26 HFT firms. Some of those firms mostly provide liquidity while others mostly take it. Although I cannot observe it directly in my data, talking with market participants, even the registered market makers will take liquidity. This is contrary to most of the theoretical literature on market makers assuming they are passive providers of liquidity. Yet, empirical papers such as Chae and Wang (2003) and Van der Wel (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.

In addition to understanding the trading behavior of HFTs at the overall day level, it is informative to understand how HFTs activities vary across firms. Panel C and D in table 3 shows the variation in HFT market makeup in different stocks on different days.

The results in Panel C show that HFTs are involved in 50.8% of the average day-firm dollar-volume traded in the sample. Their level of daily involvement varies from .22% to 100%. They demand liquidity in 50.9% of all dollar-volume traded and supply it in 51%. Panel D shows that HFTs participate in 50.8% of trades for the average firm-day and that this varies at the firm-day level from 3% to 100%. They demand liquidity in 50.9% of trades for the average firm-day and supply it in 51%. The statistics in panel C and D compared to panel A and B show that HFTs day-to-day trading level varies much less than does their firm-day trading. Although these statistics do not pick up the variation of a HFT over time for a specific firm they do show that HFT varies significantly across firms.

[Table 3 about here.]

4.2.1 HFT Trading Activity Time Series

A concern surrounding the May 6 "flash crash" was that the regular market participants, such as HFTs, stopped trading. Although the database I have does not include the May 6, 2010 data, it does span 2008 and 2009, which were volatile times in U.S. equity markets. As table 3 panel A shows, HFT daily percent of dollar-volume activity ranges from 60.4% to 75.9%. To see whether HFT percent of market trades varies significantly from day to day, and especially around time periods when the U.S. market experienced large losses, I graph the fraction of trading activity in which HFT was involved for each trading day. The results are shown in figure 1. There are three graphs. The first is a time series of 2008 and 2009 of the fraction of trades in which HFTs were involved. The second graph looks at the fraction of shares in which HFTs participated. The final graph looks at the fraction of dollar-volume activity in which HFTs were part of the transaction. In each graph there are three lines. The line labeled "All HFT" represents the fraction of exchanges in which HFTs were involved in either as a liquidty provider or as a liquidity taker; the line labeled "HFT Liquidity Supplied" represents the fraction of transactions in which HFTs were providing liquidity; the line "HFT Liquidity Demanded" represents the fraction of trades in which HFTs were demanding liquidity. All three graphs in the three measures tend to fluctuate +/- 5% on a day-to-day basis. Especially of note, there is no abnormally large drop, or increase, in HFT participation in the sample data as a whole occurring in September of 2009, when the U.S. equity markets were especially volatile. The large drops and increases occur in the following places: April 11, 2008 HFT dollar-volume liquidity provided jumped from 45% to 50% and the next day fell back to 43%. November 28, 2008 HFT dollar-volume all activity fell from 71% to 66% and the following day rose to 72%. Not displayed, but the VIX, the standard measure of market-wide volatility, is strongly positively correlated with the different measures of HFT market participation. The dollar-volume correlation with VIX is: All 0.71 , Supply 0.35, Demand 0.72. Note that at the time of this version the VIX data was only available through September 22, 2009 and so the correlation coefficients do not include the last three months of the HFT database data.

[Figure 1 about here.]

4.3 HFT Quote Activity

Table 4 provides summary statistics on HFTs quote activity. As stated in the data description section the data for quote-by-quote changes with HFT identification is only available from February 22, 2010 -

February 26, 2010 for the 120 firms. The quote data only contains the inside bid and ask for HFT and non-HFT quotes and the available sizes for each. The measures in each panel separate the quote activity into three categories based on firm size, Small Medium, and Large whereby there are 40 firms in each category. Panel A reports the percent of quote changes that were made by HFTs per firm-day. It is important to note that while this is a proxy of quote revisions and cancelations, it is not a pure measure of such activity as a quote change will also occur when a trade is executed and removes a limit order from the inside quote. 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). These results show the mean, standard deviation, minimum, 5th percentile, 25th percentile, median, 75th percentile, 95th percentile, and maximum fraction of activity in which HFTs are involved.

[Table 4 about here.]

5 HFT Market Behavior

The motivation of this paper is to begin to address what are HFTs doing and what is the result on how the market functions. In this section I examine the latter, the behavior of HFTs. I do so by addressing five questions: What drives HFTs decision to buy or sell? How profitable is HFT? Are HFTs systematically front running non-HFTs? How diverse are HFTs strategies? What do HFTs do in volatile markets? I find that HFTs trade in firms with larger market capitalizations, with lower market-to-book ratios, lower spreads, less depth and lower non-HFT volume. Their particular buy or sell decisions depend heavily on past returns and particularly on past order imbalances. I find HFTs tend to have a less diverse set of strategies than do non-HFTs. I estimate that HFTs earn gross annual profits of approximately $3 billion. I do not find evidence suggesting HFTs systematically front run non-HFTs. I find that HFTs tend to decrease their supply of liquidity moderately and increase their demand for liquidity as volatility increases at the day level.

5.1 HFTs Trading Hypothesis

$\mathbf{H_O}$: *HFTs trade in a random fashion.*

$\mathbf{H_A}$: *HFTs follow an identifiable strategy.*

The question of what drives HFTs can be partitioned into two analysis. The first is in what stocks and on what days do HFTs choose to trade more heavily. The second is what determines the specific high frequency decision of whether to buy or sell at a given moment. I look at the broader analysis in the next section and thereafter consider the high frequency determinants.

The summary statistics show HFTs activities varies across firms and time. I find that HFTs trade in firms with larger market capitalizations, with lower market-to-book ratios, lower spreads, less depth and lower non-HFT volume. HFTs do not readily disclose their trading strategy. What is known regarding HFTs is that they tend to buy and sell in very short time periods. HFTs must be basing their decision to buy and sell from short term signals such as stock price movements, spreads, or volume. I find their particular buy or sell decisions depend heavily on past returns and particularly on past order imbalances.

5.1.1 HFT Market Activity Determinants

To understand what determinants drive HFT trading I perform an OLS regression with the dependent variable $H_{i,t}$ being the percent of share volume (essentially equivalent to dollar-volume as I do a firm level analysis) in which HFTs were involved for company i on day t. I run the following regression:

$$\begin{aligned} H_{i,t} &= \alpha + MC_i * \beta_1 + MB_i * \beta_2 + NT_{i,t} * \beta_3 + \\ & NV_{i,t} * \beta_4 + Dep_{i,t} * \beta_5 + Vol_{i,t} * \beta_6 + AC_{i,t} * \beta_7 + \epsilon_{i,t}, \end{aligned}$$

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, NT is the number of non-HFT trades (trades were non-HFTs both supplied liquidity and demanded it) that occurred, scaled by market capitalization, NV is the dollar-volume of non-HFT transactions, scaled by market capitalization, Dep is the average depth of the bid and of the ask, equally weighted, Vol is the ten second realized volatility summed up over the day, AC is the absolute value of the Durbin-Watson score minus two from a regression of returns over the current and previous ten second period.

Table 5 reports the standardized regression coefficients in columns (1) - (6). That is, instead of running the typical OLS regression on the regressors, the variables, both dependent and independent, are demeaned, and are divided by their respective standard deviations so as to standardize all variables. The coefficients reported can be understood as signaling that when there is a one standard deviation change in an independent variable, the coefficient is the expected change in standard deviations that will occur in the dependent variable. This makes the regressors underlying scale of units irrelevant to interpreting the coefficients. Thus, the larger the coefficient, the more important its role in impacting the dependent variable. Columns (7) - (12) display the regular coefficients. The dependent variables for the column are: for (1), (4), (7), and (10) the percent of shares HFTs are involved in; for (2),(5),(8),(11) the percent of shares in which HFTs demand liquidity; for (3), (6), (9), and (12) the percent of shares in which HFTs supply liquidity. Columns (4) - (6) and (10) - (12) include only clearly exogenous variables.

The results in all the columns show that market capitalization is very important and has a positive relationship with HFT market activity. The market-to-book ratio is slightly statistically significant in the All and Supply columns, but with a very small negative coefficient, suggesting HFT tends to occur slightly more often in value firms. Also statistically significant and with moderate economic significance is the dollar-volume of non-HFT, which is interpreted as HFTs preferring to trade when there is less volume, all else being equal. The spread and depth variables are statistically significant and both have moderate economic significance. HFTs tend to trade in stocks with lower depth and lower bid-ask spreads, all else being equal. Volatility is only statistically significant in the Demand column with a negative coefficient. Autocorrelation is not statistically significant in any column. The number of non-HFT trades is statistically significant in the Supply (negative coefficient) and the Demand (positive coefficient) columns.

Many of the explanatory variables may be endogenously determined and as a result the OLS estimator may be biased. Thus, I run the same regression but only include as explanatory variables market capitalization and market-to-book. The results are also in table 5, with columns (4) - (6) reporting the standardized beta coefficients and column (10) - (12) reporting the standard OLS coefficients. This restricted regression changes the magnitude of both explanatory variables to a degree and results in the Market / Book being statistically significant with relatively larger magnitude coefficients.

[Table 5 about here.]

5.1.2 All Inclusive Search

I begin the analysis by performing an all-inclusive ordered logit regression into the potentially important factors; thereafter I analyze the promising strategies in more detail. There are three decisions a high frequency trader (HFTr, singular of HFTs) makes at any given moment: Does it buy, does it sell, or does it do nothing. This decision making process occurs continuously. I model this setting by using a three level ordered logit. The ordered logit 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 my case the dependent variable is a ten-second buy/do nothing/sell decision and not a transaction-by-transaction price process.

I divide the time frames in to ten second intervals throughout the trading day. [9] For each ten second interval I utilize a variety of independent variables. The regression I run is as follows:

$$
\begin{aligned}
HFT_{i,t} = \alpha \quad & +\beta_{1-11} \times Retlag_{i,0-10} && +\beta_{12-22} \times Depthbidlag_{i,0-10} \\
& +\beta_{23-33} \times Depthasklag_{i,0-10} && +\beta_{34-44} \times Spreadlag_{i,0-10} \\
& +\beta_{45-55} \times Tradeslag_{i,0-10} && \beta_{56-66} \times Dollarvlag_{i,0-10}.
\end{aligned}
$$

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. For example, $retlag_0$ represents the return for the particular stock during time period t. And, the return for time period t is defined as $retlag_{i,0} = (price_{i,t} - price_{i,t-1})/price_{i,t-1}$. Thus the betas represent row vectors of 1x11 and the explanatory variables column vectors of 11x1. $Depthbid$ is the average time weighted best bid depth for stock i in that time period. $Depthask$ is the average time weighted best offer depth for stock i in that time period. $Spread$ is the average time weighted spread for company i in that time period, where spread is the best ask price minus the best bid price. $Trades$ is the number of distinct trades that occurred for company i in that time period. $DollarV$ is the dollar-volume of shares exchanged in transactions for company i in that time period. The dependent variable, HFT, is -1, 0 or 1. It takes the value -1 if during that ten second period HFTs were, on net, selling shares for stock i, it is zero if the HFTs performed no transaction or its buys and sell exactly canceled, and it is 1 if, on net, HFTs were buying shares for stock i. Firm fixed effects are implemented.

From this ordered logit model one may expect to see a variety of potential patterns. A handful of

[9] I also tried other time intervals, such as 250 milliseconds, one second and 100 second periods. The results from these alternative suggestions are similar in significance to the results presented in that where a ten second period shows significance, so does the one second interval for ten lagged period's worth, and similarly where ten lagged ten second intervals show significance, so does the one lagged one hundred second interval. The ten second intervals has been adopted after attempting a variety of alterations but finding this one the best for keeping the results parsimonious and still being able to uncover important results.

different strategies have been suggested in which HFTs engage. For instance, momentum trading, price reversal trading, trading in high volume markets, or trading in high spread markets. It could be they base their trading decisions on the spread and so the $Spread$ variables would have considerable power in explaining when HFTs buy or sell. If HFTs are in general momentum traders, then I would expect to see them buy after prices rise, and to sell after prices fall. If HFTs are price reversal traders, then I would expect to observe them buying when prices fall and to sell when prices are rising. Table 6 shows the results.

[Table 6 about here.]

5.1.3 Focus on Lagged Returns

The results reported in table 6 are the marginal effects at the mean for the ordered logit. From the ordered logit regression's summarized results in table 6, there is sporadic significance in all but one place, the lagged values of company i's stock returns. There is a strong relationship with higher past returns and the likelihood the HFTs will be selling (and with low past returns and the likelihood the HFTs will be buying). There is some statistical significance in other locations, however no where is it consistent like that of the return coefficients. One interpretation of this result is that past or contemporaneous spread size, depth, and volume are not primary factors in HFTs trading decisions. Alternatively, it may be these variables are primary factors but due to endogeneity are not captured by the logit regression. Of the strategies discussed above, these results are consistent with a price reversal trading strategy. To further understand this potential price reversal strategy I focus on analyzing the lag returns influence on HFTs' trading behavior. I examine HFTs buy and sell logits separately, focusing on the lagged returns surrounding HFTs' buying or selling stocks and decomposing the differences in demanding versus supplying liquidity activity.

To better understand HFTs trading strategy I run logit regressions on different dependent variables. I consider a total of six different regressions: HFTs selling, HFTs selling when supplying liquidity, HFTs selling when demanding liquidity, HFTs buying, HFTs buying when supplying liquidity, and HFTs buying when demanding liquidity. The results found in table 8 are the marginal effects at the mean and the logit incorporates firm fixed effects. The first column is the results for HFT Sell, all types. The results show the strong relationship between past returns and HFTs decision to sell. prior to HFTs executing a sale of a stock, the stock tend to rise, with statistically significance up to 90 seconds prior to the trade, barring time period 8. This finding suggests HFTs in general engage in a price reversal strategy.

The next column has as the dependent variable a one if HFTs were on net supplying liquidity to the market and selling during a given ten second interval and a zero otherwise. The results are similar to the previous results, except that the magnitude and statistical significance is not as strong. There appears to be more scattered significance of past returns.

The third column in table 8 has as the dependent variable a one if HFTs were, on net, taking liquidity from the market and selling during the ten second interval and a zero otherwise. There is still strong statistical significance from the ten past return periods, barring the ninth one. The signs are the same as before, which is consistent with a price reversal strategy.

[Table 7 about here.]

The Buy regressions are also shown in table 8. The fourth column is the result for HFT Buy, all types. The results show the strong relationship between past returns and HFTs decision to buy. Prior to HFTs executing a purchase of a stock, the stock tend to fall, with statistically significance up to 100 seconds prior to the trade.

The fifth column has as the dependent variable a one if HFTs were on net supplying liquidity to the market and buying during a given ten second interval and a zero otherwise. The results in the lag returns are similar to the previous results, except that the magnitude of the coefficients are smaller.

The last column in table 8 has as the dependent variable a one if HFTs were, on net, taking liquidity from the market and buying during the ten second interval and a zero otherwise. There is still some statistical significance from the ten past return periods, but only in time periods 3 - 7 and 9.

The results in table 8 show that HFT are engaged in a price reversal strategy. This is true whether they are supplying liquidity or demanding it.

5.1.4 Order Imbalance and Lagged Returns

The above results show that on average HFTs engage in a price reversal strategy. The lagged returns are important determinants of HFTs investment strategy. The price formation process is well studied but not well understood. In the literature much emphasis has been placed on the volatility-volume relationship where when volume is higher returns tend be larger, but more recent literature has emphasized the importance of order imbalance (Chan and Fong, 2000). It may be the case that returns which come from order imbalance are more important to HFT than returns generally, especially as Chordia and Subrahmanyam (2004) shows that order imbalance can be used to generate a profitable trading strategy.

To analyze the hypothesis that HFTs rely on the reason for the lagged return I rerun the regression above but include two dependent variables, past returns and past order imbalances:

$$HFT_{i,t} = \alpha + Ret_{i,1-10} \times \beta_{1-10} + OIB_{i,1-10} \times \beta_{11-20} + \epsilon_{i,t}, \tag{1}$$

where $Ret_{i,1-10}$ is the return for firm i in period s, where s is from 1 to 10 periods prior to the time t, $OIB_{i,1-10}$ is a measure of order imbalance for firm i in period s. $OIB_{i,s}$ is defined as:

$$OIB_{i,s} = \begin{cases} \frac{BuyInitiatedShares_{i,s}}{Shares_{i,s}} & \text{if } Shares_{i,s} \neq 0 \\ .5 & \text{if } Shares_{i,s} = 0 \end{cases} \tag{2}$$

where $BuyInitiatedShares_{i,s}$ is the number of shares for firm i in period s where the liquidity demander was buying and where $Shares_{i,s}$ is the number of shares for firm i in period s that were exchanged. I divide OIB by 100 to reduce the number of leading zeros in the coefficient results.

$HFT_{i,s}$ takes on one of six definitions: it is equal to 1 if (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.

The results are in favor of HFTs decision to buy and sell being primarily driven by order-imbalance and not other types of returns. The fist three columns in the table look at HFTs decision to sell. In all three OIB has a strong positive coefficient in almost all ten lagged periods. Thus, following an order imbalance increase, when there are more marketable buy orders than marketable sell orders, HFTs tend to sell. The statistical and economic significance of the Ret variable drops for Sell-All and Sell-Supply. However, for Sell-Demand regression, these lagged returns are still positive and statistically significant, suggesting that HFTs demand to sell for both order imbalances and also for other returns.

The Buy decision regressions are not as uniform. Looking at the Ret coefficients, the Buy-All and

Buy-Demand still are negative and maintain their statistical and economic significance, but the Buy-Supply regression has no return coefficients of importance. The order imbalance coefficients for Buy-All and Buy-Supply are negative and strongly statistically significant, suggesting HFTs tend to supply liquidity after there have been more sell orders than buy orders. The Buy-Demand column has lagged order imbalances 1-3 and 5 being statistically significant. Unlike the other columns though, here HFTs move with the order-imbalance: when there had been more buy than sell orders, all else being equal, HFTs tend to demand to buy. This is inconsistent with the strategy suggested by the other results. Comparing the Buy-Demand Ret results of the returns-only regression to this regression show that now their is a strong statistically significant impact from past returns 1-3. Interestingly, this is where the OIB coefficients are statistically significant.

[Table 8 about here.]

5.2 HFTs Profitability Hypothesis

$\mathbf{H_O}$: *HFTs are not profitable.*

$\mathbf{H_A}$: *HFTs are profitable.*

HFTs engage in a price reversal strategy and they make up a large portion of the market. Given their trading amount a question of interest is how profitable is their behavior. HFTs have been portrayed as making tens of billions of dollars from other investors. Due to the limitations of the data, I can only provide an estimate of the profitability of HFTs. The HFT labeled trades come from many firms, but I cannot distinguish which HFT firm is buying and selling at a given time. Also, recall that the dataset only contains Nasdaq trades. Therefore, there will be many other trades that occur that the dataset does not include. Nasdaq makes up 20% - 30% of all trades and so two out of every three trades are unobserved. I circumvent these limitations by making estimates using the market behavior results from above to arrive at an overall annual profitability of HFT.

I consider all HFT actions to come from one trader. I take all HFT buys and sells at their respective prices and calculate how much money was spent on purchases and received from sales. HFTs regularly switch between being net long and net short throughout the day, but at the end of the day they tend to hold very few shares. With these considerations in mind, I can estimate the total profitability of these 26 firms. As many HFTs do not end the day with an exact net zero position in each stock I take any excess shares and assume they were traded at the mean price of that stock for that day. Thus, the daily profitability for each stock is calculated as:

$$
\begin{aligned}
Profit &= \sum_{t=1}^{T} [\mathbf{1}_{Sell} * Price_t * Shares_t - \mathbf{1}_{Buy} * Price_t * Shares_t] + \\
&\quad \frac{1}{\sum_{t=1}^{T} Shares_t} \sum_{t=1}^{T} [Price_t * Shares_t] \,,
\end{aligned}
$$

where 1_{Sell} is a dummy indicator that equals one if HFTs sold a stock in transaction t and zero otherwise, 1_{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. Summing up the $Profit$ for each stock on a given day results in the total HFT profitability for that day.

The result of this exercise is that on average, per day, HFTs make $298,113.1 from the 120 stocks in my sample on trades that occur on Nasdaq.

The above number substantially underestimates the actual profitability of HFTs. First, the 120 stocks have a combined market capitalization of $2,110,589.3 (million), a fraction of Compustat firms' combined market capitalization is $17,156,917.3 (million). Second, Nasdaq is one of several venues where trades occur and on average makes up between 20 - 30% of trading activity. To account for these limitations in the data I carry out the following exercise: I use the non-endogenous regression coefficient estimates in the HFT percent determinant estimation found in table 5 to estimate the percent of trades involving HFT and multiply it by the fraction of shares where HFTs are not trading with each other, as a HFTr exchange with a HFTr will have a net zero profit when considering HFTs profit in the aggregate.

$$
H\hat{F}T = .6835[.0308866 + MarketCap * .0725419 - Market/Book * .0119239],
$$

where $MarketCap$ is the log of the daily shares outstanding of company i multiplied by the closing price of company i, and $Market/Book$ is the ratio of the $MarketCap$ divided by the Compustat book value based on the the most recent preceding quarterly report, winsorized at the 99th percentile. $H\hat{F}T$ is calculated for each stock. I multiply $H\hat{F}T$ 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 by HFTs found above. Thus, I arrive at the annual estimated profits of HFTs by:

$$H\hat{FTAnnualProfit} = \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[H\hat{F}T_{i,t} * DVolume_{i,t} * .000106\right] \tag{3}$$

The .000106 value represents the profit per dollar volume HFT traded with a non-HFT. It is determined by taking the total profit of HFTs from the 120 sample firms over the sample time period and divide it by the HFT - non-HFT dollar volume traded (\$151,739,574/\$2,089,346,000,000). The result of this calculation is that HFTs gross profit is approximately \$ 2.995 billion annually.[10]

There is no adjustment made for transaction costs yet. However, such costs will be relatively small, the reason being that when HFT provide liquidity they receive a rebate from the exchange, for example Nasdaq offers \$.20 per 100 shares for which traders provided liquidity, but this is only for large volume traders like HFTs. On the other hand, Nasdaq charges \$.25 per 100 shares for which trades take liquidity. As the amount of liquidity demanded is slightly less than the liquidity supplied by HFT, these two values practically cancel themselves out. A rough estimate of the cost of trading is calculated by assuming that there is an equal number of shares demanded and supplied, thus I can estimate each trade of 100 shares costs .025 (.05 per trade, but only half of trades are demanding liquidity.) If I assume the average stock price is \$30, then using the total number of implied shares traded (including HFTr-to-HFTr transactions), the annual transaction cost for HFTs is \$344,469,548.

What is important isn't the level of profitability of HFT, but what it is relative to the alternative, a market consisting of non-HFT market makers. Thus, I compare how profitable HFTs trades are per dollar traded as compared to other market makers, specifically specialists of NYSE stocks in 2000. Hasbrouck and Sofianos (1993) and Coughenour and Harris study the trading activity and profitability of the NYSE specialists. From the above results, HFT make on average 1/100th of a penny (\$.000106) per dollar traded. Calculating from the summary data reported in Coughenour and Harris, specialists before HFT but after decimalization (before decimalization reported in parenthesis) made \$.00052 (\$.000894) per dollar traded in small stocks, \$.00036 (\$.00292) per dollar traded in medium stocks, and \$.00059 (\$.0025) per

[10]This number is less than what others have estimated. An article by The Tabb Group claimed HFTs made around \$21 billion annually. However, the \$3 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 different exchanges around the world and in equities, commodities, fixed income, and foreign exchange. Even if \$200 million of that profit were from U.S. equities it is still in line with my findings that 26 firms split profits of \$3 billion as Getco is one of the largest HFT firms.

dollar traded in large stocks. From this perspective, HFTs are less than a fourth as expensive as post-decimalization market makers.

Figure 2 displays the time series of HFT profitability per day. The graph is a five day-moving average of profitability of HFT per day for the 120 firms in the dataset. Profitability varies substantially from day to day, even after smoothing out the day to day fluctuations.

[Figure 2 about here.]

This section has shown that HFTs engage in a price reversal trading strategy, that HFT tends to occur more in large stocks with relatively low volume with narrow spreads and depth. In addition, there is little change in HFT activity during extreme market conditions and HFT slightly increases with exogenous shocks to volatility. Also, HFTs are profitable, making approximately $3 billion a year, but on a dollar traded basis they are significantly less expensive than traditional market makers, and that the profitability is related to volatility. Next, I investigate the role HFT plays in the demand and supply of liquidity.

5.3 HFTs Front Running Hypothesis

$\mathbf{H_O}$: *HFTs engage in systematic front running.*

$\mathbf{H_A}$: *HFTs do not engage in systematic front running.*

A potential investing strategy of which HFTs have been claimed to be engaged in is front running. Some believe HFTs are able to detect when other market participants hope to move a large number of shares in a firm and that the HFTs enters into the same position just before the other market participant. The result of such an action by HFTs would be to drive up the cost for non-HFTs to execute the desired transaction.[11]

[11]Front running is not itself an illegal activity. It is illegan when a firm has a fiduciary obligation to its client and that firm uses the client's information to front run its orders. In my dataset, as HFTs are propriety trading firms they do not have clients and so the front running they may be conducting would likely not be illegal. Yet, it is still a concern. If the HFT are able to position themselves in a role whereby a trade between two non-HFTs was about to occur its not clear what economic benefit this is providing and may be a way for HFT from profit from one, or both, sides of the trade. Where HFT and front running may be especially problematic is if there is market manipulation 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 bellow the bid price to be executed. Market makers 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 it 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.

To see whether or not this is occurring on a systematic basis I look at the frequency of observing different marketable order sequences. From previous results the percent of marketable orders in a stock by HFTs and non-HFTs varies substantially. Consequently, when looking for the occurrence of front running I cannot simply assume the equal probability of seeing different sequences of non-HFT and HFT marketable orders. Instead, I use a ratio taking advantage of the Panel B similarities in probabilities to cancel out the fact that seeing the t-1, t pattern of HN is similar to seeing that of HN. For each firm I analyze the probability of seeing different trading patterns. The approach I take requires the assumption that the absence of systematic front running will imply that it is equally likely to see a HFTr initiated transaction after a Non-HFTr initiated transaction as it is to see a HFTr initiated transaction after a HFTr initiated transaction.

To see whether HFTs tend to trade more just before non-HFTs rather than just after (my definition of front running) I look at the ratio of the frequency of observing different patterns. I consider five different patterns: HN, HHN, HHHN, HHHHN, and HHHHHN whereby the last letter stands for the type of trader demanding to Buy at time t, and the preceding letters represent who is demanding to buy in the prior trades (H represents HFTs, N represents non-HFTs). So depending on the number of time periods considering the sequence represents times ($t-5$, $t-4$, $t-3$, $t-2$, $t-1$, t). To account for different probabilities of seeing an N or a H, each of the five different patterns are scaled by the probability of seeing the opposite pattern, that is the probability of seeing N initiated buy followed by the different number of H initiated buy orders. If front running is regularly occurring it should be the case that the probability of seeing a H before a N would be more likely than the opposite. this would show up in the table as results > 1. A result of 1 would suggest seeing the two patterns is equally likely, and a value < 1 suggests its more likely to see an N followed by an H than the opposite. Table 9 shows the results. Column (1) shows the results for $\frac{Prob(HN)}{Prob(NH)}$, column (2) shows $\frac{Prob(HHN)}{Prob(NHH)}$, column (3) shows $\frac{Prob(HHHN)}{Prob(NHHH)}$, column (4) shows $\frac{Prob(HHHHN)}{Prob(NHHHH)}$, and column (5) shows $\frac{Prob(HHHHHN)}{Prob(NHHHHH)}$. Column (1) has 16 of the 120 firms being > 1, in column (2) 21 are > 1, in column (3) 18 are > 1, in column (4) 17 are > 1, in column (5) 21 are > 1. Finally the overall result for each column is < 1. And of those that are > 1, most are either 1.01 or 1.02.

These findings suggest HFTs as a whole are not front running as their main strategy. However, I cannot conclude there is no front running. It could be that the multiple strategies HFTs use cancel out the informativeness of this approach of looking at HFT. It could also be that when one non-HFTr marketable order executes it is a signal that other non-HFTr marketable orders are coming into the market and so HFTs quickly first place their own buy 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 the ratios would equal one.

[Table 9 about here.]

5.4 HFTs Diversity Hypothesis

$\mathbf{H_O}$: *HFTs utilize a less diverse set of strategies than non-HFTs.*

$\mathbf{H_A}$: *HFTs do not utilize a less diverse set of strategies 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 those of non-HFTs I examine the frequency at which HFTs trade with each other and compare it to a benchmark model used in Chaboud et al. (2009) that produces theoretical probabilities of different types of trades (demander - supplier) under the assumption that traders' activities are random and independent. Then I can compare the actual occurrence of different trades to the predicted amount. As above, there are four types of trades, 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, N_d be the number of non-HFT liquidity demanders. The probability that a HFTr will provide liquidity is then $\alpha_s = Prob(HFT - supply) = \frac{H_s}{N_s+H_s}$, and the probability the liquidity is supplied by a non-HFTr is $1 - \alpha_s$. The probability that a HFTr will demand liquidity is $\alpha_d = Prob(HFT - demand) = \frac{H_d}{N_d+H_d}$, and the probability the liquidity is demanded by a non-HFTr is $1-\alpha_d$. The probabilities of a specific demander and supplier can be calculated: $Prob(HH) = (\alpha_d)(\alpha_s)$, $Prob(HN) = (\alpha_d)(1 - \alpha_s)$, $Prob(NH) = (1 - \alpha_d)(\alpha_s)$, $Prob(NN) = (1 - \alpha_d)(1 - \alpha_s)$.

As a result, the follow fraction holds: $\frac{Prob(NN)}{Prob(NH)} \equiv \frac{Prob(HN)}{Prob(HH)}$. Let $RN \equiv \frac{Prob(NN)}{Prob(NH)}$ be the non-HFTr demanding liquidity ratio and $RH \equiv \frac{Prob(HN)}{Prob(HH)}$ be the HFTr demanding liquidity ratio. When non-HFTs are greater than HFTs then $Prob(NN) > Prob(NH)$ and $Prob(HN) > Prob(HH)$. However, regardless of the number of non-HFTs or HFTs, the ratio of ratios, $R \equiv \frac{RH}{RN}$ will equal one as non-HFT will take liquidity from other non-HFT in the same proportion as HFTs take liquidity from other HFTs. Therefore, if $R = 1$, it must be that HFTs and non-HFTs trade with each other as much as expected when their trading strategies are equally correlated. if $R > 1$ then it is the case that HFTs trade with each other less than expected, or that HFT trade with non-HFT more than expected.

The proxy used for R in the data is $\hat{RN} = \frac{Vol(NN)}{Vol(NH)}$ and $\hat{RH} = \frac{Vol(HN)}{Vol(HH)}$. Table 10 shows the results. The column R shows the results for each stock of the average $\frac{RH}{RN}$ per day. The column $Std.Dev.$ is the standard deviation of the R ratio for that stock over time. The column $\%DaysR < 1$ is the fraction of days in which R < 1. Starting with $\%DaysR < 1$, all stocks have the ratio $R < 1$ less than 50% of the time. 42 are in the 0% - 10%s, 18 are in the 10%s, 11 are in the 20%s, 37 are in the 30%s, and 8 are in the 40%s. Overall 20% of days have $R < 1$. Of the 120 firms, none have an average R less than 1. 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 engage in a less diverse variety of strategies than non-HFTs, whereby the diverse strategies result in one HFTr deciding to buy and another HFTr deciding to sell simultaneously. [Note: statistical analysis available soon.]

[Table 10 about here.]

5.5 HFTs Fleeing Hypothesis

$\mathbf{H_O}$: *HFTs flee in volatile markets.*

$\mathbf{H_A}$: *HFTs do not flee in volatile markets.*

5.5.1 HFT - Volatility Across Market Conditions

What has been shown so far has dealt mostly with means, but a major concern is that HFTs may be around during normal times, but during extreme market conditions, for example when volatility increases, HFTs reduce their trading activity. An alternative concern is that HFT induces heightened levels of volatility, which is specifically addressed in Section 6.3. Here I study how HFTs behave in different levels of volatility and am not attempting to show a causal effect. I am interested in understanding the relationship between HFT and volatility as volatility levels change. Therefore, I could perform the following regression with firm fixed effects using quantile regressions to examine how HFTs behavior varies with volatility:

$$HFT_{i,t} = \alpha + Volatility_{i,t} * \beta_1 + \epsilon_{i,t},$$

where $HFT_{i,t}$ is the percent of shares for firm i on day t involving HFT and $Volatility_{i,t}$ is the 15 minute realized return volatility for firm i on day t. However, even by using quantile regressions the picture would be incomplete. Instead, I build a graphical representation of the HFT-Volatility relationship presented in figure 3. Of interest is how HFTs either pull back or increase their trading activity as volatility changes. The X-axis consists of 100 bins grouped together based on the $VolatilityChange$ value:

$$VolatilityChange_{i,t} = \frac{Volatility_{i,t} - \sum_{t=1}^{T} \frac{1}{T} Volatility_{i,t}}{\sum_{t=1}^{T} \times \frac{1}{T} Volatility_{i,t}} \frac{1}{\sqrt{\sum_{t=1}^{T} \frac{1}{T}[Volatility_{i,t} - \sum_{t=1}^{T} \frac{1}{T} Volatility_{i,t}]^2}}$$

The $VolatilityChange$ variable is the scaled deviation from the mean, where it is scaled by the standard deviation of a firm's daily volatility. Without the scaling I would essentially be plotting HFTs fluctuation within firms, with more volatile stocks, which tend to be smaller firms, being further to the right on the X-axis.

Using the firm-days in each bin I calculate the abnormal HFTr activity, $HFTChange$:

$$HFTChange_j = \sum_{Volatility_{i,t} \epsilon j} \frac{1}{N_j} [\frac{HFT_{i,t} - \sum_{t=1}^{T} \frac{1}{T} HFT_{i,t}}{\frac{1}{T} HFT_{i,t}}] \quad (4)$$

where HFT is the fraction of shares in which HFTs are involved and j is the bin in which firm i at time t is grouped with based on its $VolatilityChange$. N is the number of observations in group j By considering firm by firm $HFTChange$ I am in effect controlling for firm specific effects. The outcome is similar to what the equation above would have produced but with the benefit of seeing changes in HFTs participation at a granular level. The results are in graph 3. There are three graphs, HFT % All, which looks at all HFTs activity, HFT Supply % Change, which only considers HFTs supplying liquidity activity, and HFT Demand % Change, which only considers HFTs demanding liquidity activity. In each graph there are four lines. The level-by-level HFT % change, a 9-period centered moving average of the HFT % change, and the upper and lower 95 % confidence intervals.

The first graph, HFT % All, HFTs activity appears to be almost flat across volatility levels. Even on the most volatile days HFT overall activity does not seem to increase or decrease substantially. However, when volatility is low HFTs activity is less than average. The second figure looks at HFT supply of liquidity.

HFT provide about 10% more liquidity than usual on very low volatility days. The level of HFT liquidity slowly declines as volatility picks up, at the highest volatility the HFT liquidity is about 10% less than on an average day. The third figure considers HFT taking liquidity. On the least volatile days HFT take about 7% less liquidity than normal, and on the most volatile days they take around 5% more liquidity than normal. These results show that HFT activity does change with volatility, but not precipitously. In particular on the most volatile days, HFT do not pull out of the market. On these days, there does seem to be a 5-10% transfer of HFT activity from supplying liquidity to demanding liquidity. This could be consistent with the market maker story. On volatile days there will be more situations where HFTs inventory becomes unbalanced and they have to demand liquidity to unload positions, whereas on low volatility days more of this rebalancing can be done through supplying liquidity. The change in liquidity taking is also consistent with a statistical arbitrage story. When prices are not moving around there is little deviations across markets and stocks and so HFT arbitragers aren't making marketable trades. However, when prices are volatile there are more arbitrage opportunities for which HFTs will step in and demand liquidity.

[Figure 3 about here.]

Table 3 looks at day level volatility. But higher frequencies are also of interest given that prices can fluctuate dramatically during one portion of the day but not the rest and so this short term deviation may not be picked up at a day level analysis. Therefore, I look at the data in 15 minute intervals. In addition, instead of looking at volatility I examine returns during the 15 minute period. Given this, I can separate the analysis along a variety of different criteria such as positive and negative return episodes, HFTs buying and selling activity, and HFTs supplying and demanding liquidity.

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

$$PriceChange_{i,t,m} = Ret_{i,t,m} \times \frac{1}{\sqrt{\sum_{i,t,m}^{t=T,m=M} \frac{1}{T*M} Ret_{i,t,m}^2}},$$

where $Ret_{i,t,m}$ is the return during that time period for firm i on day t during period m.

$$HFTChange_j = \sum_{PriceChange_{i,t,m} \epsilon j} \frac{1}{N_j} \left[\frac{HFT_{i,t,m} - \sum_{m=1}^{M*T} \frac{1}{M} HFT_{i,t,m}}{\frac{1}{M} HFT_{i,t,m}}\right] \tag{5}$$

HFT takes on one of five definitions: All Activity, Buy-Demand, Buy-Supply, Sell-Demand, or Sell-Supply where each defines HFT 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 0 for that period, and remove those periods where less than 30 trades occurred. The results are in tables 4 and 5. Results for price increases are in table 4 and those for price declines are in table 5. The figures show the five-period centered moving averages. Even with smoothing the results are noisy. In both the price inclines and declines, and for all cuts, there is no consistent pattern. This suggests that HFTs are not pulling back liquidity, or driving prices during large price swings.

These results suggest that during large price declines HFTs do not make unusually large sell demands, and they do not stop providing liquidity to those who are selling. Similarly with price inclines HFTs do not make unusually large buy demands, and they do not stop providing liquidity to those who are buying.

[Figure 4 about here.]

[Figure 5 about here.]

5.5.2 Volatility's Impact on HFT Trading

The above results still don't overcome the likely endogeneity between HFT and volatility. To overcome this one must find situations in which there are exogenous shocks to volatility. Exogenous shocks to volatility typically come from new information entering the public domain. Thus, a natural time to expect exogenous shocks to volatility is during quarterly firm earnings announcements. In the HFT sample dataset, days on which firms announce their quarterly earnings have higher volatility than the average non-announcement day for that stock. Thus, not only is it that the volatility is likely exogenous, coming from news and not traders' churning, it is at elevated levels. The difference is small, but statistically significant. Using OLS regression, I regress the percent of shares in which HFTs were involved on a dummy variable, $QuarterlyEADummy$, which is one for firm i if the observation is on the day of or the day after firm i reported its quarterly earnings, and zero otherwise. The dependent variable in column (1) is the percent of shares in stock i in which HFT was involved, in column (2) it is the percent of shares in stock i in which HFT was involved and was demanding liquidity, in column (3) it is the percent of shares in stock i

in which HFT was involved and was supplying liquidity. The results in table 11 Panel A show that HFT activity increases with a shock to volatility. For the quarterly earnings announcements, the increase arises from HFT supplying liquidity in a larger fraction of shares.

Another time in which there was an identifiable exogenous shock to volatility was the week of September 15 - September 19., 2008. This was the week in which Lehman Brothers collapsed, volatility spiked, and there was a high level of information uncertainty. Like the quarterly earnings announcements, the week in September when Lehman failed and a randomly chosen week in November also shows that firms in the September week have statistically significantly higher volatility. I therefore test whether there is a difference in HFT activity during the week of September 15, 2008 and the week of November 3, 2008 (this week is chosen as it is sufficiently far away to reduce the autocorrelation impact of volatility, but not too far away as for there to have been a significant change in HFTs strategies. Using OLS regression, I regress the percent of shares in which HFT were involved on a dummy variable, $LehmanWeekDummy$, which is one for all firms for observations on the dates September 15, 2008 - September 19, 2008 and zero otherwise. The dependent variable in column (1) is the percent of shares in stock i in which HFT was involved, in column (2) it is the percent of shares in stock i in which HFT was involved and was demanding liquidity, in column (3) it is the percent of shares in stock i in which HFT was involved and was supplying liquidity. The results show that HFT activity increases with a shock to volatility. For the Lehman Week increase in HFT, the increase arises from HFT supplying liquidity and demanding liquidity in a larger fraction of shares.

[Table 11 about here.]

6 Market Quality

In the previous section I addressed issues relating to HFTs activity, in this section I analyze their impact on market quality. By market quality I mean price discovery, liquidity, and volatility. I do so by addressing three questions: Do HFTs contributes to the price discovery process? What role does HFT in providing market liquidity? and Do HFTs generate or dampen volatility? Using three Hasbrouck measures, I find that HFTs play an important role in the price discovery process. I find that HFTs frequently offer better bid and offer prices than do non-HFTs. However, they provide less depth than non-HFTs. In addition, they tend to supply liquidity to trades that do not contain private information. Finally, I find weak evidence that volatility is either unaffected by HFT or that it decreases slightly.

6.1 HFT Price Discovery Hypothesis

$\mathbf{H_O}$: *HFTs do not add to the price discovery process.*

$\mathbf{H_A}$: *HFTs do add to the price discovery process.*

HFT makes up a significant portion of market activity, both on the demand side and the supply side, but that does not imply its activities increase price efficiency. In this section I utilize three of Hasbrouck's methodologies to see whether HFTs provide new information to the market. First, I utilize the impulse response function whose results can be interpreted as the amount of private information different traders bring to prices by measuring the amount of the price adjustment from the trade that is permanent. HFTs provide more private information to the market than do non-HFTs. Second, I use a variance decomposition technique that takes the results of the impulse response function and relates the different type of traders' trades to the price discovery process. The results show that HFTs are more important in the price discovery process than non-HFTs. Finally, I implement the information shares 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. HFTs provide substantially more information to the price process than do non-HFTs. The Hasbrouck methodologies utilized in this paper are similar to those found in Hendershott and Riordan (2009) and other papers.

6.1.1 Permanent Price Impact

To measure the information content of HFT and non-HFT I calculate the permanent price impact of HFTs and non-HFTs trades. Hendershott and Riordan (2009) performed a similar calculation for trader types looking at algorithmic trading, while Barclay, Hendershott, and McCormick (2003) used the technique to compare information from different markets. The HFT dataset is especially well suited for this as it is in milliseconds and thus avoids problems of multiple trades occurring in one time period, as occurs with data denoted in seconds. I estimate the model on a trade-by-trade basis using 10 lags for HFT and non-HFT trades. I estimate the model 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 HFT equation, and a non-HFT trade equation. The time index, t, is based on event time, not clock time, and so each t is an event that is a trade. q^H is defined as the signed (+1 for a buy, -1 for a sell) HFTs trades and q^N is the similarly denoted signed non-HFTs trades. r_t is defined as the quote midpoint to quote midpoint return between trade changes. The 10-lag vector auto regression (VAR) is:

$$
\begin{aligned}
r_t &= \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=0}^{10} \beta_i q^H_{t-i} + \sum_{i=0}^{10} \gamma_i q^N_{t-i} + \epsilon_{1,t}, \\
q^H_t &= \sum_{i=1}^{10} \delta_i r_{t-i} + \sum_{i=0}^{10} \rho_i q^H_{t-i} + \sum_{i=0}^{10} \zeta_i q^N_{t-i} + \epsilon_{2,t}, \\
q^N_t &= \sum_{i=1}^{10} \pi_i r_{t-i} + \sum_{i=0}^{10} \nu_i q^H_{t-i} + \sum_{i=0}^{10} \psi_i q^N_{t-i} + \epsilon_{3,t}.
\end{aligned}
$$

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

$$
\begin{bmatrix} r_t \\ q^H_t \\ q^N_t \end{bmatrix} = \begin{bmatrix} a(L) & b(L) & c(L) \\ d(L) & e(L) & f(L) \\ g(L) & h(L) & i(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}, \tag{6}
$$

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. 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. 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.

Table 12 shows the results of the HFT and non-HFT impulse response function for 10 events into the future. There are 105 firms presented as fifteen stocks do not contain enough data to calculate the VAR. Each stock is reported individually. For each stock I estimate the statistical significance of the difference of the impulse response function for the HFT and non-HFT 5 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. Also, I calculate the overall average HFT and non-HFT impulse response function, this calculation incorporates the Newey-West correction for time series and also a correction for the cross-section correlation standard errors.

Of the 105 companies represented 90 of them have the HFT impulse response function being larger than the non-HFT impulse response. None of the 15 firms where the non-HFT impulse response function is larger than HFT's are statistically significant. Of the 90 in the other direction, 26 of the differences are statistically significant. On average, HFT's impulse response function is 1.017 and Non HFT's impulse response is 0.759. The overall difference is statistically significant. This suggests that HFTs trades provide more private information than do non-HFTs trades. This is similar to the findings in Hendershott and Riordan (2009) with algorithmic trades. Thus, an innovation in HFT tends to lead to a 34% greater permanent price change than does a trade by a non-HFTr.

[Table 12 about here.]

6.1.1.1 LR - SR Price Impact The results in table 12 show that HFT has a larger price impact than does non-HFT over the ten period interval. An item of interest is whether the price impact is immediate or gradual over the ten future time periods. Similar to the methodology used in Chaboud, Hjalmarsson, Vega, and Chiquoine (2009) and Hendershott and Riordan (2009), I test whether the price process may cause an immediate overreaction to one type of trade and that over the next nine periods in the future the impact decreases. If it is the case that there is an immediate overreaction to a HFTs trade this would support the theory that HFTs increase the volatility of markets. To analyze this I report the difference between the long-run (LR; 10 event forecast horizon) and short-run (SR; immediate) impulse response functions in table 13.

Of the 105 The LR-SR impulse response is less for HFTs than for non-HFTs in 25 of the 105 firms. Of those 25 firms none are statistically significant. Of the 80 firms where the LR-SR impulse response function is greater for HFTs than non-HFTs 15 are statistically significant. Also, for each market participant column, a positive number implies that the LR impact of a trade is greater than the SR impact, and a negative number implies there is a short run overreaction and that over the next nine periods the permanent price impact falls. The results of table 13 suggest that HFTs individual innovations have more private information than non-HFTs trades and that the difference is persistent and increases beyond the immediate impact of the trade.

[Table 13 about here.]

6.1.2 Aggregate Amount of Information in HFT - Variance Decomposition

The permanent price impact section above shows that HFTr demanded 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) to decompose the variance of the efficient price into the portion of total price discovery that is correlated with HFT and non-HFT. The results indicate which trades contribute more to price discovery. The methodology decomposes the variance of the efficient price into the portion of total price discovery that is correlated with HFT and non-HFT trades.

This analysis was also in Hendershott and Riordan (2009) to determine whether algorithmic or human traders contribute more to price discovery and I follow a similar methodology. 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: $r_t = m_t + s_t$.

m_t represents the efficient price where m_t = m_{t-1} + w_t and w_t is a random walk with Ew_t = 0; s_t is the non-persistent price component. Let $\sigma^2_{\epsilon_1}$ = $E\epsilon^2_1$, $\sigma^2_{\epsilon_2}$ = $E\epsilon^2_2$, 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:

$$\sigma^2_w = (\sum_{i=0}^{10} a_i)^2\sigma^2_{\epsilon_1} + (\sum_{i=0}^{10} b_i)^2\sigma^2_{\epsilon_2} + (\sum_{i=0}^{10} c_i)^2\sigma^2_{\epsilon_3}, \tag{7}$$

where the a, b, c are as defined in the previous section as the lag coefficients found in the VMA matrix. The $(\sum_{i=0}^{10} b_i)^2\sigma^2_{\epsilon_2}$ term represents the proportion 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 from this exercise are found in table 14. I report the average contribution by HFT and by non-HFT for each company over the five days. The final column is the t-statistic for the difference between the HFT and non-HFT contribution and is adjusted for its time-series correlation with Newey-West standard errors. I also report the average overall contribution, whose t-statistic is corrected for time-series correlation and for cross-sectional correlation. The HFT column is the contribution to price discovery from HFTs, and the same interpretation is true with the non-HFT column. The contribution to the Returns component (the public information) is the public information related to price discovery, it is unreported here for lack of space, but can be easily calculated by taking the difference between 1 and the sum of the HFT and non-HFT components.

Of the 118 firms 68 of them show HFT as having a greater contribution to price discovery, and 28 of those stocks' HFT - non-HFT contribution difference is statistically significant. In the 50 stocks where the non-HFT contribution is greater than that of the HFT, the difference is statistically significant for 7 firms. On average HFT contributes 86% more to price discovery than do non-HFT.

[Table 14 about here.]

6.1.3 Information Share

This section examines the role HFTs and non-HFTs quotes play in the price discovery process, whereas the previous two sections had been analyzing the role of trades. I use the Information Shares (IS) approach introduced by Hasbrouck (1995) and that is used in, among others, Chaboud, Hjalmarsson, Vega, and Chiquoine (2009) and Hendershott and Riordan (2009). This approach 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.

The approach is as follows. I calculate the HFT and non-HFT price path. 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. From the VMA I gather the variance of the random walk and the coefficients of the VMA innovations.

The price process is calculated from the HFT and non-HFT midpoint, $MP_t^{HFT} = InsideBid_t^{HFT} + InsideAsk_t^{HFT})/2$ for HFT, and done similarly for non-HFT. Then the price process for HFT and non-HFT 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 vector of the HFT and non-HFT price process can be put into a VMA model:

$$\Delta p_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} \ldots, \tag{8}$$

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

$$\sigma_u^2 = \begin{bmatrix} \Psi_{HFT} & \Psi_{nHFT} \end{bmatrix} \begin{bmatrix} \sigma^2_{HFT} & \sigma^2_{HFT,nHFT} \\ \sigma^2_{HFT,nHFT} & \sigma^2_{nHFT} \end{bmatrix} \begin{bmatrix} \Psi_{HFT} \\ \Psi_{nHFT} \end{bmatrix}, \tag{9}$$

where Ψ represent the lag operator vector from above and the sigmas represent the $Var(\epsilon_t)$ from above.

As the quote data I have is 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 book 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.

The results are found in table 15. The information share attributable to HFTs and non-HFTs from their quote time-series process. The table shows the average information share (which sums to 1 for each stock) for each stock. The average is over the five days in the dataset. The t-statistics are based on the difference in the information share between HFT and the non-HFT and incorporates Newey West standard errors to account for time series correlation.

The results in Table 15 show which quotes contribute more to price discovery, HFT or non-HFT. The information share of a participant is measured as that participant's contribution to the total variance of the common component of the price. 73 stocks have the HFT information share being larger than the non-HFT information share. Of those 36 of the stock have HFT being statistically significantly providing more information in their quotes than non-HFT. Of the 43 companies where the non-HFT have a larger information share than HFT, twelve of the differences are statistically significant. In addition, overall HFTs contributio is .58 compared to non-HFTs .42 and the difference is statistically significant. This suggest that in quotes, like in trades, HFT are important in the price discovery process.

[Table 15 about here.]

6.2 HFTs Liquidity Hypothesis

$\mathbf{H_O}$: *HFTs do not provide liquidity.*

$\mathbf{H_A}$: *HFTs do provide liquidity.*

This section analyzes HFT and the supply of liquidity. I show that, beyond supplying liquidity in 51.4% of all trades, HFTs frequently supply the inside quotes throughout the day. I then consider the determinants that influence which stocks, and on what days, HFTs provide the inside quotes. Next, I examine the depth of liquidity provided by HFT and non-HFTs. While removing either type of trader would result in a larger price impacts of trades, removing non-HFTs has a larger impact. Finally, using the Price Impact measure from the previous section, but applying it to the case of who is supplying liquidity, I find that HFTs provide liquidity for less informative trades.

6.2.1 HFT Time at Inside Quotes

To begin analyzing HFT role in providing liquidity in the stock market I look at the amount of time HFTs supply the inside bid or offer. For each stock, on each day, I report the average number of minutes HFTs are providing the inside bid or offer. I include in the value the time during which HFTs quotes are better than non-HFTs quotes or match them. The results are shown in table 16.

Table 16 looks at, for the 120 sample stocks, whether HFTs provide the best inside quote (bid or ask). The manner in which the metric is constructed results in there being a total of 780 minutes (2*60*6.5) that a HFT could potentially be providing the inside quote. This is twice as many minutes then what actually occur during the trading day. Table 16 has three panels. Panel A is for all stocks at all times, Panel B and C divide the time at the best bid and offer based on whether spreads that day are higher then average. Panel B reports the results for days in which quotes are below their daily mean. Panel C reports the results for days in which quotes are at or above their daily mean. In each Panel the data are divided into three groups, with 40 firms each, based on firm size and the results are the different rows, Small, Medium, and Large. The total row is the unconditional results.

Panel A shows that HFT firms frequently provide the best bid and offer quotes. As the firm size increases HFTs are more competitive in their quotes, matching or beating non-HFTs quotes for a significant portion of the day.

Panel B and C divide the stocks into those that are offering lower (Panel B) spreads than average and those offering higher spreads than average (Panel C). The results between the two subsets do not differ much from one another. In fact, the Total mean is only 30 seconds more during the high spread period than the low spread period. Although this is consistent with HFTs attempting to capture liquidity supply profits as found in Foucault and Menkveld (2008) and Hendershott and Riordan (2009) make/take liquidity cycle, the difference is small.

[Table 16 about here.]

6.2.1.1 HFT Time at Inside Quotes Determinants Table 16 shows that HFTs produce the inside quotes frequently, but not as often as non-HFT. I perform an OLS regression similar to that found in table 5 to understand what determinants are related to which stocks and days HFT prduces the best quotes. Table 17 shows the results. It is very similar to table 5, with all variables being defined exactly the same as before except the dependent variable. The regression is:

$$
\begin{aligned}
L_{i,t} &= \alpha + MC_i * \beta_1 + MB_i * \beta_2 + NT_{i,t} * \beta_3 + \\
&\quad NV_{i,t} * \beta_4 + Dep_{i,t} * \beta_5 + Vol_{i,t} * \beta_6 + AC_{i,t} * \beta_7,
\end{aligned}
$$

where the variables and subscripts are defined as above, and the dependent variable, $L_{i,t}$ is the percent of the time for which HFTs provide the best inside quotes compared to all times when HFTs and non-HFTs quotes differ.

The coefficients reported, like those in table 5, are standardized beta coefficients which allows for an easy way to decide which determinants are more important. The results suggest there are several explanatory variables that matter, all except $Autocorrelation$ are statistically significant, and all except $AverageDepth$ have coefficient magnitudes greater than .16. $MarketCap.$ and $\#ofNonHFTTrades$ have positive coefficients, with $MarketCap.$ being the most important determinant of HFTs providing the best quotes. The other coefficients are negative, suggesting that HFTs prefer to provide the inside quotes for value firms, less volatility firms, firms with narrower spreads, and firms with a lower book depth.

[Table 17 about here.]

6.2.2 Book Depth from HFTs and non-HFTs

Thus far, the analysis on liquidity has been by looking at the best inside bid and ask. Another way of looking at HFT impact on liquidity is by looking at the depth of the book supplied by HFT. I analyze what difference having HFTs provide liquidity in the book provides in decreasing the price impact of a trade. That is, one can observe the book with all of the limit orders in it and then remove the liquidity provided by HFTs and see what the impact would be on the cost of executing a trade for different size trades. The results of this exercise are presented in table 18. I consider a variety of different impacts based on the number of shares hypothetically bought. The number of shares varies from 100 to 1000. Table 18 shows the price impact based on market capitalization and also for the overall sample (column All). The market capitalizations are divided so that Very Small includes firms under $ 400 million, Small are those between $400 million and $1.5 billion, Medium are those between $1.5 billion and $3 billion, and large are for firms valued at more than $3 billion. I present both the dollar impact, where a 1 represents one dollar increase in the price impact if HFT were not in the book, and a Basis impact, where a 1 represents a 1 basis percent increase if HFT were not in the book.

Panel A shows the results of removing HFT from the book. As the trade size increases, the price impact increases across firms of all sizes and for all ten trade size increases. The Small category tends to be more impacted by the withdrawal of HFT liquidity than is the Very Small category. One might expect the very small to impacted the most and their be a downward trend in impact as one moves to the large firms, but this need not be the case if HFTs did not have many orders in the book to begin with. The price impact is sizeable. For an average 1000 share trade, if HFT were not part of the book the price impact would be .19 percent higher than it actual is because of the liquidity HFTs provide.

Panel B shows the results of removing non-HFTs from the book. Across all categories the removing of non-HFTs has a much larger impact than does the removal of HFTs. This means that although HFTs supply liquidity in 41% of all dollars traded, they provide only a fraction of the depth compared to non-HFTs.

[Table 18 about here.]

A concern with this analysis is the endogeneity of limit orders (Rosu, 2009) and the information they may contain (Harris and Panchapagesan, 2005; Cao et al., 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. First, the most important part of this table is the comparison between HFT and non-HFT depth. With that being said, it is not clear whether once the market participant observed a given limit order he would be influenced to place his own limit order entry, place a marketable order, or to withhold from entering the market. Thus, the dynamics are not clear whether this increases or reduces the impact of the previous analysis. In addition, this concern should be even further dampened as market participants can always choose not to display their limit orders.

To get a better understanding of the HFT and non-HFT book depth figure 6 includes three graphs, the price impact from removing HFTs book orders, the price impact of removing non-HFTs book orders, and the ratio of the two (non-HFTs / HFTs) for all firm sizes with a 1000 share order working through the book. The X-axis has labeled the first day of five for which the data shows results. That is, The observation 01-07-08 is followed by observations on Januay 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. The difference between non-HFTs and HFTs depth is large and persistent, but appears to be decreasing in the latter part of the data sample. That is, although HFTs continue to provide less depth than non-HFTs the gap is closing. The correlation coefficient between the VIX and the non-HFTs / HFTs book ratio is -.38, so when expected volatility is high, HFTs narrow the difference in their book depth compared to non-HFTs.

[Figure 6 about here.]

6.2.3 Who Supplies Liquidity To Informed Traders

The liquidity results so far have shown that HFTs competitively provide the best bid and offer a significant portion of the day and that they provide some depth on the book. The perceived advantage of being a HFTr is the ability to quickly update one's quotes so as not to be caught providing liquidity to informed traders who are going to move prices and cause the liquidity provider to lose money.

The Hasbrouck Price Impact measure used above was a way to capture which liquidity takers were having a permanent impact on prices and this was interpreted as what type of traders had private information. Here I apply the same technique as the Price Impact measure but consider who is *supplying* liquidity to informed traders. That is, before I defined q^H and q^N based on who is demanding liquidity, now I do it for the supplier of liquidity in a trade. The q^H will be a +1 when a HFT supplier sells and -1 when a HFT supplier buys, The q^N value is similarly defined for non-HFTr supplied trades. The results can then be interpreted as determining what type of trader supplies liquidity to informed traders. The larger the result means more information is imputed into a stock price from trades for which that type of trader supplied liquidity.

The results are in table 19. 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 it is true that HFTs use their speed to avoid informative trades this would show up with the HFT column being smaller than the nHFT column. Of the 102 stocks with enough observations, this is true for 66 of them. Of these 66 firms 22 are statistically significant. Of the 48 where HFT is larger than nHFT only 4 are statistically significant. Overall though, HFT is larger than nHFT, but this difference is not statistically significant. The fact that more firms have nHFT being greater than HFT and many more of these are statistically significant than those with the other sign, this is inline with HFTs being able to more precisely pick trades with new information and avoid trades with important private information.

[Table 19 about here.]

6.3 HFT Volatility Hypothesis

$\mathbf{H_O}$: *HFTs increase market volatility.*

$\mathbf{H_A}$: *HFTs do not increase market 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. The results suggest that HFT and volatility are linked. I had found that HFTs overall activity is little changed with volatility, but that they decrease the liquidity they supply, and increase the liquidity they demand, as volatility increases. When there is an exogenous shock to volatility HFTs tend to increase their market participation. In this section I approach the question of how HFTs impact volatility. First, I use the period surrounding the short sale ban in September, 2008 as an event study and evaluate the impact on volatility of an exogenous decrease in HFT. Next, I compare the price path of stocks with and without HFT being part of the data generation process. The results are not strong but suggest that HFTs either have no impact on volatility or reduce it to a degree.

6.3.1 Exogenous Shock to HFTs

As seen in previous sections, HFT is influenced by volatility. Now I study whether HFT influences volatility. Before I used exogenous shocks to volatility to study its influence on HFTs. Here I use an exogenous shock on HFTs to study the reverse relationship. The exogenous shock I utilize is the September 19, 2008 ban on short sale trading for 799 financial firms, which was in place until October 9, 2008. Of the 120 firms in the HFT sample dataset, 13 were on the ban list. The ban did not directly stop HFTs from trading in those shares. However, after talking with HFT firms, it is clear they avoided these stocks as their strategies require them to switch freely between being long or short a stock, and I can observe in the data that HFTs activity dropped precipitously during this period (for the 13 affected stocks). Thus, the short sale was in fact a defacto ban on a portion of HFTs. In a quick-to-follow clarification, the SEC made clear that officially designated market makers were not subject to the ban and could freely short sell the 799 stocks. One reason HFT during this period does not drop further is that a portion of firms identified as HFTs are official market makers and so they did not experience the same trading limitations as their non-designated counterparts.

With the 13 effected firms I use the variation in the decline in HFT activity as different levels of treatment and study the subsequent change in volatility. As all 13 firms are in the short sale ban, there is no concern that my results are actually an implication of the ban itself. In addition, I match each of the 13 treated firms to a firm in the unaffected group based on proximity of HFT market participation in the week prior to the ban. I run the following OLS regression:

$$\Delta Vola_{i,t} = HFT\%Change_{i,t} * \beta_1 + \epsilon_{i,t},$$

where $\Delta Vola$ is the percent change in volatility for firm i between the pre- and post- ban period after differencing out the change in its comparable control firm, $\frac{Vol_{post}-Vol_{pre}}{Vol_{pre}}$. $HFT\%Change_{i,t}$ is the percent change in HFT activity pre- and post- ban after differencing out the change in its comparable control firm, $\frac{HFT_{post}-HFT_{pre}}{HFT_{pre}}$. The results are in table 20. Column (1) shows the results when looking at the one-day level activity. That is, the pre ban data are for September 18, 2008, and the post ban data are for September 19, 2008. The results in column (1) shows no relationship between an exogenous shock in HFT and volatility. Column (3) performs the same analysis but uses the week average, using the average per stock data of the five trading days prior to September 19, 2008, for the pre ban data, and the average per stock data of the five trading days after the ban for the post ban data. This approach produces a negative coefficient on $HFT\%Change$ which is interpreted as the more HFT decreased, the greater the rise in volatility. The coefficient is statistically significant only at the 10 percent level. Given the sparse number of observations, I implement a non-parametric bootstrap looping through the data 50 times (using replacement). This has no impact on the statistical significance of the Day level analysis as seen in column (2). However, using the bootstrap technique for the Week level analysis results in HFT % Change showing statistical significance at the 5% level.

[Table 20 about here.]

6.3.2 Alternative Price Path

I also take an alternative approach to studying the impact of HFT on volatility. To reduce the impact of endogeneity, I take advantage of the book data I have available in one minute increments. With this data I can estimate what the price impact would have been had there been no HFTs demanding liquidity or supplying liquidity. That is, I have the actual price series for each stock, but I can supplement that with the hypothetical price series of each stock assuming that there were no HFT in the market. However, if I remove HFT liquidity providing trades and replace it with the implied price after looking at how far through the book a marketable order would have to go to get filled, this will certainly increase volatility. Thus, instead of removing all types of HFT transactions I only remove HFT initiated trades.

There are a varying degree of ways in which this exercise can be performed. The two most plausible alternatives are: (a) remove all HFT initiated trades and generate the price path that a stock would take if the non-HFTs had made the same buy and sell decisions based on the prior non-HFTs price That is, determine the return from each non-HFTr initiated transaction in the true price path, then remove the HFTr initiated trades and recalculate the price path assuming the non-HFTr buy and selling was the same and the returns were the same. (b) leave prices untouched and simply trim out the HFTr trades, assuming that the prices would have achieved their actual levels but would simply jump around more (as there would be no HFT initiated trades). I take the latter approach. I do so as it is a more conservative technique and violates the microstructure theory to a lesser degree.

I calculate the 1-minute realized volatility,the sum over one minute increments of the absolute value of the returns over the day, with and without HFTr initiated trades using the technique described in part (b) of the previous paragraph. I do this for each for each stock on each day from February 22, 2010 - February 26, 2010. For the calculation I use the return from the trade closest to period 0, but occurring after time 0, and the trade closest to period 1 with the trade occurring on or after time 1. 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. That is, if they are increasing volatility, then they are buying at the peaks and selling at the troughs, by removing them I am leveling out the price path. If they have no impact or are decreasing volatility then removing the HFT initiated trades will either leave volatility unaffected, or will increase it as the previous HFT buy (sell) at a low (high) will be replaced in the realized volatility return by a non-HFT buy (sell) at a higher (lower) level.

Table 21 shows the results. Of the 120 firms, 72 of them have a higher volatility when HFTr initiated trades are removed. Thus, a small majority of firms experience slightly higher volatility without HFTr initiated trades. However of these 72 stocks, only one is statistically significant. Of the 48 stocks where the removal of HFTr initiated trades reduces volatility, suggestive of HFTs causing volatility, none show a statistically significant difference in volatility. The t-statistics for the individual firms use Newey-West standard errors to account for the time series correlation. the overall t-statistic also corrects for cross-sectional correlation. The overall results show that when removing HFTr initiated trades volatility increases and this difference is statistically significant. Admittedly the results are not strong in one direction or another, they lean in favor of HFT having no impact or reducing volatility.

[Table 21 about here.]

7 Conclusion

This paper examines HFT and its role in U.S. equity markets. I try to provide a better understanding of the behavior of HFTs and what their impact has been on market quality. I test eight hypotheses and find (1) HFTs tend to follow a price reversal strategy driven by order imbalances, (2) HFTs make approximately $3 billion annually, (3) HFTs do not seem to systematically front run non-HFTs, (4) HFTs rely on a less diverse set of strategies than do non-HFTs, (5) HFTs trading level changes only slightly 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, but only around one-fourth of the book depth as do non-HFTs, and (8) HFTs do not seem to increase volatility and may in fact reduce it.

References

M.J. Barclay, W.G. Christie, J.H. Harris, E. Kandel, and P.H. Schultz. Effects of market reform on the trading costs and depths of Nasdaq stocks. *The Journal of Finance*, 54(1):1–34, 1999.

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. Technical report, Working paper, State University of New York, Buffalo, 2003.

K. Chan and W.M. Fong. Trade size, order imbalance, and the volatility-volume relation* 1. *Journal of Financial Economics*, 57(2):247–273, 2000.

T. Chordia and A. Subrahmanyam. Order imbalance and individual stock returns: Theory and evidence* 1. *Journal of Financial Economics*, 72(3):485–518, 2004.

K.H. Chung, C. Chuwonganant, and D.T. McCormick. Order preferencing and market quality on NASDAQ before and after decimalization* 1. *Journal of Financial Economics*, 71(3):581–612, 2004.

J.F. Coughenour and L. Harris. Specialist profits and the minimum price increment.

J. Cvitanic and A.A. Kirilenko. High Frequency Traders and Asset Prices.

C Duhigg. Stock Traders Find Speed Pays, in Milliseconds. *New York Times*, (http://www.nytimes.com/2009/07/24/business/24trading.html), July 23, 2009.

K.Y.L. Fong and W.M. Liu. Limit order revisions. *Journal of Banking & Finance*, 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. In *Proceedings of the 16th European Conference on Information Systems (ECIS); Galway, Ireland*, 2008.

L. Harris, G. Sofianos, and J.E. Shapiro. Program trading and intraday volatility. *Review of Financial Studies*, 7(4):653–685, 1994.

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. 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 NYSEs Hybrid. *Manuscript, UC Berkeley*, 2007.

T. Hendershott and R. Riordan. Algorithmic trading and information. *SSRN*, 2009.

T. Hendershott, C.M. Jones, and A.J. Menkveld. Does Algorithmic Trading Improve Liquidity? *Journal of Finance, Forthcoming*, 2008.

R.D. Huang. The quality of ECN and Nasdaq market maker quotes. *Journal of Finance*, 57(3):1285–1319, 2002.

D. Kardos and S. Patterson. Senator Calls for More Regulation of High-Frequency Trade. *Wall Street Journal*, (http://www.nytimes.com/2009/07/24/business/24trading.html), January 18, 2010.

M. Kearns, A. Kulesza, and Y. Nevmyvaka. Empirical Limitations on High Frequency Trading Profitability. *Working Paper*, 2010.

E. Krudy. Is the May 6 ”flash crash” in US stock markets the new normal? *Reuters*, (http://www.reuters.com/article/idUSN1057063420100610), June 10, 2010.

A. Lee. Computers, high-speed trading led to flash crash: survey. *Market Watch*, (http://www.marketwatch.com/story/high-speed-trading-led-to-flash-crash-survey), August 10, 2010.

M. O'Hara. *Market microstructure theory*. Wiley, 1995.

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

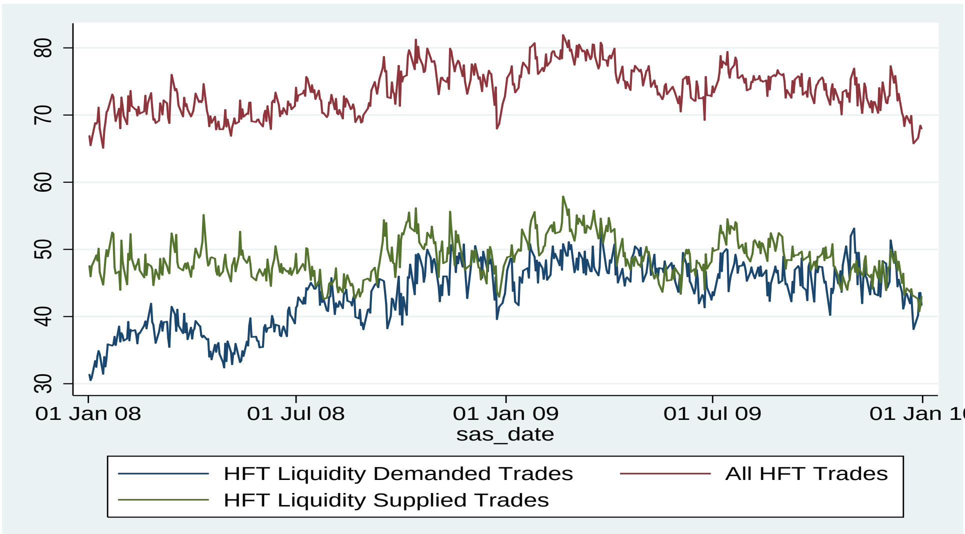
SEC. Concept Release on Equity Market Structure. (Concept Release on Equity Market Structure), January 14, 2010.

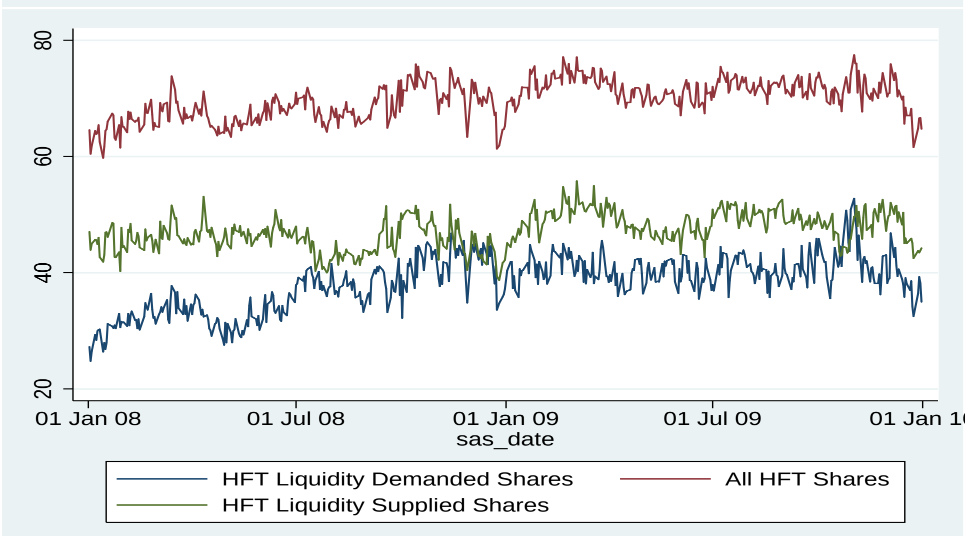
D.J. Skinner. Options markets and stock return volatility. *Journal of Financial Economics*, 23(1):61–78, 1989.

Menkveld A. Sarkar A. Van der Wel, M. Are Market Makers Liquidity Suppliers? *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 HFT Market Participation The first graph is a time series of the fraction of trades in which HFT was involved in during 2008 and 2009. The second graph looks at the fraction of shares in which HFT was involved. The final graph looks at the fraction of dollar volume in which HFT was involved. In each graph three lines appear. One line represents whether HFT was involved as either a liquidity provider or a liquidity taker; another line represents transactions in which HFT was providing liquidity; the final line represents when HFT was demanding liquidity.





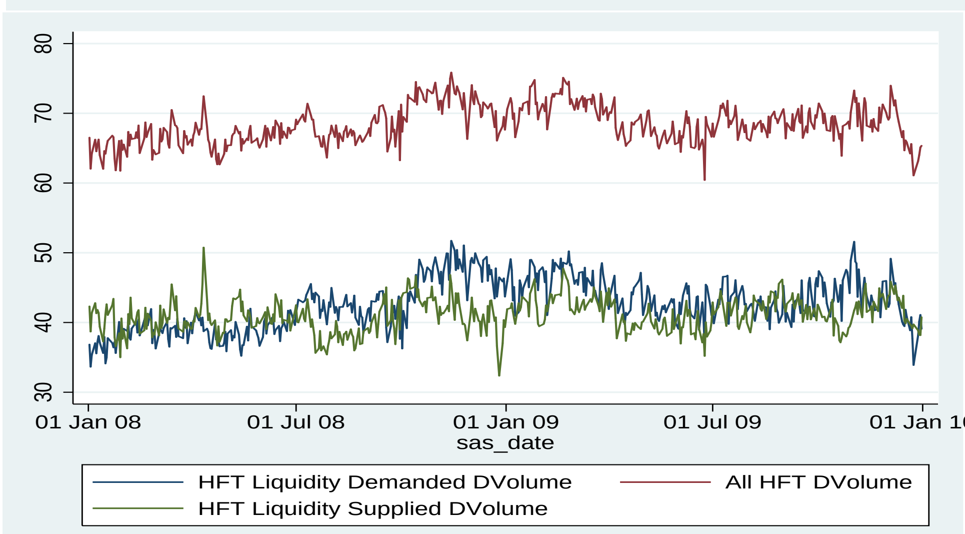


Figure 2: Time Series of HFT Profitability Per Day. The figure shows the 5-day moving average profitability for all trading days in 2008 and 2009 for trades in the HFT data set. Profitability is calculated by aggregating all HFT for a given stock on a given day and comparing the cost of shares bought and the revenue from shares sold. For any end-of-day 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.

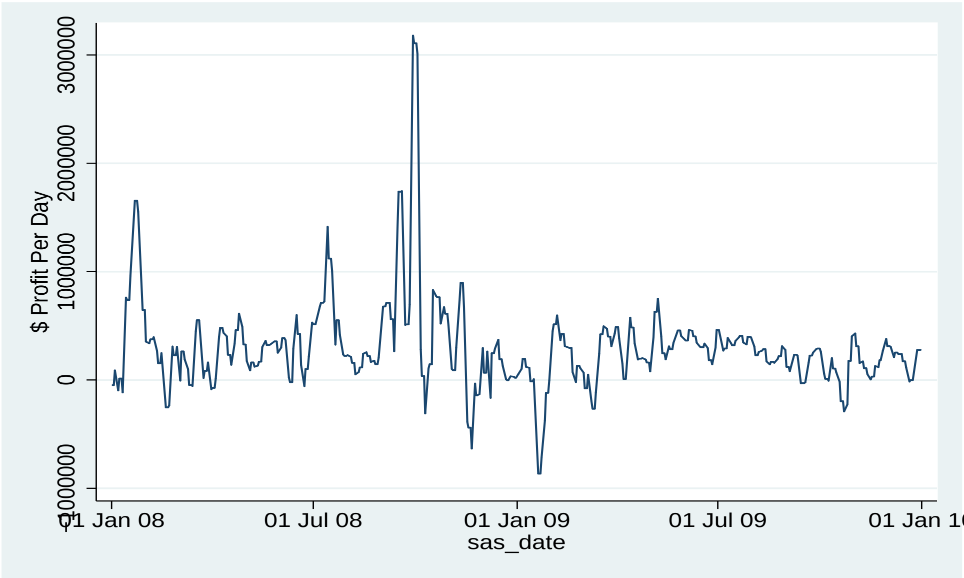


Figure 3: HFT - Day Level Volatility Relationship. This graph shows how HFT participation varies when volatility is higher or lower than its average level. On the X-axis are 100 bins based on $VolatilityChange$, $VolatilityChange_{i,t} = \frac{Volatility_{i,t} - \sum_{t=1}^{T} \frac{1}{T} Volatility_{i,t}}{\sum_{t=1}^{T} \times \frac{1}{T} Volatility_{i,t}} \frac{1}{\sqrt{\sum_{t=1}^{T} \frac{1}{T} [Volatility_{i,t} - \sum_{t=1}^{T} \frac{1}{T} Volatility_{i,t}]^2}}$. On the Y-axis is $HFTChange$ $HFTChange_j = \sum_{Volatility_{i,t} \epsilon j} \frac{1}{N_j} [\frac{HFT_{i,t} - \sum_{t=1}^{T} \frac{1}{T} HFT_{i,t}}{\frac{1}{T} HFT_{i,t}}]$. The first graph defines $HFT_{i,t}$ as the fraction of shares in which HFTs are involved in any capacity, the second defines $HFT_{i,t}$ as the fraction of shares in which HFTs are involved as the liquidity supplier, the last graph defines $HFT_{i,t}$ as the fraction of shares in which HFTs are involved as the liquidity demander.

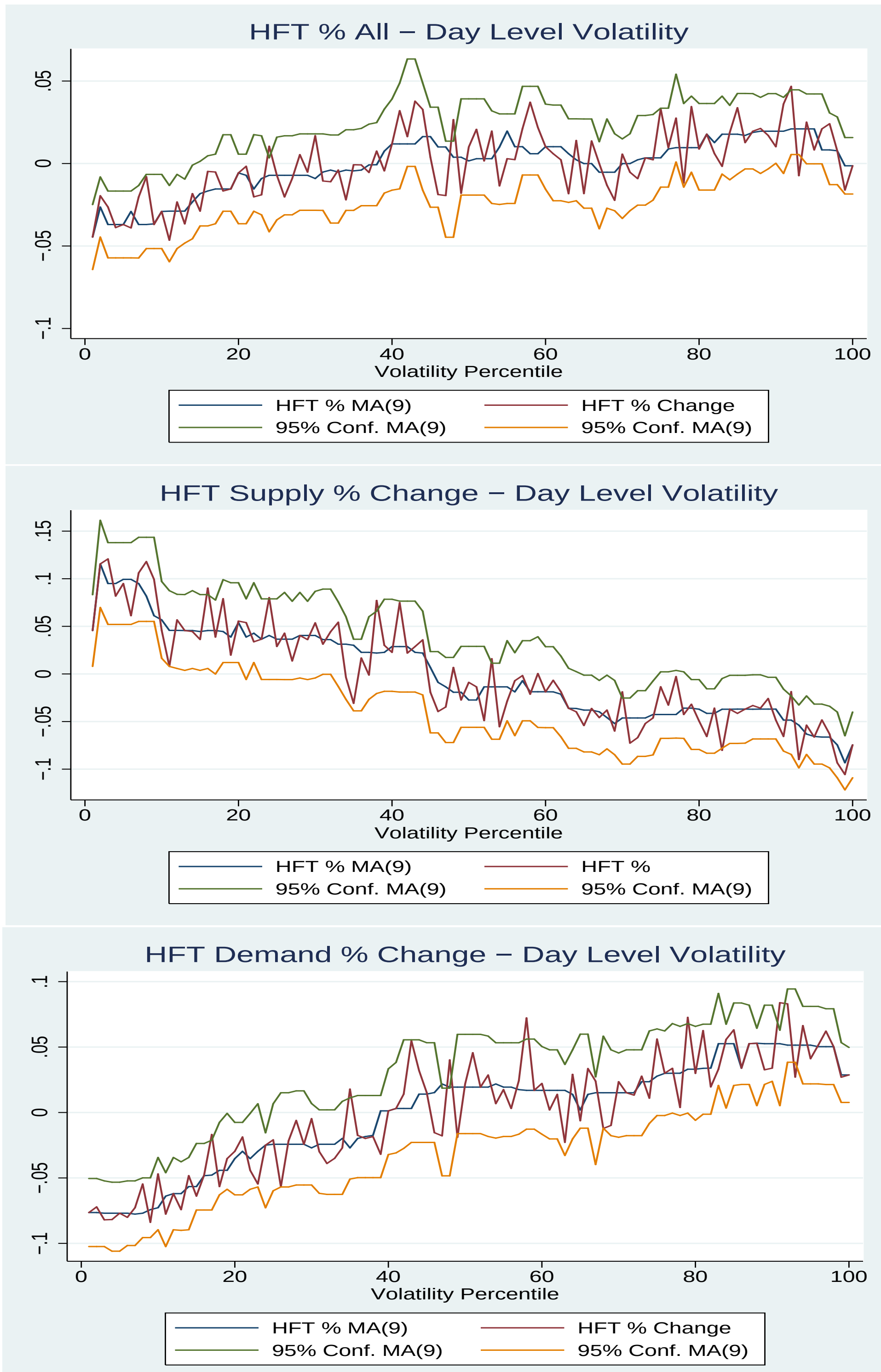


Figure 4: HFT - 15-minute Price Increases. The figures show the HFT activity for different level price increases. The X-axis is the $PriceChange$, $PriceChange_{i,t,m} = Ret_{i,t,m} \times \frac{1}{\sqrt{\sum_{i,t,m}^{t=T,m=M} \frac{1}{T*M} Ret_{i,t,m}^2}}$ where $Ret_{i,t,m}$ is the return during that time period for firm i on day t during period m. The higher the percentile, the larger the Price increase. The Y-axis is $HFTChange$, $HFTChange_j = \sum_{PriceChange_{i,t,m} \epsilon j} \frac{1}{N_j} [\frac{HFT_{i,t,m} - \sum_{m=1}^{M*T} \frac{1}{M} HFT_{i,t,m}}{\frac{1}{M} HFT_{i,t,m}}]$. HFT takes on five definitions All Activity, Buy-Demand, Buy-Supply, Sell-Demand, or Sell-Supply where each defines HFT 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 0 for that period, and remove those periods where less than 30 trades occurred. The figures show the five-period centered moving averages.

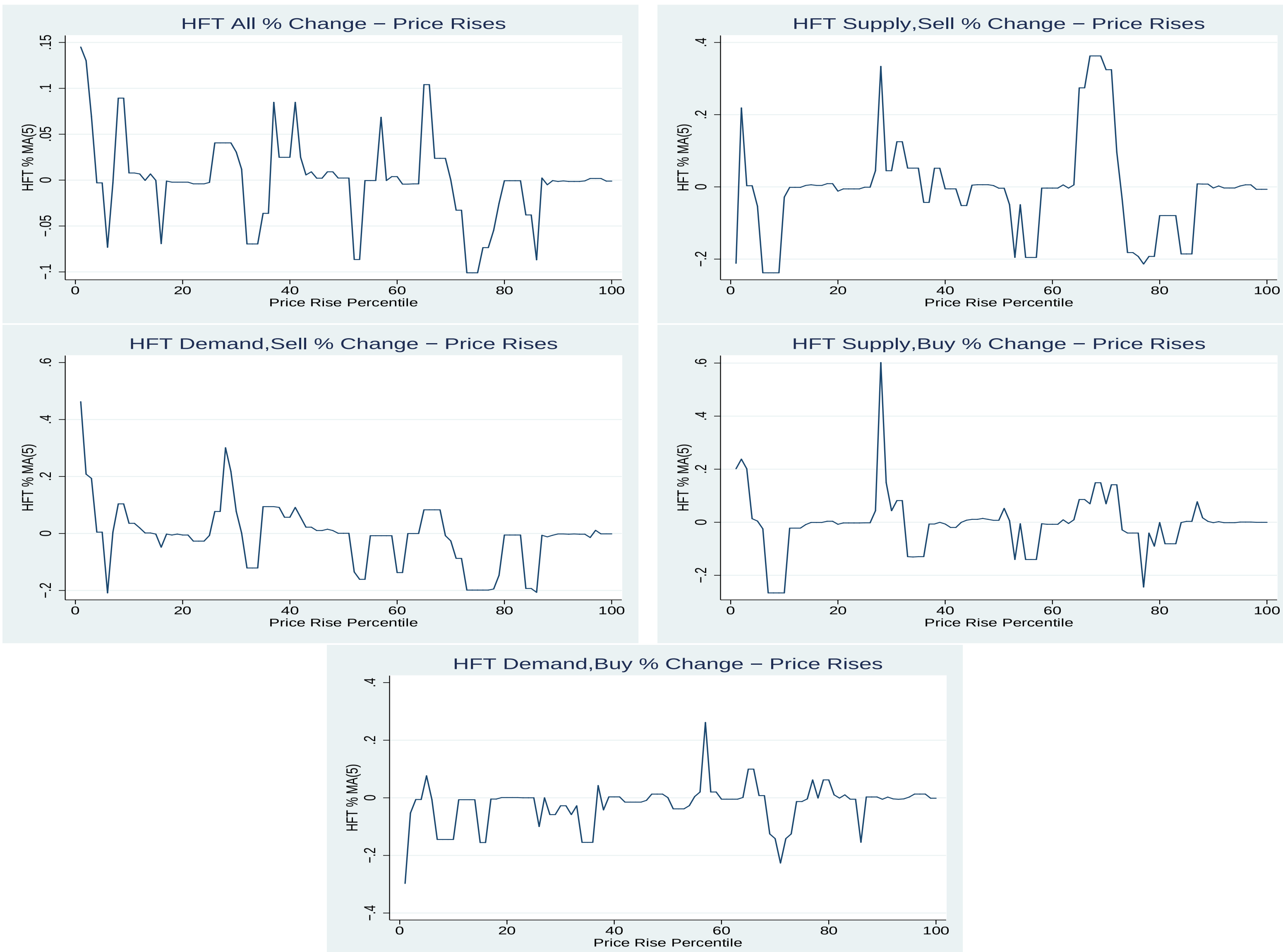


Figure 5: HFT - 15-minute Price Decreases. The figures show the HFT activity for different level price declines. The X-axis is the $PriceChange$, $PriceChange_{i,t,m} = Ret_{i,t,m} \times \frac{1}{\sqrt{\sum_{i,t,m}^{t=T,m=M} \frac{1}{T*M} Ret_{i,t,m}^2}}$ where $Ret_{i,t,m}$ is the return during that time period for firm i on day t during period m. The higher the percentile, the larger the Price decrease. The Y-axis is $HFTChange$, $HFTChange_j = \sum_{PriceChange_{i,t,m} \epsilon j} \frac{1}{N} [\frac{HFT_{i,t,m} - \sum_{m=1}^{MT} \frac{1}{M} HFT_{i,t,m}}{\frac{1}{M} HFT_{i,t,m}}]$. HFT takes on five definitions All Activity, Buy-Demand, Buy-Supply, Sell-Demand, or Sell-Supply where each defines HFT 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 0 for that period, and remove those periods where less than 30 trades occurred.

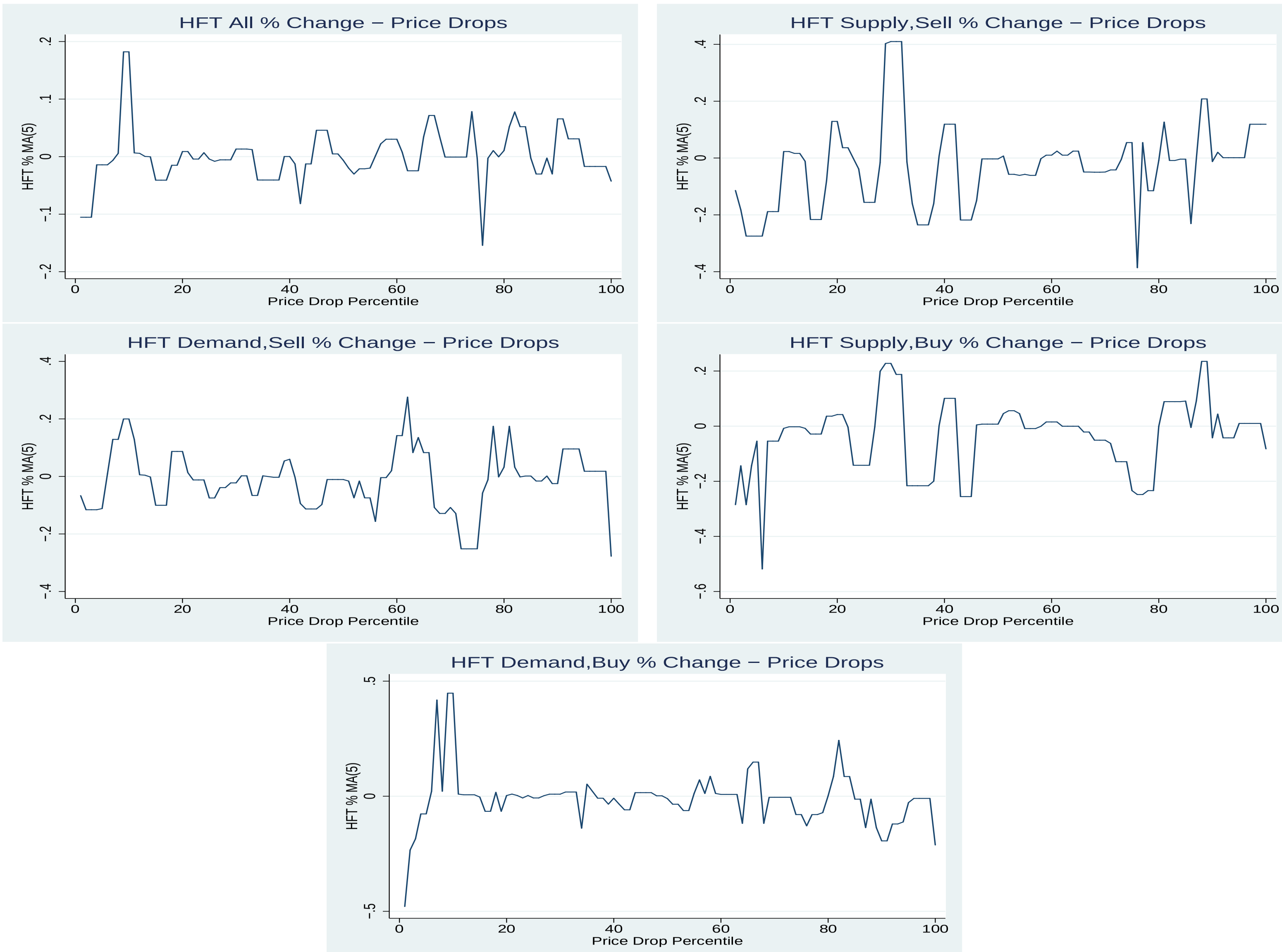


Figure 6: HFT - NHFT Time Series Book Depth. This figure includes three graphs, the price impact from removing HFTs book orders, the price impact of removing non-HFTs book orders, and the ratio of the two (non-HFTs / HFTs) for all firm sizes with a 1000 share order working through the book. The x axis has labeled the first day of five for which the data shows results. That is, The observation 01-07-08 is followed by observations on Januay 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.

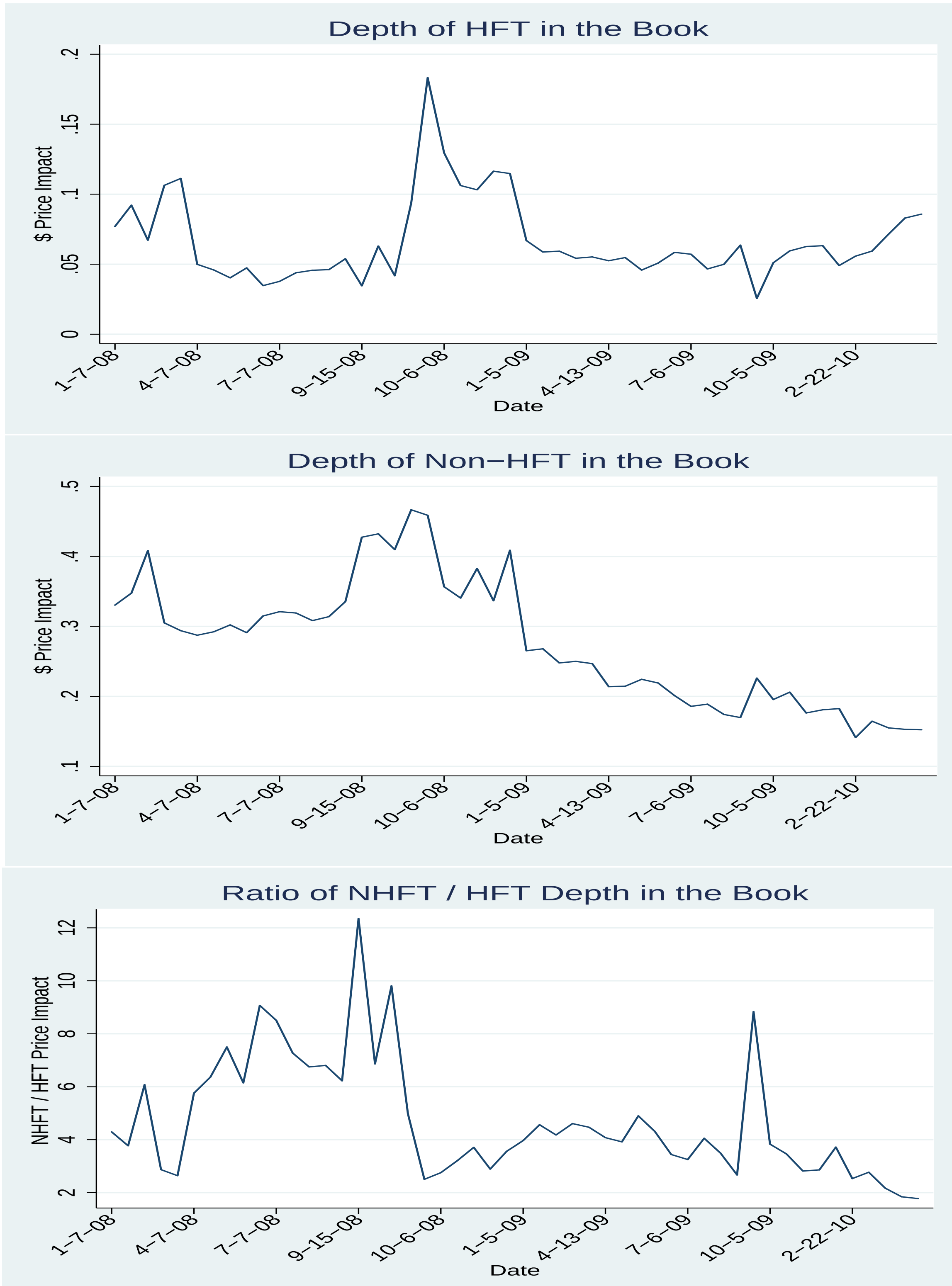


Table 1: Sample Stocks. List of tickers for the set of sample stocks containing HFT information.

AA	AAPL	ABD	ADBE	AGN	AINV	AMAT	AMED	AMGN	AMZN	ANGO	APOG
ARCC	AXP	AYI	AZZ	BARE	BAS	BHI	BIIB	BRCM	BRE	BW	BXS
BZ	CB	CBEY	CBT	CBZ	CCO	CDR	CELG	CETV	CHTT	CKH	CMCSA
CNQR	COO	COST	CPSI	CPWR	CR	CRI	CRVL	CSCO	CSE	CSL	CTRN
CTSH	DCOM	DELL	DIS	DK	DOW	EBAY	EBF	ERIE	ESRX	EWBC	FCN
FFIC	FL	FMER	FPO	FRED	FULT	GAS	GE	GENZ	GILD	GLW	GOOG
GPS	HON	HPQ	IMGN	INTC	IPAR	ISIL	ISRG	JKHY	KMB	KNOL	KR
KTII	LANC	LECO	LPNT	LSTR	MAKO	MANT	MDCO	MELI	MFB	MIG	MMM
MOD	MOS	MRTN	MXWL	NC	NSR	NUS	NXTM	PBH	PFE	PG	PNC
PNY	PPD	PTP	RIGL	ROC	ROCK	ROG	RVI	SF	SFG	SJW	SWN

Table 2: HFT Sample v. Compustat and TAQ. Panel A in table 2 describes the 120 stocks in the HFT database compared to the Compustat database. The table looks at the market capitalization, market-to-book ratio, industry, and listing exchange summary statistics and provides the t-statistic for the differences in means. The Compustat firms consist of all firms in the Compustat database with data available, that have a market capitalization greater than $10 million in 2009, and where listed on either Nasdaq or NYSE, which ammounts to 5,050 firms. The data for both the Compustat and the HFT firms are for fiscal year end on December 31, 2009. If a firm's year-end is on a different date, the fiscal year-end that is most recent, but prior to December 31, 2009, is used. The industries are categorized based on the Fama-French 10 industry groups. Panel B describes the market characteristics of the 120 stocks in the HFT sample database and compares them to the full TAQ database which includes 7,537 firms. Each firm has 5 observations. These statistics are taken for the five trading days from February 22 to February 26, 2010. The statistics considered include half spreads, stock price, bid size, offer size, daily volume traded, number of trades, and size of a trade.

Panel A: HFT Database v. Compustat Database

	HFT Database				Compustat Database				
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.	T-Test
Market Cap. (millions)	17588.24	37852.38	80.6025	197012.3	3456.773	14053.7	10.0491	322334.1	10.02
Market-to-Book	2.650261	3.134771	-11.77995	20.04067	14.17931	699.9663	-688.4559	44843.56	-0.018
Industry - Non-Durables	.0333333	.1802581			.0394059	.194578			-0.33
Industry - Durables	.025	.1567796			.0178218	.1323164			0.58
Industry - Manufacturing	.1666667	.3742406			.0819802	.2743617			3.31
Industry - Energy	.0083333	.0912871			.0417822	.2001109			-1.826
Industry - High Tech	.1583333	.3665839			.1635644	.3699164			-0.15
Industry - Telecom.	.05	.2188588			.029901	.170331			1.27
Industry - Wholesale	.0916667	.2897647			.0720792	.2586446			0.82
Industry - Health Care	.15	.3585686			.0974257	.296566			1.91
Industry - Utilities	.0333333	.1802581			.0263366	.1601502			0.47
Industry - Other	.2833333	.4525062			.4322772	.4954415			-3.26
Exchange - NYSE	.5	.5020964			.4718812	.4992581			0.61
Exchange - Nasdaq	.5	.5020964			.5281188	.4992581			-0.61
Observations	120				5050				

Panel B: HFT Database v. TAQ Database

	HFT Database		TAQ Database		
Variable	Mean	Std. Dev.	Mean	Std. Dev.	T-Test
Quoted Half Spread (Dollars)	.0711	.0863	.1744	.2503	-10.059
Stock Price (Dollars)	35.42	40.02	27	525.9	0.391
Bid Size (Hundreds of Shares)	23.88	68.69	17.59	193.7	0.791
Offer Size (Hundreds of Shares)	24.24	70.69	17.04	166.8	1.052
Daily Volume Traded (Millions of Dollars)	31.46	76.1	27.21	3911	0.027
Number of Trades	3090	3326	909.8	1724	29.966
Size of a Trade (Shares)	207.5	172.3	340.2	2224	-1.455
Observations	600		37685		

Table 3: HFT Aggregate Activity. This table looks at the prevalence of HFT in the stock market. It captures this in a variety of ways. Panel A and B look at HFT activity at a day level, ignoring firm-by-firm variations. Panel C and D look at HFT activity at a firm-day level. That is, whereas Panel A and B each have 509 (252*2+5) observations, Panel C and D have 61080 ((252*2+5)*120) observations. Panel A (C) measures HFT activity based on the percent of dollar-volume involving HFTs for each day (day-firm). Panel B (D) measures HFT activity based on the percent of trades involving HFTs for each day (day-firm). Within each Panel are three rows. The row HFT - All shows the fraction of activity where HFTs are either demanding liquidity, supplying liquidity, or both. The row HFT - Demand shows the fraction of activity where HFTs are demanding liquidity. The row HFT - Supply shows the fraction of activity where HFTs are supplying liquidity. The summary statistics include the mean, standard deviation, minimum, 5th percentile, 25th percentile, median, 75th percentile, 95th percentile, and maximum fraction of activity in which HFTs are involved.

Panel A: HFT Dollar-Volume Market-wide Participation

HFT Activity Type	Mean	Std. Dev.	Min.	5%	25%	50%	75%	95%	Max.
HFT - All	68.49	2.766	60.44	64.23	66.49	68.27	70.48	73.3	75.85
HFT - Demand	42.74	3.646	33.66	37.01	40.14	42.69	45.42	48.93	51.72
HFT - Supply	41.12	2.396	32.37	37.15	39.49	41.07	42.86	45.05	50.75

Panel B: HFT Trades Market-wide Participation

HFT Activity Type	Mean	Std. Dev.	Min.	5%	25%	50%	75%	95%	Max.
HFT - All	73.76	3.348	65.11	68.48	71.33	73.72	76.11	79.5	81.91
HFT - Demand	43.63	4.769	30.52	34.74	40.19	44.78	47.31	49.93	53.12
HFT - Supply	48.65	2.992	40.75	43.92	46.62	48.33	50.48	53.95	57.89

Panel C: HFT Dollar-Volume Firm-specific Participation

HFT Activity Type	Mean	Std. Dev.	Min.	5%	25%	50%	75%	95%	Max.
HFT - All	50.78	20	.2247	19.59	34.36	50.68	67.04	82.26	100
HFT - Demand	50.92	22.66	.2619	13.3	34.09	50.73	67.5	89.29	100
HFT - Supply	50.98	31.25	.0425	6.307	19.36	51.19	82.55	94.52	100

Panel D: HFT Trades Firm-specific Participation

HFT Activity Type	Mean	Std. Dev.	Min.	5%	25%	50%	75%	95%	Max.
HFT - All	50.75	21.62	2.994	16.97	32.55	50.79	68.92	84.25	100
HFT - Demand	50.91	21.13	.6369	14.87	36.4	50.56	65.13	87.8	100
HFT - Supply	50.96	29.88	.3628	7.214	22.03	51.31	80.24	93.69	100

Table 4: Quote Activity. This table reports summary statistics on HFTs quote activity. As stated in the data description section the data for quote-by-quote changes with HFT identification is only available from February 22, 2010 - February 26, 2010 for the 120 firms. The quote data only contains the inside bid and ask for HFT and non-HFT quotes and the available sizes for each. The measures in each panel separate the quote activity into three categories based on firm size, Small Medium, and Large whereby there are 40 firms in each category. The rows labeled Total do not condition the analysis on firm size. Panel A reports the percent of quote changes that were made by HFTs per firm-day.

Panel B: HFT Percent of Quote Revisions / Changes

Firm 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 5: Determinants of HFT Percent of the Market This table shows the result of the following OLS regression: $H_{i,t} = \alpha + MC_i * \beta_1 + MB_i * \beta_2 + NT_{i,t} * \beta_3 + NV_{i,t} * \beta_4 + Dep_{i,t} * \beta_5 + Vol_{i,t} * \beta_6 + AC_{i,t} * \beta_7 + \epsilon_{i,t}$, where i is the subscript representing the firm, t is the subscript for each day, H is the percent of share volume in which HFT are involved out of all trades, 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, NT is the number of non HFT trades that occurred, scaled by market capitalization, NV is the volume of non HFT dollars that were exchanged, scaled by market capitalization, Dep is the average depth of the bid and of the ask, equally weighted, Vol is the ten second realized volatility summed up over the day, AC is the absolute value of the Durbin-Watson score minus two from a regression of returns over the current and previous ten second period. Columns (1) - (6) shows the standardized beta coefficients and columns (7) - (12) display the regular coefficients. The dependent variables for the column are: for (1), (4), (7), and (10) the percent of shares HFTs are involved in; for (2),(5),(8),(11) the percent of shares in which HFTs demand liquidity; for (3), (6), (9), and (12) the percent of shares in which HFTs supply liquidity. Columns (4) - (6) and (10) - (12) include only clearly exogenous variables.

	Standardized Beta Coefficients						OLS Coefficients					
	All	Dem.	Sup.	All	Dem.	Sup.	All	Dem.	Sup.	All	Dem.	Sup.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Market Cap.	0.721***	0.589***	0.790***	0.757***	0.691***	0.705***	0.068***	0.050***	0.060***	0.073***	0.059***	0.054***
	(0.0035)	(0.0035)	(0.0030)	(0.0028)	(0.0027)	(0.0024)	(0.0035)	(0.0035)	(0.0030)	(0.0028)	(0.0027)	(0.0024)
Market / Book	-0.063*	-0.054	-0.072*	-0.122***	-0.092**	-0.136***	-0.006*	-0.005	-0.006*	-0.012***	-0.008**	-0.011***
	(0.0029)	(0.0029)	(0.0024)	(0.0029)	(0.0028)	(0.0025)	(0.0029)	(0.0029)	(0.0024)	(0.0029)	(0.0028)	(0.0025)
$ of Non HFT Volume	-0.138***	-0.086*	-0.165***				-0.024***	-0.013*	-0.023***			
	(0.0063)	(0.0063)	(0.0054)				(0.0063)	(0.0063)	(0.0054)			
Average Spread	-0.111***	-0.051	-0.112***				-0.043***	-0.018	-0.035***			
	(0.0110)	(0.0110)	(0.0094)				(0.0110)	(0.0110)	(0.0094)			
Average Depth	-0.132***	-0.085**	-0.101***				-0.030***	-0.017**	-0.018***			
	(0.0062)	(0.0062)	(0.0053)				(0.0062)	(0.0062)	(0.0053)			
Volatility	-0.030	-0.003	-0.091**				-0.002	-0.000	-0.005**			
	(0.0018)	(0.0018)	(0.0016)				(0.0018)	(0.0018)	(0.0016)			
Autocorrelation	-0.020	-0.023	-0.014				-0.001	-0.001	-0.001			
	(0.0015)	(0.0015)	(0.0013)				(0.0015)	(0.0015)	(0.0013)			
# of Non HFT Trades	0.042	-0.114*	0.249***				0.019	-0.047*	0.092***			
	(0.0198)	(0.0197)	(0.0169)				(0.0198)	(0.0197)	(0.0169)			
Constant	*		***		***	***	0.083*	0.029	-0.134***	0.031	-0.072***	-0.076***
	(0.0323)	(0.0323)	(0.0277)	(0.0219)	(0.0211)	(0.0189)	(0.0323)	(0.0323)	(0.0277)	(0.0219)	(0.0211)	(0.0189)
Observations	590	590	590	595	595	595	590	590	590	595	595	595
Adjusted R^2	0.575	0.472	0.522	0.533	0.447	0.458	0.575	0.472	0.522	0.533	0.447	0.458

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

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

Table 6: HFT Ordered Logit - Exploratory Regression. This table includes several explanatory variables in order to uncover in which strategies HFTs are engaged. Each explanatory variable is followed by a number between 0 and 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. For example, $retlag_0$ represents the return for the particular stock during time period t. And, the return for time period t is defined as $retlag_{i,0} = (price_{i,t} - price_{i,t-1})/price_{i,t-1}$. $Depthbid$ is the average time weighted best bid depth for stock i in that time period. $Depthask$ is the average time weighted best offer depth for stock i in that time period. $Spread$ is the average time weighted spread for company i in that time period, where spread is the best ask price minus the best bid price. $Trades$ is the number of distinct trades that occurred for company i in that time period. $DollarV$ is the dollar-volume of shares exchanged in transactions for company i in that time period. The dependent variable, HFT, is -1, 0 or 1. It takes the value -1 if during that ten second period HFTs were on net selling shares for stock i, it is zero if the HFTs performed no transaction or its buys and sell exactly canceled, and it is 1 if on net HFTs were buying shares for stock i. The regression uses firm fixed effects.

Variable	Coefficient	T-Stat	Variable	Coefficient	T-Stat
retlag0	7.461	(0.49)	depthasklag1	-8.72e-13	(-1.01)
retlag1	5.017**	(3.22)	depthasklag2	-5.15e-13	(-1.22)
retlag2	4.577***	(4.14)	depthasklag3	3.49e-13	(0.69)
retlag3	5.744***	(5.63)	depthasklag4	-2.97e-13	(-0.65)
retlag4	4.405***	(4.35)	depthasklag5	-5.88e-13	(-1.19)
retlag5	4.176***	(5.04)	depthasklag6	-5.81e-13	(-1.18)
retlag6	4.254***	(5.65)	depthasklag7	5.55e-13	(1.21)
retlag7	2.724***	(3.82)	depthasklag8	-2.03e-13	(-0.55)
retlag8	1.423*	(2.29)	depthasklag9	-1.58e-13	(-0.34)
retlag9	2.245**	(3.08)	depthasklag10	1.79e-13	(0.36)
retlag10	1.216*	(2.24)	depthasklag0	1.68e-12	(1.42)
spreadlag1	0.00528	(0.69)	tradeslag1	-0.000184	(-1.38)
spreadlag2	0.00199	(0.50)	tradeslag2	0.00000749	(0.07)
spreadlag3	-0.00549	(-1.19)	tradeslag3	0.000203**	(2.69)
spreadlag4	-0.000316	(-0.05)	tradeslag4	-0.000165*	(-1.97)
spreadlag5	0.000114	(0.03)	tradeslag5	0.000169*	(2.31)
spreadlag6	0.00456	(0.79)	tradeslag6	-0.0000886	(-0.95)
spreadlag7	0.00254	(0.30)	tradeslag7	0.0000884	(1.17)
spreadlag8	-0.00960	(-1.56)	tradeslag8	-0.000176*	(-1.99)
spreadlag9	0.00126	(0.30)	tradeslag9	-0.00000946	(-0.08)
spreadlag10	0.00870	(1.31)	tradeslag10	0.0000171	(0.14)
spreadlag0	-0.00332	(-0.47)	tradeslag0	0.000208	(1.18)
depthbidlag1	7.07e-13	(0.86)	dvolumelag1	-4.09e-14	(-0.06)
depthbidlag2	8.71e-13**	(2.74)	dvolumelag2	3.61e-13	(0.29)
depthbidlag3	6.21e-13	(1.38)	dvolumelag3	-1.68e-12*	(-2.05)
depthbidlag4	6.82e-13	(1.75)	dvolumelag4	8.88e-13	(1.31)
depthbidlag5	9.16e-13	(1.10)	dvolumelag5	-1.37e-12*	(-2.16)
depthbidlag6	-5.30e-13	(-0.98)	dvolumelag6	5.47e-13	(0.70)
depthbidlag7	-2.33e-14	(-0.06)	dvolumelag7	-3.29e-13	(-0.47)
depthbidlag8	6.43e-13	(1.77)	dvolumelag8	2.19e-13	(0.26)
depthbidlag9	-1.22e-13	(-0.21)	dvolumelag9	2.73e-13	(0.15)
depthbidlag10	-1.75e-12*	(-2.50)	dvolumelag10	-3.70e-13	(-0.26)
depthbidlag0	-1.80e-12	(-1.30)	dvolumelag0	-3.99e-13	(-0.33)
N	1281695				

Marginal effects; t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Regressions of the Buy/Sell decision, by Liquidity Type. This table reports the results from running a logit with dependent variable equal to 1 if (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 number between 1 and 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. For example, $retlag_1$ represents the return for the particular stock during time period t. And, the return for time period t is defined as $retlag_{i,t} = (price_{i,t} - price_{i,t-1})/price_{i,t-1}$. Firm fixed effects are used. The reported coefficients are the marginal effects at the mean.

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

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

Table 8: Regressions of the Buy and Sell decision, decomposition of explanatory return variable. This table reports the results from running the logit, $HFT_{i,t} = \alpha + Ret_{i,1-10} \times \beta_{1-10} + OIB_{i,1-10} \times \beta_{11-20} + \epsilon_{i,t}$, where the dependent variable is equal to 1 if (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 number between 0 and 10. This represents the number of lagged time periods away from the event occurring in the time t dependent variable. Firm fixed effects are used. The OIB variables are scaled by 100 to increase the size of the coefficients. The reported coefficients are the marginal effects at the mean.

	(1) HFT S-A	(2) HFT S-S	(3) HFT S-D	(4) HFT B-A	(5) HFT B-S	(6) HFT B-D
Ret_1	2.790*	-1.481	5.793***	-4.985**	0.424	-5.476***
	(1.394)	(0.868)	(0.787)	(1.547)	(0.864)	(1.040)
Ret_2	1.957	-0.429	3.241***	-2.572	0.704	-3.465***
	(1.369)	(0.972)	(0.822)	(1.320)	(0.927)	(0.834)
Ret_3	2.702*	0.477	2.861***	-3.980**	0.439	-4.708***
	(1.238)	(0.873)	(0.766)	(1.359)	(0.622)	(1.027)
Ret_4	1.182	-0.648	2.548**	-2.976*	0.984	-4.318***
	(1.173)	(0.901)	(0.803)	(1.250)	(0.798)	(0.926)
Ret_5	0.601	-0.237	1.191	-3.205**	1.264	-4.811***
	(1.289)	(0.981)	(0.785)	(1.197)	(0.851)	(0.712)
Ret_6	1.669	0.178	1.937*	-3.195**	0.362	-3.884***
	(1.074)	(0.758)	(0.763)	(1.098)	(0.651)	(0.814)
Ret_7	0.593	-1.266	2.547**	-0.608	1.292	-2.119**
	(0.967)	(0.750)	(0.864)	(1.071)	(0.723)	(0.727)
Ret_8	-1.245	-1.340	0.407	0.224	1.361	-1.272
	(1.072)	(0.896)	(0.705)	(1.154)	(0.755)	(0.799)
Ret_9	0.507	-1.161	2.224*	-0.206	1.098	-1.533*
	(1.217)	(0.861)	(0.925)	(1.072)	(0.867)	(0.643)
Ret_{10}	-1.003	-1.611	1.042	-0.374	-0.00465	-0.347
	(1.087)	(0.824)	(0.808)	(1.066)	(0.728)	(0.829)
OIB_1	0.513	0.0169	0.410***	-0.280	-1.182***	1.53***
	(0.307)	(0.251)	(0.102)	(0.288)	(0.196)	(0.198)
OIB_2	0.714***	0.402*	0.261**	-0.723***	-1.125***	0.586***
	(0.202)	(0.166)	(0.0798)	(0.191)	(0.116)	(0.108)
OIB_3	0.883***	0.570***	0.288***	-0.807***	-1.107***	0.330***
	(0.140)	(0.134)	(0.0575)	(0.130)	(0.0828)	(0.0773)
OIB_4	0.846***	0.470**	0.339***	-0.761***	-0.807***	0.125
	(0.166)	(0.155)	(0.0654)	(0.159)	(0.0960)	(0.0978)
OIB_5	0.839***	0.558***	0.269***	-0.710***	-0.821***	0.182*
	(0.150)	(0.137)	(0.0750)	(0.169)	(0.106)	(0.0788)
OIB_6	0.802***	0.534***	0.260***	-0.684***	-0.644***	0.0437
	(0.147)	(0.132)	(0.0675)	(0.134)	(0.0876)	(0.0731)
OIB_7	0.597***	0.441***	0.147*	-0.581***	-0.616***	0.109
	(0.131)	(0.125)	(0.0617)	(0.117)	(0.0788)	(0.0740)
OIB_8	0.446**	0.233	0.192**	-0.535***	-0.580***	0.112
	(0.148)	(0.126)	(0.0618)	(0.146)	(0.0906)	(0.0774)
OIB_9	0.463**	0.427**	0.0460	-0.589***	-0.433***	-0.0836
	(0.147)	(0.140)	(0.0625)	(0.122)	(0.0838)	(0.0728)
OIB_{10}	0.287*	0.138	0.127*	-0.435**	-0.489***	0.113
	(0.130)	(0.131)	(0.0567)	(0.147)	(0.0934)	(0.0754)
N	1377798	1377798	1343177	1377798	1366278	1377798

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

Table 9: Searching for Front Running. Column (1) shows the results for $\frac{Prob(HN)}{Prob(NH)}$, column (2) shows $\frac{Prob(HHN)}{Prob(NHH)}$, column (3) shows $\frac{Prob(HHHN)}{Prob(NHHH)}$, column (4) shows $\frac{Prob(HHHHN)}{Prob(NHHHH)}$, and column (5) shows $\frac{Prob(HHHHHN)}{Prob(NHHHHH)}$. If front running where regularly occurring it should be the case that the probability of seeing a H before a N would be more likely than the opposite. this would show up in the table as results > 1. A result of 1 would suggest seeing the two patterns is equally likely, and a value < 1 suggests its more likely to see an N followed by an H than the opposite. Depending on the number of time periods considering the pattern represents times $(t-5, t-4, t-3, t-2, t-1, t)$.

Sym.	1	2	3	4	5	Sym.	1	2	3	4	5	Sym.	1	2	3	4	5
AA	0.92	0.88	0.85	0.84	0.83	CPWR	0.92	0.88	0.86	0.84	0.83	JKHY	0.96	0.95	0.95	0.95	0.97
AAPL	0.96	0.94	0.93	0.92	0.92	CR	1.00	1.01	1.03	1.05	1.06	KMB	0.95	0.93	0.93	0.92	0.93
ABD	1.00	1.03	1.01	1.00	0.97	CRI	0.96	0.97	0.98	1.00	1.00	KNOL	1.00	1.00	0.98	0.98	1.00
ADBE	0.92	0.89	0.87	0.87	0.87	CRVL	1.01	1.00	0.98	0.98	0.97	KR	0.92	0.89	0.87	0.86	0.85
AGN	0.95	0.93	0.93	0.92	0.93	CSCO	0.94	0.90	0.87	0.84	0.83	KTII	1.00	0.92	0.94	0.94	0.96
AINV	0.95	0.93	0.91	0.91	0.90	CSE	0.92	0.89	0.87	0.86	0.86	LANC	0.99	0.99	1.00	0.99	1.01
AMAT	0.95	0.90	0.87	0.85	0.83	CSL	0.98	0.99	1.01	1.02	1.04	LECO	0.98	0.98	0.99	1.00	1.02
AMED	0.97	0.95	0.95	0.95	0.96	CTRN	0.99	0.99	0.99	0.99	1.01	LPNT	0.97	0.96	0.96	0.95	0.97
AMGN	0.93	0.90	0.89	0.88	0.88	CTSH	0.93	0.90	0.89	0.88	0.88	LSTR	0.97	0.98	0.98	0.99	1.00
AMZN	0.96	0.95	0.94	0.94	0.94	DCOM	1.00	1.00	1.01	1.00	0.99	MAKO	0.97	0.98	0.98	0.99	0.99
ANGO	1.01	1.00	0.97	0.95	0.97	DELL	0.93	0.90	0.87	0.85	0.83	MANT	0.99	0.98	1.00	0.99	1.02
APOG	0.98	0.99	1.00	1.01	1.02	DIS	0.91	0.88	0.86	0.85	0.84	MDCO	0.96	0.95	0.95	0.95	0.94
ARCC	0.95	0.94	0.93	0.92	0.92	DK	1.01	1.02	1.01	0.99	0.99	MELI	0.97	0.97	0.97	0.97	0.98
AXP	0.93	0.91	0.89	0.89	0.89	DOW	0.93	0.89	0.87	0.86	0.86	MFB	1.02	1.02	0.98	0.97	0.99
AYI	0.99	1.01	1.03	1.04	1.05	EBAY	0.92	0.88	0.86	0.84	0.83	MIG	1.01	0.98	0.99	0.98	0.98
AZZ	1.01	1.01	1.01	0.98	0.98	EBF	1.01	1.04	1.00	0.96	0.97	MMM	0.97	0.95	0.94	0.94	0.94
BARE	0.95	0.95	0.94	0.95	0.93	ERIE	1.01	1.02	1.03	1.02	1.05	MOD	1.00	1.00	0.99	0.99	0.97
BAS	0.99	1.01	1.00	0.99	1.01	ESRX	0.97	0.95	0.94	0.94	0.94	MOS	0.97	0.95	0.94	0.94	0.94
BHI	0.98	0.95	0.94	0.94	0.94	EWBC	0.94	0.92	0.91	0.91	0.91	MRTN	0.99	1.00	1.04	1.01	1.00
BIIB	0.96	0.94	0.93	0.92	0.92	FCN	0.96	0.97	0.99	0.99	1.00	MXWL	1.00	1.00	0.97	0.98	0.96
BRCM	0.92	0.89	0.87	0.86	0.85	FFIC	1.01	1.00	0.99	0.95	0.96	NC	1.07	1.04	0.97	0.95	0.95
BRE	0.97	0.98	0.98	0.99	1.01	FL	0.92	0.90	0.89	0.88	0.89	NSR	0.98	0.98	1.00	1.01	1.01
BW	1.01	1.04	1.09	1.07	1.01	FMER	0.97	0.96	0.96	0.96	0.97	NUS	1.00	1.02	1.03	1.02	1.00
BXS	0.97	0.98	0.98	0.99	1.02	FPO	1.02	1.04	1.00	0.97	0.99	NXTM	0.99	1.00	1.00	1.00	0.98
BZ	0.96	0.98	0.98	0.98	0.98	FRED	0.97	0.96	0.96	0.98	0.98	PBH	1.00	1.00	1.00	1.01	0.99
CB	0.96	0.95	0.95	0.95	0.95	FULT	0.94	0.92	0.91	0.91	0.91	PFE	0.94	0.89	0.86	0.84	0.82
CBEY	0.97	0.98	0.97	0.98	0.97	GAS	0.98	0.99	0.99	0.99	1.04	PG	0.93	0.91	0.89	0.89	0.88
CBT	0.98	0.99	0.99	1.01	1.03	GE	0.93	0.90	0.87	0.85	0.83	PNC	0.96	0.94	0.94	0.94	0.94
CBZ	0.99	1.02	0.97	0.99	0.98	GENZ	0.97	0.95	0.95	0.94	0.95	PNY	0.98	1.00	1.02	1.03	1.01
CCO	0.99	0.99	0.99	0.99	0.98	GILD	0.93	0.91	0.89	0.88	0.88	PPD	1.02	1.04	0.96	0.96	0.97
CDR	0.98	0.99	0.99	1.00	0.98	GLW	0.91	0.87	0.85	0.83	0.82	PTP	0.99	0.99	0.99	1.03	1.05
CELG	0.96	0.94	0.93	0.92	0.92	GOOG	0.97	0.95	0.94	0.93	0.93	RIGL	0.96	0.96	0.96	0.97	0.97
CETV	0.97	0.96	0.95	0.95	0.96	GPS	0.91	0.88	0.86	0.85	0.85	ROC	0.97	0.97	0.98	0.98	0.99
CHTT	0.97	0.97	0.97	0.98	0.98	HON	0.94	0.92	0.91	0.90	0.91	ROCK	0.98	0.98	0.99	1.01	0.98
CKH	1.00	1.03	1.04	1.08	1.07	HPQ	0.92	0.89	0.88	0.87	0.86	ROG	1.01	1.04	1.01	0.99	0.97
CMCSA	0.94	0.90	0.87	0.85	0.84	IMGN	0.99	0.97	1.00	0.97	0.97	RVI	0.99	1.01	1.00	0.99	0.98
CNQR	0.97	0.98	0.98	1.00	1.00	INTC	0.94	0.90	0.87	0.85	0.83	SF	1.00	1.01	1.02	1.02	1.05
COO	0.99	0.99	1.03	1.04	1.06	IPAR	1.00	1.01	0.98	0.95	0.96	SFG	0.99	1.00	1.02	1.04	1.06
COST	0.96	0.94	0.93	0.93	0.93	ISIL	0.91	0.89	0.87	0.86	0.86	SJW	1.03	1.04	1.01	0.97	0.98
CPSI	0.98	0.98	1.00	0.99	0.97	ISRG	0.96	0.96	0.95	0.95	0.95	SWN	0.94	0.92	0.91	0.91	0.91
Total	0.97	0.96	0.95	0.95	0.95												

Table 10: Diversity of HFTs Strategies. This table reports the mean per stock of the daily ratio $R = RH/RN$ where $\hat{RN} = \frac{Vol(NN)}{Vol(NH)}$ and $\hat{RH} = \frac{Vol(HN)}{Vol(HH)}$. The results are for the full sample period. $Std.Dev.$ is the standard deviation of the daily ratio R for that particular stock over time. The column $\%DaysR < 1$ is the fraction of days in which R < 1 for that stock.

Symbol	R	Std. Dev.	% Days R < 1	Symbol	R	Std. Dev.	% Days R < 1	Symbol	R	Std. Dev.	% Days R < 1
AA	1.74	0.39	0.01	CPWR	1.70	0.53	0.03	JKHY	1.56	0.53	0.12
AAPL	1.28	0.14	0.00	CR	1.25	0.74	0.42	KMB	1.48	0.40	0.05
ABD	2.28	6.96	0.32	CRI	1.54	0.85	0.19	KNOL	1.86	4.26	0.39
ADBE	1.57	0.29	0.02	CRVL	4.10	27.78	0.22	KR	1.70	0.41	0.01
AGN	1.49	0.47	0.09	CSCO	1.64	0.28	0.01	KTII	4.05	20.34	0.32
AINV	1.64	0.71	0.12	CSE	1.75	0.61	0.07	LANC	1.70	1.08	0.17
AMAT	1.55	0.27	0.01	CSL	1.59	4.98	0.35	LECO	1.73	0.81	0.15
AMED	1.76	0.80	0.08	CTRN	1.55	1.10	0.32	LPNT	1.34	0.59	0.27
AMGN	1.34	0.25	0.03	CTSH	1.62	0.33	0.01	LSTR	1.44	0.54	0.14
AMZN	1.27	0.22	0.06	DCOM	1.58	1.42	0.30	MAKO	1.20	4.35	0.14
ANGO	2.23	10.38	0.39	DELL	1.59	0.33	0.01	MANT	1.91	1.33	0.15
APOG	1.57	0.92	0.26	DIS	2.01	0.48	0.00	MDCO	1.28	0.56	0.33
ARCC	1.57	0.89	0.28	DK	1.58	2.00	0.41	MELI	1.56	0.69	0.18
AXP	1.64	0.34	0.00	DOW	1.87	0.42	0.01	MFB	1.99	5.70	0.35
AYI	1.38	0.94	0.40	EBAY	1.74	0.34	0.01	MIG	2.25	17.64	0.31
AZZ	4.88	47.16	0.35	EBF	2.82	17.26	0.37	MMM	1.54	0.38	0.02
BARE	1.34	0.57	0.29	ERIE	1.68	3.41	0.40	MOD	1.86	3.46	0.34
BAS	1.81	2.05	0.32	ESRX	1.30	0.34	0.17	MOS	1.56	0.42	0.01
BHI	1.36	0.27	0.04	EWBC	1.43	0.48	0.15	MRTN	1.58	2.53	0.41
BIIB	1.24	0.26	0.15	FCN	1.70	0.88	0.18	MXWL	1.92	8.67	0.38
BRCM	1.49	0.27	0.02	FFIC	1.61	2.46	0.41	NC	2.64	9.84	0.34
BRE	1.46	0.60	0.19	FL	1.61	0.49	0.05	NSR	1.97	10.69	0.30
BW	1.75	2.40	0.42	FMER	1.52	0.43	0.07	NUS	1.49	1.19	0.37
BXS	1.26	0.49	0.29	FPO	2.59	6.49	0.33	NXTM	2.67	21.33	0.29
BZ	1.36	2.06	0.18	FRED	1.29	0.60	0.34	PBH	2.93	8.34	0.38
CB	1.42	0.34	0.06	FULT	1.59	0.52	0.09	PFE	2.00	0.53	0.00
CBEY	1.65	6.00	0.33	GAS	1.41	0.70	0.24	PG	1.50	0.29	0.02
CBT	1.46	0.71	0.25	GE	1.81	0.39	0.00	PNC	1.40	0.29	0.05
CBZ	2.15	10.62	0.38	GENZ	1.30	0.32	0.14	PNY	1.35	0.96	0.41
CCO	2.50	9.03	0.39	GILD	1.33	0.21	0.04	PPD	2.95	12.63	0.31
CDR	3.29	26.00	0.40	GLW	2.05	0.72	0.02	PTP	1.54	2.85	0.33
CELG	1.23	0.23	0.13	GOOG	1.48	0.34	0.02	RIGL	1.28	0.86	0.35
CETV	1.81	0.99	0.18	GPS	1.86	0.49	0.01	ROC	1.55	1.03	0.29
CHTT	1.55	0.82	0.23	HON	1.85	0.51	0.02	ROCK	1.41	0.98	0.34
CKH	1.48	0.86	0.30	HPQ	1.86	0.55	0.01	ROG	3.91	36.73	0.39
CMCSA	1.66	0.32	0.01	IMGN	1.83	3.32	0.35	RVI	2.54	21.04	0.31
CNQR	1.40	0.57	0.20	INTC	1.62	0.30	0.00	SF	1.54	1.35	0.32
COO	1.33	0.88	0.37	IPAR	2.02	7.44	0.39	SFG	1.49	0.91	0.29
COST	1.31	0.25	0.07	ISIL	1.72	0.50	0.02	SJW	2.75	12.11	0.36
CPSI	2.35	13.15	0.38	ISRG	1.71	0.57	0.04	SWN	1.47	0.29	0.04
Total	1.79	8.42	0.20								

Table 11: HFT - Exogenous Volatility Relationship. This table shows the results from two different approaches of trying to understand the impact volatility has on HFT activity. The dependent variable in column (1) is the percent of shares in stock i in which HFT was involved, in column (2) it is the percent of shares in stock i in which HFT was involved and was demanding liquidity, in column (3) it is the percent of shares in stock i in which HFT was involved and was supplying liquidity. In Panel A the explanatory variable, $QuarterlyEADummy$ which is one for firm i if the observation is on the day and next day on which firm i reported its quarterly earnings, and zero otherwise. In Panel B the explanatory variable, $LehmanWeekDummy$ is one for all firms for observations on the dates September 15, 2008 - September 19, 2008 and zero otherwise. Firm fixed effects are used.

Panel A - HFT - Exogenous Volatility, Quarterly Earnings

	(1)	(2)	(3)
	HFT - ALL	HFT - Demand	HFT - Supply
Quarterly EA Dummy	0.00684*	-0.00116	0.0124***
	(0.00311)	(0.00288)	(0.00231)
Constant	0.756***	0.402***	0.545***
	(0.0167)	(0.0155)	(0.0125)
Observations	4806	4806	4806
Adjusted R^2	0.743	0.606	0.783

Standard errors in parentheses

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

Panel B - HFT - Exogenous Volatility, Lehman Failure

	(1)	(2)	(3)
	HFT - ALL	HFT - Demand	HFT - Supply
Lehman Week Dummy	0.0130**	0.0103*	0.0142***
	(0.00464)	(0.00435)	(0.00322)
Constant	0.753***	0.390***	0.542***
	(0.0257)	(0.0242)	(0.0179)
Observations	1200	1200	1200
Adjusted R^2	0.840	0.730	0.883

Standard errors in parentheses

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

Table 12: HFT and non-HFT Long-Run Impulse Response Functions. This table reports the average long-run (10 events in the future) impulse response function for HFT and non-HFT. The last column reports the T-statistics for the HFT - non-HFT difference for each security.

Stock	HFT	Non HFT	T Test	Stock	HFT	Non HFT	T Test	Stock	HFT	Non HFT	T Test
AA	1.607	1.596	0.061	CSCO	2.261	1.483	1.193	LECO	0.857	-0.130	1.482
AAPL	0.150	0.449	-1.066	CSE	1.011	0.596	5.731	LPNT	1.897	1.212	1.107
ABD	9.985	4.546	1.043	CSL	3.189	7.428	-3.511	LSTR	2.202	1.136	1.808
ADBE	1.049	0.753	2.405	CTRN	1.631	-0.003	1.491	MAKO	1.488	0.899	2.080
AGN	1.157	0.089	2.982	CTSH	13.210	-10.877	1.717	MANT	-8.091	-0.540	-0.814
AINV	3.845	2.204	2.875	DCOM	0.707	0.549	2.407	MDCO	1.872	1.872	-0.000
AMAT	1.536	1.050	2.501	DELL	4.955	4.255	0.237	MELI	5.887	3.159	2.195
AMED	2.843	1.877	1.543	DIS	1.518	0.840	2.553	MFB	3.399	0.788	3.284
AMGN	0.647	0.418	2.744	DOW	1.042	0.960	0.593	MIG	-29.172	-2.589	-0.726
AMZN	0.699	0.535	1.659	EBAY	1.565	0.904	2.480	MMM	5.719	3.783	0.356
APOG	0.323	1.658	-0.688	ERIE	1.219	1.214	0.034	MOD	0.846	0.485	4.458
ARCC	4.215	0.358	2.968	EWBC	-0.488	-3.009	0.811	MOS	7.608	3.085	1.320
AXP	2.604	1.719	1.110	FCN	2.800	1.984	1.431	MRTN	1.687	1.087	2.154
AYI	1.054	0.746	2.691	FFIC	1.854	0.900	2.675	MXWL	4.840	-3.752	1.680
BAS	0.003	-0.244	0.145	FL	5.589	-4.170	1.435	NSR	3.045	2.259	0.773
BHI	10.987	3.239	2.667	FMER	2.973	2.671	0.420	NUS	1.375	0.794	1.129
BIIB	0.659	0.621	0.260	FPO	1.781	0.889	2.979	NXTM	1.973	1.155	0.957
BRCM	1.247	0.689	9.814	FRED	15.008	-18.457	1.684	PBH	14.133	4.891	1.077
BRE	1.066	1.047	0.181	FULT	0.664	1.178	-0.291	PFE	11.802	1.541	1.793
BW	1.463	0.525	3.015	GAS	3.465	2.360	1.826	PG	1.089	1.217	-0.534
BXS	4.945	2.890	0.393	GE	1.328	0.840	1.159	PNC	0.676	0.553	2.083
BZ	1.627	0.864	0.647	GENZ	1.206	0.980	1.357	PNY	0.820	0.681	1.342
CB	3.882	4.642	-0.287	GILD	0.782	0.585	1.401	PTP	1.905	1.398	0.517
CBEY	0.813	0.581	3.384	GLW	0.644	0.670	-0.476	RIGL	2.105	1.643	0.656
CBT	3.313	1.770	1.041	GOOG	1.539	1.700	-0.637	ROC	2.768	3.592	-0.393
CCO	2.239	1.239	1.790	GPS	0.819	0.468	3.575	ROCK	2.516	1.762	0.548
CDR	10.974	2.623	1.041	HON	1.450	1.513	-0.495	SF	2.704	2.025	0.384
CELG	5.719	2.649	0.551	HPQ	1.237	0.604	4.786	SFG	2.356	2.092	0.286
CETV	0.956	0.579	3.215	IMGN	0.730	0.633	1.840	SWN	1.680	0.439	2.349
CKH	3.798	1.993	3.061	INTC	3.440	3.235	0.140				
CMCSA	0.841	0.644	0.19	IPAR	1.021	0.769	6.930				
CNQR	1.351	0.931	3.638	ISIL	3.365	2.177	0.262				
COO	2.941	0.583	4.398	ISRG	2.364	1.409	4.782				
COST	1.813	1.095	1.301	JKHY	1.849	0.940	3.497				
CPSI	0.676	0.575	1.867	KMB	2.094	0.712	2.191				
CPWR	2.911	1.782	0.519	KNOL	1.225	0.394	4.986				
CR	3.431	0.808	4.790	KR	0.357	2.713	-0.225				
CRI	2.296	-0.721	2.314	LANC	1.448	1.539	-0.431				
Overall	1.017	0.759	3.476								

Table 13: Long-Run - Short Run Impulse Response Functions. This table reports the average long-run - short run HFT and non-HFT impulse response function (IRF), where the long run is the 10 events in the future IRF minus the one period IRF. The last column reports the T-statistic for the HFT - non-HFT difference for each security.

Stock	HFT	Non HFT	T Test	Stock	HFT	Non HFT	T Test	Stock	HFT	Non HFT	T Test
AA	0.815	0.848	-0.225	CSCO	0.756	0.777	-0.044	LECO	0.158	-0.469	1.027
AAPL	-0.050	0.174	-0.886	CSE	0.633	0.405	3.433	LPNT	0.354	0.755	-0.599
ABD	4.772	0.514	1.198	CSL	2.343	2.963	-0.607	LSTR	0.849	0.505	0.527
ADBE	0.611	0.382	2.383	CTRN	0.110	-0.295	0.418	MAKO	0.494	0.199	1.202
AGN	0.146	-0.247	1.118	CTSH	5.023	-7.641	1.384	MANT	-8.150	-3.740	-0.483
AINV	2.239	1.081	2.615	DCOM	0.260	0.117	2.467	MDCO	0.709	1.233	-0.339
AMAT	1.145	0.726	3.186	DELL	0.751	11.427	-1.098	MELI	1.940	1.303	0.504
AMED	1.535	1.315	0.436	DIS	1.068	0.655	2.257	MFB	2.008	0.459	2.375
AMGN	0.300	0.099	2.767	DOW	0.493	0.417	0.757	MIG	1.961	-2.427	0.625
AMZN	0.166	0.114	1.026	EBAY	0.515	0.040	2.195	MMM	3.294	0.732	0.510
APOG	-2.298	2.489	-1.480	ERIE	0.707	0.728	-0.159	MOD	0.330	0.154	3.837
ARCC	0.591	-1.159	0.922	EWBC	-4.875	-3.751	-0.242	MOS	1.646	-0.795	1.134
AXP	1.294	0.665	1.706	FCN	1.168	0.803	1.145	MRTN	0.739	0.418	1.863
AYI	0.475	0.231	3.598	FFIC	0.372	0.247	0.364	MXWL	1.457	-3.906	1.794
BAS	-2.271	-1.567	-0.369	FL	0.698	-6.941	1.333	NSR	1.660	0.030	2.216
BHI	7.158	0.875	3.210	FMER	2.157	1.165	1.870	NUS	0.542	-0.036	1.242
BIIB	0.090	-0.021	1.084	FPO	0.847	0.194	3.179	NXTM	0.079	0.328	-0.357
BRCM	0.758	0.301	6.917	FRED	3.590	-11.955	0.965	PBH	7.416	5.846	0.169
BRE	0.505	0.481	0.303	FULT	-0.318	0.276	-0.308	PFE	9.244	-3.007	2.104
BW	0.390	0.170	0.678	GAS	2.082	1.379	1.309	PG	0.778	0.801	-0.121
BXS	0.229	-1.465	0.292	GE	0.080	0.365	-0.605	PNC	0.370	0.175	3.134
BZ	-0.051	-0.082	0.036	GENZ	0.723	0.676	0.308	PNY	0.253	0.107	1.774
CB	0.642	-0.760	0.784	GILD	0.187	0.123	0.891	PTP	0.433	0.579	-0.158
CBEY	0.327	0.259	0.972	GLW	0.337	0.302	1.145	RIGL	0.710	1.002	-0.454
CBT	1.182	0.580	0.458	GOOG	0.976	0.701	1.248	ROC	-0.292	0.839	-0.513
CCO	0.484	0.447	0.063	GPS	0.527	0.345	2.264	ROCK	0.689	1.013	-0.282
CDR	6.129	-0.344	1.082	HON	1.025	0.725	2.021	SF	-0.373	-0.611	0.180
CELG	3.466	0.150	0.543	HPQ	0.497	0.044	3.578	SFG	1.035	1.272	-0.275
CETV	0.257	0.104	0.852	IMGN	0.265	0.208	1.463	SWN	0.546	-0.347	1.868
CKH	2.284	1.073	2.266	INTC	2.013	0.171	1.588				
CMCSA	-0.285	0.343	-0.55	IPAR	0.677	0.611	1.368				
CNQR	0.994	0.677	3.392	ISIL	3.154	1.229	0.486				
COO	1.255	0.031	2.282	ISRG	1.581	0.640	5.740				
COST	0.318	0.392	-0.109	JKHY	1.200	0.662	2.471				
CPSI	0.318	0.203	1.781	KMB	0.988	0.014	1.980				
CPWR	-0.162	0.528	-0.286	KNOL	0.539	0.133	3.710				
CR	2.684	0.499	5.844	KR	-8.166	-5.971	-0.119				
CRI	0.623	-1.252	1.768	LANC	0.946	0.925	0.117				
Overall	0.515	0.341	3.563								

Table 14: HFT - non-HFT Variance Decomposition. This table reports the percentage of the variance of the efficient price correlated with HFT and non-HFT trades. The remainder is in the Return column (unreported) and is interpreted as the price discovery from publicly available information.

Stock	HFT %	Non HFT %	T Test	Stock	HFT %	Non HFT %	T Test	Stock	HFT %	Non HFT %	T Test
AA	0.366	0.113	3.790	CPWR	0.025	0.030	-1.729	JKHY	0.107	0.067	3.929
AAPL	0.002	0.002	-1.463	CR	0.027	0.020	1.907	KMB	0.002	0.003	-1.428
ABD	0.117	0.102	1.115	CRI	0.000	0.000	-1.652	KNOL	0.119	0.048	2.247
ADBE	0.053	0.029	4.171	CRVL	0.278	0.178	5.319	KR	0.002	0.004	-3.106
AGN	0.015	0.013	0.502	CSCO	0.003	0.003	-0.781	KTII	0.079	0.070	0.535
AINV	0.120	0.096	1.207	CSE	0.032	0.030	0.254	LANC	0.047	0.020	2.933
AMAT	0.016	0.014	1.511	CSL	0.002	0.002	0.738	LECO	0.021	0.020	0.189
AMED	0.147	0.111	1.261	CTRN	0.141	0.070	5.130	LPNT	0.039	0.016	2.923
AMGN	0.243	0.041	1.682	CTSH	0.002	0.018	-1.359	LSTR	0.002	0.008	-1.804
AMZN	0.005	0.010	-0.776	DCOM	0.147	0.268	-1.464	MAKO	0.021	0.055	-2.115
ANGO	0.004	0.009	-1.091	DELL	0.085	0.059	1.098	MANT	0.003	0.005	-2.266
APOG	0.010	0.035	-1.436	DIS	0.003	0.002	1.490	MDCO	0.031	0.023	2.365
ARCC	0.105	0.029	4.205	DK	0.037	0.012	5.774	MELI	0.002	0.031	-1.357
AXP	0.023	0.013	1.377	DOW	0.099	0.066	5.802	MFB	0.019	0.109	-1.018
AYI	0.008	0.009	-0.852	EBAY	0.001	0.001	-2.158	MIG	0.173	0.040	3.846
AZZ	0.073	0.573	-2.449	EBF	0.003	0.005	-1.205	MMM	0.001	0.001	-0.563
BARE	0.001	0.002	-0.613	ERIE	0.011	0.009	1.124	MOD	0.097	0.014	5.212
BAS	0.129	0.027	5.701	EWBC	0.021	0.015	2.355	MOS	0.003	0.009	-1.525
BHI	0.110	0.059	3.048	FCN	0.002	0.040	-1.125	MRTN	0.002	0.007	-1.856
BIIB	0.080	0.045	3.477	FFIC	0.010	0.012	-1.068	MXWL	0.004	0.007	-5.671
BRCM	0.053	0.026	1.749	FL	0.039	0.029	1.851	NC	0.013	0.035	-2.208
BRE	0.002	0.002	0.356	FMER	0.010	0.008	0.590	NSR	0.016	0.010	2.175
BW	0.027	0.043	-0.868	FPO	0.010	0.032	-2.187	NUS	0.000	0.001	-1.148
BXS	0.003	0.002	0.484	FRED	0.017	0.025	-1.370	NXTM	0.037	0.065	-0.996
BZ	0.197	0.059	5.198	FULT	0.049	0.011	4.218	PBH	0.166	0.079	1.494
CB	0.013	0.010	0.541	GAS	0.267	0.068	2.177	PFE	0.211	0.080	6.714
CBEY	0.022	0.007	5.016	GE	0.078	0.050	3.899	PG	0.081	0.036	9.966
CBT	0.001	0.003	-4.709	GENZ	0.139	0.109	4.564	PNC	0.014	0.017	-0.662
CBZ	0.001	0.003	-2.638	GILD	0.039	0.041	-0.124	PNY	0.002	0.004	-4.382
CCO	0.004	0.012	-2.011	GLW	0.203	0.102	2.420	PPD	0.027	0.045	-0.756
CDR	0.077	0.178	-0.716	GOOG	0.079	0.037	2.671	PTP	0.001	0.002	-3.096
CELG	0.010	0.012	-1.973	GPS	0.085	0.027	2.326	RIGL	0.020	0.008	3.759
CETV	0.044	0.100	-5.152	HON	0.082	0.037	2.897	ROC	0.003	0.002	1.005
CHTT	0.068	0.018	2.436	HPQ	0.002	0.003	-1.859	ROCK	0.000	0.000	1.947
CKH	0.210	0.182	0.817	IMGN	0.300	0.168	5.207	ROG	0.003	0.003	-1.350
CMCSA	0.025	0.024	0.150	INTC	0.001	0.002	-2.164	RVI	0.032	0.034	-0.502
CNQR	0.026	0.020	2.054	IPAR	0.043	0.029	1.534	SF	0.030	0.024	1.352
COO	0.139	0.097	1.838	ISIL	0.112	0.073	2.436	SFG	0.001	0.000	3.414
COST	0.007	0.007	-0.074	ISRG	0.024	0.023	0.381	SJW	0.073	0.017	3.701
CPSI	0.059	0.250	-1.799								
Overall	.195	.105	2.654								

Table 15: HFT and non-HFT Information Shares: This table reports the Hasbrouck (1995) information shares for HFT and non-HFT.

Firm	HFT	nHFT	Tstat	Firm	HFT	nHFT	Tstat	Firm	HFT	nHFT	Tstat
AA	0.445	0.555	-3.838	CR	0.574	0.426	0.792	KMB	0.930	0.070	7.960
AAPL	0.499	0.501	-0.036	CRI	0.803	0.197	3.080	KNOL	0.376	0.624	-0.900
ABD	0.431	0.569	-0.521	CRVL	0.780	0.220	11.649	KR	0.414	0.586	-1.765
ADBE	0.450	0.550	-0.385	CSCO	0.509	0.491	0.320	LANC	0.937	0.063	61.284
AGN	0.869	0.131	7.494	CSE	0.357	0.643	-3.640	LECO	0.698	0.302	1.910
AINV	0.432	0.568	-2.113	CSL	0.777	0.223	2.140	LPNT	0.767	0.233	2.322
AMAT	0.539	0.461	2.874	CTRN	0.878	0.122	8.472	LSTR	0.847	0.153	10.627
AMED	0.974	0.026	96.815	CTSH	0.490	0.510	-0.097	MAKO	0.419	0.581	-0.843
AMGN	0.606	0.394	0.595	DCOM	0.562	0.438	0.550	MANT	0.547	0.453	0.316
AMZN	0.352	0.648	-2.287	DELL	0.514	0.486	0.636	MDCO	0.423	0.577	-1.028
ANGO	0.352	0.648	-1.494	DIS	0.258	0.742	-9.594	MELI	0.750	0.250	1.678
APOG	0.566	0.434	1.176	DK	0.770	0.230	2.865	MFB	0.581	0.419	0.427
ARCC	0.354	0.646	-1.914	DOW	0.630	0.370	1.877	MIG	0.104	0.896	-69.826
AXP	0.233	0.767	-4.253	EBAY	0.518	0.482	0.178	MMM	0.854	0.146	4.984
AYI	0.640	0.360	0.791	EBF	0.429	0.571	-0.623	MOD	0.293	0.707	-3.182
AZZ	0.492	0.508	-0.072	ERIE	0.565	0.435	0.830	MOS	0.777	0.223	4.443
BAS	0.470	0.530	-0.245	EWBC	0.861	0.139	15.832	MRTN	0.692	0.308	1.806
BHI	0.710	0.290	1.229	FCN	0.568	0.432	0.575	MXWL	0.497	0.503	-0.021
BIIB	0.672	0.328	2.521	FFIC	0.266	0.734	-2.581	NC	0.864	0.136	11.814
BRCM	0.282	0.718	-8.553	FL	0.414	0.586	-1.077	NSR	0.354	0.646	-1.011
BRE	0.770	0.230	3.299	FMER	0.703	0.297	4.198	NUS	0.607	0.393	0.526
BW	0.803	0.197	4.292	FPO	0.820	0.180	4.232	NXTM	0.776	0.224	2.865
BXS	0.511	0.489	0.109	FRED	0.612	0.388	0.985	PBH	0.350	0.650	-1.257
BZ	0.290	0.710	-2.013	FULT	0.451	0.549	-1.223	PFE	0.516	0.484	0.531
CB	0.825	0.175	4.986	GAS	0.548	0.452	0.377	PG	0.662	0.338	1.245
CBEY	0.416	0.584	-0.931	GE	0.518	0.482	1.073	PNC	0.832	0.168	6.445
CBT	0.720	0.280	1.892	GENZ	0.835	0.165	8.274	PNY	0.502	0.498	0.016
CBZ	0.750	0.250	1.691	GILD	0.424	0.576	-1.083	PPD	0.253	0.747	-1.645
CCO	0.529	0.471	0.320	GLW	0.157	0.843	-7.506	PTP	0.666	0.334	1.435
CDR	0.524	0.476	0.255	GOOG	0.937	0.063	19.317	RIGL	0.462	0.538	-0.417
CELG	0.864	0.136	4.487	GPS	0.332	0.668	-1.103	ROC	0.601	0.399	0.752
CETV	0.735	0.265	5.398	HON	0.685	0.315	1.549	ROCK	0.339	0.661	-1.671
CKH	0.499	0.501	-0.020	HPQ	0.430	0.570	-0.556	ROG	0.869	0.131	3.987
CMCSA	0.582	0.418	2.760	IMGN	0.321	0.679	-1.341	RVI	0.445	0.555	-0.480
CNQR	0.845	0.155	12.404	INTC	0.512	0.488	0.933	SF	0.367	0.633	-1.104
COO	0.767	0.233	3.353	IPAR	0.809	0.191	3.178	SFG	0.828	0.172	3.396
COST	0.555	0.445	0.588	ISIL	0.546	0.454	0.493	SWN	0.853	0.147	3.182
CPSI	0.508	0.492	0.058	ISRG	0.735	0.265	3.964				
CPWR	0.382	0.618	-1.577	JKHY	0.648	0.352	1.096				
Overall	0.5812	0.4188	6.1554								

Table 16: HFT Time at Best Quotes. This table reports the number of minutes HFTs are at the best bid or offer. The time at which both HFTs and non-HFTs are both at the best quotes is included in the value. Panel A is for all stocks at all times, Panel B and C divide the time at the best bid and offer based on whether spreads that day are higher then average. Panel B reports the results for days in which quotes are below their daily mean. Panel C reports the results for days in which quotes are at or above their daily mean. In each Panel the data are divided into three groups, with 40 firms each, based on firm size and the results are the different rows, Small, Medium, and Large. The total row is the unconditional results.

Panel A: All

Firm Size	Mean	Std. Dev.	Min.	5%	25%	50%	75%	95%	Max.
Small	390.3	209.6	.0905	15.87	290.4	423.7	527	725.1	759.7
Medium	482.4	182.3	27.67	120.3	396	467.6	601.1	774.8	779.6
Large	658.9	160.5	168.3	253.1	619.4	735.2	773.8	779.1	779.6
Total	509.3	215.9	.0905	63.9	391.5	506	719.5	778.5	779.6

Panel B: Low Spread

Firm Size	Mean	Std. Dev.	Min.	5%	25%	50%	75%	95%	Max.
Small	392.6	196.8	.0905	10.13	320.6	407.8	513.7	703.9	759.7
Medium	486.8	184.7	27.67	77.46	407.4	467.6	618.8	776.4	779.5
Large	644.4	170.3	198	250.4	574.9	728.5	773.6	779.1	779.5
Total	509	211.8	.0905	77.46	388.8	485.8	727.8	778.5	779.5

Panel C: High Spread

Firm Size	Mean	Std. Dev.	Min.	5%	25%	50%	75%	95%	Max.
Small	388.1	222.2	.6929	16.56	150.6	437.1	529.8	733.2	749
Medium	478.7	181	50.51	131.9	391	468	595.8	773.3	779.6
Large	673.8	149.2	168.3	253.1	639.7	738.5	775.3	779.1	779.6
Total	509.5	220.1	.6929	50.51	393.9	511.7	718.2	778.4	779.6

Table 17: Determinants of HFT Percent of Liquidity Supplying The dependent variable is the ratio of number of minutes HFTs provides the inside bid or ask divided by the total number of minutes of when the inside bid and ask differ between HFTs and non-HFTs. Columns (1) and (2) display the standardized beta coefficients, columns (3) and (4) show regular OLS coefficients. Columns (3) and (4) include only coefficients that are clearly exogenous.

	(1)	(2)	(3)	(4)
Market Cap.	0.654***	0.565***	0.083***	0.073***
	(0.0055)	(0.0046)	(0.0055)	(0.0046)
Market / Book	-0.163***	-0.253***	-0.021***	-0.033***
	(0.0045)	(0.0047)	(0.0045)	(0.0047)
$ of Non HFT Volume	-0.162***		-0.380***	
	(0.0995)		(0.0995)	
Average Spread	-0.165***		-0.086***	
	(0.0175)		(0.0175)	
Average Depth	-0.086**		-0.026**	
	(0.0098)		(0.0098)	
Volatility	-0.241***		-0.021***	
	(0.0029)		(0.0029)	
Autocorrelation	-0.006		-0.433	
	(2.3699)		(2.3699)	
# of Non HFT Trades	0.217***	s	0.134***	
	(0.0314)		(0.0314)	
Constant	*		-0.131*	-0.036
	(0.0514)	(0.0360)	(0.0514)	(0.0360)
Observations	590	595	590	595
Adjusted R^2	0.410	0.299	0.410	0.299

Standardized beta coefficients in columns (1) and (2) ; Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Liquidity Book Impact. This table looks at the liquidity depth of HFT and non-HFT by analyzing the price impact for different size firms if a varying range of trade-sizes were to hit the book. The two-column wide labels, Very Small to Large refer to the firm size. The columns labeled Dollars in Panel A (Panel B) is the dollar difference as a result of HFTs (Non-HFTs) being in the market. The column labeled Basis in Panel A (Panel B) is the percent basis points change in price as a result of HFTs (Non-HFTs) being in the market. The different rows represent a varying number of shares traded. This table is for when a buy order hits the book (offer depth), similar but unreported results occur for when a sell order hits the book (bid depth).

Panel A: HFT Withdrawal Impact

Trade Size	Large		Medium		Small		Very Small		All	
	Basis	Dollars	Basis	Dollars	Basis	Dollars	Basis	Dollars	Basis	Dollars
100	1.065	0.004	2.474	0.008	8.074	0.026	12.789	0.020	5.176	0.013
200	1.260	0.005	3.686	0.012	9.734	0.030	17.049	0.028	6.739	0.016
300	1.331	0.007	4.151	0.014	11.450	0.035	19.328	0.032	7.683	0.019
400	1.428	0.008	4.619	0.016	13.233	0.040	22.283	0.037	8.784	0.022
500	1.592	0.011	5.161	0.019	16.307	0.051	25.041	0.042	10.147	0.027
600	1.663	0.013	6.042	0.023	19.934	0.065	28.749	0.048	11.866	0.033
700	1.770	0.016	7.162	0.028	22.472	0.075	31.133	0.052	13.176	0.038
800	1.855	0.017	9.394	0.037	26.358	0.088	34.587	0.058	15.250	0.045
900	1.955	0.018	11.539	0.045	29.677	0.098	38.725	0.064	17.330	0.050
1000	2.036	0.019	13.204	0.052	33.324	0.108	42.900	0.072	19.352	0.056

Panel B: Non-HFT Withdrawal Impact

Trade Size	Large		Medium		Small		Very Small		All	
	Basis	Dollars	Basis	Dollars	Basis	Dollars	Basis	Dollars	Basis	Dollars
100	3.602	0.018	32.028	0.105	44.349	0.095	53.770	0.102	29.128	0.072
200	4.400	0.022	32.322	0.105	42.481	0.091	51.532	0.108	28.656	0.074
300	4.825	0.024	33.716	0.108	46.595	0.111	50.145	0.105	29.706	0.079
400	5.617	0.028	37.564	0.121	47.966	0.116	51.790	0.107	31.574	0.085
500	6.521	0.033	45.801	0.151	49.979	0.125	60.458	0.119	36.147	0.099
600	8.168	0.041	55.427	0.186	53.886	0.136	70.117	0.130	41.910	0.115
700	9.451	0.048	62.154	0.208	57.413	0.141	77.518	0.138	46.272	0.126
800	10.429	0.054	66.734	0.223	58.548	0.138	81.071	0.143	48.727	0.132
900	11.041	0.059	69.913	0.234	63.340	0.146	82.834	0.145	51.068	0.139
1000	11.530	0.062	71.606	0.239	66.057	0.148	81.785	0.143	52.006	0.141

Table 19: Price Impact for HFT and non-HFT Liquidity Providing Trades This table reports the average long-run (10 events in the future) impulse response function for HFT and non-HFT supplied trades. That is, in table 12 I defined q^H and q^N based on who is demanding liquidity, now I do it for the supplier of liquidity in a trade. The q^H will be a +1 when a HFT supplier sells and -1 when a HFT supplier buys, The q^N value is similarly defined for non-HFTr supplied trades. The last column reports the T-statistics for the HFT - non-HFT difference for each security.

Firm	HFT	nHFT	Tstat	Firm	HFT	nHFT	Tstat	Firm	HFT	nHFT	Tstat
AA	1.111	2.159	-6.304	CPWR	1.047	2.507	-3.682	JKHY	0.551	1.317	-1.416
AAPL	0.380	0.099	1.276	CR	-0.683	1.867	-1.891	KMB	0.439	0.589	-1.103
ABD	4.190	1.650	0.511	CRI	0.541	0.404	0.199	KNOL	4.470	-2.686	0.849
ADBE	0.544	0.806	-2.710	CSCO	0.463	0.972	-5.551	KR	1.212	1.373	-1.327
AGN	0.406	0.628	-0.924	CSE	4.439	1.859	2.384	LANC	0.795	0.959	-0.323
AINV	1.980	1.996	-0.029	CSL	0.589	1.267	-0.859	LECO	1.063	0.704	0.503
AMAT	1.036	1.643	-3.242	CTSH	0.323	0.439	-1.876	LPNT	1.178	1.045	0.346
AMED	2.878	1.178	2.017	DCOM	-0.401	2.104	-1.088	LSTR	0.231	0.487	-1.385
AMGN	0.330	0.446	-2.629	DELL	0.876	1.353	-2.134	MAKO	0.575	-0.689	0.299
AMZN	0.316	0.327	-0.425	DIS	0.515	1.000	-6.312	MANT	1.620	0.141	1.202
APOG	2.831	2.410	0.225	DOW	0.381	1.244	-4.292	MDCO	2.794	3.065	-0.222
ARCC	1.549	1.146	0.584	EBAY	0.980	1.124	-1.139	MELI	1.517	1.711	-0.397
AXP	0.412	0.939	-4.882	ERIE	3.082	-0.272	1.127	MFB	9.192	0.764	1.007
AYI	-0.248	0.027	-0.192	EWBC	1.263	1.187	0.142	MIG	4.118	2.886	0.294
BHI	0.252	0.430	-1.390	FCN	0.697	0.606	0.190	MMM	0.242	0.574	-4.707
BIIB	0.336	0.794	-5.663	FFIC	7.920	-1.203	2.170	MOD	8.375	-0.361	1.840
BRCM	0.570	0.816	-2.679	FL	2.026	1.740	0.465	MOS	0.662	1.070	-1.918
BRE	0.340	0.900	-1.653	FMER	0.833	1.546	-1.708	MRTN	3.987	5.396	-0.203
BW	8.531	0.506	2.836	FPO	0.690	3.215	-0.231	MXWL	2.128	1.230	0.751
BXS	0.622	0.615	0.006	FRED	0.272	0.430	-0.133	NSR	0.253	0.361	-0.279
BZ	2.563	2.157	0.170	FULT	1.986	1.900	0.278	NUS	3.005	1.828	0.888
CB	0.444	0.656	-1.511	GAS	0.677	0.886	-0.387	NXTM	-1.210	0.882	-0.082
CBEY	1.658	1.424	0.158	GE	0.713	1.488	-8.903	PBH	-4.313	1.479	-1.114
CBT	0.900	1.134	-0.312	GENZ	0.429	0.460	-0.440	PFE	0.954	1.407	-3.268
CCO	3.363	6.504	-0.608	GILD	0.479	0.481	-0.051	PG	0.271	0.539	-6.687
CDR	1.255	2.984	-0.410	GLW	1.113	1.401	-1.597	PNC	0.328	0.495	-1.442
CELG	0.434	0.488	-0.436	GOOG	-0.386	1.755	-1.461	PNY	1.127	0.321	0.596
CETV	1.761	1.456	0.342	GPS	1.006	1.210	-1.453	PTP	0.467	0.430	0.047
CKH	0.834	0.357	1.667	HON	0.173	0.891	-6.481	RIGL	2.280	0.957	0.885
CMCSA	0.774	1.275	-2.52	HPQ	0.349	0.588	-3.771	ROC	1.554	1.415	0.136
CNQR	0.618	1.541	-1.445	IMGN	2.971	3.499	-0.667	ROCK	2.464	1.455	0.740
COO	1.060	0.666	0.449	INTC	0.640	1.071	-3.330	SF	1.064	0.156	0.711
COST	0.371	0.482	-2.159	ISIL	1.113	1.775	-3.536	SFG	-0.190	0.947	-1.800
CPSI	-0.291	0.697	-0.345	ISRG	0.730	1.095	-1.680	SWN	0.676	0.786	-1.000
Overall	1.893	1.490	1.000								

Table 20: Exogenous HFT - Volatility Relationship. This table shows the results of an exogenous removal of HFT and its impact on volatility. It uses the short sale ban as the source of exogenous shock. 13 firms are impacted. I run the following OLS regression: $\Delta Vola_{i,t} = HFT\%Change_{i,t} * \beta_1 + \epsilon_{i,t}$, where $\Delta Vola$ is the percent change in volatility for firm i between the pre- and post- ban period after differencing out the change in its comparable control firm, $\frac{Vol_{post}-Vol_{pre}}{Vol_{pre}}$. $HFT\%Change_{i,t}$ is the change in HFT activity pre- and post- ban after differencing out the change in its comparable control firm, $\frac{HFT_{post}-HFT_{pre}}{HFT_{pre}}$. Column (1) shows the results using the day before and day after data; column (3) shows the results using the average values from the week before and week after for pre and post data points. Columns (2) and (4) utilize a non-parametric bootstrap looping through the data 50 times (and using replacement).

	(1) 1 Day	(2) 1 Day - Bootstrap	(3) 1 Week	(4) 1 Week - Bootstrap
HFT % Change	-0.171	-0.171	-0.716	-0.716
	(0.411)	(0.332)	(0.338)	(0.331)*
Constant	0.182	0.182	-0.00825	-0.00825
	(0.178)	(0.151)	(0.103)	(0.0924)
Observations	13	13	13	13
Adjusted R^2	-0.074	-0.074	0.226	0.226

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

Table 21: HFT Impact on Volatility - No Demand of Liquidity. This table looks at the impact of HFT on volatility. I sum the one minute realized volatility and compare its actual value with what it would be if HFT trading and liquidity had not occurred. To calculate the alternative price path I leave prices untouched and trim out the HFTr trades, assuming that the prices would have achieved their actual levels but would simply jump around more (as there would be no HFT initiated trades). 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. That is, if they are increasing volatility, then they are buying at the peaks and selling at the troughs, by removing them I am leveling out the price path. If they have no impact or are decreasing volatility then removing the HFT initiated trades will either leave volatility unaffected, or will increase it as the previous HFT buy (sell) at a low (high) will be replaced in the realized volatility return by a non-HFT buy (sell) at a higher (lower) level.

Firm	H RV	no H RV	T-Stat	Firm	H RV	no H RV	T-Stat	Firm	H RV	no H RV	T-Stat
AA	0.235	0.243	-0.741	CPWR	0.193	0.197	-0.401	JKHY	0.117	0.119	-0.330
AAPL	0.154	0.154	0.031	CR	0.117	0.121	-1.073	KMB	0.118	0.121	-0.579
ABD	0.210	0.209	0.033	CRI	0.164	0.163	0.025	KNOL	0.159	0.161	-0.138
ADBE	0.156	0.154	0.331	CRVL	0.187	0.189	-0.187	KR	0.130	0.128	0.108
AGN	0.138	0.137	0.221	CSCO	0.158	0.158	0.206	KTII	0.003	0.003	0.000
AINV	0.181	0.185	-0.656	CSE	0.270	0.262	0.185	LANC	0.101	0.102	-0.924
AMAT	0.207	0.208	-0.064	CSL	0.127	0.128	-0.156	LECO	0.205	0.210	-0.446
AMED	0.269	0.268	0.087	CTRN	0.101	0.101	-0.087	LPNT	0.187	0.188	-0.181
AMGN	0.123	0.125	-2.543	CTSH	0.142	0.142	-0.052	LSTR	0.143	0.144	-0.134
AMZN	0.200	0.199	0.181	DCOM	0.100	0.103	-0.253	MAKO	0.140	0.142	-0.140
ANGO	0.083	0.082	0.031	DELL	0.164	0.165	-0.257	MANT	0.114	0.116	-0.152
APOG	0.149	0.151	-0.315	DIS	0.141	0.143	-0.315	MDCO	0.246	0.239	0.338
ARCC	0.175	0.175	0.052	DK	0.076	0.076	0.000	MELI	0.347	0.346	0.082
AXP	0.178	0.183	-0.739	DOW	0.281	0.280	0.150	MFB	0.102	0.103	-0.326
AYI	0.098	0.100	-0.424	EBAY	0.194	0.197	-0.643	MIG	0.088	0.089	-0.186
AZZ	0.126	0.128	-0.532	EBF	0.137	0.138	-0.295	MMM	0.147	0.147	0.059
BARE	0.012	0.012	0.088	ERIE	0.096	0.100	-0.356	MOD	0.238	0.241	-0.148
BAS	0.253	0.253	-0.002	ESRX	0.228	0.227	0.118	MOS	0.296	0.298	-0.179
BHI	0.212	0.214	-0.416	EWBC	0.260	0.256	0.403	MRTN	0.101	0.100	0.137
BIIB	0.160	0.161	-0.841	FCN	0.193	0.194	-0.048	MXWL	0.186	0.186	0.001
BRCM	0.181	0.182	-0.304	FFIC	0.117	0.117	0.000	NC	0.081	0.081	0.161
BRE	0.144	0.142	0.337	FL	0.141	0.141	0.020	NSR	0.088	0.090	-0.458
BW	0.210	0.212	-0.402	FMER	0.143	0.146	-0.465	NUS	0.133	0.135	-0.444
BXS	0.179	0.183	-0.103	FPO	0.110	0.111	-0.075	NXTM	0.323	0.321	0.068
BZ	0.296	0.294	0.069	FRED	0.103	0.102	0.343	PBH	0.116	0.117	-0.060
CB	0.105	0.105	0.091	FULT	0.194	0.195	-0.125	PFE	0.176	0.174	0.295
CBEY	0.167	0.165	0.218	GAS	0.126	0.129	-0.294	PG	0.120	0.118	0.860
CBT	0.209	0.211	-0.295	GE	0.174	0.172	0.348	PNC	0.211	0.208	0.556
CBZ	0.076	0.078	-0.263	GENZ	0.149	0.150	-1.153	PNY	0.092	0.093	-0.070
CCO	0.174	0.171	0.193	GILD	0.145	0.145	-0.024	PPD	0.085	0.086	-0.095
CDR	0.146	0.146	-0.041	GLW	0.208	0.203	0.873	PTP	0.070	0.070	-0.023
CELG	0.188	0.189	-0.464	GOOG	0.139	0.140	-0.043	RIGL	0.225	0.227	-0.096
CETV	0.259	0.261	-0.059	GPS	0.147	0.147	0.005	ROC	0.237	0.237	0.042
CHTT	0.001	0.001	0.000	HON	0.172	0.172	-0.091	ROCK	0.390	0.391	-0.020
CKH	0.110	0.113	-0.333	HPQ	0.128	0.131	-1.272	ROG	0.102	0.102	-0.019
CMCSA	0.184	0.181	0.287	IMGN	0.197	0.198	-0.105	RVI	0.119	0.122	-0.372
CNQR	0.150	0.155	-0.560	INTC	0.173	0.171	0.386	SF	0.068	0.068	-0.019
COO	0.150	0.150	0.507	IPAR	0.105	0.105	0.037	SFG	0.128	0.132	-0.683
COST	0.110	0.108	0.531	ISIL	0.163	0.163	0.061	SJW	0.097	0.099	-0.149
CPSI	0.122	0.123	-0.145	ISRG	0.155	0.157	-0.650	SWN	0.288	0.282	0.493
Overall	0.159	0.160	-1.99								