HIGH FREQUENCY TRADING AND ITS IMPACT ON MARKET QUALITY

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

July 16, 2010

*I would like to thank my advisors, Tom Brennan, Robert Korajczyk, Robert McDonald, Annette Vissing-Jorgensen for the considerable amount of time and energy they spent discussing this topic with me. I would like to thank Nasdaq OMX for making available the data used in this project. Also, I would like to thank 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 equities markets. 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 efficiency, and volatility. I find that in my sample HFTs participate in 77% of all trades and that they tend to engage in a price-reversal strategy. I find no evidence suggesting HFT withdraw from markets in bad times or that they engage in abnormal front-running of large non-HFT trades. The 26 HFT firms in the sample earn approximately $3 billion in profits annually. HFTs demand liquidity for 50.4% of all trades and supply liquidity for 51.4% of all trades. HFTs tend to demand liquidity in smaller amounts, and trades before and after a HFT demanded trade occur more quickly than other trades. HFTs provide the inside quotes approximately 50% of the time. In addition if HFTs were not part of the market, the average trade of 100 shares would result in a price movement of $.013 more than it currently does, while a trade of 1000 shares would cause the price to move an additional $.056. HFTs are an integral part of the price discovery process and price efficiency. Utilizing a variety of measures introduced by Hasbrouck (1991a, 1991b, 1995), I show that HFT trades and quotes contribute more to price discovery than do non-HFT activity. Finally, HFT reduces volatility. By constructing a hypothetical alternative price path that removes HFTs from the market, I show that the volatility of stocks is roughly unchanged when HFT initiated trades are eliminated and significantly higher when all types of HFT trades are removed.

1 Introduction

1.1 Motivation

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". The next day, some media blamed high frequency traders (HFTs; HFT is also used to refer to high frequency trading) for driving the market down (Krudy, June 10, 2010). Others in the media blamed the temporary withdrawal of HFTs from the market as causing the precipitous fall (Creswell, May 16, 2010).[1] HFTs have come to make up a large portion of the U.S. equity markets, yet the evidence of their role in the financial markets has come from news articles and anecdotal stories. The SEC has also been interested in the issue. It 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 (Securities and Commission, January 14, 2010).

In addition, the Dodd Frank Wall Street Reform and Consumer Protection Act calls for an in depth study on HFT (Section 967(2)(D)). In this paper I examine the empirical consequences of HFT on market functionality. I utilize a unique dataset from Nasdaq OMX that distinguishes HFT from non-HFT quotes and trades. This paper provides an analysis of HFT behavior and its impact on financial markets. Such an analysis is necessary since to ensure properly functioning financial markets the SEC and exchanges must set appropriate rules for traders. These rules should be based on the actual behavior of actors and not on hearsay and anecdotal stories. It is equally important that institutional and retail investors understand whether or not they are being manipulated or exploited by sophisticated traders, such as HFT.

This paper studies HFT from a variety of viewpoints. First, it describes the activities of HFTs, showing that HFTs make up a large percent of all trading and that they both provide liquidity and demand liquidity. Their activities tend to be stable over time. Second, it examines HFT strategy and profitability. HFTs generally engage in a price reversal strategy, buying after price declines and selling after price gains. They are profitable, making around $3 billion each year on trading volume of $30 trillion dollars traded. Third, it considers the impact of HFTs on the market, focusing on three areas - liquidity, price discovery, and volatility. HFTs increase market liquidity: using a variety of Hasbrouck measures, I find that HFTs appear to add to the efficiency of the markets. Finally, I find that HFTs tend to decrease volatility.

[1] To date, the true cause of the flash crash has not been determined.

Given these results, HFT appear to be a new form of market makers. HFTs appear to make markets operate better (i.e. increase liquidity and price efficiency, and reduce volatility) for all market participants.

HFT is a recent phenomenon. Tradebot, a large player in the field who frequently makes up over 5% of all trading activity, states that the strategy of HFT has only been around for the last ten years (starting in 1999). Whereas only recently an average trade on the NYSE took ten seconds to execute, (Hendershott and Moulton, 2007), now some firms entire trading 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.

1.2 Definitions

To date, there lacks a clear definition for many of the terms in rapid trading and in computer controlled trading. Even the Securities and Exchange recognizes this and says that high frequency trading "does not have a settled definition and may encompass a variety of strategies in addition to passive market making" (Securities and Commission, January 14, 2010). High frequency trading is a type of strategy that is engaged in buying and selling shares rapidly, often in terms of milliseconds and seconds. This paper takes the definition from the SEC: HFT refers to, "professional traders acting in a proprietary capacity that engages in strategies that generate a large number of trades on a daily basis" (Securities and Commission, January 14, 2010). By some estimates HFT makes up over 50% of the total volume on equity markets daily (Securities and Commission, January 14, 2010; Spicer, December 2, 2009).

Other terms of interest when discussing HFT include "pinging" and "algorithmic trading."

The SEC defines pinging as, "an immediate-or-cancel order that can be used to search for and access all types of undisplayed liquidity, including dark pools and undisplayed order types at exchanges and ECNs. The trading center that receives an immediate-or-cancel order will execute the order immediately if it has available liquidity at or better than the limit price of the order and otherwise will immediately respond to the order with a cancelation" (Securities and Commission, January 14, 2010). The SEC goes on to clarify, "[T]here is an important distinction between using tools such as pinging orders as part of a normal search for liquidity with which to trade and using such tools to detect and trade in front of large trading interest as part of an 'order anticipation' trading strategy" (Securities and Commission, January 14, 2010).

A type of trading that is similar to HFT, but fundamentally different is algorithmic trading. Algorithmic Trading 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). Algorithmic and HFT are similar in that they both use automatic computer generated decision making technology. However, they differ in that algorithmic trading 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 and to close the trading day in a neutral position. Thus, HFT must be a type of algorithmic trading, but algorithmic trading need not be HFT.

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. The only academic paper regarding HFT is one by Kearns, Kulesza, and Nevmyvaka (2010), and this paper shows that the maximum amount of profitability that HFT can make based on TAQ data under the implausible 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. Although my research is the first to look at the impact of HFT on the stock market, it touches on a variety of related fields of research, the most relevant being algorithmic trading.

2.1 Algorithmic Trading

In principal algorithmic trading 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. Nonetheless several papers have studied algorithmic trading (AT).

Hendershott and Riordan (2009) use data from the firms listed on the Deutsche Boerse DAX. 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 does human trading. 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 between 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. 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.

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 decomposing the AT type traders into short-horizon traders and others and focusing on the impact of the short-horizon traders on market quality.

2.2 Theory

There is an extant literature in theoretical asset pricing. Of these papers only a handful try to understand what the impact on market quality will be of having investors with different investment time horizons. Two papers directly address the scenario when there are short and long term investors in a market: "Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation" (Froot, Scharfstein, and Stein, 1992); and "Short-Term Investment and the Informational Efficiency of the Market" (Vives, 1995).

Froot, Scharfstein, and Stein (1992) find that short-term speculators may put too much emphasis on some (short term) information and not enough on fundamentals. The result being a decrease in the informational quality of asset prices. Although the paper does not extend its model in the following direction, a decrease in the informational quality suggests a decrease in price efficiency and an increase in volatility.

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 suggest that HFT may be beneficial to market quality or may be harmful to it.

3 Data

3.1 Standard Data

The data in this paper comes from a variety of sources. It uses in standard fashion CRSP data when considering daily data not included in the Nasdaq dataset. Compustat data is used to incorporate firm characteristics in the analysis.

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, excluding trades that occurred at the opening, closing, and during intraday crosses. The trade date 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 (HFT or not) is providing or taking liquidity.

The Quote data is from February 22, 2010 to February 26, 2010. It includes the best bid and ask that is being offered by HFT firms and by non-HFT firms 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 is provided by a HFT or a non-HFT, and whether the liquidity was displayed or hidden.

The Nasdaq dataset consists of 26 traders that have been identified 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 HFT firms that are identified are the following: They engage in proprietary trading; that is, the firm does not have customers but instead trades its own capital. The HFT use sophisticated trading tools such as high-powered analytics and computing co-location services located near exchanges to reduce latency. The HFT engage in sponsored access providers whereby they have access to the co-location services and can obtain large-volume discounts. HFT tend to use OUCH protocol whereas non-HFT tend to use RASH.

The HFT firms tend to switch between long and short net positions several times throughout the day, whereas non-HFT labeled 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 for non-HFT firms.

Firms that others may define as HFT are not labeled as HFT firms here if they satisfy one of the following: firms like Lime Brokerage and Swift Trade who provide direct market access and other powerful trading tools to its customers, who are likely engaging in HFT and thus are likely HFT traders but are not labeled so; proprietary trading firms that are a desk of a larger, integrated firm, like Goldman Sachs or JP Morgan; an independent firm that is engaged in HFT activities, but who routes its trades through a MPID of a non-HFT type firm; firms that engage in HFT activities but because they are small are not considered in the study as being labeled a HFT firm.

The data is for a sample of 120 Nasdaq stocks where the ticker symbols are listed in Table 1. These sample stocks were selected by a group of academics. The stocks consist of a varying degree of market capitalization, market-to-book ratios, industries, and listing venues.

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 show they are typical stocks, I then compare the firm characteristics to the Compustat database and show they are on average larger firms, but otherwise a relatively close match to an average Compustat firm. Finally, I provide general statistics on the percent of the market trades in which HFT are involved, considering all types of trades, supplying liquidity trades, and demanding liquidity trades.

Table 2 describes the 120 stocks in the Nasdaq sample data set. These statistics are taken for the five trading days from February 22 to February 26, 2010. This table shows that these stocks are quite average and provide a reasonable subsample of the market. The price of the stocks is on average 39.57 and ranges between 4.6 and 544. The daily trading volume on Nasdaq for these stocks averages 1.064 million shares, and ranges from as small as 2,000 shares to 14 million shares. This is done on average over 5,150 trades, whereas some stock trade just 8 times on a given day while others trade as many as 59,799 times. The 120 stocks are quite liquid. Quoted half-spreads are calculated when trades occur. the average

Table 1: List of Stocks

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

quoted half-spread of 1.82 cents is comparable to large and liquid stocks in other markets. The average trade size, in shares is 139.6. The average depth of the inside bid and ask, measured by summing the depth at the bid and at the ask times their respective prices, dividing by two and taking the average per day, is $71,550.

Table 2: Summary Statistics. Summary statistics for the HFT dataset from February 22, 2010 to February 26, 2010.

Variable	Mean	Std. Dev.	Min.	Max.
Price	39.573	60.336	4.628	544.046
Daily Trading Volume (Millions)	1.064	2.137	0.002	14.857
Daily Number of Trades per Day	5150.983	7591.812	8	59799
Quoted Half Spread (cents)	1.838	4.956	0.5	42.5
Trade Size	139.617	107.631	37	1597
Depth (Thousand Dollars)	71.55	196.421	1.161	2027.506
N	600			

Table 3 describes the 120 stocks in the HFT database compared to the Compustat database. The table shows that the HFT database is on average larger then the average Compustat firm. The Compustat firms consist of all firms in the Compustat database with data available and that have a market capitalization of greater than $10 million in 2009. The Compustat database statistics include the firms that are found in the HFT database. The data for both the Compustat and the HFT firms are for firms 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 $2.6 billion, the average HFT database firm has a market capitalization of $17.59 billion. However the sample does span a large size variation of firms, from the very small with a market capitalization of only $80 million, to the the very large with market capitalization of $175.9 billion. Compustat includes many very small firms that reduce the mean market capitalization, making the HFT sample be overweighted with larger firms. The market-to-book ratio also differs between Compustat and the HFT sample. Whereas HFT have a mean market-to-book of 2.65, the Compustat data has one of 10.9. Based on industry, the HFT sample is a relatively close match to the Compustat database. The industries are determined based on the Fama-French 10 industry designation from SIC identifiers. The HFT database tends to overweight Manufacturing, Telecommunications, Healthcare, and underweight Energy and Other. The HFT firms are all listed on the NYSE or

Nasdaq exchange, with half of the firms listed on each exchange. Whereas about one-third of Compustat firms are listed on other exchanges. [2] The HFT database provides a robust variety of industries, market capitalization, and market-to-book values.

Table 4 looks at the prevalence of HFT in the stock market. It captures this in a variety of ways: the number of trades, shares, and dollar-volume that have a HFT involved compared to trades where no HFT participates. The table provides summary statistics for the involvement of HFT traders in the market. Three different statistics are calculated for each split of the data. The column "Trades" reports the number of trades, "Shares" reports the number of shares traded, and "Dollar" reports the dollar value of those shares traded. Panel A - HFT Involved In Any Trade splits the data based on whether a HFT was involved in any way in a trade or not. The results show that HFT make up over 77% of all trades. HFT tend to trade in smaller shares as per-share traded they make up just under 75%. Finally, based on a dollar-volume basis of trade, they make up 73.8% of the trading volume.

The next two panels separate HFT transactions into what side of the trade they are on based on liquidity. Panel B - HFT Involved As Liquidity Taker groups trades into HFT only when the HFT is demanding the liquidity in the transaction. HFT takes liquidity in 50.4% of all trades, worth a dollar amount of just about the same percentage of all transactions on a dollar basis. They make up only 47.6% of shares trading, suggesting they provide liquidity in stocks that are slightly higher priced.

Panel C - HFT Involved As Liquidity Supplier groups trades into HFT only when the HFT is supplying the liquidity in the transaction. The amount of liquidity supplied is only slightly more than that demanded by HFT at 51.4% of all trades having a liquidity supplier being a HFT. Based on number of shares this value falls to 50.8% of all shares traded; and based on dollar-volume, it drops to 45.5% of all trades.

5 HFT Strategy

Before analyzing HFT impact on market quality, it is insightful to understand more about what drives HFT activity. To research this, I use an ordered logit regression to show their trading strategy is heavily dependent on past returns. I further identify that they engage in a price reversal strategy, whereby they tend to

[2] Comparing Compustat firms that are listed on NYSE or Nasdaq reduces the number of firms to 5050 with an average market capitalization of $3.46 billion, and with the industries more closely matching those in the HFT dataset.

Table 3: HFT Sample v. Compustat. This table compares the HFT-identified dataset with the Compustat dataset. The Compustat data consists of all firms in the Compustat database with a market capitalization of $10 million or more. The industries are categorized based on the Fama-French 10 industry groups.

	HFT Dataset				**Compustat Dataset**			
	mean	sd	min	max	mean	sd	min	max
Market Cap. (millions)	17588.24	37852.38	80.602	197012.3	2613.01	12057.34	10.001	322334.1
Market-to-Book	2.65	3.134	-11.779	20.040	10.919	598.126	-2489.894	44843.56
Industry - Non-Durables	.0333	.180			.034	.181		
Industry - Durables	.025	.156			.014	.120		
Industry - Manufacturing	.1667	.374			.071	.257		
Industry - Energy	.0083	.091			.049	.217		
Industry - High Tech	.1583	.366			.124	.330		
Industry - Telecom.	.05	.218			.024	.153		
Industry - Wholesale	.0917	.289			.058	.235		
Industry - Health Care	.15	.358			.080	.272		
Industry - Utilities	.0333	.180			.034	.183		
Industry - Other	.2833	.452			.509	.499		
Exchange - NYSE	.5	.502			.288	.453		
Exchange - Nasdaq	.5	.502			.322	.467		
Exchange - Other	0	0			.388	.487		
Observations	120				8260			

Table 4: HFT Aggregate Activity. The table provides summary statistics for the involvement of HFT traders in the market. Three different statistics are calculated for each split of the data. The column Trades reports the number of trades, Shares reports the number of shares traded, and Dollar reports the dollar value of those shares traded. Panel A - HFT Involved In Any Trade splits the data based on whether a HFT was involved in any way in a trade or not. Panel B - HFT Involved As Liquidity Taker groups trades into HFT only when the HFT is demanding the liquidity in the transaction. Panel C - HFT Involved As Liquidity Supplier groups trades into HFT only when the HFT is supplying the liquidity in the transaction.

Panel A - HFT Involved In Any Trade

Type of Trader	Trades (Sum)	Trades (%)	Shares (Sum)	Shares (%)	Dollar (Sum)	Dollar (%)
HFT	2,387,851	77.3%	477,944,435	74.9%	$19,427,424,121	73.8%
Non HFT	702,739	22.7%	160,337,476	25.1%	$6,879,817,754	26.2%
Total	3,090,590	100.0%	638,281,911	100.0%	$26,307,241,875	100.0%

Panel B - HFT Involved As Liquidity Taker

HFT	1,556,766	50.4%	303,971,478	47.6%	$13,169,044,493	50.1%
Non HFT	1,533,824	49.6%	334,310,433	52.4%	$13,138,197,383	49.9%
Total	3,090,590	100.0%	638,281,911	100.0%	$26,307,241,875	100.0%

Panel C - HFT Involved As Liquidity Supplier

HFT	1,588,157	51.4%	324,221,557	50.8%	$11,959,264,046	45.5%
Non HFT	1,502,433	48.6%	314,060,354	49.2%	$14,347,977,829	54.5%
Total	3,090,590	100.0%	638,281,911	100.0%	$26,307,241,875	100.0%

buy stocks at short-term troughs and they tend to sell stocks at short-term peaks. This is true regardless of whether they are supplying or demanding liquidity. Also, HFT tend to trade in larger, value firms, with lower volume and lower spreads and depth. Finally, based on their trading activities at the aggregate level I estimate they earn approximately $3 billion a year.

5.1 Investment Strategy

HFT do not readily share their trading strategies. However, the anecdotal stories of HFT firms suggest they have essentially replaced the role of market makers by providing liquidity and a continuous market into which other investors can trade.

What is known regarding HFT is that they tend to buy and sell in very short time periods. Therefore, rather than changes in firm fundamentals, HFT firms must be basing their decision to buy and sell from short term signals such as stock price movements, spreads, or volume.

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 HFT firm 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.

Before getting to the ordered logit, I summarize the theoretical reason for why an ordered logit is appropriate in this setting, as first discussed by Hausman, Lo, and MacKinlay (1992).

HFT trading behavior consist of a sequence of actions $Z(t_1), Z(t_2), \ldots, Z(t_\eta)$ observed at regular time intervals $t_0, t_1, t_2, \ldots, t_\eta$. Let Z_k^* be an unobservable continuous random variable where

$$Z_k^* = X_k'\beta + \varepsilon_k,\ E[\varepsilon_k|X_k] = 0,\ \varepsilon_k\ i.n.i.d.\ N(0, \sigma_k^2) \tag{1}$$

where 'i.n.i.d.' stands for the assumption that the ε_k's are independent but not identically distributed, and X_k is a $q \times 1$ vector of predetermined variables that sets the conditional mean of Z_k^*. Whereas Hausman, Lo, and MacKinlay (1992) deal with tick by tick stock price data, the scenario in this paper deals with HFT trade behavior data that is aggregated into ten second intervals. Therefore, the subscripts are used to denote ten second period, not transaction time.

The essence of the ordered logit model is the assumption that observed HFT behavior Z_k are related to the continuous variable Z_k^* in the following mapping:

$$Z_k = \begin{cases} s_1 & \text{if} \quad Z_k^* \in A_1, \\ s_2 & \text{if} \quad Z_k^* \in A_2, \\ \vdots & \vdots \\ s_m & \text{if} \quad Z_k^* \in A_m, \end{cases}$$

where the sets A_j form a partition of the state space ζ^* of Z_k^*. The partition will have the properties that $\zeta^* = \bigcup_{j=1}^{m} A_j$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, and the s_j's are the discrete values that comprise the state space ζ of Z_k. The ordered logit specification allows an investigator to understand the link between ζ^* and ζ and relate it to a set of economic variables used as explanatory variables that can be used to understand the HFT trading strategy. In this application the s_j's are Sell, Do Nothing, Buy. Note, the observable actions could also be split into size, for example, Sell 1000 + shares, Sell 500 - 1000, etc., but I restrict the ζ partition to these three natural breaks. The alternative fine tuned separation, for instance, by subdividing the buys and selling into the number of shares exchanged, is beyond the needs of this analysis.

I assume the error terms in ε_k's in equation 1 are conditionally independently, but not identically, distributed, conditioned on the X_k's and the other explanatory variables, W_k, that are omitted from equation 1, which allows for heteroscedasticity in σ_k^2.

The conditional distribution of observed return changes Z_k, conditioned on X_k and W_k, is determined by the partition boundaries calculated from the ordered logit regression. As stated in Hausman, Lo, and MacKinlay (1992), for a Gaussian ε_k, the conditional distribution is

$$\begin{aligned} & P(Z_k = s_i | X_k, W_k) \\ & = P(X_k^{'}\beta + \varepsilon_k \in A_i | X_k, W_k) \end{aligned}$$

$$= \begin{cases} P(X_k^{'}\beta + \varepsilon_k \leq \alpha_1 | X_k, W_k) & \text{if } i = 1 \\ P(\alpha_{i-1} < X_k^{'}\beta + \varepsilon_k \leq \alpha_i | X_k, W_k) & \text{if } 1 < i < m, \\ P(\alpha_{m-1} < X_k^{'}\beta + \varepsilon_k | X_k, W_k) & \text{if } i = m, \end{cases} \tag{2}$$

$$= \begin{cases} \Phi(\frac{\alpha_1 - X_k^{'}\beta}{\sigma_k}) & \text{if } i = 1 \\ \Phi(\frac{\alpha_i - X_k^{'}\beta}{\sigma_k}) - \Phi(\frac{\alpha_{i-1} - X_k^{'}\beta}{\sigma_k}) & \text{if } 1 < i < m, \\ 1 - \Phi(\frac{\alpha_{m-1} - X_k^{'}\beta}{\sigma_k}) & \text{if } i = m, \end{cases} \tag{3}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

The intuition for the ordered logit model is that the probability of the type of behavior by the HFT is determined by where the conditional mean lies relative to the partition boundaries. Therefore, for a given conditional mean $X_k^{'}\beta$, shifting the boundaries will alter the probabilities of observing each state, Sell, Do Nothing, or Buy. The order of the outcomes could be reversed with no real consequence except for the coefficients changing signs as the ordered logit only takes advantage of the fact there is some natural ordering of the events. The explanatory variables then allow one to analyze the different effects of relevant economic variables to understand HFT behavior . As the data determines where the partition boundaries the ordered logit model creates an empirical mapping between the unobservable ζ^* state space and the observable ζ state space. Here, the empirical relationship between HFT behavior can be analyzed with respect to the economic variables X_k and W_k.

I divide the time frames in to ten second intervals throughout the trading day. [3] 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 and its associated beta coefficient 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 contempo-

[3] 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.

raneous 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 HFT were on net selling shares for stock i, it is zero if the HFT performed no transaction or its buys and sell exactly canceled, and it is 1 if on net HFT were buying shares for stock i.

From this ordered logit model one may expect to see a variety of potential patterns. A handful of different strategies have been suggested in which HFT 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 srpead and so the $Spread$ variables would have a lot of power in explaining when HFT 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 5 shows the results.

The results reported in table 5 are the marginal effects at the mean for the ordered probit. From the ordered logit regression's summarized results in table 5, 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 HFT will be selling (and with low past returns and the likelihood the HFT will be buying). There is some statistical significance in other locations, however no where is it consistent like that of the return coefficients. This suggests that past spread size, depth, and volume are not primary factors in HFT trading decisions. 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 HFT's trading behavior.

It appears that HTF engage in a price reversal strategy. To analyze this further, I analyze the HFT buy and sell logits separately, focusing on the lagged returns surrounding a HFT firm's buying or selling stocks, analyzing the differences in

Table 5: HFT Ordered Logit - Exploratory Regression. This table includes several explanatory variables in order to uncover which HFT strategies are evidenced within the data. 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

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

demanding versus supplying liquidity.

To better understand the HFT trading strategy I run logit regressions on different dependent variables. I consider a total of six different regressions: HFT selling, HFT selling when supplying liquidity, HFT selling when demanding liquidity, HFT buying, HFT buying when supplying liquidity, and HFT buying when demanding liquidity. The results found in tables 6 and 7 are the marginal effects at the mean. The Sell logit regressions are shown in table 6. The first column is the results for HFT Sell, all types. The results show the strong relationship between past returns and HFT decision to sell. prior to HFT 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 HFT in general engage in a price reversal strategy.

The next column has as the dependent variable a one if HFT 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 last column in table 6 has as the dependent variable a one if HFT 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 nineth one. The signs are the same as before, which is consistent with a price reversal strategy. One large difference is the fact that the contemporaneous period return coefficient is large and negative. It is not clear from the logit model whether this means that HFT initiate a sale once prices have started to fall, or that after they start selling prices fall. This cannot be determined from this regression as the contemporaneous return will include within its time period HFT transactions, but I cannot determine whether HFT were selling before prices fell or after they fell within this ten second increment.

The Buy regressions are shown in table 7. The first columns is the result for HFT Buy, all types. The results show the strong relationship between past returns and HFT decision to buy. Prior to HFT executing a purchase of a stock, the stock tend to fall, with statistically significance up to 100 seconds prior to the trade.

The next column has as the dependent variable a one if HFT 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. There is an especially large relationship with the contemporaneous period return and the HFT decision to supply liquidity and buy in a trade.

Table 6: Regressions of the Sell decision, split based on Liquidity Type. This table reports the results from running a logit with dependent variable equal to 1 if (1) HFT on net sell in a given ten second period, (2) HFT on net sell and supply liquidity, and (3) HFT on net sell and demand liquidity, and 0 otherwise. Firm fixed effects are used. The reported coefficients are the marginal effects at the mean.

	(1)	(2)	(3)
	HFT Sell - ALL	**HFT Sell - Supply**	**HFT Sell - Demand**
retlag0	4.925	16.35***	-16.48***
	(0.66)	(3.69)	(-8.38)
retlag1	5.145**	-0.934	7.594***
	(3.13)	(-0.96)	(9.80)
retlag2	5.230***	1.179	4.619***
	(3.87)	(1.44)	(5.54)
retlag3	6.521***	2.684***	4.276***
	(5.87)	(4.03)	(6.03)
retlag4	4.881***	1.234	4.209***
	(4.13)	(1.60)	(5.41)
retlag5	4.194***	2.038*	2.498***
	(3.82)	(2.46)	(3.45)
retlag6	5.098***	2.278***	2.989***
	(4.91)	(3.65)	(4.26)
retlag7	3.380***	0.497	3.439***
	(3.84)	(0.83)	(4.21)
retlag8	0.923	0.0277	1.087
	(0.99)	(0.04)	(1.68)
retlag9	2.521*	0.270	2.610**
	(2.53)	(0.43)	(3.14)
retlag10	0.351	-0.854	1.603*
	(0.39)	(-1.28)	(2.22)
N	1377798	1377798	1343177

Marginal effects; t statistics in parentheses

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

The last column in table 7 has as the dependent variable a one if HFT 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 0 and 3 - 6. The signs for the lag returns are negative as expected, except for the contemporaneous period return, which is large and positive. Like in the HFT Sell - Demand scenario, it is not clear from this logit model how to interpret this.

The results in table 6 and 7 show that HFT are engaged in a price reversal strategy. This is true whether they are supplying liquidity or demanding it.

5.1.1 Front Running

A potential investing strategy of which HFT have been claimed to be engaged in is front running. That is, the anecdotal evidence charges HFT with detecting when other market participants hope to move a large number of shares in a company and that the HFT enters into the same position just before the other market participant. It is in this context where the HFT pinging, as defined in the Definitions section, and the SEC's concern with it apply. That is, some claim HFT ping stock prices to detect large orders being executed. If they detect a large order coming through they may increase their trading activity. The result of such an action by the HFT would be to drive up the cost for the non-HFT market participant to execute the desired transaction.

To see whether or not this is occurring on a systematic basis I perform the following exercise: For each stock over the database time series I create twenty bins based on trade size for trades initiated by non-HFT. Each bin has roughly the same number of observations. Next, I look at the average percent of trades that were initiated by a HFT for different number of trades prior to the non-HFT initiated trade (for prior trades 1 - 10).

I graph the results in figure 5. The x-axis is the 20 different non-HFT initiated trade size bins; the y-axis is the fraction of trades for different non-HFT trade size bins for different prior trade periods that were initiated by a HFT; the z-axis is the different prior trade periods.

The figure suggests front running by HFT before large orders is not systematically occurring. In fact, it appears that larger trades, relative to each stock, tend to be preceded by fewer HFT initiated trades. The non-HFT trades that are preceded by the highest number of HFT initiated trades are those that are small and those are of moderate size. Also, it is interesting that the immediately preceding trades tend to have fewer HFT initiated trades than those further out. As will be shown later, trades initiated by one type of market participant have a greater probability

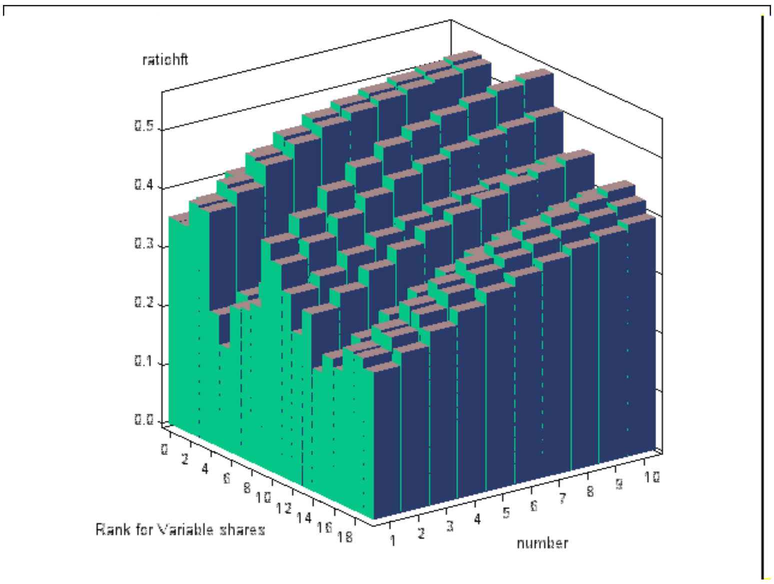
Table 7: Regressions of the Buy decision, split based on Liquidity Type. This table reports the results from running a logit with dependent variable equal to 1 if (1) HFT on net buy in a given ten second period, (2) HFT on net buy and supply liquidity, and (3) HFT on net buy and demand liquidity, and 0 otherwise. Firm fixed effects are used. The reported coefficients are the marginal effects at the mean.

	(1)	(2)	(3)
	HFT Buy - ALL	**HFT Buy - Supply**	**HFT Buy - Demand**
retlag0	-2.793	-48.10***	53.48***
	(-0.37)	(-14.21)	(18.93)
retlag1	-6.490***	-4.910***	-0.874
	(-3.87)	(-4.25)	(-0.69)
retlag2	-5.763***	-4.533***	-1.408
	(-4.44)	(-4.38)	(-1.80)
retlag3	-7.460***	-4.257***	-2.906**
	(-6.29)	(-6.80)	(-2.89)
retlag4	-6.291***	-2.802***	-3.202***
	(-5.25)	(-3.75)	(-3.84)
retlag5	-6.384***	-2.572***	-3.023***
	(-6.34)	(-3.32)	(-4.32)
retlag6	-6.110***	-3.042***	-2.766***
	(-6.14)	(-4.28)	(-3.85)
retlag7	-3.260**	-2.001**	-1.022
	(-3.13)	(-2.67)	(-1.45)
retlag8	-2.274*	-1.553*	-0.226
	(-2.43)	(-2.50)	(-0.28)
retlag9	-2.770**	-1.513*	-1.395*
	(-2.77)	(-2.08)	(-2.01)
retlag10	-2.049*	-1.908**	0.445
	(-2.26)	(-2.88)	(0.67)
N	1377798	1366278	1377798

Marginal effects; t statistics in parentheses

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

Figure 1: HFT Front Running. The graph shows the percent of trades initiated by HFT for different prior time periods that precede different size non-HFT initiated trades. The x-axis is the 20 different non-HFT initiated trade size bins; the y-axis is the fraction of trades for different non-HFT trade size bins for different prior trade periods that were initiated by a HFT; the z-axis is the different prior trade periods.



of being preceded by the same type of market participant.

5.1.2 HFT Market Activity

In addition to understanding the trading behavior of HFT at the trade by trade level, it is informative to understand what drives HFT to trade in certain stocks on certain days. Table 8 shows the variation in HFT market makeup in different stock on different days. Panel A is the percent of trading variation of non-HFT and HFT in a certain stock on a given day. Panel B is the percent of trading variation of HFT trading and non-HFT in supplying liquidity for a particular stock on a given day. Panel C is the percent of trading variation of HFT trading and non-HFT in demanding liquidity for a particular stock on a given day.

Panel A shows that HFT's share of the market varies a great deal depending on the stock and the day. Its percent of all trades varies from 10.8% to 93.6% based on number of trades. They average being involved in 61.8% of all trades, which compared to the numbers seen in the descriptive statistics from table 4, suggests that they trade more in stocks that trade frequently, as they make up 77% of all trades in the entire market.

Panel B looks at HFT supplying liquidity. HFT supply liquidity in 35.5% of trades in the average stock per day. This number is substantially smaller than the 50% they were found to supply in the market as a whole in table 4. Thus, HFT must supply liquidity in stocks that trade more frequently. Also, notice the wide variation in the supply of liquidity, in some stocks they provide no liquidity, while in others they supply 74%.

Panel C looks at HFT demanding liquidity. They demand liquidity in 39.6% of trades in the average stock per day. So HFT must be taking liquidity in stocks that trade more frequently. Also, the HFT demand for liquidity varies substantially ranging from 3.6% to 79.9%, but less than when they supply liquidity.

The results in table 8 show there is a large variation in the degree HFT trading in different stocks over time, the next step is to consider which determinants result in HFT increasing or decreasing their activity.

5.1.3 HFT Market Activity Determinants

Table 9 examines which determinants drive HFT trading. I perform an OLS regression, with the dependent variable being the percent of share volume, in which HFT were involved in for a given company on a given day. I run the following regression:

Table 8: Summary statistics 1 This table shows the variation in HFT market makeup. Panel A is the percent of trading variation of non-HFT and HFT in a certain stock on a given day. Panel B is the percent of trading variation of HFT trading and non-HFT in supplying liquidity for a particular stock on a given day. Panel C is the percent of trading variation of HFT trading and non-HFT in demanding liquidity for a particular stock on a given day.

	Trades					Shares				
Type of Trader	Mean	Median	Std. Dev.	Min	Max	Mean	Median	Std. Dev.	Min	Max
Panel A - HFT Involved In A Stock										
HFT	61.8%	64.0%	18.25	10.8%	93.6%	58.4%	59.4%	17.99	7.8%	90.9%
Non HFT	39.3%	37.1%	19.24	6.4%	92.2%	42.7%	41.6%	18.83	9.1%	93.1%
Total	100.0%	100.0%				100.0%	100.0%			
Panel B - HFT Involved In A Stock As Liquidity Supplier										
HFT	36.8%	35.5%	15.99	0%	74.4%	33.4%	32.7%	14.54	0.2%	66.4%
Non HFT	64.1%	65.3%	16.63	25.6%	100.0%	67.5%	67.9%	15.13	33.6%	100.0%
Total	100.0%	100.0%				100.0%	100.0%			
Panel C - HFT Involved In A Stock As Liquidity Taker										
HFT	39.6%	40.3%	16.43	3.6%	79.9%	37.8%	37.7%	16.49	2.6%	78.9%
Non HFT	61.1%	60.3%	16.70	20.1%	96.4%	62.8%	62.7%	16.73	21.1%	97.4%
Total	100.0%	100.0%				100.0%	100.0%			

$$\begin{aligned} H_{i,t} &= \alpha + MC_i * \beta_i + MB_i * \beta_i + NT_{i,t} * \beta_{i,t} + \\ &\quad NV_{i,t} * \beta_{i,t} + Dep_{i,t} * \beta i, t + Vol_{i,t} * \beta_{i,t} + AC_{i,t} * \beta i, t, \end{aligned}$$

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.

Table 9 reports the standardized regression coefficients. That is, instead of running the typical OLS regression on the regressors, the variables, both dependent and independent, are de-meaned, 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.

The results show that market capitalization is very important and has a positive relationship with HFT market percent. The market to book ratio is slightly statistically significant, but with a very small negative coefficient, suggesting HFT tend to slightly prefer value firms. Also statistically significant and with moderate economically significant is the dollar volume of non HFT trading, which is interpreted as HFT preferring to trade when there is less volume, all else being equal. The spread and depth variables are statistically significant and both have medium economically significance. HFT prefer to trade when there is less depth and lower spreads between bids and asks, all else being equal. Volatility, autocorrelation, and the number of non HFT trades are not statistically significant.

Table 9: Determinants of HFT Percent of the Market This table has as the dependent variable the percent (in dollar volume) of trades involving a HFT for a given stock on a given day.

	(1) Economic Impact
Market Cap.	0.722***
	(19.51)
Market / Book	-0.063*
	(-2.13)
$ of Non HFT Volume	-0.138***
	(-3.82)
Average Spread	-0.111***
	(-3.88)
Average Depth	-0.132***
	(-4.79)
Volatility	-0.031
	(-1.07)
Autocorrelation	-0.017
	(-0.62)
# of Non HFT Trades	0.042
	(0.98)
Constant	*
	(2.54)
Observations	590
Adjusted R^2	0.575

Standardized beta coefficients; t statistics in parentheses

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

5.1.4 HFT Market Activity Time Series

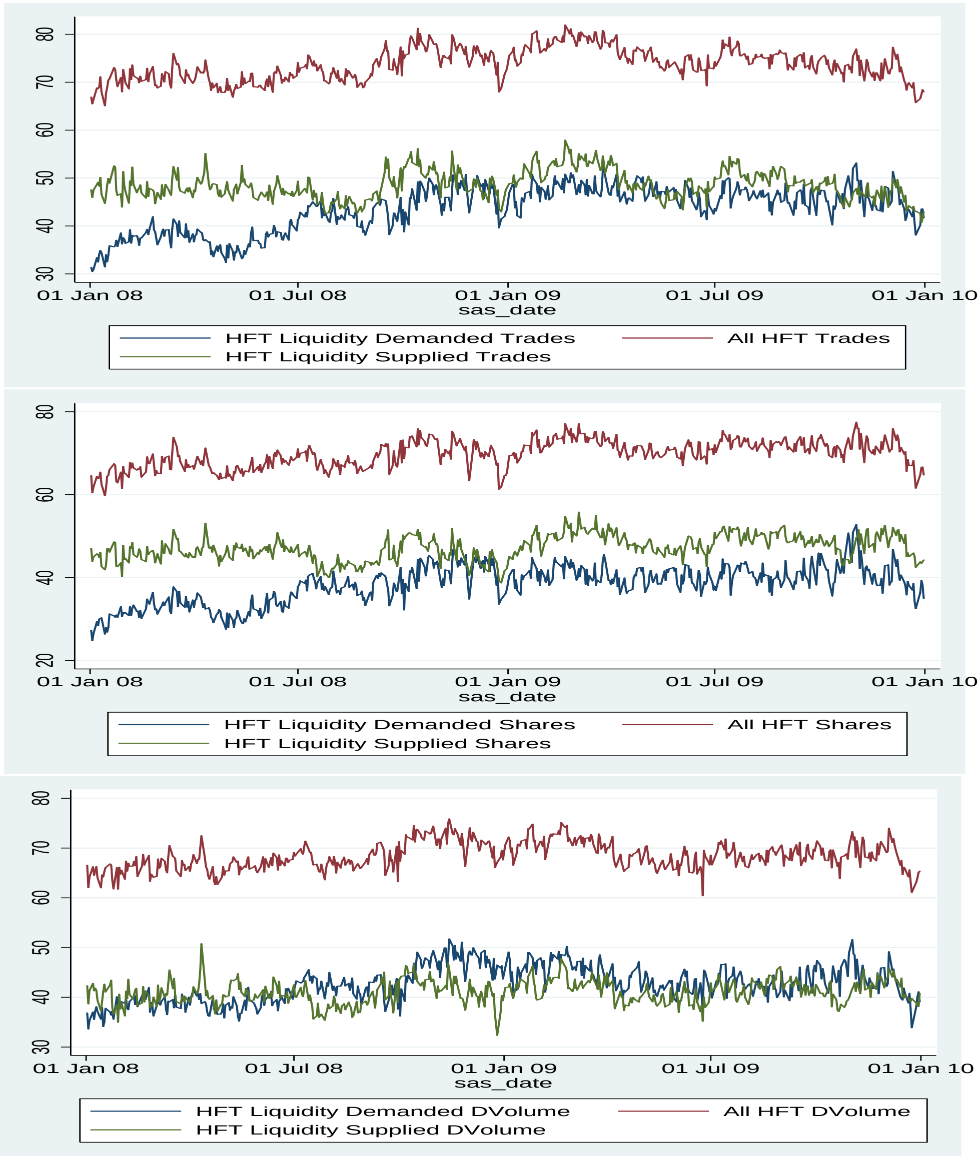
A concern surrounding the May 6 "flash crash" was that the regular market participants, such as HFT, stopped trading. Although the database I have does not include the May 6, 2010 date, it does span 2008 and 2009, which were volatile times in U.S. equity markets. To see whether HFT percent of market trades varies significantly from day to day, and especially around time periods when the U.S. market experienced large losses, I look at each trading day and count what fraction of trades in which HFT were involved. The results are shown in figure 2. There are three graphs. The first is a time series of the fraction of trades HFT were involved in during 2008 and 2009. The second graph looks at the fraction of shares in which HFT were involved during this period. The final graph looks at the fraction of dollar volume in which HFT were involved during this period. In each graph there are three lines. The line labeled "All HFT" represents the fraction of exchanges in which HFT were involved either as a liquidity provider or a liquidity taker; the line labeled "HFT Liquidity Supplied" represents the fraction of transactions in which HFT were providing liquidity; the line "HFT Liquidity Demanded" represents the fraction of trades in which HFT were demanding liquidity. All three graphs have minimal volatility among the three measures. Especially of note, there is no abnormally large drop, or increase, in HFT participation occurring in September of 2009, when the U.S. equity markets were especially volatile.

5.2 Profitability

HFT 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. HFT 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 HFT. 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 the dataset only contains Nasdaq trades. Therefore, there will be many other trades that occur that the dataset does not include. Nasdaq makes up about 20% of all trades and so 4 out of every 5 trades are not part of the data set.

I consider all HFT to be one trader. I take all the buys and sells at the respective prices of the HFT and calculate how much money was spent on purchases and received from sales. HFT tend to 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 calculate an estimate of the total prof-

Figure 2: Time Series of HFT Market Participation The first graph is a time series of the fraction of trades HFT where involved in during 2008 and 2009. The second graph looks at the fraction of shares in which HFT were involved. The final graph looks at the fraction of dollar volume in which HFT were involved. In each graph three lines appear. One line represents whether HFT were involved as either a liquidity provider or a liquidity taker; another line represents transactions in which HFT were providing liquidity; the final line represents when HFT were demanding liquidity.



itability of these 26 firms. As many stocks do not end the day with an exact net zero buying and selling by HFT, I take any excess shares and assume they were traded at the mean price of that stock for that day. The result of this exercise is that on average, per day, HFT 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 HFT. First, the 120 stocks have a combined market capitalization of \$2,110,589.3 (million), whereas all compustat firms' combined market capitalization is \$17,156,917.3 (million), and so I should multiply the profitability by 8.13, raising the per day HFT profitability from all stocks to \$2,423,659.5 per day. The other large factor to be incorporated is that Nasdaq trades make up approximately 20% of all trades, so assuming HFT trade on other exchanges as they do on Nasdaq, the previous number should be multiplied by five. Thus the estimated daily profit of these 26 firms is \$12,118,297.5. Per year that is \$3,029,574,380. Although this is a large absolute number, relatively it is small, especially given that HFT trade around \$30 trillion annually.

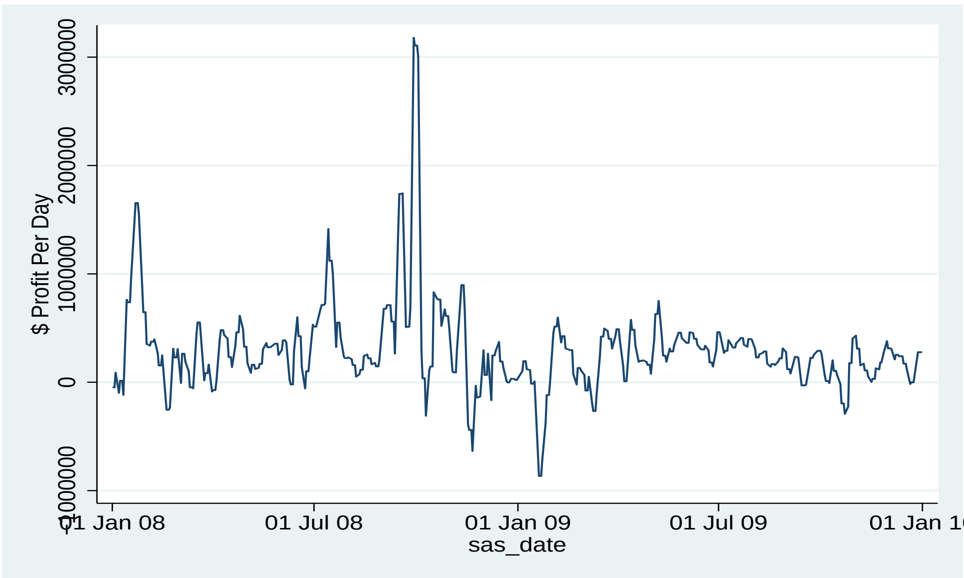
There is no adjustment made for transaction costs yet. However, such costs will be negligible, the reason being that when HFT provide liquidity they receive a rebate from the exchange, for example Nasdaq offers \$.20 per 100 shares for which traders provided liquidity, but this is only for large volume traders like the HFT. On the other hand, Nasdaq charges something like \$.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.

Figure 3 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.

To try to understand what drives the changes in profitability per day I look at the determinants for what stocks on different days are the most profitable. I regress the profitability on several potentially important variables, the same ones used in the regression to determine HFT percent of the market. I run the following standardized regression (to obtain the economic impact):

$$\begin{aligned} Profit_{i,t} &= \alpha + H_{i,t} * \beta_{i,t} + MC_i * \beta_i + MB_i * \beta_i + NT_{i,t} * \beta_{i,t} + \\ &\quad NV_{i,t} * \beta_{i,t} + Dep_{i,t} * \beta i, t + Vol_{i,t} * \beta_{i,t} + AC_{i,t} * \beta i, t, \end{aligned}$$

Figure 3: 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 compaing 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.



where all variables are defined as before, and the dependent variable $Profit$ takes on three different definitions. The results are displayed in Table 10. In the first column $Profit$ is defined as the profit per HFT share traded averaged over stock i on day t; in the second column it is the amount of money HFT made for stock i on day t; in the third column it is the number of HFT shares traded for stock i on day t. The second and third regression decompose the parts of the first regression's dependent variable. Again, the reported coefficients have been standardized so that the coefficient value represents a one standard deviation movement in a particular variable's impact on $Profit$.

The Profit per HFT Share Traded regression has no statistically or economically significant variables and has a negative r-squared. The second regression, with the dependent variable as profits, has two coefficients that are statistically significant. Autocorrelation and Volatility. $Autocorrelation$ has a smaller coefficient and is negative, implying the less predictable price movements in a stock the more profitable is that stock for HFT. The $Volatility$ measure has a large positive economic impact and is highly statistically significant.

The third regression, HFT shares traded, has three statistically significant and economically significant variables. $MarketCap.$ is positive with a coefficient of 0.21, the $AverageDepth$ coefficient is positive and has a coefficient of 0.098, and the $Volatility$ coefficient, which also has a positive relationship with the dependent variable, shows the largest coefficient magnitude of 0.622.

The results in this section have shown that HFT engage in a price reversal trading strategy, that HFT tend to trade more in large stocks with relatively low volume with narrow spreads and depth. Also, HFT are profitable, making approximately $3 billion a year, and that the profitability is driven by volatility. Next, I investigate the role HFT play in demanding and supplying liquidity.

6 Market Quality

The following section analyzes HFT impact on market quality. Market quality refers to liquidity, price discovery, and volatility. Each analysis uses different techniques to study the relationship between HFT and each type of market quality.

6.1 HFT Liquidity

Liquidity supply and demand in the microstructure literature refers to which side of the transaction entered the marketable order and which side had a limit order in place that was executed. The side with the limit order is the liquidity supplier, and the marketable order side is the liquidity taker. In this section I look at the de-

Table 10: Determinants of HFT Profits Per Stock Per Day The dependent variable for the first column is defined as the profit per HFT share traded averaged over each stock i on day t; in the second column it is the amount of money HFT made for each stock on each day; in the third column it is the number of HFT shares traded.

	Profit per HFT Share Traded	Profits	HFT Shares Traded
HFT Percent	-0.087	-0.024	0.062
	(-1.04)	(-0.40)	(1.42)
Market Cap.	-0.015	-0.059	0.210***
	(-0.16)	(-0.86)	(4.20)
Market / Book	0.014	-0.003	0.013
	(0.23)	(-0.07)	(0.43)
$ of Non HFT Volume	-0.058	0.019	-0.073
	(-0.76)	(0.36)	(-1.90)
Average Spread	-0.004	-0.015	0.000
	(-0.06)	(-0.35)	(0.01)
Average Depth	-0.009	-0.008	0.098***
	(-0.16)	(-0.19)	(3.33)
Volatility	0.038	0.359***	0.622***
	(0.67)	(8.67)	(20.70)
Autocorrelation	-0.033	-0.078*	-0.036
	(-0.62)	(-1.97)	(-1.24)
# of Non HFT Trades	0.027	-0.025	0.031
	(0.29)	(-0.41)	(0.69)
Constant			***
	(0.96)	(1.45)	(-3.66)
Observations	360	590	590
Adjusted R^2	-0.014	0.111	0.532

Standardized beta coefficients; t statistics in parentheses

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

scriptive statistics of how HFT demand liquidity, then I examine how they supply liquidity, finally I analyze how much liquidity they provide in the quotes and the book, not just for trades.

6.1.1 HFT Liquidity Demand

The results in table 4 show that liquidity is demanded by HFT in 50.4% of all trades. This section will analyze how HFT initiated trades tend to behave compared to non-HFT trades. HFT tend to demand liquidity in similar dollar size trades as do non HFT. There appears to be clustering in trades, whereby if a previous trade is a buy, it is much more likely the next trade will also be a buy, and the same is true for sales, and this clustering is stronger for HFT than for Non-HFT. Trades that either proceed a HFT or follow a HFT tend to occur more quickly than those proceeding or following a non-HFT. As trade size increases, the time between trade decreases, and this is true regardless of the size of the firm. Finally, HFT demands are quite consistent across the day, but they make up a significantly smaller portion of trades at the opening and close of the trading day.

Table 11 looks at the percent of all transactions for different size trades, in dollar terms, and with different HFT and non-HFT liquidity providers and demanders. The first column of Table 11 reports the fraction of trading volume for different combinations of HFT firms and non-HFT firms as liquidity providers and takers. For small trades, those worth less than \$1,000 HFT are not as involved as Non HFT, this is consistent with the previous results that show HFT tended to trade more in stocks with large market caps, which typically have stock prices in the double digits. Most trades occur in the value range of \$1,000 to \$4,999. The HFT in two of their three categories are the most engaged in these transactions. HFT's share of trades engaged in falls in the \$5,000 to \$14,999 category, except for when they are demanding liquidity. In the \$30,000 plus category of trades, HFT provide the least amount of liquidity, but tend to demand the most. This suggest that HFT are liquidity takers in large trades and liquidity providers in small shares, which is consistent with the theory that HFT are concerned with informed traders in big trades.

The previous table analyzed the frequency of different types of trades, the next table examines the conditional frequency and occurrence of different types of trades. Table 12, similar to that in Biais, Hillion, and Spatt (1995) and Hendershott and Riordan (2009), provides evidence on the clustering of HFT trades in trade sequences. In the table, H stands for HFT and N stands for non-HFT. The first letter in the rows for Panel A and B is who is demanding liquidity at Time t-1. The second letter in these two panels is who is demanding liquidity at time

Table 11: HFT Volume by Trade-size Category. This table reports dollar-volume participation by HFT and non-HFT in 5 dollar-trade size categories. The first letter in the column labels represents the liquidity seeking side. The second letter in the column labels represents the passive party. H represents a HFT, N represents a non-HFT.

	Type of Liquidity Taker and Liquidity Supplier					
Dollar Size Categories	**HH**	**HN**	**NH**	**NN**	**Total**	**N**
0-1,000	21.5%	25.7%	25.3%	27.5%	100.0%	245,401
1,000-4,999	25.8%	24.5%	28.3%	21.5%	100.0%	1,473,047
5,000-14,999	23.7%	27.0%	26.3%	23.0%	100.0%	940,634
15,000-29,999	25.1%	25.9%	26.2%	22.8%	100.0%	308,102
30,000 +	17.4%	32.2%	19.6%	30.7%	100.0%	141,609
Total	24.4%	25.8%	26.8%	23.0%	100.0%	3,108,793
N	757,864	803,177	833,883	713,869	3,108,793	

t. Panel A reports the unconditional frequency of observing HFT and non-HFT trades. Seeing a HFT demand liquidity in time t-1 followed by a HFT demanding liquidity in time t is as common as seeing any other time t-1, t sequence. Panel B reports the conditional frequency of observing HFT and non-HFT trades after observing trades of other participants. In Panel B, the columns are whether the liquidity taker is buying (B) or selling (S). The first letter represents what the liquidity taker is doing in the time t-1 trade. The second letter represents what the liquidity taker is doing in the time t trade. In column and row headings t indexes trades, not time. The results suggest that one tends to see liquidity demanders purchase shares follow a previous trade of a liquidity demander purchasing shares, and the same with sales, regardless of what type of trader was demanding the liquidity. The clustering affect is stronger, in both buying and selling, for HFT demanders than it is for Non-HFT demanders.

Panel C provides conditional probabilities based on the previous trade's size and type of trader. The rows represent the type of trader taking liquidity at time t-1, either H for HFT or N for non-HFT. In addition, the rows are further partitioned based on the size of the trade, measured by the dollar size of shares exchanged in the t-1 trade. 1 represents a trade of size \$0 -\$999; 2 represents a trade of size \$1,000 - \$4,999; 3 represents a trade of size \$5,000 - \$14,999; 4 represents a trade of size \$15,000 - \$29,999; and 5 represents a trade of size greater than \$30,000. The columns identify who was the liquidity demander at time t (H or N) and is further partitioned along the size categories discussed above. The results show that trades of size and type of liquidity demander are highly dependent on the previous trade type. HFT tend to trade with HFT, and the larger the dollar size of a trade the higher the likelihood the next trade will be large.

The next set of results regarding type of trader initiating trading looks at the time between trades. Table 13 reports the average time between trades dependent on different trade characteristics. All times reported are in seconds. Panel A reports the average amount of time between two trades, two HFT liquidity demanding trades, and two non-HFT liquidity demanding trades, and between a trade where the t-1 trade was initiated by a trader who was a HFT, or a non-HFT. Both trades when the liquidity demander is HFT at both t-1 and t, and when HFT is the liquidity demander at t-1, regardless of who demands liquidity at time t, are more rapidly executed.

Panel B provides the average amount of time between two different trade orderings and total dollar-volume and per trade dollar-volume categories. The first two columns in Panel B is for some trade type at t-1 and at time t there is a liquidity taker of H or N, where the columns are separated based on the time t liquidity

Table 12: Trade Frequency Conditional on Previous Trade. Panel A reports the unconditional frequency of observing HFT and non-HFT trades. Panel B reports the conditional frequency of observing HFT and non-HFT trades after observing trades of other participants. In column and row headings t index trades. Panel C provides conditional probabilities based on the previous trade's size and participant. The first letter in the rows for Panel A and B is who is demanding liquidity at Time t-1. The second letter in these two panels is who is demanding liquidity at time t.

Panel A

T-1 Type and T Type	%
HH	24.4%
HN	25.8%
NH	26.8%
NN	23.0%
Total	100.0%

Panel B

	T-1 Buy or Sell and T Buy or Sell				
T-1 Type and T Type	**BB**	**BS**	**SB**	**SS**	**Total**
	%	%	%	%	%
HH	44.3%	5.9%	5.9%	43.9%	100.0%
HN	44.1%	6.5%	6.3%	43.1%	100.0%
NH	41.9%	7.5%	7.7%	42.9%	100.0%
NN	41.5%	7.7%	7.8%	42.9%	100.0%
Total	43.0%	6.9%	6.9%	43.2%	100.0%

Panel C

	LD_Size										
lag_LD_Size	**H1**	**H2**	**H3**	**H4**	**H5**	**N1**	**N2**	**N3**	**N4**	**N5**	**Total**
	%	%	%	%	%	%	%	%	%	%	%
H1	25.9%	32.2%	13.3%	2.5%	1.0%	7.5%	10.8%	5.4%	0.8%	0.4%	100.0%
H2	5.1%	58.2%	14.8%	3.8%	1.6%	1.3%	11.4%	3.0%	0.7%	0.3%	100.0%
H3	3.2%	23.1%	46.2%	6.1%	2.3%	0.9%	4.3%	12.0%	1.3%	0.5%	100.0%
H4	1.9%	18.2%	18.7%	35.0%	7.6%	0.5%	2.9%	4.5%	8.6%	2.1%	100.0%
H5	1.7%	17.0%	16.2%	17.3%	27.8%	0.5%	2.7%	3.9%	5.2%	7.6%	100.0%
N1	6.8%	7.4%	3.5%	0.7%	0.3%	39.6%	29.5%	9.6%	1.7%	0.8%	100.0%
N2	1.7%	11.5%	2.8%	0.7%	0.3%	5.3%	57.9%	14.5%	3.6%	1.8%	100.0%
N3	1.3%	4.6%	12.4%	1.6%	0.6%	2.7%	23.1%	44.8%	6.0%	2.7%	100.0%
N4	0.6%	3.4%	3.9%	9.0%	2.5%	1.5%	17.8%	18.6%	34.5%	8.1%	100.0%
N5	0.7%	3.7%	4.0%	4.8%	7.9%	1.5%	18.3%	18.0%	17.3%	23.9%	100.0%
Total	3.7%	23.9%	15.4%	5.1%	2.3%	4.2%	23.6%	14.9%	4.8%	2.2%	100.0%

taker. The last two columns is similar except that its columns are distinguished based on the time it takes when the time t-1 liquidity taker is a certain type (H or N). Rows S1 through M5 represent different types of stocks. The first character, S,M, or L, represents the dollar volume traded in a given stock on a given day, with S being for trades in small stocks with total dollar volume under $800 Million, M for medium stocks with dollar volume between $800 Million and $1.2 Billion, and L for large stocks with dollar volume greater than $1.2 Billion. The second character ,the number 1 through 5 represents the size of the particular trade. If the trade was less than $1000 then it is a 1, if its between $1,000 and $4,999 its a 2, if between $5,000 and $14,999 its a 3, if between $15,000 and $29,999 its a 4, and if its greater than $30,000 it is a 5. The results suggest that as more dollar volume is traded ,the time between trades decreases. Also, within each day dollar volume category, the larger the trade, usually the shorter the time before another trade occurs. This is the opposite of what Hendershott and Riordan (2009) find; they see that small orders for Algorithmic Traders tend to execute faster. This is evidence that HFT actively monitor the market for liquidity, but that they focus their trading strategy around price pressures from large trades. Finally, for most of the different categories, HFT tend to trade more rapidly, whether looking at time t-1 or time t.

Finally, I examine the intraday pattern of HFT supply and demand of liquidity. If HFT do try and end the day with a near net zero position in stocks then they should wind down their trading before the end of the trading day in order to prevent getting stuck with shares in their position they do not want to hold overnight. Similarly, at the beginning of the trading day they will have few positions in which they are trying to maintain a near net zero position in and so trading should be less prevalent. To analyze this I create a time series of the type of traders throughout the day. I take all trades that occur on February 22, 2010 - February 26, 2010 and put them in to ten second bins based on the time of day they occurred, regardless of the day. Then, I split them into the types of trades based on who was supplying liquidity and demanding liquidity and calculate the percent of each type of transaction per time period bin. Figure 4 shows the make up of different types of trades throughout the day. The four different patterns, HH, HN, NH, NN refer to the type of liquidity demander (first letter) and liquidity supplier (second letter). The figure is stacked so that each time period sums to one. During the day the trading ratios are quite stable, except at the beginning and end of day. During these periods HFT tend to trade with each other much less frequently and HFT tend to initiate fewer trades and to provide liquidity in fewer trades. This is consistent with the scenario of HFT trying to end the day near net-zero in their equity positions.

Table 13: Average Time Between Trades. All values are in seconds. Panel A reports the average amount of time between two trades, two HFT liquidity demanding trades, and two non-HFT liquidity demanding trades, and between a trade where the initial trade had a HFT, or a non-HFT liquidity demander. Panel B provides the average amount of time between two different trade orderings and trade-size categories (refer to the previous table for the different trade-size categories) The first two columns in Panel B is for some trade type at t-1 and at time t there is a liquidity taker of H or N, where the columns are separated based on the time t liquidity taker. The last two columns is similar except that its columns are distinguished based on the time it takes when the time t-1 liquidity taker is a certain type (H or N).

Panel A

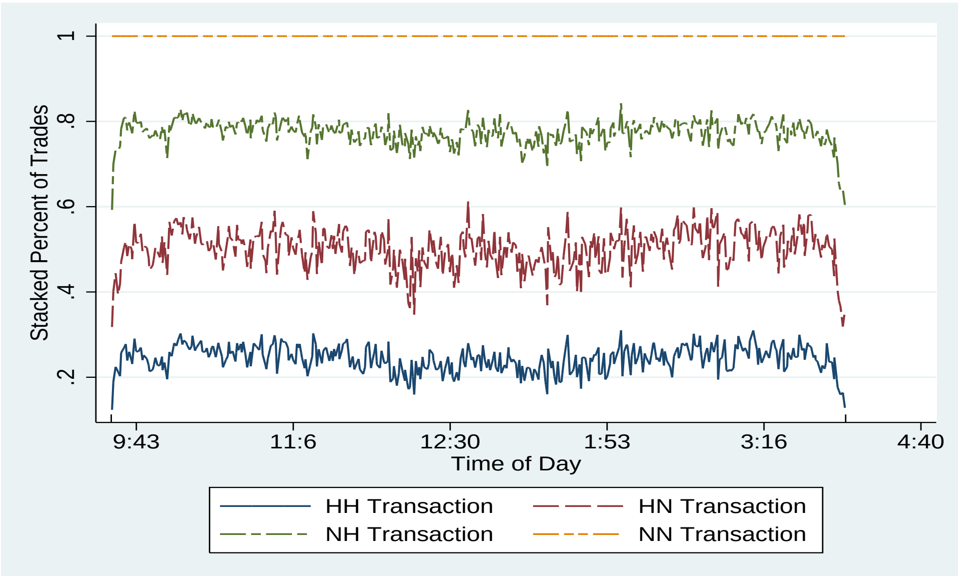
	HFT	**Non-HFT**
Unconditional Time Between Trades	3.351	5.667
Time Between Trades of Same Type of Trader	8.696	9.081

Panel B

	Time t Liquidity Taker		**Time t-1 Liquidity Taker**	
	HFT	**Non-HFT**	**HFT**	**Non-HFT**
S1	30.746	25.449	18.384	35.843
S2	9.620	10.578	7.582	12.657
S3	5.069	5.513	4.485	6.134
S4	3.426	3.354	2.742	4.067
S5	4.576	4.065	3.572	5.056
M1	0.988	1.538	1.183	1.378
M2	1.219	1.399	1.064	1.572
M3	1.095	1.435	1.112	1.434
M4	1.098	1.166	0.998	1.273
M5	1.136	0.989	0.851	1.290
L1	0.746	1.266	0.899	1.118
L2	1.135	1.119	0.841	1.447
L3	0.746	1.149	0.842	1.065
L4	0.780	0.724	0.668	0.833
L5	0.729	0.639	0.602	0.760

Figure 5 also looks at trades throughout the day, but only charts the percent of dollar volume in which HFT are demanding liquidity (top graph) and supplying liquidity (bottom graph). Again this shows that HFT significantly reduce both their supply and their demand for liquidity at the start and end of trading hours.

Figure 4: Type of Liquidity Providers / Takers throughout the day. The figure shows the make up of traders throughout the day. It shows that HFT tend to reduce their trading activity at the opening and closing of the trading day. the first letter is the liquidity taker, the second letter is the liquidity provider.

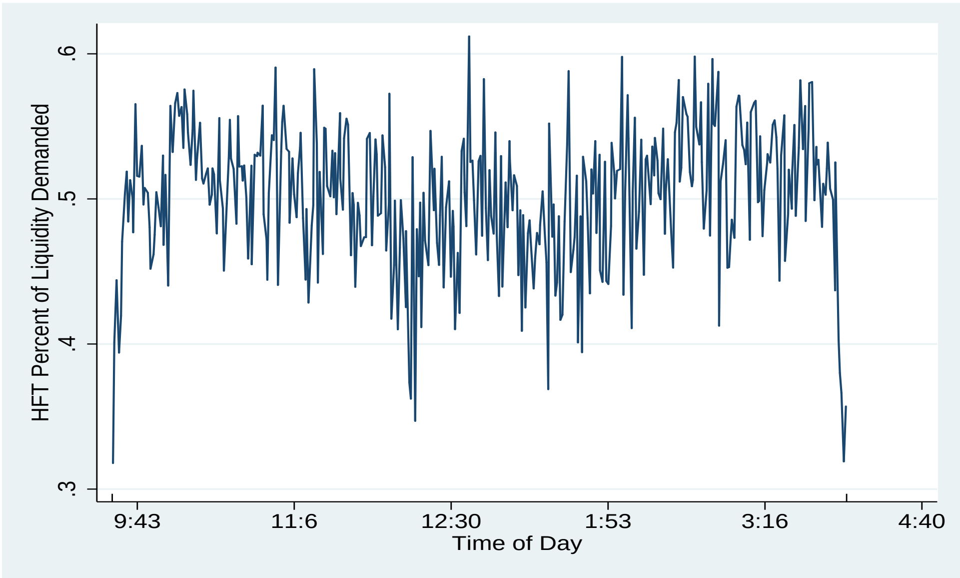


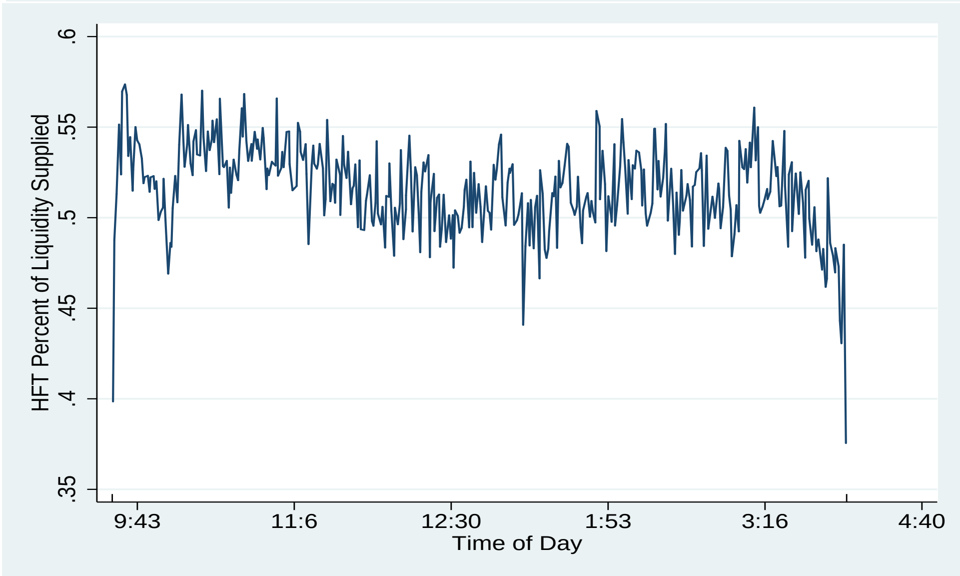
6.1.2 HFT Liquidity Supply

This section analyzes how HFT supply liquidity. I focus on the amount of liquidity HFT supply. I show that, beyond supplying liquidity in 51.4% of all trades, they also often supply the inside quotes throughout the day. I then consider the determinants that determine which stocks, and on what days, do HFT provide the inside quotes. Finally, I examine what the additional price impact would be on stocks if HFT where not part of the book. There is a sizeable impact, which shows the importance of HFT providing liquid markets.

6.1.2.1 HFT Time at Inside Quotes To begin analyzing HFT's role in providing liquidity in the stock market I look at the amount of time HFT supply the inside bid or ask compared to non-HFT firms. For each stock, on each day, I take

Figure 5: HFT Liquidity Demander or Supplier throughout the day. The graph shows the make up of HFT throughout the day. The first graphs shows the HFT demand for liquidity throughout the day. The second graph shows the HFT supply of liquidity throughout the day.





the number of minutes HFT are providing the inside bid or ask and, either: (a) subtract the number of minutes non-HFT are providing the best inside bid or ask (ties are dropped), these are the "Minutes" results, or (b), divide this value by the total number of minutes where HFT and non-HFT did not have the same inside quotes, these are the "Percent" results. The results are in Table 14.

Table 14 looks at, for the 120 stocks, whether the HFT provide more liquidity than non-HFT traders by considering how often they are providing the inside quote (bid or ask). The way the metric is constructed there are a total of 900 minutes (2*60*7.5) that a HFT could potentially be providing the inside quote. This is twice as many minutes then what actually occur during the trading day. The table is separated into two categories - the category "HFT -" is when the HFT provides fewer inside quotes than the non-HFT for a stock on a given day; category "HFT +" is when the HFT is more frequently the inside quote provider for a stock on a given day. The reason to separate out the two types is that it may be that some stocks HFT do not actively try and provide inside quotes for, thus just taking the average across all stocks would underweight the liquidity they provide in stocks for which they are activily trying to place competitive quotes. Table 14 has three panels, and within each panel either a category called "Minutes" or "Percent." Panel A considers quotes for all stocks; Panel B considers quotes for stocks on days they are below their average spread; and Panel C considers quotes for stocks on days they are above their average spread. The Minutes results display the number of minutes HFT provide the inside quotes more than non-HFT through the following calculation: sum the number of minutes HFT provide the best bid or ask, subtract the number of minutes non-HFT are providing the best inside bid or ask, and drop ties. A problem with this is that since the ties will vary across days and stocks, the Minutes approach does not necessarily capture the frequency that HFT provide better inside quotes than non-HFT. The Percent results avoids this issue by dividing the number of minutes HFT provide the best quotes by the total number of minutes where HFT and non-HFT did not have the same inside quotes.

It does not appear there is a significant difference between the stocks in which HFT decide to place aggressive bid/ask orders and those in which it does not. The very low value for the mean of Net - by itself may imply that the HFT do not attempt to match or out-price the quotes of non-HFT for some stocks. But looking at the "Percent" data, shows that the Net - and Net + results are about equally distant from .5. Thus, the "Minutes" Net - results must be biased downwards as a result of a large number of periods where HFT and non-HFT provide the same prices. Looking at the Panel A - Percent results, on average HFT provide the best inside quotes 45% of the time, a significant portion of the trading day. This suggests that HFT act as market makers and are competitive.

Panel B and C divide the stocks into those that are offering higher spreads than usual and those offering lower spreads than usual. Panel B reports the low spread stock days, Panel C the high spread stock days. The results between the two subsets do not differ much from one another. The average time HFT offer the best quotes is slightly higher when spreads are high at 71.7% compared to 70.1% when it is low. Also, HFT provide the best quotes more often than non-HFT slightly more often when spreads are high, doing so 46% of the time as opposed to 45.1% when spreads are low. This is consistent with HFT attempting to capture liquidity supply profits as found in Foucault and Menkveld (2008) and Hendershott and Riordan (2009) make/take liquidity cycle, but as the difference is small does not provide much support for it.

6.1.2.2 HFT Time at Inside Quotes Determinants Tables 14 shows that HFT are at the inside quotes frequently, but not as much as non-HFT. I perform an OLS regression similar to that found in table 9 to understand what determinants are related to which stocks and days HFT decide to provide the best quotes. Table 15 shows the results. It is very similar to table 9, with all variables being defined exactly the same as before except the dependent variable. The regression is:

$$\begin{aligned} L_{i,t} &= \alpha + MC_i * \beta_i + MB_i * \beta_i + NT_{i,t} * \beta_{i,t} + \\ &\quad NV_{i,t} * \beta_{i,t} + Dep_{i,t} * \beta_{i,t} + Vol_{i,t} * \beta_{i,t} + AC_{i,t} * \beta_{i,t}, \end{aligned}$$

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

The coefficients reported, like those in table 9, 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 determining HFT providing the best quotes. The other coefficients are negative, suggesting that HFT prefer to provide the inside quotes for value firms, less volatility firms, firms with narrower spreads, and firms with a lower book depth.

Table 14: HFT Time at Best Quotes. This table reports the number of minutes HFT are at the best bid or ask compared to non HFT. The remainder of time both HFT and non-HFT are both at the best quotes is not considered. Panel A is for all stocks at all times, Panel B is for days when the spread is below average for that stock, Panel C is for days when the spread is above average.

Panel A All *-Minutes*

HFT	mean	min	p25	p50	p75	max	N
Net -	-255.6	-779.7	-368.5	-190.3	-75.8	-0.4	353.0
Net +	65.8	0.1	14.3	55.1	94.0	343.0	242.0
Average	-124.9	-779.7	-223.5	-41.2	33.1	343.0	595.0

-Percent

Net -	0.281	0.000	0.162	0.307	0.410	0.499	353.000
Net +	0.709	0.502	0.587	0.708	0.802	0.964	242.000
Average	0.455	0.000	0.272	0.446	0.668	0.964	595.000

Panel B Low Spread *-Minutes*

Net -	-243.1	-779.7	-367.0	-183.0	-70.9	-0.4	187.0
Net +	67.7	0.1	17.5	58.6	94.0	319.8	125.0
Average	-118.5	-779.7	-205.9	-40.4	39.8	319.8	312.0

-Percent

Net -	0.283	0.000	0.162	0.309	0.404	0.498	187.000
Net +	0.701	0.509	0.587	0.704	0.786	0.960	125.000
Average	0.451	0.000	0.267	0.435	0.650	0.960	312.000

Panel C High Spread *-Minutes*

Net -	-269.8	-779.1	-401.2	-204.0	-84.9	-1.1	166.0
Net +	63.6	1.1	13.3	50.4	91.5	343.0	117.0
Average	-131.9	-779.1	-227.1	-41.2	28.9	343.0	283.0

-Percent

Net -	0.278	0.000	0.158	0.296	0.415	0.499	166.000
Net +	0.717	0.502	0.593	0.714	0.814	0.964	117.000
Average	0.460	0.000	0.274	0.455	0.680	0.964	283.000

Table 15: Determinants of HFT Percent of Liquidity Supplying The dependent variable is the ratio of number of minutes HFT provides the inside bid or ask divided by the total number of minutes of when the inside bid and ask differ between HFT and non-HFT.

	(1)
	Economic Impact
Market Cap.	0.654***
	(15.05)
Market / Book	-0.163***
	(-4.72)
$ of Non HFT Volume	-0.162***
	(-3.82)
Average Spread	-0.165***
	(-4.90)
Average Depth	-0.086**
	(-2.65)
Volatility	-0.241***
	(-7.14)
Autocorrelation	-0.006
	(-0.18)
# of Non HFT Trades	0.217***
	(4.28)
Constant	*
	(-2.55)
Observations	590
Adjusted R^2	0.410

Standardized beta coefficients; t statistics in parentheses

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

6.1.2.3 Price Impact Reduction from HFT Liquidity 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 HFT providing 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 HFT 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 16. I consider a variety of different impacts based on the number of shares hypothetically bought. The number of shares varies from 100 to 1000. I show 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.

As the trade size increases, the price impact increases across firms of all sizes and for all ten trade size increases. Interestingly 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. The price impact is substantial. 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 HFT provide.

6.2 HFT Price Discovery

HFT makes up a significant portion of the market, both on the demand and 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 HFT 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. HFT provide more private information to the market than do non-HFT traders. 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 HFT are more important in the price discovery process than non-HFT trades. Finally, I implement the information shares approach which takes the innovations in HFT and non-HFT quotes

Table 16: Liquidity Book Impact. This table looks at the liquidity depth of HFT and non-HFT traders 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 column label Dollars is the dollar difference as a result of HFT being in the market. The column label Basis is the percent basis points change in price as a result of HFT being in the market. The different rows represent a varying number of shares traded.

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

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. The HFT provide substantially more information to the price process than do non-HFT. The Hasbrouck methodologies utilized in this paper are similar to those found in Hendershott and Riordan (2009) and other papers.

6.2.1 Permanent Price Impact

To measure the information content of HFT and non-HFT trades I calculate the permanent price impact of HFT and non-HFT 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, an 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 or quote change. q^H is defined as the signed (+1 for a buy, -1 for a sell) HFT trades and q^N is the similarly denoted signed non-HFT 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:

$$\begin{bmatrix} r_t \\ q_t^H \\ q_t^N \end{bmatrix} = \begin{bmatrix} a(L) & b(L) & c(L) \\ d(L) & e(L) & f(L) \\ g(L) & h(L) & i(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}, \tag{4}$$

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 trades. 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 trades. The impulse response function is a technology first used in the macroeconomic 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 17 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 HFT provide more private information than do non-HFT trades. This is similar to the findings in Hendershott and Riordan (2009) with algorithmic trades. Thus, an innovation in HFT trading tends to lead to a 34% greater permanent price change than does a trade by a non HFT.

Table 17: 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								

6.2.1.1 LR - SR Price Impact The results in table 17 show that HFT has a larger price impact than does non-HFT over the 10 period intervals. 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 HFT trade this would support the theory that HFT 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 18.

Of the 105 The LR-SR impulse response is less for HFT than for non-HFT 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 HFT than non-HFT 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 18 suggest that HFT individual innovations have more private information than non-HFT trades and that the difference is persistent and increases beyond the immediate impact of the trade.

6.2.2 Aggregate Amount of Information in HFT - Variance Decomposition

The permanent price impact section above shows that HFT demanded trades add important information to the market, but the methodology does not directly estimate the importance of HFT and non-HFT trading 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 trades. 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 methodlogy. To perform the variance decomposition the return series r_t(using midpoint returns to avoid the bid-ask bounce) 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

Table 18: 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								

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:

$$\sigma^2_w = (\sum_{i=0}^{10} a_i)^2 \sigma^2_{\epsilon_1} + (\sum_{i=0}^{10} b_i)^2 \sigma^2_{\epsilon_2} + (\sum_{i=0}^{10} c_i)^2 \sigma^2_{\epsilon_3}, \tag{5}$$

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 19. 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 is the contribution to price discovery from HFT, 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.

6.2.3 Information Share

This section examines the role HFT and non-HFT 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.

Table 19: 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								

The approach is as follows. I calculate 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:

$$\Delta p_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} \ldots, \tag{6}$$

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:

$$\sigma_u^2 = \begin{bmatrix} \Psi_{HFT} & \Psi_{nHFT} \end{bmatrix} \begin{bmatrix} \sigma_{HFT}^2 & \sigma_{HFT,nHFT}^2 \\ \sigma_{HFT,nHFT}^2 & \sigma_{nHFT}^2 \end{bmatrix} \begin{bmatrix} \Psi_{HFT} \\ \Psi_{nHFT} \end{bmatrix}, \tag{7}$$

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 20. The information share attributable to HFT and non-HFT 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 20 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. 103 stocks have the HFT information share being larger than the non-HFT information share. Of those 63 of the stock have HFT being statistically significantly providing more information in their quotes than non-HFT. Of the 17 companies where the non-HFT have a larger information share than HFT, only two of the differences are statistically significant. This suggest that in quotes, like in trades, HFT are important in the price discovery process.

6.3 Volatility

The final market quality measure analyzed is the relationship between HFT and volatility. I first do an OLS regression to observe whether there is any relationship between HFT and volatility. The results suggest that HFT and volatility are not highly related, especially contemporaneously. Next I compare the price path of stocks with and without HFT being part of the data generation process. The results suggest that HFT reduces volatility to a degree.

I begin this analysis by doing two simple regressions. The first is a regression with the dependent variable being volatility, calculated in terms of 10 second realized volatility for each stock over the five trading days February 22 - 26, 2010, and the explanatory variables are the total shares traded during that 10 second period and the percent of trades involving a HFT during that ten second period, as well as leads and lags for these two variables, as well as the volatility, for the previous ten periods.

Similarly, I switch the regression so that the dependent variable is the HFT percent of trades in that ten second window, and the volatility is one of the explanatory variables, along with the others previous included in the regression. The two regressions are as follows:

$$
\begin{aligned}
Vol_{i,t} &= \alpha + \beta_{1-11} \times rvlag_{i,0-10} &+ \\
&\beta_{12-22} \times totshareslag_{i,0-10} &+ \\
&\beta_{23-33} \times Hperclag_{i,0-10} &+ \\
HPerc_{i,t} &= \alpha + \beta_{1-11} \times rvlag_{i,0-10} &+ \\
&\beta_{12-22} \times totshareslag_{i,0-10} &+ \\
&\beta_{23-33} \times Hperclag_{i,0-10}
\end{aligned}
$$

Each explanatory variable has a subscript 0-10, this represents the number of ten-second time periods prior to the dependent variable time t event that the variable represents. Subscript 0 represents the contemporaneous value for that

Table 20: 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	0.911	0.089	5.879	CDR	0.994	0.006	131.673	FL	0.698	0.302	3.199
AAPL	0.706	0.294	1.275	CELG	0.864	0.136	4.053	FMER	0.875	0.125	3.175
ABD	0.541	0.459	0.242	CETV	0.827	0.173	1.889	FPO	0.646	0.354	1.236
ADBE	0.999	0.001	835.842	CHTT	0.958	0.042	11.002	FRED	0.468	0.532	-0.564
AGN	0.411	0.589	-1.365	CKH	0.499	0.501	-0.013	FULT	0.962	0.038	12.070
AINV	0.529	0.471	0.933	CMCSA	0.913	0.087	8.335	GAS	0.520	0.480	0.988
AMAT	0.999	0.001	1588.193	CNQR	0.961	0.039	14.953	GE	0.984	0.016	36.818
AMED	0.881	0.119	5.472	COO	0.575	0.425	0.400	GENZ	0.616	0.384	0.601
AMGN	0.772	0.228	2.581	COST	0.836	0.164	2.047	GILD	0.476	0.524	-0.111
AMZN	0.649	0.351	0.824	CPSI	0.387	0.613	-1.356	GLW	0.999	0.001	1603.170
ANGO	0.574	0.426	1.047	CPWR	0.941	0.059	8.288	GOOG	0.285	0.715	-1.237
APOG	0.358	0.642	-1.511	CR	0.998	0.002	216.047	GPS	0.939	0.061	7.608
ARCC	0.164	0.836	-5.393	CRI	0.722	0.278	7.383	HON	0.550	0.450	0.274
AXP	0.996	0.004	122.520	CRVL	0.509	0.491	0.326	HPQ	0.347	0.653	-1.021
AYI	0.864	0.136	2.665	CSCO	0.527	0.473	0.205	IMGN	0.548	0.452	1.658
AZZ	0.068	0.932	-7.616	CSE	0.983	0.017	32.191	INTC	1.000	0.000	10364.989
BARE	0.811	0.189	1.671	CSL	0.995	0.005	96.045	IPAR	0.511	0.489	0.094
BAS	0.992	0.008	103.425	CTRN	0.601	0.399	0.788	ISIL	1.000	0.000	4.59e+10
BHI	0.987	0.013	56.934	CTSH	0.561	0.439	0.323	ISRG	0.553	0.447	0.271
BIIB	0.491	0.509	-0.053	DCOM	0.526	0.474	1.141	JKHY	1.000	0.000	1368.332
BRCM	0.991	0.009	67.936	DELL	0.386	0.614	-0.739	KMB	0.570	0.430	0.377
BRE	0.999	0.001	837.482	DIS	0.999	0.001	429.917	KNOL	0.706	0.294	1.701
BW	0.957	0.043	10.772	DK	0.994	0.006	199.784	KR	1.000	0.000	1.30e+07
BXS	0.473	0.527	-0.151	DOW	0.635	0.365	0.914	KTII	0.984	0.016	31.776
BZ	0.929	0.071	6.698	EBAY	0.588	0.412	0.397	LANC	0.883	0.117	4.721
CB	0.482	0.518	-0.138	EBF	0.952	0.048	9.526	LECO	0.952	0.048	20.212
CBEY	0.986	0.014	35.665	ERIE	0.919	0.081	9.941	LPNT	0.734	0.266	3.995
CBT	0.975	0.025	46.213	EWBC	0.942	0.058	7.595	LSTR	0.656	0.344	0.828
CBZ	0.805	0.195	2.822	FCN	0.689	0.311	1.074	MAKO	0.850	0.150	2.332
CCO	0.540	0.460	0.267	FFIC	0.683	0.317	1.368	MANT	0.505	0.495	0.034
MDCO	0.990	0.010	66.948	MELI	0.938	0.062	7.411	MFB	0.291	0.709	-1.178
MIG	1.000	0.000	1557.624	MMM	0.982	0.018	28.253	MOD	0.983	0.017	29.100
MOS	0.858	0.142	3.874	MRTN	0.856	0.144	2.656	MXWL	0.984	0.016	30.346
NC	0.346	0.654	-0.925	NSR	0.879	0.121	4.341	NUS	1.000	0.000	2.40e+10
NXTM	0.542	0.458	0.220	PBH	0.511	0.489	0.308	PFE	0.666	0.334	0.984
PG	0.987	0.013	51.060	PNC	0.631	0.369	0.734	PNY	0.848	0.152	2.292
PPD	0.655	0.345	1.158	PTP	0.455	0.545	-0.494	RIGL	0.988	0.012	60.841
ROC	0.870	0.130	3.134	ROCK	1.000	0.000	5.22e+10	ROG	0.780	0.220	1.603
RVI	0.715	0.285	1.097	SF	0.987	0.013	37.185	SFG	1.000	0.000	8.55e+10
Overall	0.757	0.258	19.923								

variable. Volatility is defined as, for example using the fourth lag, $rvlag_4 = (log(price_{i,t-5}/price_{i,t4}))^2$. The betas represent row vectors of 1x11 and the explanatory variables column vectors of 11x1. rv is the squared price change for company i for the respective time period. $totshares$ is the number of shares that were traded for a company i in that ten second time period. $Hperc$ is the percent of trades for stock i in that time period for which HFT were involved.

Table 21 shows the results of the two regressions and only reports the variables of interest HFT trading percent "$HFT\%$" for the first regression and "RV" for the second regresion. In the first two columns are the results of the first regression with volatility as the dependent variable. The last two columns display the result for the second regression with HFT percent of trading as the dependent variable. For both, only the results of the variables of interest are shown. The results suggest that there is some statistically significant relationship between the two variables. when Volatility is the dependent variable, the Percent of HFT trading coefficient is statistically significant and negative in the two period lag period.

In the second regression, with HFT Percent as the dependent variable, many of the prior volatility coefficients are statistically significant. The periods lag 1, 2, 5, 9, and 10 statistically significant. All of the statistically significant lag coefficients are positive except for period 9. This suggests that after volatility has been elevated HFT tend to make up more of the market trades. Of course, because of the endogeneity problem and econometric issues not much weight should be put on these results. I include them only to show that there might be a relationship between HFT and volatility. The next section attempts to avoid the econometric issues and to reduce the endogeneity problem.

6.3.0.1 HFT Impact on Volatility I next try to disentangle the HFT - Volatility relationship and minimize the endogeneity problem. 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 HFT 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. Table 22, 23, and 24 shows the results. For each stock I calculate the realized volatility, the sum over one minute increments of the absolute value of the returns over the day. I perform this calculation for each stock on each day and do it for the actual price path, and also for an alternative price path based on the role of HFT. In table 22 I remove HFT initiated trades and also HFT liquidity providing trades. Thus, were the realized volatility calculation would have used a trade by

Table 21: HFT - RV Relationship. This table tries to capture whether there is a relationship between HFT and short-term market volatility.

Dep = RV			Dep = HFTPercent		
Variable	**Coefficient**	**Std. Err.**	**Variable**	**Coefficient**	**Std. Err.**
HFT % lag0	-0.00025	0.00022	RV lag0	-0.00917	0.00822
HFT % lag1	-0.00048*	0.00027	RV lag1	0.12203***	0.04379
HFT % lag2	-0.00031	0.00026	RV lag2	0.09531***	0.03580
HFT % lag3	0.00004	0.00025	RV lag3	-0.03216	0.02459
HFT % lag4	-0.00001	0.00025	RV lag4	0.00072	0.02362
HFT % lag5	-0.00005	0.00025	RV lag5	0.03959*	0.02196
HFT % lag6	0.00003	0.00025	RV lag6	-0.00092	0.01904
HFT % lag7	0.00010	0.00025	RV lag7	-0.02628	0.01746
HFT % lag8	0.00006	0.00025	RV lag8	0.02731	0.01675
HFT % lag9	0.00017	0.00025	RV lag9	-0.04089**	0.01608
HFT % lag10	-0.00018	0.00025	RV lag10	0.05887***	0.01552
N	552845		N	552845	
R^2	0.12254		R^2	0.89658	
$F_{(187,552657)}$	1135.02718		$F_{(187,552657)}$	70461.35190	

Significance levels : $*$: 10% $**$: 5% $***$: 1%

a HFT initiated trade, it instead has to grab the price from the next trade that is initiated by a non-HFT. Also, when the realized volatility would have had a trade where a HFT was providing the liquidity, I adjust the price based on the size of the trade and the price impact it would have on the book after removing the HFT book entries. Table 23 does the same calculation but only removes the HFT liquidity providing trades. Finally, table 24 removes only the HFT initiated trades from the alternative price path.

In table 22 of the 120 stocks, only one exhibits that volatility would not be reduced if HFT had not been in the market. Of those where HFT reduced the volatility, 85 of them have volatility that is statistically significantly less than what it would be if HFT had not been part of the market. 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. These results suggest that HFT help to reduce the volatility in the market.

Table 23 looks at what happens when HFT is only removed from providing liquidity. Only one firm shows that volatility is increased by removing HFT from providing liquidity. Of the 119 that show HFT is reduced, 82 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. This is the mechanical portion of the HFT price reduction: when liquidity is removed, the only way prices can move is further away from their previous path, thus increasing volatility.

Table 24 again compares the realized volatility of the 120 firms, but it compares the volatility of the actual price path with the volatility of the price path if only HFT initiated trades are removed. Table 24, unlike the previous table, may show a positive, negative, or no direction in its impact on volatility. Of the 120 firms, 72 of them have a higher volatility when HFT initiated trades are present. Thus, a small majority of firms experience slightly higher volatility with HFT initiated trades. However of these 72 stocks, only one is statistically significant. Of the 48 stocks where the presence of HFT initiated trades reduces 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. These results suggest that HFT initiated trades do not result in increased volatility.

Table 22: HFT Impact on Volatility - No Demand or Supply 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.

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	0.231	0.253	-2.802	CPWR	0.192	0.201	-1.060	JKHY	0.109	0.142	-5.075
AAPL	0.154	0.160	-0.791	CR	0.106	0.126	-3.851	KMB	0.116	0.131	-3.779
ABD	0.193	0.305	-7.614	CRI	0.152	0.187	-2.056	KNOL	0.148	0.237	-3.280
ADBE	0.152	0.166	-2.534	CRVL	0.175	0.226	-3.161	KR	0.120	0.135	-1.673
AGN	0.126	0.143	-5.291	CSCO	0.156	0.164	-1.867	KTII	0.002	0.002	-1.207
AINV	0.176	0.216	-3.275	CSE	0.256	0.337	-1.644	LANC	0.088	0.101	-7.361
AMAT	0.202	0.209	-1.223	CSL	0.103	0.119	-3.095	LECO	0.191	0.237	-2.663
AMED	0.255	0.311	-2.763	CTRN	0.093	0.124	-3.318	LPNT	0.175	0.195	-3.006
AMGN	0.124	0.129	-4.391	CTSH	0.140	0.163	-3.646	LSTR	0.131	0.156	-4.271
AMZN	0.198	0.210	-2.400	DCOM	0.093	0.127	-2.753	MAKO	0.137	0.206	-5.506
ANGO	0.073	0.085	-1.577	DELL	0.161	0.164	-0.421	MANT	0.110	0.130	-1.660
APOG	0.123	0.148	-4.428	DIS	0.135	0.157	-3.288	MDCO	0.231	0.293	-3.133
ARCC	0.169	0.208	-3.760	DK	0.070	0.092	-2.912	MELI	0.340	0.377	-1.461
AXP	0.177	0.199	-6.516	DOW	0.271	0.311	-3.835	MFB	0.097	0.167	-12.196
AYI	0.086	0.097	-1.824	EBAY	0.192	0.202	-2.186	MIG	0.084	0.142	-9.840
AZZ	0.111	0.144	-7.824	EBF	0.119	0.158	-5.954	MMM	0.139	0.154	-3.200
BARE	0.011	0.011	0.000	ERIE	0.080	0.098	-3.536	MOD	0.206	0.252	-2.448
BAS	0.235	0.286	-1.628	ESRX	0.225	0.238	-0.886	MOS	0.265	0.289	-3.305
BHI	0.204	0.235	-5.428	EWBC	0.241	0.282	-3.748	MRTN	0.087	0.102	-1.889
BIIB	0.159	0.173	-9.241	FCN	0.183	0.218	-1.034	MXWL	0.178	0.257	-3.335
BRCM	0.181	0.197	-3.911	FFIC	0.106	0.138	-2.857	NC	0.069	0.093	-3.697
BRE	0.105	0.134	-10.304	FL	0.131	0.161	-1.816	NSR	0.084	0.120	-5.166
BW	0.175	0.222	-9.960	FMER	0.138	0.165	-2.777	NUS	0.114	0.136	-3.199
BXS	0.165	0.193	-0.612	FPO	0.089	0.151	-3.698	NXTM	0.308	0.446	-2.305
BZ	0.288	0.485	-5.311	FRED	0.097	0.144	-11.893	PBH	0.101	0.125	-1.755
CB	0.096	0.125	-6.023	FULT	0.187	0.231	-4.391	PFE	0.167	0.180	-2.433
CBEY	0.150	0.214	-3.612	GAS	0.102	0.133	-3.033	PG	0.116	0.128	-3.568
CBT	0.165	0.203	-3.116	GE	0.165	0.181	-2.462	PNC	0.206	0.231	-6.084
CBZ	0.075	0.157	-12.996	GENZ	0.149	0.165	-5.688	PNY	0.079	0.098	-3.192
CCO	0.159	0.202	-2.751	GILD	0.145	0.155	-3.194	PPD	0.079	0.123	-4.604
CDR	0.140	0.194	-5.620	GLW	0.193	0.213	-3.532	PTP	0.063	0.081	-3.221
CELG	0.188	0.208	-4.976	GOOG	0.137	0.155	-5.110	RIGL	0.215	0.283	-3.812
CETV	0.253	0.287	-2.291	GPS	0.140	0.149	-0.695	ROC	0.188	0.256	-11.827
CHTT	0.001	0.001	-0.708	HON	0.165	0.187	-3.044	ROCK	0.363	0.503	-2.069
CKH	0.071	0.120	-5.628	HPQ	0.128	0.139	-2.567	ROG	0.092	0.126	-3.921
CMCSA	0.180	0.188	-0.948	IMGN	0.184	0.236	-4.469	RVI	0.106	0.165	-4.984
CNQR	0.139	0.167	-4.738	INTC	0.169	0.179	-2.050	SF	0.057	0.095	-10.003
COO	0.135	0.161	-7.339	IPAR	0.093	0.142	-5.570	SFG	0.103	0.131	-3.190
COST	0.106	0.119	-4.397	ISIL	0.153	0.175	-3.379	SJW	0.060	0.092	-2.094
CPSI	0.114	0.140	-3.345	ISRG	0.144	0.164	-4.538	SWN	0.262	0.311	-4.077
Overall	0.148	0.182	-17.562								

Table 23: HFT Impact on Volatility - No Supply 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.

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	0.224	0.253	-2.625	CPWR	0.190	0.201	-1.891	JKHY	0.108	0.142	-7.230
AAPL	0.154	0.160	-0.734	CR	0.103	0.126	-6.593	KMB	0.113	0.131	-4.500
ABD	0.194	0.305	-7.452	CRI	0.153	0.187	-1.648	KNOL	0.147	0.237	-3.391
ADBE	0.154	0.166	-2.062	CRVL	0.172	0.226	-2.180	KR	0.122	0.135	-1.157
AGN	0.127	0.143	-3.597	CSCO	0.157	0.164	-1.300	KTII	0.002	0.002	-2.051
AINV	0.172	0.216	-4.085	CSE	0.263	0.337	-1.547	LANC	0.088	0.101	-9.613
AMAT	0.202	0.209	-1.225	CSL	0.102	0.119	-2.630	LECO	0.187	0.237	-2.593
AMED	0.257	0.311	-3.176	CTRN	0.093	0.124	-3.169	LPNT	0.175	0.195	-3.416
AMGN	0.122	0.129	-4.382	CTSH	0.140	0.163	-2.878	LSTR	0.131	0.156	-3.527
AMZN	0.199	0.210	-1.714	DCOM	0.091	0.127	-2.190	MAKO	0.136	0.206	-4.452
ANGO	0.073	0.085	-2.309	DELL	0.160	0.164	-0.587	MANT	0.108	0.130	-1.846
APOG	0.121	0.148	-4.928	DIS	0.133	0.157	-3.353	MDCO	0.237	0.293	-2.563
ARCC	0.169	0.208	-2.967	DK	0.070	0.092	-3.459	MELI	0.342	0.377	-1.360
AXP	0.173	0.199	-4.682	DOW	0.273	0.311	-3.260	MFB	0.096	0.167	-7.880
AYI	0.083	0.097	-3.725	EBAY	0.190	0.202	-2.265	MIG	0.083	0.142	-7.604
AZZ	0.109	0.144	-5.839	EBF	0.117	0.158	-9.442	MMM	0.140	0.154	-2.341
BARE	0.011	0.011	0.352	ERIE	0.078	0.098	-3.157	MOD	0.204	0.252	-2.601
BAS	0.237	0.286	-1.567	ESRX	0.227	0.238	-1.011	MOS	0.262	0.289	-2.192
BHI	0.202	0.235	-5.558	EWBC	0.246	0.282	-2.814	MRTN	0.089	0.102	-2.428
BIIB	0.158	0.173	-10.984	FCN	0.182	0.218	-1.357	MXWL	0.178	0.257	-3.394
BRCM	0.179	0.197	-3.092	FFIC	0.106	0.138	-2.839	NC	0.070	0.093	-4.464
BRE	0.109	0.134	-9.684	FL	0.133	0.161	-1.692	NSR	0.083	0.120	-7.097
BW	0.171	0.222	-9.447	FMER	0.135	0.165	-3.270	NUS	0.111	0.136	-2.480
BXS	0.162	0.193	-0.665	FPO	0.089	0.151	-3.798	NXTM	0.311	0.446	-1.948
BZ	0.292	0.485	-5.894	FRED	0.099	0.144	-13.175	PBH	0.099	0.125	-2.074
CB	0.096	0.125	-7.138	FULT	0.186	0.231	-4.287	PFE	0.169	0.180	-1.572
CBEY	0.153	0.214	-3.199	GAS	0.101	0.133	-4.145	PG	0.118	0.128	-3.726
CBT	0.165	0.203	-3.702	GE	0.167	0.181	-2.696	PNC	0.208	0.231	-3.471
CBZ	0.073	0.157	-7.267	GENZ	0.148	0.165	-8.960	PNY	0.079	0.098	-6.363
CCO	0.162	0.202	-2.691	GILD	0.145	0.155	-2.644	PPD	0.077	0.123	-6.712
CDR	0.139	0.194	-2.153	GLW	0.200	0.213	-2.863	PTP	0.063	0.081	-3.135
CELG	0.187	0.208	-5.246	GOOG	0.137	0.155	-6.208	RIGL	0.213	0.283	-3.663
CETV	0.251	0.287	-1.806	GPS	0.141	0.149	-0.658	ROC	0.189	0.256	-11.980
CHTT	0.001	0.001	-0.823	HON	0.165	0.187	-3.041	ROCK	0.359	0.503	-2.047
CKH	0.070	0.120	-3.876	HPQ	0.125	0.139	-2.518	ROG	0.091	0.126	-5.113
CMCSA	0.182	0.188	-0.705	IMGN	0.185	0.236	-3.249	RVI	0.103	0.165	-6.791
CNQR	0.135	0.167	-7.345	INTC	0.171	0.179	-2.499	SF	0.057	0.095	-6.475
COO	0.136	0.161	-6.727	IPAR	0.092	0.142	-2.600	SFG	0.100	0.131	-5.423
COST	0.108	0.119	-3.535	ISIL	0.153	0.175	-2.949	SJW	0.059	0.092	-2.487
CPSI	0.113	0.140	-2.659	ISRG	0.141	0.164	-10.118	SWN	0.267	0.311	-3.582
Overall	0.148	0.182	-15.705								

Table 24: 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.

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								

7 Conclusion

This paper examines high frequency trading and its role in financial markets. HFT make up a large majority of all trades. They supply liquidity in about half of all trades and demand liquidity in about half as well. Their activities i nthe market, both in initiating trades and in providing liqidity, are stable over time. They tend to engage in a price reversal strategy, and this is stronger when they are demanding liquidity. There is no evidence of abusive front running occurring. The HFT firms are profitable, making around \$3 billion between the 26 of them. HFT prefer to trade in large stocks with lower volume, lower spreads and depth, and companies that are considered value firms. They tend to make more money in volatile times. HFT prefer to demand liquidity in small amounts, usually in value between \$1,000 and \$4,999, and they tend to have lower time between trades than non-HFT. They provide the best quotes about 45% of the time. They provide more inside quotes for larger value firms, with lower volume, lower volatility, lower spreads and depth, and with greater number of trades. From the different Hasbrouck measures, the evidence suggest HFT play a very important role in price efficiency and the price discovery process. In fact, they provide more useful information to the price generation process than do non-HFT. Finally, HFT activity either has no impact on volatility or tends to decrease it.

References

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

B. Biais, P. Hillion, and C. Spatt. An empirical analysis of the limit order book and the order flow in the Paris Bourse. *Journal of Finance*, 50(5):1655–1689, 1995.

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. Creswell. Speedy New Traders Make Waves Far From Wall Street. *New York Times*, (http://www.nytimes.com/2010/05/17/business/17trade.html?dbk), May 16, 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.

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.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.

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

Securities and Exchange Commission. Concept Release on Equity Market Structure. (Concept Release on Equity Market Structure), January 14, 2010.

Lash H. Spicer, J. Who's afraid of high-frequency trading. *Reuters*, (http://www.reuters.com/article/idUSN173583920091202), December 2, 2009.

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