The Dissertation Committee for Nicholas Hauschel Hirschey
certifies that this is the approved version of the following dissertation:

Predictability in Order Flow

Committee:

John Griffin, Supervisor

Sheridan Titman, Supervisor

Carlos Carvalho

Clemens Sialm

Laura Starks

Predictability in Order Flow

by

Nicholas Hauschel Hirschey, B.A.; M.S.Fin.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2012

Predictability in Order Flow

Publication No. ____________________

Nicholas Hauschel Hirschey, Ph.D.
The University of Texas at Austin, 2012

Supervisors: John Griffin
Sheridan Titman

High-frequency traders (HFTs) accounted for roughly forty percent of trading volume on the NASDAQ Stock Market in 2009, but there is little evidence on the type of information these investors trade on. This study tests the hypothesis that HFTs anticipate and trade ahead of other investors' order flow. I find that HFTs' aggressive purchases predict future aggressive buying by non-HFTs, and their aggressive sales predict future aggressive selling by non-HFTs. The positive correlation between trading by HFTs and future trading by other investors is robust to the exclusion of trading around news releases, indicating the effect is not driven by HFTs reacting to news announcements faster than other investors. The effects are stronger in the morning and on high volume days. There are also persistent differences among HFTs in the tendency of their trades to predict future order flow. These findings have implications for the speed at which prices adjust to new information, incentives to acquire information, and the price impact of traditional asset managers' trades.

Table of Contents

Abstract **iv**

List of Tables **vii**

List of Figures **viii**

Chapter 1. Do High-frequency traders anticipate buying and selling pressure? **1**

1.1 Introduction 2
1.2 Market Structure Overview 6
1.3 Data 9
1.3.1 Identification of high-frequency traders 9
1.3.2 Sample stocks 10
1.3.3 Trade imbalances 12
1.3.4 Intra-day returns 13
1.4 Do trades from HFTs lead trades from non-HFTs? 14
1.5 Alternative explanations 18
1.5.1 Controlling for serial correlation and returns 18
1.5.2 Excluding trading around news events 23
1.5.3 Inventory management 26
1.6 Why is non-HFT order flow predictable? 27
1.7 Conditioning on times when HFTs are likely to be impatient 30
1.7.1 VAR estimates near the market open and close 31
1.7.2 VAR estimates on high volume and high imbalance days 33
1.7.3 VAR estimates in high versus low spread stocks 34
1.8 Cross-sectional differences in prediction ability 35
1.9 Conclusion 38

Appendices **69**

Appendix A. Supplementary Tables **70**

Appendix B. Heteroskedasticity in VAR estimates **89**

Bibliography **93**

List of Tables

1.1 Summary of Stock Characteristics 47
1.2 Summary of Stock-Day Observations 48
1.3 Summary of News Data 49
1.4 Non-HFT Net Marketable Buying for Stocks Sorted by HFT Net Marketable Buying 50
1.5 Returns for Stocks Sorted by HFT Net Marketable Buying 52
1.6 Intra-day VAR Estimates for Individual Stock-day Observations 54
1.7 VAR Estimates on Days with and without News 57
1.8 Predicting Order Flow with Returns and Quotes 59
1.9 VAR Estimates at Different Times of the Day 62
1.10 VAR Estimates on High Volume and High Imbalance Days 64
1.11 VAR Estimates for Stocks with High versus Low Spreads 66
1.12 Examining Cross-sectional Differences in Prediction Ability 68

A.1 Summary of CRSP Universe 71
A.2 Sample Universe 72
A.3 Detailed Summary of Stock-Day Observations 73
A.4 Returns for Stocks Sorted by High-Frequency Trader Net Buying 74
A.5 Robustness of VAR to Price Feed Latency 77

B.1 Different weighting schemes for VAR coefficients 92

List of Figures

1.1 Cumulative HFT vs. non-HFT Net Marketable Buying. 41
1.2 Returns 42
1.3 Post-sort Returns By Size Portfolios. 43
1.4 Response of non-HFT Net Marketable Buying to a One Standard Deviation Shock to HFT Net Marketable Buying 44
1.5 HFT Net Buying and Returns 60 minutes before and after intense HFT Net Marketable Buying 45
1.6 Transition Probabilities 46

A.1 HFTs' Share of NASDAQ Dollar Volume 80
A.2 Liquidity Removing Trades as a Percent of HFT Dollar Volume 81
A.3 Market Share By Trading Venue 82
A.4 Coefficients from VAR with 30 lags 83
A.5 Cumulative HFT Net Buying vs. non-HFT Net Marketable Buying 84
A.6 HFT Net Buying and Returns 60 minutes before and after intense HFT Net Marketable Buying By Size Portfolio 87
A.7 Response of non-HFT Net Marketable Buying to a One Standard Deviation Shock to HFT Net Marketable Buying: Different Times of the Day 88

B.1 Heteroskedasticity in stock-day VAR coefficients 91

Chapter 1

Do High-frequency traders anticipate buying and selling pressure?

1.1 Introduction

Most trading in equity markets today is automated, and a large portion of these automated trades originate from short-term investors known as high-frequency traders (HFTs). These HFTs account for a substantial fraction of equity market trading volume, including roughly 40% of NASDAQ dollar volume in 2009.[1] Since HFTs are market makers on many exchanges, one would expect market making to account for much of their trading activity. A simple way to get a rough lower bound on the share of their trading that is not due to market making is to split HFT volume into the fraction due to liquidity providing versus liquidity removing trades. Doing so reveals that a little over half of HFTs' dollar volume is actually due to trades in which HFTs remove liquidity from the order book with a marketable order.[2,3] While HFTs' liquidity providing trades very likely benefit other investors, the effect of their liquidity removing trades is unclear.

This paper examines whether HFTs' liquidity removing trades arise from strategies that anticipate and trade ahead of traditional asset manager order flow. Since HFTs' expertise lies in extracting information from price and order data feeds, they seem well suited to the type of analysis that could reveal information about future order flow. If, indeed, an HFT were able to forecast a traditional asset managers' order flow, then the HFT may have an incentive to trade ahead of the traditional asset manager in order to profit from their

[1] Appendix Figure A.1 displays HFTs' share of dollar volume for all stocks trading on the NASDAQ Stock Market.

[2] All orders on the NASDAQ Stock Market are limit orders. A marketable buy order is a buy order with a limit price at least as high as the best ask, and a marketable sell order is a sell order with a limit price at least as low as the best bid. Thus, marketable orders are functionally equivalent to market orders.

[3] Appendix Figure A.2 reports HFT dollar volume in liquidity removing executions as a percent of total HFT dollar volume.

subsequent price impact.

Anticipatory trading of this form has the potential to affect both liquidity and price efficiency. An HFT taking liquidity that non-HFTs intend to access could cause stock prices to move against non-HFTs right before they trade, thereby increasing non-HFT trading costs. Moreover, if HFTs anticipate informed non-HFT trades, then the result is equivalent to an increase in the number of informed investors, which speeds the rate at which prices adjust to the non-HFTs' information. But since HFTs anticipating informed non-HFT trades capture profits that would otherwise accrue to non-HFTs, the long-run effect could be a reduction in information production by non-HFTs. For these reasons, it is important to understand the extent to which HFTs anticipate and trade ahead of other investors' order flow.

To examine these issues, I analyze return and trade patterns around periods of aggressive buying and selling by HFTs using unique trade data from the NASDAQ Stock Market. Specifically, I focus on HFTs' aggressive trades, that is, trades where an HFT initiates the transaction by submitting a marketable buy or sell order, because it is a simple way to screen out liquidity providing trades. I test whether HFTs' aggressive share purchases predict future aggressive buying by non-HFTs, and whether HFTs' aggressive sales predict future aggressive selling by non-HFTs.

I find evidence consistent with HFTs being able to anticipate order flow from other investors. In tests where stocks are sorted by HFT net marketable buying at the one second horizon, the stocks bought most aggressively by HFTs have cumulative standardized non-HFT net marketable buying of 0.66 over the following thirty seconds. For the median stock, this equates to non-HFTs buying roughly 28 more shares with marketable orders than they sell with marketable orders over the next thirty seconds.[4] The figures for stocks HFTs sell most aggressively are similar, but in the opposite direction. Moreover, the stocks HFTs buy aggressively have positive future returns, and the stocks they sell aggressively have negative future returns.

I consider several explanations for these findings. One possibility is the results are driven by HFTs responding to news faster than other investors. I test this hypothesis by examining the lead-lag relationship between HFT and non-HFT net marketable during periods containing no stock-specific news. HFT net marketable buying continues to lead non-HFT net marketable buying, even when there is no news about a stock. A second explanation is that HFT and non-HFT trading are driven by the same underlying serially correlated process (i.e., same trading signals), so HFT trading predicts non-HFT order flow only because it is a proxy for lagged non-HFT trading. However, HFT net marketable buying remains positively correlated with non-HFT net marketable buying after controlling for serial correlation in non-HFT trading. A third explanation is that if non-HFTs chase price trends, HFTs might actually cause future trading by non-HFTs through their effect on returns. But controls for lagged returns do not drive the relationship between HFT and non-HFT trading to zero, which is inconsistent with this third hypothesis.

I also examine whether there are cross-sectional differences in how well different HFTs' trades forecast future order flow. Perhaps some HFTs are more skilled or focus more on strategies that anticipate order flow, while others focus on market making or index arbitrage. There are indeed persistent differences among HFTs. Trades from HFTs that were the most

[4]The median standard deviation of a stock's non-HFT net marketable buying, available in Table 1.2, is 42. The 28 share figure comes from multiplying 0.66 by 42.

highly correlated with future order flow in a given month have trades that also exhibit stronger than average correlation with future non-HFT order flow in later months.

While other papers have examined HFTs, this is the first paper to find that 1) HFTs anticipate buying and selling pressure from other investors, 2) some HFTs are better than others at anticipating order flow, and 3) while HFTs are contrarian at longer horizons, their aggressive trades chase returns over the prior three seconds. Brogaard (2011a) studies the determinants and characteristics of HFT activity, including the relative probability an HFT execution is immediately followed by a non-HFT execution in the same direction, but he finds conflicting results for whether HFTs trade ahead of other investors. Hendershott and Riordan (2011b) find that HFTs tend to buy prior to permanant price increases and sell prior to permanent price declines. Kirilenko, Kyle, Samadi, and Tuzun (2011) study HFTs' trades in the E-mini futures contract during the May 6$^{\text{th}}$, 2010, "Flash Crash" and find that HFTs neither took large long nor large short positions as the market crashed. Brogaard (2011b) studies how HFT activity is related to volatility. O'Hara, Yao, and Ye (2011) examine price discovery through odd-lot trades, which they show are commonly used by HFTs. This paper builds on this literature by exploring whether HFTs use information about future order flow to predict price changes.

This study also contributes to a small but growing literature on how market participants' adoption of algorithmic trading strategies affects markets. Hendershott, Jones, and Menkveld (2011) find that an increase in electronic message traffic on the NYSE is associated with lower bid-ask spreads and less price discovery through trades. Hendershott and Riordan (2011a) show that trades and quotes entered by algorithms on the Deutsche Boerse are more informative about permanent price changes than those entered by humans, and algorithms supply relatively more liquidity when spreads are wide. Chaboud, Chiquoine, Hjalmarsson, and Vega (2009) study liquidity provision and price discovery by algorithmic traders in foreign exchange markets. Hasbrouck and Saar (2011) find that trade and quote activity by traders using low-latency strategies on NASDAQ decrease spreads, increase depth, and lower volatility. In contrast to the present paper and the aforementioned HFT studies, since the above studies do not have investor identifiers, they cannot distinguish algorithmic trading by HFTs from that of other investors. This paper contributes to this literature by specifically focusing on lead-lag relationships between trades of these two types of investors.

The structure of the paper is as follows. Section 1.2 provides an overview of NASDAQ market structure and how it relates to other trading venues. Section 1.3 discusses the trade and return data. Section 1.4 examines whether HFT net marketable buying leads non-HFT net marketable buying. Section 1.5 examines alternative explanations for the results on the relationship between HFT and non-HFT net marketable buying. Section 1.6 explores the hypothesis that HFTs might use past returns and quotes to predict non-HFT order flow. Section 1.7 examines lead-lag effects between HFT and non-HFT trading in scenarios when non-HFTs are hypothesized to be impatient. Section 1.8 examines cross-sectional differences in the correlation between HFT firms' trades and non-HFT net marketable buying. Finally, section 1.9 concludes.

1.2 Market Structure Overview

There have been a number of changes to equity market structure since the mid-1990s. In light of these changes, it is helpful to explain NASDAQ's structure and how it compares to other trading venues.

One of the biggest changes is the increase in competition among trading venues. Equities are no longer restricted to trading on their listing exchange, and this has spurred entry by new exchanges and other trading venues. Evidence of the increased competition may be seen in the fact that NASDAQ's share of dollar volume in 2009 was roughly 36% in NASDAQ-listed securities and 17% in securities listed on the NYSE.[5] The remainder of U.S. equity trading is spread among trading venues with displayed order books, such as the NYSE and BATS, and trading venues with non-displayed order books, such as ITG's POSIT Marketplace, Credit Suisse's Crossfinder, and Knight Capital.

A number of papers examine how this competition among various trading venues affects liquidity. O'Hara and Ye (2011) show that stocks with proportionately more trading occuring in off-exchange venues have lower spreads. Jovanovic and Menkveld (2010) find that when an alternative trading venue catering to HFTs, Chi-X, begins trading Euronext stocks, spreads decline. Foucault and Menkveld (2008) find that market depth increases after the London Stock Exchange's entry into the market for Dutch stocks traditionally traded on Euronext. These recent studies, which focus on entry by trading venues that cater to electronic trading, build on an existing literature studying competition among trading venues more generally (e.g., Battalio, Greene, and Jennings 1997, Mayhew 2002).

Another consequence of the increase in trading venues is some degree of market segmentation. Executions in displayed markets predominately come from professional traders. Few retail orders reach the displayed markets directly. Most retail brokerages have contracts with market making firms who pay for the right to fill retail orders. For example, in the third

[5]Appendix Figure A.3 shows the time series of NASDAQ's market share by listing venue.

quarter of 2009, Charles Schwab routed more than 90% of its customers' orders in NYSE-listed and NASDAQ-listed stocks to UBS's market making arm for execution (Schwab 2009). Similarly, E*Trade routed nearly all its customers' market orders and over half its customers' limit orders to either Citadel or E*Trade's market making arms (E*Trade 2009). However, when there is a large imbalance between retail buy and sell orders in a stock, market making firms likely offload the imbalance by trading in displayed markets, so there is some interaction between retail trading demand and the displayed markets. This means most trades sent to NASDAQ come from professional traders.

Another broad theme has been exchanges' replacement of open outcry trading pits with electronic limit order books. Even the NYSE is now essentially fully electronic. This transition to electronic trading seems to have improved liquidity. One line of research utilizes improvements in exchange trading technology to instrument for increases in automated trading. Riordan and Storkenmaier (2009) study the effects of an upgrade to the Deutsche Boerse that reduced latency and find that quoted and effective spreads decline. Hendershott and Moulton (2011) find that a change to the NYSE's trading system that improved the speed of electronic trading and reduced the advantage of floor-based traders relative to electronic traders increased spreads and improved price efficiency. A general theme in this literature is that faster trading technology allows electronic market makers to quote narrower spreads, because they are able to update their quotes faster in response to new information. This shift towards electronic limit order books also means that NASDAQ, itself an electronic market, now has essentially the same market structure as all other displayed trading venues.

HFTs are among some of the most active participants on electronic exchanges. HFTs are typically proprietary trading firms using high-turnover automated trading strategies.

While estimates of their share of equity trading vary among sources, all estimates indicate HFTs are a large part of the market. The TABB Group LLC, for example, estimated that that HFTs accounted for 61% of U.S. Equity share volume in 2009 (Tabb 2009). HFTs are active outside the U.S. as well, with some estimates suggesting HFTs account for as much as 77% of U.K. trading (Sukumar 2011). Examples of such traders include Tradebot Systems, Inc., and GETCO. These firms are remarkably active traders. On their websites, Tradebot says they often account for more than 5% of total U.S. equity trading volume, and GETCO says they are "among the top 5 participants by volume on many venues, including the CME, Eurex, NYSE ARCA, NYSE ARCA Options, BATS, NASDAQ, NASDAQ Options, Chi-X, BrokerTec, and eSpeed" (Tradebot 2010, GETCO 2010). Such firms likely engage in some combination of market making and statistical arbitrage.

1.3 Data

This study primarily uses intra-day transactions data obtained from the NASDAQ Stock Market, which covers all equities traded on NASDAQ, including listings from the NASDAQ, NYSE, AMEX, and ARCA exchanges. The sample period is January 1 through December 31, 2009.[6] This section describes the data and its characteristics.

1.3.1 Identification of high-frequency traders

Trade records on exchanges include a Market Participant Identifier (MPID) indicating the broker/dealer making the trade. A broker/dealer may have multiple MPIDs that are used by different business lines or customers. A typical reason for a customer to have

[6] I exclude January 27$^{\text{th}}$, because quote data for NYSE-listed stocks is missing.

their own MPID is they have sponsored access, an arrangement where the customer handles connectivity with the exchange and typically has limited interaction with the broker/dealer's trading system. These MPIDs allow exchanges and regulators to identify which trading firms execute each trade.

The data from NASDAQ classifies market participants as either an HFT or a non-HFT. Firms were classified as HFT firms using a variety of qualitative and quantitative criteria. The firms classified as HFTs typically use low-latency connections and trade more actively than other investors. Their orders have shorter durations than other investors, and they show a greater tendency to flip between long and short positions in a stock during a day.

1.3.2 Sample stocks

The sample stocks are chosen to be representative of those in which actively managed traditional asset managers invest. The sample is constructed from CRSP common stocks, identified by stocks having share code 10 or 11. Dual-class stocks are eliminated, because differences in ticker symbol conventions across databases make matching stock observations from different databases based on ticker symbols harder for dual-class stocks.[7] A reasonable definition of the set of stocks in which traditional asset managers invest is those in either the Russell 3000 or the MSCI Investable Market 2500. These indexes contain the top 3,000 and top 2,500 stocks ranked on market capitalization. This roughly corresponds to the top eight size deciles, so I exclude stocks in the bottom two size deciles from the sample. These

[7] Appendix Table A.1 summarises stock-day observations of CRSP common stocks with dual-class shares removed.

restrictions limit the sample to 2,792 common stocks at the end of 2008. To ensure sample stocks are fairly liquid, I require average daily dollar volume in December 2008 to be greater than $1 million and that the stock price at the end of 2008 is greater than $5. These two liquidity restrictions further reduce the sample universe to 1,882 stocks.[8] Another benefit of the stock price restriction is that it ensures the minimum tick-size is not too large a percentage of the stock price, which can be important when examining short-term returns. A large tick-size relative to stock price would require a relatively large change in stock value for a change in price to occur. From the sample universe of 1,882 stocks, I create the sample of 96 stocks used in this study by randomly selecting 6 NASDAQ-listed and 6 NYSE-listed stocks from each of the eight size deciles.

Table 1.1 reports summary statistics for all stock days. Stocks are removed from the sample any day during which the prior day's closing price is less than $1. They are permanently removed if daily dollar volume falls below $100,000. Thus, the sample averages 93 stocks per trading day. Market capitalization ranges from $22.1 million to $125,330.6 million. The median small-cap stock's price is $14.77, compared to $25.04 for mid-cap stocks and $31.37 for large-cap stocks. Share and dollar volume increase as market capitalization rises as well. Median dollar volume for small-cap stocks, for example, is $1.9 million, compared to $120.2 million for large-cap stocks. As expected, small-cap stocks are more volatile than mid and large-cap stocks. The standard deviation of small-cap stocks' daily returns is 4.5%, compared to 3.5% for mid-cap stocks and 2.5% for large-cap stocks. On average, 27.2% of the sample stocks' dollar volume trades on NASDAQ, and this value is fairly constant across size portfolios.

[8] Appendix Table A.2 summarises stock-day observations for this sample of stocks.

HFTs are relatively more active in large-cap stocks. Their median share of total dollar volume is 14.8% in small-cap stocks, 29.2% in mid-cap stocks, and 40.9% in large-cap stocks. It is conceivable that since HFTs' comparative advantage is reacting quickly to market events, they find more profit opportunities in stocks for which quoted prices and depths update frequently.

1.3.3 Trade imbalances

The study uses two measures of trade imbalances: marketable imbalances and buy-sell imbalances. The marketable imbalance is a common measure of buying and selling pressure from the existing literature (e.g., Chorida, Roll, and Subrahmanyam 2002). The simplest explanation of marketable imbalances is that they are shares in buyer-initiated trades minus shares in seller-initiated trades. If investors with resting limit orders in the order book are passive liquidity suppliers, then the marketable imbalance is an intuitive measure of trading demand. The buy-sell imbalance is simply shares bought minus shares sold and has been used to measure position changes of different investor groups (e.g., Griffin, Harris, and Topaloglu 2003). To put trade imbalances on a similar scale across stocks, I normalize all imbalance measures by a stock's 20-day trailing volume from CRSP.

Table 1.2 summarizes trade imbalances for the sample stocks. The table describes the distribution of the stock-day standard deviations of HFTs' net buying, their net marketable buying, their net marketable buying when it is the same direction as their net buying, and non-HFTs' net marketable buying. For HFTs' net marketable buying in the same direction as their net buying, positive values of HFT net marketable buying are set to zero if net buying is less than the fourth quintile, and negative values are set to zero if net buying is greater than the second quintile. The purpose of the net marketable buying measure that requires net buying to be in the same direction is that it ensures variation is driven by times when HFTs are either on net buying and buying aggressively or on net selling and selling aggressively.

In practice, there is little difference between the two HFT net marketable buying measures. The mean standard deviation of HFTs' net buying among all stock days is 83 shares, compared to 80 shares for their net marketable buying and 76 shares for their net marketable buying when it is the same direction as their net buying. The average standard deviation of non-HFTs' net marketable buying, at 125 shares, is somewhat higher than that of HFTs. Predictably, the standard deviation of the imbalances measures is larger for stocks with larger market capitalizations. The mean standard deviation of HFTs' net marketable buying, for example, is 11 shares in small-cap stocks and 178 shares in large-cap stocks.

1.3.4 Intra-day returns

Intra-day returns are calculated using bid-ask midpoints from two sources. The primary source for quotes is the National Best Bid and Best Offer (NBBO). The NBBO aggregates quotes from all displayed order books.[9] The second source for quotes is the NASDAQ best bid and best offer (NASDAQ BBO). While the NASDAQ BBO excludes quotes on other exchanges, it has the advantage that the timestamps are precisely aligned with the

[9] The largest displayed order books are the NYSE, NASDAQ, AMEX, Archepelago, BATS, and Direct-Edge.

trade data. These quote data are filtered to remove anomalous observations.[10]

Table 1.2 reports the distribution of the standard deviation of NBBO bid-ask midpoint returns across all stock days. The mean standard deviation is 0.05% among small-cap stocks, 0.09% among mid-cap stocks, and 0.11% among large-cap stocks. These returns may be converted to daily standard deviations by mulitiplying them by the square root of the number of one-second periods in the trading day, $\sqrt{6.5 \times 3600}$. The mean daily standard deviations of returns are then 8.01% among small-cap stocks, 13.01% among mid-cap stocks, and 17.44% among large-cap stocks. These standard deviations seem quite large, but the means are inflated by volatile stock-day observations.

1.4 Do trades from HFTs lead trades from non-HFTs?

This section begins the examination of whether HFTs anticipate buying and selling pressure from other investors. HFTs may anticipate the trades of a mutual fund, for instance, if the mutual fund splits large orders into a series of smaller ones and the initial trades reveal information about the mutual funds' future trading intentions. HFTs might also forecast order flow if traditional asset managers with similar trading demands do not all trade at the same time, allowing the possibility that the initiation of a trade by one mutual fund could forecast similar future trades by other mutual funds. If an HFT were able to forecast a traditional asset managers' order flow by either these or some other means, then the HFT could potentially trade ahead of them and profit from the traditional asset manager's

[10] I remove quote updates where the bid is greater than the ask or where the bid-ask spread is more than 20% greater than the bid-ask midpoint. To remedy bad pre-market quotes in the NYSE data, the last of which is used to proxy for the opening price, I throw out the last price before the open if there is more than a 20% difference between the last pre-open bid-ask midpoint and the first post-open bid-ask midpoint.

subsequent price impact.

There are two main empirical implications of HFTs engaging in such a trading strategy. The first implication is that HFT trading should lead non-HFT trading—if an HFT buys a stock, non-HFTs should subsequently come into the market and buy those same stocks. Second, the HFT's whole objective is to profit from the non-HFT's subsequent price impact, so it must be the case that HFT trades precede profitable price changes. These two patterns, together, are consistent with HFTs trading stocks in order to profit from non-HFTs' future buying and selling pressure.

Table 1.4 reports results from tests of the above hypothesis using portfolio sorts. The table reports average non-HFT net marketable buying for stocks sorted on HFTs' net marketable buying. Stocks are sorted into deciles at time t based on HFT net marketable buying in the same direction as net buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. The results are reported for all stocks as well as separately for small, mid, and large-cap stocks. Non-HFTs' net marketable buying is averaged across all observations for a day, and the mean of the daily time series is reported in the table. Parentheses indicate Newey and West (1994) t-statistics for the time-series means.

The results show that HFT net marketable buying is positively correlated with lagged, contemporaneous, and future net marketable buying from non-HFTs. In small-cap stocks, in the thirty seconds before the sort period, investment fund net marketable buying is -1.12 for stocks sold aggressively by HFTs and 1.11 for stocks bought aggressively. Going forward, in the five minutes after the sort, non-HFT net marketable buying is -3.22 times the one-second standard deviation of net marketable buying in the stocks sold aggressively and 2.55 times the one-second standard deviation in the stocks bought aggressively. The patterns are the same in mid and large-cap stocks, though the magnitude of non-HFT net marketable buying declines as market capitalization increases.

The top panel in Figure 1.1 shows this relationship in graphical form. The figure plots cumulative net marketable buying for HFTs and for non-HFTs for stocks in net marketable buying portfolios one and ten. The portfolios are formed in the same way as they are in Table 1.4. The figure shows that at time zero, the sort period, net marketable buying spikes for both investor groups. Afterwards, net marketable buying for HFTs is relatively flat, whereas for non-HFTs, cumulative net marketable buying continues to increase in portfolio ten and fall in portfolio one. The figure clearly shows that HFTs' net marketable buying leads net marketable buying from non-HFTs.

Table 1.4 and Figure 1.1 show that HFTs' net marketable buying leads future net marketable buying by other investors. If HFTs make these trades in anticipation of non-HFT buying and selling pressure, then we should also see that the stocks that are bought aggressively have postive future returns and that the stocks that are sold aggressivley have negative future returns.

The top panel in Figure 1.2 shows returns for stocks in HFT net marketable buying portfolios one and ten around the sort period. The figure shows that stocks bought aggressively by HFTs subsequently have positive returns, while those sold aggressively subsequently have negative returns. These post-sort return patterns are likely caused by the buying and selling pressure from non-HFTs illustrated in Figure 1.1.

Table 1.5 reports the magnitude of these returns across all stocks and for each size portfolio. Thirty seconds after the sort period, basis point returns for small-cap stocks that were bought versus sold are 2.54 versus -2.59, compared to 1.52 and -1.28 for mid-cap stocks and 0.41 and -0.14 for large-cap stocks. These returns are all significantly different from zero, though the magnitude of small and mid-cap stocks' post-sort returns and t-stats are much larger than those of large-cap stocks. Since small-cap stocks are generally more illiquid, they may provide a greater opportunity for predicting price pressure. By five minutes after the sort period, prices of large-cap stocks sold by HFTs are actually higher than at the end of the sort period, and prices of the stocks they bought are essentially the same as at the end of the sort period. Cumulative returns for large-cap stocks that were bought versus sold aggressively are 0.08 versus 0.67. However, in small and mid-cap stocks, cumulative returns for stocks that were bought aggressively remain positive, and returns of stocks that were sold aggressively remain negative. Cumulative returns in the five-minutes after the sort period for small cap stocks that were bought versus sold aggressively are 1.32 versus -2.71, compared to 1.17 versus -0.34 for mid-cap stocks. Figure 1.3 illustrates post-sort returns across size portfolios in more detail. The figure indicates that the reversal in large-cap stocks begins immediately after the initial post-sort price change. Small and mid-cap stocks that were bought aggressively continue to briefly rise after they are bought aggressively and fall after they are sold aggressively by HFTs. By about fifty seconds after the sort period, prices of small and mid-cap stocks begin to partially reverse. These return patterns, in combination with the patterns in net marketable buying shown in Table 1.4 and Figure 1.1, are consistent with HFTs anticipating buying and selling pressure from traditional asset managers that moves prices.

1.5 Alternative explanations

Section 1.4 presented results showing HFT trades lead trades from non-HFTs as well as returns. While these findings are consistent with HFTs anticipating buying and selling pressure, there are other potential explanations for these results. This section tests several alternative explanations for why HFT net marketable buying leads non-HFT net marketable buying.

1.5.1 Controlling for serial correlation and returns

This section evaluates two alternative explanations for the lead-lag relationship between HFT and non-HFT net marketable buying. The first alternative is that HFT and non-HFT trading might be driven by the same serially correlated process, in which case HFT trading predicts non-HFT trading because it is a noisy proxy for lagged non-HFT trading. This might be the case if, for example, HFTs and non-HFTs use the same trading signals. If this explanation is driving the lead-lag relationship between HFT and non-HFT trading, then the lead-lag relationship between the two variables will go away after controlling for lagged non-HFT trading. A second alternative is that the lead-lag relationship between HFT and non-HFT trading is due to a predictable relationship with past returns. This is essentially a reverse causality story. If non-HFTs follow trend-chasing strategies, then purchases by HFTs could actually cause future non-HFT trading through their effect on returns. This explanation predicts that HFT trading will be uncorrelated with future non-HFT trading after controlling for lagged returns. This section controls for these confounding effects using vector autoregressions (VAR).

The VAR is a system of three equations in which lags of returns, HFT net marketable buying, and non-HFT net marketable buying are all used to explain each other. The equation with non-HFTs' net marketable buying as the dependent variable is the primary focus. This equation isolates the predictive ability of HFTs' aggressive trades, controlling for serial correlation in non-HFTs' net marketable buying and past returns.

The VAR is estimated separately for each stock every day and includes ten lags of each variable. All variables are divided by their standard deviation among all stocks for that day to ease interpretation. Panel A in Table 1.6 summarizes coefficient estimates from these VARs. The panel reports the average of each coefficient as well as the percent that are positive or negative and significant. This is a simple way to summarise the VAR results, but it does not distinguish between effects that are consistent across days and effects that exist on only a few days. To check the consistency of effects across days, I also calculate the mean of each coefficient every day, and perform a t-test on the time-series mean of daily mean coefficients. These results are reported in Panel B.

This section is motivated by concerns about the confounding effects of serial correlation and trend-chasing by non-HFTs. If these effects are present, then the coefficients on lagged non-HFT net marketable buying and lagged returns in the equation where non-HFT net marketable is the dependent variable will be positive. Indeed, the coefficients on lagged non-HFT net-marketable buying are positive, indicating positive serial correlation. Coefficients on lagged non-HFT net marketable buying decline from 0.074 at lag one to 0.012 at lag ten.[11] All coefficients are more likely to be positive and significant than negative and sig-

[11] High-frequency traders' net marketable buying is also serially correlated, though to a lesser degree than non-HFTs'. A one standard deviation increase in HFT net marketable buying leads to a 0.026 standard deviation increase in the same variable the next period. Coefficients decline with additional lags to 0.001 at lag 10.

nificant, and the time-series means of average daily coefficients are all significantly different from zero, with t-statistics ranging from 27.63 to 41.18. Turning to returns, the coefficient on lag one returns is the largest of all those in the VAR specification. A one standard deviation increase in returns leads to a 0.86 standard deviation increase in the next period's non-HFT net marketable buying imbalance. Coefficients on returns at lags two through ten are much smaller. The large positive coefficient on lag one returns suggests non-HFTs are trend-chasing at short horizons.[12,13] In summary, the coefficients on lagged non-HFT trading and lagged returns in Table 1.6 show controls for serial correlation and trend chasing by non-HFTs are warranted.

The main question, then, is whether HFT net marketable buying is still correlated with future non-HFT net marketable buying after these controls. In fact, as was the case for the portfolio sorts in Table 1.4, HFT net marketable buying is positively correlated with future net marketable buying from other investors in Table 1.6. A one standard deviation increase in HFT net marketable buying on average leads to a 0.0023 standard deviation change in non-HFT net marketable buying the next period.[14] The lag one coefficient is positive and significant 24.92 percent of the time and negative and significant 16.59 percent

[12]Other interpretations include market makers anticipating a forthcoming net marketable imbalance and adjust prices accordingly or traders submitting aggressive limit orders prior to submitting marketable orders, thereby moving the bid-ask midpoint in the direction of future marketable trades.

[13] One concern might be that the apparent trend-chasing behavior could be driven by misaligning trade and NBBO quote time-stamps. Appendix Table A.5, which uses NQBBO quotes, shows using precisely aligned timestamps does not change any of these conclusions.

[14] There is also short-horizon correlation between non-HFT net marketable buying and future HFT net marketable buying. When HFT net marketable buying is the dependent variable, the lag one coefficient on non-HFT net marketable buying is 0.005 and positive and significant 22.9 percent of the time. The lead-lag relationship is less persistent than that between HFT net marketable buying and future non-HFT net marketable buying—by lag five, the coefficients are much smaller and the time-series means of the coefficients in Panel B are not consistently significantly different from zero.

of the time. The average coefficient on lags two through ten declines slowly, to a minimum of 0.0016 at lag ten.[15] The lag two through ten coefficients are between 1.7 and 2.1 times more likely to be positive and significant than negative and significant. Panel B shows that the time-series of daily means is positive and significantly different from zero at all lags. These findings indicate aggressive buying by HFTs is followed by aggressive buying by non-HFTs, and vice versa for aggressive selling, even after controlling for past returns and past non-HFT aggressive buying.

Similarly, the results for the relationship between HFT net marketable buying and future returns is the same as in the sorts in Section 1.4. A one-standard deviation increase in lag one HFT net marketable buying leads to a 0.018 standard deviation increase in the next period return. The coefficient is much more likely to be positive and significant than negative and significant, and the time-series mean of the daily average coefficients are significantly different from zero. Coefficients on additional lags of HFT net marketable buying are also positive, though by lag ten they are no longer significantly different from zero.

The economic magnitude of the individual coefficients on HFTs' net marketable buying looks small at first glance. But the important thing is the effect of a shock to HFT net marketable buying on non-HFT net marketable buying accumulated over all lags. This cumulative effect can be determined by calculating impulse response functions.

Figure 1.4 uses impulse response functions to plot the response of non-HFT net marketable buying to a one standard deviation shock to HFT net marketable buying. Impulse

[15] Appendix Figure A.4, which plots coefficients for the VAR using thirty lags, shows that the coefficients on HFT net marketable buying continue to decline towards zero at higher lags.

response functions are first calculated for all stocks separately each day.[16] The stock-day impulse response functions are then averaged across all stocks on a day to create a time-series of daily cross-sectional average impulse response functions. The figure plots the time-series mean of the daily cross-sectional impulse response functions as well a 95% confidence interval calculated using standard errors from the daily time series of mean impulse response functions. The figure indicates the average cumulative effect on non-HFT net marketable buying after thirty seconds is 0.052 times the one-second standard deviation.

An additional point to note is that HFTs also exhibit short horizon trend chasing. Coefficients in Table 1.6 on returns at lags one through four are positive, while those on lags five through ten are negative. A one standard deviation increase in returns leads to a 1.631 standard deviation increase in HFT net marketable buying the next period. The coefficient is positive and significant 85.6 percent of the time. Coefficients on lags two through four are 0.034, 0.001, and -0.007. Coefficients five through ten are negative, though only lags eight, nine, and ten have time-series means of daily coefficients that are significantly different from zero. These results indicate HFTs chase very short-term price trends, but at longer horizons they are contrarian.

This section used a VAR framework to test whether HFT trading leads non-HFT trading because either HFT trading is a noisy proxy for serially correlated non-HFT trading or non-HFTs are chasing returns caused by HFT trades. Consistent with the sort results in

[16]The impulse response function is orthogonalized to allow for contemporanous effects among the variables. Contemporaneous effects are included, because HFT and non-HFT trading affect contemporaneous returns. The calculation is structured such that HFT net marketable buying has a contemporaneous effect on non-HFT net marketable buying and returns, non-HFT net marketable buying has a contemporaneous effect on returns, and returns do not have a contemporaneous effect on either of the trading variables.

Section 1.4, HFT trading is positively correlated with future non-HFT trading in the VAR specification. Thus, the explanations examined in this section do not appear to be driving the lead-lag relationship between HFT and non-HFT trading.

1.5.2 Excluding trading around news events

Another alternative explanation for the finding that HFT net marketable buying leads non-HFT net marketable buying is that HFTs simply react to news faster than other investors. To rule out this alternative hypothesis, I reexamine the sort and VAR results after excluding periods around news announcements. First, I redo the sorts after excluding the five minutes before and after intra-day news announcements. Next, I examine VAR estimates on days with and without news, where a news day is alternately defined as either a day when a news article about the stock is published or a day when the absolute value of a stock's return is greater than two percent. All three methods indicate the lead-lag relationship between HFT and non-HFT trading is not attributable to HFTs simply reacting faster to news announcements.

The news articles used in this section were collected from the Factiva news archive. The Factiva archive is remarkably complete. Factiva contains news from over 35,000 sources, including most major newswires, newpapers, and magazines. Factiva tags articles with identifiers indicating which firms are covered in an article, and these identifiers are used to match articles to the firms in this study's sample.

Table 1.3 describes the sample of news articles. The are an average of 762 articles per stock, and the average number of days a stock has an article about it published is 130. All articles contain the date an article is published, but some also include a time stamp for the article. The sample of articles with time stamps is somewhat smaller and predominately consists of newswires. If the sample of articles is restricted to those with time stamps occuring during the trading day, the number of articles per stock and days with news fall to 70 and 29. Panel B reports the top sources for time-stamped articles, and, not surprisingly, the list consists entirely of newswires.

The first test reexamines the sort results in Table 1.4 after excluding periods that are within five minutes of a minute during which an article about the stock is published. Doing so restricts the news sample to articles with timestamps. This restriction unfortunately removes news that is only revealed in articles without time stamps, but if a news article contains timely information on which investors will immediately trade, then the news should be picked up by newswires, which contain timestamps and so are in the sample. Panel B in Table 1.4 reports sorts of non-HFT net marketable buying after periods near intra-day news are removed, and the results are nearly identical to those in Panel A. Similarly, Panel B in Figure 1.1, which excludes trading in periods near news announcements, shows the same pattern as the full sample. Thus, the sorts provide no evidence that the lead-lag relationship between HFT and non-HFT trading is driven by HFTs reacting faster to news.

However, excluding trading in the five minutes before and after Factiva news articles may not be sufficient to exclude all news trading events. This methodology will miss trading around news events if either Factiva timestamps are wrong or if Factiva doesn't include all types of news. To address the possibility that Factiva timestamps might be wrong, I estimate the VAR from Table 1.6 after excluding a stock any day a news article about it is published. To address the possibility that HFTs are reacting faster than non-HFTs to news that isn't in Factiva (e.g., analyst forecasts), I also estimate the VAR excluding stocks any day when the absolute value of the stock's return is greater than two percent. The idea is that extreme returns are a proxy for news that captures events missing from standard databases.

Panel A in Table 1.7 reports VAR estimates for days with and without stock news, where news is defined as articles in the Factiva news archive. The panel reports estimates for coefficients on lags of HFT net marketable buying in regressions where the dependent variable is non-HFT net marketable buying. The primary focus is determining whether coefficients on lags of non-HFT net marketable buying are positive after excluding trading on days with news. The middle columns contain estimates for non-news days. The average non-news day lag one coefficient on HFT net marketable buying is 0.0027 and, with a t-statistic of 7.06, significantly different from zero. Lags two through ten and the sum of all ten lags are also positive and significantly different from zero. In general, the coefficients on news and non-news days are similar and not significantly different from each other. These results are consistent with HFT net marketable buying forecasting non-HFT net marketable buying on days when there is no news for a stock.

Panel B has the same structure, except news days are defined as days when the absolute value of a stock's market-adjusted return is greater than one percent. The coefficient estimates on days with small returns are all positive and significantly different from zero. These findings are inconsistent with the hypothesis that the lead-lag relationship between HFT and non-HFT trading is driven by trading on news events that are not in Factiva.

This section tested whether the explanation for why HFT trading forecasts non-HFT trading is that HFTs react faster than non-HFTs to news announcements. I identified trading during times with no news in three different ways, and in all three cases, HFT net marketable buying remains positively correlated with future non-HFT net marketable buying. These findings are inconsistent with the lead-lag relationship between HFT and non-HFT trading being driven by HFTs reacting faster to news announcements.

1.5.3 Inventory management

Inventory management is another potential explanation for HFT trading leading non-HFT trading. This explanation implies that HFTs happen to dispose of inventory positions accumulated in the process of making markets immediately before non-HFT buying and selling pressure.

A counter argument is the only way for HFTs to systematically dispose of inventory positions immediately prior to adverse price moves caused by non-HFT buying and selling pressure is if the HFTs anticipate the impending non-HFT trading. Inventory management based on anticipated adverse price changes caused by non-HFT trading is the same as HFTs trading ahead of anticipated non-HFT order flow.

Another way to address this alternative is to examine whether there is evidence HFTs build inventory positions prior to their aggressive trades. Figure 1.5 plots cumulative HFT net buying for the first and tenth net marketable buying portfolios in Table 1.4 from 60 minutes before to 60 minutes after the sort period. If HFTs' trades in the sort period are unwinding previous positions accumulated while making markets, then for the stocks HFTs are buying aggressively at time zero, one would expect them to have previously been net sellers of those stocks. If this is going on, then in the figure, the dotted line indicating net position changes for stocks HFTs sell aggressively at time zero should rise, and the solid line indicating position changes for stocks HFTs buy at time zero should fall. The lines in the hour before the sort period are mostly flat and close to zero, providing no evidence HFTs are building inventory positions that they later dispose of at time zero.

Another way to evaluate whether HFTs are building inventory positions in the pre-sort period that offset their trades in the sort period is to look at returns. The HFT net buying measure is calculated from NASDAQ data, so it is possible the calculation misses inventory positions being built on other trading venues. Inferences based on returns, illustrated in the bottom panel of Figure 1.5, do not have this problem. If sales during the sort period are disposing of shares acquired due to liquidity provision during the prior 60 minutes, then one would expect negative returns during the prior 60 minutes. But returns from 60 minutes to 1 minute before aggressive HFT sales are positive, which is not what one would expect if HFTs were providing buy-size liquidity during that period. The returns for stocks HFTs buy aggressively at time zero are analogous. Thus, the HFT net buying and return patterns in the 60 minutes before the sort period do not support the hypothesis that the HFT trades at time zero are disposing of inventory positions built in the prior 60 minutes in the course of providing liquidity. Therefore, there seems to be little support for the argument that the lead-lag relationship between HFT and non-HFT trading is caused by inventory management that is not triggered by HFTs anticipating price changes caused by future non-HFT buying and selling pressure.

1.6 Why is non-HFT order flow predictable?

One explanation for why HFT trades forecast non-HFT trades is HFTs are able to model non-HFT order flow as a function of past returns and quotes. This section explores this explanation by examining the extent to which non-HFT order flow may be predicted by a simple linear model of past returns and bid-ask depth imbalances. Past returns could predict future non-HFT order flow if non-HFTs chase price trends, follow return reversal strategies, or have persistent order flow that causes price impact. The bid-ask imbalance could predict future order flow if, for instance, a large bid size relative to the ask size indicates pent up buying demand from non-HFTs.

Table 1.8 reports results from these tests. Panel A displays results from regressions in which non-HFT net marketable buying is alternately regressed on four lags of the bid-ask quote imbalance, returns, HFT net marketable buying in the same direction as net buying, and non-HFT net marketable buying. Subsequent panels examine regressions including various combinations of these lagged variables.

The hypothesis that a positive bid-ask imbalance might reflect built up buying pressure implies positive coefficients on lags of the bid-ask quote imbalance in the first regression in Panel A. While coefficients on the second and third lags of the bid-ask imbalance are positive and significantly different from zero, the coefficient on the first lag is negative, significant, and much larger than the coefficients on the other lags. This negative coefficient contradicts the hypothesis that a positive bid-ask imbalance forecasts future non-HFT buying pressure.

There are a couple possible explanations for the negative lag one coefficient in regression one. One reason a small bid size relative to the ask size might forecast buying pressure is if the small bid indicates the bid was just raised. A second possibility is that the small bid could mean limit buy orders at the best bid were just cancelled and are being switched to marketable buy orders. Regardless the explanation, the low average adjusted R^2 relative to the other regressions in Panel A indicates the bid-ask size imbalance is a fairly weak predictor of non-HFT net marketable buying.

Returns turn out to be a better predictor of non-HFT net marketable buying. Coefficients on all four lags of returns in regression two are positive and significant, and the average adjusted R^2, 0.0006, is much larger than that of the bid-ask size imbalance regression. Since non-HFT net marketable buying is serially correlated, as indicated in regression four, and likely to cause price impact, it is possible the predictability picked up by past returns is due only to price impact from past non-HFT net marketable buying. Regression six in Panel B, which adds non-HFT net marketable buying as a control, shows that much of the predictive power of returns is independent of the information in past non-HFT net marketable buying. The fact that the predictive power of lagged returns and lagged non-HFT net marketable buying are independent can be seen by the fact that the adjusted R^2 in regression six is nearly the sum of the adjusted R^2 in regressions two and four. Thus, returns seem a good candidate for the foundation of a model used by HFTs to predict non-HFT net marketable buying. If HFTs mainly use returns to predict non-HFT net marketable buying, then returns and HFT net marketable buying will largely contain redundent information. In other words, adding HFT net marketable buying to regression two will not meaningfully increase the adjusted R^2 if they contain redundant information. In fact, as regression five shows, the adjusted R^2 from the regression having both returns and HFT net marketable buying is nearly the sum of the adjusted R^2 in regressions two and three, indicating much of the preditive ability of HFT net marketable buying is independent of that in returns. Similarly, regression eight shows that the contribution of HFT net marketable buying and returns to the adjusted R^2 relative to the regression with only lags of non-HFT net marketable buying, regression four, is the same as the sum of the marginal contributions of the two variables in regressions six and seven.

Table 1.8 reports regressions testing the hypothesis that returns and the lagged bid-ask quote size imbalance predict non-HFT net marketable buying and that the information in these variables is correlated with the information in lagged HFT net marketable buying. The results show that while the bid-ask quote size imbalance is correlated with future non-HFT net marketable buying, the correlation is opposite that expected and the explanatory power of the variable is orders of magnitude smaller than that of returns and HFT net marketable buying. It turns out that returns and HFT net marketable buying have similar predictive power, but the variation explained by returns is fairly independent of the variation explained by HFT net marketable buying. Thus, while returns in part explain future non-HFT net marketable buying, much of the predictive power of HFT net marketable buying comes from something other than returns.

1.7 Conditioning on times when HFTs are likely to be impatient

Prior sections demonstrated that HFT net marketable buying leads both non-HFT net marketable buying and returns. These findings suggest HFTs are able to anticipate and trade ahead of order flow from non-HFTs. If this is true, then perhaps it may be easier to anticipate non-HFT trades when non-HFTs are relatively impatient. This section uses three methods for identifying times when non-HFTs are hypothesized to be relatively impatient and examines whether HFT trades are more strongly correlated with future trades from non-HFTs at these times.

The methodology involves comparing estimates of the VAR in section 1.5.1 at times when non-HFTs are hypothesized to be relatively impatient to estimates from normal times. The focus is comparing the size of coefficients on lagged HFT net marketable buying in the regression where the dependent variable is non-HFT net marketable buying. Larger positive coefficients at times when non-HFTs are hypothesized to be impatient are consistent with HFTs having an easier time anticipating order flow.

1.7.1 VAR estimates near the market open and close

The first set of tests uses the open and close of trading to instrument for times when non-HFTs are impatient. To see why investors might be impatient at the open, imagine an investor who receives a signal overnight. The investor knows that either other investors received the same signal or will receive it shortly. Therefore, the investor knows they need to trade early in order to profit from that information. Investors may be impatient near the close for related reasons. An investor may need to trade a position before the close, because a) they have private information about an impending news announcement or b) they are facing a liquidity shock. If non-HFTs are more impatient at the open or the close, then I expect HFTs will have an easier time forecasting their order flow.

Table 1.9 reports results comparing trading in the first and last half hours of the trading day to trading in the middle of the day. Panel A looks at results for the first half hour of the trading day. The average coefficient on the first lag of HFT net marketable buying during the morning is 0.0040, compared to 0.0012 in the middle of the trading day. The difference is significantly different from zero, with a t-statistic of 6.05 In fact, all the coefficients are larger in the morning than in the middle of the trading day. One way to get a sense for the overall difference is to look at the sum of the coefficients on all ten lags of HFT net marketable buying. The sum of all lags in the morning is 0.0224, compared to 0.0147 in the middle of the day. These results indicate HFT net marketable buying is more strongly correlated with future non-HFT net marketable buying in the morning than in the middle of the trading day.

Panel B examines results during the last half hour of the trading day. In contrast to earlier results, the coefficient on the first lag of HFT net marketable buying near the close is negative, -0.0029. The median of the daily mean coefficients, -0.0030, is also negative and close to the mean, indicating the result is not driven by an outlier. The difference with the average coefficient in the middle of the day, -0.0041, is significantly different from zero, with a t-statistic of -8.87. Coefficients on the next few lags of HFT net marketable buying, though positive, are also less than those in the middle of the trading day. Thus, HFT net marketable buying at the close does not exhibit a stronger positive correlation with non-HFT net marketable buying than that during the middle of the trading day.

The results in the morning are consistent with non-HFTs being more impatient at the open, making it easier for HFTs to forecast their order flow. However, the results at the close are inconsistent with the hypothesis that HFTs are better able to forecast order flow. There is something different about the close. One possibility is that HFTs are less aggressive near the close, because they do not want to build an inventory position that they do not have time to dispose of before the close. This explains weaker effects at the close, but not the negative coefficient on the first lag of HFT net marketable buying. An explanation for the negative coefficient could be that HFTs still anticipate order flow, but that rather than selling aggressively in anticipation of selling pressure, they buy aggressively to dispose of an inventory position. That said, this explanation for the negative coefficient on the first lag of HFT net marketable buying near the close is not fully satisfying.

1.7.2 VAR estimates on high volume and high imbalance days

High volume days are another possible time when non-HFTs might be more impatient. High volume days are likely days when certain investors are trading large postions. When an investor needs to trade a large position, it is potentially harder for them to hide with noise traders. In other words, they may stick out more, making it easier for HFTs to forecast their order flow. Thus, high volume or large marketable imablance days could be used to identify days when a non-HFT is impatient because they are trading a large position.

Table 1.10 compares VAR estimates on high volume or high imbalance days to normal days. As in section 1.7.1, the coefficients being compared are those on lags of HFT net marketable buying from the equation where the dependent variable is non-HFT net marketable buying. High volume and high imbalance days are identified using a methodology similar to that of Gervais, Kaniel, and Mingelgrin (2001). A day's CRSP volume or the absolute value of the day's aggregate net marketable buying imbalance is ranked relative to the prior 19 trading days. If the day's rank is among the two highest during the 20-day ranking period, then the day is marked a high volume or high imbalance day. All other days are considered normal days.[17]

Panel A examines high volume days, and Panel B examines high imbalance days. Both panels tell a similar story. The coefficient on the first lag of HFT net marketable buying is 0.0049 on high volume days and 0.0016 on normal volume days. The 0.0033 difference between the two is significantly different from zero, with a t-statistic of 4.27. The next few lags on high volume days remain higher than those on normal volume days, but at higher lags

[17] The volume and imbalance rankings are completely independent of each other, so a normal volume day in the volume tests, for example, could be a high imbalance day in the imbalance tests.

there is no significant difference between the coefficients on high volume and normal volume days. Similarly, Panel B shows that on high imbalance days, the coefficient on the first lag of HFT net marketable buying is 0.0036, which is 0.0018 higher than on normal imbalance days. The t-statistic from the test that the two coefficients are equal is 2.48, indicating we can reject the hypothesis that there is no difference between the two. Looking at the sum of all ten lags, the difference between the sums on high-volume days and normal volume days is significantly different from zero, but the t-statistic from the test of the difference between the sum of all lags on high imbalance and normal imbalance days is only 1.89. Overall, there appears to be a stronger correlation between HFT net marketable buying and future net marketable buying by non-HFTs on both high volume and high imbalance days.

1.7.3 VAR estimates in high versus low spread stocks

HFTs may also have an easier time forecasting order flow in illiquid stocks. The intuition is that if non-HFTs do not perfectly scale position sizes relative to liquidity, then in illiquid stocks, they will have larger relative positions than in liquid stocks. When non-HFTs trading illiquid stocks enter and exit these larger relative positions, it may be harder to hide future trading demand than would be the case in liquid stocks (i.e., it is harder to hide when one is a bigger part of the market).

Table 1.11 tests this hypothesis by comparing VAR estimates of how strongly HFT net marketable buying is correlated with future net marketable buying from non-HFTs in high bid-ask spread versus low bid-ask spread stocks. Bid-ask spreads are calculated in two ways: Panel A uses bid-ask spreads, while Panel B uses relative bid-ask spreads, which are spreads divided by the bid-ask midpoint. Normal bid-ask spreads are in some ways more intuitive, but they do not account for the fact that liquid high-priced stocks may have wide nominal spreads. In dividing by the bid-ask midpoint, relative spreads address this issue.

The results for spreads and relative spreads both indicate the correlation between HFT net marketable buying and future non-HFT net marketable buying is stronger in illiquid stocks. The lag one coefficient in high-spread stocks is 0.0053, compared to 0.0000 in low-spread stocks. The difference between these coefficients is statistically significant, with a t-statistic of 11.18. The sum of coefficients on lags one through ten is also larger for high versus low spread stocks. Similarly, the lag one coefficient for high relative spread stocks is 0.0046, is higher than the lag one coefficient in low relative spread stocks, 0.0012. The sum of coefficients on the first ten lags in high relative spread stocks, 0.0214, is also higher than that in low relative spread stocks, 0.0179. As was the case in Panel A, it is really the case that coefficients only on the first few lags of HFT net marketable buying are higher in high spread stocks. These results are consistent with HFTs being able to better forecast non-HFT order flow in illiquid stocks.

1.8 Cross-sectional differences in prediction ability

Prior results indicate aggregate HFT net marketable buying leads non-HFT net marketable buying. It is possible that among HFTs, some firms' trades are strongly correlated with future non-HFT order flow, while other firms' trades have little or no correlation with non-HFT order flow. This may be the case if certain HFTs focus more on strategies that anticipate order flow or if some HFTs are more skilled than other firms. To examine this issue, this section tests whether trades from HFTs whose trades are most strongly correlated with future non-HFT order flow in one month continue to have higher than average correlation with future non-HFT order flow in future months. The advantage of looking at the question in this way, that is, in looking at persistence in the ability to predict buying and selling pressure rather than looking at full sample cross-sectional differences in ability, is that it accounts for the fact that in any given period, some HFTs will look better than others due to chance.

High-frequency traders' ability to predict buying and selling pressure is calculated using regressions similar to those used in the VAR analysis in section 1.5.1. Each day, for each HFT, I estimate two regressions. In the first regression, I regress non-HFT net marketable buying on ten lags of the HFT's net marketable buying, ten lags of non-HFT net marketable buying, and ten lags of returns. High-frequency traders' net marketable buying is required to be in the same direction as their net buying in the stock; If the HFT's net buying is negative when net buying is positive, or net buying is positive when net marketable buying is negative, then net marketable buying for the period is set to zero. Returns and non-HFT net marketable buying are divided by their standard deviation, and HFTs' net marketable buying is divided by the standard deviation of aggregate HFT net marketable buying that day. The second regression is the same, except the HFT's net buying is substituted for their net marketable buying. The heading for Table 1.12 contains the regression equation.

High-frequency traders' ability to predict buying and selling pressure is measured in two ways: first, by the average coefficient on the first lag of the HFT's net marketable buying or net buying, and, second, by the average sum of the coefficients on all ten lags of their net marketable buying or net buying. A positive coefficient means the HFT's trades are positively correlated with future non-HFT order flow. I take the mean of each ability measure across all days in a month for each HFT and sort the sample into three groups based on the magnitude of the HFTs' ability measures.

One simple way to look at persistence is to look at the probably an HFT in the highest correlation group remains in that group in future months. Figure 1.6 plots the probability an HFT who is in the highest-correlation group will again be in the highest-correlation group one, two, and three months later. Since there are three groups, if being in the highest-correlation group is random, then the probably a firm that was in highest-correlation group one month will be in the highest-correlation group the next month is 33.3%. So under the null hypothesis of no persistent difference among HFTs, in the first month after the sort period, only 33.3% of the HFTs should still be in the highest-correlation group. In fact, regardless whether HFTs are sorted by only the first or by all lags of HFT net marketable buying or net buying, between 57% and 78% of the HFTs are still in the highest-correlation group one month later. This simple test illustrates that some HFTs' trades are consistently more strongly correlated with non-HFT order flow than are trades from other HFTs.

Another way to examine persistence is to compare post-sort month ability measures for the three HFT groups. If the ability measures are persistent, then the highest-correlation group should continue to have larger average coefficients than the lowest-correlation group in the post-sort month. Table 1.12 reports average post-sort month ability measures for the three HFT groups. Results for regressions using HFTs' net marketable buying are in the first three columns and those for regressions using HFTs' net buying are in the last three columns.

The results in Table 1.12 indicate there are persistent differences among HFTs in both the correlation between their net marketable buying and non-HFTs' future net marketable buying and the correlation between their net buying and non-HFTs' future net marketable buying. The first group of columns in the top half of the table examine the persistence of the coefficient on the first lag of HFTs' net marketable buying, $\overline{\gamma_{d,t,1}}$. The average $\overline{\gamma_{d,t,1}}$ for the highest-correlation group is 0.014, compared to 0.002 for the lowest-correlation group. The p-value from a test of the hypothesis that the the time-series of monthly differences between the two groups equals zero is 0.006, indicating the difference between the two groups is persistent. The first three columns in the bottom half of Table 1.12 show results using ten lags of HFTs' net marketable buying, rather than just the first lag. The average sums of the first ten lags are greater than the average first lags for all three groups, indicating lags two through ten are on average positive. As was the case for the test using just the first lag, the difference between the highest and lowest-correlation groups in the post-sort month is significantly different from zero. The last three columns in Table 1.12 report results from the tests using HFTs' net buying rather than net marketable buying. The results from these tests are the same as for the net marketable buying tests—there are persistent differences between the highest and lowest-correlation groups. One may conclude from these results that trades from some HFTs are more correlated with future order flow than are trades from other HFTs.

1.9 Conclusion

This study examines the relationship between high-frequency traders' aggressive trades and future order flow from other investors. I find that aggressive buying of a stock by HFTs predicts future aggressive buying by non-HFTs, and aggressive selling by HFTs predicts future aggressive selling by non-HFTs. This finding is consistent with HFTs trading ahead of anticipated price changes caused by non-HFTs' future buying and selling pressure. I explore several alternative explanations for these findings, including their being driven by serial correlation in non-HFT order flow, non-HFTs chasing return trends, HFTs reacting faster than non-HFTs to news, and inventory management. However, the anticipatory trading hypothesis proves robust to all these alternative explanations. In addition, I examine whether there are differences among HFTs in how well their trades forecast future order flow. There are, in fact, persistent differences among HFTs in terms of how well their trades forecast future buying and selling pressure. These findings suggest HFTs trade on forecasted price changes caused by buying and selling pressure from traditional asset managers.

These findings have implications for price efficiency and stock liquidity. If HFTs anticipate and trade ahead of informed trades, then they will cause that information to become incorporated into prices more quickly (Holden and Subrahmanyam 1992). However, in doing so, HFTs capture some of the informed traders' trading profits and, consequently, decrease their incentive to become informed. Thus, a benefit due to an increase in the speed at which information is reflected in prices would be reduced by the fact it decreases investors' incentives to acquire new information. Alternately, if HFTs anticipate and trade ahead of liquidity trades, as in Brunnermeier and Pedersen (2005), then in these instances they will increase price impact and, consequently, increase traditional asset managers' transaction costs.

On net, it is probable HFTs have a positive impact on market quality. The emergence of HFTs has coincided with a substantial decrease in quoted and realized spreads on exchanges (Castura, Litzenberger, Gorelick, and Dwivedi 2010). They make markets in numerous financial securities and drove floor traders out of business by quoting narrower bid-ask spreads. But the results of this study suggest that in the process of predicting future returns, HFTs at times acquire information about other investors' future order flow, and in these instances likely increase those investors' trading costs.

This study suggests multiple areas for further research. While the results suggest HFTs forecast non-HFT order flow, it is still unclear what information may be used to predict order flow. One possibility is order flow predictability is driven by cross-correlations caused by delayed reaction to common information. It is also possible a more sophisticated analysis of the order book could reveal information useful for predicting order flow. Additionally, this paper demonstrates that some HFTs' trades have a stronger tendency to predict future order flow. This may indicate that the HFTs whose trades are most strongly correlated with future non-HFT order flow are more skilled. One way to test this would be to see if trades from the HFTs among those whose trades are most strongly correlated with future order flow forecast larger returns than do trades from other HFTs. It would also be worthwhile knowing whether HFTs are predominately anticipating buying and selling pressure driven by liquidity shocks or by information about fundamentals. This can be determined by looking at whether initial price changes following aggressive buying and selling by HFTs eventually reverse. The evidence in this study is consistent with HFTs anticipating informed trades in small and mid-cap stocks, but the topic warrants further research. Another area to explore is the effect of changing the way time is measured. All current tests use clock time, but since prices and orders change much more quickly in active stocks than in inactive stocks, it is possible marking off time by a set number of trades, known as trade time, is a more useful notion of time. Research into these questions would contribute both to a better understanding of the results in this study and of the characteristics of high-frequency returns and liquidity in general.

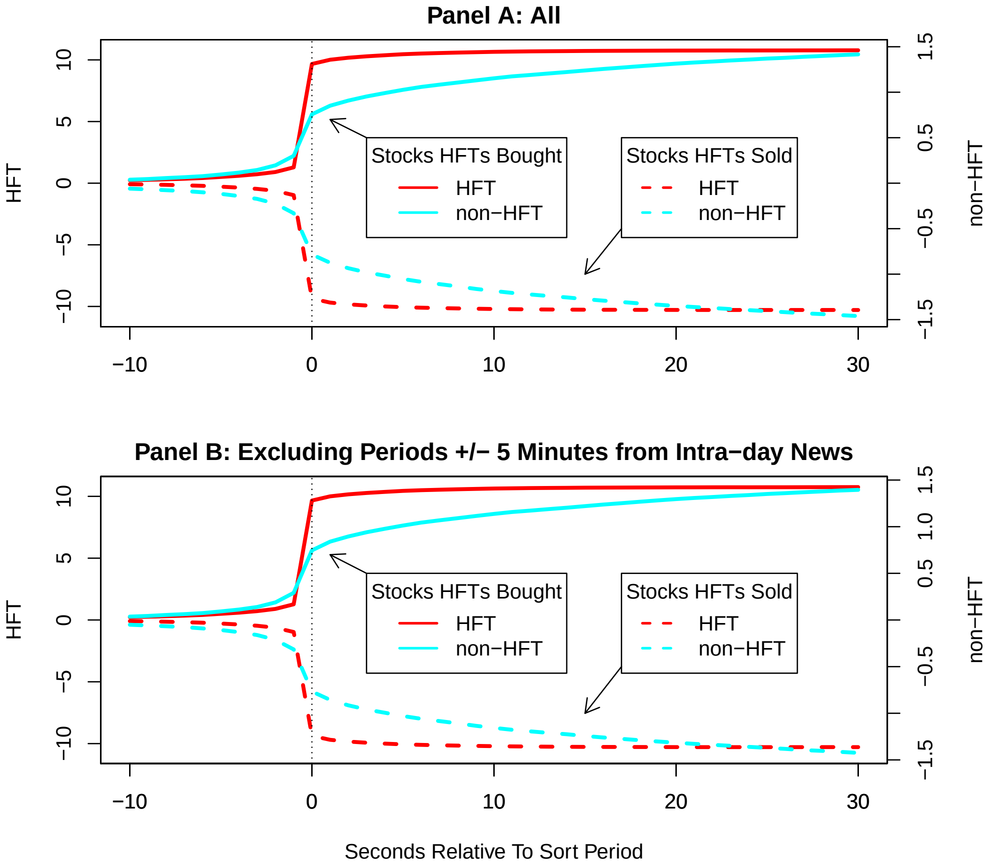


Figure 1.1: Cumulative HFT vs. non-HFT Net Marketable Buying.
This figure plots cumulative standardized net marketable buying for stocks sorted into portfolios by HFTs' net marketable buying. Panel A uses all intra-day periods. Panel B excludes stocks that have a news article about them published within five minutes of the sort period. Table 1.3 describes the news data. The left y-axis is for HFT net marketable buying in the same direction as their net buying, HFT_{NMBSD}, and the right y-axis is for non-HFT net marketable buying, $non\text{-}HFT_{NMB}$. Table 1.2 describes construction of these imbalance measures. Stocks are sorted into deciles based on HFT net marketable buying using decile breakpoints from the prior trading day. Stocks in decile ten and for which HFT_{NMBSD} is greater than zero are marked as those HFTs bought. Stocks in decile one and for which HFT_{NMBSD} is less than zero are marked as those HFTs sold. The reason for conditioning on HFT_{NMBSD} rather than just HFT_{NMB} is that it ensures variation is driven by times when HFTs are either on net buying and buying aggressively or on net selling and selling aggressively.

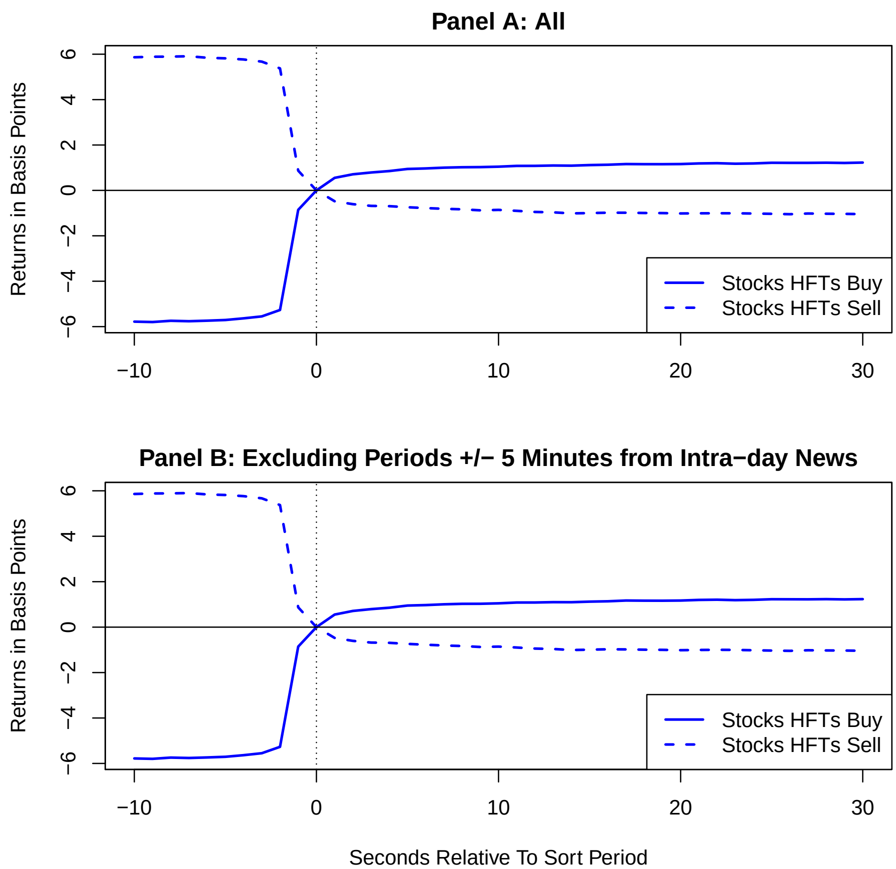


Figure 1.2: Returns

This figure plots returns for stocks sorted into portfolios by HFTs' net marketable buying. Panel A uses all intra-day periods. Panel B excludes stocks that have a news article about them published within five minutes of the sort period. Table 1.3 describes the news data. Stocks are sorted into deciles based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten and for which HFT_{NMBSD} is greater than zero are marked as those HFTs bought. Stocks in decile one and for which HFT_{NMBSD} is less than zero are marked as those HFTs sold. The reason for conditioning on HFT_{NMBSD} rather than just HFT_{NMB} is that it ensures variation is driven by times when HFTs are either on net buying and buying aggressively or on net selling and selling aggressively. Table 1.2 describes construction of these imbalance measures.

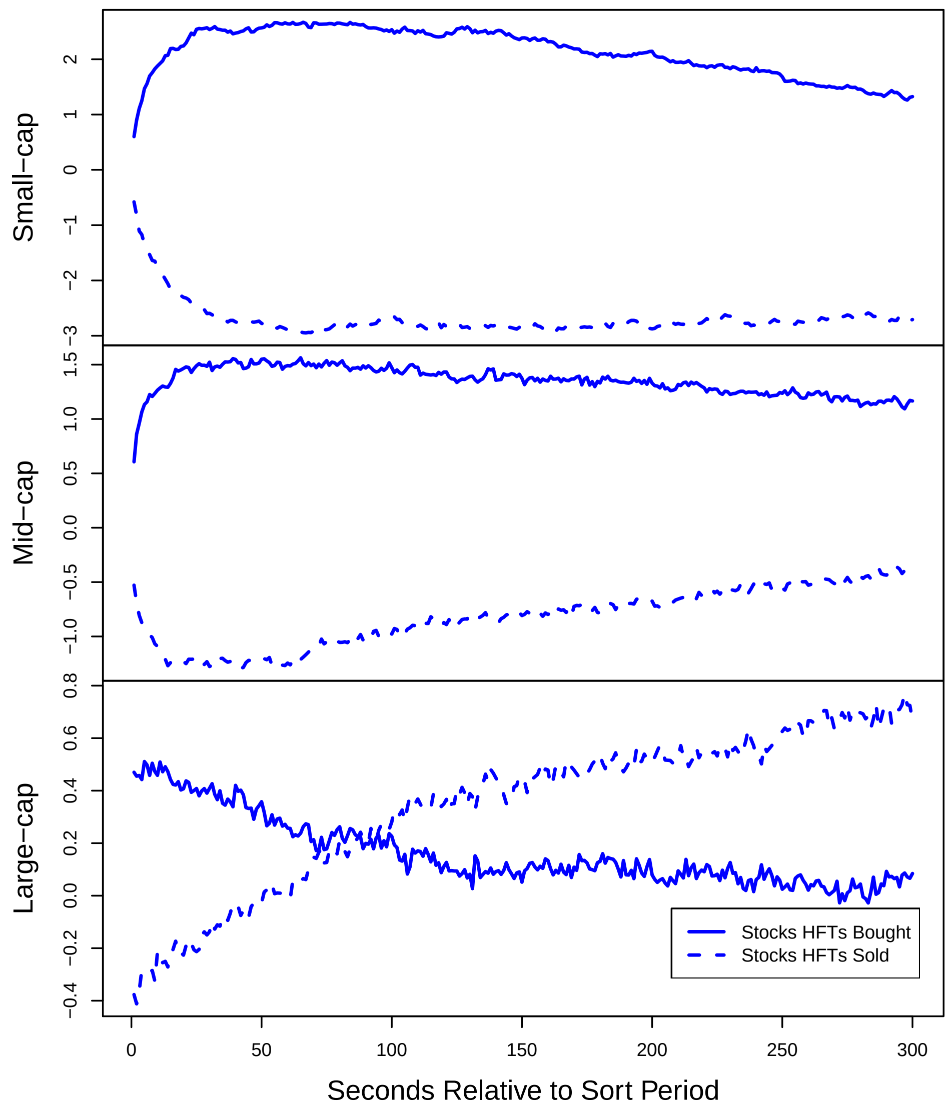


Figure 1.3: Post-sort Returns By Size Portfolios.

The y-axis scale is returns in basis points. Stocks are sorted into deciles based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten and for which HFT_{NMBSD} is greater than zero are marked as those HFTs bought. Stocks in decile one and for which HFT_{NMBSD} is less than zero are marked as those HFTs sold. The reason for conditioning on HFT_{NMBSD} rather than just HFT_{NMB} is that it ensures variation is driven by times when HFTs are either on net buying and buying aggressively or on net selling and selling aggressively. Table 1.2 describes construction of these imbalance measures.

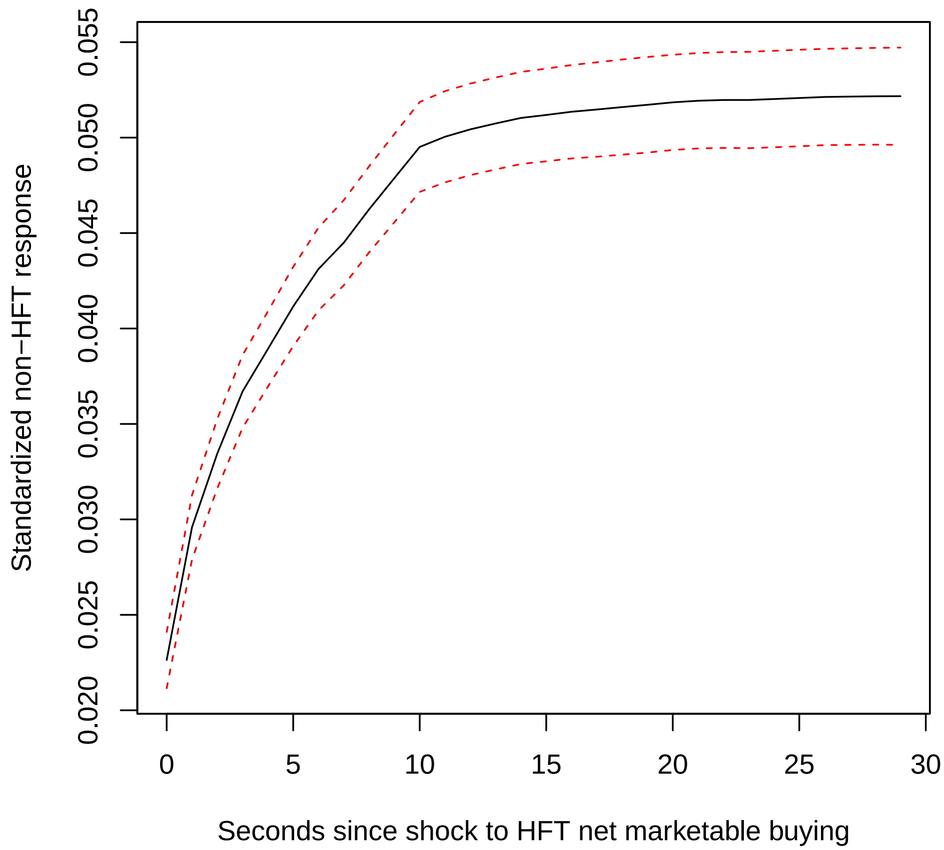


Figure 1.4: Response of non-HFT Net Marketable Buying to a One Standard Deviation Shock to HFT Net Marketable Buying

This figure plots the impulse response function describing the response of non-HFT net marketable buying, $non\text{-}HFT_{NMB}$, to a one standard deviation shock to HFT net marketable buying in the same direction as net buying, HFT_{NMBSD}. Table 1.2 describes construction of these imbalance measures. The response is expressed in standard deviations. The results are based on the vector autoregression (VAR) in Table 1.6. Stock-day observations are excluded if any of the variables fail an augmented Dickey-Fuller test for stationarity. The impulse response function is orthogonalized to allow for contemporaneous effects. The ordering of the variables is such that HFT_{NMBSD} has a contemporaneous effect on $non\text{-}HFT_{NMB}$ and returns. $non\text{-}HFT_{NMB}$ has a contemporaneous effect on returns but does not contemporaneously affect HFT_{NMBSD}. Returns are assumed to have no contemporaneous effect on either trading measure. Impulse response functions are estimated by stock each day, and then the daily cross-sectional mean is calculated. The solid line is the mean of the daily time series, and the dotted lines indicate 95% confidence intervals using standard errors calculated from the daily time-series of mean impulse response functions.

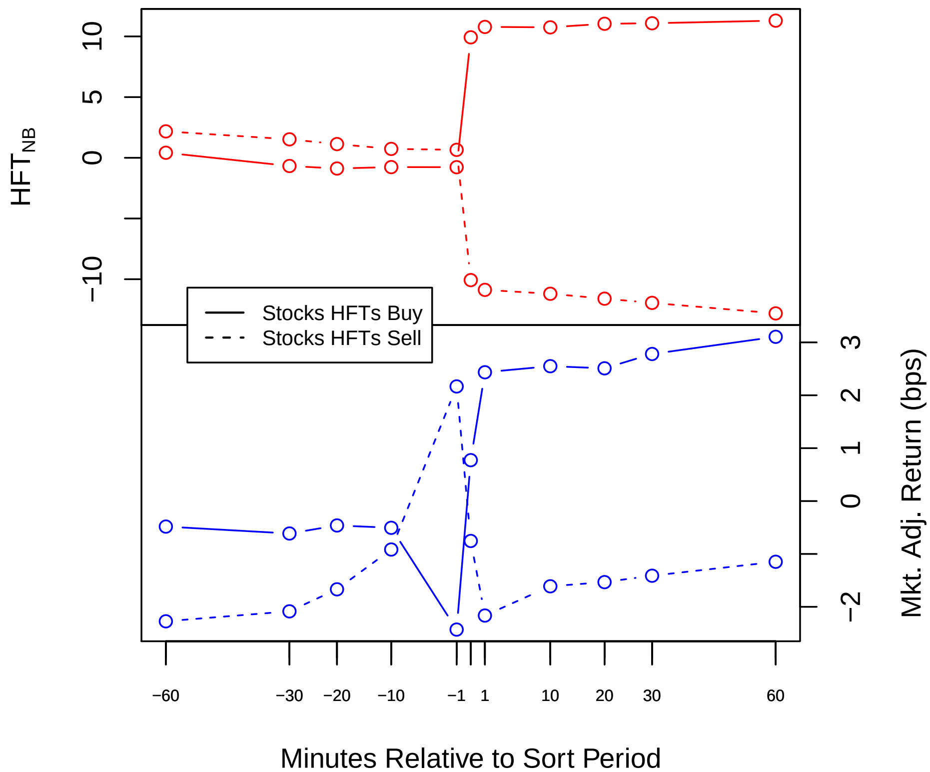


Figure 1.5: HFT Net Buying and Returns 60 minutes before and after intense HFT Net Marketable Buying

The figure examines HFT net buying from 60 minutes before to 60 minutes after periods of intense HFT net marketable buying or selling. HFT_{NB} is cumulative standardized HFT net buying. Buy and hold returns are market adjusted using contemporaneous returns on SPY. Stocks are sorted into deciles based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten and for which HFT_{NMBSD} is greater than zero are marked as those HFTs bought. Stocks in decile one and for which HFT_{NMBSD} is less than zero are marked as those HFTs sold. The reason for conditioning on HFT_{NMBSD} rather than just HFT_{NMB} is that it ensures variation is driven by times when HFTs are either on net buying and buying aggressively or on net selling and selling aggressively. To handle clustering of observations, observations are first averaged by stock-day, then by day, and then finally across the complete time-series. Observations must have data from 60 minutes before to 60 minutes after the sort period, so the figure excludes the first and last hour of the trading day.

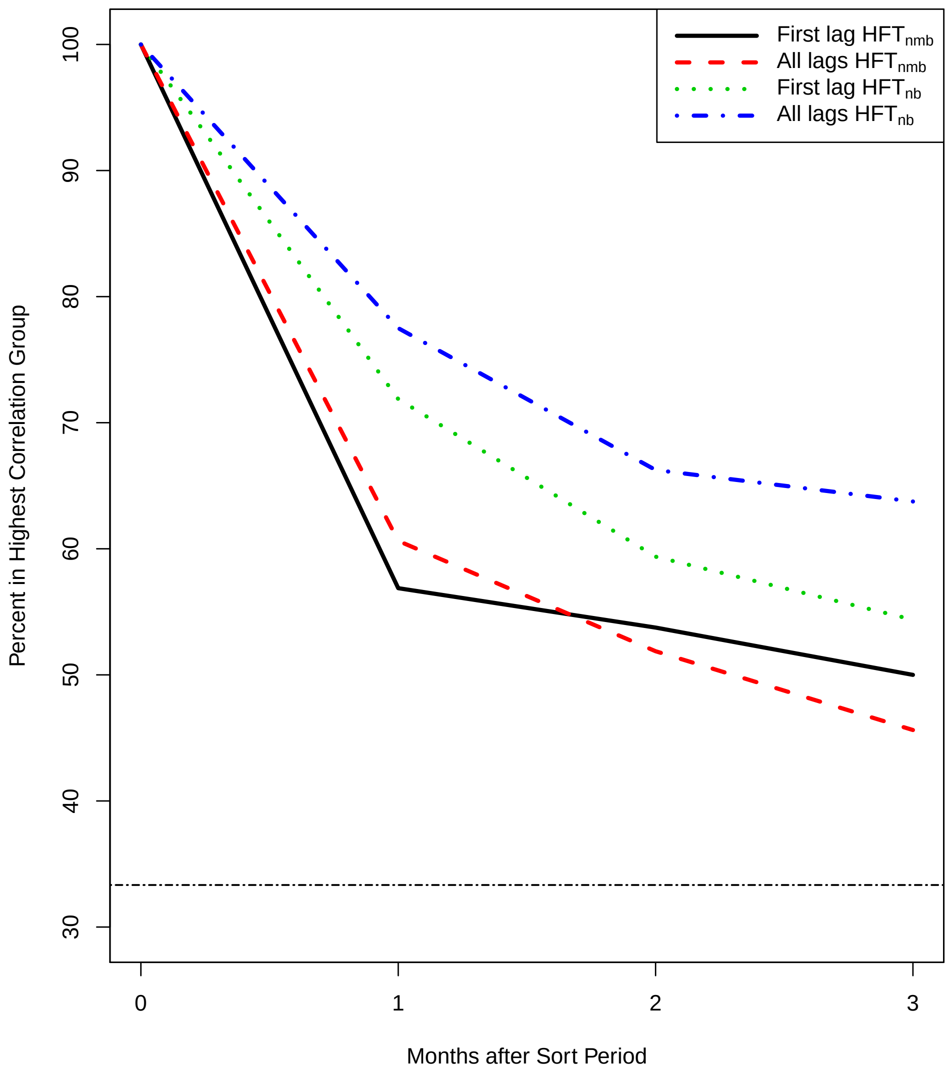


Figure 1.6: Transition Probabilities
This figure plots the probability an HFT who is in the group of HFTs whose trades are most strongly correlated with future non-HFT order flow in month 0 will be in that group in later months. The percents are calculated for each sort month and then averaged across all sort months in the sample. HFTs that leave the sample after the sort period are assigned to the lowest correlation group. The four lines indicate different sorting methods discussed in Table 1.12. The dotted line at 33.3% is what would be expected in months 1–3 if there were no persistence in which HFTs' trades are most strongly correlated with future non-HFT trades.

Table 1.1: Summary of Stock Characteristics

This table summarises stock characteristics. The characteristics are calculated from the pooled time-series of all stock-day observations. Size portfolio breakpoints are computed among NYSE-listed stocks. Size portfolios for year t are formed on December 31st of year $t-1$. Deciles one through five are small-cap, six through eight are mid-cap, and nine through ten are large-cap. MV is market value, PRC is the closing price, VOL is daily share volume, DOLVOL is daily dollar volume, RET is the daily total return, NQ is the share of total dollar volume that traded on NASDAQ, HFT is the fraction of NASDAQ dollar volume traded by HFTs, and N is stocks per day.

	MV	PRC	VOL	DOLVOL	RET	NQ	HFT	N
	Mil $	$	Mil Shr.	Mil $	%	%	%	
All								
mean	5301.9	26.38	2.413	58.138	0.16	27.2	27.6	93.2
sd	12908.9	19.96	5.419	111.438	3.69	13.5	13.7	1.6
0%	22.1	0.91	0.007	0.075	−69.96	0.7	0.0	89.0
25%	402.4	13.58	0.165	2.728	−1.51	15.6	16.1	93.0
50%	1301.4	22.05	0.601	12.627	0.06	25.0	27.7	93.0
75%	4027.3	33.73	2.389	69.827	1.75	38.1	37.9	94.0
100%	125330.6	166.82	113.013	2153.084	40.92	80.7	78.4	96.0
Small-cap								
mean	366.7	16.52	0.296	4.000	0.15	26.3	16.7	33.4
sd	269.2	9.93	0.694	7.815	4.51	15.1	9.6	1.3
0%	22.1	0.91	0.007	0.075	−69.96	0.8	0.0	30.0
25%	188.9	9.29	0.074	1.025	−1.99	13.3	9.4	33.0
50%	293.4	14.77	0.144	1.894	0.00	22.5	14.8	33.0
75%	448.5	20.51	0.277	3.869	2.20	38.3	22.4	34.0
100%	1920.8	71.69	21.938	363.294	40.92	78.8	78.4	36.0
Mid-cap								
mean	1900.1	26.10	2.288	34.188	0.17	27.1	28.9	35.8
sd	1191.5	13.76	5.278	50.672	3.52	12.6	11.1	0.4
0%	238.3	1.02	0.009	0.234	−30.59	0.7	0.0	35.0
25%	1011.5	14.50	0.303	7.538	−1.45	16.3	21.3	36.0
50%	1565.5	25.04	0.741	15.524	0.08	26.0	29.2	36.0
75%	2461.7	35.65	1.869	40.599	1.71	36.7	36.6	36.0
100%	8932.0	77.22	110.098	1270.319	34.40	80.7	71.2	36.0
Large-cap								
mean	17252.2	40.55	5.551	169.288	0.15	28.5	40.7	24.0
sd	21241.6	28.16	7.467	164.436	2.46	12.2	9.0	0.0
0%	2344.1	6.79	0.143	7.258	−13.36	3.1	8.1	24.0
25%	6043.6	22.78	1.829	70.931	−1.12	17.3	34.3	24.0
50%	9413.5	31.37	3.557	120.235	0.08	26.9	40.9	24.0
75%	18532.6	49.58	6.140	203.584	1.37	39.4	47.2	24.0
100%	125330.6	166.82	113.013	2153.084	23.26	62.8	66.9	24.0

Table 1.2: Summary of Stock-Day Observations

This table provides summary statistics describing returns and net buying measures for the sample stocks. Each day, for every stock, the standard deviation of NBBO bid-ask midpoint returns (ret), HFTs' net buying (HFT_{NB}), HFTs' net marketable buying (HFT_{NMB}), HFTs' net marketable buying when it is the same direction as their net buying (HFT_{NMBSD}), and non-HFTs' net marketable buying ($non\text{-}HFT_{NMB}$) are calculated. Net buying is shares bought minus shares sold. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make net-buying measures comparable across stocks, they are divided by 20-day trailing average daily volume. For HFT_{NMBSD}, I require that HFT net buying is in the same direction as net marketable buying. Specifically, positive values of HFT net marketable buying are set to zero if net buying is less than the fourth quintile, and negative values are set to zero if net buying is greater than the second quintile. For this table only, imbalance measures are expressed in shares rather than as a fraction of trailing volume to ease interpretation. Rows describe the distribution of that stock-level statistic across all stock days.

	$\sigma(ret)$ %	$\sigma(HFT_{NB})$ shares	$\sigma(HFT_{NMB})$ shares	$\sigma(HFT_{NMBSD})$ shares	$\sigma(non\text{-}HFT_{NMB})$ shares
All Stocks					
mean	0.08	83	80	76	125
sd	0.15	177	157	147	294
0%	0.00	0	0	0	0
25%	0.02	10	8	8	18
50%	0.03	28	26	26	42
75%	0.07	86	86	83	115
100%	2.70	3387	2351	2235	5681
Small-cap					
mean	0.05	13	11	11	27
sd	0.07	19	17	16	39
0%	0.00	0	0	0	0
25%	0.02	5	4	3	10
50%	0.03	9	7	7	18
75%	0.05	16	13	13	30
100%	0.88	553	543	537	1001
Mid-cap					
mean	0.09	89	81	78	138
sd	0.17	193	155	145	325
0%	0.00	0	0	0	1
25%	0.01	17	16	15	25
50%	0.02	35	34	33	49
75%	0.08	74	71	69	110
100%	2.29	3387	2351	2235	5681
Large-cap					
mean	0.11	175	178	169	247
sd	0.21	228	211	196	391
0%	0.00	13	13	12	12
25%	0.01	67	69	66	81
50%	0.02	116	123	118	138
75%	0.13	185	199	188	247
100%	2.70	2511	1915	1769	4909

Table 1.3: Summary of News Data

This table summarizes news data. News for a stock comes from the Factiva news archive. Panel A shows the distribution among sample stocks in the total number of articles and number of trading days with news. The left two columns in Panel B show the top 10 sources for time-stamped news, and the right three columns shows the number of stamped vs. all articles from three major business news publications.

Panel A: Distribution of Articles

Per Stock	**Time-stamped Articles**		**All Articles**	
	Articles	Days with Articles	Articles	Days with Articles
mean	70	29	762	130
sd	147	33	1611	78
0%	1	1	14	12
25%	12	8	128	60
50%	22	16	227	105
75%	83	40	884	207
100%	1106	158	11877	251

Panel B: Source Summary

Top Time-stamped News Sources		**Stamp/No-stamp Breakdown for Major Sources**		
	Articles		Stamped	All
Dow Jones News Service	871	Dow Jones	4413	6597
Associated Press Newswires	751	Reuters	1317	3234
MidnightTrader	604	Wall Street Journal	174	1556
PR Newswire (U.S.)	488			
Reuters News	365			
Regulatory News Service	341			
Business Wire	337			
MarketWatch	259			
Market News Publishing	234			
DJ em Portugu??s	219			

Table 1.4: Non-HFT Net Marketable Buying for Stocks Sorted by HFT Net Marketable Buying

This table shows investment fund net marketable buying for stocks sorted on HFTs' net marketable buying at the one-second horizon. Panel B excludes stocks that have a news article about them published within five minutes of the sort period. Table 1.3 describes the news data. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. Stocks are sorted into deciles at time t based on HFT net-marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. To make net-buying measures comparable across stocks, they are divided by 20-day trailing average daily volume. I require that HFT net buying is in the same direction as net marketable buying. Specifically, portfolios one and two require net buying to be in net buying quintiles one or two, and portfolios nine and ten require net buying to be in quintiles four or five. Net buying is shares bought minus shares sold. Size portfolio breakpoints are computed among NYSE-listed stocks. Size portfolios for year t are formed on December 31$^{\text{st}}$ of year $t-1$. Deciles one through five are small-cap, six through eight are mid-cap, and nine through ten are large-cap. Non-HFTs' net marketable buying is averaged across all observations for a day, and the mean of the daily time series is reported in the table. Parentheses indicate Newey and West (1994) t-statistics for the time-series means.

Panel A: All intra-day periods

Decile	**Seconds**					
	$[t-30, t-1]$	$t-1$	t	$t+1$	$[t+1, t+30]$	$[t+1, t+300]$
All Stocks						
10 (HFT Buying)	0.30	0.10	0.46	0.09	0.66	1.22
	(5.88)	(22.55)	(24.57)	(25.35)	(15.67)	(3.06)
9	0.18	0.04	0.10	0.03	0.26	0.50
	(15.41)	(29.68)	(23.66)	(20.16)	(21.74)	(7.26)
2	− 0.16	− 0.05	− 0.11	− 0.03	− 0.26	− 0.60
	(−16.10)	(−26.13)	(−24.40)	(−28.10)	(−25.95)	(−8.47)
1 (HFT Selling)	− 0.33	− 0.11	− 0.45	− 0.09	− 0.68	− 1.76
	(−6.02)	(−22.36)	(−37.69)	(−19.93)	(−11.59)	(−4.84)
Small-cap						
10 (HFT Buying)	1.11	0.25	0.96	0.21	1.41	2.55
	(5.70)	(17.69)	(18.11)	(14.46)	(9.62)	(1.75)
1 (HFT Selling)	− 1.12	− 0.26	− 0.95	− 0.18	− 1.28	− 3.22
	(−5.14)	(−11.73)	(−26.07)	(−9.83)	(−5.69)	(−2.46)
Mid-cap						
10 (HFT Buying)	0.28	0.11	0.47	0.09	0.62	1.22
	(6.07)	(19.12)	(27.28)	(22.99)	(16.37)	(4.81)
1 (HFT Selling)	− 0.36	− 0.11	− 0.48	− 0.09	− 0.68	− 1.73
	(−7.49)	(−21.32)	(−29.30)	(−21.44)	(−14.33)	(−6.21)
Large-cap						
10 (HFT Buying)	− 0.00	0.03	0.23	0.05	0.39	0.86
	(−0.00)	(9.75)	(27.92)	(20.49)	(18.17)	(6.78)
1 (HFT Selling)	0.03	− 0.03	− 0.23	− 0.05	− 0.38	− 0.92
	(1.56)	(−15.65)	(−30.10)	(−17.55)	(−23.70)	(−8.22)

Table 1.4—continued

Panel B: Periods that are not within +/− 5 minutes of a news article

Decile	**Seconds**					
	$[t-30, t-1]$	$t-1$	t	$t+1$	$[t+1, t+30]$	$[t+1, t+300]$
10 (HFT Buying)	0.29	0.10	0.45	0.09	0.65	1.26
	(6.45)	(22.46)	(24.69)	(27.70)	(16.80)	(3.99)
9	0.18	0.04	0.10	0.03	0.26	0.51
	(15.41)	(29.79)	(23.86)	(20.10)	(21.71)	(7.23)
2	− 0.16	− 0.05	− 0.11	− 0.03	− 0.26	− 0.61
	(−15.56)	(−25.85)	(−24.57)	(−28.26)	(−25.22)	(−8.82)
1 (HFT Selling)	− 0.32	− 0.10	− 0.45	− 0.09	− 0.66	− 1.68
	(−7.36)	(−26.62)	(−38.42)	(−22.62)	(−15.62)	(−6.03)

Table 1.5: Returns for Stocks Sorted by HFT Net Marketable Buying

This table shows returns in basis points for stocks sorted on HFTs' net marketable buying at the one-second horizon. Panel B excludes stocks that have a news article about them published within five minutes of the sort period. Table 1.3 describes the news data. Stocks are sorted into deciles at time t based on HFT net-marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. To make net-buying measures comparable across stocks, they are divided by 20-day trailing average daily volume. I require that HFT net buying is in the same direction as net marketable buying. Specifically, portfolios one and two require net buying to be in net buying quintiles one or two, and portfolios nine and ten require net buying to be in quintiles four or five. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. Net buying is shares bought minus shares sold. Size portfolio breakpoints are computed among NYSE-listed stocks. Size portfolios for year t are formed on December 31st of year $t-1$. Deciles one through five are small-cap, six through eight are mid-cap, and nine through ten are large-cap. Returns are averaged across all observations for a day, and the mean of the daily time series is reported in the table. Parentheses indicate Newey and West (1994) t-statistics for the time-series means.

Panel A: All intra-day periods

Decile	**Seconds**					
	$[t-30, t-1]$	$t-1$	t	$t+1$	$[t+1, t+30]$	$[t+1, t+300]$
All Stocks						
10 (HFT Buying)	4.56	4.47	0.92	0.55	1.23	0.62
	(27.90)	(27.29)	(15.33)	(16.91)	(12.40)	(1.41)
9	3.47	3.11	0.66	0.48	0.68	0.51
	(29.18)	(24.50)	(9.77)	(15.48)	(11.16)	(3.26)
2	− 3.28	− 3.02	− 0.51	− 0.33	− 0.49	0.14
	(−30.27)	(−27.70)	(−9.32)	(−10.32)	(−6.66)	(0.84)
1 (HFT Selling)	− 4.40	− 4.45	− 0.80	− 0.48	− 1.04	− 0.41
	(−28.10)	(−28.22)	(−14.16)	(−14.41)	(−13.63)	(−1.24)
Small-cap						
10 (HFT Buying)	8.20	6.56	1.32	0.60	2.54	1.32
	(19.43)	(20.18)	(13.09)	(10.58)	(8.39)	(0.85)
1 (HFT Selling)	− 7.77	− 6.52	− 1.21	− 0.58	− 2.59	− 2.71
	(−15.39)	(−21.55)	(−15.14)	(−12.22)	(−14.84)	(−2.57)
Mid-cap						
10 (HFT Buying)	4.63	4.56	1.02	0.61	1.52	1.17
	(26.23)	(27.00)	(13.87)	(12.23)	(15.90)	(4.31)
1 (HFT Selling)	− 4.54	− 4.65	− 0.79	− 0.53	− 1.28	− 0.34
	(−31.22)	(−31.24)	(−12.21)	(−9.90)	(−13.39)	(−1.41)
Large-cap						
10 (HFT Buying)	3.01	3.49	0.65	0.47	0.41	0.08
	(35.83)	(30.13)	(12.71)	(10.92)	(5.14)	(0.49)
1 (HFT Selling)	− 2.83	− 3.39	− 0.64	− 0.38	− 0.14	0.67
	(−27.62)	(−30.88)	(−10.21)	(−7.40)	(−2.07)	(4.34)

Table 1.5—continued

Panel B: Periods that are not within +/− 5 minutes of a news article

Decile	Seconds					
	$[t-30, t-1]$	$t-1$	t	$t+1$	$[t+1, t+30]$	$[t+1, t+300]$
10 (HFT Buying)	4.54	4.47	0.92	0.55	1.23	0.65
	(28.45)	(27.25)	(15.48)	(16.75)	(12.02)	(1.51)
9	3.46	3.11	0.66	0.49	0.68	0.50
	(29.20)	(24.48)	(9.73)	(15.76)	(11.46)	(3.17)
2	− 3.29	− 3.02	− 0.51	− 0.33	− 0.49	0.13
	(−29.96)	(−27.56)	(−9.13)	(−10.53)	(−6.77)	(0.75)
1 (HFT Selling)	− 4.39	− 4.45	− 0.80	− 0.48	− 1.04	− 0.41
	(−28.53)	(−28.37)	(−14.31)	(−14.19)	(−13.04)	(−1.21)

Table 1.6: Intra-day VAR Estimates for Individual Stock-day Observations

For each stock-day observation, the following vector autoregressions (VARs) with ten lages are estimated:

$$R_t = \alpha_t + \sum_{i=1}^{10} \gamma HFT_{NMBSD,t-i} + \sum_{i=1}^{10} \beta non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda R_{t-i} + \epsilon_{t,R} \tag{1.1}$$

$$HFT_{NMBSD,t} = \alpha_t + \sum_{i=1}^{10} \gamma HFT_{NMBSD,t-i} + \sum_{i=1}^{10} \beta non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda R_{t-i} + \epsilon_{t,HFT} \tag{1.2}$$

$$non\text{-}HFT_{NMB,t} = \alpha_t + \sum_{i=1}^{10} \gamma HFT_{NMBSD,t-i} + \sum_{i=1}^{10} \beta non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda R_{t-i} + \epsilon_{t,non\text{-}HFT} \tag{1.3}$$

where R_t is the one-second return, $HFT_{NMBSD,t}$ is one-second HFT net marketable buying in the same direction as net buying, and $non\text{-}HFT_{NMB,t}$ is one-second non-HFT net marketable buying. Table 1.2 describes construction of these imbalance measures. Panel A reports the average coefficients and percent of stock days with positive and negative coefficients that are significantly different from zero at the five percent confidence level. In Panel B, coefficients are averaged across all observations for a day, and the mean of the daily time series is reported in the table. Parentheses indicate Newey and West (1994) t-statistics for the time-series means. I require at least two non-zero observation for each variable. This limits the sample to 23,072 stock-day observations. When constructing cross-sectional means, stock-days are weighted by the minimum number of non-zero observations among the three variables. All variables are divided by their standard deviation among all stocks that day.

Table 1.6—continued

Panel A: Summary of stock-day observations

lag	γ (HFT)			β (non-HFT)			λ (R)		
	μ	% +	% -	μ	% +	% -	μ	% +	% -
$y = non\text{-}HFT_t$									
1	0.0023	24.9	16.6	0.0740	71.0	5.1	0.8595	90.3	0.7
2	0.0025	18.4	8.8	0.0256	49.8	7.4	−0.0171	30.2	14.0
3	0.0023	14.7	7.1	0.0188	39.7	6.3	−0.0015	18.3	10.0
4	0.0020	13.0	6.6	0.0156	34.9	6.7	−0.0025	15.3	8.1
5	0.0018	12.4	6.5	0.0146	33.6	5.8	−0.0032	12.3	8.0
6	0.0018	11.4	5.9	0.0115	28.8	6.4	0.0045	12.1	6.5
7	0.0017	10.3	6.0	0.0094	25.3	6.4	0.0017	9.9	6.9
8	0.0017	10.6	5.7	0.0088	24.1	7.1	0.0008	8.7	6.3
9	0.0016	10.3	5.3	0.0090	23.8	6.2	−0.0001	7.9	6.5
10	0.0016	10.0	5.6	0.0119	28.3	5.6	−0.0022	7.2	6.2
$y = R_t$									
1	0.0177	40.5	3.4	0.0285	44.7	3.3	−0.1440	20.1	65.5
2	0.0109	28.1	3.4	0.0164	30.7	3.8	−0.1230	11.2	63.2
3	0.0079	21.2	4.0	0.0125	23.9	4.5	−0.1005	10.1	60.1
4	0.0064	16.8	4.6	0.0107	19.9	4.8	−0.0851	12.1	54.6
5	0.0054	13.8	4.4	0.0084	16.8	4.6	−0.0702	11.2	53.5
6	0.0038	11.8	4.5	0.0068	13.5	4.9	−0.0605	11.7	50.7
7	0.0035	9.7	4.4	0.0036	11.7	5.3	−0.0490	10.8	48.8
8	0.0019	8.5	4.6	0.0042	10.8	4.9	−0.0395	12.3	45.8
9	0.0015	8.1	4.9	0.0021	11.5	5.2	−0.0290	11.9	44.7
10	−0.0002	6.5	4.7	−0.0002	9.0	5.0	−0.0179	13.0	44.8
$y = HFT_t$									
1	0.0259	49.0	3.7	0.0053	22.9	12.5	1.6312	85.6	2.1
2	0.0090	24.9	3.9	0.0014	13.1	8.0	0.0337	27.3	7.8
3	0.0055	17.8	3.5	0.0004	9.9	7.2	0.0011	14.1	6.8
4	0.0038	14.6	4.1	−0.0005	8.7	6.9	−0.0067	10.4	6.2
5	0.0040	14.4	3.7	−0.0008	8.3	6.7	−0.0076	8.5	6.0
6	0.0025	11.0	4.5	−0.0008	7.7	6.7	−0.0076	7.2	6.2
7	0.0021	9.7	4.1	−0.0009	6.9	6.3	−0.0071	6.5	5.8
8	0.0012	8.9	4.9	−0.0005	6.9	6.2	−0.0078	5.5	5.5
9	0.0013	8.5	4.7	−0.0009	6.5	6.1	−0.0095	4.9	5.2
10	0.0005	8.7	5.3	−0.0015	6.7	6.5	−0.0081	4.6	5.3

Table 1.6—continued

Panel B: Time-series average of mean daily coefficients

lag	γ (HFT)		β (non-HFT)		λ (R)	
	μ	t-stat	μ	t-stat	μ	t-stat
$y = non\text{-}HFT_t$						
1	0.0020	6.64	0.0756	38.36	0.9452	13.60
2	0.0024	18.06	0.0251	41.18	−0.0200	−4.97
3	0.0022	13.69	0.0188	40.46	−0.0020	−1.27
4	0.0019	14.09	0.0154	33.41	−0.0035	−2.23
5	0.0016	15.02	0.0145	38.71	−0.0040	−2.72
6	0.0017	15.06	0.0112	28.10	0.0047	4.61
7	0.0016	14.18	0.0094	32.96	0.0015	1.43
8	0.0016	17.22	0.0085	27.63	0.0005	0.66
9	0.0016	14.73	0.0089	31.16	−0.0006	−0.62
10	0.0015	14.00	0.0118	31.61	−0.0027	−2.69
$y = R_t$						
1	0.0160	11.22	0.0264	15.31	−0.1438	−44.42
2	0.0098	12.58	0.0152	13.70	−0.1219	−58.42
3	0.0071	13.48	0.0114	11.54	−0.0996	−66.64
4	0.0058	9.30	0.0099	12.46	−0.0842	−62.88
5	0.0048	10.55	0.0077	11.14	−0.0697	−61.21
6	0.0034	12.20	0.0064	13.08	−0.0603	−65.25
7	0.0031	6.45	0.0034	6.43	−0.0488	−63.82
8	0.0017	8.97	0.0040	10.11	−0.0394	−66.43
9	0.0014	4.60	0.0017	3.18	−0.0290	−75.91
10	−0.0002	−1.09	−0.0003	−0.56	−0.0180	−59.27
$y = HFT_t$						
1	0.0247	31.88	0.0048	5.24	1.8031	10.52
2	0.0088	25.87	0.0016	3.45	0.0382	6.18
3	0.0056	23.70	0.0007	1.68	0.0022	0.87
4	0.0039	19.36	−0.0003	−1.14	−0.0068	−3.74
5	0.0042	16.35	−0.0009	−2.52	−0.0077	−4.68
6	0.0025	11.99	−0.0008	−2.41	−0.0074	−4.54
7	0.0021	11.43	−0.0009	−2.79	−0.0073	−5.23
8	0.0014	6.77	−0.0006	−2.11	−0.0081	−6.38
9	0.0014	8.14	−0.0010	−2.84	−0.0103	−7.04
10	0.0006	3.34	−0.0016	−4.73	−0.0086	−6.18

Table 1.7: VAR Estimates on Days with and without News

This table reports coefficients on HFTs' net marketable buying from the VAR in Table 1.6 conditional on whether there is news for the stock on a given day. The table includes results for lags one through ten as well as for the sum of those ten lags. In Panel A, a news day is the day an article about the stock appears in the Faciva news archive. In Panel B, a news day is any day when the absolute value of market-adjusted returns is greater than 1%. Every day, the average coefficient is calculated for each lag of HFTs' net marketable buying in the VAR. Each panel reports the time-series mean and median of daily cross-sectional means, For t-tests, the null hypothesis is that the mean of the daily time series equals zero. The difference column group also reports the p-value from a Wilcoxan rank sum test that the time-series medians are equal.

Panel A: News day defined as a day with a Factiva article

lag	**News days**			**Non-news days**			**Difference**		
	mean	t-stat	median	mean	t-stat	median	mean	t-stat	rank sum p-value
1	0.0019	8.39	0.0019	0.0027	7.06	0.0026	-0.0008	-1.79	0.06
2	0.0025	16.03	0.0023	0.0022	8.29	0.0019	0.0003	0.97	0.06
3	0.0022	13.31	0.0020	0.0021	10.85	0.0018	0.0002	0.76	0.21
4	0.0019	15.18	0.0018	0.0017	8.68	0.0014	0.0002	0.91	0.02
5	0.0017	13.60	0.0017	0.0016	6.18	0.0013	0.0001	0.48	0.06
6	0.0018	13.88	0.0017	0.0012	5.53	0.0011	0.0006	2.33	0.01
7	0.0016	12.73	0.0014	0.0016	6.93	0.0011	0.0000	0.05	0.05
8	0.0018	14.34	0.0016	0.0014	7.16	0.0009	0.0004	1.82	0.00
9	0.0016	12.26	0.0014	0.0016	8.73	0.0014	0.0000	0.08	0.46
10	0.0016	11.46	0.0013	0.0014	8.44	0.0009	0.0002	0.74	0.00
Σ 1-10	0.0187	29.39	0.0178	0.0174	17.95	0.0156	0.0013	1.08	0.04

Table 1.7—continued

Panel B: News day defined as a day with $|return| > 1\%$

lag	$\|return\| > 1\%$			$\|return\| \leq 1\%$			**Difference**		
	mean	t-stat	median	mean	t-stat	median	mean	t-stat	rank sum p-value
1	0.0026	10.01	0.0024	0.0013	4.74	0.0011	0.0014	3.60	0.00
2	0.0026	14.96	0.0024	0.0022	11.74	0.0019	0.0004	1.51	0.03
3	0.0023	11.61	0.0021	0.0020	12.52	0.0019	0.0003	1.02	0.10
4	0.0020	13.28	0.0020	0.0017	11.06	0.0016	0.0003	1.19	0.02
5	0.0015	10.52	0.0016	0.0018	10.75	0.0015	-0.0003	-1.25	0.81
6	0.0016	10.48	0.0015	0.0019	11.05	0.0016	-0.0003	-1.29	0.67
7	0.0016	10.63	0.0013	0.0015	10.82	0.0014	0.0001	0.25	0.59
8	0.0018	13.29	0.0016	0.0015	9.81	0.0013	0.0003	1.55	0.04
9	0.0014	10.90	0.0013	0.0018	11.44	0.0015	-0.0003	-1.69	0.11
10	0.0015	10.23	0.0014	0.0017	10.52	0.0013	-0.0002	-0.96	0.84
Σ 1-10	0.0188	27.52	0.0180	0.0173	22.63	0.0155	0.0015	1.45	0.03

Table 1.8: Predicting Order Flow with Returns and Quotes

This table examines the extent to which non-HFT order flow can be predicted using past quotes and returns. The table reports results from regressions of the form

$$non\text{-}HFT_{i,t} = \alpha + \sum_{j=1}^{J} \sum_{l=1}^{4} \beta X_{j,i,t-l} + \epsilon_{i,t},$$

where *non-HFT* is non-HFT net marketable buying, X_j is some variable, i indexes stocks, t indexes seconds. These regressions are calculated each day using the pooled cross-section of all observations that day. The table reports the time-series average of daily coefficients and adjusted R^2 as well as Newey and West (1994) t-statistics from tests of the hypothesis that the time-series mean of the daily coefficients equals zero. Qt is the bid-ask quote imbalance, $\frac{BidSize-AskSize}{BidSize+AskSize}$, calculated using the national best bid and best offer (NBBO). R is the bid-ask midpoint return. HFT is HFT net marketable buying in the same direction as net buying.

Panel A: Regressions with one independent variable

lag	**(1)**	**(2)**	**(3)**	**(4)**
	Qt	R	*HFT*	*non-HFT*
1	−0.0136	0.0223	0.0118	0.0601
	(−14.66)	(16.96)	(14.72)	(21.98)
2	0.0019	0.0088	0.0055	0.0258
	(3.51)	(17.68)	(10.07)	(17.38)
3	0.0003	0.0056	0.0049	0.0181
	(0.68)	(19.23)	(14.03)	(13.11)
4	0.0013	0.0035	0.0034	0.0155
	(3.01)	(18.18)	(8.48)	(21.52)
Adj. R^2	0.0000	0.0006	0.0006	0.0080

Panel B: Regressions with two independent variables

lag	(5)		(6)		(7)	
	HFT	R	*non-HFT*	R	*HFT*	*non-HFT*
1	0.0115	0.0221	0.0599	0.0214	0.0079	0.0595
	(14.40)	(16.93)	(21.90)	(16.87)	(11.12)	(21.74)
2	0.0053	0.0084	0.0256	0.0068	0.0032	0.0254
	(9.74)	(17.65)	(17.25)	(16.76)	(6.16)	(17.22)
3	0.0048	0.0053	0.0180	0.0041	0.0032	0.0178
	(13.67)	(18.51)	(13.03)	(18.18)	(10.43)	(12.91)
4	0.0034	0.0033	0.0155	0.0025	0.0019	0.0153
	(8.39)	(18.18)	(21.44)	(16.78)	(5.13)	(21.18)
Adj. R^2	0.0011		0.0085		0.0084	

Table 1.8—continued

Panel C: Regression with three independent variables

lag	**(8)**		
	HFT	*non-HFT*	R
1	0.0077	0.0593	0.0212
	(10.82)	(21.66)	(16.85)
2	0.0030	0.0253	0.0065
	(5.91)	(17.11)	(16.81)
3	0.0031	0.0177	0.0039
	(10.18)	(12.84)	(18.32)
4	0.0019	0.0153	0.0023
	(5.07)	(21.11)	(16.85)
Adj. R^2	0.0089		

Table 1.9: VAR Estimates at Different Times of the Day

This table reports coefficients on HFTs' net marketable buying from the VAR in Table 1.6 conditional on the time of day. The table includes results for lags one through ten as well as for the sum of those ten lags. Panel A compares estimates from the open (9:30 a.m. to 10:30 a.m.) to the middle of the day (10:30 a.m. to 3:30 p.m.), and Panel B compares estimates from the close (3:30 p.m. to 4:00 p.m.) to the middle of the day. Every day, the average coefficient is calculated for each lag of HFTs' net marketable buying in the VAR. Each panel reports the time-series mean and median of daily cross-sectional means, For t-tests, the null hypothesis is that the mean of the daily time series equals zero. The difference column group also reports the p-value from a Wilcoxan rank sum test that the time-series medians are equal.

Panel A: Comparing the open to the middle of the day

lag	**9:30–10:00 a.m.**			**10:00 a.m.–3:30 p.m.**			**Difference**		
	mean	t-stat	median	mean	t-stat	median	mean	t-stat	rank sum p-value
1	0.0040	9.52	0.0038	0.0012	6.28	0.0010	0.0028	6.05	0.00
2	0.0036	9.71	0.0031	0.0021	16.84	0.0019	0.0015	3.80	0.00
3	0.0025	8.35	0.0023	0.0019	18.59	0.0017	0.0006	1.79	0.07
4	0.0019	6.69	0.0017	0.0017	17.24	0.0015	0.0002	0.73	0.84
5	0.0017	6.92	0.0013	0.0015	17.18	0.0014	0.0002	0.94	0.95
6	0.0021	7.91	0.0015	0.0013	16.50	0.0013	0.0008	2.97	0.02
7	0.0018	7.37	0.0015	0.0013	14.73	0.0010	0.0005	2.03	0.21
8	0.0013	5.79	0.0013	0.0012	14.51	0.0010	0.0002	0.65	0.52
9	0.0016	5.84	0.0015	0.0012	15.15	0.0011	0.0004	1.52	0.50
10	0.0019	7.88	0.0015	0.0015	16.59	0.0012	0.0004	1.64	0.49
Σ 1-10	0.0224	21.14	0.0189	0.0147	25.99	0.0132	0.0077	6.39	0.00

Table 1.9—continued

Panel B: Comparing the close to the middle of the day

lag	3:30–4:00 p.m.			10:00 a.m.–3:30 p.m.			Difference		
	mean	*t*-stat	median	mean	*t*-stat	median	mean	*t*-stat	rank sum p-value
1	-0.0029	-6.88	-0.0030	0.0012	6.28	0.0010	-0.0041	-8.87	0.00
2	0.0009	2.88	0.0005	0.0021	16.84	0.0019	-0.0012	-3.43	0.00
3	0.0012	3.37	0.0009	0.0019	18.59	0.0017	-0.0007	-2.08	0.00
4	0.0015	5.36	0.0015	0.0017	17.24	0.0015	-0.0002	-0.57	0.19
5	0.0011	4.15	0.0010	0.0015	17.18	0.0014	-0.0003	-1.13	0.05
6	0.0016	5.30	0.0015	0.0013	16.50	0.0013	0.0004	1.16	0.46
7	0.0015	5.22	0.0011	0.0013	14.73	0.0010	0.0003	0.91	0.80
8	0.0018	5.97	0.0013	0.0012	14.51	0.0010	0.0007	2.08	0.52
9	0.0022	5.72	0.0012	0.0012	15.15	0.0011	0.0010	2.54	0.56
10	0.0013	5.01	0.0009	0.0015	16.59	0.0012	-0.0001	-0.39	0.23
Σ 1-10	0.0103	8.00	0.0076	0.0147	25.99	0.0132	-0.0044	-3.13	0.00

Table 1.10: VAR Estimates on High Volume and High Imbalance Days

This table reports coefficients on HFTs' net marketable buying from the VAR in Table 1.6 conditional on the stock's volume or absolute value of net marketable buying imbalance level. The table includes results for lags one through ten as well as for the sum of those ten lags. Panel A compares days with high volume to normal days, and Panel B compares days with high imbalances to normal days. High volume and high imbalance days are calculated using a methodology similar to that of Gervais, Kaniel, and Mingelgrin (2001). A day's volume or imbalance is ranked relative to the prior nineteen days, and if the rank is nineteen or above, the day is considered to be a high volume or high imbalance day. Every day, the average coefficient is calculated for each lag of HFTs' net marketable buying in the VAR. Each panel reports the time-series mean and median of daily cross-sectional means, For t-tests, the null hypothesis is that the mean of the daily time series equals zero. The difference column group also reports the p-value from a Wilcoxan rank sum test that the time-series medians are equal.

Panel A: Comparing high volume to normal days

lag	High Volume Days			Normal Days			Difference		
	mean	t-stat	median	mean	t-stat	median	mean	t-stat	rank sum p-value
1	0.0049	6.64	0.0032	0.0016	7.63	0.0015	0.0033	4.27	0.00
2	0.0045	9.25	0.0038	0.0021	15.53	0.0021	0.0023	4.63	0.00
3	0.0031	7.74	0.0024	0.0020	13.71	0.0019	0.0010	2.45	0.03
4	0.0022	5.75	0.0022	0.0018	16.96	0.0016	0.0004	1.00	0.08
5	0.0022	5.18	0.0017	0.0016	15.91	0.0015	0.0005	1.22	0.43
6	0.0020	5.57	0.0018	0.0017	14.77	0.0015	0.0003	0.87	0.69
7	0.0015	3.42	0.0008	0.0015	14.33	0.0013	0.0000	0.01	0.02
8	0.0019	4.31	0.0015	0.0016	15.62	0.0015	0.0003	0.69	0.59
9	0.0020	5.44	0.0017	0.0015	13.45	0.0014	0.0005	1.23	0.46
10	0.0019	5.37	0.0013	0.0014	12.55	0.0013	0.0005	1.28	0.99
Σ 1-10	0.0261	14.74	0.0223	0.0169	30.11	0.0159	0.0092	4.94	0.00

Table 1.10—continued

Panel B: Comparing high imbalance to normal days

lag	**High Imbalance Days**			**Normal Days**			**Difference**		
	mean	*t*-stat	median	mean	*t*-stat	median	mean	*t*-stat	rank sum p-value
1	0.0036	5.27	0.0035	0.0018	8.75	0.0015	0.0018	2.48	0.00
2	0.0036	7.15	0.0032	0.0023	16.82	0.0021	0.0013	2.57	0.00
3	0.0026	5.34	0.0030	0.0021	14.78	0.0019	0.0005	0.99	0.01
4	0.0022	5.94	0.0019	0.0018	16.53	0.0017	0.0003	0.91	0.38
5	0.0014	2.35	0.0019	0.0017	16.59	0.0015	-0.0003	-0.51	0.37
6	0.0027	4.64	0.0019	0.0016	15.14	0.0015	0.0010	1.78	0.17
7	0.0008	0.62	0.0013	0.0015	14.85	0.0014	-0.0008	-0.61	0.54
8	0.0019	4.22	0.0013	0.0016	16.47	0.0015	0.0003	0.61	0.37
9	0.0018	3.73	0.0019	0.0015	13.19	0.0014	0.0002	0.47	0.41
10	0.0017	4.01	0.0012	0.0015	13.98	0.0013	0.0002	0.49	0.56
Σ 1-10	0.0222	9.22	0.0237	0.0175	31.01	0.0165	0.0047	1.89	0.00

Table 1.11: VAR Estimates for Stocks with High versus Low Spreads

This table reports coefficients on HFTs' net marketable buying from the VAR in Table 1.6 conditional on the stock's bid-ask spread the prior day. Daily spreads are calculated by duration-weighting intra-day spread observations. The table includes results for lags one through ten as well as for the sum of those ten lags. Panel A compares stocks sorted on spreads, and Panel B compares stocks sorted on relative spreads, which is the spread divided by the bid-ask midpoint. Every day, the average coefficient is calculated for each lag of HFTs' net marketable buying in the VAR. Each panel reports the time-series mean and median of daily cross-sectional means, For t-tests, the null hypothesis is that the mean of the daily time series equals zero. The difference column group also reports the p-value from a Wilcoxan rank sum test that the time-series medians are equal.

Panel A: Comparing high spread to low spread stocks

lag	High Spread Stocks			Low Spread Stocks			Difference		
	mean	t-stat	median	mean	t-stat	median	mean	t-stat	rank sum p-value
1	0.0053	14.60	0.0046	0.0000	0.03	0.0001	0.0053	11.18	0.00
2	0.0032	13.15	0.0026	0.0020	9.28	0.0019	0.0012	3.57	0.00
3	0.0025	10.13	0.0022	0.0020	10.46	0.0017	0.0005	1.49	0.13
4	0.0018	12.03	0.0017	0.0018	10.78	0.0014	0.0000	0.07	0.44
5	0.0018	11.16	0.0015	0.0016	10.64	0.0015	0.0002	0.70	0.81
6	0.0015	12.53	0.0015	0.0018	11.77	0.0016	-0.0002	-1.14	0.57
7	0.0016	9.73	0.0014	0.0017	11.34	0.0013	-0.0001	-0.53	0.64
8	0.0013	8.32	0.0011	0.0019	12.73	0.0016	-0.0005	-2.51	0.01
9	0.0013	8.09	0.0011	0.0018	10.32	0.0016	-0.0005	-2.17	0.01
10	0.0015	8.85	0.0013	0.0016	9.32	0.0015	-0.0001	-0.61	0.16
Σ 1-10	0.0218	25.56	0.0207	0.0162	22.09	0.0145	0.0056	4.95	0.00

Table 1.11—continued

Panel B: Comparing high relative spread to low relative spread stocks

lag	High RelSpread Stocks			Low RelSpread Stocks			Difference		
	mean	t-stat	median	mean	t-stat	median	mean	t-stat	rank sum p-value
1	0.0046	7.81	0.0037	0.0012	4.68	0.0007	0.0034	5.22	0.00
2	0.0036	8.39	0.0030	0.0023	13.21	0.0022	0.0014	2.93	0.02
3	0.0032	8.38	0.0022	0.0020	11.78	0.0019	0.0011	2.76	0.14
4	0.0014	6.37	0.0011	0.0020	14.56	0.0018	-0.0006	-2.18	0.00
5	0.0019	6.32	0.0015	0.0016	12.61	0.0017	0.0002	0.73	0.61
6	0.0014	6.62	0.0010	0.0019	13.06	0.0018	-0.0005	-2.00	0.01
7	0.0010	3.97	0.0008	0.0016	11.94	0.0013	-0.0006	-1.98	0.00
8	0.0013	4.67	0.0009	0.0018	13.09	0.0015	-0.0005	-1.65	0.00
9	0.0012	4.55	0.0009	0.0018	12.32	0.0017	-0.0006	-2.01	0.01
10	0.0018	5.33	0.0014	0.0016	11.52	0.0014	0.0002	0.56	0.69
Σ 1-10	0.0214	17.66	0.0190	0.0179	25.03	0.0169	0.0035	2.52	0.10

Table 1.12: Examining Cross-sectional Differences in Prediction Ability

This table tests whether some HFTs consistently predict future buying and selling pressure better than others. The test examines whether the HFTs who predict buying and selling pressure the best one month continue to do so the next month. Each day, the following regression is run for each HFT:

$$non\text{-}HFT_{d,s,t} = \alpha_{d,i} + \sum_{l=1}^{10} \gamma_{d,i,l} HFT_{d,s,i,t-l} + \sum_{l=1}^{10} \beta_{d,i,l} non\text{-}HFT_{d,s,t-l} + \sum_{i=1}^{10} \lambda_{d,i,l} R_{d,s,t-l} + \epsilon_{d,s,t}, \quad (1.4)$$

where d indexes days, s indexes stocks, t indexes seconds, and i indexes HFTs. *non-HFT* is non-HFT net marketable buying, *HFT* is either the HFT's net marketable buying or their net buying, and R is the stock's return. The individual HFTs' net marketable buying and net buying measures are divided by the standard deviation of aggregate HFT net marketable buying and net buying, respectively. The left two columns in the table below report results from regressions where *HFT* is HFTs' net marketable buying, and the right two columns report results where *HFT* is their net buying. For regression (1.4), I require there to be more than 100 non-zero net marketable buying observations to ensure relatively precise coefficient estimates. Then for each month, among HFTs for whom regression (1.4) could be estimated at least 15 days during the current and following month, HFTs are split into three groups based on their γ_i coefficients. There are two groupings: in the first grouping, HFTs are split based on their $\overline{\gamma_{d,i,1}}$ for that month, and in the second grouping, the split is based on their $\overline{\sum_{l=1}^{10} \gamma_{d,i,l}}$. The cross-sectional average $\overline{\gamma_{d,i,1}}$ or $\overline{\sum_{l=1}^{10} \gamma_{d,i,l}}$ coefficients are then calculated the following month (i.e., the post-sort month). The table reports the time-series mean, t-stat, and p-value from t-tests of the monthly time-series of cross-sectional means for the three groups. The p-values are included, because since the number of months, and so degrees of freedom, is so small, standard rules of thumb for determining statistical significance (e.g., $|t\text{-stat}| > 1.96$) do not apply. High-frequency traders go in and out of the sample, so months are weighted by the number of HFTs in that month's group.

	HFT = **Net Mkt. Buying**			HFT = **Net Buying**		
	μ_{t+1}	t-stat	p-value	μ_{t+1}	t-stat	p-value
$\overline{\gamma_{d,i,1}}$						
High in month t	0.014	5.97	0.000	0.022	12.07	0.000
Mid in month t	0.011	7.61	0.000	0.010	3.28	0.008
Low in month t	0.002	0.64	0.538	0.005	3.59	0.005
High minus Low	0.012	3.51	0.006	0.017	8.18	0.000
$\overline{\sum_{l=1}^{10} \gamma_{d,i,l}}$						
High in month t	0.080	12.15	0.000	0.098	6.91	0.000
Mid in month t	0.034	7.00	0.000	0.033	8.67	0.000
Low in month t	0.012	1.91	0.085	0.009	2.65	0.025
High minus Low	0.068	8.71	0.000	0.089	6.26	0.000

Appendices

Appendix A

Supplementary Tables

Table A.1: Summary of CRSP Universe

This table summarises 2009 stock-day observations for CRSP common stocks with dual-class stocks removed. The table summarises market capitization, *mv*, dollar volume, *dolvol*, and price, *prc*. Market value and dollar volume are in millions. The column *szp* denotes size deciles. Size portfolio breakpoints are computed among NYSE-listed stocks. Size portfolios for year t are formed on December 31st of year $t-1$. Deciles one through five are small-cap, six through eight are mid-cap, and nine through ten are large-cap.

szp	nstocks	avg	sd	min	q1	q2	q3	max
mv								
1	690	18	23	0	7	12	19	436
2	753	58	56	1	30	44	64	996
3	589	139	100	4	81	114	165	1,525
4	575	305	216	13	188	262	360	5,695
5	414	576	308	24	389	509	672	3,526
6	312	977	447	31	698	883	1,126	4,194
7	275	1,677	728	101	1,193	1,538	1,981	6,643
8	212	2,871	1,092	373	2,164	2,672	3,341	10,955
9	213	5,645	2,450	523	3,926	5,208	6,876	33,010
10	202	35,668	43,505	2,375	13,512	19,839	34,675	415,274
dolvol								
1	690	0.15	1.99	0.00	0.00	0.01	0.04	324.23
2	753	0.45	2.88	0.00	0.01	0.05	0.19	231.69
3	589	1.27	5.30	0.00	0.12	0.34	0.89	590.10
4	575	3.34	13.92	0.00	0.59	1.36	2.97	1,533.59
5	414	7.41	17.36	0.00	1.85	3.63	7.49	1,704.35
6	312	14.23	26.80	0.00	4.14	7.70	15.22	1,674.99
7	275	26.72	34.07	0.06	9.90	17.49	31.94	2,655.60
8	212	44.17	57.97	0.07	17.25	30.10	54.75	7,143.57
9	213	90.52	120.99	0.52	40.24	65.87	106.64	7,129.60
10	202	374.36	556.13	7.96	128.37	218.11	381.70	19,972.16
prc								
1	690	2.13	3.24	0.01	0.54	1.24	2.63	78.00
2	753	4.99	5.10	0.01	1.65	3.51	6.71	62.00
3	589	7.87	7.92	0.05	3.10	5.96	10.06	329.79
4	575	12.59	10.99	0.10	5.74	9.71	16.33	155.72
5	414	17.14	14.09	0.06	8.32	14.09	22.36	195.98
6	312	20.60	12.78	0.26	11.30	17.94	26.66	103.86
7	275	29.12	71.06	0.25	14.26	22.42	32.79	1,549.00
8	212	38.63	58.21	0.75	17.76	26.94	39.49	731.00
9	213	32.17	22.92	0.35	18.77	27.70	39.77	306.58
10	202	45.08	43.05	1.02	24.63	37.39	52.40	622.87

Table A.2: Sample Universe

This table summarises 2009 stock-day observations for the set of stocks from which the sample is constructed. The stocks consist of CRSP common equities with dual-class stocks removed. Stocks are also excluded from the sample universe if they fall in the bottom two size deciles, if their price at the end of 2008 is less than \$5, or if average daily dollar volume in December 2008 is less than \$1 dollars. The table summarises market capitization, *mv*, dollar volume, *dolvol*, and price, *prc*. Market value and dollar volume are in millions. The column *szp* denotes size deciles. Size portfolio breakpoints are computed among NYSE-listed stocks. Size portfolios for year t are formed on December 31st of year $t-1$. Deciles one through five are small-cap, six through eight are mid-cap, and nine through ten are large-cap.

szp	nstocks	avg	sd	min	q1	q2	q3	max
mv								
3	44	175	90	22	114	160	216	674
4	322	319	162	24	215	289	381	2,151
5	347	554	244	28	392	506	655	2,499
6	293	956	421	31	692	875	1,105	4,194
7	261	1,656	695	169	1,189	1,529	1,961	5,977
8	205	2,858	1,057	373	2,169	2,672	3,329	10,955
9	209	5,568	2,216	523	3,915	5,176	6,835	17,693
10	201	35,774	43,588	2,375	13,525	19,917	34,872	415,274
dolvol								
3	44	1.88	3.86	0.00	0.60	1.15	2.16	229.51
4	322	3.36	7.18	0.00	1.02	1.88	3.56	888.89
5	347	6.59	13.04	0.00	1.98	3.64	7.07	1,704.35
6	293	13.31	22.07	0.00	4.11	7.51	14.46	1,283.42
7	261	25.71	31.28	0.06	9.82	17.16	30.99	2,655.60
8	205	43.30	46.58	0.07	17.20	29.88	54.12	1,969.46
9	209	85.76	84.23	0.52	39.87	65.17	105.01	2,761.70
10	201	375.33	557.28	7.96	128.60	218.68	382.83	19,972.16
prc								
3	44	12.66	11.78	0.91	6.84	10.60	15.58	329.79
4	322	14.82	10.17	0.73	7.95	12.12	18.88	111.85
5	347	18.17	11.53	0.53	9.78	15.62	23.60	120.33
6	293	21.51	12.55	0.26	12.62	18.69	27.31	103.86
7	261	30.37	72.71	1.04	15.38	23.34	33.45	1,549.00
8	205	39.33	58.76	1.02	18.28	27.27	39.79	731.00
9	209	32.65	22.84	1.03	19.25	27.98	40.08	306.58
10	201	45.28	43.07	1.02	24.85	37.53	52.49	622.87

Table A.3: Detailed Summary of Stock-Day Observations

This table provides summary statistics describing returns and net buying measures for the sample stocks. Each day, for every stock, the average, standard deviation, minimum, maximum, and number of non-zero observations of HFTs' net marketable buying, HFTs' net marketable buying where it is the same direction as their net buying, non-HFTs' net marketable buying, and returns are calculated. For HFTs' net marketable buying in the same direction as net buying, observations are set to zero if net marketable buying is negative and net buying is above the second quintile, or if net marketable buying is positive, and net buying is below the fourth quintile. Columns indicate the stock-level statistic, and rows describe the distribution of that stock-level statistic across all stock-days.

	avg	sd	min	max	non-zero
HFT Net Buy (shares)					
mean	−0.05	82.92	−3208.92	3143.25	778.79
sd	3.22	176.82	7417.72	6587.39	991.21
0%	−75.25	0.00	−374962.00	0.00	0.00
25%	−0.29	9.84	−3117.00	500.00	81.00
50%	0.00	28.22	−1200.00	1200.00	334.00
75%	0.22	86.07	−500.00	3100.00	1152.00
100%	97.37	3386.79	0.00	129526.00	9255.00
HFT Net Mkt. Buy (shares)					
mean	−0.03	79.92	−3120.39	3104.63	629.65
sd	2.35	156.69	6426.73	6395.86	829.02
0%	−65.81	0.00	−152374.00	0.00	0.00
25%	−0.20	8.20	−3100.00	400.00	53.00
50%	0.00	26.25	−1100.00	1100.00	266.00
75%	0.18	86.37	−400.00	3100.00	908.00
100%	71.76	2351.48	0.00	119277.00	6874.00
HFT Net Mkt. Buy Same Dir (shares)					
mean	−0.03	76.14	−3098.06	3086.62	373.29
sd	2.19	146.77	6382.88	6375.01	463.04
0%	−52.77	0.00	−152374.00	0.00	0.00
25%	−0.20	7.95	−3084.00	400.00	45.00
50%	0.00	25.52	−1100.00	1100.00	188.00
75%	0.18	82.99	−400.00	3087.00	535.00
100%	50.79	2234.93	0.00	119277.00	4802.00
non-HFT Net Mkt. Buy (shares)					
mean	−0.06	124.70	−5946.96	6108.01	822.36
sd	4.61	294.22	15288.41	16937.88	906.77
0%	−116.22	0.00	−347535.00	0.00	0.00
25%	−0.48	18.10	−5000.00	800.00	194.00
50%	−0.01	41.74	−1900.00	1900.00	456.00
75%	0.44	115.43	−800.00	5100.00	1160.00
100%	89.91	5681.36	0.00	590702.00	9957.00
Ret (%)					
mean	0.00	0.08	−2.09	2.20	2078.62
sd	0.00	0.15	2.47	2.72	1638.14
0%	0.00	0.00	−36.86	0.00	0.00
25%	0.00	0.02	−2.83	0.42	987.00
50%	0.00	0.03	−0.92	0.94	1639.00
75%	0.00	0.07	−0.42	2.92	2672.00
100%	0.04	2.70	0.00	71.35	13934.00

Table A.4: Returns for Stocks Sorted by High-Frequency Trader Net Buying

This table shows returns for stocks sorted on HFTs' net marketable buying (Panels A and B) and net buying (Panel C). Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. Net buying is shares bought minus shares sold. Stocks are sorted into deciles at time t based on HFT net marketable buying (Panels A and B) or net buying (Panel C). Decile breakpoints are calculated from non-zero observations during the prior trading day. To make net-buying measures comparable across stocks, they are divided by 20-day trailing average daily volume. In Panel B, I require that HFT net buying is in the same direction as net marketable buying. Specifically, portfolios one and two require net buying to be in net buying quintiles one or two, and portfolios nine and ten require net buying to be in quintiles four or five. Sorts in each panel are performed for net-buying measures aggregated over one second, thirty seconds, and thirty-minute horizons. Returns are shown for periods $t-3$ through $t+3$, from the end of the sort period to five minutes later, $[t, t+5m]$, and from the end of the sort period to thirty minutes later, $[t, t+30m]$. Returns are averaged across all observations for a day, and the mean of the daily time series is reported in the table. Parentheses indicate Newey and West (1994) t-statistics for the time-series means.

Panel A: Sorts on HFT net *marketable* buying

Deciles	**Ret (in basis points)**								
	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$	$[t, t+5m]$	$[t, t+30m]$
1 sec.									
1	−0.06	−0.25	−4.38	−0.79	−0.45	−0.11	−0.05	−1.40	−0.53
	(−3.51)	(−9.38)	(−26.91)	(−14.38)	(−14.61)	(−5.23)	(−2.10)	(−3.38)	(−0.40)
2	−0.04	−0.20	−3.04	−0.50	−0.34	−0.01	0.03	−0.48	0.64
	(−2.02)	(−9.24)	(−27.29)	(−8.98)	(−11.23)	(−0.58)	(1.16)	(−2.93)	(0.79)
9	0.11	0.19	3.13	0.64	0.46	0.07	0.05	1.14	1.92
	(4.60)	(9.40)	(23.79)	(9.47)	(14.55)	(3.26)	(2.01)	(6.97)	(2.52)
10	0.09	0.29	4.40	0.91	0.54	0.16	0.12	1.47	1.54
	(5.18)	(10.92)	(24.57)	(16.20)	(18.02)	(7.96)	(5.28)	(3.00)	(1.28)
30 sec.									
1	0.65	0.83	0.62	−8.23	−0.21	0.11	0.25	0.53	2.43
	(10.22)	(10.69)	(7.37)	(−17.74)	(−3.67)	(2.06)	(4.42)	(2.85)	(2.93)
2	0.38	0.62	0.54	−5.98	−0.20	0.22	0.16	0.61	2.18
	(6.95)	(8.23)	(9.38)	(−18.20)	(−3.74)	(3.60)	(3.75)	(4.11)	(2.84)
9	−0.25	−0.47	−0.44	6.21	0.25	0.01	−0.06	−0.14	0.77
	(−4.19)	(−8.38)	(−7.76)	(21.13)	(4.70)	(0.27)	(−1.16)	(−0.77)	(0.99)
10	−0.50	−0.72	−0.54	8.55	0.33	−0.01	−0.05	−0.20	0.35
	(−7.54)	(−10.22)	(−6.76)	(18.09)	(6.20)	(−0.20)	(−1.02)	(−0.92)	(0.44)
30 min.									
1	3.98	4.04	4.84	−12.26	4.17	2.25	0.93	0.77	4.17
	(3.72)	(3.71)	(3.93)	(−9.67)	(4.24)	(2.26)	(1.03)	(1.34)	(4.24)
2	1.23	1.75	2.74	−9.73	2.05	1.22	0.36	0.52	2.05
	(1.01)	(1.69)	(2.62)	(−8.77)	(2.32)	(1.49)	(0.40)	(1.01)	(2.32)
9	−1.05	−0.89	−1.65	11.88	−0.03	−0.53	1.05	0.12	−0.03
	(−0.94)	(−0.78)	(−1.48)	(10.57)	(−0.03)	(−0.61)	(1.29)	(0.26)	(−0.03)
10	−1.16	−0.55	−3.26	14.56	−0.12	0.15	−0.14	−0.86	−0.12
	(−1.06)	(−0.48)	(−2.59)	(10.85)	(−0.14)	(0.15)	(−0.14)	(−1.77)	(−0.14)

Table A.4 — continued

Panel B: Sorts on HFT net marketable buying requiring net buying to be in the same direction.

Deciles	Ret (in basis points)								
	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$	$[t, t+5m]$	$[t, t+30m]$
1 sec.									
1	− 0.07	− 0.26	− 4.36	− 0.79	− 0.47	− 0.10	− 0.05	− 1.22	− 0.23
	(−3.86)	(−9.32)	(−27.51)	(−14.16)	(−14.25)	(−4.24)	(−2.21)	(−3.61)	(−0.20)
2	− 0.07	− 0.26	− 2.96	− 0.50	− 0.33	− 0.01	0.02	− 0.39	0.83
	(−3.94)	(−10.57)	(−27.01)	(−9.25)	(−10.50)	(−0.51)	(0.78)	(−2.30)	(1.01)
9	0.13	0.24	3.04	0.65	0.47	0.07	0.06	1.11	1.95
	(5.49)	(11.74)	(23.10)	(9.76)	(15.38)	(2.71)	(2.15)	(6.68)	(2.50)
10	0.10	0.31	4.36	0.91	0.54	0.17	0.11	1.50	1.64
	(5.68)	(11.29)	(26.33)	(15.62)	(16.84)	(7.81)	(4.36)	(3.46)	(1.47)
30 sec.									
1	0.68	0.83	0.57	− 8.05	− 0.16	0.11	0.28	0.64	2.60
	(11.19)	(9.77)	(6.79)	(−18.72)	(−2.56)	(1.93)	(4.72)	(3.75)	(3.18)
2	0.46	0.61	0.41	− 5.60	− 0.16	0.24	0.18	0.69	2.17
	(7.62)	(7.68)	(6.79)	(−18.10)	(−2.88)	(3.57)	(3.52)	(4.39)	(2.81)
9	− 0.29	− 0.50	− 0.33	5.89	0.21	− 0.02	− 0.06	− 0.19	0.81
	(−4.74)	(−8.49)	(−5.87)	(21.56)	(3.76)	(−0.28)	(−0.97)	(−1.03)	(1.02)
10	− 0.53	− 0.70	− 0.48	8.38	0.30	− 0.03	− 0.05	− 0.29	0.20
	(−7.78)	(−9.81)	(−6.04)	(18.03)	(5.34)	(−0.61)	(−0.82)	(−1.43)	(0.24)
30 min.									
1	3.85	4.04	5.52	− 10.07	3.86	2.41	1.05	0.54	3.86
	(3.65)	(3.70)	(4.67)	(−8.28)	(4.06)	(2.45)	(1.11)	(1.19)	(4.06)
2	1.45	1.93	2.76	− 6.38	1.74	1.04	0.27	0.40	1.74
	(1.09)	(1.98)	(2.63)	(−5.64)	(1.87)	(1.20)	(0.29)	(0.81)	(1.87)
9	− 1.27	− 1.45	− 2.13	8.68	− 0.25	− 0.40	0.91	0.24	− 0.25
	(−1.05)	(−1.23)	(−1.80)	(7.51)	(−0.29)	(−0.45)	(1.05)	(0.50)	(−0.29)
10	− 1.30	− 0.57	− 3.74	12.29	− 0.34	0.26	− 0.07	− 0.91	− 0.34
	(−1.17)	(−0.48)	(−2.93)	(9.08)	(−0.36)	(0.27)	(−0.06)	(−1.89)	(−0.36)

Table A.4 — continued

Panel C: Sorts on HFT net buying

Deciles	Ret (in basis points)								
	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$	$[t, t+5m]$	$[t, t+30m]$
1 sec.									
1	− 0.06	− 0.18	− 1.61	− 0.27	− 0.19	− 0.06	0.00	0.08	0.84
	(−3.76)	(−7.18)	(−12.04)	(−7.06)	(−6.55)	(−3.18)	(0.23)	(0.23)	(0.71)
2	− 0.05	− 0.18	− 0.95	− 0.16	− 0.11	− 0.02	0.05	0.45	1.32
	(−2.81)	(−8.80)	(−12.76)	(−5.16)	(−4.02)	(−0.57)	(2.78)	(2.73)	(1.63)
9	0.11	0.19	1.07	0.25	0.23	0.06	0.03	0.35	1.19
	(5.86)	(10.44)	(11.83)	(6.91)	(6.94)	(3.16)	(1.69)	(2.33)	(1.59)
10	0.08	0.22	1.58	0.40	0.28	0.10	0.07	0.15	0.17
	(5.68)	(9.74)	(10.20)	(10.28)	(12.30)	(6.80)	(4.23)	(0.34)	(0.14)
30 sec.									
1	0.54	0.56	0.25	− 3.48	0.13	0.23	0.30	1.21	2.89
	(11.71)	(10.05)	(5.21)	(−10.75)	(2.56)	(5.00)	(6.00)	(7.38)	(3.44)
2	0.40	0.43	0.30	− 2.61	0.08	0.21	0.21	0.81	2.18
	(8.81)	(7.09)	(6.24)	(−12.00)	(1.95)	(3.97)	(5.51)	(5.40)	(2.85)
9	− 0.16	− 0.41	− 0.16	2.80	0.07	− 0.05	− 0.04	− 0.34	0.65
	(−3.08)	(−7.51)	(−3.25)	(13.75)	(1.51)	(−1.30)	(−0.81)	(−2.21)	(0.85)
10	− 0.40	− 0.42	− 0.16	3.69	− 0.01	− 0.12	− 0.10	− 0.67	− 0.08
	(−7.50)	(−8.83)	(−3.02)	(10.17)	(−0.18)	(−2.76)	(−1.92)	(−3.33)	(−0.10)
30 min.									
1	2.99	3.36	5.73	3.48	1.33	1.43	0.58	− 0.11	1.33
	(2.66)	(3.19)	(4.65)	(2.67)	(1.41)	(1.68)	(0.58)	(−0.27)	(1.41)
2	1.52	2.62	2.71	1.10	2.46	0.59	0.60	0.71	2.46
	(1.20)	(2.33)	(2.80)	(1.05)	(2.94)	(0.69)	(0.71)	(1.69)	(2.94)
9	− 1.39	− 0.88	− 1.84	2.19	0.70	0.58	0.28	0.30	0.70
	(−1.07)	(−0.86)	(−1.66)	(2.10)	(0.78)	(0.68)	(0.30)	(0.61)	(0.78)
10	− 0.47	0.13	− 3.61	− 2.38	0.06	0.80	0.52	− 0.16	0.06
	(−0.46)	(0.12)	(−2.66)	(−1.70)	(0.06)	(0.80)	(0.59)	(−0.31)	(0.06)

Table A.5: Robustness of VAR to Price Feed Latency

This table tests whether potentially mismatched NASDAQ trade and NBBO quote timestamps affect the VAR results in Table 1.6 by rerunning the VAR using NASDAQ BBO midpoint returns rather than NBBO midpoint returns. The NASDAQ trade and NASDAQ BBO timestamps are precisely aligned. Since calculating the NASDAQ BBO is computationally intensive, the VAR uses only a subset of the sample period, January 1$^{\text{st}}$ to March 4$^{\text{th}}$ of 2009. Panel A reports results with NASDAQ BBO midpoint returns, and Panel B reports results with NBBO midpoint returns over the same time period. In both panels, coefficients are averaged across all observations for a day, and the mean of the daily time series is reported in the table. Parentheses indicate Newey and West (1994) *t*-statistics for the time-series means. I require at least one non-zero observation for each variable. When constructing cross-sectional means, stock-days are weighted by the minimum number of non-zero observations among the three variables. All variables are divided by their standard deviation among all stocks that day.

Table A.5—continued

Panel A: NASDAQ BBO time-series average of mean daily coefficients

lag	γ (HFT)		β (non-HFT)		λ (R)	
	μ	t-stat	μ	t-stat	μ	t-stat
$y = R_t$						
1	0.0209	19.01	0.0318	43.14	−0.0512	−26.73
2	0.0124	19.08	0.0181	31.32	−0.0347	−25.42
3	0.0082	20.74	0.0145	27.07	−0.0249	−27.70
4	0.0070	18.64	0.0111	20.86	−0.0204	−22.72
5	0.0047	17.86	0.0096	15.57	−0.0169	−26.68
6	0.0037	12.00	0.0067	12.68	−0.0143	−21.59
7	0.0027	12.44	0.0050	10.09	−0.0109	−24.67
8	0.0018	4.33	0.0052	13.94	−0.0088	−25.46
9	0.0015	3.38	0.0062	10.95	−0.0072	−16.68
10	−0.0002	−0.65	0.0027	7.47	−0.0070	−10.49
$y = non\text{-}HFT_t$						
1	0.0027	8.47	0.0515	26.94	0.1714	29.84
2	0.0028	13.44	0.0217	33.69	−0.0005	−0.86
3	0.0024	9.72	0.0159	40.11	−0.0001	−0.30
4	0.0023	10.07	0.0140	19.01	0.0004	1.85
5	0.0023	9.70	0.0128	31.97	−0.0002	−0.60
6	0.0021	8.97	0.0101	21.19	0.0006	1.74
7	0.0017	10.97	0.0073	27.48	0.0007	2.61
8	0.0015	10.33	0.0074	17.01	0.0006	3.24
9	0.0016	7.62	0.0075	16.62	0.0003	1.79
10	0.0018	10.64	0.0098	31.90	0.0000	0.15
$y = HFT_t$						
1	0.0288	26.61	0.0073	8.21	0.3067	34.31
2	0.0070	5.52	0.0002	0.31	0.0160	23.13
3	0.0032	5.39	0.0008	1.63	0.0051	8.45
4	0.0034	10.57	0.0003	0.66	0.0028	5.84
5	0.0023	6.30	−0.0001	−0.13	0.0015	4.06
6	0.0014	5.78	−0.0005	−1.06	0.0011	2.68
7	0.0011	4.29	−0.0008	−1.65	0.0006	2.03
8	0.0003	0.93	−0.0007	−2.01	0.0002	0.56
9	0.0001	0.47	−0.0005	−0.71	−0.0002	−0.76
10	0.0000	0.14	−0.0001	−0.22	−0.0006	−2.55

Table A.5—continued

Panel B: NBBO time-series average of mean daily coefficients

lag	γ (HFT)		β (non-HFT)		λ (R)	
	μ	t-stat	μ	t-stat	μ	t-stat
$y = R_t$						
1	0.0199	14.55	0.0283	20.68	−0.1257	−29.90
2	0.0126	12.42	0.0168	23.97	−0.1077	−39.60
3	0.0093	16.94	0.0148	11.31	−0.0881	−42.24
4	0.0073	12.52	0.0098	12.89	−0.0752	−36.22
5	0.0061	6.74	0.0084	10.40	−0.0593	−38.77
6	0.0048	6.80	0.0070	13.12	−0.0531	−47.93
7	0.0042	9.94	0.0044	3.85	−0.0420	−31.96
8	0.0021	4.62	0.0035	6.12	−0.0357	−28.71
9	0.0022	5.66	0.0028	3.72	−0.0264	−29.15
10	−0.0004	−1.18	0.0000	−0.07	−0.0179	−26.95
$y = non\text{-}HFT_t$						
1	0.0022	6.23	0.0515	27.79	0.5565	20.79
2	0.0028	13.22	0.0219	30.71	0.0021	0.97
3	0.0023	10.04	0.0160	38.98	0.0001	0.04
4	0.0022	8.40	0.0140	18.35	0.0008	0.68
5	0.0023	8.92	0.0129	31.93	−0.0001	−0.14
6	0.0022	7.19	0.0102	21.73	0.0023	1.84
7	0.0016	9.56	0.0074	27.82	0.0023	2.52
8	0.0015	9.15	0.0074	18.50	0.0025	2.48
9	0.0015	7.07	0.0075	17.00	0.0017	1.98
10	0.0018	9.20	0.0099	28.69	0.0010	1.46
$y = HFT_t$						
1	0.0301	28.11	0.0096	10.14	0.9359	31.86
2	0.0079	5.84	0.0017	2.15	0.0295	14.69
3	0.0035	6.45	0.0015	2.50	0.0061	3.58
4	0.0035	10.82	0.0006	1.46	0.0026	1.80
5	0.0024	6.81	0.0004	0.84	0.0003	0.22
6	0.0014	7.05	−0.0003	−0.64	−0.0018	−1.71
7	0.0011	4.64	−0.0006	−1.27	−0.0013	−1.60
8	0.0001	0.40	−0.0005	−1.51	−0.0023	−6.86
9	0.0001	0.24	−0.0005	−0.72	−0.0029	−2.09
10	0.0000	−0.17	0.0002	0.36	−0.0031	−3.07

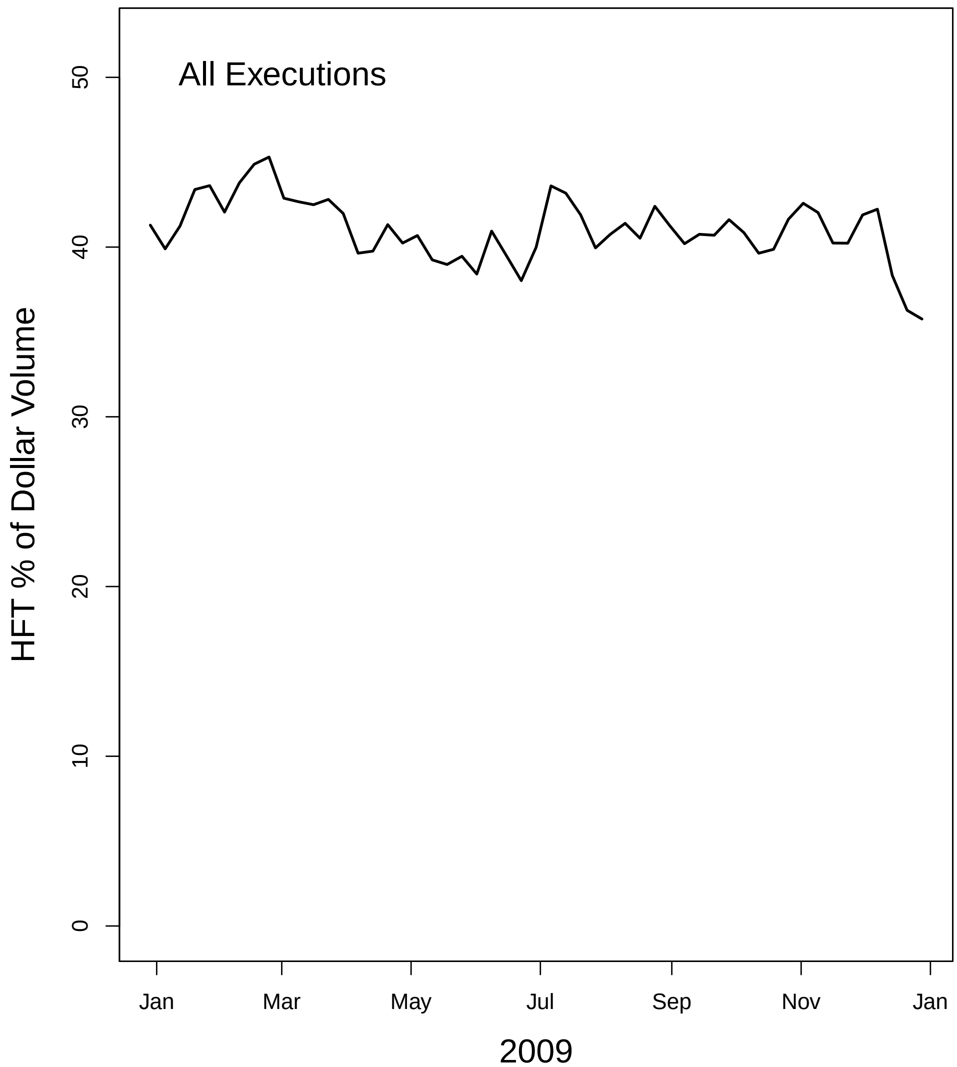


Figure A.1: HFTs' Share of NASDAQ Dollar Volume

This figure shows HFTs' share of dollar volume on the NASDAQ Stock Market. The calculation includes all stocks with CRSP share code 10 or 11 trading on NASDAQ, regardless of listing venue.

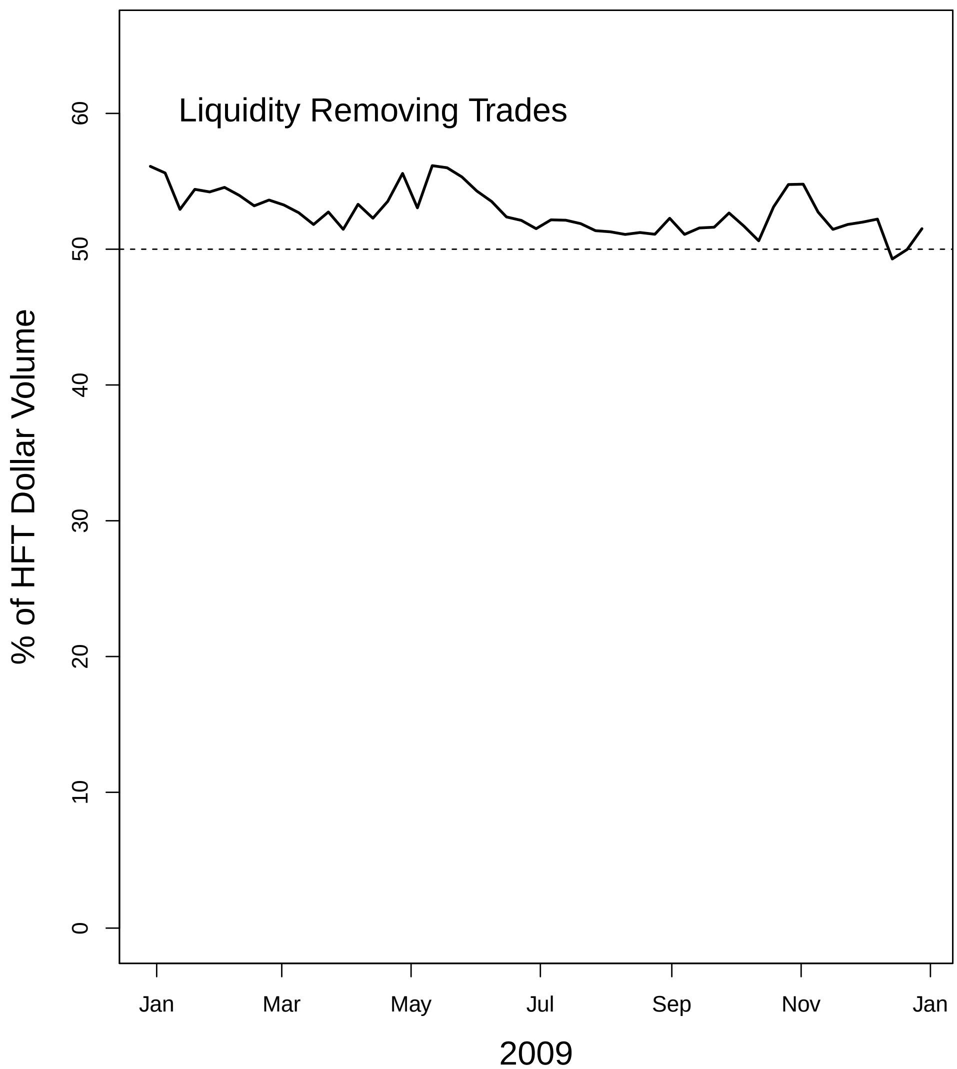


Figure A.2: Liquidity Removing Trades as a Percent of HFT Dollar Volume
This figure shows liquidity removing trades as a percent of HFTs' dollar volume on the NASDAQ Stock Market. Liquidity removing trades are those in which the HFT initiates the trade with a marketable order, which is functionally equivalent to a market order. The calculation includes all stocks with CRSP share code 10 or 11 trading on NASDAQ, regardless of listing venue.

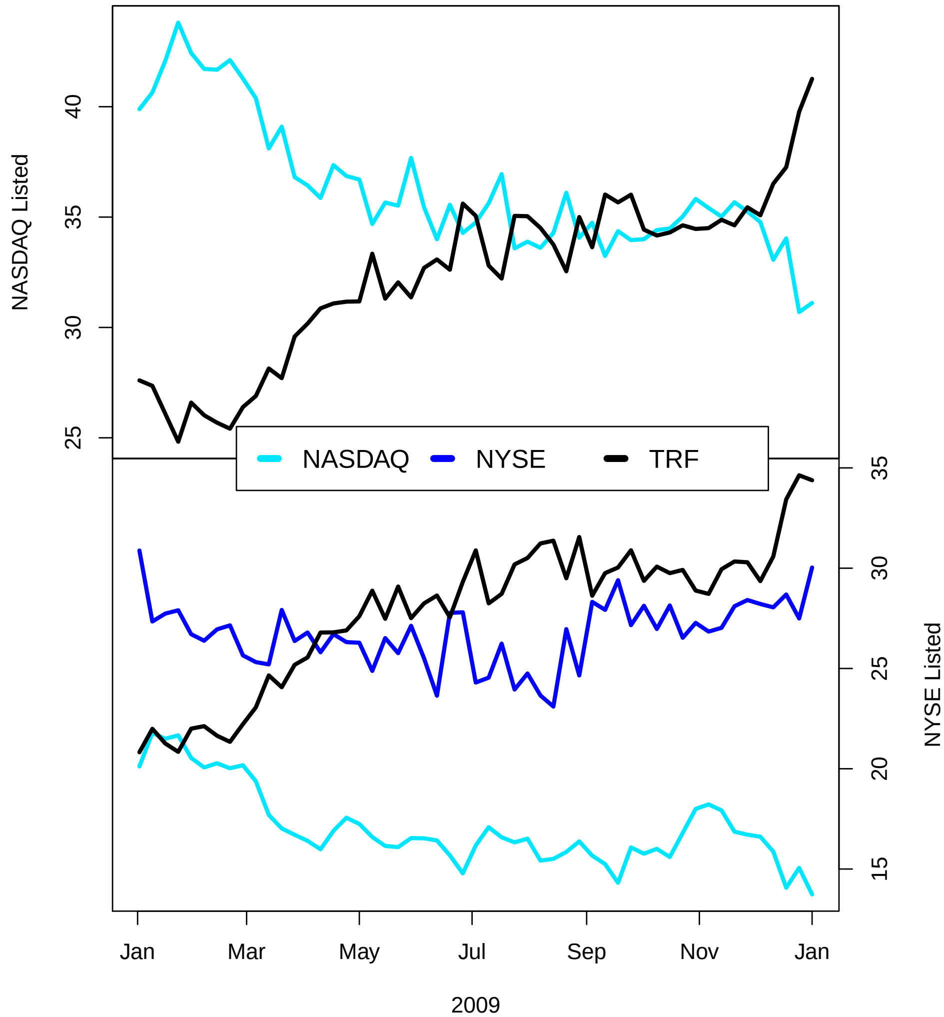


Figure A.3: Market Share By Trading Venue

Market share is reported as percent of dollar volume. NASDAQ is the NASDAQ Stock Market, NYSE is the New York Stock Market, and TRF is the FINRA Trade Reporting Facility that includes trades that do not occur on a stock exchange (e.g., trades executed in dark pools or by off-exchange market making firms).

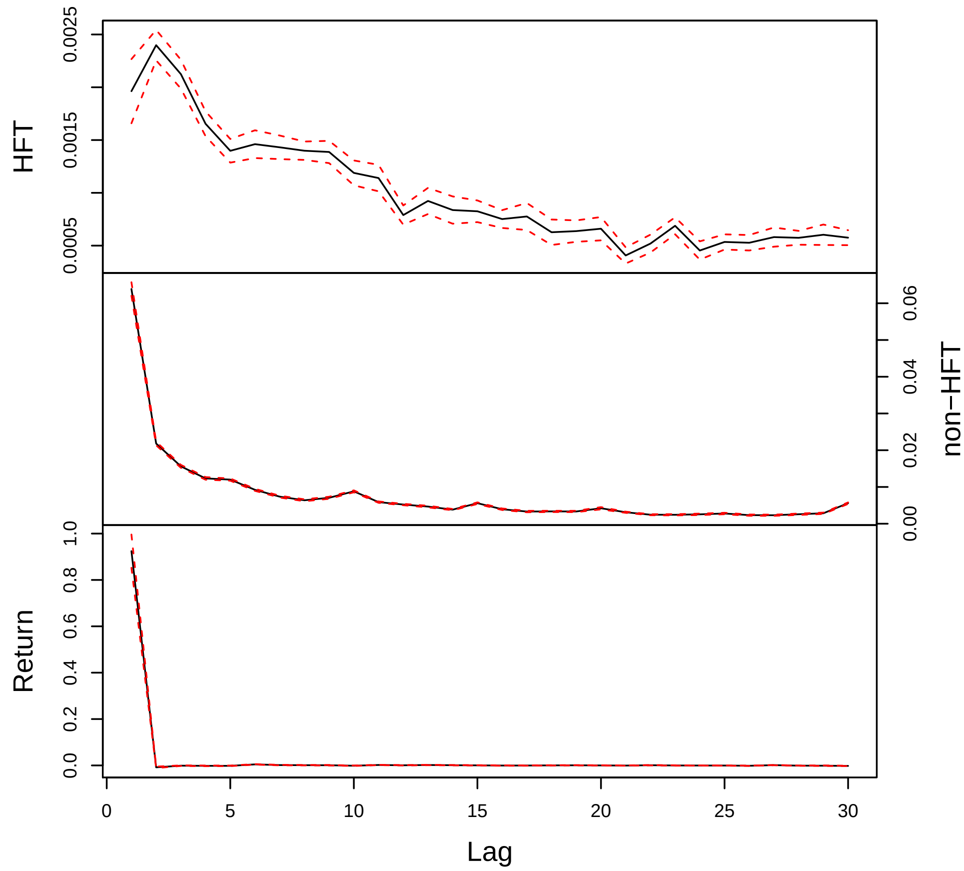


Figure A.4: Coefficients from VAR with 30 lags

This figure plots the mean coefficients (solid line) and 95% confidence intervals (dotted lines) from a VAR system similar to the one in Table 1.6, with the difference being that it uses thirty lags rather than the ten used in Table 1.6. The coefficients are only plotted for the equation where non-HFT trading is the dependent variable. Coefficients are averaged across all observations for a day, and the mean in the figure is the average of the daily time series. Standard errors for the time-series mean are calculated following Newey and West (1994).

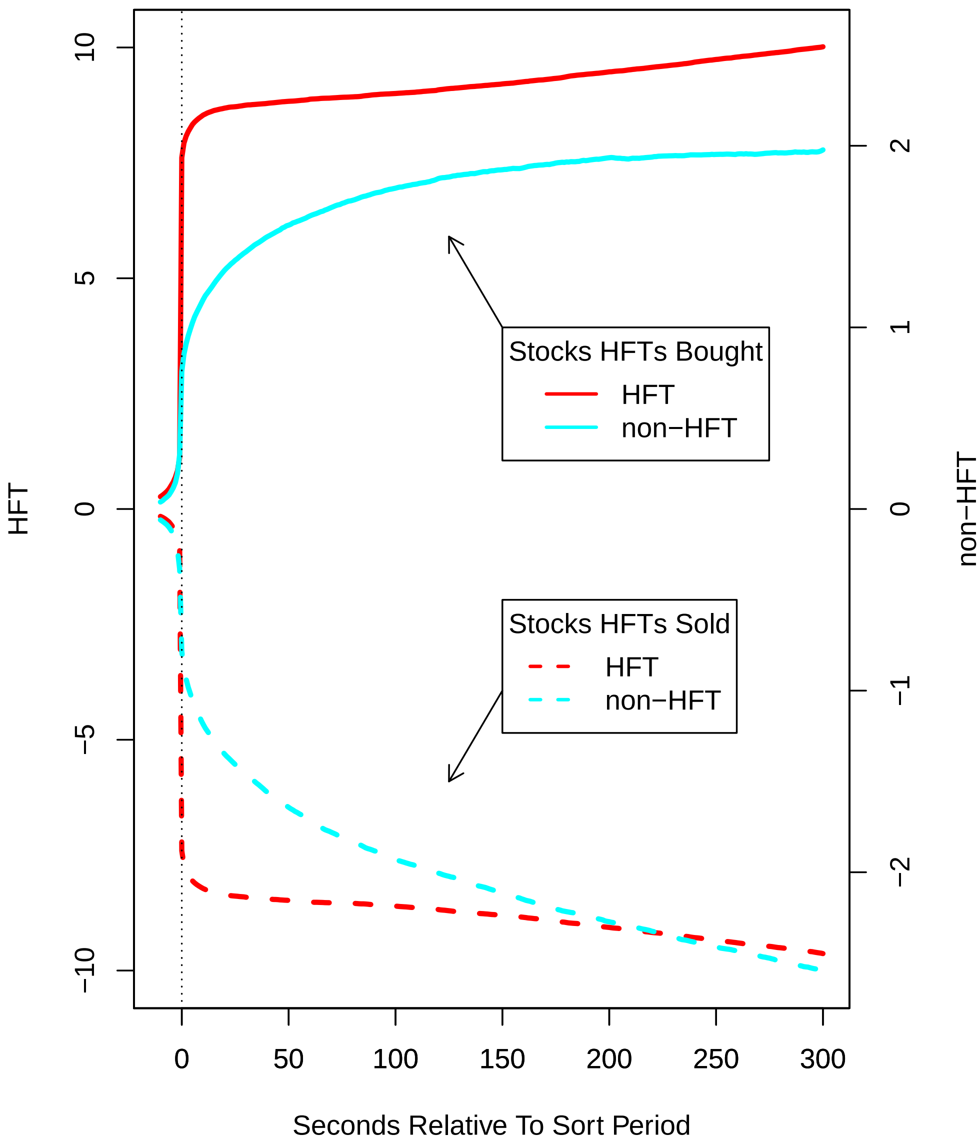
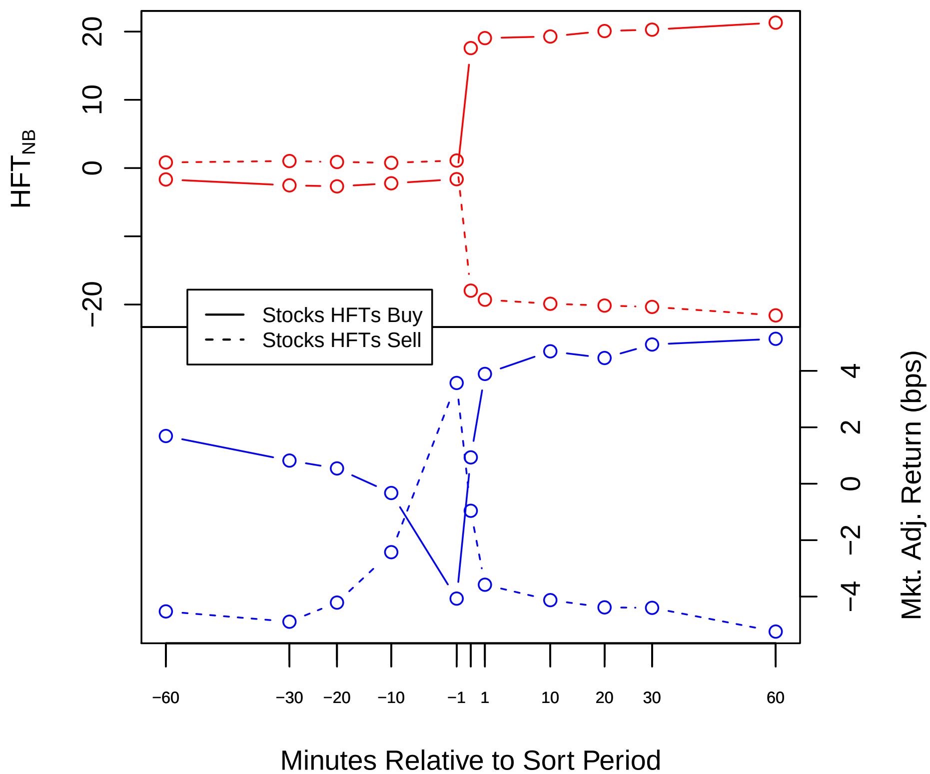
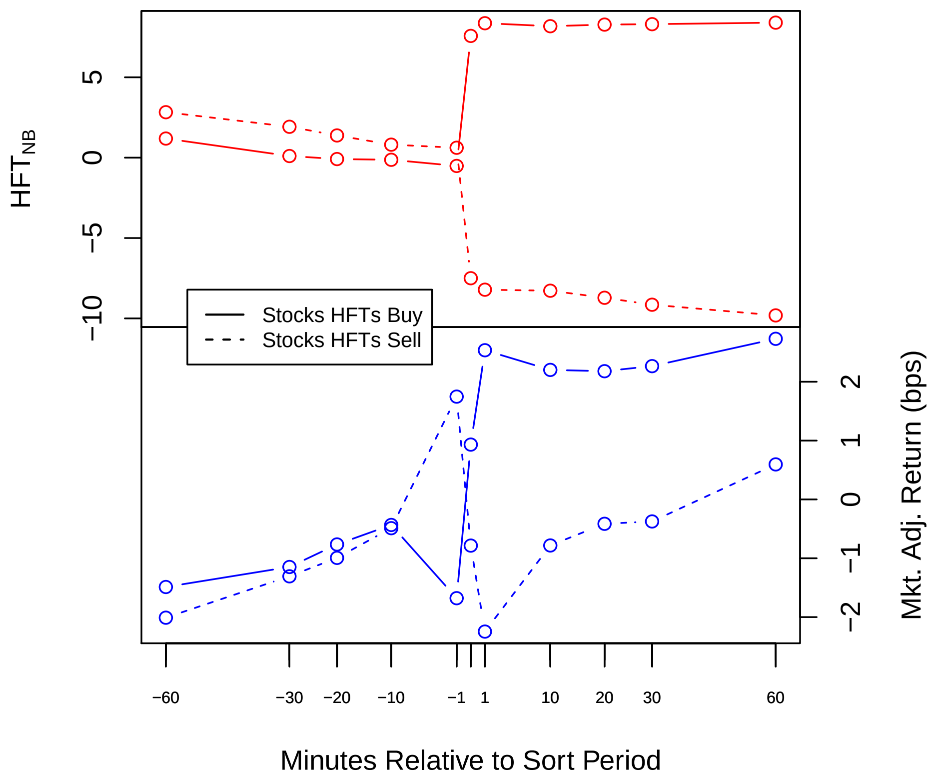


Figure A.5: Cumulative HFT Net Buying vs. non-HFT Net Marketable Buying
This figure examines whether HFTs reverse their positions after a period of intense net marketable buying. The left y-axis scale is cumulative standardized HFT net buying, HFT_{NB}, and the right y-axis scale is cumulative standardized non-HFT net marketable buying, $non\text{-}HFT_{NMB}$. Stocks are sorted into deciles based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten and for which HFT_{NMBSD} is greater than zero are marked as those HFTs bought. Stocks in decile one and for which HFT_{NMBSD} is less than zero are marked as those HFTs sold. The reason for conditioning on HFT_{NMBSD} rather than just HFT_{NMB} is that it ensures variation is driven by times when HFTs are either on net buying and buying aggressively or on net selling and selling aggressively. Table 1.2 describes construction of these imbalance measures.

Panel A: Small−cap Stocks



Panel B: Mid−cap Stocks



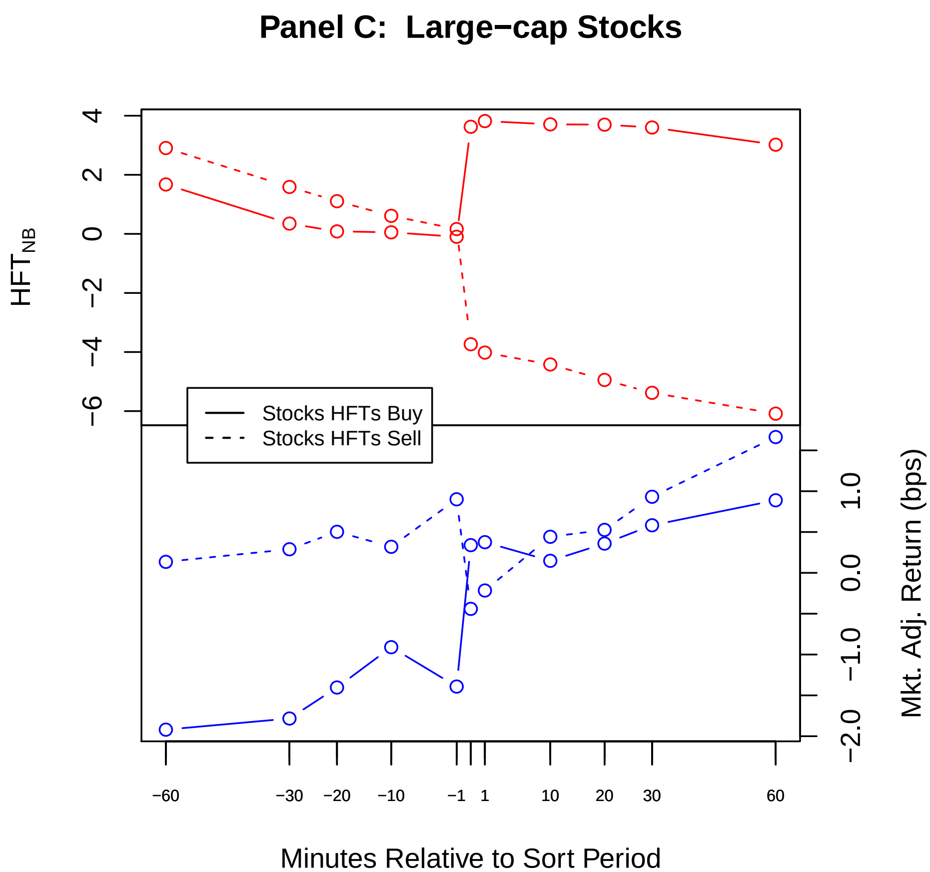


Figure A.6: HFT Net Buying and Returns 60 minutes before and after intense HFT Net Marketable Buying By Size Portfolio

The figure examines HFT net buying from 60 minutes before to 60 minutes after periods of intense HFT net marketable buying or selling. HFT_{NB} is cumulative standardized HFT net buying. Buy and hold returns are market adjusted using contemporaneous returns on SPY. Stocks are sorted into deciles based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten and for which HFT_{NMBSD} is greater than zero are marked as those HFTs bought. Stocks in decile one and for which HFT_{NMBSD} is less than zero are marked as those HFTs sold. The reason for conditioning on HFT_{NMBSD} rather than just HFT_{NMB} is that it ensures variation is driven by times when HFTs are either on net buying and buying aggressively or on net selling and selling aggressively. To handle clustering of observations, observations are first averaged by stock-day, then by day, and then finally across the complete time-series. Observations must have data from 60 minutes before to 60 minutes after the sort period, so the figure excludes the first and last hour of the trading day.

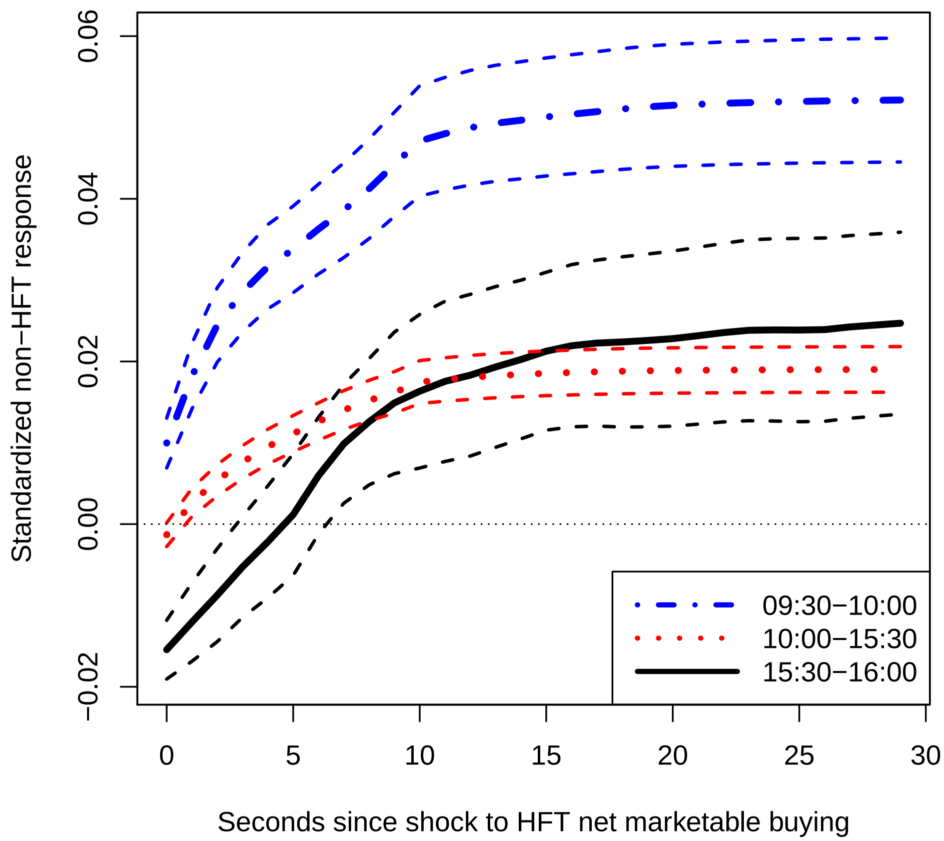


Figure A.7: Response of non-HFT Net Marketable Buying to a One Standard Deviation Shock to HFT Net Marketable Buying: Different Times of the Day
This figure plots the impulse response function describing the response of non-HFT net marketable buying, *non-*HFT_{NMB}, to a one standard deviation shock to HFT net marketable buying in the same direction as net buying, HFT_{NMBSD}. Table 1.2 describes construction of these imbalance measures. The response is expressed in standard deviations. The results are based on the vector autoregression (VAR) in Table 1.6. The VARs are estimated using 10 lags. Stock-day observations are excluded if any of the variables fail an augmented Dickey-Fuller test for stationarity. The impulse response function is orthogonalized to allow for contemporaneous effects. The ordering of the variables is such that HFT_{NMBSD} has a contemporaneous effect on *non-*HFT_{NMB} and returns. *non-*HFT_{NMB} has a contemporaneous effect on returns but does not contemporaneously affect HFT_{NMBSD}. Returns are assumed to have no contemporaneous effect on either trading measure. Impulse response functions are estimated by stock each day, and then the daily cross-sectional mean is calculated. The solid line is the mean of the daily time series, and the dotted lines indicate 95% confidence intervals using standard errors calculated from the daily time-series of mean impulse response functions. The impulse response functions are calculated separately for the first 30 minutes of the trading day, from 30 minutes after the open to 30 minutes before the close, and for the last 30 minutes. For a stock to be included on a given day, there must be at least 25 non-zero observations per variable in each of the three time periods.

Appendix B

Heteroskedasticity in VAR estimates

Since the vector autoregressions are estimated at the stock level, differences in the precision of coefficient estimates among stocks may warrent different weightings for stocks when creating summary statistics. The precision of the coefficient estimates are going to be related to the number of non-zero observations for the stock during the day. While there are many seconds in the trading day, during most seconds, returns and imbalances are zero. If a stock has more seconds when returns and imbalances are non-zero, then the coefficients in the VAR will be estimated more precisely.

To get a sense for how the precision of these VAR estimates varies with the number of non-zero observations, Figure B.1 plots the distribution of individual stock-day coefficients on the first lag of HFT net marketable buying from the equation in which non-HFT net marketable buying is the dependent variable. The figure clearly shows that when the number of non-zero observations for a stock on a given day is small, the coefficient estimates are more variable. Consequently, giving more weight to stock days with more non-zero observations may improve estimates of the estimate for the full population.

Table B.1 compares how different weighting schemes affect estimates of the population average. The table reports time-series means and standard errors of daily weighted cross-sectional mean coefficient estimates. There are four different weighting schemes that are functions of the minimum number of non-zero observations among the independent variables for the stock that day, η_j. The first weights observations equally, which assumes all coefficients are estimated with equal precision. Figure B.1 suggests coefficients estimated with a larger η_j are more precise, and the three other weighting schemes account for this by giving more weight to stocks for which η_j is larger. The second weighting scheme assigns a weight of zero to stocks for which $\eta_j \leq 100$ and 1 to those for which $\eta_j > 100$. This weighting scheme just excludes the least precise estimates. The third schemes weights stocks by η_j, and the fourth weights stocks by $\sqrt{\eta_j}$. The fourth assumes precision is increasing in η_j, but at a decreasing rate, perhaps because of autocorrelation among nearby stock-day observations. The scheme that weights coefficient estimates by η_j has the lowest White and Newey-West standard error estimates, so that weighting scheme appears to be the most efficient.

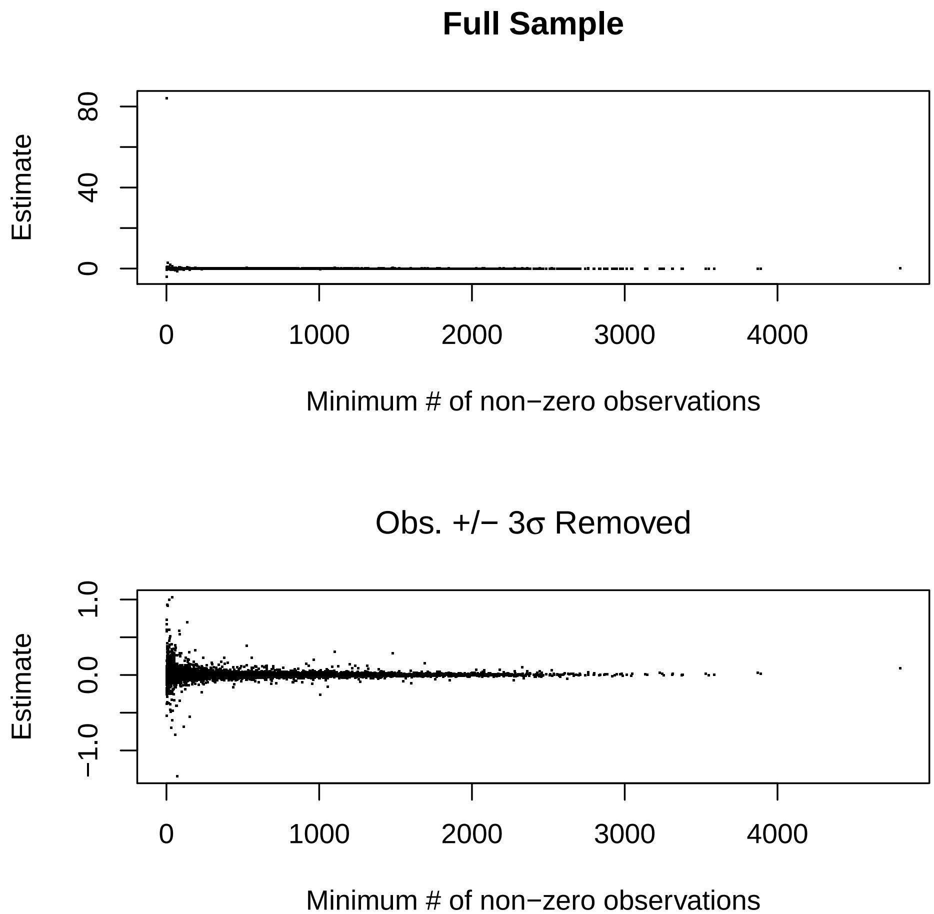


Figure B.1: Heteroskedasticity in stock-day VAR coefficients

The figure shows how the precision of coefficient estimates in the VAR varies by the number of non-zero observations for the stock the day the VAR is estimated. The figure plots coefficients on the first lag of HFT net marketable buying from a regression where the dependent variable is non-HFT net marketable buying and the independent variables are 10 lags of HFT net marketable buying, non-HFT net marketable buying, and returns. The x-axis is the minimum number of non-zero observations among the HFT, non-HFT, and return time series for the stock that day.

Table B.1: Different weighting schemes for VAR coefficients

This table compares different schemes for weighting stock-day VAR coefficient estimates. The schemes are evaluated using estimates of the coefficient on the first lag of HFT net marketable buying in the following OLS regression estimated separately for each stock each day,

$$non\text{-}HFT_{j,t} = \alpha_{j,t} + \sum_{i=1}^{10} \gamma_j HFT_{j,t-i} + \sum_{i=1}^{10} \beta_j non\text{-}HFT_{j,t-i} + \sum_{i=1}^{10} \lambda_j R_{j,t-i} + \epsilon_{j,t,non\text{-}HFT} \tag{B.1}$$

Where j indexes stocks and t indexes seconds. Estimates are averaged each day, and the table reports means and standard errors for the time series of daily cross-sectional means. The weightings refer to how the cross-sectional mean each day is calculated. The weights are functions of η_j, the minimum number of non-zero observations among the independent variables in the regression for stock j that day. Three standard error estimates are reported: standard OLS estimates, estimates following White (1980), and estimates following Newey and West (1994).

Weight $w_j =$	μ	SE_{OLS}	SE_{White}	$SE_{NeweyWest}$
1	0.00667	0.003634	0.003642	0.003603
If $\eta_j \leq 100 \Rightarrow 0$ else 1	0.00195	0.000220	0.000220	0.000312
η_j	0.00205	0.000213	0.000213	0.000307
$\sqrt{\eta_j}$	0.00253	0.000306	0.000306	0.000345

Bibliography

BATS Exchange, 2011, *US Equity/Options Connectivity Manual* 5.3.4 edn.

Battalio, Robert, Jason Greene, and Robert Jennings, 1997, Do competing specialists and preferencing dealers affect market quality?, *Review of Financial Studies* 10, 969.

Bertsimas, Dimitris, and Andrew W. Lo, 1998, Optimal control of execution costs, *Journal of Financial Markets* 1, 1–50.

Brogaard, Jonathan A., 2011a, The activity of high frequency traders, *Working Paper.*

——— , 2011b, High frequency trading and volatility, *Working Paper.*

Brunnermeier, M.K., and L.H. Pedersen, 2005, Predatory trading, *The Journal of Finance* 60, 1825–1863.

Cai, Fang, 2003, Was there front running during the ltcm crisis?, *Manuscript, Board of Governors of the Federal Reserve System.*

Castura, Jeff, Robert Litzenberger, Richard Gorelick, and Yogesh Dwivedi, 2010, Market efficiency and microstructure evolution in U.S. equity markets: A high-frequency perspective, *Manuscript, RGM Advisors, LLC.*

Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2009, Rise of the machines: Algorithmic trading in the foreign exchange market, *Board of Governors of the Federal Reserve System, mimeo.*

Chen, J., S. Hanson, H. Hong, and J.C. Stein, 2009, Do hedge funds profit from mutual-fund distress?, *Manuscript, NBER.*

Chorida, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2002, Order imbalance, liquidity, and market returns, *Journal of Financial Economics* 65, 111–130.

Creswell, Julie, 2010, Speedy new traders make waves far from Wall St., *New York Times* p. A1.

Dow Jones, 2011, Algorithmic trading and quantitative trading for the enterprise, `http://www.itg.com/offerings/posit-crossing/posit-marketplace/` Accessed: 9/28/2011.

Easley, David, Marcos M. Lopez de Prado, and Maureen O'Hara, 2011, The microstructure of the 'flash crash': Flow toxicity, liquidity crashes and the probability of informed trading, *The Journal of Portfolio Management* 37, 118–128.

E*Trade, 2009, Etrade q3 2009 rule 606 filing, `www.etrade.com`.

Foucault, Thierry, and Albert J. Menkveld, 2008, Competition for order flow and smart order routing systems, *The Journal of Finance* 63, 119–158.

Gervais, S., R. Kaniel, and D.H. Mingelgrin, 2001, The high-volume return premium, *The Journal of Finance* 56, 877–919.

GETCO, 2010, Our firm: Getco, `http://www.getcollc.com/index.php/getco/tertiary/our_firm/` Accessed: 6/2/2010.

Griffin, J.M., J.H. Harris, and S. Topaloglu, 2003, The dynamics of institutional and individual trading, *The Journal of Finance* 58, 2285–2320.

Hasbrouck, Joel, and Gideon Saar, 2011, Low-latency trading, *Manuscript, Cornell University.*

Hawkins, O. Mason, Richard W. Hussey, Andrew McCarroll, Jeffrey D. Engelberg, and W. Douglas Schrank, 2010, Comments & analysis on equity market structure, `http://www.sec.gov/comments/s7-02-10/s70210-228.pdf` Accessed: 4/27/2010.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does algorithmic trading improve liquidity?, *The Journal of Finance* 66, 1–33.

Hendershott, Terrence, and Pamela C. Moulton, 2011, Automation, speed, and stock market quality: The NYSE's hybrid, *Journal of Financial Markets, forthcoming.*

Hendershott, Terrence, and Ryan Riordan, 2011a, Algorithmic trading and information, *Working Paper.*

——— , 2011b, High-frequency trading and price discovery, *Manuscript, University of California at Berkeley.*

Holden, Craig W., and Avanidhar Subrahmanyam, 1992, Long-lived private information and imperfect competition, *Journal of Finance* pp. 247–270.

Jegadeesh, N., and S. Titman, 1995, Short-horizon return reversals and the bid-ask spread, *Journal of Financial Intermediation* 4, 116–132.

Jovanovic, Boyan, and Albert J. Menkveld, 2010, Middlemen in limit order markets, *Manuscript, VU University Amsterdam.*

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tuzun Tuzun, 2011, The flash crash: The impact of high frequency trading on an electronic market, *Manuscript, U of Maryland*.

Kisling, Whitney, 2010, GETCO buys 350 designated market maker slots at NYSE, *Bloomberg* February 11.

Kraus, A., and H.R. Stoll, 1972, Parallel trading by institutional investors, *Journal of Financial and Quantitative Analysis* 7, 2107–38.

Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1385.

Lakonishok, J., A. Shleifer, and R.W. Vishny, 1992, The impact of institutional trading on stock prices, *Journal of Financial Economics* 32, 23–43.

Longstaff, F.A., 2001, Optimal portfolio choice and the valuation of illiquid securities, *Review of financial studies* 14, 407.

Mayhew, Stewart, 2002, Competition, market structure, and bid-ask spreads in stock option markets, *The Journal of Finance* 57, 931–958.

McCarthy, John A., 2010, Concept release on equity market structure, `http://www.getcollc.com/images/uploads/Getco_comment_letter_re_Concept_Release-_clear_copy.pdf` Accessed: 4/27/2010.

Newey, Whitney K., and Kenneth D. West, 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies* 61, 631.

O'Hara, Maureen, Chen Yao, and Mao Ye, 2011, What's not there: The odd-lot bias in taq data, *Working Paper*.

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality?, *Journal of Financial Economics*.

POSIT, 2010, Posit marketplace, `http://www.itg.com/offerings/posit-crossing/posit-marketplace` Accessed: 5/30/2010.

Riordan, Ryan, and Andreas Storkenmaier, 2009, Latency, Liquidity and Price Discovery, *Manuscript, Karlsruhe Institute of Technology* presented at 21st Australasian Finance and Banking Conference 2008, 9th Annual London Business School Transatlantic Doctoral Consortium 2009, 16th Annual Meeting of the German Finance Association 2009.

Schwab, 2009, Schwab q3 2009 rule 606 filing, `www.schwab.com` Accessed: 12/21/2009.

Spicer, Jonathan, and Herbert Lash, 2009, Who's afraid of high-frequency trading?, *Reuters*.

Sukumar, Nandini, 2011, High-frequency trading is 77% of u.k. market, tabb group says, *Bloomberg*.

Tabb, Larry, 2009, High frequency trading: What is it & should i be worried?, `http://www.world-exchanges.org/files/file/LarryTABB-HFT-part1.ppt` Accessed: 8/9/2012.

Tradebot, 2010, About us: Tracebot systems, `http://www.tradebotsystems.com` Accessed: 6/2/2010.

Wermers, Russ, 1999, Mutual fund herding and the impact on stock prices, *The Journal of Finance* 54, 581–622.

White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica: Journal of the Econometric Society* pp. 817–838.