Do High-Frequency Traders Anticipate Buying and Selling Pressure?

Nicholas H. Hirschey *
London Business School

April 1, 2013

Abstract

High-frequency traders (HFTs) accounted for 40% of NASDAQ volume in 2009, but we know little about how their trading affects liquidity. This study examines one means by which HFTs could increase non-HFTs' trading costs—namely, by anticipating and trading ahead of their order flow. I find that HFTs' aggressive purchases and sales lead those of other investors. The effect is stronger at times when non-HFTs may be more impatient, such as near the market open, on high volume and high imbalance days, and for stocks with wide bid-ask spreads. I explore whether these results are explained by HFTs reacting faster to news, positive-feedback trading by non-HFTs, or HFTs and non-HFTs trading on the same signals, but the results are best explained by the anticipatory trading hypothesis. Consistent with the idea that such trading is related to HFT skill, there is persistence in which HFTs' trades best forecast order flow, and these HFTs' trades are more highly correlated with future returns. While it is probable HFTs on net improve liquidity, these findings support the existence of an anticipatory trading channel through which HFTs may increase non-HFT trading costs.

*Contact information: Department of Finance, London Business School, Regent's Park, London NW1 4SA, United Kingdom; Email: nhirschey@london.edu. I am grateful to my PhD committee, Carlos Carvalho, John Griffin (co-Chair), Clemens Sialm, Laura Starks, and Sheridan Titman (co-Chair), for their advice and suggestions. I would like to thank NASDAQ OMX for providing the data and Frank Hatheway, Claude Courbois, and Jeff Smith for many useful discussions about the data and market structure. I would also like to thank Andres Almazan, Fernando Anjos, Robert Battalio, Jonathan Cohn, Shane Corwin, Andres Donangelo, Cesare Fracassi, Terry Hendershott, Paul Irvine, Richard Lowery, Tim Loughran, Jeffrey Pontiff, Mark Seasholes, and seminar participants at Boston College, the Georgia Institute of Technology, the Hong Kong University of Science and Technology, the London Business School, Rice University, the University at Buffalo, the University of California Irvine, the University of Georgia, the University of Melbourne, the University of Miami, the University of Notre Dame, the University of Texas at Austin, and the Western Finance Association meeting for helpful comments and suggestions.

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. However, there is fairly little known about how their trading affects liquidity. To the extent that HFTs act simply as market makers, they will tend to improve liquidity. But HFTs also search trade and order data for clues about where prices will go in the future, and when they trade on this information, they may compete with long-term investors for liquidity, thereby increasing those investors' trading costs.

This paper studies one aspect of HFTs' effect on liquidity by examining whether HFTs' liquidity removing trades arise from strategies that anticipate and trade ahead of traditional asset manager order flow. An HFT may anticipate the trades of a mutual fund, for instance, if that mutual fund splits large orders into a series of smaller ones and their initial trades reveal information about their future trading intentions. If, indeed, an HFT were able to forecast a traditional asset manager's order flow, then the HFT may have an incentive to trade ahead of the traditional asset manager in order to profit from their subsequent price impact.

Anticipatory trading of this form has the potential to affect both liquidity and price efficiency. An HFT buying stock that non-HFTs intend to buy could cause stock prices to rise right before the non-HFTs trade, thereby increasing their trading costs. If the non-HFT is trading to fund a liquidity shock, the HFT's purchase reduces efficiency by temporarily driving the stock's price above its fundamental value (Brunnermeier and Pedersen 2005). If in contrast the HFT is anticipating an informed non-HFT trade, then their purchase pushes prices towards fundamental values faster than would otherwise be the case (e.g., Holden and Subrahmanyam 1992). But in capturing some of the non-HFT's information rent, the HFT reduces the non-HFT's profits and, in consequence, their incentives to do fundamental research. Thus, the long-run effect could be a decline in information production (Grossman and Stiglitz 1980). For these reasons, it is important to understand the extent to which HFTs anticipate and trade ahead of other investors' order flow.

Given the potential costs imposed on non-HFTs if HFTs are able to anticipate their trades, non-HFTs employ execution algorithms designed to acquire or dispose of positions without revealing their trading intentions. HFTs' ability to anticipate buying and selling pressure, then, depends on the outcome of competition between algorithms used by HFTs and non-HFTs. To the extent that non-HFTs are constrained by their desire to enter or exit their position, they will be at a disadvantage in this competition.

To examine these issues, I analyze return and trade patterns around periods of aggressive buying and selling by HFTs using an entire year of unique trade and trader-level 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. I hypothesize that if HFTs anticipate and trade ahead of non-HFT order flow, then when an HFT buys a stock aggressively, this should forecast future aggressive buying by non-HFTs as well as an increase in price.

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, non-HFT net marketable buying for the stocks bought most aggressively by HFTs rises by a cumulative 66% of its one-second standard deviation 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. 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., the 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. Instead, the evidence is most consistent with the anticipatory trading hypothesis. Specifically, the evidence points to there being periods of semi-persistent non-HFT buying or selling pressure that HFTs recognize in real time. Since non-HFT liquidity providers do not notice and update their quotes, the result is short-term price momentum, which is taken advantage of by HFTs. Particularly in small and mid-cap stocks, these shocks to non-HFT order flow seem to reflect information.

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. Indeed, trades from HFTs that were the most 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. Consistent with the idea that these HFTs are more skilled, their marketable trades are also more strongly correlated with future returns.

If HFT trades flow primarily from market making activities, then it might follow that marketable imbalances studied in this paper arise from market makers disposing of inventory positions. Thus, HFTs selling ahead of anticipated non-HFT order flow would be doing so to avoid losses on inventory positions rather than to profit on directional bets. To evaluate this issue, I examine cumulative HFT net buying and stock returns from one hour before to one hour after seconds with intense HFT net marketable buying. If a period with large marketable purchases is disposing of a previously acquired short inventory position, then one would expect HFTs to have previously been providing liquidity to buyers of the stock. But neither prior cumulative net buying nor prior returns provide evidence trades in the sort period are disposing of previously acquired market-maker inventories.

This study contributes to research examining how evolutions in market structure, including greater competition among trading venues and increasing automation, affect stock trading. The literature shows that trading costs on the NYSE tend to be lower when specialists compete with liquidity providers on non-listing exchanges (Battalio, Greene, and Jennings 1997, Brown, Mulherin, and Weidenmier 2008) or broker-dealers executing trades off the NYSE (Battalio 1997). Similarly, stocks with proportionately more trading occurring in off-exchange venues have lower spreads (O'Hara and Ye 2011), and spreads for stocks listed on Euronext decline when they begin trading on additional market centers (Foucault and Menkveld 2008, Jovanovic and Menkveld 2011). The findings of these studies, which generally suggest competition among equity trading venues lowers trading costs, are supported by evidence from options markets (Mayhew 2002, Battalio, Shkilko, and Van Ness 2011). Comerton-Forde and Putnins (2012) find that Australian stocks' prices deviate more from a random walk when a greater share of their trading occurs outside the ASX's central limit-order book. Biais, Foucault, and Moinas (2013) find there can be too much investment in fast trading technology when investors endogenously choose how fast to trade. Research into the effects of automation by Hendershott, Jones, and Menkveld (2011) shows that an increase in electronic message traffic on the NYSE is associated with lower bid-ask spreads and less price discovery through trades. Boehmer, Fong, and Wu (2012) find in an international sample that the introduction of co-location facilities is associated with lower bid-ask spreads, less autocorrelation in prices, and higher volatility. Hendershott and Riordan (2012) show that trades and quotes entered by algorithms on the Deutsche Boerse supply relatively more liquidity when spreads are wide. 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. Hasbrouck (2013) examines very short-term volatility of bid and ask quotes. But the empirical studies above cannot identify which trades and quotes come from HFTs, so they are unable to specifically examine how HFTs affect other market participants—a topic of immense interest to both investors and regulators. In contrast, this paper observes investor identities, which allows for studying interaction between HFTs and non-HFTs as well as differences among HFTs themselves.

While other papers have examined HFTs, this is the first paper to find that HFTs anticipate buying and selling pressure from other investors, that some HFTs are better than others at anticipating order flow, and that HFTs chase very short-term price trends. In doing so, this paper contributes to a growing literature studying how market participants' adoption of algorithmic trading strategies affects price discovery and liquidity in financial markets. Brogaard, Hendershott, and Riordan (2013) find that HFT purchases executed using marketable orders precede price increases, while purchases executed using non-marketable orders precede price declines (and vice versa, for sales). O'Hara, Yao, and Ye (2012) show that trades of 100 or fewer shares, sizes commonly used by HFTs and broker execution algorithms, account for a substantial portion of price discovery. Menkveld (2013) shows that bid-ask spreads fall when a large HFT begins trading on Chi-X. Kirilenko, Kyle, Samadi, and Tuzun (2011) argue that while HFTs did not cause the market crash on May 6, 2010, they did contribute to the heightened volatility. This paper builds on this literature by providing evidence that one reason HFTs appear informed is they predict price changes caused by other investors' buying and selling pressure. An implication for informed non-HFTs is that because HFTs trade ahead of them, the non-HFTs trade fewer shares before prices adjust to their information. From the perspective of non-HFTs, the effect is analogous to increasing their price impact, which is an important component of trading costs. So while it is probable HFTs on net improve liquidity, this study's findings support the existence of an anticipatory trading channel through which HFTs may at times increase non-HFT trading costs.

HFTs trading ahead of informed non-HFT order flow implies two opposing effects on price efficiency. While the non-HFTs' information gets into prices faster, HFTs capture some of the informed non-HFTs' profits and, consequently, decrease non-HFTs' incentives to become informed. Thus, a benefit due to an increase in the speed at which information is incorporated into prices would be reduced by the fact it decreases investors' incentives to acquire new information. This highlights the point that when evaluating whether HFTs make prices more efficient, it is important to take into account the source of the information they use to trade. If the information on which HFTs trade would get into prices soon by some other means, such as through trading by non-HFTs, then this moderates the welfare benefits of HFT participation in the price discovery process.

The structure of the paper is as follows. Section 1 provides background on electronic trading and discusses the data. Section 2 explains and tests empirical implications of the anticipatory trading hypothesis, while Section 3 examines alternative explanations. Section 4 examines whether certain HFTs are more skilled at predicting order flow. Section 5 examines whether non-HFT order flow is easier to predict at times when non-HFTs are hypothesized to be impatient. Finally, Section 6 concludes.

1 Background and Data

1.1 Background on Electronic Trading and HFTs

Since the data primarily come from NASDAQ, it is helpful to explain NASDAQ's structure and relation to other trading venues. NASDAQ is structured as an electronic limit-order book. This is essentially the same market structure as all other stock exchanges (e.g., the NYSE, ARCA, BATS, and DirectEdge). Executions on exchanges such as NASDAQ predominately come from professional traders, because most retail brokerages have contracts with market making firms who pay for the right to fill retail orders.[1] NASDAQ trades both NASDAQ and NYSE-listed stocks, and its share of dollar volume in 2009 was roughly 36% in NASDAQ-listed securities and 17% in securities listed on the NYSE.[2] The remainder of U.S. equity trading was spread among other exchanges and off-exchange trading venues such as broker crossing networks, market making firms, and dark pools.[3]

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

[1]For example, in the third 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. See Battalio and Loughran (2008) for a discussion of these relationships.

[2]Appendix Figure A3 shows the time series of the market share volume breakdown by listing listing venue.

[3]Examples include ITG's POSIT Marketplace, Credit Suisse's Crossfinder, and Knight Capital.

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" (Tradebot 2010, GETCO 2010).

1.2 Sample construction

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.[4] The trade 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.

The sample stocks are chosen to be representative of those in which actively managed mutual funds invest. The sample is constructed from CRSP common stocks, identified by stocks having share code 10 or 11.[5] I exclude the bottom two NYSE size deciles from the sample to roughly match common definitions of active funds' investable universe (e.g., Russell 3000 or MSCI Investable Market 2500). These 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

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

[5] 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. Appendix Table A1 summarises stock-day observations of CRSP common stocks with dual-class shares removed.

the sample universe to 1,882 stocks.[6] 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 reports summary statistics for all stock days. The sample averages 93 stocks per trading day.[7] Market capitalization ranges from $22 million to $125,331 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. Dollar volume increases 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. On average, 27.2% of the sample stocks' dollar volume trades on NASDAQ, and this value is fairly constant across size portfolios.

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 Trade imbalances and returns

This study uses net marketable buying and net buying imbalances. A net marketable buying imbalance, defined as shares in buyer-initiated trades minus shares in seller-initiated trades, is a common measure of buying and selling pressure from the existing literature (e.g., Chorida, Roll, and Subrahmanyam 2002). Though Hasbrouck and Saar (2009) show limit orders sitting in the order book are used in increasingly active strategies, their use is still generally consistent with passive liquidity provision. Hence, the marketable imbalance is an intuitive measure of trading demand. The net buying imbalance is simply shares bought

[6] Appendix Table A2 summarises stock-day observations for this sample of stocks.

[7] The number of stocks varies because stocks are removed from the sample any day during which the prior day's closing price is less than $1 and permanently removed if daily dollar volume falls below $100,000.

minus shares sold and has previously been used to measure position changes of different investor groups (e.g., Griffin, Harris, and Topaloglu 2003). For HFTs, I predominately use a modified net marketable buying measure that sets HFT net marketable buying to zero if HFT net marketable buying and net buying are in opposite directions.[8] The purpose of this modified HFT net marketable buying measure is to ensure that when the HFT net marketable buying imbalance is positive, HFTs are on net buying more shares than they are selling. 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.[9]

Panel B in Table 1 summarizes trade imbalances for the sample stocks prior to their being standardized by trailing volume. 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. In practice, there is little difference between the two HFT net marketable buying measures. The mean standard deviation of HFTs' net buying (HFT_{NB}) among all stock days is 83 shares, compared to 80 shares for their net marketable buying (HFT_{NMB}) and 76 shares for their net marketable buying when it is the same direction as their net buying (HFT_{NMBSD}). These figures are slightly smaller than the 100 shares that O'Hara, Yao, and Ye (2012) report as the median trade size on NASDAQ in 2008 and 2009. The average standard deviation of non-HFTs' net marketable buying ($non\text{-}HFT_{NMB}$), at 125 shares, is somewhat higher than that of HFTs. The wide variation in imbalance standard deviations among size portfolios motivates the normalization by trailing volume in later results.

Intra-day returns are calculated using bid-ask midpoints from the National Best Bid and Best Offer (NBBO). The NBBO aggregates quotes from all displayed order books, so it is the

[8]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.

[9] One could also adjust by the second or minute of the trading day to account for intra-day volume patterns (Jain and Joh 1988), but such estimates for thinly traded stocks can by noisy. Potential effects related to the time of day are examined in Table 9.

best measure of a stock's quoted price.[10] These quote data are filtered to remove anomalous observations.[11] Table 1 reports the distribution of the standard deviation of NBBO bid-ask midpoint returns across all stock days. The median standard deviation is 0.031% among small-cap stocks, 0.022% among mid-cap stocks, and 0.023% among large-cap stocks.

1.4 News articles

Certain tests use articles from the Factiva news archive. Factiva contains news from over 35,000 sources, including most major newswires, newspapers, and magazines. Prior studies provide evidence these articles contain value-relevant information (e.g., Tetlock 2007, Tetlock, Saar-Tsechansky, and Macskassy 2008, Griffin, Hirschey, and Kelly 2011). Factiva tags articles with identifiers indicating which firms are covered in an article, and these identifiers are used to match articles to the sample firms.[12]

2 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

[10] The largest displayed order books are the NYSE, NASDAQ, AMEX, Archepelago, BATS, and DirectEdge.

[11] 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.

[12] Table A3 summarises the frequency of coverage and top sources for sample firms in the Factiva archive.

HFT could potentially trade ahead of them and profit from the traditional asset manager's 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, since the HFT's objective would be to profit from non-HFTs' subsequent price impact, it should be the case that the prices of the stocks they buy rise and those of the stocks they sell fall. These two patterns, together, are consistent with HFTs trading stocks in order to profit from non-HFTs' future buying and selling pressure.

The analysis begins with portfolio sorts to identify the stocks HFTs are buying or selling aggressively. Stocks are sorted each second into decile portfolios 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. By this method, each stock is assigned to one of ten portfolios each second. Then, a daily mean of the variable of interest for each portfolio is calculated by taking an average among all stock-second observations for the portfolio that day:

$$V_{d,p} = \frac{1}{N_p} \sum_{i,t} V_{d,p,i,t},$$

where V is the variable of interest, d indexes days, p indexes portfolios, N_p is the number of stock-second observations in the portfolio that day, i indexes stocks, and t indexes seconds. Hypothesis tests are based on the means of these daily time series. This methodology is used to examine non-HFT and HFT net marketable buying as well as returns of stocks at different times relative to the sort period.

Figure 1 plots cumulative net marketable buying for the decile portfolios HFTs are buying and selling most intensely during the sort period. The solid lines are for decile ten, the stocks HFTs are buying most intensely, and the dotted lines are for decile one, the stocks

HFTs are selling most intensely. The dark red lines indicate HFTs' cumulative net marketable buying in the same direction as their net buying plotted on the left y-axis. This is the cumulated version of the variable used to sort the stocks. The light blue lines indicate non-HFTs' cumulative net marketable buying plotted on the right y-axis. The objective for plotting cumulative net marketable buying for the two investor groups is to determine whether aggressive buying by HFTs forecasts aggressive buying by non-HFTs. If this relationship exists, then after the sort period, cumulative non-HFT net marketable buying will increase for the stocks HFTs bought most aggressively and decrease for the stocks HFTs sold most aggressively.

The figure shows that prior to the sort period, for the stocks HFTs buy most aggressively, net marketable buying for both investor groups is positive. At time zero, the sort period, cumulative net marketable buying for HFTs spikes upward and then is relatively flat afterwards. Cumulative net marketable buying from non-HFTs also spikes during the sort period, and then, importantly, it continues to increase afterwards. This increase afterwards means non-HFTs are buying more with market orders than they are selling. The picture for the stocks HFTs sold most aggressively is symmetric. Thus, the figure illustrates results consistent with HFT net marketable buying leading non-HFT net marketable buying.

Table 2 presents this sort data in a form conducive to hypothesis tests. The table reports non-HFT net marketable buying from thirty seconds before to five minutes after the sort period for all stocks and for stocks split by size portfolios. As indicated in the figure, the table shows HFT net marketable buying is positively correlated with lagged, contemporaneous, and future net marketable buying from non-HFTs. The purpose of the table is to determine whether positive non-HFT net marketable buying for the stocks HFTs buy aggressively and negative non-HFT net marketable buying for the stocks HFTs sell aggressively are significantly different from zero in the post-sort periods. The table shows non-HFT net marketable buying for stocks HFTs are buying most aggressively is 0.09 times the one-second standard deviation of net marketable buying in the first second after the sort period. The median stock's one-second standard deviation of non-HFT net marketable buying is 42 shares, so this implies non-HFTs buy 3.8 more shares with marketable orders than they sell in the second after intense marketable buying by HFTs. In the first thirty seconds and five minutes after the sort period, cumulative non-HFT net marketable buying rises to 0.66 and 1.22 times the one-second standard deviation (or 28 and 51 shares for the median stock). For the stocks HFTs sell most aggressively, the figures are −0.09, −0.68, and −1.76 times the one-second standard deviation of non-HFT net marketable buying. All six values are significantly different from zero. The same holds for deciles two and nine, which are the portfolios with the next most extreme levels of HFT aggressive buying and selling. These tests indicate the post-sort changes in cumulative non-HFT net marketable buying illustrated in Figure 1 are significantly different from zero.

Table 2 also suggests the relationship between HFT trading and future non-HFT trading is stronger among small-cap stocks. In small-cap stocks, non-HFT net marketable buying in the five minutes after the sort period is 2.55 times the one-second standard deviation of net marketable buying in the stocks HFTs buy most aggressively and −3.22 times the one-second standard deviation in the stocks they sell most aggressively, compared to 1.22 and −1.73 in mid-cap stocks and 0.86 and −0.92 in large-cap stocks. If the pattern is due to HFTs anticipating non-HFT order flow, then the larger post-sort levels of non-HFT net marketable buying in small-cap stocks may be due, for example, to non-HFTs being more impatient when trading relatively more illiquid stocks. Later sections examine this hypothesis in more detail.

If the patterns in Figure 1 and Table 2 are due to HFTs anticipating non-HFT buying and selling pressure, then we should also see that the stocks that are bought aggressively have positive future returns and that the stocks that are sold aggressively have negative future returns. Figure 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. The spread between the average returns of these two groups of stocks widens throughout the first thirty seconds after the sort period. These post-sort return patterns coincide with the buying and selling pressure from non-HFTs illustrated in Figure 1.

Table 3 reports the magnitude of these returns across all stocks and for each size portfolio. As indicated in the figure, among all sample stocks, average returns in the thirty seconds after the sort period are positive for the stocks bought most intensely by HFTs, 1.23 basis points, and negative for the stocks sold most intensely by HFTs, −1.04 basis points. Both portfolios' return changes are significantly different from zero. Over the next four and a half minutes, there is some reversal in these returns, to 0.62 basis points for the stocks HFTs buy and −0.41 basis points for the stocks HFTs sell, but the spread in the return between these two portfolios nonetheless remains positive. The finding that price changes forecasted by HFT marketable trades last at least five minutes is consistent with Brogaard, Hendershott, and Riordan's (2013) findings that HFT marketable trades forecast permanent price changes using a state-space model. The return patterns here, in combination with the patterns in net marketable buying shown in Table 2 and Figure 1, are consistent with HFTs anticipating price changes caused by buying and selling pressure from traditional asset managers.

As might be expected given prior results that the magnitude of post-sort non-HFT net marketable buying is larger among stocks with lower market capitalizations, small-cap stocks exhibit larger post-sort return spreads. Thirty seconds after the sort period, returns for small-cap stocks that were bought versus sold are 2.54 versus −2.59 basis points, compared to 1.52 and −1.28 basis points for mid-cap stocks and 0.41 and −0.14 basis points for large-cap stocks. Figure 3 illustrates these return differences in more detail. In small and mid-cap stocks, the spread between the stocks bought versus sold by HFTs widens between the first and the thirtieth second after the sort period, whereas in large-cap stocks the return spread begins to narrow over that time span. By five minutes after the sort period, prices of large-cap stocks have fully reversed, with prices of stocks bought by HFTs essentially the same as at the end of the sort period and prices of stocks sold by HFTs actually higher than at the end of the sort period. But among 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 basis points, compared to 1.17 versus -0.34 basis points for mid-cap stocks. Thus, the full-sample finding that HFT net marketable buying predicts future non-HFT net marketable buying and returns is strongest among small and mid-cap stocks.

One may be interested in learning whether HFTs' aggressive trades, which analysis thus far suggests tend to forecast non-HFT buying and selling pressure, generally result from HFTs entering or exiting risky positions. The anticipatory trading story that likely first comes to mind is one in which an HFT enters a position and then soon afterwards sells that position to aggressive non-HFT buyers. This story implies that when HFTs aggressively purchase a stock, they should subsequently be net sellers of that same stock. Another possible story is one in which HFT marketable trades arise from market-making HFTs anticipating order flow to manage inventory risk. This story implies aggressive purchases by HFTs will tend to occur when HFTs have large short positions in a stock. The first story is one in which HFTs anticipate order flow to make directional bets, whereas the second story is more about anticipating order flow to manage risk.

There is justification for believing HFT marketable imbalances are largely attributable to directional bets rather than inventory management. The business of a market maker is to earn the bid-ask spread on a stock through buying at the bid and selling at the ask. This is done using non-marketable orders. Marketable trades to manage inventory risk, on the other hand, pay the bid-ask spread, reducing profits. It follows that market makers would want to minimize the share of their volume attributable to inventory management via marketable transactions (Harris 2002). By extension, if HFT marketable trades are mostly for inventory management, then they should account for a relatively small share of total HFT volume. But over half of HFT dollar volume on NASDAQ is due to marketable orders, suggesting directional bets account for much of the variation in HFT net marketable buying.[13]

The top panel of Figure 4 examines these issues by plotting cumulative HFT net buying for the first and tenth net marketable buying portfolios in Table 2 from 60 minutes before to 60 minutes after the sort period. This test is related to Brogaard, Hendershott, and Riordan's (2013) examination of whether aggregate HFT net buying is stationary in general, though here the focus is net buying around periods of intense HFT net marketable buying. Cumulated net buying provides an estimate of HFTs' position in a stock. Positions must be estimated from transactions, because, unlike research using data on NYSE specialists (e.g., Hasbrouck and Sofianos 1993, Madhavan and Smidt 1993, Hendershott and Seasholes 2008, Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes 2010), the data from NASDAQ do not contain position information. If HFTs' trades in the sort period are unwinding positions accumulated over the prior hour, then for the stocks HFTs are buying aggressively at time zero, one would expect HFTs to have previously been net sellers of those stocks. If this is going on, then in the figure, the solid line indicating cumulative net buying for stocks HFTs buy aggressively at time zero should be falling in the pre-sort period, and the dotted line indicating cumulative net buying for stocks HFTs sell at time zero should be rising. However, the lines in the hour before the sort period are mostly flat and close to zero, providing no discernible evidence that HFT trades in the sort period are disposing of previously acquired positions. Looking at the hour after the sort may give some indication

[13] See Figure A2 for a plot of marketable trades as a percent of HFT dollar volume.

of whether HFTs are entering a position in the sort period. Specifically, if trades in the sort period are due to HFTs entering positions, one would expect them to subsequently reverse those trades. This implies the lines marking cumulative HFT net buying should reverse towards zero in the post-sort period. However, there is no evidence of such a reversal in the following 60 minutes and, if anything, there is actually a slight continuation of net buying.

The failure of Figure 4 to provide evidence of HFT sales either before or after periods of intense net marketable buying is perhaps not very surprising given limitations of the data. For one thing, since this study only uses transactions occurring on NASDAQ, position estimates using cumulative net buying are bound to be imprecise. For example, if HFTs purchase a share of MSFT on NASDAQ and then sell the share on the NYSE, they appear to be long one share of MSFT based on NASDAQ trade data, but in reality have no position. In fact Menkveld (2013) provides evidence using European data of a low correspondence between HFT positions estimated from cumulated transactions on one exchange and those estimated using data across all exchanges. Given NASDAQ's average share of volume among the sample stocks is 27.2%, there may be many cases where only one leg of the trade occurs on NASDAQ. Moreover, if a purchase on NASDAQ is hedged with positions in other instruments, then the HFT may have no need to sell in a nearby time period.

Another way to evaluate whether HFT trades in the sort period offset previously built inventory positions is to look at returns. If sales on NASDAQ during the sort period are disposing of shares previously acquired on another venue while making a market in the stock, then one would expect the stock to have negative returns over that time span. One expects negative returns, because if market makers have built a long inventory position, then there have probably been many marketable sellers, which suggests prices would be pushed down. The advantage of looking at the question in this way is that there is no assumption that inventories calculated using only NASDAQ data accurately reflect market-wide inventory positions. It turns out returns in the bottom panel of Figure 4 from 60 minutes to 1 minute before aggressive HFT sales are positive, which is not what one would expect if HFTs were providing buy-side liquidity on other trading venues during that period. The returns for stocks HFTs buy aggressively at time zero show analogous patterns. 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.

3 Alternative explanations

Section 2 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 HFT trading leading non-HFT trading. These explanations include HFT and non-HFT trading being driven by the same serially correlated process, non-HFTs chasing past price trends, and HFTs reacting faster than non-HFTs to news stories. This section evaluates these alternatives.

3.1 Is the explanation correlated signals or trend chasing?

If there are common trading patterns among firms, then the HFT and non-HFT trading measures will be contemporaneously correlated. This might be the case if, for instance, firms in the HFT and non-HFT samples use the same trading signals. If there is also serial correlation in these trading patterns, then HFT trading will predict non-HFT trading simply because it is a noisy proxy for lagged non-HFT trading. If this explanation is driving the lead-lag relationship between HFT and non-HFT trading and the form is such that HFT trading is non-HFT trading plus noise, 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. Specifically, purchases by HFTs would cause prices to rise, which would trigger new purchases by trend-chasing non-HFTs. This explanation predicts that HFT trading will be uncorrelated with future non-HFT trading after controlling for lagged returns.

This section controls for these two 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 Table 4 heading displays these equations. 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 4 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. Panel B reports these results.

This section is motivated by concerns about the confounding effects of serial correlation in non-HFT trading 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. Average coefficients on lagged non-HFT net marketable buying in Panel A decline from 0.074 at lag one to 0.012 at lag ten.[14] All coefficients are more likely to be positive and significant than negative and significant, and the time-series means of average daily coefficients in Panel B 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.[15,16] In summary, the coefficients on lagged non-HFT trading and lagged returns in Table 4 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 2, HFT net marketable buying is positively correlated with future net marketable buying from other investors in Table 4. A one standard deviation increase in HFT net marketable buying on average leads to a 0.0023 standard deviation change in

[14] 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.

[15] Other interpretations include market makers anticipating a forthcoming net marketable imbalance and adjusting 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.

[16] One concern might be that the apparent trend-chasing behavior could be driven by misaligning trade and NBBO quote time-stamps. Appendix Table A4, which uses NQBBO quotes, shows using precisely aligned timestamps does not change any of these conclusions.

non-HFT net marketable buying the next period.[17] The lag one coefficient is positive and significant 24.92 percent of the time and negative and significant 16.59 percent of the time. The average coefficient on lags two through ten declines slowly, to a minimum of 0.0016 at lag ten.[18] 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 2. 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. These findings are related to Hasbrouck's (1991) use of a VAR to measure the price impact of trades. However, in the present setting, the intent is to capture quote updates correlated with HFT trading and caused by future non-HFT order flow.

To get a sense for the economic magnitude of the conditional lead-lag relationship between HFT and non-HFT trading, Figure 5 uses impulse response functions to plot the

[17] 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.

[18] Appendix Figure A4, which plots coefficients for the VAR using thirty lags, shows that the coefficients on HFT net marketable buying continue to decline towards zero at longer lags.

response of non-HFT net marketable buying to a one standard deviation shock to HFT net marketable buying. Impulse response functions are first calculated for all stocks separately each day.[19] 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 as a 95% confidence interval calculated using standard errors from the daily time series of mean impulse response functions. The figure indicates the average 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 4 on returns at lags one through three are positive, while those on lags four 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 and three are 0.034 and 0.001. Coefficients on lags four through ten are negative and 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 Section 2, 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

[19]The impulse response function is orthogonalized to allow for contemporaneous 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. These assumptions allow for non-HFT net marketable buying to affect future HFT net marketable buying, which would be the case if HFTs make markets.

the lead-lag relationship between HFT and non-HFT trading.

3.2 Are HFTs simply reacting faster to news?

Another potential 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. The ability to react to market events faster than other investors is undoubtedly an important HFT skill (Ye, Yao, and Gai 2012). In fact, news agencies provide machine-readible news feeds to enable exactly this type of trading and Riordan, Storkenmaier, and Wagener (2012) show measures of information asymmetry increase around these news releases. Foucault, Hombert, and Rosu (2012) study the theoretical implications of HFTs having a speed advantage when trading on news and, along with Hasbrouck (1991), discuss problems with using a VAR to measure price impact when trades are correlated with future news releases. While price impact is not the focus of the present study, HFTs trading on news slightly faster than other investors can still be a problem, because a VAR could show HFTs trading ahead of non-HFT order flow when the HFTs are simply reacting faster. However, the question is not whether HFTs ever react faster to news, but whether news-based trading is driving the full-sample results presented earlier. To evaluate 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 one percent. All three methods indicate the lead-lag relationship between HFT and non-HFT trading is not attributable to HFTs reacting faster to news announcements.

The first test reexamines the sort results in Table 2 after excluding stocks that have a Factiva news article about them published within five minutes of the sort period. Panel A in

Table 5 reports sorts of non-HFT net marketable buying after these periods near intra-day news are removed, and the results are nearly identical to those in Table 2. In the first five minutes after the sort period, cumulative non-HFT net marketable buying for the stocks HFTs buy versus sell most aggressively is 1.26 and −1.68 after excluding the five minutes around news announcements, compared to 1.22 and −1.76 for the full sample in Table 2. Returns outside news release periods, shown in the bottom panel of Table 5 are also very similar to results from the full sample. For the stocks HFTs buy most aggressively, returns in the five minutes after the sort period are 0.65 basis points in the restricted sample versus 0.62 basis points in the full sample. Similarly, for the stocks HFTs sell most aggressively, returns in the five minutes after the sort period are −0.41 basis points in the restricted sample versus −0.41 basis points in the full sample.[20] 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. The above test necessarily requires an article have a timestamp in order to filter out nearby trading periods. While wire services typically include timestamps with articles, they are less common in articles from news magazines and daily papers. One concern this restriction raises is that if there are intra-day periods when a news item is only published in articles without timestamps, then they will not be excluded from the above sorts. The test will also miss news-trading periods if timestamps in the articles are wrong or if Factiva does not include all types of news.

To address these concerns, Table 6 uses more aggressive criteria to filter out periods when HFTs may potentially be reacting faster than non-HFTs to news announcements. Rather than excluding just the five minutes before and after a news announcement, the table looks at non-news days, meaning days when there is no news at any point during the

[20]Figure A5 also confirms the plots for the restricted sample are very similar to those of the full sample.

day. The definition of news is also more general than in Table 5. Panel A again uses Factiva to identify news announcements, but since all articles include publication dates, there is no reliance on potentially missing or inaccurate timestamps. Panel B uses absolute market-adjusted stock returns of greater than one percent to identify days when there is news that may not show up in Factiva (e.g., analyst forecasts). In both panels, the table reports estimates for coefficients on lags of HFT net marketable buying in regressions where the dependent variable is non-HFT net marketable buying. These estimates are simply those from equation 3 in the VAR of Table 4 conditioned on whether they come from a day with or without news. 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 groups of columns contain estimates for non-news days. In Table 6, Panel A, 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. Similarly, in Panel B, 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.

4 Is predicting order flow related to HFT skill?

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 examines differences among HFTs in how strongly their trades are correlated with future non-HFT net marketable buying and returns.

4.1 Are some HFTs better at forecasting order flow?

This section tests whether some HFTs are better at forecasting order flow by examining 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 later months. The advantage in looking at persistence 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 3.1. 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 7 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 probability an HFT in the highest correlation group remains in that group in future months. Figure 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 the 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, 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. Similarly, in months two and three, more HFTs are still in the high group than would be the expectation under the null hypothesis of no persistence. 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 7 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 7 indicate there are persistent differences among HFTs in how non-HFT net marketable buying is correlated with both HFTs' lagged net marketable buying and lagged net 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 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 7 show results using ten lags of HFTs' net marketable buying, rather than just the first lag. 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 7 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—whether on looks at the coefficients on the first lag or on all ten lags, there are persistent differences between the highest and lowest-correlation groups. One may conclude from these results that trades from some HFTs are more highly correlated with future order flow than are trades from other HFTs.

4.2 Are the HFTs who forecast order flow best also better at predicting returns?

One potential explanation for some HFTs' trades being more strongly correlated with non-HFT order flow is that they are more skilled. If these HFTs are more skilled, then one would expect their trades to also be more strongly correlated with future returns. This explanation is related to work by Anand, Irvine, Puckett, and Venkataraman (2012), who provide evidence some institutional trading desks are more skilled than others. Baron, Brogaard, and Kirilenko (2012) also examine skill differences among HFTs, but they focus on differences in performance rather than the relationship between differences in order flow and return predictability.

Table 8 examines whether trades from the HFTs whose trades are the most strongly correlated with future non-HFT order flow also forecast larger returns. To do so, HFTs are split into two groups depending on whether their trades' correlation with future non-HFT order flow is above or below the median using the methodology discussed in Table 7. Then trades are aggregated among HFTs in each group, resulting in one time-series of aggregated trades from either of the two HFT groups. Returns are then alternately regressed on ten lags of each aggregate HFT series, controlling for ten lags of returns and ten lags of non-HFT net marketable buying. Thus, the regressions identify the two HFT groups' ability to forecast returns that is independent of information in past returns and non-HFT order flow. These regressions are estimated separately for each stock each day. A weighted cross-sectional average is calculated from the stock-level estimates each day, and then Table 8 reports the mean and median of the daily time-series of coefficients for the above and below median groups.

The results indicate trades from HFTs whose trades are more strongly correlated with future non-HFT order flow are also more strongly correlated with future returns. The coefficients on the first four lags of HFT net marketable buying in the above-median regressions are 0.026, 0.017, 0.013, and 0.009, compared to 0.018, 0.012, 0.008, and 0.007 in the below-median regressions. Both the means and medians of the two coefficient time-series are significantly different from each other. This indicates that trades from the HFTs in the above-median group have a stronger positive correlation with future returns over the next few seconds. At longer lags, there is generally no difference between the two series. These findings are consistent with the above-median HFTs, the HFTs whose trades are most strongly correlated with future non-HFT order flow, being more skilled at predicting returns.

5 When is non-HFT order flow more predictable?

Prior sections demonstrated that HFT net marketable buying leads both non-HFT net marketable buying and returns. If these findings are due to HFTs anticipating and trading ahead of non-HFT order flow, then perhaps the effects will be stronger when non-HFTs are relatively impatient. At such times, non-HFTs may not hide their order flow as well, making it easier for HFTs to anticipate their trades. 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 non-HFT trades.

The methodology involves comparing estimates of the VAR in section 3.1 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 is consistent with HFTs having an easier time anticipating order flow.

5.1 VAR estimates near the market open and close

The first set of tests uses the open and close of trading to proxy 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. They may have private information about a post-close news announcement or be facing a liquidity shock that needs to be funded before the close. Panel A in Table 9 reports results comparing trading in the first and last half hours of the trading day to trading in the middle of the day.

The first two columns contain estimates from 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 between the morning and mid-day estimate 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.

HFT net marketable buying at the close does not exhibit a stronger positive correlation with non-HFT net marketable buying than during the middle of the trading day. In contrast to results from the open, the coefficient on the first lag of HFT net marketable buying near the close, -0.0029, is actually negative. 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.

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 at the close. One explanation 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. One potential, though not fully satisfying, explanation for the negative coefficient near the close 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.

5.2 VAR estimates on high volume and high imbalance days

Days when a stock's volume or absolute net marketable buying imbalance are high could also be good proxies for times when non-HFTs are relatively impatient. High volume or imbalance days are likely days when certain investors are trading large positions. 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.

Panel B in Table 9 compares VAR estimates on high volume or high imbalance days to normal days. As in section 5.1, 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.[21]

The columns on the left in Panel B examine high volume days, and the columns on the right examine 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 longer lags there is no significant difference between the coefficients on high volume and normal volume days. Similarly, 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.

5.3 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

[21]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.

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).

Panel C in Table 9 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: the left column group uses bid-ask spreads, while the right column group uses relative bid-ask spreads, which are spreads divided by the bid-ask midpoint. Normal bid-ask spreads are easier to interpret, 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 an estimate of roughly zero 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, 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. For both spread measures, only coefficients 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 near term non-HFT order flow in illiquid stocks.

6 Conclusion

This study examines the relationship between high-frequency traders' aggressive trades and future order flow from other investors. I find that aggressive trades by HFTs lead those of other investors. Specifically, if HFTs buy a stock aggressively during a particular second, then this forecasts future aggressive buying by non-HFTs that continues up through five-minutes into the future. I explore several explanations for these findings, including their being driven by serial correlation in non-HFT order flow, non-HFTs chasing return trends, and HFTs reacting faster than non-HFTs to news. However, the findings are best explained by HFTs trading ahead of anticipated price changes caused by non-HFTs' future buying and selling pressure. Consistent with this anticipatory trading hypothesis, the effects are stronger at times when non-HFTs may be impatient, such as at the market open, on high volume or imbalance days, and in stocks with wide bid-ask spreads. I also find that HFTs vary in their skill at predicting non-HFT order flow, and that trades from the HFTs who are most skilled at predicting order flow also predict larger price changes than do trades from other HFTs. These findings provide evidence supporting the existence of an anticipatory trading channel through which HFTs may increase non-HFT trading costs.

One of the primary reasons for interest in research on HFTs is to understand the effect their trading has on other market participants. Likely benefits from the existence of HFTs acting as hyper-efficient market makers include lower bid-ask spreads and reduced return reversals (Castura, Litzenberger, Gorelick, and Dwivedi 2012). It is harder to get a handle on the potential costs HFTs impose on others. This study takes a step in that direction by providing evidence indicating HFTs anticipate other investors' order flow. But there is still much to learn about HFT information acquisition and its effect on other market participants. One topic that is not covered in this study, but that would be particularly interesting to non-HFTs, is how HFTs anticipate non-HFT order flow. Future research could examine whether the predictability arises, for example, from cross-correlations caused by delayed reaction to common information (Mech 1993, Hou and Moskowitz 2005) or from sophisticated analysis of the order book (Cao, Hansch, and Wang 2009, Brogaard, Hendershott, and Riordan 2013). Additionally, research that specifically focuses on HFT trading around news releases could determine whether HFTs are simply faster at reacting to news or if they are also perhaps better than other investors at interpreting the information content of a news release. Research into these questions would improve our understanding of the ways in which HFTs acquire information and in doing so, inform evaluations of the welfare costs in addition to the benefits of HFT participation in the price discovery process.

References

Anand, Amber, Paul Irvine, Andy Puckett, and Kumar Venkataraman, 2012, Performance of institutional trading desks: An analysis of persistence in trading costs, *Review of Financial Studies* 25, 557–598.

Baron, Matthew, Jonathan Brogaard, and Andrei Kirilenko, 2012, The trading profits of high frequency traders, *Working paper, University of Washington*.

Battalio, Robert H., 1997, Third market broker-dealers: Cost competitors or cream skimmers?, *Journal of Finance* 52, 341–352.

———, Jason Greene, and Robert Jennings, 1997, Do competing specialists and preferencing dealers affect market quality?, *Review of Financial Studies* 10, 969–993.

Battalio, Robert H., and Tim Loughran, 2008, Does payment for order flow to your broker help or hurt you?, *Journal of Business Ethics* 80, 37–44.

Battalio, Robert H., Andriy Shkilko, and Robert A. Van Ness, 2011, To pay or be paid? The impact of taker fees and order flow inducements on trading costs in US options markets, *Working paper, University of Notre Dame*.

Biais, Bruno, Thierry Foucault, and Sophie Moinas, 2013, Equilibrium fast trading, *Working paper, Toulouse School of Economics*.

Boehmer, Ekkehart, Kingsley Fong, and Julie Wu, 2012, International evidence on algorithmic trading, *Available at SSRN 2022034*.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2013, High-frequency trading and price discovery, *Working paper, University of California, Berkeley*.

Brown, William O., J. Harold Mulherin, and Marc Weidenmier, 2008, Competing with the New York Stock Exchange, *Quarterly Journal of Economics* 123, 1679–1719.

Brunnermeier, Markus, and Lasse Pedersen, 2005, Predatory trading, *Journal of Finance* 60, 1825–1863.

Cao, Charles, Oliver Hansch, and Xiaoxin Wang, 2009, The information content of an open limit-order book, *Journal of Futures Markets* 29, 16–41.

Castura, Jeff, Robert Litzenberger, Richard Gorelick, and Yogesh Dwivedi, 2012, High-frequency and algorithmic trading, *Annual Review of Financial Economics* 4.

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

Comerton-Forde, Carole, Terrence Hendershott, Charles M. Jones, Pamela C. Moulton, and Mark S. Seasholes, 2010, Time variation in liquidity: The role of market-maker inventories and revenues, *The Journal of Finance* 65, 295–331.

Comerton-Forde, Carole, and Talis Putnins, 2012, Dark trading and price discovery, *Available at SSRN*.

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

Foucault, Thierry, Johan Hombert, and Ioanid Rosu, 2012, News trading and speed, *Working paper, HEC Paris*.

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

Gervais, Simon, Ron Kaniel, and Dan H. Mingelgrin, 2001, The high-volume return premium, *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, John M., Jeffrey H. Harris, and Selim Topaloglu, 2003, The dynamics of institutional and individual trading, *Journal of Finance* 58, 2285–2320.

Griffin, John M., Nicholas H. Hirschey, and Patrick J. Kelly, 2011, How important is the financial media in global markets?, *Review of Financial Studies* 24, 3941–3992.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American Economic Review* pp. 393–408.

Harris, Larry, 2002, *Trading and exchanges: Market microstructure for practitioners* (Oxford University Press, USA).

Hasbrouck, Joel, 1991, Measuring the information content of stock trades, *The Journal of Finance* 46, 179–207.

——— , 2013, High frequency quoting: Short-term volatility in bids and offers, *Working paper, New York University*.

——— , and Gideon Saar, 2009, Technology and liquidity provision: The blurring of traditional definitions, *Journal of financial Markets* 12, 143–172.

——— , 2011, Low-latency trading, *Working paper, NYU*.

Hasbrouck, Joel, and George Sofianos, 1993, The trades of market makers: An empirical analysis of nyse specialists, *Journal of Finance* 48, 1565–1593.

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

Hendershott, Terrence, and Ryan Riordan, 2012, Algorithmic trading and the market for liquidity, *Journal of Financial and Quantitative Analysis* p. forthcoming.

Hendershott, Terrence, and Mark S. Seasholes, 2008, Market maker inventories and stock prices, *American Economic Review* 97, 210–214.

Holden, Craig W., and Avanidhar Subrahmanyam, 1992, Long-lived private information and imperfect competition, *Journal of Finance* 47, 247–270.

Hou, Kewei, and Tobias J. Moskowitz, 2005, Market frictions, price delay, and the cross-section of expected returns, *Review of Financial Studies* 18, 981–1020.

Jain, Prem, and Gun-Ho Joh, 1988, The dependence between hourly prices and trading volume, *Journal of Financial and Quantitative Analysis* 23, 269–283.

Jovanovic, Boyan, and Albert J. Menkveld, 2011, Middlemen in limit order markets, *Working paper, NYU*.

Kirilenko, Andrei, Albert Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2011, The Flash Crash: The impact of high frequency trading on an electronic market, *Working paper, University of Maryland*.

Madhavan, Ananth, and Seymour Smidt, 1993, An analysis of changes in specialist inventories and quotations, *The Journal of Finance* 48, 1595–1628.

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

Mech, Timothy S., 1993, Portfolio return autocorrelation, *Journal of Financial Economics* 34, 307–344.

Menkveld, Albert J., 2013, High frequency trading and the new-market makers, *Working paper, VU University Amsterdam*.

Newey, Whitney K., and Kenneth D. West, 1994, Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* 61, 631–653.

O'Hara, Maureen, Chen Yao, and Mao Ye, 2012, What's not there: The odd-lot bias in TAQ data, *Working paper, Cornell University*.

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality?, *Journal of Financial Economics* 100, 459–474.

Riordan, Ryan, Andreas Storkenmaier, and Martin Wagener, 2012, Public information arrival: Price discovery and liquidity in electronic limit order markets, *Journal of Banking and Finance* p. forthcoming.

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

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.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, *Journal of Finance* 62, 1139–1168.

——— , Maytal Saar-Tsechansky, and Sofus Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* 63, 1437–1467.

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

Ye, Mao, Chen Yao, and Jiading Gai, 2012, The externalities of high-frequency trading, *Working paper, University of Illinois at Urbana-Champaign*.

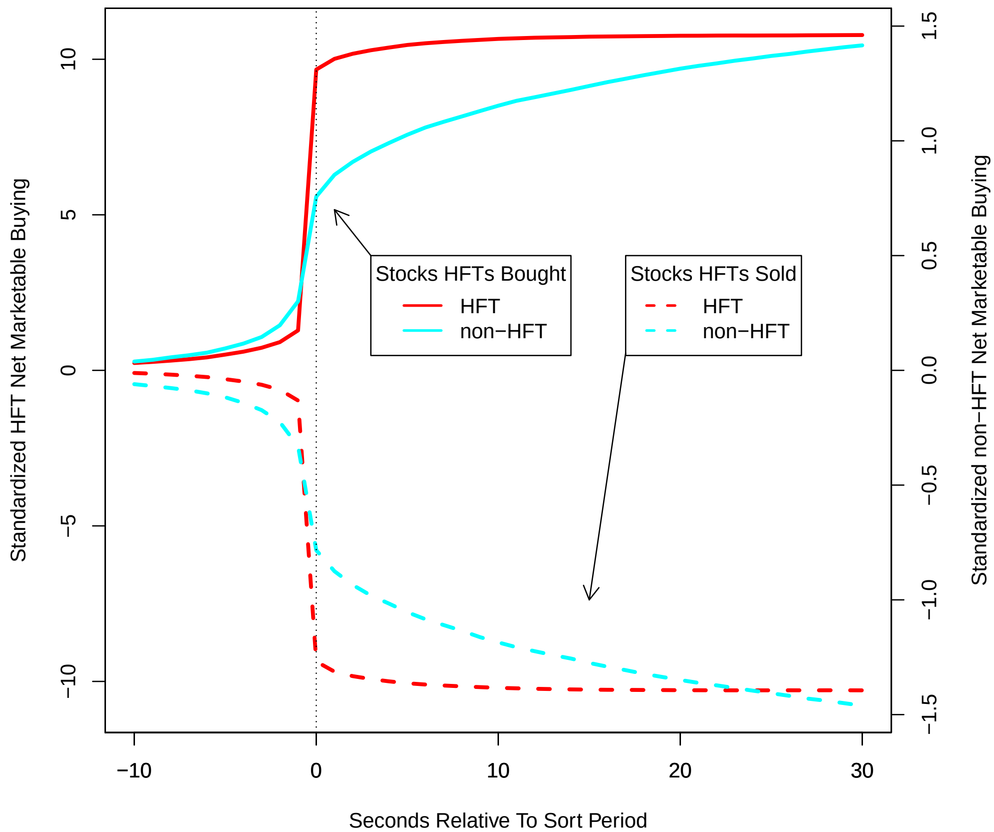

Figure 1: Cumulative HFT vs. non-HFT Net Marketable Buying.
This figure plots cumulative standardized net marketable buying for stocks sorted into portfolios by HFTs' net marketable buying. The left y-axis is for HFT net marketable buying in the same direction as their net buying, HFT_{NMBSD}, and the right y-axis is for non-HFT net marketable buying, $non\text{-}HFT_{NMB}$. Table 1 describes construction of these imbalance measures. 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.

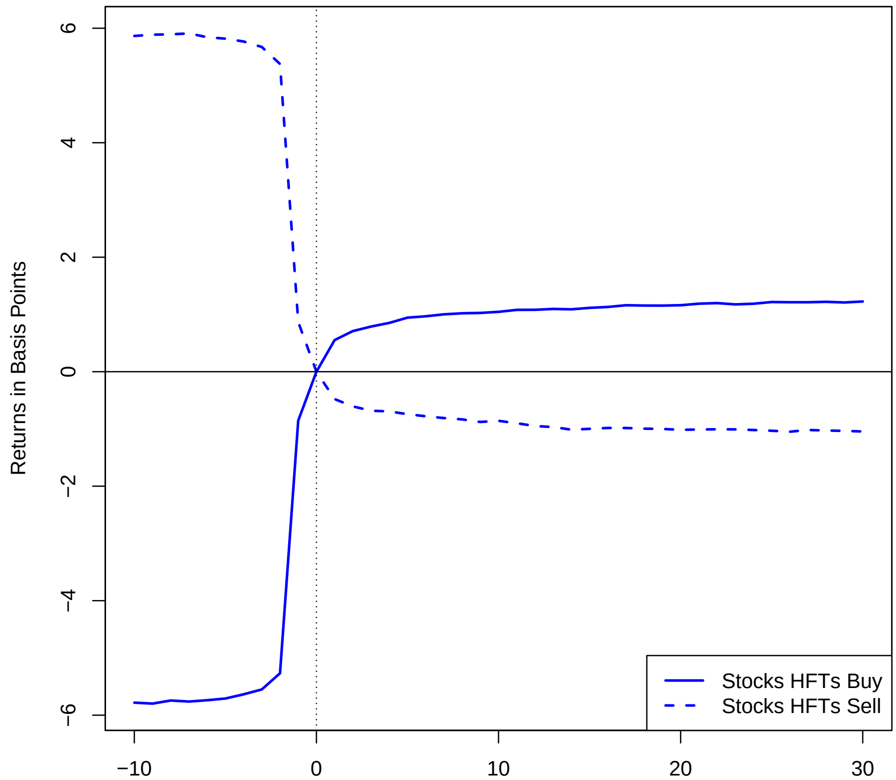

Figure 2: Returns
This figure plots returns for stocks sorted into portfolios by HFTs' net marketable buying. Stocks are sorted into deciles based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten and for which 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 describes construction of these imbalance measures.

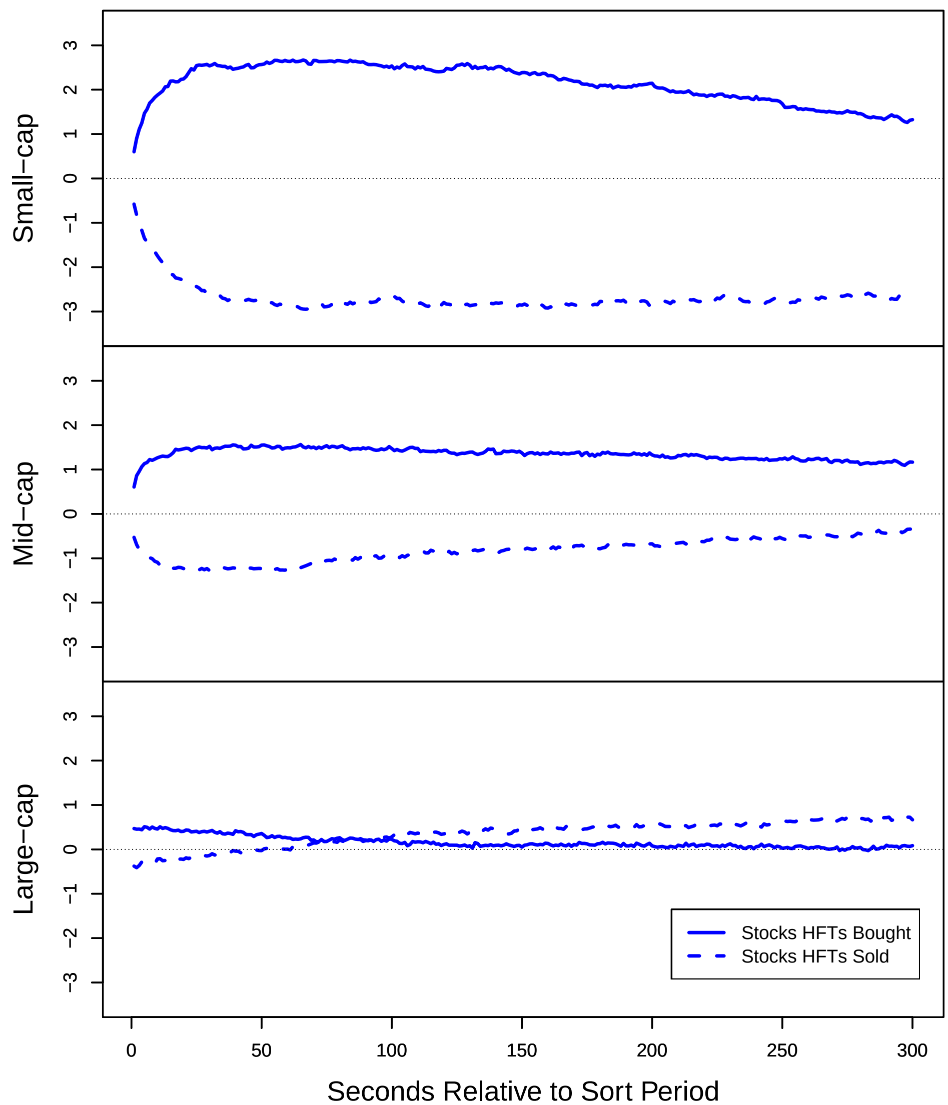

Figure 3: Post-sort Returns By Size Portfolios.
The y-axis scale is returns in basis points. Stocks are sorted into deciles based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten and for which HFT_{NMBSD} is 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 describes construction of these imbalance measures.

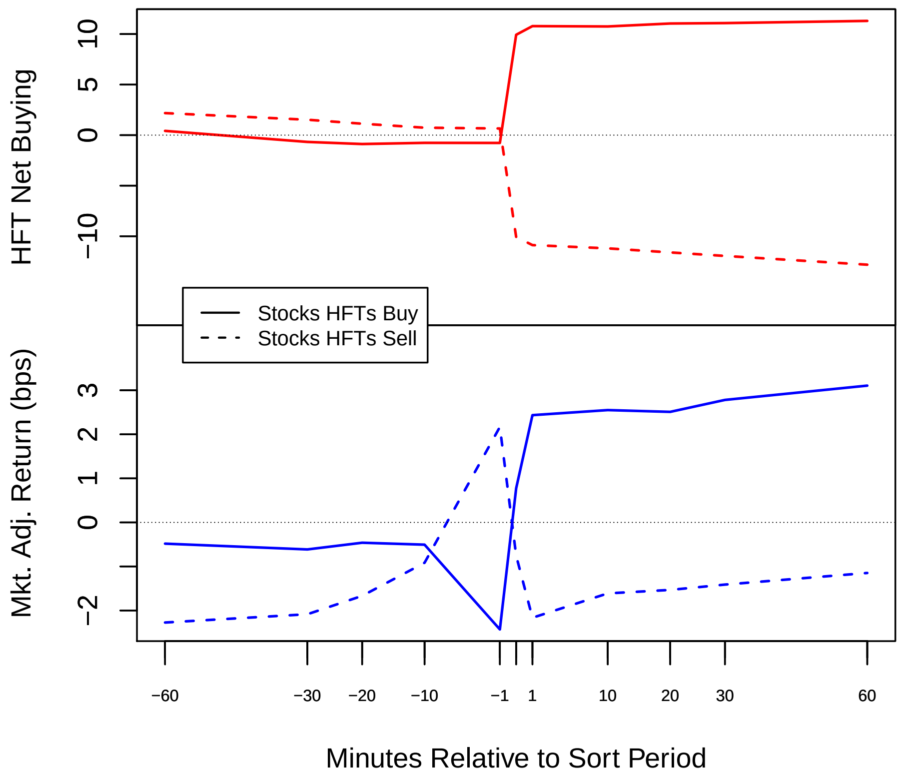

Figure 4: HFT Net Buying and Returns 60 minutes before and after intense HFT Net Marketable Buying

The figure examines HFT net buying from 60 minutes before to 60 minutes after periods of intense HFT net marketable buying or selling. The top panel plots cumulative standardized HFT net buying, while the bottom panel plots cumulative buy and hold returns. 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.

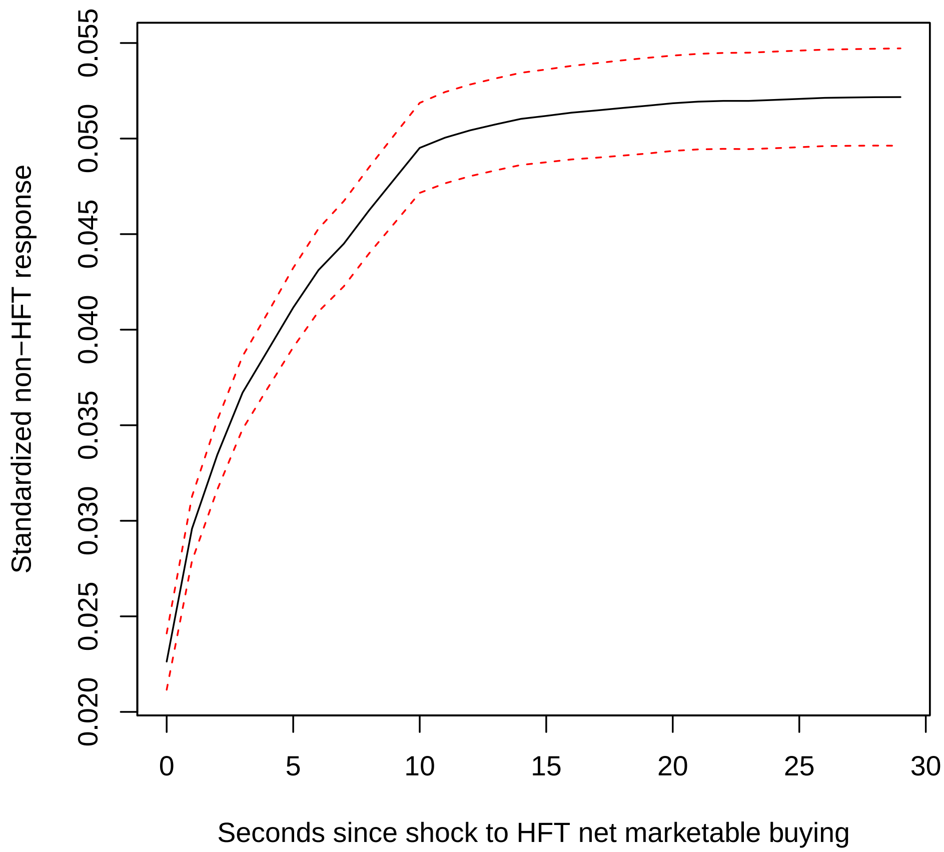

Figure 5: Response of non-HFT Net Marketable Buying to a One Standard Deviation Shock to HFT Net Marketable Buying

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 describes construction of these imbalance measures. The response is expressed in standard deviations. The results are based on the vector autoregression (VAR) in Table 4. 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.

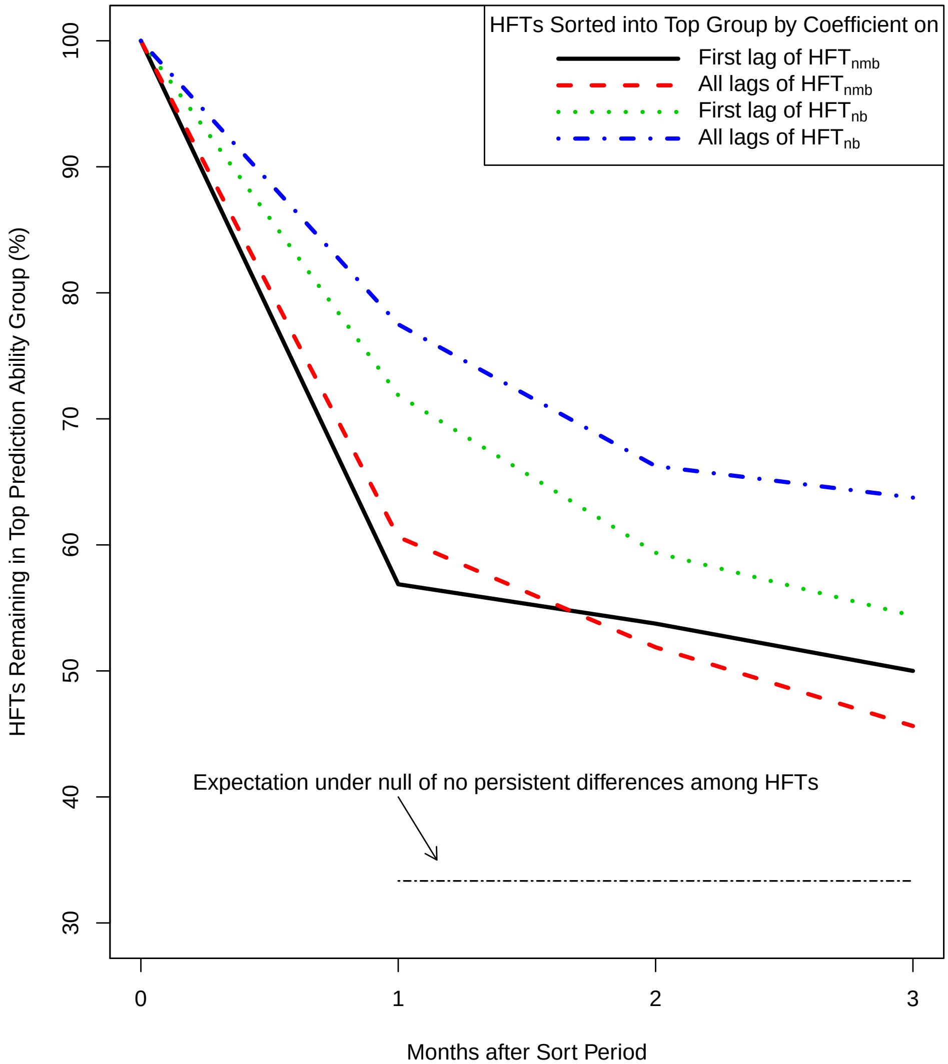

Figure 6: Persistence of Cross-sectional Differences in Prediction Ability
This figure plots the probability an HFT who is in the group of HFTs whose trades are most strongly correlated with future non-HFT order flow in month 0 will be in that group in later months. The percents are calculated for each 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 7. 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
Summary Statistics

Panel A summarises stock characteristics calculated from the pooled time-series of all stock-day observations. Market capitalization, price, and dollar volume are end of day values. 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 in the sample. Panel B summarises intra-day returns and net buying measures by reporting the mean and median daily standard deviation of these variables across all stock days. Each day, for every stock, the following are calculated: 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-*HFT_{NMB}). 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. 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. In later tables imbalances are divided by 20-day trailing average daily volume to make net buying measures comparable across stocks. 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.

Panel A: Daily Stock Characteristics

		Mkt Cap Mil. $	Price $	Volume Mil. $	NQ %	HFT %	N
All Stocks	mean	5,302	26.38	58.1	27.2	27.6	93.2
	median	1,301	22.05	12.6	25.0	27.7	93.0
	std dev	12,909	19.96	111.4	13.5	13.7	1.6
	min	22	0.91	0.1	0.7	0.0	89.0
	max	125,331	166.82	2,153.1	80.7	78.4	96.0
Small-cap	mean	367	16.52	4.0	26.3	16.7	33.4
	median	293	14.77	1.9	22.5	14.8	33.0
Mid-cap	mean	1,900	26.10	34.2	27.1	28.9	35.8
	median	1,565	25.04	15.5	26.0	29.2	36.0
Large-cap	mean	17,252	40.55	169.3	28.5	40.7	24.0
	median	9,413	31.37	120.2	26.9	40.9	24.0

Panel B: Standard Deviation of Intra-day Variables

		Ret %	HFT_{NB} shares	HFT_{NMB} shares	HFT_{NMBSD} shares	*non-*HFT_{NMB} shares
All Stocks	mean	0.080	83	80	76	125
	median	0.027	28	26	26	42
Small-cap	mean	0.052	13	11	11	27
	median	0.031	9	7	7	18
Mid-cap	mean	0.085	89	81	78	138
	median	0.022	35	34	33	49
Large-cap	mean	0.114	175	178	169	247
	median	0.023	116	123	118	138

Table 2
Non-HFT Net Marketable Buying for Stocks Sorted by HFT Net Marketable Buying

This table shows non-HFT net marketable buying for stocks sorted on HFTs' net marketable buying at the one-second horizon. Stocks are sorted into deciles at time t based on HFT net-marketable buying. Decile break-points 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. Stocks in deciles nine and ten must have HFT_{NMBSD} greater than zero, while stocks in deciles one and two must have HFT_{NMBSD} less than zero. 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.

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 3
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. 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. Stocks in deciles nine and ten must have HFT_{NMBSD} greater than zero, while stocks in deciles one and two must have HFT_{NMBSD} less than zero. 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.

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 4
Intra-day VAR Estimates for Individual Stock-day Observations

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

$$R_t = \alpha_1 + \sum_{i=1}^{10} \gamma_{1,i} HFT_{NMBSD,t-i} + \sum_{i=1}^{10} \beta_{1,i} non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_{1,i} R_{t-i} + \epsilon_{1,t} \tag{1}$$

$$HFT_{NMBSD,t} = \alpha_2 + \sum_{i=1}^{10} \gamma_{2,i} HFT_{NMBSD,t-i} + \sum_{i=1}^{10} \beta_{2,i} non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_{2,i} R_{t-i} + \epsilon_{2,t} \tag{2}$$

$$non\text{-}HFT_{NMB,t} = \alpha_3 + \sum_{i=1}^{10} \gamma_{3,i} HFT_{NMBSD,t-i} + \sum_{i=1}^{10} \beta_{3,i} non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_{3,i} R_{t-i} + \epsilon_{3,t} \tag{3}$$

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 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 4 — 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 4 — 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 5
Sorts Excluding Periods Within 5 Minutes of Intra-day News

This table shows sorts of non-HFT net marketable buying and returns in periods when there is no news for a stock. Specifically, the sorts exclude stocks that have a news article about them published within five minutes of the sort period. 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. Stocks in deciles nine and ten must have HFT_{NMBSD} greater than zero, while stocks in deciles one and two must have HFT_{NMBSD} less than zero. Non-HFT net marketable buying in Panel A and returns in Panel B 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: Non-HFT net marketable buying

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)

Panel B: Returns

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 6
VAR Estimates on Days with and without News

This table reports coefficients on HFTs' net marketable buying from the VAR in Table 4 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 Factiva 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

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 7
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:

$$non\text{-}HFT_{d,s,t} = \alpha_{d,i} + \sum_{l=1}^{10} \gamma_{d,i,l} HFT_{d,s,i,t-l} + \sum_{l=1}^{10} \beta_{d,i,l} non\text{-}HFT_{d,s,t-l} + \sum_{i=1}^{10} \lambda_{d,i,l} R_{d,s,t-l} + \epsilon_{d,i,s,t}, \quad (4)$$

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 (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 (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 the small number of months imply 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

Table 8
Differences among HFTs in How Strongly Their Trades Forecast Returns

This table examines whether trades from the HFTs whose trades are most strongly correlated with future non-HFT order flow also predict larger future returns. HFTs are split into two groups each month: HFTs who are above the median in terms of the correlation between their trades and future non-HFT order flow, and HFTs who are below the median. The split is based on the average $\gamma_{d,i,1}$ coefficient from regression 4 in Table 7 estimated during the prior month. Table 7 provides more detail on the methodology for sorting HFTs. Trades are then aggregated among above and below-median HFTs in the post-sort month. The table below compares coefficients on the above and below-median HFT groups' aggregate net marketable buying in regressions of the following form:

$$R_t = \alpha + \sum_{i=1}^{10} \gamma_i HFT^G_{NMBSD,t-i} + \sum_{i=1}^{10} \beta_i non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_i R_{t-i} + \epsilon_{t,R} \tag{5}$$

where R_t is the one-second return, $HFT^G_{NMBSD,t}$ is one-second HFT net marketable buying in the same direction as net buying for either the above or below-median group, and $non\text{-}HFT_{NMB,t}$ is one-second non-HFT net marketable buying. Returns and non-HFT net marketable are divided by their respective one-second standard deviations, whereas the above and below-median HFT groups' net marketable buying measure is divided by the one-second standard deviation of HFT_{NMBSD} aggregated among all (i.e., not only above or below-median) HFTs. Table 1 describes construction of these imbalance measures. The values reported are the time-series mean and median of daily cross-sectional mean coefficients. 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.

lag	**Above-median HFTs' γ_i**			**Below-median HFTs' γ_i**			**Difference**		
	mean	t-stat	median	mean	t-stat	median	mean	t-stat	rank sum p-value
1	0.0260	9.28	0.0219	0.0181	15.18	0.0154	0.0079	2.59	0.00
2	0.0170	8.67	0.0135	0.0119	14.19	0.0097	0.0052	2.42	0.00
3	0.0126	9.47	0.0102	0.0078	15.00	0.0069	0.0048	3.36	0.00
4	0.0089	11.16	0.0071	0.0065	11.98	0.0053	0.0024	2.44	0.01
5	0.0071	6.51	0.0058	0.0057	13.14	0.0046	0.0015	1.26	0.27
6	0.0052	8.86	0.0051	0.0044	13.11	0.0038	0.0009	1.26	0.05
7	0.0040	7.46	0.0045	0.0038	6.05	0.0031	0.0002	0.25	0.09
8	0.0023	4.20	0.0023	0.0022	7.64	0.0022	0.0001	0.09	0.79
9	0.0021	2.47	0.0015	0.0015	3.45	0.0015	0.0006	0.68	0.65
10	-0.0005	-0.92	-0.0003	-0.0001	-0.19	0.0000	-0.0005	-0.72	0.55
Σ 1-10	0.0848	10.44	0.0710	0.0618	14.77	0.0472	0.0230	2.52	0.00

Table 9
Conditioning on Times when non-HFTs are Hypothesized to be Impatient

This table reports coefficients on HFTs' net marketable buying from the VAR in Table 4 conditional on times when non-HFTs are hypothesized to be relatively more impatient. The estimates are from the equation where the dependent variable is non-HFT net marketable buying,

$$HFT_{NMBSD,t} = \alpha_2 + \sum_{i=1}^{10} \gamma_{2,i} HFT_{NMBSD,t-i} + \sum_{i=1}^{10} \beta_{2,i} non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_{2,i} R_{t-i} + \epsilon_{2,t},$$

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 describes construction of these imbalance measures. The regression is estimated separately for each stock each day, and then among a set of stocks on a given day, the average coefficient is calculated for each lag of HFTs' net marketable buying. Panels report the time-series mean of daily cross-sectional means. For t-tests, the null hypothesis is that the mean of the daily time series equals zero. 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.) and the close (3:30 p.m. to 4:00 p.m.) to the middle of the day (10:30 a.m. to 3:30 p.m.). In Panel B, 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. In Panel C, daily spreads are calculated by duration-weighting intra-day spread observations, and relative spreads are calculated by dividing the spread by the bid-ask midpoint. High spread or relative spread stocks are those in the top third of the sample based on the prior day's spread or relative spread, and low spread or relative spread stocks are those in the bottom third.

Panel A: Comparing γ estimates from the open and close to the middle of the day

							Difference with Mid-day			
	9:30–10:00am		10:00am–3:30pm		3:30–4:00pm		9:30–10:00am		3:30–4:00pm	
lag	mean	t-stat	mean	t-stat	mean	t-stat	mean	t-stat	mean	t-stat
1	0.0040	9.52	0.0012	6.28	−0.0029	−6.88	0.0028	6.05	−0.0041	−8.87
2	0.0036	9.71	0.0021	16.84	0.0009	2.88	0.0015	3.80	−0.0012	−3.43
3	0.0025	8.35	0.0019	18.59	0.0012	3.37	0.0006	1.79	−0.0007	−2.08
4	0.0019	6.69	0.0017	17.24	0.0015	5.36	0.0002	0.73	−0.0002	−0.57
5	0.0017	6.92	0.0015	17.18	0.0011	4.15	0.0002	0.94	−0.0003	−1.13
6	0.0021	7.91	0.0013	16.50	0.0016	5.30	0.0008	2.97	0.0004	1.16
7	0.0018	7.37	0.0013	14.73	0.0015	5.22	0.0005	2.03	0.0003	0.91
8	0.0013	5.79	0.0012	14.51	0.0018	5.97	0.0002	0.65	0.0007	2.08
9	0.0016	5.84	0.0012	15.15	0.0022	5.72	0.0004	1.52	0.0010	2.54
10	0.0019	7.88	0.0015	16.59	0.0013	5.01	0.0004	1.64	−0.0001	−0.39
Σ 1-10	0.0224	21.14	0.0147	25.99	0.0103	8.00	0.0077	6.39	−0.0044	−3.13

Table 9 — continued

Panel B: Comparing γ estimates on high volume/imbalance days to normal days

	Volume				Imbalance			
	High	Normal	Difference		High	Normal	Difference	
lag	mean	mean	mean	t-stat	mean	mean	mean	t-stat
1	0.0049	0.0016	0.0033	4.27	0.0036	0.0018	0.0018	2.48
2	0.0045	0.0021	0.0023	4.63	0.0036	0.0023	0.0013	2.57
3	0.0031	0.0020	0.0010	2.45	0.0026	0.0021	0.0005	0.99
4	0.0022	0.0018	0.0004	1.00	0.0022	0.0018	0.0003	0.91
5	0.0022	0.0016	0.0005	1.22	0.0014	0.0017	−0.0003	−0.51
6	0.0020	0.0017	0.0003	0.87	0.0027	0.0016	0.0010	1.78
7	0.0015	0.0015	0.0000	0.01	0.0008	0.0015	−0.0008	−0.61
8	0.0019	0.0016	0.0003	0.69	0.0019	0.0016	0.0003	0.61
9	0.0020	0.0015	0.0005	1.23	0.0018	0.0015	0.0002	0.47
10	0.0019	0.0014	0.0005	1.28	0.0017	0.0015	0.0002	0.49
Σ 1-10	0.0261	0.0169	0.0092	4.94	0.0222	0.0175	0.0047	1.89

Panel C: Comparing γ estimates from high spread stocks to low spread stocks

	Spread				Relative Spread			
	High	Low	Difference		High	Low	Difference	
lag	mean	mean	mean	t-stat	mean	mean	mean	t-stat
1	0.0053	0.0000	0.0053	11.18	0.0046	0.0012	0.0034	5.22
2	0.0032	0.0020	0.0012	3.57	0.0036	0.0023	0.0014	2.93
3	0.0025	0.0020	0.0005	1.49	0.0032	0.0020	0.0011	2.76
4	0.0018	0.0018	0.0000	0.07	0.0014	0.0020	−0.0006	−2.18
5	0.0018	0.0016	0.0002	0.70	0.0019	0.0016	0.0002	0.73
6	0.0015	0.0018	−0.0002	−1.14	0.0014	0.0019	−0.0005	−2.00
7	0.0016	0.0017	−0.0001	−0.53	0.0010	0.0016	−0.0006	−1.98
8	0.0013	0.0019	−0.0005	−2.51	0.0013	0.0018	−0.0005	−1.65
9	0.0013	0.0018	−0.0005	−2.17	0.0012	0.0018	−0.0006	−2.01
10	0.0015	0.0016	−0.0001	−0.61	0.0018	0.0016	0.0002	0.56
Σ 1-10	0.0218	0.0162	0.0056	4.95	0.0214	0.0179	0.0035	2.52

A Internet Appendix

Table A1
Summary of CRSP Universe

This table summarises 2009 stock-day observations for CRSP common stocks with dual-class stocks removed. The table summarises market capitalization, *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 A2
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 capitalization, *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 A3
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 A4
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 4 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 1st to March 4th 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 A4 — 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 A4 — 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

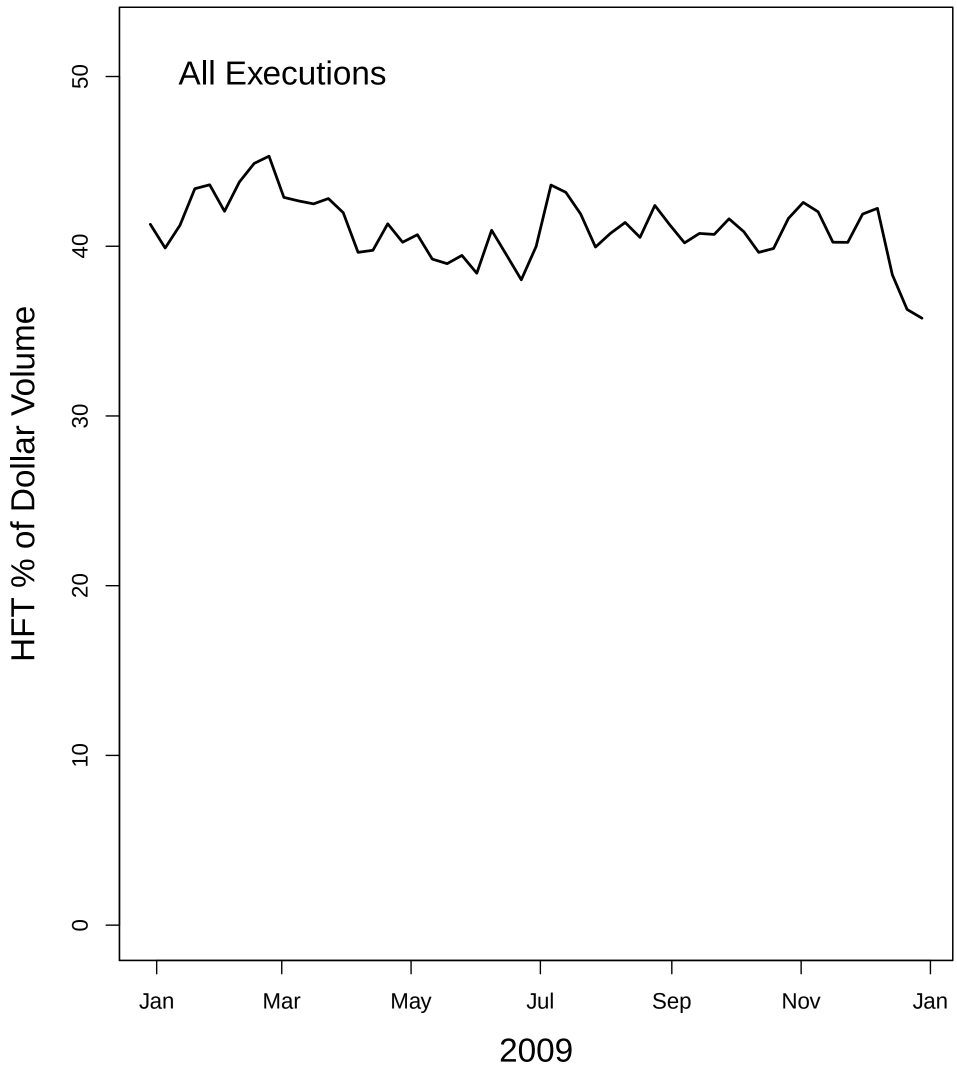

Figure A1: HFTs' Share of NASDAQ Dollar Volume
This figure shows HFTs' share of dollar volume on the NASDAQ Stock Market. The calculation includes all stocks with CRSP share code 10 or 11 trading on NASDAQ, regardless of listing venue.

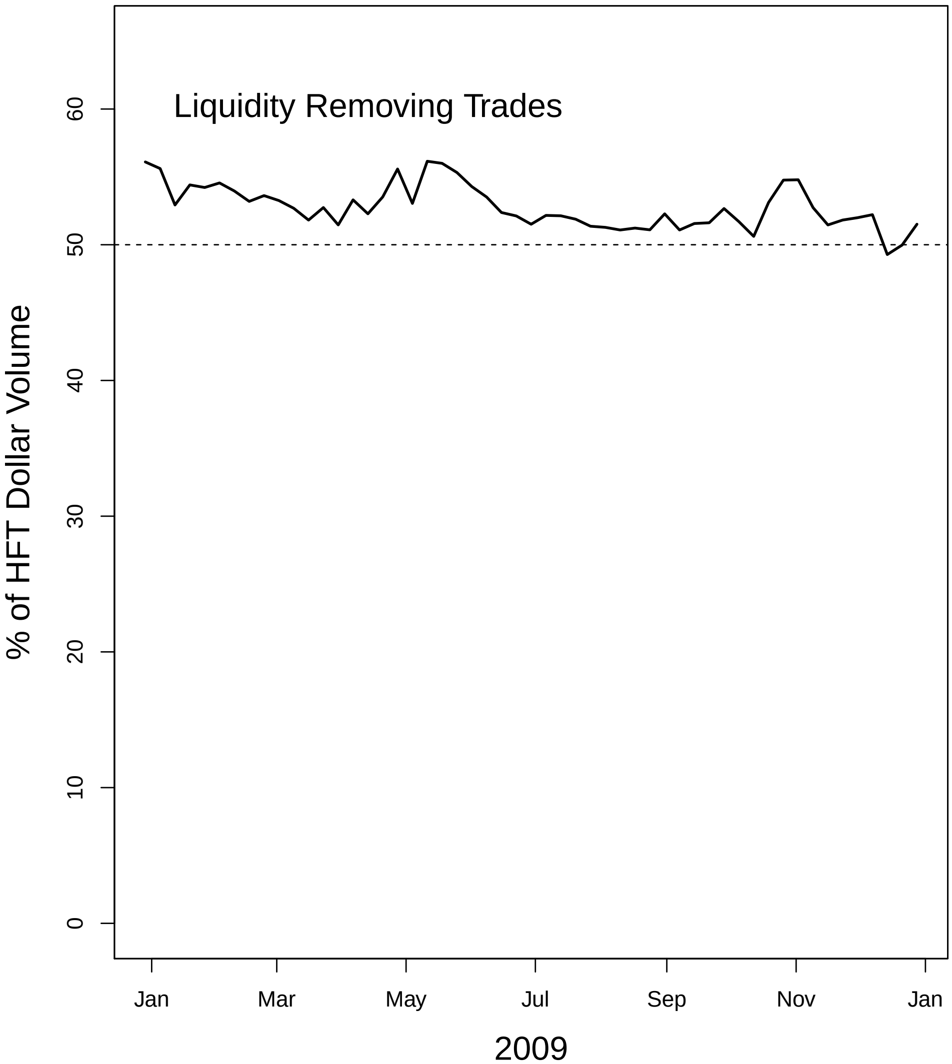

Figure A2: Liquidity Removing Trades as a Percent of HFT Dollar Volume
This figure shows liquidity removing trades as a percent of HFTs' dollar volume on the NASDAQ Stock Market. Liquidity removing trades are those in which the HFT initiates the trade with a marketable order, which is functionally equivalent to a market order. The calculation includes all stocks with CRSP share code 10 or 11 trading on NASDAQ, regardless of listing venue.

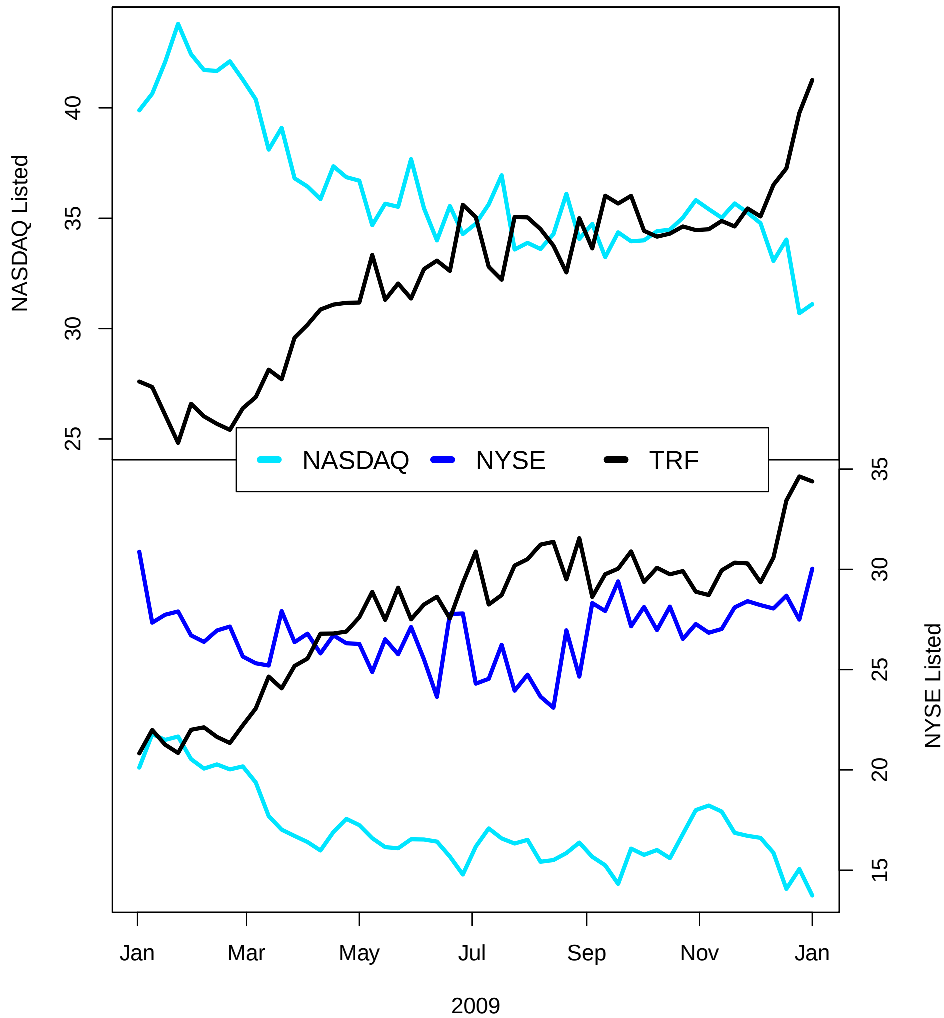

Figure A3: Market Share By Trading Venue
Market share is reported as percent of dollar volume. NASDAQ is the NASDAQ Stock Market, NYSE is the New York Stock Market, and TRF is the FINRA Trade Reporting Facility that includes trades that do not occur on a stock exchange (e.g., trades executed in dark pools or by off-exchange market making firms).

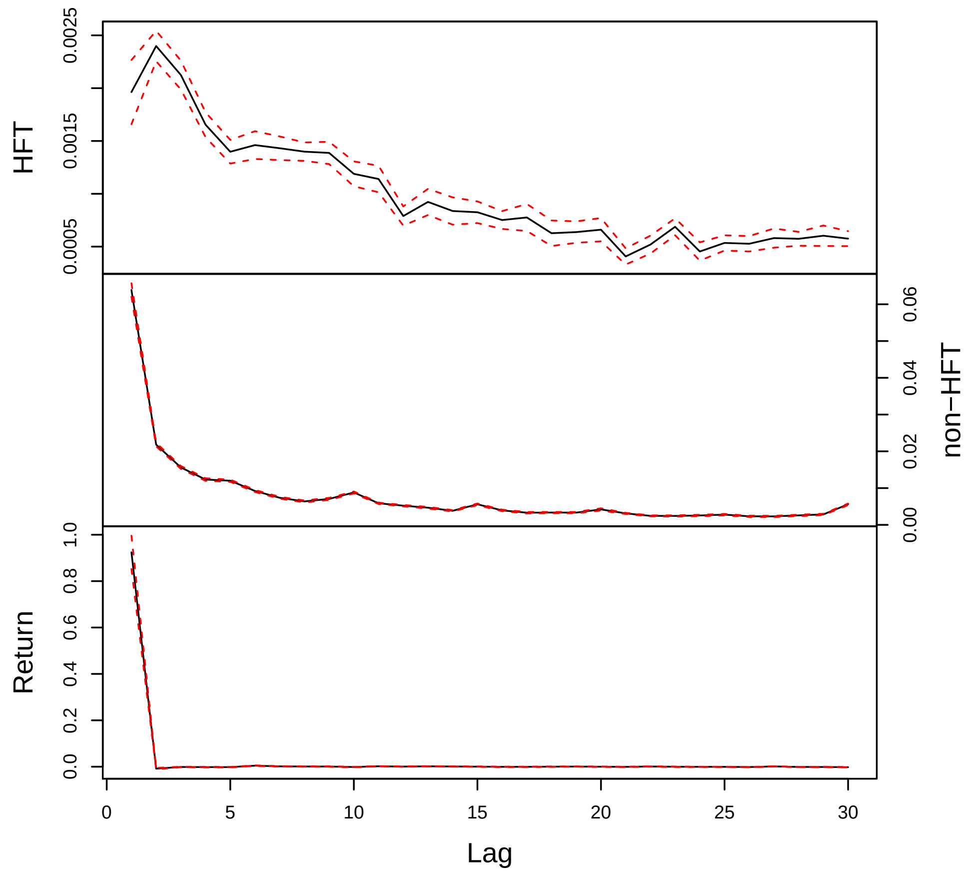

Figure A4: Coefficients from VAR with 30 lags

This figure plots the mean coefficients (solid line) and 95% confidence intervals (dotted lines) from a VAR system similar to the one in Table 4, with the difference being that it uses thirty lags rather than the ten used in Table 4. 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).

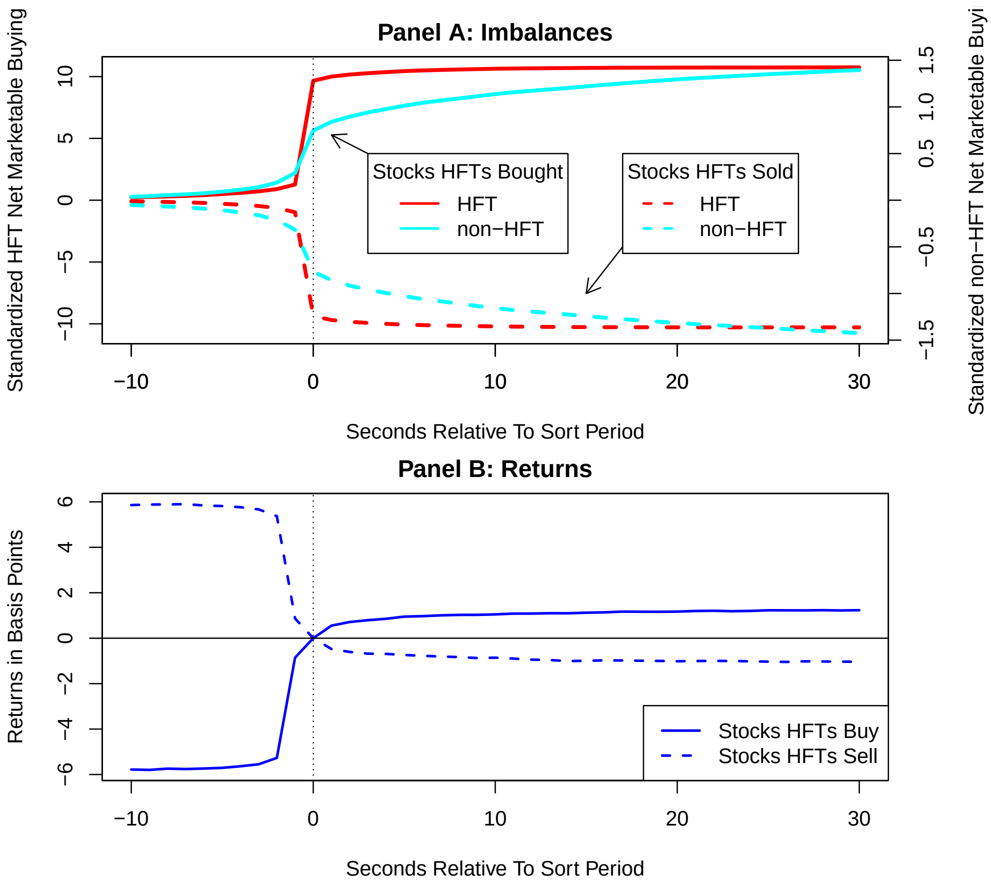

Figure A5: Sorts Excluding Periods +/- 5 Minutes from Intra-day News
This figure plots imbalances and returns for stocks sorted into portfolios by HFTs' net marketable buying after excluding stocks that have a news article about them published within five minutes of the sort period. In panel A, the left 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, $non\text{-}HFT_{NMB}$. 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 describes construction of these imbalance measures.