Do High-Frequency Traders Anticipate Buying and Selling Pressure?

Nicholas Hirschey *
London Business School

May 2018

Abstract: This study provides evidence that high-frequency traders (HFTs) identify patterns in past trades and orders that allow them to anticipate and trade ahead of other investors' order flow. Specifically, HFTs' aggressive purchases and sales lead those of other investors, and this effect is stronger at times when it is more difficult for non-HFTs to disguise their order flow. There is also persistence in which HFTs' trading predicts non-HFT order flow the best, indicating a subset of HFTs are either more skilled or more focused on anticipatory strategies. The results are not explained by HFTs reacting faster to news or past returns, by contrarian or trend-chasing behavior by non-HFTs, or by trader misclassification. These findings support the existence of an anticipatory trading channel through which HFTs 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 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, Johannes Breckenfelder, Carlos Carvalho, Jonathan Cohn, Shane Corwin, Andres Donangelo, Cesare Fracassi, John Griffin, Terry Hendershott, Paul Irvine, Pete Kyle, Richard Lowery, Tim Loughran, Jeffrey Pontiff, Mark Seasholes, Clemens Sialm, Laura Starks, Sheridan Titman, and seminar participants at Boston College, the European Finance Association meeting, the Georgia Institute of Technology, the Hong Kong University of Science and Technology, the London Business School, the London School of Economics, Rice University, the University at Buffalo, the University of California Irvine, the University of Georgia, Manchester Business School, 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.

Despite immense interest from investors and regulators, little is known about the automated strategies used by high-frequency traders (HFTs) and how they affect other investors. Some trading by HFTs has clear benefits. HFTs are market makers on many exchanges, and when providing liquidity at the bid or ask, they do so at narrower spreads than the human floor traders they displaced. Yet much of HFTs' trading is informed trading rather than liquidity provision; more than half of their dollar volume on NASDAQ comes from marketable trades that must predict price changes to be profitable.[1] This paper studies one potential information-based HFT strategy with negative welfare implications.

This is the first paper to provide evidence that HFTs identify patterns in past trades and orders that allow them to anticipate and trade ahead of non-HFT order flow. This strategy can be profitable for an HFT who is more skilled than some liquidity providers at learning from trade and order data. The HFT could identify a pattern indicating a large mutual fund will be selling a stock's shares in the near future. This pattern could be any number of things, for example a series of marketable sell orders. Institutions use execution algorithms to remove simple forms of predictability, but it is nonetheless possible some patterns remain. If liquidity providers have not updated their bid prices to account for the impending selling, then the HFT can sell shares and profit when the price subsequently falls. Consequently, anticipatory trading implies marketable HFT trades positively forecast returns and, importantly, non-HFT order flow.

This strategy makes the non-HFT whose trades are anticipated worse off. If the non-HFT is selling to fund a liquidity shock, the HFT's selling depresses the non-HFT's liquidation price (Brunnermeier and Pedersen 2005).[2] If instead the non-HFT is selling because they observed a signal indicating the stock is overvalued, the HFT is effectively reverse engineering the non-HFT's signal and capturing some of their information rent (Madrigal 1996, Yang and Zhu 2017, Baldauf and Mollner 2015)[3]. These exact concerns formed the basis for Michael Lewis's (2014) book, *Flash Boys*, and led to the founding of the deliberately slowed down IEX trading platform.

To examine these issues, I analyze return and trade patterns around periods of marketable buying and selling by HFTs using an entire year of unique trade and trader-level data from the

[1] See Figure IA.2.

[2] Brunnermeier and Pedersen (2005) do not model the source of information about liquidations, but the signal could come from observing past order flow.

[3] The non-HFT's profit is lower, because from that point forward the non-HFT is competing with the HFT to sell the stock before the price falls to its fundamental value.

NASDAQ Stock Market.[4] The imbalance between HFTs' marketable purchases and sales is a simple measure of their directional bets. The goal is to see if some of the information motivating these trades is related to non-HFT order flow. Though the data is from one exchange, it is representative of exchange trading—NASDAQ was the biggest U.S. exchange by volume during the sample period, and the sample includes an equal number of randomly selected NYSE and NASDAQ-listed firms with market capitalizations ranging from \$105 million to \$89 billion at the time of sample construction.[5] It is also an important sample for analyzing HFTs, given they accounted for 40% of total NASDAQ dollar volume.[6]

The evidence is 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 65% of its one-second standard deviation over the following thirty seconds. For the median stock, this equates to a 100 share HFT imbalance predicting non-HFTs to buy roughly 13 more shares with marketable orders than they sell over the next thirty seconds, growing to 22 shares after 5 minutes. Impulse response functions show similar magnitudes. Though the sort is on HFT net marketable buying, HFT net buying is also positive in the surrounding seconds. Thus at these times HFTs in aggregate take rather than provide liquidity. 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. These returns only slightly reverse, consistent with HFTs anticipating informed trades as opposed to liquidity trades.

I also examine whether some HFTs are more skilled at anticipating order flow and whether HFTs are better able to forecast order flow when non-HFTs are impatient. Consistent with HFTs having heterogenous skills, there is persistence in which individual HFTs' trades most strongly predict non-HFT order flow. Interestingly, estimates from the top third of HFTs are 6 times larger than for the bottom third. These top HFTs' trades are also more strongly correlated with future

[4] A marketable limit order is functionally equivalent to a market order: a buy order with a limit price at or above the best ask or a sell order with a limit price at or below the best bid at the time it enters the order book. NASDAQ requires all orders to have a limit price, and a trade is initiated when a marketable limit order crosses either the best bid or best ask.

[5] See Section 2 for more details on sample construction.

[6] See Figure IA.1 for HFTs' share of NASDAQ dollar volume aggregated among all listed securities.

returns. This skill persistence result holds regardless whether individual HFTs are sorted based on correlations with their net marketable buying or total net buying (summing marketable and non-marketable trades). Furthermore, the correlation between HFT and future non-HFT trades is stronger at times non-HFTs are impatient and thus less focused on disguising order flow: at the market open, on days with high volume, and when trading illiquid or small-cap stocks. This is consistent with Goldman Sachs' acknowledgement their algorithms leak more information in small-caps (Traders Magazine 2013).

I consider several explanations for the strong evidence that marketable HFT trades lead marketable non-HFT trades. The effect could be caused by HFTs reacting faster to a signal non-HFTs also observe, but the most probable signals they would both utilize in close succession (e.g., past returns and news[7]) do not fully explain the results. Controls for returns also show the effect is not due to a reverse causality story in which HFT purchases cause price increases that then cause trend chasing non-HFTs to purchase shares as well. I also demonstrate the results are not caused by a particular month or quarter during the year. Another explanation is HFT trades are actually forecasting trades from other HFTs following similar strategies, but whose trades are incorrectly labeled by NASDAQ as coming from non-HFTs. I evaluate this explanation by comparing the way HFT trading forecasts itself to how it forecasts non-HFT trading. Once you look more than seven seconds into the future, HFT marketable trades forecast non-HFT marketable trades more strongly than HFT marketable trades. By 15 seconds later, HFT marketable trading no longer forecasts itself, whereas it continues to forecast non-HFT marketable trading for more than 30 seconds. As misclassification implies the non-HFT predictability comes from it being a noisy proxy for HFT trading, the explanation is inconsistent with this finding of non-HFT trades being more predictable than HFT trades at longer lags.

Taken together, the evidence is most consistent with HFTs using anticipatory trading as one of their information-based strategies. It points to HFTs recognizing semi-persistent informed non-HFT order flow in real time. Since liquidity providers are slow to update their quotes, HFTs trade ahead of the impending non-HFT order flow and associated price change to earn a profit.

Another question is whether the results are more consistent with HFTs trading ahead of order

[7] Von Beschwitz, Keim, and Massa (2015) and Chordia, Green, and Kottimukkalur (2016) study high-frequency reactions to news announcements.

flow to avoid losses on market making inventory positions or to profit from speculation. To evaluate this question, I examine cumulative HFT net buying for stocks in extreme HFT net marketable buying portfolios before and after the sort second. There is no evidence marketable HFT trades are predominately caused by inventory management. It is possible the marketable trades are disposing of inventory acquired on exchanges other than NASDAQ. However, additional evidence is more consistent with a speculative interpretation. Marketable trades are more than half of HFT volume, which is inconsistent with a pure risk management motive. Additionally, HFT trades lead non-HFT trades more strongly early in the trading day and in small-cap stocks, whereas inventory risk is more binding towards the end of the day and HFT market making is more prevalent in large-cap than small-cap stocks (Yao and Ye 2015).

This is the first paper to provide evidence that HFTs anticipate buying and selling pressure from other investors. In a lower frequency context, Barbon, Maggio, Franzoni, and Landier (2017) and Maggio, Franzoni, Kermani, and Sommavilla (2017) show evidence of non-HFT brokers leaking information about one non-HFT client's trades to other non-HFT clients. Clark-Joseph (2013) studies a form of anticipatory trading in S&P 500 futures, though his focus is whether HFTs use small trades to learn about price impact. Recent work using data from Canada (Korajczyk and Murphy 2017) and Sweden (Van Kervel and Menkveld 2017) shows HFTs initially providing liquidity at the start of large multi-hour institutional orders before switching to trade in the same direction as the institutional order after 30 or more minutes.[8] Using U.S. equities data, this paper complements those studies by showing that if you condition on higher frequency second-by-second data, you observe HFTs taking liquidity ahead of non-HFTs at more favorable prices. I also provide evidence that HFTs' ability to predict order flow is not limited to the latter part of large orders; HFT marketable trades predict non-HFT marketable trades more strongly the first half hour of trading, which is necessarily early in a multi-hour order, and the second-level spikes in HFT marketable imbalances coincide with spikes in non-HFT marketable imbalances, which shows the patterns in this paper are distinct from longer-term trends. The higher-frequency analysis also allows examination of the trend-chasing and faster reaction to news alternatives.

These findings improve our understanding of HFTs' effect on liquidity, which relates to a broader

[8] Reiss and Werner's (1998) evidence on dealers trading ahead of worked orders is also related, though in their setting the presumption is the customer intends for the dealer to acquire the position beforehand.

debate about the merits of modern market structure.[9] Research generally shows that faster trading technology is associated with improvements in common liquidity measures.[10] Some of these benefits likely come from HFTs displacing less efficient incumbent liquidity providers.[11] There is also evidence that HFTs are net liquidity providers during extremely large idiosyncratic price moves.[12] This paper does not contradict evidence that the net effect of HFTs and electronic trading on liquidity is likely positive. Rather, the net effect on liquidity combines positive and negative channels, and this paper highlights one negative channel. A non-HFT whose trades are anticipated either trades fewer shares than they desire or obtains executions at a worse average price. These effects are hard to detect directly with standard liquidity measures, because doing so requires conditioning on the magnitude of a non-HFT's liquidity shock or, if the non-HFT is informed, the magnitude of their signal. In this way, the paper adds to other research examining potential negative effects of fast trading on liquidity provision, such as increased liquidity provider losses to cross-market arbitrage (Budish, Cramton, and Shim 2015) and to news trading (Menkveld and Zoican 2017).

This paper also contributes to the literature examining how automated trading affects price efficiency. There is evidence a fair portion of price discovery comes from algorithmic traders in general (O'Hara, Yao, and Ye 2014) and HFTs in particular (Brogaard, Hendershott, and Riordan 2014, 2016). More efficient prices are a good outcome. However, the welfare benefits of price discovery by HFTs depends in part on the information they use to trade. This paper provides evidence one reason HFTs appear informed is they extract information from informed non-HFT order flow. While this can speed price adjustment to non-HFTs' signals,[13] it also implies moderated benefits of HFT participation in price discovery for two reasons. First, it implies some information arriving via HFT trades would soon be incorporated into prices by non-HFT trades anyway. Second,

[9] Other issues include the effects of alternative trading venues (Foucault and Menkveld 2008, O'Hara and Ye 2011, Comerton-Forde and Putniņš 2015), tick size (Chao, Yao, and Ye 2016, Yao and Ye 2015, O'Hara, Saar, and Zhong 2015), high-frequency volatility (Kirilenko, Kyle, Samadi, and Tuzun 2017, Hasbrouck 2015, Egginton, Van Ness, and Van Ness 2016), overinvestment in technology (Biais, Foucault, and Moinas 2015), and agency conflicts caused by exchange fee rebates (Battalio, Corwin, and Jennings 2016).

[10] Examples include Hendershott, Jones, and Menkveld (2011), Boehmer, Fong, and Wu (2015), Hendershott and Riordan (2013), Hasbrouck and Saar (2013), Conrad, Wahal, and Xiang (2015), Conrad and Wahal (2016), and Brogaard, Hagstromer, Norden, and Riordan (2016).

[11] See Jovanovic and Menkveld (2015), Menkveld (2013), Breckenfelder (2013), Brogaard and Garriott (2015), and Ait-Sahalia and Saglam (2016).

[12] See Brogaard, Carrion, Moyaert, Riordan, Shkilko, and Sokolov (2017) and Griffith, Van Ness, and Van Ness (2017).

[13] Non-HFTs may at first trade less aggressively to hide from HFTs, slowing price adjustment. But once the HFT has learned how the non-HFT is trading, competition between them ultimately pushes prices closer to fundamental value (Madrigal 1996, Yang and Zhu 2017).

a non-HFT's profit conditional on a value-relevant signal is lower, which reduces incentives to acquire information (Grossman and Stiglitz 1980, Stiglitz 2014, Weller 2017).[14]

Finally, this paper contributes to our understanding of heterogeneity among HFTs. There is evidence HFTs follow a variety of different strategies (Boehmer, Li, and Saar 2016) and fast HFTs are more profitable (Baron, Brogaard, Hagströmer, and Kirilenko 2016). This paper shows a subset of HFTs appear to consistently focus more on anticipatory trading. These particular HFTs' marketable trades also predict returns more strongly. Accordingly, in addition to helping us understand why some HFTs are more profitable, the paper shows potential for limiting anticipatory trading while leaving liquidity providing HFTs unaffected.

1 Hypothesis Development

1.1 Theory of anticipatory trading.

There is substantial empirical evidence that marketable purchases lead to price increases and marketable sales lead to price declines (e.g., Hasbrouck (1991)). Two common explanations are that trades may contain information about fundamental values (e.g., Glosten and Milgrom (1985) and Kyle (1985)) or cause liquidity providers to take risky positions (Stoll 1978).[15]

It follows that a speculator able to predict marketable trade imbalances could use this information to help predict returns and earn a profit (e.g., Madrigal (1996), Brunnermeier and Pedersen (2005), and Yang and Zhu (2017)). This is especially true if the speculator is better at learning about future order flow than some liquidity providers. To see this, imagine the speculator identifies a signal indicating that a large trader will be selling in the near future. If the liquidity provider has not conditioned on this signal, then the liquidity provider's bid will be too high. The speculator can earn a profit by selling shares now and buying them back after the large trader's selling pushes prices down.

This theory has three types of traders in mind: a large trader who has price impact, a speculator

[14] These incentive effects arise when HFTs learn from the endogenous order flow of non-HFTs who pay a cost for their signals. This is absent from models in which the speculative HFT's trades do not affect the signal itself, such as Foucault, Hombert, and Rosu (2016).

[15] With the information story, selling shares causes the stock price to fall, because it indicates the seller thought the stock was overvalued. In contrast, the liquidity mechanism is sales cause the price to fall, because risk-averse liquidity providers require a price concession. This price concession compensates the liquidity provider for the additional risk borne as a result of adding the shares to their portfolio.

skilled at learning from order flow, and a liquidity provider who is less skilled than the speculator. Real world analogues for the large trader could be an active mutual fund who thinks a stock is mispriced or a pension fund liquidating assets to cover liabilities. HFTs are good candidates for the speculator. Trading on forecasts of order flow requires conditioning on a near continuous stream of quote updates and trades. This requires fast automated analysis and order routing technology, and HFTs specifically invest in these technologies. The liquidity providers could be non-HFTs or perhaps HFTs who are less skilled or who concentrate on different strategies.

Large traders consequently have an incentive to disguise their order flow, and in reality we see them deploy execution algorithms to achieve this goal. Yet even in theory (Madrigal 1996, Yang and Zhu 2017), some predictability remains. From a practical perspective, dark pool operator ITG writes, “some traders are more willing to increase fill rate at the expense of execution quality—the risk of information leakage and impact” (ITG 2013). HFTs’ ability to anticipate order flow, 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.

Information about order flow can be revealed in several ways. One way is through the practice of splitting a large order into small trades executed over a period of time. This practice is generally optimal for a trader who has discretion over when to trade, and it is useful for both informed traders (Kyle 1985) and liquidity traders (Admati and Pfleiderer (1988) and Vayanos (2001)). But splitting trades leaves open the possibility that in executing the first part of an order, the large trader might reveal more than he intends about the remaining parts. A similar type of information leakage can occur when a trader attempts to simultaneously send orders for a stock to different exchanges (e.g., NYSE and NASDAQ), but routing complications cause the orders to arrive at those venues at different times (Lewis 2014). A third way for predictability to arise is for traders with correlated trading demands to trade at different times; for example, a trade by one mutual fund might forecast future trades by similar funds. Additional sources of predictability could be changes to the order book (via new or cancelled orders) or trades in one stock predicting trades in related stocks (e.g., selling in GM predicts selling in Ford). These are several plausible examples, but any number of patterns could reveal information about trading demand.

1.2 Testable hypotheses

This theory can be tested by comparing the timing of HFT marketable trades to non-HFT marketable trades. An advantage of testable hypotheses relying only on transaction timing is that they are not biased by only having data from NASDAQ. A finding that, on NASDAQ, HFTs make marketable purchases and sales just before non-HFTs do the same is true on NASDAQ regardless what is happening on other exchanges and regardless the HFT's net position. To reverse such a finding for all markets, it would be necessary to find the reverse pattern elsewhere. However, while it is hard to know exactly how similar transactions are across exchanges, factors such as market structure convergence to electronic limit order books and the fact mostly the same HFTs and non-HFTs trade on all exchanges suggest NASDAQ is representative. Moreover, even if a finding were only valid on NASDAQ, it is an important subsample given it is the largest exchange in the sample period.

This motivates the first testable hypothesis. When an HFT anticipates marketable selling from non-HFTs that implies the current best bid is too high, the HFT should sell the stock with a marketable limit order that hits the bid. Afterwards, other investors will similarly sell with marketable orders, driving the price lower and creating a profit for the HFT. Thus, the HFT's marketable imbalance will be positively correlated with future non-HFT marketable imbalances. The HFT could also sell with a non-marketable order, but much of the variation in HFTs' non-marketable orders is caused by liquidity provision. By focusing on HFTs' marketable orders, I instead focus on times when their trades are more likely to be caused by information.

The HFT could be anticipating non-HFTs selling for information or liquidity reasons. Price changes after the HFT's trade provide insight into which of these explanations is more typical. The information explanation says the price will decline after the HFTs trade and stay down. The liquidity explanation says the price will fall but then reverse back to its initial price after the non-HFT's marketable selling subsides.

Hypothesis 1a *If HFTs anticipate and trade ahead of other investors' order flow, then HFT net marketable buying will be positively correlated with future non-HFT net marketable buying.*

Hypothesis 1b *Anticipatory trading implies HFTs' marketable trades will be positively correlated with future returns. If HFTs anticipate informed order flow, the price changes will be permanent,* *whereas if they anticipate uninformed order flow, the price changes will be temporary.*

One natural question is whether HFT trading forecasts non-HFT trading more strongly than the reverse pattern–non-HFT trading forecasting HFT trading. If HFTs learn about order flow from past trades, then there will be some correlation between non-HFT trades and future HFT trades. However, the evidence for HFTs trading ahead of predicted order flow is strongest if HFT trading predicts non-HFT trading more strongly than non-HFT trading predicts HFT trading.[16]

Hypothesis 2 *Anticipatory HFT trading is most consistent with HFT trades forecasting non-HFT trades more strongly than non-HFT trades forecast HFT trades.*

Anticipatory trading may be concentrated in a subset of the HFT firms. This may be the case if certain HFTs focus more on anticipatory strategies while others focus on liquidity provision, or if some HFTs are more skilled at anticipating order flow than other firms.

Hypothesis 3a *If certain HFTs are better at forecasting order flow or focus more on the strategy, then trades from these HFTs will be consistently more strongly correlated with future non-HFT trades than trades from other HFTs.*

Hypothesis 3b *If the HFTs whose trades consistently forecast non-HFT order flow the best are more skilled, then their trades should be more strongly correlated with future returns.*

The end of Section 1.1 discussed the a trade-off between disguising order flow and trading a large position quickly. An implication is that when non-HFTs are impatient, they may not hide their order flow as well, making it easier for HFTs to anticipate their trades.

Hypothesis 4 *The correlation between HFT trades and future non-HFT trades will be stronger at times when non-HFTs are more impatient.*

[16] This is how the theory has been presented. Nonetheless, it is worth noting that HFTs do not need to trade earlier than non-HFTs to profit from figuring out what the non-HFTs will be buying or selling (Brunnermeier and Pedersen 2005). For instance, if an HFT learns that informed non-HFTs are buying a stock, the HFT can buy the stock at the same time as the informed non-HFTs. In this scenario, the HFT is profiting from anticipating informed non-HFT order flow, but they are not necessarily trading before the non-HFTs.

2 Data

2.1 Sample construction

This study uses intra-day transactions data obtained from the NASDAQ Stock Market. NASDAQ was the largest U.S. exchange by volume during the January to December 2009 sample period,[17] with a 21.3% share of total volume (NASDAQ 10K 2009).[18] It trades all listed equities, including listings from the NYSE, ARCA, and AMEX exchanges. Like all other exchanges, it is structured as an electronic limit-order book and executions primarily come from professional traders.[19]

The 96 sample stocks (48 NASDAQ-listed and 48 NYSE-listed) are randomly selected from a sample universe intended to be representative of the stocks typically traded by actively managed investment funds. I use a random sample instead of all common stocks because of computational constraints. The sample is constructed from single-class CRSP common stocks (share code 10 or 11) at the end of 2008. I exclude the bottom two NYSE size deciles to roughly match common definitions of active funds' investable universe (e.g., Russell 3000 or MSCI Investable Market 2500). 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.[20] From this sample universe of 1,882 stocks, I then randomly select 6 NASDAQ-listed and 6 NYSE-listed stocks from each of the eight remaining size deciles.

Table 1 reports summary statistics for all stock days. The sample averages 93 stocks per trading day.[21] 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.

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

[18] This includes 33.0% of NASDAQ-listed volume and 15.7% of NYSE-listed volume.

[19] Most retail brokerages have contracts with market making firms who pay for the right to fill retail orders. 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.

[20] Appendix Table IA.2 summarizes stock-day observations for this sample of stocks.

[21] The number of stocks varies for two reasons: first, to ensure tick size does not constrain price movements, stocks are temporarily removed any day the prior close is less than $1; second, to remove stocks that stop trading due to bankruptcy or acquisition, stocks are permanently removed from the sample after a day with dollar volume less than $100,000. These filters use past rather than future data to avoid look-ahead biases. Similarly, I do not constrain the sample to stocks that pass these filters every day of the sample or add stocks to the sample when one of the original stocks exits, because doing so introduces look-ahead biases.

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.

2.2 Background on HFTs

The firms typically described as HFTs are proprietary trading firms using high-turnover automated trading strategies. Examples include Tradebot Systems, Inc., and GETCO. These firms are remarkably active traders. On their websites, Tradebot says they often account for more than 5% of total U.S. equity trading volume, and GETCO says they are "among the top 5 participants by volume on many venues" (Tradebot 2010, GETCO 2010). While estimates of their share of equity trading vary among sources, all estimates indicate HFTs are a large part of the market.[22]

2.3 HFT identification

The trade data from NASDAQ classifies trading firms as either an HFT or a non-HFT. The identities of the HFT firms and exact criteria used to determine the classification is not disclosed by NASDAQ. While the dataset is different from that in Brogaard, Hendershott, and Riordan (2014), the HFT versus non-HFT classification is the same. NASDAQ identifies 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.

A disadvantage of the classification scheme is that it is largely done at the broker level. As a result, HFT orders routed through the standard trading system of a broker that handles non-HFT orders, such as Goldman Sachs or Morgan Stanley, will be misclassified as non-HFT (Brogaard, Hendershott, and Riordan 2014). This is a concern primarily for HFTs that are internal divisions of larger investment banks. Independent HFTs who act as their own broker or use specialist brokers in order to be co-located and retain direct control of their routing infrastructure should be correctly

[22] 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).

classified. Section 4.3 evaluates whether potential misclassification has an important effect on the empirical results.

Table 1 shows 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.

2.4 Construction of returns and trade imbalances

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 best measure of a stock's quoted price.[23] When constructing the NBBO, quote data are filtered to remove anomalous observations.[24] 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.

This study uses net marketable buying and net buying imbalances. I use all trades during regular continuous trading hours (no open or closing cross). Marketable trades consist of all liquidity-removing sides of transactions. Non-marketable trades includes all liquidity-providing sides of transactions (both visible and hidden orders). The NASDAQ matching engine makes this marketable/non-marketable designation at the same time the trade is executed. 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 or-

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

[24] These filters are based on Hasbrouck's (2010) guidelines for constructing the NBBO. First, I remove quote updates where the bid on one exchange is greater than the ask on another. This can be a reporting artifact due to latency or a temporary arbitrage that gets corrected within fractions of a second. Second, I remove quotes where the bid-ask spread is more than 20% greater than the bid-ask midpoint. The purpose of this 20% filter is to remove very rarely occuring stub quotes (e.g., bid at $0.01 and ask at $20.00) that, if included, would cause large fluctuations in the bid-ask midpoint that is not reflective of a true price change. The 20% relative spread cutoff for what constitutes reasonable quotes comes from the SEC's initial single-stock circuit breaker rule. The rule halted stocks after a 10% return in 5 minutes to prevent "clearly erroneous transactions" (SEC 2010). This is consistent with a 20% relative spread limit, because a stock with successive trades first at a negligable spread and then second at a 20% relative spread (i.e., bid/ask 10% away from prior trade) would trip the circuit breaker. When an invalid quote is removed due to either filter, I use the last valid bid-ask midpoint for the price until a new valid quote is available. A new valid quote is typically available within a few milliseconds. 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.

der 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 liquidity demand. The net buying imbalance is simply shares bought minus shares sold and has previously been used to measure position changes of different investor groups (e.g., Griffin, Harris, and Topaloglu 2003). To put trade imbalances on a similar scale across stocks, I normalize all imbalance measures by a stock's 20-day trailing volume from CRSP.[25]

Panel B in Table 1 summarizes trade imbalances for the sample stocks prior to their being normalized by trailing volume. The table describes the distribution of the stock-day standard deviations of HFTs' net buying, their net marketable buying, and non-HFTs' net marketable buying. The mean standard deviation of HFTs' net buying ($\boldsymbol{HFT_{NB}}$) among all stock days is 83 shares, compared to 80 shares for their net marketable buying ($\boldsymbol{HFT_{NMB}}$). These figures are slightly smaller than the 100 shares that O'Hara, Yao, and Ye (2014) report as the median trade size on NASDAQ in 2008 and 2009. The average standard deviation of non-HFTs' net marketable buying ($\boldsymbol{non\text{-}HFT_{NMB}}$), 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.

2.5 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 such as *The New York Times*, *Wall Street Journal*, *Dow Jones newswire*, and *Reuters*. 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.[26]

[25] One could also adjust by the second or minute of the trading day to account for intra-day volume patterns, but such estimates for thinly traded stocks can by noisy. Potential effects related to the time of day are examined in Table 7.

[26] Table IA.3 summarizes the frequency of coverage and top sources for sample firms in the Factiva archive.

3 Tests of the anticipatory trading hypotheses

This section tests whether HFT trades forecast non-HFT trades and then examines additional empirical implications of the pattern being caused by anticipatory trading.

Throughout the paper, I use a method similar to Fama and MacBeth (1973) to account for cross-sectional correlation across stocks and autocorrelation through time. Observations in nearby seconds are likely highly correlated, so they should not be treated as independent. To be conservative, I collapse all estimates on a day to a single average estimate for that day. This gives me a time-series of the average effect each day. I then base hypothesis tests on the mean of that daily time series. In effect, I am asking whether the effect in question is consistently different from zero throughout the days in the sample. I am not assuming any independence among observations occuring on the same day. I only assume different days are independent. Of course, there is likely some residual autocorrelation between nearby days, so I use Newey and West t-statistics with the optimal lag length for the daily series chosen following Newey and West (1994). Despite the conservative approach to standard errors, many of the tests show high t-statistics. This happens when the estimated effect is consistent across the days in the sample.[27]

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

An immediate empirical implication of the anticipatory trading strategy is that an increase in HFT net marketable buying should forecast an increase in non-HFT net marketable buying (Hypothesis 1a) and an increase in returns (Hypothesis 1b). The first tests use portfolio sorts to examine the hypothesized relationships in a simple and easy to interpret form. Afterward, vector autoregressions (VARs) evaluate whether HFT trading has any predictive power beyond that of past returns and serial correlation in non-HFT trading.

The idea behind the sorts is to identify stocks with large HFT net marketable buying imbalances in a given second and then look at non-HFT net marketable buying and returns afterward. Every stock is assigned to one of ten portfolios each second based on the stock's HFT net marketable buying imbalance. Portfolio breakpoints are calculated from non-zero observations during the prior trading day. Then, a daily mean of the variable of interest for each portfolio is calculated by taking

[27] This precision in the estimate of each day's average implies that in reality there is a lot of independence among the second-level observations behind each daily estimate.

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.

Figure 1 plots cumulative net marketable buying for sort deciles ten (stocks HFTs bought) and one (stocks HFTs sold). Because the figure plots cumulative values, positive net marketable buying causes the line in that second to rise, while negative net marketable buying causes the line to fall. The dark red lines indicate HFTs' cumulative net marketable buying plotted on the left y-axis. Looking at the stocks HFTs buy, HFT net marketable buying is slightly positive before the sort, spikes in the sort period, and then is relatively flat afterwards. The sort-period spike is expected, because the line is the cumulative value of the variable used to form the portfolios. The real question is what happens to non-HFT net marketable buying.

Consistent with Hypothesis 1a, Figure 1 shows that in the seconds after these sorts on HFT net marketable buying, non-HFT net marketable buying is positive for the stocks HFTs bought and negative for the stocks HFTs sold. Cumulative non-HFT net marketable buying is depicted by dashed blue lines plotted on the right y-axis. In the seconds after the sort, cumulative non-HFT net marketable buying for the stocks HFTs bought steadily increases. This indicates consistently more marketable purchases than sales—non-HFTs buy the stocks HFTs previously bought.

Table 2 presents this sort data in a form conducive to hypothesis tests. In the first thirty seconds after the sort, cumulative non-HFT net marketable buying rises to 0.65 times the one-second standard deviation (or 27 shares for the median stock). Over the next four and a half minutes, it continues rising to 1.15 (or 48 shares for the median stock). These values are both significantly different from zero, with t-statistics of 14.35 and 2.41. The picture for the stocks HFTs sold most aggressively is symmetric; non-HFT net marketable buying in decile one is -0.69 and -1.89 times the one-second standard deviation. These results are not driven by a particular month or quarter in the year.[28]

[28] Compare for example these full sample results to Appendix Table IA.4, which is the same except it spans only the first two months of the year and uses NASDAQ BBO quotes for returns. Later results splitting HFTs into groups also

If the order flow patterns in Figure 1 and Table 2 are due to HFTs anticipating non-HFT buying and selling, then we should also see that the stocks HFTs bought aggressively have positive future returns and the stocks they sold have negative future returns (Hypothesis 1b). The returns for stocks in these portfolios are shown in Figure 2 and Table 3. Indeed, Figure 2 shows that in the 30 seconds after the sort, prices of stocks in HFT net marketable buying decile ten increase by 1.23 basis points, while the prices of decile one stocks decline by −1.07 basis points. Table 3 shows that both portfolios' return changes are significantly different from zero, with t-statistics of 11.78 and −13.48.

These price changes will be permanent if HFTs are anticipating order flow from informed traders and temporary if the order flow is from liquidity traders (Hypothesis 1b). The last two columns in Table 3 show there is some reversal in these returns over the next four and a half minutes. Decile ten reverses from 1.23 basis points 30 seconds after the sort to 0.57 basis points five minutes afterwards; decile one reverses from −1.07 to −0.58 basis points. Nonetheless, the return spread between these two portfolios remains positive. This is consistent with Brogaard, Hendershott, and Riordan's (2014) findings, using a state-space model, that HFT marketable trades forecast permanent price changes. Thus, the evidence is more consistent with HFTs anticipating informed rather than uninformed order flow.

The order flow and return patterns are consistent with the following anticipatory trading interpretation. Figure 1 shows that for the stocks HFTs buy, non-HFT net marketable buying picks up in the few seconds before the sort. HFTs know it is a good bet this buying will continue; marketable buying is known to have positive serial correlation (Hasbrouck and Ho 1987, Hasbrouck 1988, Chorida, Roll, and Subrahmanyam 2002), and this is also true of non-HFT net marketable buying in the sample. Conditioning on this serial correlation and more complex signals, HFTs infer the price is too low. Liquidity providers also realize the stock price is low, so a second before the sort they raise quotes and Table 3 shows the bid-ask midpoint rising by 4.51 basis points, or $0.01 for the median stock. But we are conditioning on times HFTs believe the quote update is insufficient, so in the sort second HFTs and non-HFTs buy the stock with marketable trades. Then, after the sort, non-HFT net marketable buying continues and the price rises an additional 1.23 basis points the next 30 seconds, creating a profit for the HFTs. HFTs also buy shares after the sort,

---

shows monthly results in which HFT trades are consistently positively correlated with non-HFT liquidity demand.

but in comparison to non-HFTs, they are buying relatively more in the sort second than afterward. This is of course not intended as an exact description of all marketable HFT trades. HFTs use marketable trades in a variety of strategies. Rather, the point is the patterns are consistent with anticipatory trading being one of their strategies.

The sorts provide a few additional insights. First, the order flow and return results are stronger in small-cap stocks. In the first 30 seconds after the sort, Table 2 shows standardized non-HFT net marketable buying for small-cap stocks HFTs bought (decile ten) is 1.39 versus −1.27 for stocks they sold (decile one). This compares to 0.61 and −0.69 in mid-cap stocks and 0.39 and −0.39 in large-cap stocks. This greater order flow predictability in small-caps persists through the next four and a half minutes. Similarly, looking at returns the first 30 seconds after the sort in Table 3, the decile ten minus decile one return spread is 5.18 basis points (2.54 −(−2.64)) for small-cap stocks, compared to 2.80 for mid-cap stocks and 0.55 for large-cap stocks. Four and a half minutes later, the decile ten minus decile one return spreads for small and mid-cap stocks remain positive (4.45 and 1.52 basis points), while prices of large-cap stocks have fully reversed. The larger post-sort spread in smaller stocks could be due to non-HFTs having a harder time disguising order flow when trading relatively illiquid stocks. This intuition is consistent with a *Traders Magazine* article on Goldman Sachs small-cap execution algorithms that emphasizes the difficulty of trading illiquid securities without leaking information (Traders Magazine 2013). Section 3.4.2 examines this illiquidity hypothesis in more detail.

Second, Table 3 provides evidence HFTs chase short-term trends with their marketable orders. Stocks HFTs buy with marketable orders have higher returns the prior 30 seconds (4.43 basis points) than the stocks they sell (−4.30 basis points). The trend chasing is notable, because HFTs are often characterized as liquidity providers, and liquidity provision is typically assocated with contrarian trading. Thus, the actions of HFTs using these market orders are inconsistent with standard notions of liquidity provision.

Third, that HFT trades predict non-HFT trades at horizons much longer than a second shows the result is not an artifact of cross-market arbitrage. This type of arbitrage is understood to occur at sub-second horizons (see Budish, Cramton, and Shim 2015).

Table 4 builds on the sort results using a vector autoregression (VAR) similar to Hasbrouck (1991). The portfolio sorts show evidence of positive serial correlation and trend-chasing in non-

HFT net marketable buying. These effects provide a potentially simple way for an HFT to predict non-HFTs' trades. The VAR isolates the predictive ability of HFTs' marketable trades in excess of anything coming from these simple signals.

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:

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

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

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

where $\boldsymbol{R_t}$ is the one-second NBBO bid-ask midpoint return, $\boldsymbol{HFT_{NMB,t}}$ is one-second HFT net marketable buying, and $\boldsymbol{non\text{-}HFT_{NMB,t}}$ is one-second non-HFT net marketable buying. The equations are estimated for each stock every day. All variables are standardized by their standard deviation among all stocks that day.

Table 4 reports the VAR results in two different formats. Panel A first averages all stocks' coefficients on the same day and then reports significance tests based on the mean of the resulting daily time series (Fama and MacBeth 1973). Panel B shows additional information about the full cross-sectional coefficient distribution pooled across all stock days: the mean stock-day coefficient and the percent that were either positive and significant or negative and significant in the stock-day regression. The conclusions from either panel are the same, so I focus the discussion on Panel A for brevity.

First, the VAR confirms the positive serial correlation and trend chasing in non-HFT net marketable buying that motivated the test. In the equation where non-HFT net marketable buying is the dependent variable, average coefficients on all lags of itself are positive and significantly different from zero, starting at 0.0754 at lag one and declining to 0.0117 at lag ten.[29] With respect to trend chasing, return coefficients imply a one standard deviation increase in returns leads to a 0.9181 standard deviation increase in non-HFT net marketable buying one second later. While six of the lag two through lag ten return coefficients are negative, these coefficients are much smaller

[29] The serial correlation in HFTs' net marketable buying is lower. A one standard deviation increase in HFT net marketable buying leads to a 0.0243 standard deviation increase in the same variable the next second. Coefficients decline with additional lags to 0.0007 at lag 10.

than the first lag, so the trend-chasing effect dominates.[30,31]

The main question, then, is whether HFT net marketable buying is still positively correlated with future non-HFT net marketable buying after controlling for returns and serial correlation. In fact, Table 4 shows the positive correlation still exists in the VAR. In the equation forecasting non-HFT net marketable buying, the lag one coefficient on HFT net marketable buying is 0.0007, rising to 0.0021 at lag two and then declining slowly to 0.0016 at lag ten. Though the first lag coefficient is the smallest, they are all significantly different from zero, and the decay in predictability from lags two to ten is slow.

Similarly, the VAR confirms the sort evidence that HFT net marketable buying positively predicts future returns. A one-standard deviation increase in lag one HFT net marketable buying leads to a 0.0185 standard deviation increase in the next second's return, with a t-statistic of 10.92. Coefficients remain positive and significantly different from zero up through the ninth lag.

The VAR also shows that non-HFT net marketable buying is positively correlated with future returns and, as in the sorts, HFT net marketable buying chases short-term returns. The result for non-HFT marketable imbalances predicting returns is shown by the positive coefficients on lagged non-HFT variables in the return forecasting regression. This shows predicting non-HFT net marketable buying can help predict returns.[32] HFTs chasing short term trends with marketable orders is shown in the regression with HFT net marketable buying as the dependent variable. Coefficients on the first two lags of returns are large and positive. At higher lags, the coefficients are negative, though substantially smaller. This means there is strong trend-chasing on very short-term returns, followed by a shift to weaker contrarian trading at longer horizons. As discussed in the sorts, the strong short-horizon trend chasing is more consistent with these trades demanding liquidity than providing it.

This section evaluated lead-lag relationships between HFT and non-HFT net marketable buying

[30] 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.

[31] One concern might be that the apparent trend-chasing behavior could be driven by misaligning trade and NBBO quote time-stamps. Appendix Table IA.4, which uses NASDAQ BBO quotes from the same computer that assigns trade timestamps, shows there is still evidence of trend-chasing when using precisely aligned timestamps.

[32] To be precise, the VAR conditions on HFT net marketable buying and returns, so the return predictability denoted by these non-HFT coefficients is coming from variation in non-HFT net marketable buying uncorrelated with the HFT and return variables. However, non-HFT net marketable buying also predicts returns if these controls are left out.

using portfolio sorts and VARs. Both sets of tests provide results consistent with Hypotheses 1a and 1b. Specifically, an increase in HFT net marketable buying is followed by an increase in non-HFT net marketable buying and returns. These effects are strongest in small and mid-cap stocks and more consistent with HFTs anticipating informed order flow. We also saw that both HFT and non-HFT net marketable buying chase short-term trends and positively predict future returns.

3.2 Do HFT trades predict non-HFT trades better than non-HFT trades predict HFT trades?

This section uses the VAR from Table 4 to test whether HFT trades predict non-HFT trades more strongly than non-HFT trades predict HFT trades (Hypothesis 2). As discussed in the hypothesis section, this result is not a necessary requirement, but it is most consistent with the anticipatory trading story.

Table 4 shows HFT net marketable buying does in fact predict non-HFT net marketable buying more strongly than the reverse. The lag one coefficients at first seem to contradict this statement; in the HFT forecasting equation, the coefficient on non-HFT net marketable buying is bigger (0.0028) than the HFT coefficient in the non-HFT forecasting equation (0.0007). However, this flips at higher lags. In the HFT forecasting regression, the non-HFT coefficient is near zero at lag two and becomes negative at higher lags. In contrast, in the non-HFT forecasting regression, the HFT coefficient rises to 0.0021 at lag two and then declines slowly but remains positive through lag ten. Putting all ten lags together, the non-HFT coefficients sum to -0.005 when predicting HFT net marketable buying, compared to a sum of 0.016 for the HFT coefficients when predicting non-HFT net marketable buying.

Impulse response functions provide another way to evaluate magnitudes. They estimate the effect a shock to one variable in the VAR system has on other variables over time.[33] Figure 3 reports these functions for HFT and non-HFT trading. The light blue line plots the response of non-HFT net marketable buying to a 100 share shock to HFT net marketable buying. The dark red line plots the opposite effect—the response of HFT net marketable buying to a 100 share shock to non-HFT net marketable buying.

[33] The impulse response functions assume no contemporaneous effects among the variables, so the ordering of the variables in the VAR does not matter. These assumptions allow for non-HFT net marketable buying to affect future HFT net marketable buying, which would be the case if some HFTs make markets.

Using the the ten lag VAR, the top panel of Figure 3 shows that by ten seconds after an initial shock, the response of non-HFT trading to HFT trading is higher than the response of HFT trading to non-HFT trading. From this point forward, the two response functions are outside each other's 95% confidence intervals. The non-HFT response to HFT trading stabilizes at 11 shares, while the HFT response to non-HFT trading stabilizes at 9 shares. If the VAR is extended to 30 lags, the difference becomes even more pronounced; the non-HFT response stabilizes at 17 shares and the HFT response stabilizes at 10 shares.

Both the coefficient comparisions and impulse response functions support Hypothesis 2. HFT net marketable buying predicts non-HFT net marketable buying more strongly than non-HFT net marketable buying predicts HFT net marketable buying.

3.3 Is predicting order flow related to HFT skill?

This section tests first whether some HFTs are better at predicting order flow (Hypothesis 3a) and second if the HFTs who predict order flow the best are also better at predicting returns (Hypothesis 3b).

3.3.1 Are some HFTs better at forecasting order flow?

Different levels of skill or focus on anticipatory strategies will cause trades from some HFTs to have stronger correlations with future non-HFT trades. This section examines this issue with regressions similar to the non-HFT forecasting regression in Table 4. The difference here is the regression is estimated separately for each HFT in the sample.

There are two regression estimates for each HFT every day. In the first, the HFT variable is the HFT's net buying. Since these estimates are calculated for each HFT separately, we do not have to worry that variation in net buying by HFTs anticipating order flow might be obscured by variation in net buying caused by HFTs providing liquidity.[34] In the second regression, the HFT variable is the HFT's net marketable buying, with the slight adjustment that if one of either the HFT's net marketable buying or net buying is positive and the other is negative, then net marketable buying is set to zero. The adjustment is included in this test due to a limitation in how the data was

[34] This is a concern when examining aggregate HFT trading, which is why those results condition on marketable trades to more precisely focus on trades likely to forecast order flow.

constructed, but it is not important.[35] Those concerned about the adjustment to net marketable buying can focus on the firm-level HFT net buying results. The heading for Table 5 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. 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 consistency is to look at the probability an HFT in the highest correlation group remains in that group in future months. Figure 4 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, under the null hypothesis of no persistent difference among HFTs, only 33.3% of the HFTs should still be in the highest-correlation group one month after the sort. In fact, whether HFTs are sorted into high and low correlation groups by coefficients on only the first or on 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 consistently have a stronger positive correlation with future non-HFT order flow.

Another way to examine persistence is to compare post-sort month coefficients for the three HFT groups. If correlations 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 5 reports post-sort month coefficients for the three HFT groups. The first group of columns examines the persistence of the coefficient on the first lag of HFTs' net marketable buying,

[35] Applying this adjustment to aggregate HFT net marketable buying data does not affect the aggregate results. Unfortunately, the firm level data used in this section is not available both ways. However, given the adjustment did not matter in aggregate results, it should not matter in these firm level results. The reason is adjusted and unadjusted HFT net marketable buying will be even more similar to each other at the firm level than in aggregate, because an HFT will not provide liquidity to itself. As a result, the adjustment is rarely applied in practice because it is almost always the case an HFT's net buying and net marketable buying is the same direction within a second. The historical reason for this adjustment is that when the dataset was constructed on NASDAQ's servers, it was assumed the adjustment gave a slightly better measure of directional trading intent. The aggregate data was generated using both regular and adjusted net marketable buying, but the firm-level data was generated using only the adjusted version. I now focus the aggregate results on regular net marketable buying, because it is easier to interpret and the adjustment has no effect on results.

$\overline{\gamma_{d,t,1}}$. The average $\overline{\gamma_{d,t,1}}$ for the highest-correlation group is 0.014, compared to 0.002 for the lowest-correlation group. Both the Newey and West (1994) p-value and non-parametric rank sum p-value indicate this difference is significantly different from zero. The left group of columns in the bottom half of Table 5 show results using ten lags of HFTs' net marketable buying. 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 right group of columns in Table 5 report results from tests using HFTs' net buying rather than net marketable buying. Results from these tests are the same as for the net marketable buying tests—whether one looks at the coefficients on the first lag or on all ten lags, there are persistent differences between the highest and lowest-correlation groups. These differences are also consistent throughout the months in the sample, as shown in Figure 5. These results indicate that some HFTs' trades are consistently more strongly correlated with future non-HFT order flow.

3.3.2 Are the HFTs who are the best at forecasting order flow 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. Hypothesis 3b is that 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, Hagströmer, and Kirilenko (2016) also examine skill differences among HFTs, but they focus on differences in performance rather than the relationship between ability to predict order flow and ability to predict returns.

Table 6 examines whether marketable trades from the HFTs whose marketable 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 5. Then, trades are aggregated among HFTs in each group, resulting in one time-series of aggregated trades from the above-median group and one from the below-median group. 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 6 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 0.0288, 0.0190, 0.0142, and 0.0099, compared to 0.0207, 0.0134, 0.0093, and 0.0075 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 greater than four lags, there is generally no significant difference between the two series. These findings are consistent with the above-median HFTs being more skilled at predicting short-term returns.

3.4 Are non-HFTs' trades more predictable when they are impatient?

Hypothesis 4 proposes that non-HFTs' order flow is more predictable when they are impatient. This section tests the hypothesis by comparing VAR estimates from normal times to estimates from times when non-HFTs are hypothesized to be impatient: at the market open and close, on days with high volume or high trade imbalances, and when trading illiquid stocks. The VAR is the same as in Section 3.1. I focus on the equation where the dependent variable is non-HFT net marketable buying. If HFTs are doing relatively more anticipatory trading, then I expect larger positive coefficients on the lagged HFT imbalance variables.

3.4.1 Theory for the impatience proxies

Non-HFTs may be impatient at the open or close for a variety of reasons. Imagine an investor who receives a signal overnight. This signal could be a major news announcement or the output from overnight investment analytics. The investor knows it is possible other investors either received the same signal or will receive it shortly. Therefore, the investor knows they need to trade near the open 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.

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

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 orders that are harder to hide. This intuition is consistent with the Goldman Sachs anecdote about small-cap execution algorithms discussed on page 17.

3.4.2 VAR estimates when non-HFTs are hypothesized to be impatient

The first impatience proxy is time of day. Panel A in Table 7 compares estimates of how HFT trading predicts non-HFT trading in the first and last half hours of the trading day to estimates from the middle of the day. In the morning, both the coefficient on the first lag of HFT net marketable buying (0.0026) and the sum of all ten lag coefficients (0.0207) are greater than those from the middle of the day (0.0003 and 0.0131). The differences are significantly different from zero. In contrast, near the market close, the first lag coefficient (-0.0050) is actually negative. Coefficients on the next few lags, though positive, are also less than those from the middle of the trading day. Accordingly, the predictability is stronger at the open but not the close.

The second impatience proxy is days with high trading activity. Panel B in Table 7 analyzes VAR estimates for stocks when volume or the absolute value of total net marketable buying for the stock is high compared to other days. The methodology, inspired by Gervais, Kaniel, and Mingelgrin (2001), classifies a day as high volume or high imbalance if the volume/imbalance is among the top 10% of days the past month (i.e., rank relative to prior 19 trading days is 19 or 20).[36]

Overall, marketable HFT trades appear to predict marketable non-HFT trades more strongly on high volume days, but there is not a big difference on high imbalance days. On high volume

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

days, both the coefficient on the first lag of HFT net marketable buying (0.0036) and the sum of all ten lag coefficients (0.0235) are significantly greater than those from other days (0.0004 and 0.0149). On high imbalance days, only the second lag coefficient is consistently higher than other days (mean difference is 0.0012 with t-statistic of 2.30), and the sum of all ten lags is not significantly different from other days (t-statistic is 1.53).

Table 7 Panel C contains the final impatience test. It examines illiquid stocks with large bid-ask spreads. Bid-ask spreads are calculated in two ways: the left column group uses regular quoted spreads, while the right column group uses relative bid-ask spreads, which are the quoted spread divided by the bid-ask midpoint.[37]

Results are the same using either spread definition—HFT marketable trades predict non-HFT marketable trades more strongly in illiquid stocks. Using regular spreads, the first lag coefficient is 0.0069 higher in high spread stocks than low spread stocks, and the sum of all ten lag coefficients is 0.0088 higher in high spread stocks. Similarly, using relative spreads the first lag and sum of all lag coefficients are 0.0045 and 0.0049 higher in high spread stocks.

The above three tests show that at times when non-HFTs are hypothesized to be relatively impatient, HFT net marketable buying is generally more strongly correlated with future non-HFT net marketable buying. Predictability is reliably stronger at the market open, on high volume days, and in illiquid stocks. Predictability near the close is weaker, perhaps because HFTs do not want to build an inventory position they cannot unwind before the end of the day.[38] The difference between the high impatience times and other times is mostly concentrated in the coefficients on the first few lags of HFT net marketable buying. Overall, the evidence is consistent with HFTs having an easier time anticipating order flow when non-HFTs are more constrained when trying to disguise their trades.

4 Alternative explanations

While the anticipatory trading hypothesis predicts that HFT liquidity demand will be positively correlated with future non-HFT liquidity demand, there are other possible explanations. Any signal

[37]By putting spreads in terms of a percent of the stock's price, relative spreads adjust for the fact that a liquid stock may have a wide nominal spread if the price is high enough.

[38] This explains weaker effects at the close, but not the negative coefficient on the first lag of HFT net marketable buying.

that HFTs and non-HFTs separately observe and then trade on could result in HFT liquidity demand positively forecasting non-HFT liquidity demand, particularly if HFTs react faster. This contrasts with anticipatory trading, in which the HFT does not directly observe the same signal as the non-HFT but instead infers it from the non-HFT's trading activity.

It is difficult to completely rule out this alternative, because it is not possible to fully examine all inputs to HFT and non-HFT trading processes. Nevertheless, if the relation between HFT and non-HFT trading is still strong after accounting for the most likely signals they would both utilize in close succession, then it makes this correlated signal explanation less probable. To examine the issue, this section explores the two most likely signals HFTs and non-HFTs could both be reacting to: news and past returns. It then evaluates whether the results could be caused by trader misclassification, another correlated signal explanation.

4.1 Are HFTs simply reacting faster to news?

Reacting to news faster than other investors is an important HFT skill (Ye, Yao, and Gai 2013, Foucault, Hombert, and Rosu 2016) and plausible explanation for HFT marketable trades leading non-HFT marketable trades.

Alternative Hypothesis 1 *If HFT trades lead non-HFT trades because HFTs are reacting faster to firm news, then the lead-lag relationship should go away after excluding periods surrounding publication of news articles about the firm.*

Table 8 tests this hypothesis using the Table 4 VAR estimated seperately on days with and without news. The table uses two definitions of a news day for a firm. The first definition (Panel A) is any day when an article in the Factiva database mentions the firm. The second definition (Panel B) is any day when the absolute market-adjusted stock return is greater than one percent. The return-based definition captures news events that may be excluded from the Factiva database (e.g., analyst forecasts not covered by the media). The objective is to determine whether the previously identified correlation between HFT and non-HFT net marketable buying exists even on days with no news.

The middle group of columns in each panel of Table 8 contains estimates for non-news days. In Panel A, the average non-news day lag one coefficient on HFT net marketable buying is 0.0019 and, with a t-statistic of 4.95, is significantly different from zero. Lags two through ten and the sum of all ten lags are also positive and significantly different from zero. In Panel B, the lag one coefficient estimate on days with small returns is approximately zero, but all higher lags are positive and significantly different from zero. In both panels, the news day and non-news day coefficient estimates are similar. Some individual coefficients are different, but the overall pattern is the same.

These results showed HFT net marketable buying positively forecasts non-HFT net marketable buying even when there is no news. This finding is inconsistent with the lead-lag relationship between HFT and non-HFT net marketable buying being solely caused by HFTs reacting faster to news.

4.2 Is the explanation non-HFTs and HFTs both chase price trends?

Another hypothesis for why HFT trading might lead non-HFT trading is that non-HFT trading might have a simple predictable relationship with past returns. Section 3.1 showed HFTs and non-HFTs both chase short-term price trends with their marketable trades. It is possible HFT trades predict non-HFT trades because they are both reacting to lagged returns, except HFTs react faster. Additionally, with non-HFTs following trend-chasing strategies, another possibility is purchases by HFTs could actually cause future non-HFT buying through their effect on returns.

Alternative Hypothesis 2 *If HFT net marketable buying leads non-HFT net marketable buying because non-HFTs chase past price trends, then there should be no correlation between HFT and non-HFT trading after controlling for lagged returns.*

This alternative hypothesis is addressed by the VAR in Table 4. The coefficients in regressions of non-HFT net marketable buying on HFT net marketable buying are positive controlling for lagged returns. HFT trades may cause some non-HFT trend chasing through their effect on returns, but it is not sufficient to explain the lead-lag correlation between HFT and non-HFT trading.

4.3 Are HFT trades forecasting misclassified HFT trades rather than true non-HFT trades?

Finally, it is possible HFT trading predicts non-HFT trading because some HFT trades are misclassified. Misclassification can occur if an HFT routes orders through brokers that handle non-HFT orders, such as Goldman Sachs and Morgan Stanley (Brogaard, Hendershott, and Riordan 2014).[39] The problem with misclassification is that HFT trading is serially correlated. This serial correlation could be caused by some HFTs being slower or by trade splitting. Either way, if you take some of these later HFT trades and throw them into the non-HFT bucket, then current HFT trading will forecast those later misclassified trades. Consequently, misclassification could lead to an erroneous conclusion that HFT trades were forecasting non-HFT trades.

To get a sense for how HFT trades would forecast misclassified trades, you can look at how HFT trading forecasts itself. If later correctly classified HFT trades are representative of later misclassified trades, then the way that correctly classified HFT trading forecasts itself should be proportional to the way it forecasts misclassified trades. For example, a 100 share purchase by correctly classified HFTs might forecast that correctly classified HFTs purchase 20 shares 5 seconds from now and 10 shares 10 seconds from now, while misclassified HFTs purchase half those amounts 5 and 10 seconds from now. The essential idea is that the decay in predictability will be approximately the same. Importantly, the above hypothesis can be true even if correctly classified trades are not entirely representative of misclassified ones. It is only necessary that the later correctly classified HFT trades that generate the serial correlation are representative of later misclassified trades.

We can make additional predictions under the assumption that most HFT trades are correctly classified. This is a reasonable assumption given NASDAQ's classification scheme attributes roughly 40% of dollar volume on the exchange to the HFT trades they identified.[40] If correctly classified HFT trades account for the majority of HFT trades, then HFT trading should predict itself more than it predicts non-HFT trading. This is because misclassification implies non-HFT trading is a noisy HFT proxy, so lead-lag correlations should be attenuated relative to a sample with no noise (i.e., the correctly classified HFT trades).

Alternative Hypothesis 3a *Misclassification implies the decay in the correlation between non-HFT trading and lagged HFT trading should be proportional to the decay in the correlation between HFT trading and lagged values of itself.*

Alternative Hypothesis 3b *Misclassification implies HFT trading should forecast itself more strongly than non-HFT trading.*

[39] The sample construction section (2.1) discusses potential misclassification in more detail.
[40] See Figure IA.1.

The sort (Figure 1 and Table 2) showed that if you condition on a spike in HFT net marketable buying, non-HFT net marketable buying spikes and then continues to drift in the same direction for up to five minutes into the future. At the same time, the spike in HFT net marketable buying trails off much faster towards zero. This is evidence that the correlation between HFT net marketable buying and itself decays faster than its correlation with non-HFT net marketable buying, inconsistent with Alternative Hypothesis 3a.

Figure 6 tests the misclassification hypotheses more precisely by comparing the VAR coefficients on HFT net marketable buying in Table 4. The dark red line plots coefficients from the regression where the dependent variable is HFT net marketable buying, while the light blue line is for the regression where the dependent variable is non-HFT net marketable buying. The VAR regression coefficients are standardized, so to facilitate comparison they are transformed to show the number of shares change in the dependent variable for a 100 share change in HFT net marketable buying.[41] The y-axis is capped to zoom in on higher lags.

First, the figure shows the decay of the HFT to lagged HFT correlation is faster than the decay of the non-HFT to lagged HFT correlation. Using the VAR with 10 lags, Panel A shows coefficients for the first few lags are much higher when HFT net marketable buying is the dependent variable (dark red lines) than when non-HFT net marketable buying is the dependent variable (light blue lines). However, as the lags increase, the coefficients when HFT net marketable buying is the dependent variable drop steadily, while the coefficients when non-HFT net marketable buying is the dependent variable remain comparatively stable. This finding contradicts Alternative Hypotheses 3a.

Second, the fact HFT net marketable buying predicts non-HFT net marketable buying better than itself at higher lags conflicts with Alternative Hypothesis 3b. Specifically, after seven lags, the HFT coefficient is actually bigger in the regression where the dependent variable is non-HFT net marketable buying. The coefficients in Panel A imply a 100 share increase in HFT net marketable buying ten seconds ago is associated with a 0.066 share increase in HFT net marketable buying, compared to a 0.244 share increase in non-HFT net marketable buying. Using the VAR with 30 lags, Panel B shows the coefficient in the HFT regression is roughly zero by the fifteenth lag, while the coefficient in the non-HFT regression remains positive and significantly different from zero even

[41] For example, the lag 1 HFT coefficient in the non-HFT regression in Table 4 is 0.0007. Table 1 shows the standard deviations of $\boldsymbol{HFT_{NMB}}$ (80) and $\boldsymbol{non\text{-}HFT_{NMB}}$ (125). The coefficient in terms of a 100 share change in HFT net marketable buying is 0.0007×125×100/80=0.1

after 30 lags.

If the explanation for why HFT marketable trades predict non-HFT marketable trades is misclassification, then trades from correctly classified HFTs should not forecast misclassified HFTs' trades *more* strongly than correctly classified HFTs' trades. It is also unclear why misclassified HFTs' trades would be much slower moving (predictable 30 seconds later) than correctly classified HFTs (not predictable 30 seconds later). The simpler explanation is that HFT marketable trades are mostly forecasting correctly classified non-HFT marketable trades, especially when looking 10 seconds or more into the future.

5 Relation to inventory management

Order anticipation typically comes up in the context of speculation. Being able to anticipate order flow or, equivalently, predict prices conditional on order flow, is also useful for market makers managing the risk of their inventory positions. This is relevant because market making is a well known HFT strategy. Consider a market making HFT who acquires a short position in a stock by selling to liquidity demanding non-HFT buyers. If the HFT subsequently realizes these non-HFT purchases will continue and push prices up, then the HFT can avoid a loss by exiting the short position quickly with a marketable buy order. In this scenario, the HFT buys ahead of non-HFTs to get out of a losing position.

Inventory management and speculation have different implications for how HFTs' marketable buying is related to existing positions. Marketable trades that reverse existing positions are consistent with inventory management.[42] Speculation is consistent with the other leg of a marketable purchase occuring beforehand or afterwards, though a common expectation is to see a non-marketable sale afterward.

The NASDAQ data does not contain position information. However, HFT positions can be estimated by cumulating their net buying across all NASDAQ trades. A downside to relying on data from a single exchange is it misses position changes caused by trades on other venues (Menkveld 2013, Reiss and Werner 1998). In particular, if an HFT purchases a share of GE on

[42] For research on inventory management by NYSE specialists, see Hasbrouck and Sofianos (1993), Madhavan and Smidt (1993), Hendershott and Seasholes (2008), and Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010).

NASDAQ and sells the share on the NYSE, they will appear to be long one share of GE in the NASDAQ data when they actually have no position. Nevertheless, the estimates can be informative.

Figure 7 plots cumulative HFT net buying for the first and tenth net marketable buying decile portfolios. The portfolios are the same as in Figure 1, only the plotted variable is net buying instead of net marketable buying. The objective is to look at estimated position changes before and after these sorts on HFT net marketable buying. The solid line is for decile ten, the decile of stocks HFTs buy most aggressively. The dotted line is for decile one.

If HFTs' marketable purchases in the sort period are unwinding short inventory positions accumulated over the prior 30 seconds, then the solid line indicating cumulative net buying for decile ten should be falling in the pre-sort period. However, net buying for decile 10 is slightly increasing in the 30 seconds before the sort. Similarly, inventory management implies the dotted line showing cumulative net buying for decile one should be rising in the pre-sort period, but instead it is slightly declining. Looking at net buying after the sort, we also see no evidence of position reversal there. Therefore, in expectation the other leg of the sort-period trades occurs on another venue or on NASDAQ more than 30 seconds before or after the sort.

This test shows no evidence HFT trades in the sort period dispose of inventory positions previously acquired on NASDAQ. It is possible they are disposing of inventory acquired on another exchange. An HFT might sell to liquidity demanding non-HFTs on the NYSE and then, realizing non-HFT buying will continue and push the price up, turn around and buy the shares back on NASDAQ. It could be necessary from a welfare perspective to allow HFTs to unload their inventories in these instances so they can provide liquidity at other times. However, it is important to keep in mind this scenario still implies the HFT correctly predicts the non-HFT's order flow, trades ahead of it to increase their profits, and is taking liquidity the non-HFT would have otherwise accessed. It is still anticipatory trading.

That said, the evidence is more consistent with a speculative trading interpretation. Marketable trades are too large a share of HFT volume to be primarily due to inventory management. They account for over half of HFT dollar volume in the sample and, since they pay the bid-ask spread and take fee, market makers want to minimize their use (Harris 2002). Consequently, it is probable a lot of the variation in marketable trades is from non-market making strategies. Speculation is also more consistent with the evidence in Table 7 that HFT trades predict non-HFT trades more strongly early in the trading day. Market makers typically like to end the trading day with minimal inventory positions. Therefore, the end of the trading day, when the results are weakest, is also the time when market makers' desire to exit inventory positions will be strongest. The stronger results in small-cap stocks (Table 2) is also less consistent with inventory management, because these are the stocks in which HFT market making is least prevalent (Yao and Ye 2015).

Turning back to Figure 7, the net buying pattern is consistent with speculation where the HFT is equally likely to be getting into or out of the position and has no immediate need to get back to a flat inventory level. This is plausible given the median sort period position change is $3,609. Positions this small are not risky.

While a holding period longer than a few seconds contradicts some notions of high-frequency trading, it is perhaps important to distinguish between holding periods of market making and speculative strategies. To this point, the HFT firm Hudson River Trading told the *Wall Street Journal* it trades 5% of U.S. volume using both market making and speculative strategies, with 60% of their strategies taking liquidity (Hope 2014). They said their average holding period was about five minutes and positions held overnight commonly accounted for 25% of firm capital. Thus, holding periods of more than a few seconds are not unusual. It should be said the firm explicitly states in the article they "... don't try to race ahead of an insitution's order ..." Nonetheless, for some subset of HFTs, the evidence in this paper shows those HFTs' marketable trades do just that.

6 Conclusion

This study examines whether HFTs anticipate and then trade ahead of other investors' order flow. I find that if HFTs on net buy a stock with marketable orders during a particular second, then non-HFTs will subsequently buy this same stock with marketable orders. This forecasted non-HFT net marketable buying continues for at least five minutes after the sort second. I explore several explanations for these findings, including their being driven by HFTs reacting faster to news, non-HFTs chasing return trends, and trader misclassification. However, the findings are best explained by HFTs learning which stocks non-HFTs will be buying or selling and then trading ahead of that order flow and its associated price impact.

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 days, and in stocks with wide bid-ask spreads. I also find evidence consistent with HFTs varying in their skill at predicting non-HFT order flow, and that trades from HFTs who appear to be the most skilled at predicting order flow predict larger price changes. 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. 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 trade ahead of other investors' order flow. Future research could examine which characteristics of past trades and the order book allow HFTs to predict the stocks non-HFTs will buy and sell. 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 better at interpreting it. 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

Admati, Anat R., and Paul Pfleiderer, 1988, A theory of intraday patterns: volume and price variability, *Review of Financial Studies* 1, 3–40.

Ait-Sahalia, Yacine, and Mehmet Saglam, 2016, High frequency market making, *Available at SSRN 2331613*.

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.

Baldauf, Markus, and Joshua Mollner, 2015, High-frequency trading and market performance, *Available at SSRN 2674767*.

Barbon, Andrea, Marco Di Maggio, Francesco Franzoni, and Augustin Landier, 2017, Brokers and order flow leakage: Evidence from fire sales, Discussion paper, National Bureau of Economic Research.

Baron, Matthew D., Jonathan Brogaard, Björn Hagströmer, and Andrei A Kirilenko, 2016, Risk and return in high frequency trading, *Available at SSRN 2433118*.

Battalio, Robert, Shane A Corwin, and Robert Jennings, 2016, Can brokers have it all? on the relation between make-take fees and limit order execution quality, *The Journal of Finance*.

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.

Biais, Bruno, Thierry Foucault, and Sophie Moinas, 2015, Equilibrium fast trading, *Journal of Financial Economics* 116, 292–313.

Boehmer, Ekkehart, Kingsley Fong, and Julie Wu, 2015, International evidence on algorithmic trading, *Available at SSRN 2022034*.

Boehmer, Ekkehart, Dan Li, and Gideon Saar, 2016, Correlated high-frequency trading, *Working Paper*.

Breckenfelder, Johannes, 2013, Competition between high-frequency traders and market quality, *Available at SSRN 2264858*.

Brogaard, Jonathan, Allen Carrion, Thibaut Moyaert, Ryan Riordan, Andriy Shkilko, and Konstantin Sokolov, 2017, High-frequency trading and extreme price movements, *Journal of Financial Economics* forthcoming.

Brogaard, Jonathan, and Corey Garriott, 2015, High-frequency trading competition, *Working Paper*.

Brogaard, Jonathan, Bjorn Hagstromer, Lars Norden, and Ryan Riordan, 2016, Trading fast and slow: Colocation and liquidity, *Review of Financial Studies* forthcoming.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2014, High-frequency trading and price discovery, *Review of Financial Studies* 27, 2267–2306.

——— , 2016, Price discovery without trading: Evidence from limit orders, *Available at SSRN 2655927*.

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

Budish, Eric, Peter Cramton, and John Shim, 2015, The high-frequency trading arms race: Frequent batch auctions as a market design response, *The Quarterly Journal of Economics* 130, 1547–1621.

Chao, Yong, Chen Yao, and Mao Ye, 2016, What drives price dispersion and market fragmentation across u.s. stock exchanges?, *Working Paper*.

Chordia, Tarun, T. Clifton Green, and Badrinath Kottimukkalur, 2016, Do high frequency traders need to be regulated? evidence from trading on macroeconomic announcements, *Working Paper*.

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

Clark-Joseph, Adam, 2013, Exploratory trading, *Working Paper*.

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 Tālis J Putniņš, 2015, Dark trading and price discovery, *Journal of Financial Economics* 118, 70–92.

Conrad, Jennifer, and Sunil Wahal, 2016, The term structure of liquidity provision, Discussion paper, Working Paper.

——— , and Jin Xiang, 2015, High-frequency quoting, trading, and the efficiency of prices, *Journal of Financial Economics* 116, 271–291.

Egginton, Jared F, Bonnie F Van Ness, and Robert A Van Ness, 2016, Quote stuffing, *Financial Management* 45, 583–608.

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

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of political economy* 81, 607–636.

Foucault, Thierry, Johan Hombert, and Ioanid Rosu, 2016, News trading and speed, *The Journal of Finance* 71, 335–382.

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.

Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71–100.

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.

Griffith, Todd G, Bonnie F Van Ness, and Robert A Van Ness, 2017, High-frequency trading patterns around short-term volatility spikes, *The Journal of Trading* 12, 48–68.

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, 1988, Trades, quotes, inventories, and information, *Journal of Financial Economics* 22, 229–252.

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

——— , 2010, The best bid and offer: A short note on programs and practices, *Working Paper.*

——— , 2015, High frequency quoting: Short-term volatility in bids and offers, *Working paper, New York University.*

——— , and Thomas SY Ho, 1987, Order arrival, quote behavior, and the return-generating process, *The Journal of Finance* 42, 1035–1048.

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

——— , 2013, Low-latency trading, *Journal of Financial Markets* 16, 646–679.

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, 2013, Algorithmic trading and the market for liquidity, *Journal of Financial and Quantitative Analysis* pp. 1001–1024.

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

Hope, Bradley, 2014, Inside hudson river, the firm that does 5% of all stock trading, https://blogs.wsj.com/moneybeat/2014/10/15/inside-hudson-river-the-firm-that-does-5-of-all-stock-trading.

ITG, 2013, Dark liquidity aggregation state of play, http://www.itg.com/2012/11/19/dark-liquidity-aggregation-state-of-play/print/ Accessed: 2/4/2013.

Jovanovic, Boyan, and Albert J. Menkveld, 2015, Middlemen in limit order markets, *Working paper*.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2017, The flash crash: high frequency trading in an electronic market, *Journal of Finance*.

Korajczyk, Robert A., and Dermot Murphy, 2017, High-frequency market making to large institutional trades, *Working Paper*.

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

Lewis, Michael, 2014, *Flash boys: a wall street revolt* (WW Norton & Company).

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

Madrigal, Vicente, 1996, Non-fundamental speculation, *The Journal of Finance* 51, 553–578.

Maggio, Marco Di, Francesco Franzoni, Amir Kermani, and Carlo Sommavilla, 2017, The relevance of broker networks for information diffusion in the stock market, Discussion paper, National Bureau of Economic Research.

Menkveld, Albert J., 2013, High frequency trading and the new-market makers, *Journal of Financial Markets*.

Menkveld, Albert J, and Marius A Zoican, 2017, Need for speed? exchange latency and liquidity, *Review of Financial Studies* forthcoming.

NASDAQ 10K, 2009, Our firm: Getco, http://www.getcollc.com/index.php/getco/tertiary/our_firm/ Accessed: 2/22/2017.

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, Gideon Saar, and Zhuo Zhong, 2015, Relative tick size and the trading environment, *Working Paper*.

O'Hara, Maureen, Chen Yao, and Mao Ye, 2014, What's not there: Odd lots and market data, *Journal of Finance* 69, 2199–2236.

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

Reiss, Peter C., and Ingrid M. Werner, 1998, Does risk sharing motivate interdealer trading?, *The Journal of Finance* 53, 1657–1703.

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

SEC, 2010, Self-regulatory organizations; financial industry regulatory authority, inc.; order granting approval of proposed rule change relating to clearly erroneous transactions, https://www.sec.gov/rules/sro/finra/2010/34-62885.pdf Accessed: 3/20/2017.

Stiglitz, Joseph E, 2014, Tapping the brakes: Are less active markets safer and better for the economy?, in *Federal Reserve Bank of Atlanta 2014 Financial Markets Conference Tuning Financial Regulation for Stability and Efficiency, April* vol. 15.

Stoll, Hans R., 1978, The supply of dealer services in securities markets, *The Journal of Finance* 33, 1133–1151.

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.

Traders Magazine, 2013, New goldman sachs algo takes aim at small caps, http://www.tradersmagazine.com/news/goldman-sachs-algorithm-small-109890-1.html Accessed: 2/5/2013.

Van Kervel, Vincent, and Albert J. Menkveld, 2017, High-frequency trading around large institutional orders, *Journal of Finance* forthcoming.

Vayanos, Dimitri, 2001, Strategic trading in a dynamic noisy market, *The Journal of Finance* 56, 131–171.

Von Beschwitz, Bastian, Donald B Keim, and Massimo Massa, 2015, First to'read'the news: News analytics and institutional trading, .

Weller, Brian M., 2017, Does algorithmic trading deter information acquisition?, *Review of Financial Studies* forthcoming.

Yang, Liyan, and Haoxiang Zhu, 2017, Back-running: seeking and hiding fundamental information in order flows, *Available at SSRN 2583915*.

Yao, Chen, and Mao Ye, 2015, Why trading speed matters: A tale of queue rationing under price controls, .

Ye, Mao, Chen Yao, and Jiading Gai, 2013, The externalities of high-frequency trading, *Available at SSRN 2066839*.

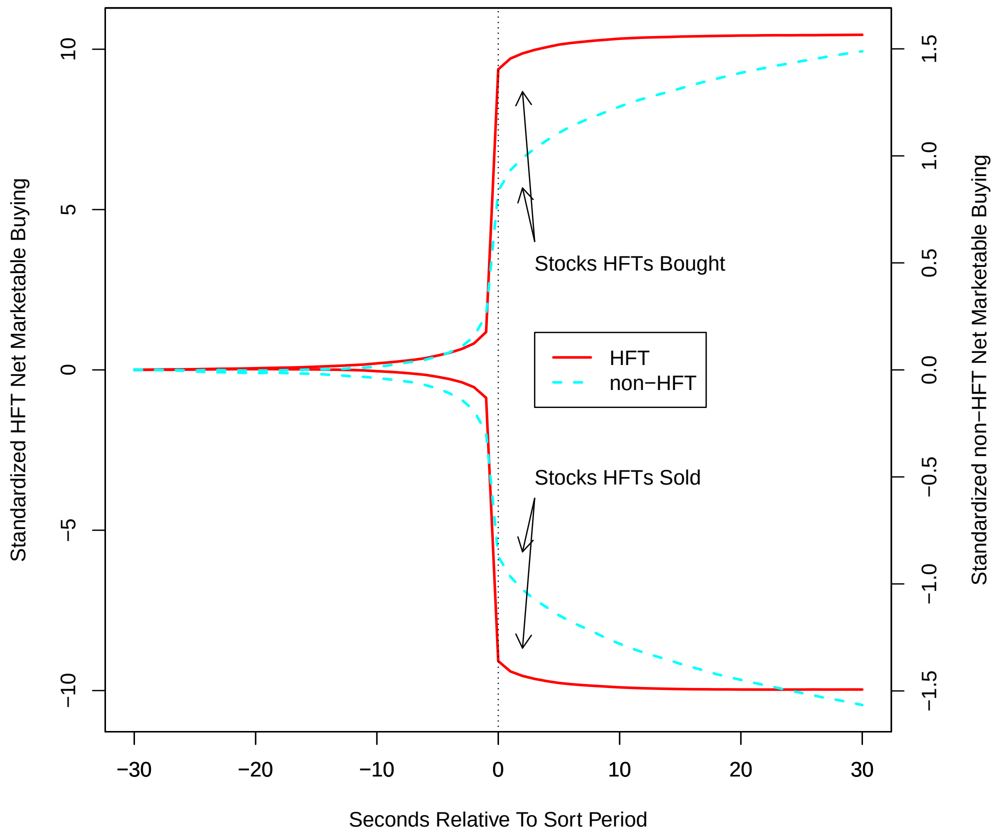

Figure 1: Cumulative HFT vs. non-HFT Net Marketable Buying.
This figure plots cumulative standardized net marketable buying for stocks sorted into portfolios by HFTs' net marketable buying at the one-second horizon. The focus is whether the sort forecasts changes in cumulative non-HFT imbalances after the sort period. The time t cumulative imbalance is the time $t-1$ cumulative imbalance plus the time t imbalance. The left y-axis is for HFT net marketable buying, $\boldsymbol{HFT_{NMB}}$, and the right y-axis is for non-HFT net marketable buying, $\boldsymbol{non\text{-}HFT_{NMB}}$. Stocks are sorted into deciles each second based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten are marked as those HFTs are buying and those in decile one are marked as those HFTs are selling. Portfolio means are calculated by first averaging across all observations for a day and then averaging across all days in the sample. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. Both variables in the figure are divided by their respective standard deviation among all sample stocks that day.

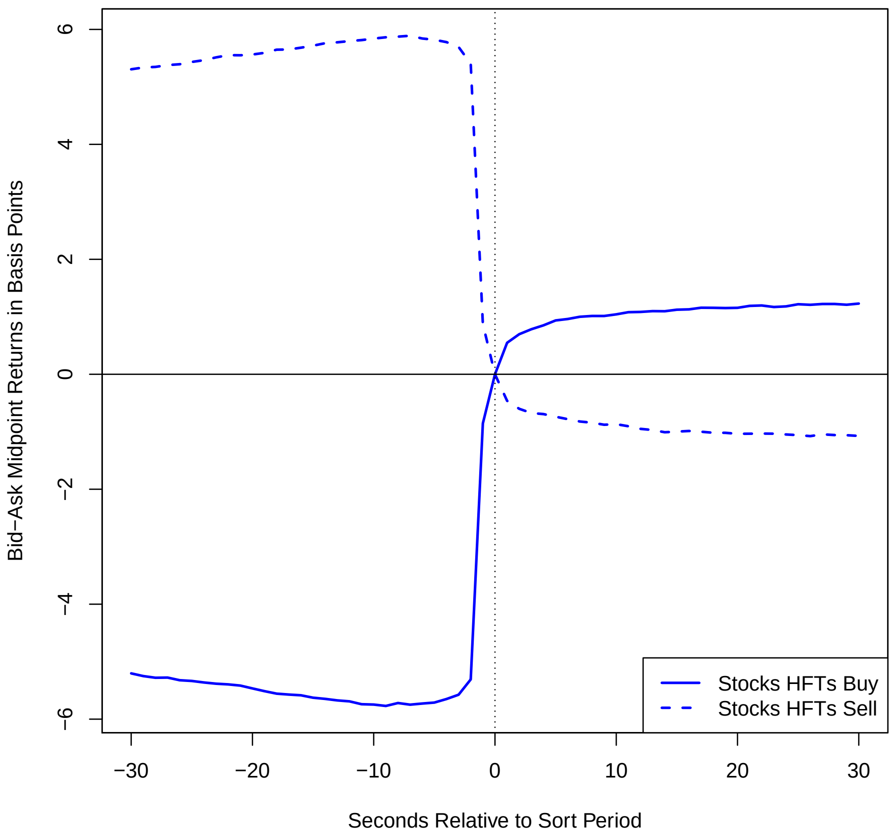

Figure 2: Bid-Ask Midpoint Returns
This figure plots NBBO bid-ask midpoint returns in basis points for stocks sorted into portfolios by HFTs' net marketable buying. Stocks are sorted into deciles each second based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten are marked as those HFTs are buying and those in decile one are marked as those HFTs are selling. Portfolio means are calculated by first averaging across all observations for a day and then averaging across all days in the sample. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume.

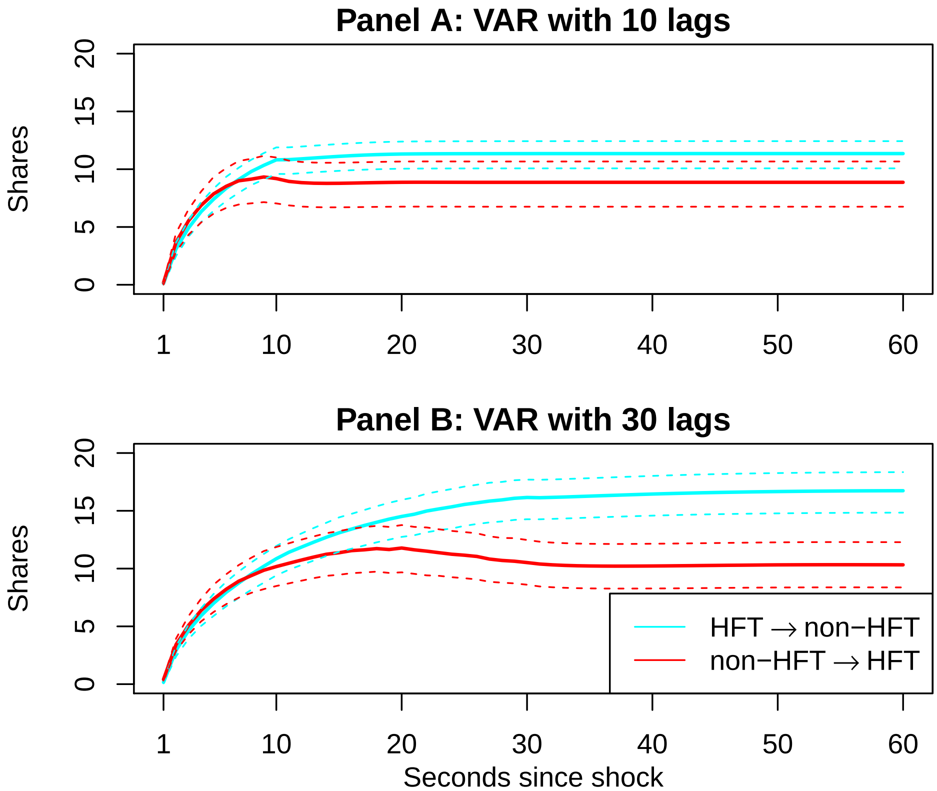

Figure 3: Response of non-HFT and HFT Net Marketable Buying to a 100 Share Shock to the Other Trader Type
This figure plots the impulse response function of non-HFT net marketable buying to a 100 share shock to HFT net marketable buying, and then the response of HFT net marketable buying to a 100 share shock to non-HFT net marketable buying. The solid line is the estimated response function, and dotted lines indicate 95% confidence intervals. Estimates come from the VAR in Table 4:

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

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

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

where R_t is the one-second NBBO bid-ask midpoint return, $HFT_{NMB,t}$ is one-second HFT net marketable buying, $non\text{-}HFT_{NMB,t}$ is one-second non-HFT net marketable buying, and J is the number of lags. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. All variables are divided by their standard deviation among all stocks that day. The equations are estimated for each stock every day. Coefficients are then averaged across all stocks for a day, with stocks weighted by the minimum number of non-zero observations among the three variables. The full-sample estimate is the mean of that daily coefficient time-series. Standard errors correct for autocorrelation using the Newey and West (1994) optimal lag selection method. The impulse response functions assume no contemporaneous effects among variables, so the ordering of the variables does not matter.

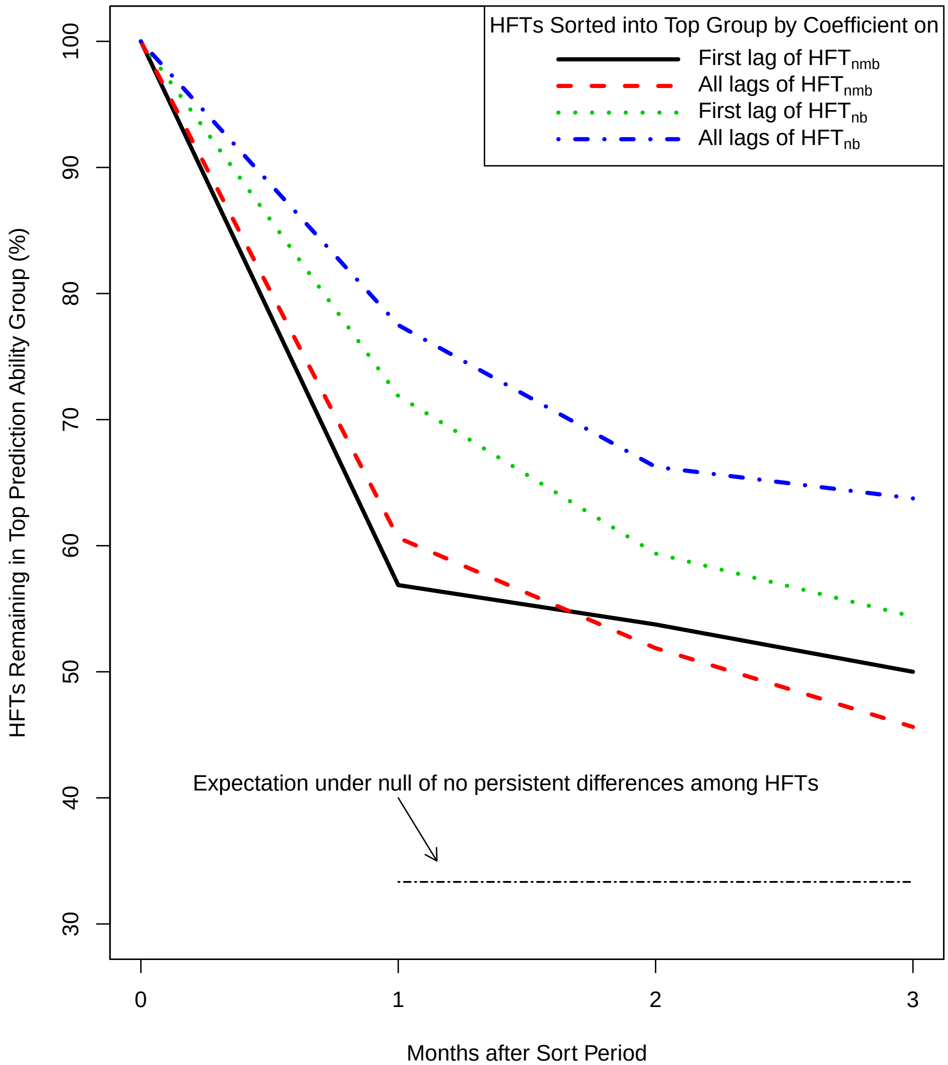

Figure 4: Persistence of Differences in Individual HFTs' Prediction Ability
Given an HFT is among the top third of HFTs in terms of how their trades predict non-HFT order flow, the figure plots the probability they will still be in the top third one, two, and three months later. The dotted line at 33.3% 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. The four lines indicate different sorting methods discussed in Table 5. $\boldsymbol{HFT_{NB}}$ is HFT net buying. $\boldsymbol{HFT_{NMB}}$ is the HFT's net marketable buying with the slight adjustment that if one of either the HFT's net marketable buying or net buying is positive and the other is negative, then its net marketable buying is set to zero (see footnote 35 on page 22 for details). Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. Net buying is shares bought minus shares sold. The probabilities 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.

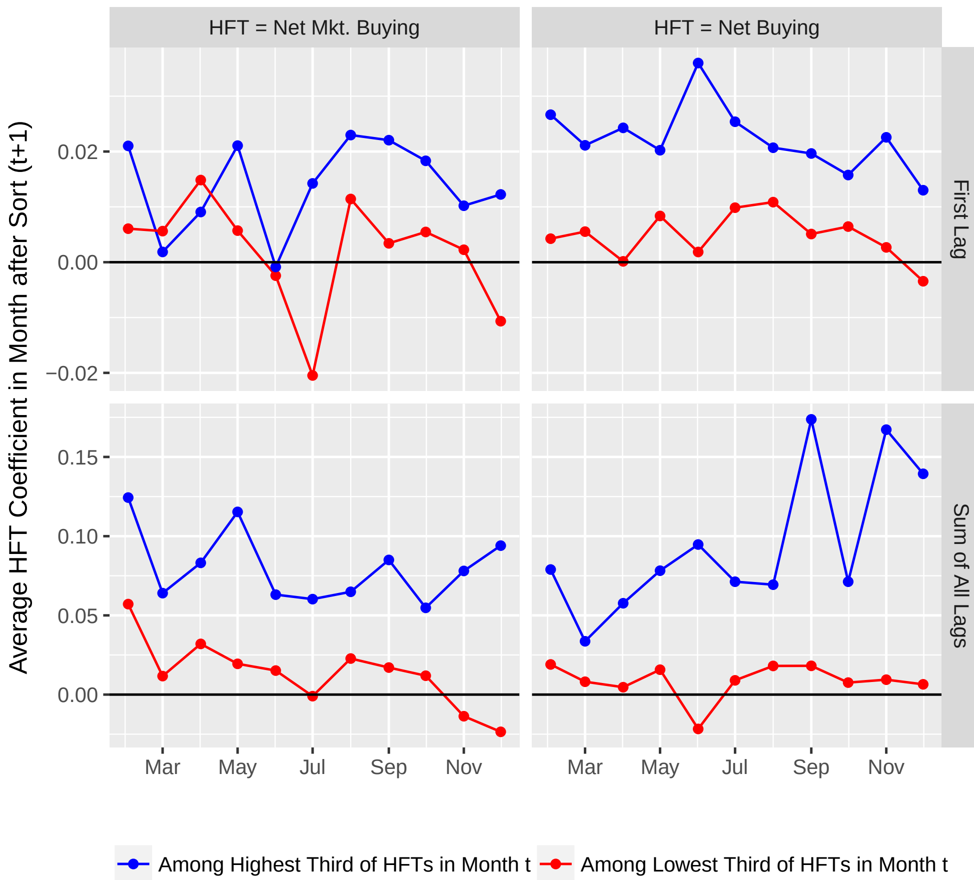

Figure 5: Time Series of Differences in Individual HFTs' Prediction Ability
This test examines whether the HFTs who predict order flow the best one month continue to do so the next month. HFTs are sorted into three groups in month t based on coefficients from regressions of non-HFT trading on lagged HFT trading, where HFT trading is either the HFT's net marketable buying or net buying (including all marketable and non-marketable trades). The figure then shows the average coefficient on HFT trading in month $t+1$ for the top and bottom HFT groups. More precisely, the following regression is run each day for each HFT:

$$\textit{non-HFT}_{d,s,t} = \alpha_{d,i} + \sum_{l=1}^{10} \gamma_{d,i,l} \textit{HFT}_{d,s,i,t-l} + \sum_{l=1}^{10} \beta_{d,i,l} \textit{non-HFT}_{d,s,t-l} + \sum_{i=1}^{10} \lambda_{d,i,l} R_{d,s,t-l} + \epsilon_{d,i,s,t}, \tag{4}$$

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 buying ($\boldsymbol{HFT_{NB}}$) or net marketable buying ($\boldsymbol{HFT_{NMB}}$), and R is the stock's NBBO bid-ask midpoint return. The firm-level HFT net marketable buying is slightly adjusted such that if one of either the HFT's net marketable buying or net buying is positive and the other is negative, then its net marketable buying is set to zero (see footnote 35 on page 22 for details). 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. HFTs are then sorted into three groups based on either their average $\overline{\gamma_{d,i,1}}$ or $\overline{\sum_{l=1}^{10} \gamma_{d,i,l}}$ for the month. 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). Table 5 contains additional details.

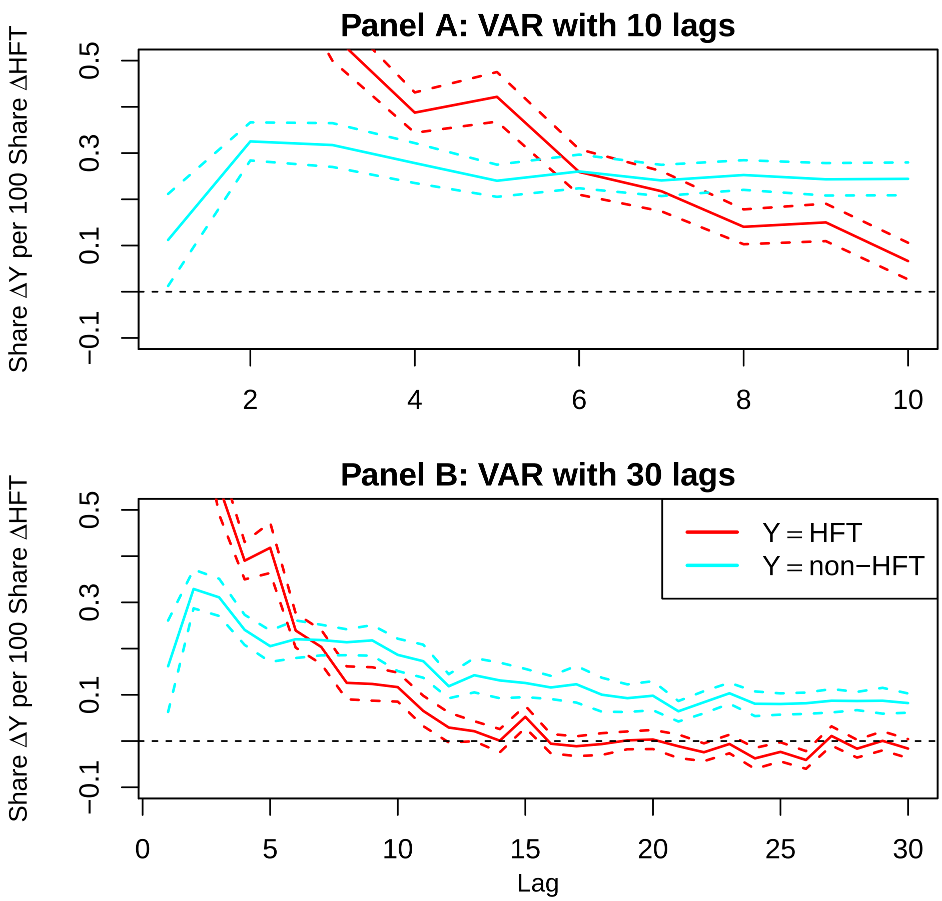

Figure 6: Does HFT net marketable buying predict itself or non-HFT net marketable buying better?

Comparison of HFT coefficients ($\gamma_{2,i}$ and γ_{3_i} below) from regressions where HFT and non-HFT trading are the dependent variables. The solid line depicts the estimate and dotted lines show 95% confidence intervals. To make the comparison in the same units, the standardized coefficients are converted to show the share change in the dependent variable for a 100 share change in the HFT variable. The conversion uses the mean standard deviation of the variables in Table 1. Y-axis is cut off above **0.5** to zoom in on higher lags. The regressions are equations 6 and 7 from Table 4:

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

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

where R_t is one-second NBBO bid-ask midpoint return, $HFT_{NMB,t}$ is one-second HFT net marketable buying, $non\text{-}HFT_{NMB,t}$ is one-second non-HFT net marketable buying, and J is the number of lags. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. All variables are divided by their standard deviation among all stocks that day. The equations are estimated for each stock every day. Coefficients are then averaged across all stocks for a day, with stocks weighted by the minimum number of non-zero observations among the three variables. The full-sample estimate is the mean of that daily coefficient time-series. Standard errors correct for autocorrelation using the Newey and West (1994) optimal lag selection method.

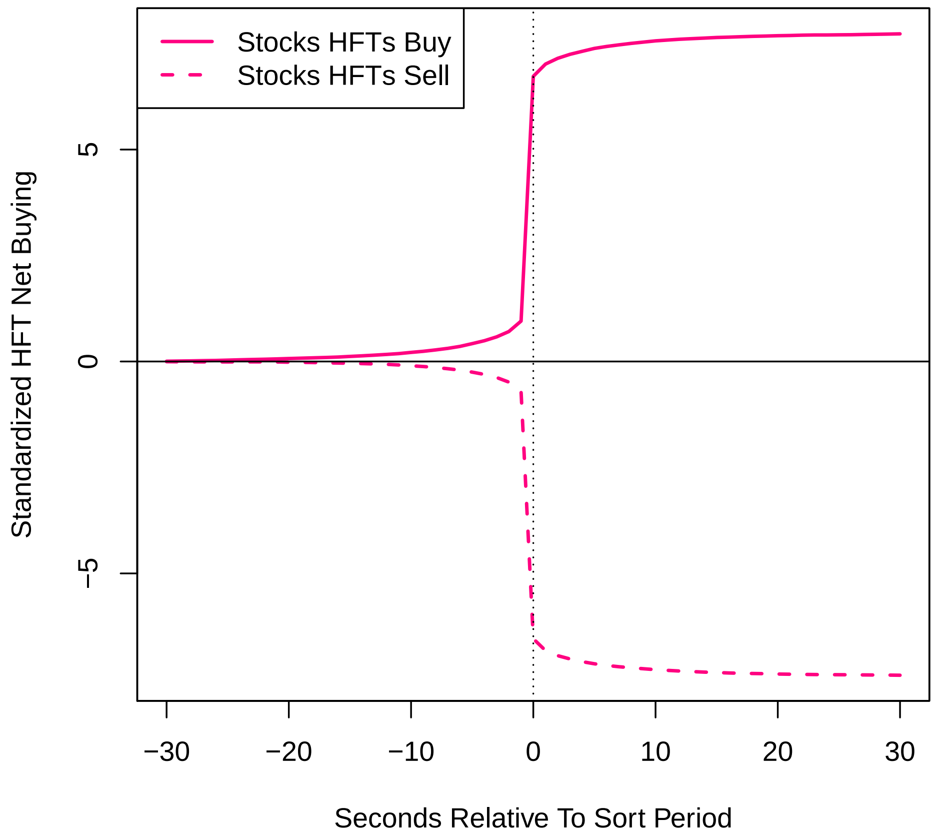

Figure 7: HFT Net Buying around the HFT Net Marketable Buying Sorts
The figure examines cumulative standardized HFT net buying from 30 seconds before to 30 seconds after sorts on HFT net marketable buying. The time t cumulative imbalance is the time $t-1$ cumulative imbalance plus the time t imbalance. Stocks are sorted into deciles each second based on HFT net marketable buying. Decile breakpoints are calculated from non-zero observations during the prior trading day. Stocks in decile ten are marked as those HFTs are buying and those in decile one are marked as those HFTs are selling. Portfolio means are calculated by first averaging across all observations for a day and then averaging across all days in the sample. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. Net buying is shares bought minus shares sold. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. HFT net buying is divided by its standard deviation among all sample stocks that day to ease interpretation.

Table 1
Summary Statistics

The sample consists of 96 stocks chosen by randomly selecting six NASDAQ-listed and six NYSE-listed stocks from each of the top eight NYSE size deciles of CRSP common stocks passing liquidity filters described in Section 2. Size portfolio breakpoints are computed among NYSE-listed stocks. Size portfolios for year t are formed on December 31$^{\text{st}}$ of year $t-1$. Deciles one through five are small-cap, six through eight are mid-cap, and nine through ten are large-cap. Panel A summarizes 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 summarizes 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 (*BAM Ret*), HFTs' net buying ($\boldsymbol{HFT_{NB}}$), HFTs' net marketable buying ($\boldsymbol{HFT_{NMB}}$), and non-HFTs' net marketable buying (***non-***$\boldsymbol{HFT_{NMB}}$). 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 this table only, imbalance measures are expressed in shares.

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

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

Table 2

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

This table shows standardized 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 breakpoints are calculated from non-zero observations during the prior trading day. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. Non-HFT net marketable buying portfolios are averaged across all observations for a day, and the mean of the daily time series is reported in the table. Parentheses indicate t-statistics for the time-series means, with autocorrelation corrections following the optimal automatic lag selection method of Newey and West (1994). The reported imbalance is divided by its standard deviation among all stocks that day to ease interpretation.

Decile	**Seconds**					
	$[t-30,t-1]$	$t-1$	t	$t+1$	$[t+1,t+30]$	$[t+1,t+300]$
All Stocks						
10 (HFT Buying)	0.26	0.10	0.58	0.10	0.65	1.15
	(4.39)	(22.01)	(24.92)	(26.17)	(14.35)	(2.41)
9	0.12	0.04	0.22	0.03	0.26	0.50
	(10.35)	(25.24)	(24.86)	(21.91)	(23.47)	(7.28)
2	−0.11	−0.04	−0.22	−0.03	−0.26	−0.62
	(−10.69)	(−24.92)	(−26.79)	(−29.18)	(−25.85)	(−8.59)
1 (HFT Selling)	−0.29	−0.10	−0.58	−0.09	−0.69	−1.89
	(−4.52)	(−20.24)	(−32.28)	(−19.66)	(−10.15)	(−4.26)
Small-cap						
10 (HFT Buying)	1.09	0.25	1.09	0.21	1.39	2.42
	(5.49)	(18.14)	(20.60)	(13.91)	(9.68)	(1.54)
1 (HFT Selling)	−1.04	−0.26	−1.07	−0.19	−1.27	−3.27
	(−4.68)	(−11.71)	(−27.77)	(−10.16)	(−5.40)	(−2.27)
Mid-cap						
10 (HFT Buying)	0.22	0.10	0.62	0.09	0.61	1.20
	(4.70)	(19.20)	(22.82)	(23.48)	(15.92)	(4.65)
1 (HFT Selling)	−0.30	−0.10	−0.63	−0.10	−0.69	−1.83
	(−5.64)	(−22.49)	(−24.15)	(−20.15)	(−14.24)	(−5.75)
Large-cap						
10 (HFT Buying)	−0.02	0.03	0.31	0.06	0.39	0.89
	(−1.09)	(9.52)	(24.51)	(20.37)	(18.59)	(6.91)
1 (HFT Selling)	0.05	−0.03	−0.32	−0.05	−0.39	−0.96
	(2.84)	(−15.78)	(−26.57)	(−18.14)	(−23.93)	(−8.55)

Table 3

NBBO Bid-Ask Midpoint Returns for Stocks Sorted by HFT Net Marketable Buying

This table shows NBBO bid-ask midpoint 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 imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. NBBO bid-ask midpoint returns are averaged across all observations for a day, and the mean of the daily time series is reported in the table. Parentheses indicate t-statistics for the time-series means, with autocorrelation corrections following the optimal automatic lag selection method of Newey and West (1994).

Decile	**Seconds**					
	$[t-30, t-1]$	$t-1$	t	$t+1$	$[t+1, t+30]$	$[t+1, t+300]$
All Stocks						
10 (HFT Buying)	4.43	4.51	0.92	0.55	1.23	0.57
	(25.26)	(25.49)	(15.91)	(17.98)	(11.78)	(1.16)
9	3.24	3.20	0.65	0.47	0.70	0.55
	(29.04)	(24.28)	(9.41)	(14.53)	(14.03)	(3.63)
2	−3.04	−3.11	−0.51	−0.34	−0.51	0.05
	(−30.62)	(−27.98)	(−8.95)	(−11.07)	(−7.42)	(0.30)
1 (HFT Selling)	−4.30	−4.47	−0.81	−0.46	−1.07	−0.58
	(−26.30)	(−27.66)	(−14.28)	(−14.58)	(−13.48)	(−1.42)
Small-cap						
10 (HFT Buying)	7.98	6.63	1.31	0.61	2.54	1.36
	(19.77)	(19.64)	(13.75)	(11.48)	(8.41)	(0.86)
1 (HFT Selling)	−7.59	−6.55	−1.19	−0.57	−2.64	−3.09
	(−15.25)	(−21.68)	(−15.07)	(−13.08)	(−14.37)	(−2.58)
Mid-cap						
10 (HFT Buying)	4.46	4.61	1.01	0.59	1.52	1.14
	(23.64)	(27.20)	(14.01)	(13.07)	(16.18)	(4.19)
1 (HFT Selling)	−4.35	−4.64	−0.81	−0.51	−1.28	−0.38
	(−30.73)	(−30.33)	(−12.86)	(−10.66)	(−13.06)	(−1.47)
Large-cap						
10 (HFT Buying)	2.93	3.48	0.64	0.47	0.40	0.11
	(33.48)	(29.45)	(14.42)	(10.80)	(5.20)	(0.64)
1 (HFT Selling)	−2.76	−3.40	−0.64	−0.36	−0.15	0.64
	(−27.66)	(−31.04)	(−10.24)	(−7.57)	(−2.29)	(4.31)

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

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

$$R_t = \alpha_1 + \sum_{i=1}^{10} \gamma_{1,i} HFT_{NMB,t-i} + \sum_{i=1}^{10} \beta_{1,i} non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_{1,i} R_{t-i} + \epsilon_{1,t} \quad (5)$$

$$HFT_{NMB,t} = \alpha_2 + \sum_{i=1}^{10} \gamma_{2,i} HFT_{NMB,t-i} + \sum_{i=1}^{10} \beta_{2,i} non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_{2,i} R_{t-i} + \epsilon_{2,t} \quad (6)$$

$$non\text{-}HFT_{NMB,t} = \alpha_3 + \sum_{i=1}^{10} \gamma_{3,i} HFT_{NMB,t-i} + \sum_{i=1}^{10} \beta_{3,i} non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_{3,i} R_{t-i} + \epsilon_{3,t} \quad (7)$$

where $\boldsymbol{R_t}$ is the one-second NBBO bid-ask midpoint return, $\boldsymbol{HFT_{NMB,t}}$ is one-second HFT net marketable buying, and $\boldsymbol{non\text{-}HFT_{NMB,t}}$ is one-second non-HFT net marketable buying. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. All variables are divided by their standard deviation among all stocks that day. The equations are estimated for each stock every day. In Panel A, coefficients are averaged across all stocks for a day, with stocks weighted by the minimum number of non-zero observations among the three variables. The full-sample estimate is the mean of that daily coefficient time-series. Standard errors correct for autocorrelation using the Newey and West (1994) optimal lag selection method. Panel B reports the average coefficients and percent of stock days with positive and negative coefficients significantly different from zero at the five percent confidence level in the stock-day regression.

Table 4 — continued

Panel A: 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	0.0007	2.21	0.0754	37.32	0.9181	13.68
2	0.0021	15.47	0.0246	39.42	−0.0145	−4.28
3	0.0020	13.12	0.0184	38.72	−0.0008	−0.53
4	0.0018	12.62	0.0153	33.37	−0.0028	−1.95
5	0.0015	13.52	0.0145	38.66	−0.0032	−2.41
6	0.0017	14.00	0.0111	25.98	0.0047	4.68
7	0.0015	13.97	0.0093	33.33	0.0016	1.54
8	0.0016	15.41	0.0084	28.33	0.0005	0.73
9	0.0016	13.59	0.0089	30.21	−0.0002	−0.22
10	0.0016	13.48	0.0117	30.19	−0.0025	−2.59
$y = HFT_t$						
1	0.0243	25.55	0.0028	2.64	1.9956	10.19
2	0.0088	24.58	0.0007	1.49	0.0290	4.79
3	0.0056	19.07	−0.0001	−0.21	−0.0047	−2.16
4	0.0039	17.41	−0.0008	−2.32	−0.0124	−6.52
5	0.0042	15.42	−0.0013	−3.34	−0.0131	−7.84
6	0.0026	10.34	−0.0011	−3.11	−0.0125	−7.26
7	0.0022	9.79	−0.0014	−3.98	−0.0117	−7.40
8	0.0014	7.30	−0.0007	−2.35	−0.0127	−9.49
9	0.0015	7.26	−0.0012	−3.34	−0.0148	−8.26
10	0.0007	3.28	−0.0019	−5.56	−0.0121	−7.83
$y = R_t$						
1	0.0185	10.92	0.0269	14.82	−0.1551	−55.30
2	0.0115	11.82	0.0155	12.64	−0.1310	−63.87
3	0.0085	11.93	0.0119	9.85	−0.1070	−68.61
4	0.0067	9.60	0.0102	11.30	−0.0906	−66.86
5	0.0056	9.94	0.0078	10.86	−0.0749	−71.81
6	0.0042	10.60	0.0066	11.69	−0.0645	−72.24
7	0.0035	7.67	0.0037	6.65	−0.0522	−70.93
8	0.0023	10.21	0.0041	10.01	−0.0420	−77.02
9	0.0017	5.68	0.0018	3.24	−0.0308	−77.50
10	−0.0001	−0.38	−0.0003	−0.73	−0.0188	−67.19

Table 4 — continued

Panel B: Summary of stock-day coefficient estimates

lag	γ (HFT)			β (non-HFT)			λ (R)		
	μ	% +	% -	μ	% +	% -	μ	% +	% -
$y = non\text{-}HFT_t$									
1	0.0009	24.5	18.4	0.0738	71.0	5.1	0.8454	90.2	0.7
2	0.0022	18.5	9.3	0.0251	49.7	7.4	−0.0127	30.4	13.8
3	0.0021	15.2	7.5	0.0185	39.5	6.3	−0.0006	18.3	9.9
4	0.0019	13.1	7.1	0.0154	34.7	6.7	−0.0020	15.3	8.1
5	0.0016	12.6	7.0	0.0145	33.4	5.9	−0.0026	12.3	7.9
6	0.0017	11.7	6.1	0.0114	28.6	6.4	0.0045	12.1	6.6
7	0.0016	10.6	6.3	0.0094	25.2	6.4	0.0018	9.8	7.0
8	0.0017	10.7	6.0	0.0087	24.0	7.2	0.0008	8.8	6.3
9	0.0016	10.6	5.5	0.0090	23.8	6.2	0.0001	7.9	6.6
10	0.0016	10.4	5.8	0.0119	28.2	5.6	−0.0022	7.2	6.2
$y = HFT_t$									
1	0.0255	50.3	4.5	0.0033	22.6	14.3	1.8224	86.7	2.0
2	0.0090	26.4	4.7	0.0005	13.1	9.0	0.0257	28.2	9.0
3	0.0055	18.7	4.1	−0.0004	10.1	8.0	−0.0053	15.1	7.9
4	0.0037	15.5	5.0	−0.0009	8.9	7.5	−0.0119	11.2	7.1
5	0.0041	15.1	4.2	−0.0011	8.4	7.4	−0.0126	9.0	7.0
6	0.0025	11.7	5.3	−0.0011	7.8	7.1	−0.0122	7.4	7.0
7	0.0021	10.3	4.8	−0.0013	7.0	6.8	−0.0110	6.7	6.6
8	0.0013	9.2	5.5	−0.0007	7.0	6.6	−0.0120	5.6	6.1
9	0.0014	9.1	5.2	−0.0011	6.7	6.5	−0.0136	5.1	5.8
10	0.0005	9.1	5.9	−0.0018	7.0	6.7	−0.0112	4.7	5.8
$y = R_t$									
1	0.0206	42.6	3.5	0.0290	43.5	3.4	−0.1566	20.0	65.6
2	0.0127	29.9	3.3	0.0167	29.9	4.0	−0.1331	10.9	63.6
3	0.0095	23.0	4.0	0.0130	23.3	4.6	−0.1089	9.8	60.5
4	0.0075	18.5	4.7	0.0110	19.4	4.8	−0.0921	11.9	54.8
5	0.0063	15.1	4.5	0.0084	16.4	4.7	−0.0760	11.1	53.6
6	0.0047	12.5	4.5	0.0070	13.3	4.9	−0.0653	11.6	50.8
7	0.0039	10.3	4.4	0.0038	11.5	5.3	−0.0528	10.8	48.9
8	0.0025	9.2	4.6	0.0044	10.7	4.9	−0.0423	12.2	45.9
9	0.0018	8.6	5.0	0.0021	11.5	5.2	−0.0310	12.0	44.8
10	−0.0001	6.7	4.8	−0.0003	8.9	5.0	−0.0188	12.9	44.9

Table 5
Examining Differences in Individual HFTs' Prediction Ability

This table tests whether trades from 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:

$$non\text{-}HFT_{d,s,t} = \alpha_{d,i} + \sum_{l=1}^{10} \gamma_{d,i,l} HFT_{d,s,i,t-l} + \sum_{l=1}^{10} \beta_{d,i,l} non\text{-}HFT_{d,s,t-l} + \sum_{i=1}^{10} \lambda_{d,i,l} R_{d,s,t-l} + \epsilon_{d,i,s,t}, \tag{8}$$

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 buying ($\boldsymbol{HFT_{NB}}$) or net marketable buying ($\boldsymbol{HFT_{NMB}}$), and R is the stock's NBBO bid-ask midpoint return. The firm-level HFT net marketable buying is slightly adjusted such that if one of either the HFT's net marketable buying or net buying is positive and the other is negative, then its net marketable buying is set to zero (see footnote 35 on page 22 for details). 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. 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 set of columns in the table below report results from regressions where *HFT* is HFTs' net marketable buying, and the right set report results where *HFT* is their net buying. For regression (8), 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 (8) 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). For each group's monthly time series, the table reports the time-series mean, p-value for the test that the mean is different from zero using standard errors corrected for autocorrelation following the Newey and West (1994) optimal lag selection method, time-series median, number of months out of 11 that the value is greater than zero, and p-value from a non-parametric Wilcoxon rank sum test that the time-series median is equal to zero. I report p-values for the t-test because the small number of months means standard rules of thumb (e.g., t-stat > 1.96) for determining statistical significance 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}	p-value	median	#>0 of 11	Wilcoxon p-value	μ_{t+1}	p-value	median	#>0 of 11	Wilcoxon p-value
$\overline{\gamma_{d,i,1}}$										
H in month$_t$	0.014	< 0.01	0.014	10	< 0.01	0.022	< 0.01	0.021	11	< 0.01
M in month$_t$	0.011	< 0.01	0.009	11	< 0.01	0.010	< 0.01	0.010	10	0.01
L in month$_t$	0.002	0.51	0.005	8	0.32	0.005	0.01	0.005	10	0.01
High−Low	0.012	< 0.01	0.013	9	0.01	0.017	< 0.01	0.016	11	< 0.01
$\overline{\sum_{l=1}^{10} \gamma_{d,i,l}}$										
H in month$_t$	0.080	< 0.01	0.078	11	< 0.01	0.098	< 0.01	0.078	11	< 0.01
M in month$_t$	0.034	< 0.01	0.033	11	< 0.01	0.033	< 0.01	0.033	11	< 0.01
L in month$_t$	0.012	0.15	0.015	8	0.10	0.009	0.01	0.009	10	0.05
High−Low	0.068	< 0.01	0.061	11	< 0.01	0.089	< 0.01	0.063	11	< 0.01

Table 6

Differences among HFTs in How Strongly Their Marketable 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 NBBO bid-ask midpoint returns. HFTs are split into two groups each month: HFTs who are above the median in terms of the correlation between their marketable 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 8 in Table 5 estimated during the prior month; for only this regression determining the split, the HFT variable is the HFT's net marketable buying with the slight adjustment that if one of either the HFT's net marketable buying or net buying is positive and the other is negative, then its net marketable buying is set to zero (see footnote 35 on page 22 for details). Group-level HFT net marketable buying in Regression 9 below is unadjusted. Table 5 provides more detail on the methodology for sorting HFTs. Once above/below median HFTs are determined, all their 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:

$$R_t = \alpha + \sum_{i=1}^{10} \gamma_i HFT^G_{NMB,t-i} + \sum_{i=1}^{10} \beta_i non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_i R_{t-i} + \epsilon_{t,R} \tag{9}$$

where $\boldsymbol{R_t}$ is the one-second NBBO bid-ask midpoint return, $\boldsymbol{HFT^G_{NMB,t}}$ is one-second HFT net marketable buying for either the above or below-median group, and $\boldsymbol{non\text{-}HFT_{NMB,t}}$ is one-second non-HFT net marketable buying. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. Returns and non-HFT net marketable buying are divided by their respective one-second standard deviations, whereas the above and below-median HFT groups' net marketable buying is divided by the one-second standard deviation of $\boldsymbol{HFT_{NMB}}$ aggregated among all (i.e., not only above or below-median) HFTs. 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 non-parametric Wilcoxon 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	0.0288	11.32	0.0239	0.0207	16.11	0.0172	0.0081	2.84	0.00
2	0.0190	10.38	0.0154	0.0134	14.41	0.0110	0.0055	2.68	0.00
3	0.0142	8.81	0.0111	0.0093	14.07	0.0078	0.0048	2.77	0.00
4	0.0099	14.02	0.0083	0.0075	13.58	0.0065	0.0024	2.73	0.01
5	0.0081	9.31	0.0067	0.0064	13.78	0.0056	0.0017	1.70	0.12
6	0.0063	10.95	0.0057	0.0050	11.78	0.0043	0.0013	1.76	0.02
7	0.0050	8.36	0.0050	0.0043	7.32	0.0033	0.0007	0.83	0.03
8	0.0029	5.37	0.0029	0.0027	8.52	0.0025	0.0002	0.25	0.72
9	0.0025	3.48	0.0019	0.0019	4.29	0.0019	0.0006	0.74	0.65
10	-0.0005	-0.89	-0.0002	-0.0001	-0.32	-0.0001	-0.0004	-0.60	0.70
Σ 1-10	0.0960	12.26	0.0755	0.0712	15.15	0.0549	0.0248	2.72	0.00

Table 7
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,

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

where R_t is the one-second NBBO bid-ask midpoint return, $HFT_{NMB,t}$ is one-second HFT net marketable buying, and $non\text{-}HFT_{NMB,t}$ is one-second non-HFT net marketable buying. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. All variables are divided by their standard deviation among all stocks that day. The equations are estimated for each stock every day. Coefficients are then averaged across all stocks for a day, with stocks weighted by the minimum number of non-zero observations among the three variables. The full-sample estimate is the mean of that daily coefficient time-series. Standard errors correct for autocorrelation using the Newey and West (1994) optimal lag selection method. 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 has high volume or high absolute daily aggregate net marketable buying imbalance if the volume/imbalance is among the top 10% of days the past month (i.e., rank relative to prior 19 trading days is 19 or 20). In Panel C, relative spreads are calculated by dividing the bid-ask spread by the bid-ask midpoint. The daily average spread and relative spread measures are calculated by duration-weighting intra-day observations. For either spread measure, the high/low spread groups consist of stocks in the top/bottom third of the sample the prior day.

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	0.0026	5.70	0.0003	1.48	−0.0050	−11.29	0.0023	4.71	−0.0053	−10.98
2	0.0033	9.23	0.0018	15.02	0.0004	1.30	0.0015	3.92	−0.0014	−4.20
3	0.0023	8.08	0.0018	16.60	0.0009	2.68	0.0006	1.78	−0.0009	−2.63
4	0.0020	7.23	0.0016	15.91	0.0013	4.35	0.0004	1.42	−0.0003	−0.97
5	0.0016	6.32	0.0014	16.89	0.0010	3.72	0.0002	0.65	−0.0004	−1.31
6	0.0021	8.04	0.0012	15.67	0.0016	5.31	0.0009	3.18	0.0003	1.14
7	0.0019	8.06	0.0012	15.07	0.0013	4.39	0.0006	2.59	0.0000	0.11
8	0.0014	6.00	0.0011	14.42	0.0018	5.42	0.0002	0.89	0.0006	1.86
9	0.0016	5.86	0.0011	14.61	0.0018	5.93	0.0004	1.55	0.0007	2.17
10	0.0019	8.03	0.0015	17.40	0.0014	5.33	0.0005	1.85	−0.0001	−0.25
Σ 1-10	0.0207	19.47	0.0131	24.01	0.0065	5.29	0.0076	6.34	−0.0067	−4.99

Table 7 — continued

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

	Volume				Imbalance			
	High	Not High	Difference		High	Not High	Difference	
lag	mean	mean	mean	t-stat	mean	mean	mean	t-stat
1	0.0036	0.0004	0.0033	4.22	0.0019	0.0006	0.0012	1.64
2	0.0039	0.0018	0.0021	4.10	0.0032	0.0020	0.0012	2.30
3	0.0029	0.0019	0.0010	2.32	0.0024	0.0020	0.0004	0.84
4	0.0024	0.0017	0.0007	1.69	0.0022	0.0017	0.0004	1.18
5	0.0020	0.0015	0.0004	1.01	0.0011	0.0016	−0.0005	−0.89
6	0.0019	0.0016	0.0003	0.89	0.0025	0.0016	0.0009	1.57
7	0.0014	0.0015	−0.0001	−0.24	0.0008	0.0015	−0.0007	−0.59
8	0.0017	0.0015	0.0001	0.22	0.0018	0.0016	0.0002	0.41
9	0.0019	0.0015	0.0004	0.95	0.0017	0.0015	0.0002	0.48
10	0.0018	0.0014	0.0004	1.01	0.0018	0.0015	0.0003	0.66
Σ 1-10	0.0235	0.0149	0.0086	4.75	0.0193	0.0156	0.0037	1.53

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	0.0050	−0.0019	0.0069	14.75	0.0042	−0.0003	0.0045	6.83
2	0.0031	0.0015	0.0015	4.91	0.0034	0.0019	0.0016	3.50
3	0.0025	0.0018	0.0007	2.35	0.0029	0.0019	0.0010	2.65
4	0.0018	0.0015	0.0003	1.44	0.0014	0.0018	−0.0004	−1.48
5	0.0019	0.0014	0.0004	1.89	0.0017	0.0015	0.0002	0.74
6	0.0016	0.0017	−0.0001	−0.63	0.0013	0.0018	−0.0005	−2.06
7	0.0016	0.0016	0.0000	0.16	0.0010	0.0016	−0.0006	−2.11
8	0.0013	0.0017	−0.0004	−1.84	0.0012	0.0017	−0.0005	−1.52
9	0.0013	0.0016	−0.0004	−1.70	0.0011	0.0017	−0.0006	−2.05
10	0.0015	0.0017	−0.0002	−0.91	0.0019	0.0016	0.0003	0.74
Σ 1-10	0.0215	0.0127	0.0088	7.96	0.0202	0.0153	0.0049	3.59

Table 8

VAR Estimates on Days with and without News

This table reports coefficients on HFTs' net marketable buying from the VAR equation in Table 4 where the dependent variable is non-HFT net marketable buying, conditional on news days,

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

where R_t is the one-second return, $HFT_{NMB,t}$ is one-second HFT net marketable buying, and $non\text{-}HFT_{NMB,t}$ is one-second non-HFT net marketable buying. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. All variables are divided by their standard deviation among all stocks that day. The equations are estimated for each stock every day. Coefficients are then averaged across all stocks for a day, with stocks weighted by the minimum number of non-zero observations among the three variables. The full-sample estimate is the mean of that daily coefficient time-series. Standard errors correct for autocorrelation using the Newey and West (1994) optimal lag selection method. The difference column group also reports the p-value from a non-parametric Wilcoxon rank sum test that the time-series medians are equal. See Section 2.5 on page 13 for a description of the Factiva data.

Panel A: News day defined as a day with a Factiva article for the stock

lag	News days			Non-news days			Difference		
	mean	t-stat	median	mean	t-stat	median	mean	t-stat	rank sum p-value
1	0.0005	1.93	0.0003	0.0019	4.95	0.0016	-0.0014	-3.18	0.00
2	0.0021	13.88	0.0019	0.0020	7.36	0.0016	0.0001	0.44	0.18
3	0.0021	13.01	0.0020	0.0019	9.95	0.0017	0.0003	1.04	0.18
4	0.0018	14.64	0.0017	0.0016	7.64	0.0014	0.0002	0.95	0.03
5	0.0016	12.42	0.0015	0.0014	5.91	0.0010	0.0002	0.61	0.04
6	0.0017	13.31	0.0016	0.0013	6.06	0.0012	0.0004	1.78	0.07
7	0.0015	12.79	0.0015	0.0015	6.81	0.0010	0.0000	0.18	0.09
8	0.0017	13.77	0.0018	0.0012	6.35	0.0009	0.0005	2.12	0.00
9	0.0016	12.02	0.0013	0.0014	7.40	0.0013	0.0002	0.79	0.26
10	0.0016	12.11	0.0015	0.0014	7.93	0.0009	0.0003	1.17	0.00
Σ 1-10	0.0163	26.21	0.0158	0.0155	16.38	0.0144	0.0008	0.70	0.27

Table 8 — continued

Panel B: News day defined as a day with |market-adjusted return| > 1%

lag	$\lvert return \rvert > 1\%$			$\lvert return \rvert \leq 1\%$			Difference		
	mean	t-stat	median	mean	t-stat	median	mean	t-stat	rank sum p-value
1	0.0014	5.01	0.0014	-0.0001	-0.20	-0.0000	0.0014	3.65	0.00
2	0.0022	12.90	0.0021	0.0019	10.26	0.0017	0.0003	1.08	0.06
3	0.0021	10.82	0.0019	0.0019	11.66	0.0018	0.0002	0.90	0.21
4	0.0019	12.68	0.0019	0.0016	9.79	0.0014	0.0003	1.56	0.01
5	0.0014	9.75	0.0015	0.0017	9.96	0.0013	-0.0003	-1.19	0.86
6	0.0016	10.62	0.0015	0.0018	9.88	0.0015	-0.0003	-1.20	0.66
7	0.0016	10.92	0.0013	0.0015	10.14	0.0014	0.0001	0.60	0.84
8	0.0017	11.95	0.0016	0.0014	9.42	0.0013	0.0003	1.33	0.09
9	0.0014	10.53	0.0013	0.0017	10.35	0.0015	-0.0003	-1.23	0.21
10	0.0015	10.66	0.0014	0.0017	10.79	0.0014	-0.0001	-0.66	0.77
Σ 1-10	0.0168	24.20	0.0166	0.0151	19.50	0.0138	0.0017	1.67	0.04

A Internet Appendix

Table IA.1
Summary of CRSP Universe

This table summarizes 2009 stock-day observations for CRSP common stocks with dual-class stocks removed. The table summarizes 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 31$^{\text{st}}$ of year $t-1$. Deciles one through five are small-cap, six through eight are mid-cap, and nine through ten are large-cap.

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 IA.2
Sample Universe

This table summarizes 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 $1 million dollars. The table summarizes market capitalization, *mv*, dollar volume, *dolvol*, and price, *prc*. Market value and dollar volume are in millions. The column *szp* denotes size deciles. Size portfolio breakpoints are computed among NYSE-listed stocks. Size portfolios for year t are formed on December 31$^{\text{st}}$ of year $t-1$. Deciles one through five are small-cap, six through eight are mid-cap, and nine through ten are large-cap.

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 IA.3
Summary of News Data

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

Panel A: Distribution of Articles

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

Panel B: Source Summary

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

Table IA.4
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 re-running 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. For each stock-day observation, the following vector autoregressions (VARs) with ten lags are estimated:

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

$$HFT_{NMB,t} = \alpha_2 + \sum_{i=1}^{10} \gamma_{2,i} HFT_{NMB,t-i} + \sum_{i=1}^{10} \beta_{2,i} non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_{2,i} R_{t-i} + \epsilon_{2,t} \tag{11}$$

$$non\text{-}HFT_{NMB,t} = \alpha_3 + \sum_{i=1}^{10} \gamma_{3,i} HFT_{NMB,t-i} + \sum_{i=1}^{10} \beta_{3,i} non\text{-}HFT_{NMB,t-i} + \sum_{i=1}^{10} \lambda_{3,i} R_{t-i} + \epsilon_{3,t} \tag{12}$$

where R_t is the bid-ask midpoint return, $HFT_{NMB,t}$ is one-second HFT net marketable buying, and $non\text{-}HFT_{NMB,t}$ is one-second non-HFT net marketable buying. Net marketable buying is shares bought in buyer-initiated trades minus shares sold in seller-initiated trades. To make imbalance measures comparable across stocks, they are divided by 20-day trailing average daily volume. All variables are divided by their standard deviation among all stocks that day. The equations are estimated for each stock every day. Coefficients are then averaged across all stocks for a day, with stocks weighted by the minimum number of non-zero observations among the three variables. The full-sample estimate is the mean of that daily coefficient time-series. Standard errors correct for autocorrelation using the Newey and West (1994) optimal lag selection method.

Table IA.4 — 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
$y = non\text{-}HFT_t$						
1	0.0018	5.99	0.0515	26.96	0.1714	29.85
2	0.0025	17.08	0.0217	33.62	−0.0002	−0.30
3	0.0022	9.64	0.0159	39.73	−0.0002	−0.44
4	0.0022	8.78	0.0139	19.04	0.0004	1.62
5	0.0021	10.02	0.0127	31.66	−0.0002	−0.82
6	0.0020	8.63	0.0101	20.81	0.0006	1.59
7	0.0015	9.56	0.0073	27.62	0.0006	2.37
8	0.0014	7.84	0.0073	16.94	0.0006	3.01
9	0.0015	8.11	0.0075	16.57	0.0003	1.69
10	0.0017	9.67	0.0097	31.78	0.0000	−0.27
$y = HFT_t$						
1	0.0292	26.26	0.0059	6.92	0.3536	33.51
2	0.0069	5.20	−0.0005	−0.60	0.0148	18.34
3	0.0029	5.11	0.0002	0.43	0.0042	5.48
4	0.0032	9.88	−0.0004	−0.76	0.0020	4.05
5	0.0021	5.47	−0.0006	−0.90	0.0009	2.24
6	0.0011	4.72	−0.0009	−1.73	0.0006	1.24
7	0.0008	2.71	−0.0011	−1.94	0.0000	0.03
8	0.0001	0.19	−0.0010	−2.59	−0.0004	−1.15
9	−0.0001	−0.34	−0.0009	−1.20	−0.0005	−1.56
10	−0.0004	−1.26	−0.0005	−0.93	−0.0011	−4.00
$y = R_t$						
1	0.0222	18.79	0.0310	42.20	−0.0514	−26.91
2	0.0135	18.97	0.0177	29.16	−0.0359	−26.36
3	0.0088	21.64	0.0142	26.74	−0.0257	−28.86
4	0.0074	20.69	0.0109	18.32	−0.0209	−22.87
5	0.0049	17.89	0.0095	15.41	−0.0172	−27.38
6	0.0039	12.20	0.0067	12.52	−0.0145	−22.60
7	0.0028	12.24	0.0050	10.11	−0.0111	−24.67
8	0.0020	4.77	0.0051	13.58	−0.0089	−26.60
9	0.0017	3.90	0.0062	11.09	−0.0073	−17.33
10	0.0000	−0.10	0.0028	6.42	−0.0070	−10.37

Table IA.4 — 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
$y = non\text{-}HFT_t$						
1	0.0013	4.58	0.0515	27.82	0.5567	20.80
2	0.0025	18.78	0.0218	30.73	0.0031	1.43
3	0.0022	9.84	0.0160	38.39	0.0000	−0.02
4	0.0020	7.45	0.0140	18.35	0.0007	0.59
5	0.0021	9.08	0.0129	31.71	−0.0003	−0.24
6	0.0020	6.91	0.0102	21.32	0.0022	1.76
7	0.0014	8.32	0.0073	28.05	0.0021	2.21
8	0.0013	6.79	0.0074	18.35	0.0025	2.53
9	0.0014	8.33	0.0074	16.87	0.0016	1.92
10	0.0017	8.79	0.0098	28.75	0.0008	1.16
$y = HFT_t$						
1	0.0306	26.21	0.0080	8.35	1.0500	32.56
2	0.0076	5.59	0.0008	0.94	0.0287	10.59
3	0.0030	5.97	0.0007	1.55	0.0041	1.80
4	0.0031	10.42	−0.0001	−0.21	0.0011	0.76
5	0.0021	5.36	−0.0001	−0.11	−0.0015	−1.14
6	0.0011	4.90	−0.0007	−1.44	−0.0036	−3.19
7	0.0006	2.51	−0.0009	−1.71	−0.0027	−3.98
8	−0.0002	−0.67	−0.0009	−2.29	−0.0039	−10.39
9	−0.0003	−0.69	−0.0009	−1.30	−0.0042	−2.77
10	−0.0005	−1.81	−0.0001	−0.15	−0.0044	−4.26
$y = R_t$						
1	0.0219	13.64	0.0263	20.90	−0.1260	−30.00
2	0.0138	12.19	0.0156	24.01	−0.1087	−40.03
3	0.0102	14.88	0.0140	10.84	−0.0888	−42.30
4	0.0081	13.95	0.0092	12.38	−0.0758	−36.43
5	0.0067	7.40	0.0079	10.38	−0.0597	−38.59
6	0.0055	8.85	0.0067	11.90	−0.0534	−47.89
7	0.0046	11.02	0.0042	3.64	−0.0423	−32.13
8	0.0026	5.82	0.0034	5.85	−0.0359	−28.80
9	0.0022	6.88	0.0029	3.66	−0.0265	−29.25
10	−0.0002	−0.44	0.0001	0.58	−0.0180	−26.73

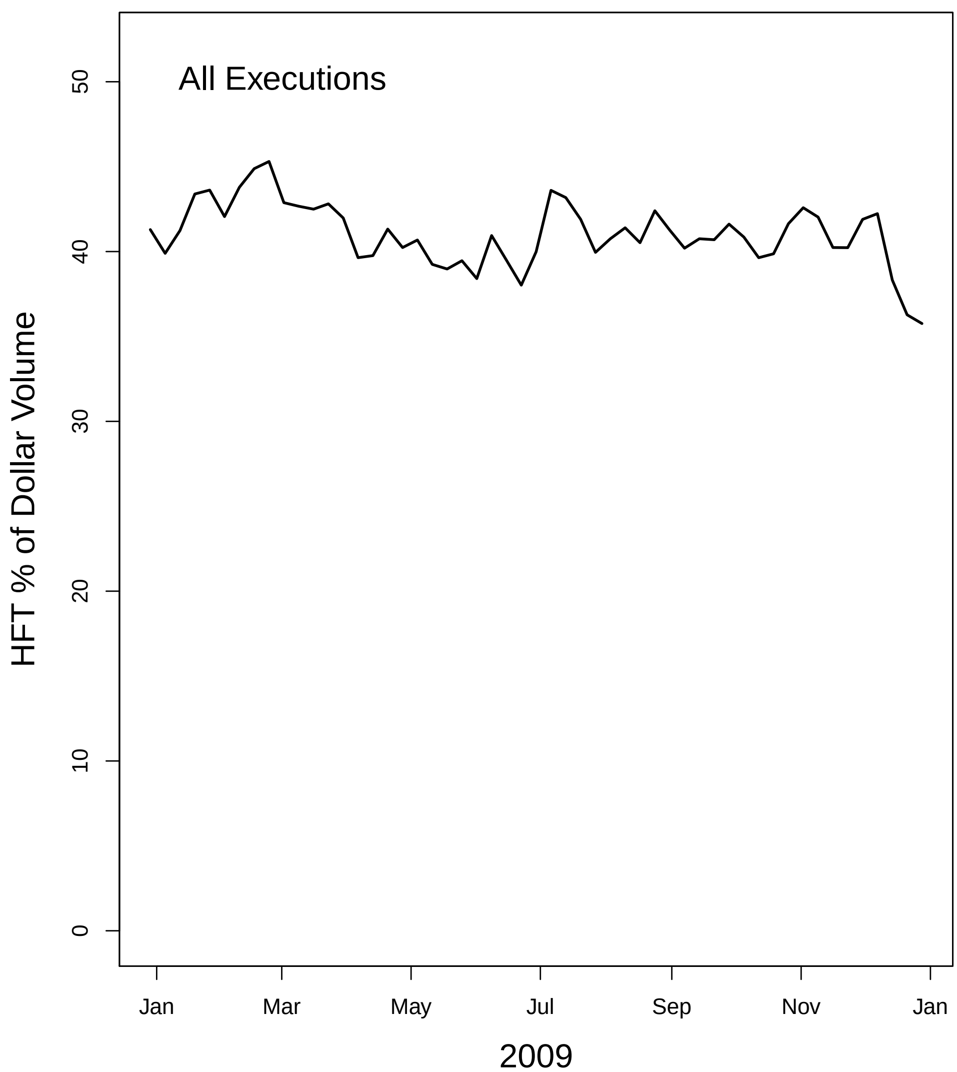

Figure IA.1: HFTs' Share of NASDAQ Dollar Volume
This figure shows HFTs' share of dollar volume on the NASDAQ Stock Market. The calculation includes all stocks with CRSP share code 10 or 11 trading on NASDAQ, regardless of listing venue.

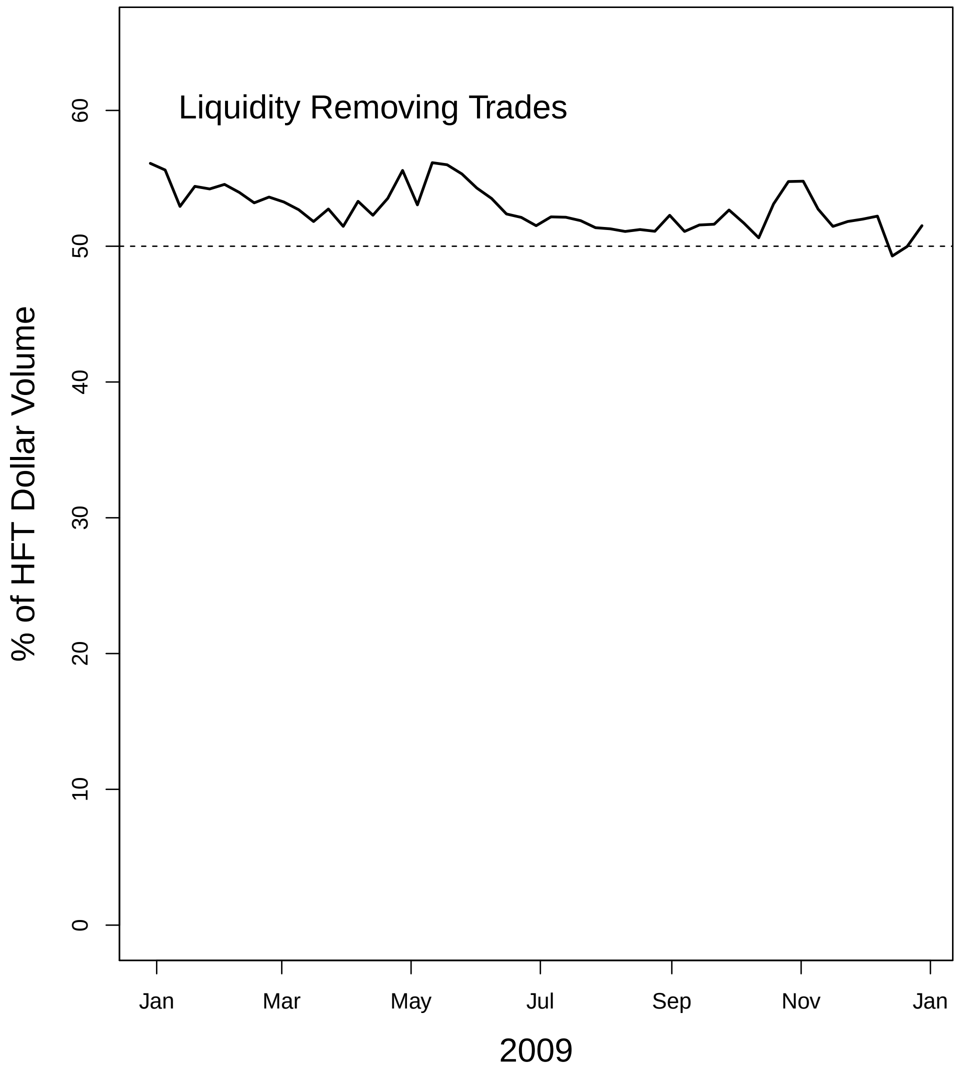

Figure IA.2: Liquidity Removing Trades as a Percent of HFT Dollar Volume
This figure shows liquidity removing trades as a percent of HFTs' dollar volume on the NASDAQ Stock Market. Liquidity removing trades are those in which the HFT initiates the trade with a marketable order, which is functionally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

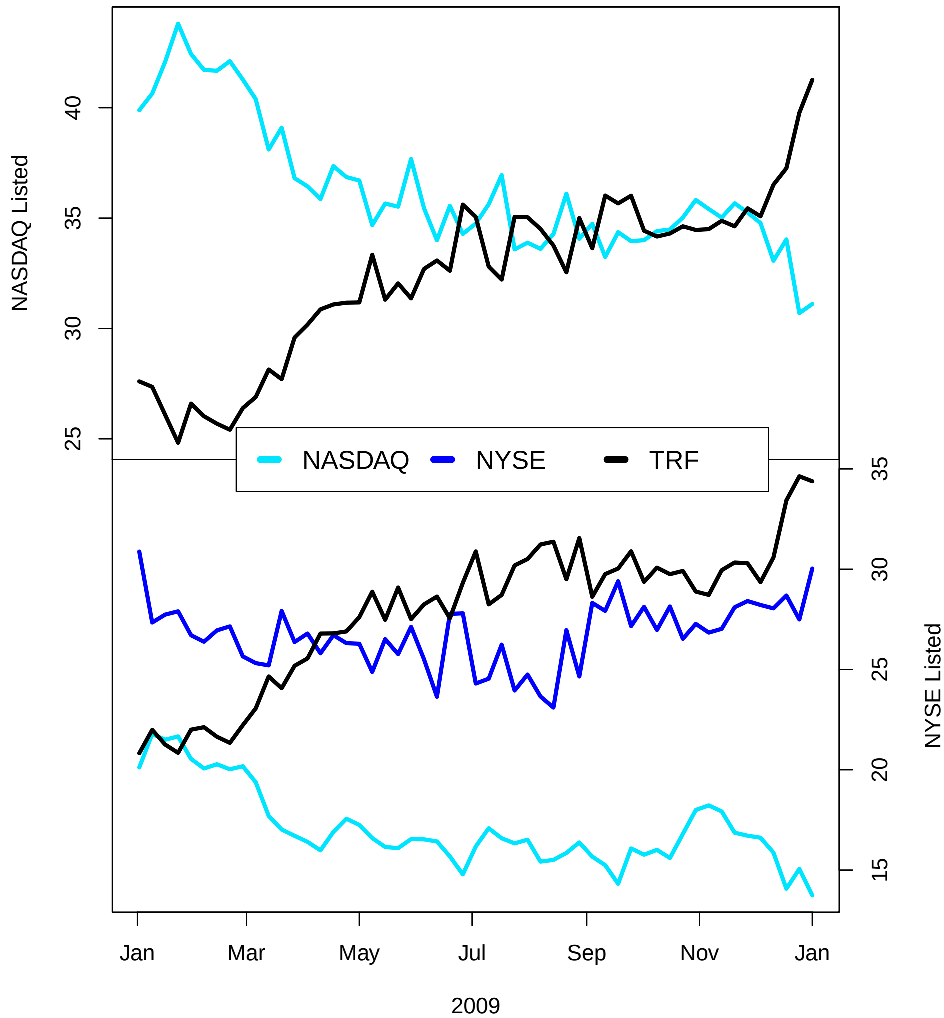

Figure IA.3: Market Share By Trading Venue
Market share is reported as percent of dollar volume. NASDAQ is the NASDAQ Stock Market, NYSE is the New York Stock Market, and TRF is the FINRA Trade Reporting Facility that includes trades that do not occur on a stock exchange (e.g., trades executed in dark pools or by off-exchange market making firms).