High Frequency Trading and Price Discovery*

By

Jonathan Brogaard
(University of Washington)

Terrence Hendershott
(University of California at Berkeley)

Ryan Riordan
(University of Ontario Institute of Technology)

* We thank Frank Hatheway and Jeff Smith at NASDAQ OMX for providing data. NASDAQ makes the data freely available to academics providing a project description and signing a non-disclosure agreement. We thank participants at the Fifth Erasmus Liquidity Conference, the University of Notre Dame & NASDAQ OMX Conference on Current Topics in Financial Regulation, and the Workshop on High Frequency Trading: Financial and Regulatory Implications. Additionaly, we thank Álvaro Cartea, Frank de Jong, Richard Gorelick, Charles Jones, Andrei Kirilenko, Albert Menkveld, Adam Nunes, Roberto Pascal, Gideon Saar, Stephen Sapp, and Cameron Smith for helpful comments. All errors are our own.

High Frequency Trading and Price Discovery

Abstract

We examine the role of high-frequency traders (HFT) in price discovery and price efficiency. Overall HFT facilitate price efficiency by trading in the direction of permanent price changes and in the opposite direction of transitory pricing errors on average days and the highest volatility days. This is done through their marketable orders. In contrast, HFT liquidity-supplying non-marketable orders are adversely selected in terms of the permanent and transitory components as these trades are in the direction opposite to permanent price changes and in the same direction as transitory pricing errors. HFT predicts price changes in the overall market over short horizons measured in seconds. HFT is correlated with public information, such as macro news announcements, market-wide price movements, and limit order book imbalances.

(for internet appendix click: http://goo.gl/vyOEB)

Financial markets have two important functions for asset pricing: liquidity and price discovery for incorporating information in prices (O'Hara (2003)). Historically, financial markets relied on intermediaries to facilitate these goals by providing immediacy to outside investors. Stock exchanges becoming fully automated (Jain (2005)) increased markets' trading capacity and enabled intermediaries to expand their use of technology. This reduced roles for traditional human market makers and led to the rise of a new class of intermediaries, typically referred to as high frequency traders (HFT; HFT also refers to high frequency trading). This paper examines the role of HFT in the price discovery process using data from NASDAQ that identifies the participation of a large group of HFT in each transaction.

Like traditional intermediaries, HFT are central to the trading process, have short holding periods, and trade frequently. Unlike traditional intermediaries, however, HFT are not granted privileged access to the market, otherwise unavailable to others.[1] Without such privileges, there is no clear basis for imposing the traditional obligations of market makers (e.g., Panayides (2007)) on HFT. These obligations were both positive and negative. Typically the positive obligations required intermediaries to always stand ready to provide liquidity and the negative obligations limited intermediaries' ability to demand liquidity. Limiting traders closest to the market from demanding liquidity mitigates the adverse selection costs imposed by having better information about the trading process and being able to react faster to public news.

The substantial, largely negative media coverage of HFT[2] and the "flash crash" on May 6th, 2010 raise significant interest and concerns about HFT's role in the stability and price efficiency of markets. Paul Krugman represents the negative view of high frequency trading:

> "It's hard to imagine a better illustration (of social uselessness) than high-frequency trading. The stock market is supposed to allocate capital to its most productive uses, for example by helping companies with good ideas raise money. But it's hard to see how traders who place their orders one-thirtieth of a second faster than anyone else do anything to improve that social function."[3]

[1] Traditional intermediaries were often given special status and located on the trading floor of exchanges. The optional value inherent in providing firm quotes and limit orders which other traders can execute against makes it difficult for liquidity providers to not be located closest to the trading mechanism. HFT typically utilize co-located servers at exchanges and purchase market data directly from exchanges. These services are also available to other investors and their brokers.

[2] For example, see Duhigg (2009) and the October 10, 2010 report on CBS News' 60 Minutes.

[3] New York Times, August 2, 2009.

While germane, this criticism presumes HFT does not generate any additional price discovery and does not help facilitate private gains from trade, e.g., trades for liquidity or risk-sharing. Our results are consistent with concerns about HFT imposing adverse selection on other investors, but we also find evidence that HFT play a beneficial role in price efficiency and continue to provide liquidity at stressful times such as the most volatile days and around macroeconomic news announcements.

We use a dataset NASDAQ makes available to academics that identifies a subset of HFT trading. The dataset also includes information on whether the initiating (liquidity-demanding) and passive (liquidity-supplying) side of each trade is a HFT. The dataset includes trading data on a stratified sample of stocks in 2008 and 2009. We use a state space model to decompose price movements into permanent and temporary components and to relate changes in both to HFT. The permanent component is normally interpreted as information and the transitory component as pricing errors, also referred to as transitory volatility or noise. The state space model incorporates the interrelated concepts of price discovery (how information is impounded into prices) and price efficiency (the informativeness of prices).

We find that overall HFT play a beneficial role in price efficiency by trading in the direction of permanent price changes and in the opposite direction of transitory pricing errors. This is done through their marketable (liquidity-demanding) orders and is true on both average days and the most volatile days. In contrast, HFT non-marketable (liquidity-supplying) orders are adversely selected in terms of the permanent and transitory components as these trades are in the direction opposite to permanent price changes and in the same direction as transitory pricing errors. HFT liquidity-demanding orders' informational advantage is sufficient to overcome the bid-ask spread and trading fees to generate positive trading revenues. For non-marketable liquidity-supplying limit orders, the costs associated with adverse selection are smaller than revenues from the bid-ask spread and liquidity rebates.

In its concept release on equity market structure, one of the Securities and Exchange Commission's (SEC (2010)) primary concerns is HFT. On p. 36-37 the SEC expresses concern regarding short-term volatility, particularly "excessive" short-term volatility. Such volatility could result from long-term institutional investors breaking large orders into a sequence of small individual trades that result in a substantial cumulative temporary price impact (Keim and Madhavan (1995, 1997)). While each trade pays a narrow bid-ask spread the overall order faces substantial transaction costs. This causes noise in prices due to price pressures arising from liquidity demand by long-term investors. If HFT trade against this transitory pricing error it can be viewed as reducing long-term investors' trading costs. If HFT trade in the direction of the pricing error it can be viewed as increasing the costs to those investors. HFT trading in the direction of pricing errors could arise from predatory trading or attempts to manipulate prices while HFT following various arbitrage strategies could lead to HFT trading in the opposite direction to pricing errors. We find that overall HFT benefits price efficiency, suggesting that the efficiency enhancing activities of HFT play a greater role. Our data represent an equilibrium outcome in the presence of HFT, so the counterfactual of how other market participants would behave in the absence of HFT is not known.

To examine HFT's informational advantage we examine the duration over which HFT predicts returns and how they use public information. HFT predicts price changes over horizons of less than 3 to 4 seconds. HFT trade on two sources of public information: macroeconomic news announcements (Andersen, Bollerslev, Diebold, and Vega (2003)) and imbalances in the limit order book (Cao, Hansch, and Wang (2009)). Given the short duration and public availability of this information, the benefit of improving price efficiency in this way is less obvious, particularly when HFT do so by demanding liquidity.[4]

The paper is structured as follows. Section 2 discusses related literature. Section 3 describes the data, institutional details, and descriptive statistics. Section 4 examines the lead-lag correlation between HFT trading and returns and uses a state space model to decompose prices into their permanent/efficient component and transitory/noise component and analyzes the role of HFT trading in each component. It also relates HFT's role in price discovery to HFT profitability. Section 5 focuses on HFT trading on high-permanent-volatility days. Section 6 analyzes the different sources of information used by HFT. Section 7 discusses the market structure and welfare implications of the findings. Section 8 concludes.

[4] Jovanovic and Menkveld (2011) examine the implications of HFT liquidity suppliers with access to public information.

2. Literature

This paper fits within the expanding literature on algorithmic trading and HFT.[5] Biais and Woolley (2011) provide background and survey related research. Hirschey (2011) uses data from NASDAQ unavailable to other researchers, which identifies trading by individual HFT firms. He finds that aggressive positions taken by HFTs predict subsequent liquidity demand by non-HFTs. If the non-HFT have information about subsequent returns, then such predictability is consistent with our findings that HFT demanding trades help incorporate information into prices. While all traders attempt to forecast each others' actions, whether or not such order anticipation is good for markets is unclear.

Kirilenko, Kyle, Samadi, and Tuzun (2011) study HFT in the E-mini S&P 500 futures market during the May 6th flash crash and suggest that HFT may have exacerbated volatility.[6] Jovanovic and Menkveld (2011) model HFT as middlemen in limit order markets to examine their welfare effects. Menkveld (2011) studies how one HFT firm enabled a new market to gain market share, as well as the role the firm had in the price discovery process along with its profitability and risk taking. Martinez and Rosu (2011) model HFT liquidity-demanding activities and present a rationale for HFT. Their results suggest a positive and stabilizing role for HFT in prices by incorporating information into prices as soon as it is revealed.

This paper also relates to algorithmic trading, of which HFT is a subset. Using an instrument, Hendershott, Jones, and Menkveld (2011) show that algorithmic trading (AT) improves liquidity and makes quotes more informative. Boehmer, Fong, and Wu (2012) provide international evidence on algorithmic trading in equity markets. Chaboud, Chiquoine, Hjalmarsson, and Vega (2009) relate AT to volatility and find little relation. Hendershott and Riordan (2012) focus on the monitoring capabilities of AT and study the relationship between AT and liquidity supply and demand dynamics. They find that AT demand liquidity when it is cheap and supply liquidity when it is expensive, smoothing liquidity over time. Hasbrouck and Saar (2010) study low-

[5] Zhang (2010) measures HFT using trading volume relative to institutional portfolio changes in quarterly 13f filings. This measure captures trading frequencies higher than those of long term investors, but does not identify well the more recent developments in HFT as defined by the SEC (2010).

[6] See Easley, Lopez de Prado, and O'Hara (2011, 2012) for analysis of order flow and price dynamics on May 6, 2010.

latency trading—substantial activity in the limit order book over very short horizons—on NASDAQ in 2007 and 2008 and find that increased low-latency trading is associated with improved market quality. Biais, Foucault, and Moinas (2011) and Pagnotta and Philippon (2011) provide models where investors and markets compete on speed.[7] Foucault, Hombert, and Rosu (2012) examine a model where a trader receives information one period ahead of the rest of the market. Our findings on macroeconomic announcements are consistent with at least some HFT having such an informational/speed advantage.

3. Data, Institutional Details, and Descriptive Statistics

NASDAQ provides the HFT data used in this study to academics under a non-disclosure agreement. The data is for a stratified sample of 120 randomly selected stocks listed on NASDAQ and the New York Stock Exchange (NYSE). The sample contains trading data for all of 2008 and 2009. Trades are time-stamped to the millisecond and identify the liquidity demander and supplier as a high-frequency trader or non-high-frequency trader (nHFT). Firms are categorized as HFT based on NASDAQ's knowledge of their customers and analysis of firms' trading, such as how often their net trading in a day crosses zero, their order duration, and their order to trade ratio. One limitation of the data is that NASDAQ cannot identify all HFT. Firms not included are those that also act as brokers for customers and engage in proprietary lower-frequency trading strategies, e.g., Goldman Sachs, Morgan Stanley, and other large integrated firms. HFT who route their orders through these large integrated firms cannot be clearly identified so they are also excluded. The 26 HFT firms in the NASDAQ data are best thought of as independent proprietary trading firms.[8]

The sample categorizes stocks into three market capitalization groups, large, medium and small. Each size group contains 40 stocks. Half of the stocks in each size category are NASDAQ-listed the other half NYSE-listed. The top 40 stocks are composed of 40 of the largest market

[7] See Boehmer (2005) for an empirical examination of differing speeds across markets.

[8] Some HFT firms were consulted by NASDAQ in the decision to make data available. No HFT firm played any role in which firms were identified as HFT and no firms that NASDAQ considers to be HFT are excluded. While these 26 firms represent a significant amount of trading activity and according to NASDAQ fit the characteristics of HFT, determining the representativeness of these firms regarding total HFT activity is not possible. Hirschey (2011) has access to more detailed data and uses the same classification approach.

capitalization stocks, the medium-size category consists of stocks around the 1000th largest stock in the Russell 3000, and the small-size category contains stocks around the 2000th largest stock in the Russell 3000.[9]

The HFT dataset is provided by NASDAQ and contains the following data:

(1) Symbol
(2) Date
(3) Time in milliseconds
(4) Shares
(5) Price
(6) Buy Sell indicator
(7) Type (HH, HN, NH, NN)

Symbol is the NASDAQ trading symbol for a stock. The Buy Sell indicator captures whether the trade was buyer- or seller-initiated. The type flag captures liquidity-demanding and liquidity-supplying participants in a transaction. The type variable can take one of four values, HH, HN, NH or NN. HH (NN) indicates that a HFT (nHFT) demands liquidity and another HFT (nHFT) supplies liquidity in a trade. HN trades indicate that an HFT firm demands and a nHFT supplies liquidity, the reverse is true for NH trades. The remainder of the paper denotes HFT demanding liquidity trades (HH or HN) as HFT^D and HFT supplying liquidity trades (HH or NH) as HFT^S. Total HFT trading activity (HFT^D + HFT^S) is labeled as HFT^{All}. This notion is used for HFT trading volume (buy volume plus sell volume) and HFT order flow (net trading: buy volume minus sell volume).

The NASDAQ HFT dataset is supplemented with the National Best Bid and Offer (NBBO) from the Trade and Quote database (TAQ). The NBBO measures the best prices prevailing across all markets to focus on market-wide price discovery and is available for all of 2008 and 2009. When combining the two data sets two small-cap stocks are dropped because their symbols do not appear in TAQ at the beginning of the sample period: BZ and MAKO. The HFT trading data and the NBBO do not have synchronized time stamps. Market capitalization data is based on the end-of-year 2009 data retrieved from Compustat. We focus on continuous trading

[9] See the internet appendix for a complete list of sample stocks and size categories.

during normal trading hours and remove opening and closing crosses as well as trading that occurs before 9:30 or after 16:00.

Table 1 reports the descriptive statistics overall and by size category. The average market capitalization of sample stocks is $18.23 billion. The range across size categories is high with an average of $52.47 billion in large stocks and $410 million in small stocks. Unsurprisingly, prices are highest and return volatility lowest in large stocks, with the reverse holding for small stocks.

Insert Table 1 here

Table 1 reports the time-weighted bid-ask spread in dollars and as a percentage of the prevailing quote midpoint using the TAQ NBBO data sampled at one minute frequencies. On average, spreads are greater in both dollar and percentage terms for small stocks than large stocks, with percentage spreads in small stocks roughly 10 times higher than large stocks. Spreads likely play an important role in HFT profitability and behavior, e.g., the decision to demand or supply liquidity. However, spreads calculated based on visible liquidity may overestimate the effective spreads actually paid or received by HFT due to non-displayed orders.

The average per stock NASDAQ daily trading volume is $62.28 million. Trading volume is highest in large stocks at $179.01 million traded per stock-day and lowest in small stocks with roughly $1.11 million traded per stock-day. HFT^D is responsible for roughly 43% of trading volume in large stocks and 23% of volume in small stocks. HFT^S makes up 41% of trading volume in large stocks and only 10% of the trading volume in small stocks. These numbers confirm the conjecture that HFT is concentrated in large liquid stocks. In Table 1 the HFT variables measure total trading volume by summing HFT buying and selling. For the remainder of the paper the HFT trading variables are in terms of HFT order flow (net trading): HFT buy volume minus HFT sell volume.

The SEC (2010) concept release lists a number of characteristics of HFT. One important characteristic is the mean reversion of their trading positions, which NASDAQ reports is true based on internal analysis of individual HFT firms in our sample. However, the aggregation of all

HFT firms trading on one of many market centers may not clearly exhibit mean reversion.[10] The results for an augmented Dickey-Fuller (ADF) test for each stock-day are provided in the internet appendix. The results of the ADF test do not suggest that the aggregate HFT inventories measurable in our data are stationary. Therefore, we use order flow rather than constructed inventory levels in the statistical analysis of HFT trading behavior.

4. HFT Trading and Returns

The correlations between HFT and returns relate HFT trading to price changes at different horizons. Figure 1 plots the correlation between HFT order flow and contemporaneous and subsequent returns at 1-second frequencies over a 10-second period.

Insert Figure 1 here

The figure shows that the correlations between HFT^D and subsequent returns are positive, die out quickly, and are essentially zero after three or four seconds. This is consistent with HFT demanding liquidity on information about short-term price movements. The correlations between HFT^S and returns are similar in that they die out quickly but are in the opposite direction as those for HFT^D. The negative HFT^S correlations suggest that HFT is supplying liquidity to earn the spread and manage risk and in doing so are exposed to adverse selection costs. Overall HFT is contemporaneously positively correlated with returns, but in contrast to HFT^D and HFT^S, the correlation essentially diminishes to zero one-second into the future.

While Figure 1 illustrates the relation between HFT and subsequent returns, it may be the case that returns also predict subsequent HFT. Figure 2 shows this relation using the same format as above using contemporaneous returns to predict subsequent HFT.

Insert Figure 2 here

[10] See Menkveld (2011) for evidence on cross-market inventory management by one HFT.

Contemporaneously the relation is identical to Figure 1. Returns positively predict HFT overall, however the correlations are considerably lower than in Figure 1. The correlations are all essentially zero after one-second. This suggests that while returns appear to possess some predictive power, the directional relation between HFT and returns is stronger from HFT to returns than the reverse.[11] The figures suggest HFT play a role in price discovery at very short horizons.

4.1 State Space Model of HFT and Prices

The results of the correlation analysis suggest that HFT^D and HFT^S have distinct relations with the price process. To better understand the relation between the HFT variables, permanent price changes, and transitory price changes, we estimate a state space model.[12] The state space model assumes that a stock's price can be decomposed into a permanent component and a transitory component (Menkveld, Koopman, and Lucas (2007)):

$$p_{i,t}=m_{i,t}+s_{i,t}$$

where $p_{i,t}$ is the (log) midquote at time interval t for stock i and is composed of a permanent component $m_{i,t}$ and a transitory component $s_{i,t}$. The permanent (efficient) component is modeled as a martingale:

$$m_{i,t}=m_{i,t-1}+w_{i,t}.$$

where $w_{i,t}$ represents innovation in the permanent price component. To capture the overall impact of HFT and the individual impacts of HFT^D and HFT^S we estimate two models. One model incorporates HFT^{All} while a second distinguishes HFT^D and HFT^S activity. Following

---

[11] Internet appendix Figures A1, A2, and A3 depict the auto and cross-autocorrelation of HFT.

[12] Hendershott and Menkveld (2011) provide several reasons why the state space methodology is preferable to other approaches, such as autoregressive models. First, maximum likelihood estimation is asymptotically unbiased and efficient. Second, the model implies that the differenced series is an invertible moving average time series model which implies an infinite lag autoregressive model. When estimating using an alternative approach such as a vector autoregression (Hasbrouck (1991) and following work), the econometrician must truncate the lag structure. Third, after estimation, the Kalman smoother (essentially a backward recursion after a forward recursion with the Kalman filter) facilitates a series decomposition where at any point in time the efficient price and the transitory deviation are estimated using all observations, i.e., past prices, the current price, and future prices.

Hendershott and Menkveld (2011) and Menkveld (2011) we specify $w_{i,t}$ for the aggregate model as:

$$w_{i,t} = \kappa_i^{All} \widetilde{HFT}_{i,t}^{All} + \mu_{i,t}$$

where $\widetilde{HFT}_{i,t}^{All}$ is the surprise innovation in HFT^{All}, which is the residual of an autoregressive model to remove autocorrelation. For the disaggregated model, $w_{i,t}$ is formulated as:

$$w_{i,t} = \kappa_i^{D} \widetilde{HFT}_{i,t}^{D} + \kappa_i^{S} \widetilde{HFT}_{i,t}^{S} + \mu_{i,t}$$

where $\widetilde{HFT}_{i,t}^{D}$ and $\widetilde{HFT}_{i,t}^{S}$ are the surprise innovations in the corresponding variables calculated analogously to the previous model. A lag length of ten seconds is used as determined by standard techniques. The trading variables are designed to capture informed trading and its role in the permanent component of prices. The changes in $w_{i,t}$ unrelated to trading are captured by $\mu_{i,t}$.

The state space model assumes that the transitory component of prices (pricing error) is stationary. To identify the transitory component of prices we include an autoregressive component and the raw trading variables in the equation. We formulate $s_{i,t}$ for the aggregate model as:

$$s_{i,t} = \phi s_{i,t-1} + \psi_i^{All} HFT_{i,t}^{All} + \upsilon_{i,t}$$

and the disaggregate model as:

$$s_{i,t} = \phi s_{i,t-1} + \psi_i^{D} HFT_{i,t}^{D} + \psi_i^{S} HFT_{i,t}^{S} + \upsilon_{i,t}.$$

The inclusion of $HFT_{i,t}^{D}$, $HFT_{i,t}^{S}$ and $HFT_{i,t}^{All}$ enables measurement of the aggregate and disaggregate roles HFT play in the transitory component of prices. As is standard, the identification assumption is that conditional on the trading variables, the innovations in the permanent and transitory components are uncorrelated: $\mathrm{Cov}(\mu_t, \upsilon_t) = 0$.[13]

[13] See the internet appendix for additional implementation details.

4.2 HFT Strategies and Predictions for the State Space Model

The HFT strategies outlined in the SEC (2010, p.48)—passive market marking, arbitrage, structural, and directional—are expected to have different impacts on the transitory and permanent components of prices. The state space model enables inference on the overall roles of HFT in prices and the differential role of liquidity-supplying and demand HFT. Before presenting the estimation results we discuss how the HFT strategies presented in the SEC concept release should impact the estimated state space model coefficients. As the individual HFT firm strategies are not observable in the NASDAQ data, the estimation results provide evidence on which strategies have the largest effects and do not necessarily demonstrate whether specific strategies exist at any HFT firm.

The predictions regarding HFT^D and HFT^S for the permanent component of prices are relatively clear. When utilizing arbitrage and directional strategies we expect HFT^D to be positively correlated with subsequent prices changes. Based on passive market making strategies we generally expect HFT^S to be negatively correlated with subsequent price changes, as better-informed liquidity demanders adversely select them, although it is possible that passive HFT trades could completely avoid adverse selection. The predictions on the transitory component of prices are less clear, with some strategies decreasing pricing errors, e.g., arbitrage, while others increase pricing errors, e.g., manipulation. The SEC does not specify whether or not arbitrage and directional strategies tend to be implemented by demanding or supplying liquidity.

The SEC concept release provides little discussion of risk management which is integral to all short-horizon trading strategies. Risk management typically involves paying transaction costs to reduce unwanted positions. For initiated trades the costs are directly observable in terms of the bid-ask spread and any transitory price impact. For passive limit orders, risk management involves the skewing of quotes, possibly past the fundamental value, e.g., placing limit orders to sell below the fundamental value when the HFT firm has a long position (see Amihud and Mendelson (1980), Ho and Stoll (1981), and others).[14] HFT applying price pressure either by

[14] See Madhavan and Sofianos (1998) for an analysis of trading and risk management strategies by designated market makers on the New York Stock Exchange (specialists).

demanding or supplying liquidity to limit risk could result in HFT order flow being positively associated with transitory pricing errors. However, a positive relation between HFT trading and pricing errors could also be evidence of attempts at manipulation. Distinguishing between risk management and manipulation is not possible without being able to identify individual HFT firms' trades and positions across related instruments.

Turning to the state space model, the interpretation of coefficients on HFT^D is straightforward, as the decision to trade lies solely with the trade initiator. A positive κ^D is interpreted as HFT^D positively contributing to the discovery of the efficient price. In the transitory equation, a positive (negative) ψ^D is interpreted as increasing (decreasing) the noise in prices. Order anticipation and manipulation strategies should result in a positive ψ^D. Identifying pricing errors and profiting from trading against them would yield a negative ψ^D.

The liquidity-supplying HFT order flow variables are interpreted in a similar way; although it is important to keep in mind that while the passive party sets the potential terms of trade, the decision to trade lies with the initiator of the trade. A negative κ^S would provide evidence that HFT are adversely selected. A positive ψ^S indicates that HFT are supplying liquidity in the direction of the pricing error. This could arise from adverse selection regarding the transitory component. HFT could pay to manage risk by applying price pressure via limit order prices to induce other investors to reduce their inventory position. This risk management would lead to a positive ψ^S. Positive ψ^S could also result from order anticipation or attempts to manipulate prices as suggested in the SEC concept release. However, manipulation via passive trading is more difficult to envision as non-marketable orders execute by narrowing the quotes to induce other traders to hit them. This is the opposite of momentum ignition, whereby the initiating trade is responsible for causing the pricing errors.

4.3 State Space Model Estimation

To estimate the state space model for each of the 23,400 one-second time intervals in a trading day, for each stock we use the NBBO midquote price, the HFT liquidity-demanding order flow (HFT dollar buying volume minus HFT dollar selling volume), the HFT liquidity-supplying order flow, and overall HFT order flow (sum of HFT liquidity demand and HFT liquidity supply order flows).[15] The state space model is estimated on a stock-day-by-stock-day basis using maximum likelihood via the Kalman filter.

The sample contains 118 stocks on 510 trading days. To estimate the state space model, we require more than 10 seconds where price changes, HFT^D, HFT^S and HFT^{All} are non-zero. We remove observations for all stocks for that day where this is not the case for any stock, resulting in 504 days for which we have enough observations for all stocks. These stock-days are used in all analyses for the remainder of the paper. We winsorize all estimates at the 1% level. Statistical inference is conducted on the average stock-days estimates by calculating standard errors controlling for contemporaneous correlation across stocks, and time series correlation within stocks, using the clustering techniques in Petersen (2009) and Thompson (2011).

Table 2 reports the results of the HFT^{All} state space model estimation for each size category and overall. Overall HFT^{All} is positively correlated with efficient price changes and negatively correlated with pricing errors. HFT demonstrate an ability to predict both permanent price changes and transitory price changes, and in general these results suggest a positive role in price discovery for HFT.

The κ and ψ coefficients are in basis points per $10,000 traded. The 4.13 value for the overall κ coefficient implies that $10,000 of positive surprise HFT order flow (buy volume minus sell volume) is associated with a 4.13 basis point increase in the efficient price. The positive coefficient is consistent with the findings of O'Hara, Yao, and Ye (2011). The aggregate proportion of efficient price variance correlated with overall HFT order flow is about 9%: 39.44 basis points squared in the 1-second permanent price variance of 438.55 basis points squared. The negative ψ coefficients show that HFT is generally trading in the direction opposite to pricing errors. The pricing errors are persistent with an AR(1) coefficient between 0.45 and 0.48.

[15] The internet appendix contains estimation of the state space model in event-time using a NASDAQ best bid and offer. This BBO is market-specific and is only available for roughly one tenth of the sample period, but the BBO does not suffer from any potential time stamp discrepancies between the HFT trading data and the quoted prices. The coefficient estimates for the event-time BBO model are qualitatively similar to the one-second calendar time NBBO model presented here.

Insert Table 2 here

Table 3 reports the results of the disaggregated model of HFT. It includes the HFT^D and HFT^S trading variables to better model their different impacts and to provide insight into the strategies HFT employ. The two key findings in Panel A on the permanent price component are: (1) HFT liquidity-demanding trades are correlated with changes in the unobserved permanent price component, and, (2) HFT liquidity-supplying trades are adversely selected, as they are negatively correlated with changes in the permanent price component. The first finding follows from κ^D being positive overall and in each size category. Such relations are typically associated with informed trading. The negative coefficients on κ^S show that HFT passive trading occurs in the direction opposite to permanent price movements. This relation exists in models of uninformed liquidity supply, where suppliers earn the spread but lose to informed traders.

Insert Table 3 here

As discussed in Section 4.2, the expected relations between the transitory component of prices and HFT, presented in Panel B of Table 3, are less clear. The state space model estimation results show that ψ^D is negatively related to pricing errors. The negative relation between liquidity-demanding HFT and the transitory component holds across size categories. This indicates that when HFT demand liquidity they trade in the opposite direction to the transitory component of prices, consistent with their trading reducing pricing errors. The natural interpretation is that when prices deviate from their fundamental value, HFT initiate trades to push prices back to their efficient levels. This reduces the distance between quoted prices and the efficient/permanent price of a stock.

The coefficients on ψ^S are positive indicating that HFT^S trading is associated with a larger transitory component of prices across stock size categories. The coefficient is largest for small stocks and smallest for large stocks. One interpretation of the positive ψ^S is that HFT are adversely selected due to being uninformed about the transitory price component. Alternatively, liquidity-supplying HFT could also be fully aware of the transitory component in prices and be intentionally using price pressure for risk management. Finally, HFT could be anticipating subsequent order flow, or attempting to passively manipulate prices, in order to demand liquidity from other market participants. Section 6 discusses these possibilities in more detail.

Liquidity-demanding HFT is associated with more information being incorporated into prices and smaller pricing errors. It is unclear whether or not the liquidity-demanding HFT order submitters know which role any individual trade plays. HFT strategies typically focus on identifying predictability, something we focus on in the following sections. Whether that predictability arises from the permanent or transitory component is less important. We find that in the disaggregate model the variance of the permanent component is 441.73 $bps.^2$ and the transitory component is 243.38 $bps.^2$. In relative terms, permanent price variance is roughly four times greater than transitory price variance in large stocks, and roughly two times greater in small stocks. This suggests that prices of large stocks are considerably less noisy than those of small stocks.

4.4 HFT Revenues

The state space model characterizes the role of HFT in the price process. HFT^D gain by trading in the direction of permanent price changes and against transitory changes. HFT^S lose due to adverse selection and trading in the direction of pricing errors. These possible gains and losses occur before taking into account trading fees and the bid-ask spread. Liquidity-supplying trading earns the spread that liquidity-demanding traders pay. In addition, NASDAQ pays liquidity rebates to liquidity suppliers and charges fees to liquidity-demanding trades.

Using the stock-day panel from the state space model we analyze revenues of overall, liquidity-demanding, and liquidity-supplying HFT. Given that HFT is short-term speculation, it must be profitable or it should disappear. We observe neither all of HFT trading nor all HFT costs, e.g., investments in technology, data and collocation fees, salaries, clearing fees, etc. Hence, we focus on HFT trading revenues incorporating NASDAQ trading maker/taker fees and rebates. We assume that HFT firms are in the highest volume categories for liquidity demand and supply. NASDAQ fees and rebates are taken from the NASDAQ Equity Trader Archive on

NasdaqTrader.com. In 2008 and 2009 we identify six fee and rebate changes affecting the top volume bracket.[16] Fees for liquidity-demanding trades range from \$0.0025 to \$0.00295 per share while rebates for passive trades range from \$0.0025 to \$0.0028 per share.

Another concern highlighted by the SEC (2010) is that HFT supply liquidity to earn fee rebates. However, if liquidity supply is competitive then liquidity rebates should be incorporated in the endogenously determined spread (Colliard and Foucault (2012)). The revenue results also show that HFT liquidity-supplying activities are unprofitable without fee rebates, suggesting that at least some of the rebates are being passed on to liquidity demanders in the form of tighter spreads. Distortions in brokers' routing decisions based on displayed prices versus prices net of fees may be problematic.[17]

We estimate HFT revenues following Sofianos (1995) and Menkveld (2011). Both analyze primarily liquidity supply trading. We decompose total HFT revenue into two components: revenue attributable to HFT^D trading activity, and revenue associated with HFT^S trading activity. We assume that for each stock and each day in our sample, HFT start and end the day without inventories. HFT^D trading revenue for an individual stock for one day is calculated as

$$\bar{\pi}^{*D} = \sum_n^N -(HFT_n^D) + INV\_HFT_N^D * P_T,$$

with each of the *N* transactions within each stock-day subscripted by *n*, and where $INV\_HFT_N^D$ is the daily closing inventory in shares and P_T is the closing quote midpoint. The first term captures trading revenues throughout the day and the second term values the terminal inventory at the closing midquote.[18] $\bar{\pi}^{*S}$ is calculated analogously.

[16] It is difficult to ensure that every fee and rebate change was identified in the archive. However, discrepancies are likely small and on the order of 0.5 to 1 cent per 100 shares traded.

[17] Reg-NMS requires that exchanges route to the exchange with the best posted price. Markets with high market maker rebates may attract more traders willing to supply liquidity at better prices than in markets without rebates. Liquidity demanders trading on these venues may be receiving the best posted price but may be paying more in terms of the spread and taker fees, and are thereby worse off.

[18] Because we do not observe HFT trading across all markets, and HFT likely use both liquidity-demanding and liquidity-supplying orders in the same strategy, the end-of-day inventory could be an important factor in revenues. For large stocks the end-of-day inventories are roughly five to seven percent of trading volume. For smaller stocks the end-of-day inventories are closer to 30 percent of volume. For robustness we calculate, but do not report, profitability using a number of alternative prices for valuing closing inventory: the volume-weighted average price, time-weighted average price, and average of open and close prices. All of these prices yielded qualitatively similar results.

$$\bar{\pi}^{*S} = \sum_{n}^{N} -(HFT_n^S) + INV\_HFT_N^S * P_T$$

Total revenue, $\bar{\pi}^{*All}$, is:

$$\bar{\pi}^{*All} = \bar{\pi}^{*D} + \bar{\pi}^{*S}$$

Table 4 presents the average HFT revenue results overall and for liquidity-demanding and liquidity-supplying trading with and without NASDAQ fees, per stock and day, and per $10,000 traded. Panel A provides the average revenue per stock-day overall and across size categories. Panel B presents the average revenue per $10,000 traded and across size categories.

Insert Table 4 here

HFT^{All} and HFT^{D} have positive revenues overall and in each size category. In Panel C, on a per stock-day basis, liquidity-demanding HFT, after fees, earn $2,433 overall and $6,643, $293 and $38 in large, medium, and small stocks, respectively. The overall, large, and medium revenue results are all statistically significantly different from zero at the 1% level using standard errors double-clustered on stock and day. HFT earn over 100 times more in large stocks than in small stocks. Menkveld (2011) also finds significantly higher revenues in larger stocks for one HFT firm.

Comparing Panels A and B with Panels C and D shows that fees make a substantial difference in revenues. Without accounting for fees, liquidity-supplying HFT have negative revenues. After accounting for fees, both liquidity-demanding and liquidity-supplying HFT have positive revenues. Liquidity-demanding trading's informational advantage is sufficient to overcome the bid-ask spread and fees. Liquidity supply trading's informational disadvantage is overcome by revenues from the bid-ask spread and fees.

Panel D of Table 4 shows that the revenues per dollar traded are low. For large stocks the average revenue per $10,000 traded is small at $0.02. For medium and small stocks the revenue is higher at $0.53 and $0.97 per $10,000, respectively, but given the lower trading volume in those stocks, they have a small impact on overall revenues. The ratios of revenues per $10,000 traded for liquidity-supplying HFT versus liquidity-demanding HFT ($0.50: $1.40) are much greater than for revenues in dollars per stock and day (\$874: \$1,476). This suggests that liquidity-demand revenues are low in high-volume stocks or on high-volume days.

The revenue analysis suggests that while HFT have positive revenues, their revenues per dollar of capital traded are small. This implies considerable competition between HFT for profitable trading opportunities.

5. State Space Model on High-Permanent-Volatility Days

The SEC (2010, p.48) and others express concern about market performance during times of stress. To better understand HFT's role in price discovery during such times, we analyze the subsample of the highest permanent volatility days. The underlying assumption is that is associated with market stress. To identify high-permanent-volatility days we place stocks based on the level of $\sigma^2(w_{i,t})$ into percentiles and examine the stock-days above the 90th percentile. We then compare those days to the remaining 90% of days.

Table 5 reports descriptive statistics, as in Table 1, for high-permanent-volatility days. Statistical inference is conducted on the difference between high-permanent volatility days and other days. The volatility of returns is considerably higher for this subsample, which is expected as total volatility is simply the sum of permanent and transitory volatility. Both dollar and relative spreads are higher on high-permanent-volatility days, consistent with inventory and adverse selection costs being higher for liquidity suppliers on high-permanent-volatility days.

Insert Table 5 here

Trading volume is higher both in total and for HFT on high-permanent-volatility days. Overall, total trading volume increases by \$32.42 million and by \$15.23, \$14.61 and \$29.86 million for HFT^D, HFT^S and HFT^{All}, respectively. As a percentage of total trading volume, HFT^D and HFT^S increase their participation. The fact that HFT^S increases their participation on high-permanent-volatility days demonstrates that HFT continue to supply liquidity in times of market stress.

Table 6 reports the state space model estimates on high-permanent-volatility days for the aggregate model. As in Table 2, Panel A reports results for the permanent price component and Panel B for the transitory price component. Statistical inference is conducted on the difference between high-permanent-volatility days and other days.

Insert Table 6 here

Comparing Tables 2 and 6, the coefficients in the state space model on high-permanent-volatility days have the same signs and are generally of larger magnitudes than on all days. The differences between high-permanent-volatility days and other days are statistically significant for most coefficients.

Table 7 presents the results of the disaggregate model's estimates structured as in Table 3. Similar to the aggregate model results in Table 3, Table 7 finds that the coefficients have the same signs and are larger in magnitude on high-permanent-volatility days. The coefficients on HFT^D and HFT^S for both permanent and transitory components are roughly two to three times greater than the estimates in Table 3. These show that HFT's role in price discovery is qualitatively similar on high-permanent-volatility days, which can be interpreted as high market stress times.[19]

Insert Table 7 here

6. Sources of Public Information

The preceding sections suggest that HFT are informed about subsequent short-term price movements and more so on high information (permanent volatility) days than on other days. These analyses provide little insight into what sources of information drive HFT trading. In this section we analyze sources of public information that HFT may use to predict subsequent price movements.

[19] Revenue analysis as in Table 4 for high-permanent-volatility days is available in the internet appendix.

Information comes from many sources and in many forms. It can be market-wide or stock specific, long-term or short-term, soft or hard, or distinguished among numerous other dimensions.[20] We focus on three types of information identified in prior literature: macroeconomic news announcements, market wide returns, and imbalances in the limit order book.[21]

6.1 Macro News Announcements

Macroeconomic news receives significant attention as a source of market wide information, e.g., Andersen, Bollerslev, Diebold, and Vega (2003). To examine this we analyze eight key macro announcements that occur during trading hours from Bloomberg: Construction Spending, Consumer Confidence, Existing Home Sales, Factory Orders, ISM Manufacturing Index, ISM Services, Leading Indicators, and Wholesale Inventories.

While the expected date and time of a report is announced in advance, the announcements occasionally occur slightly before or after the designated time. For instance, many announcements are reported to be made at 10:00:00 A.M. eastern time. However, the actual announcement may be made at 10:00:10 A.M. Therefore, instead of using the anticipated report time, we use the time second stamp of the first news announcement from Bloomberg. While this usually matches the anticipated report time, there are several occasions where it differs.

Figures 3 and 4 plot the cumulative HFT order flow and the return on a value-weighted portfolio of the stocks in our sample around positive and negative macroeconomic news. A macro announcement is considered positive if the announced value is greater than the average forecast as reported by Bloomberg, and negative if the announcement is below the forecasted average.

Insert Figures 3 and 4 here

[20] See Jovanovic and Menkveld (2011) for a discussion of the differences in types of information employed by HFT and non-HFT investors.

[21] We also obtained the Thompson Reuters News Analytics database to examine HFT and idiosyncratic news. However, the accuracy of the time stamps does not correspond to when news reaches the market and is incorporated into prices (Groß-Klußmann and Hautsch (2010)).

Both figures show that at time zero, prices begin to move in the direction of the macroeconomic announcement. As expected, when the announcement is negative, prices fall, and when the announcement is positive, prices rise. Figures 3 and 4 show that HFT^D buy on positive and sell on negative macroeconomic news; the reverse is true for HFT^S. Overall, HFT^S trading in the opposite direction of macroeconomic news is larger, which results in overall HFT (HFT^{All}) trading in the opposite direction of macroeconomic news. We cannot determine whether HFT trade on the news directly or trade on the price movements in other related securities, e.g., the index futures.

The figures show that macroeconomic announcements contain information and that HFT trade on this information. The liquidity-demanding trades impose adverse selection. The social value of trading quickly on such public information is not clear. The HFT liquidity-supplying trades are adversely selected. The fact that the liquidity-supply effect is greater than the liquidity-demand effect shows that HFT remain actively supplying liquidity under potentially stressful market conditions.

Figures 3 and 4 show that information is not fully incorporated into prices immediately, as returns continue to drift for a number of seconds after the announcement. HFT demand follows a similar drift, but given the graphs are aggregates across all the stocks in the sample, this does not directly establish that HFT demand improves price discovery. For example, it could be the case that higher HFT is associated with prices overshooting in the cross-section of stocks.

For HFT to push prices beyond their efficient level following announcements, HFT demand would need to have a transitory price impact. If this is the case, past HFT order flow should negatively predict subsequent returns. To examine this we estimate the following regression for liquidity-demanding and liquidity-supplying HFT order flow as well as overall HFT:

$$Ret_{i,t+2,t+10} = \alpha + \beta\, HFT^{D,S,All}_{i,t-1,t+1} + \varepsilon_{i,t}$$

where $HFT^{,D,S,All}_{i,t-1,t+1}$ is the HFT order flow as above from one-second before to one-second after a macroeconomic announcement becomes publicly available; $Ret_{t+2,t+10}$ is the return in basis points from two seconds after the macroeconomic announcement through ten seconds afterwards. The regression pools all 209 announcements for each stock. Statistical significance is calculated controlling for contemporaneous correlation across stocks by clustering on announcement days.

The coefficients in Table 8 capture whether HFT is associated with the incorporation of information into prices or transitory price movements. Positive coefficients imply HFT improving price discovery and negative coefficients suggest HFT causing inefficient prices. Panel A reports the HFT^D results, Panel B the HFT^S results, and Panel C the results for HFT^{All}.

Insert Table 8 here

Consistent with the state space model, HFT demand liquidity in the same direction as subsequent price movements, suggesting that they are trading on information in the announcement. This is compatible with the view that at least some component of HFT liquidity demand relates to soon to be public information as in the Foucault, Hombert, and Rosu (2012) model.

HFT supply liquidity in the opposite direction to subsequent price changes, suggesting they are adversely selected. The negative coefficient on HFT liquidity supply is consistent with a positive association with price errors, as in the state space model. The coefficient on overall HFT is positive, although the statistical significance is weak.

6.2 Market-wide Returns

Section 6.1 shows that HFT trading is impacted by macroeconomic news announcements. Jovanovic and Menkveld (2011) find that one HFT firm trades more when there is higher market-wide volatility. To examine this market-wide interaction between the trading of HFT and returns, Figures 5 and 6 extend the stock-specific cross autocorrelations between HFT and returns in Figures 1 and 2 to the sample portfolio. Market returns are for the value-weighted portfolio. Market-wide HFT is the sum of HFT order flow across all stocks.

Insert Figures 5 and 6 here

As in the individual stock correlations in Figures 1 and 2, there is a large positive contemporaneous correlation between HFT^D and returns and a negative correlation between HFT^S and returns. Unlike the individual stock results, the liquidity supply effect is greater than the liquidity demand effect so HFT^{All} is negatively correlated with returns. Another interesting difference in the market-wide results is that the correlations die out less quickly than for the individual stocks. This suggests that HFT plays a somewhat more important and longer lasting role in market-wide price discovery, although their role is still over short time horizons. This is also consistent with the Jovanovic and Menkveld (2011) finding that one HFT is more active when there is more market-wide volatility.

6.3 Limit Order Book

Macroeconomic news announcements and market returns are examples of publicly available information that HFT may use to predict short-term price movements. Another source of information is the state of the limit order book. Cao, Hansch, and Wang (2009) find that imbalances between the amount of liquidity available for buying and selling predict short-run price movements. To test the hypothesis that HFT use order book information to predict short-term subsequent price movements, we calculate limit order book imbalances (LOBI) using the NBBO TAQ best bid and best offer size as in Cao, Hansch, and Wang (2009):

$$LOBI_{i,t} = (Size_{\mathrm{i,t}}^{Offer} - Size_{\mathrm{i},t}^{Bid})/(Size_{\mathrm{i},t}^{Offer} + Size_{\mathrm{i},t}^{Bid}),$$

where *Size* is the dollar volume of orders available at the NBBO. *LOBI* is scaled by 10,000. To test whether or not *LOBI* predicts returns in our sample, we first regress the return in period *t*+1, Ret_{t+1} on *LOBI* in period *t*:

$$Ret_{\,i,t+1} = \alpha + \beta\; LOBI_{i,t} + \varepsilon_{i,t},$$

Similarly, to test if HFT is trading in the direction of limit order book imbalances, we estimate the following regression:

$$HFT_{i,t+1}^{D,S,All} = \alpha + \beta\; LOBI_{i,t} + \varepsilon_{i,t},$$

where $HFT_{i,t+1}^{D,S,All}$ is the HFT order flow in period t+1 for Demand, Supply, and All HFT activity, respectively.

Table 9 reports the mean coefficient estimates for large, medium, small, and all stocks. Panel A contains the subsequent return regression; Panels B, C, and D report the regressions with the dependent variable as HFT Demand, Supply, and All, respectively. Negative coefficients represent HFT trading in the direction of the imbalance, e.g., buying when there are fewer shares offered to buy than shares offered to sell. Positive coefficients indicate HFT is supplying liquidity on the thin side of the book or that HFT demand is trading with the thicker side of the book. As with the state space model, the regressions are conducted for each stock-day and statistical significance is based on the averages of these stock-day estimates clustering on day and stock.

Insert Table 9 here

The negative coefficients in the return regression in Panel A indicate that LOBI contains predictive power about subsequent price movements as in Cao, Hansch, and Wang (2009). The negative coefficients in the HFT^{D} and HFT^{All} regressions suggest that HFT uses the information in the limit order book to predict and profit from subsequent short-term price movements. The positive coefficient in the HFT^{S} regression suggests that HFT is often supplying liquidity on the thin side of the limit order book. This involves incurring adverse selection costs by supplying liquidity in the direction where less liquidity is available. Such liquidity supply is generally interpreted as beneficial.

Overall *LOBI* predicts liquidity demand more than liquidity supply, so HFT trade on the thinner side of the book. HFT demanders appear to use the easily interpretable public information in limit order books to trade. It is possible that the limit order submitter is aware of this, but prefers placing an aggressive limit order rather than paying the spread. In this case, the adverse selection is the limit order submitter's conscious payment to the liquidity demander to avoid paying the spread.

7. Discussion

Overall HFT has a beneficial role in the price discovery process in terms of information being impounded into prices and smaller pricing errors. Traditionally this has been viewed positively, as more informative stock prices can lead to better resource allocation in the economy. However, the information HFT use is short-lived at less than 3-4 seconds. If this information would become public without HFT, the potential welfare gains may be small or negative if HFT imposes significant adverse selection on longer-term investors.[22] Our evidence on HFT liquidity demand immediately following macroeconomic announcements may fall into this category. However, HFT liquidity supply at this time is greater than HFT liquidity demand, so overall HFT is not imposing net adverse selection on others around macroeconomic news.

The fact that HFT predicts price movements for mere seconds does not demonstrate that the information would inevitably become public. It could be the case that HFT compete intensely with each other to get information not obviously public into prices. If HFT were absent, it is unclear how such information would get into prices unless some other market participant played a similar role. This is a general issue in terms of how to define what information is public and how it gets into prices, e.g., the incentives to invest in information acquisition in Grossman and Stiglitz (1980). As Hasbrouck (1991, p. 190) writes, "the distinction between public and private information is more clearly visible in formal models than in practice."

Reducing pricing errors improves the efficiency of prices. Just as with the short-term nature of HFT's informational advantage, it is unclear whether or not intraday reductions in pricing errors facilitate better financing decisions and resource allocations by firms and investors. One important positive role of smaller pricing errors would be if these corresponded to lower implicit transaction costs by long-term investors. Examining non-public data from long-term investors' trading intentions would help answer this.

[22] Jovanovic and Menkveld (2011) show how HFT trading on soon-to-be-public information can either enhance welfare by increasing gains from trade or lower welfare by imposing adverse selection costs on other investors. They focus largely on HFT liquidity supply. Foucault, Hombert, and Rosu (2012) use a model to examine the issue for HFT liquidity demand based on soon-to-be public information.

The negative association of overall HFT with pricing errors fails to support HFT generally engaging in manipulation. However, liquidity-supplying HFT is positively associated with pricing errors. This could be due to adverse selection in the transitory component, risk management, order anticipation, or manipulation. The SEC (2010, p. 53) suggests one manipulation strategy based on liquidity supply: "A proprietary firm could enter a small limit order in one part of the market to set up a new NBBO, after which the same proprietary firm triggers guaranteed match trades in the opposite direction."[23] If the limit order is executed before being cancelled, it could result in HFT liquidity supply being positively associated with pricing errors.

As is often the case, one can argue whether the underlying problem in possible manipulation would lie with the manipulator or the market participant who is manipulated. In the SEC example, if there is no price matching, the liquidity supply manipulation could not succeed. While we think risk management is a more plausible explanation for the positive relation between liquidity-supplying HFT and pricing errors, further investigation is warranted. Cartea and Penalva (2011) present a scenario in which HFT intermediation leads to increased price volatility. The adverse selection in the transitory component, risk management, and manipulation stories are testable with more detailed data identifying each market participant's orders, trading, and positions in all markets.

8. Conclusion

We examine the role of HFT in price discovery. Overall HFT increase the efficiency of prices by trading in the direction of permanent price changes and in the opposite direction of transitory pricing errors. This is done through their marketable orders. In contrast, HFT liquidity-supplying non-marketable orders are adversely selected on both the permanent and transitory component of prices. HFT marketable orders' informational advantage is sufficient to overcome the bid-ask spread and trading fees to generate positive trading revenues. For non-marketable limit orders the costs associated with adverse selection are less than the bid-ask spread and liquidity rebates. HFT predicts price changes occurring a few seconds in the future. The short-

[23] This is the basic behavior that the Financial Industry Regulatory Authority (FINRA) fined Trillium Brokerage Services for in 2010 (http://www.finra.org/Newsroom/NewsReleases/2010/P121951). Trillium is not one of the 26 firms identified as HFT in this paper.

lived nature of HFT information raises questions about whether the informational efficiency gains outweigh adverse selection costs imposed on non-HFT.[24]

One important concern about HFT is their role in market stability.[25] Our results provide no evidence that HFT contribute directly to market instability in prices. To the contrary, HFT overall trade in the direction of reducing transitory pricing errors, both on average days and on the most volatile days, during a period of relative market turbulence (2008-2009). The fact that HFT impose adverse selection costs on liquidity suppliers, overall and at times of market stress, could lead non-HFT liquidity suppliers to withdraw from the market as discussed in Biais, Foucault, and Moinas (2011). This could indirectly result in HFT reducing market stability despite the fact that HFT liquidity suppliers remain active during these stressful periods.

Our results are one step towards better understanding how HFT trade and affect market structure and performance. We identify several types of public information related to HFT: macroeconomic announcements and limit order book imbalances. Studies examining HFT around individual stocks' news announcements, stocks' earnings, and other events could provide further identification and understanding. Our analysis is for a single market for a subset of HFT. Better data for both HFT and long-term investors may enable more general conclusions. The cross-stock, cross-market, and cross-asset behavior of HFT are also important areas of subsequent research.

HFT are a type of intermediary. When thinking about the role HFT plays in markets it is natural to compare the new market structure to the prior market structure. Some primary differences are that there is free entry into HFT, HFT do not have a designated role with special privileges, and HFT do not have special obligations. When considering the optimal industrial organization of the intermediation sector, HFT more resembles a highly competitive environment than traditional market structures. A central question is whether there were possible benefits from the old, more highly regulated intermediation sector, e.g., requiring

[24] HFT adverse selection due to marginally faster reaction can lead other investors to make significant technology investments. Another related cost for exchanges, investors and brokers of HFT activity is the significant flow of market data generated.

[25] See, for example, the speech "Race to Zero" by Andrew Haldane, Executive Director, Financial Stability, of the Bank of England, at the International Economic Association Sixteenth World Congress, Beijing, China, on July 8, 2011.

continuous liquidity supply and limiting liquidity demand, that outweigh lower innovation and higher entry costs typically associated with regulation.

References:

Amihud, Y., and H. Mendelson. (1980). Dealership Market: Market-Making with Inventory. *Journal of Financial Economics* 8:31–53.

Andersen, T., T. Bollerslev, F. Diebold, and C. Vega. (2003). Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange. *American Economic Review* 93: 38-62.

Biais,B., and P. Woolley. (2011). High Frequency Trading. *Working Paper.*

Biais, B., T. Foucault, and S. Moinas (2011). Equilibrium High Frequency Trading. *Working Paper.*

Boehmer, E. (2005). Dimensions of Execution Quality: Recent Evidence for U.S. Equity Markets. *Journal of Financial Economics* 78: 463–704.

Boehmer, E., K. Fong, and J. Wu (2012). International Evidence on Algorithmic Trading. *Working Paper*.

Cao, C., O., Hansch, and X. Wang. (2009). The Information Content of an Open Limit-Order Book. *Journal of Futures Markets* 29:16-41.

Cartea, A., and J. Penalva. (2011). Where is The Value in High Frequency Trading? *Working Paper*.

Colliard, J.-E., and T. Foucault (2012). Trading Fees and Efficiency in Limit Order Markets. *Review of Financial Studies,* forthcoming.

Duhigg, C. (2009). Stock Traders Find Speed Pays, in Milliseconds. *New York Times* July 25, 2009, A1.

Chaboud, A., E. Hjalmarsson, C. Vega, and B. Chiquoine. (2009). Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market. *Working Paper*.

Easley, D., M. M. Lopez de Prado, and M. O'Hara. (2011). The Microstructure of the 'Flash Crash': Flow Toxicity, Liquidity Crashes, and the Probability of Informed Trading. *Journal of Portfolio Management* 37:118-128

Easley, D., M. M. Lopez de Prado, and M. O'Hara. (2012). Flow Toxicity and Liquidity in a High Frequency World. *Review of Financial Studies* 25:1457-1493.

Foucault, T., Hombert, J., and I. Rosu (2012). News Trading and Speed. *Working Paper.*

Groß-Klußmann, A., and N. Hautsch. (2010). When Machines Read the News: Using Automated Text Analytics to Quantify High Frequency News-Implied Market Reactions. *Journal of Empirical Finance* 18:321-340.

Grossman, S., and J. Stiglitz. (1980). On the Impossibility of Informationally Efficient Markets. *American Economic Review* 70:393-408.

Hasbrouck, J. (1991). The Information Content of Stock Trades. *Journal of Finance* 46:179-207.

Hasbrouck, J., and G. Saar. (2010). Low Latency Trading. *Working Paper*.

Hendershott, T., C. Jones, and A.J. Menkveld. 2011. Does Algorithmic Trading Increase Liquidity? *Journal of Finance* 66:1-33.

Hendershott, T., and A.J. Menkveld. 2011. Price Pressures. *Working Paper*.

Hendershott, T., and R. Riordan. 2012. Algorithmic Trading and the Market for Liquidity. *Journal of Financial and Quantitative Analysis*, forthcoming.

Hirschey, N. (2011). Do High-Frequency Traders Anticipate Buying and Selling Pressure? *Working Paper.*

Ho, T., and H. R. Stoll (1981). Optimal Dealer Pricing under Transaction Cost and Return Uncertainty. *Journal of Financial Economics* 9:47–73.

Jain, P. (2005). Financial Market Design and the Equity Premium: Electronic vs. Floor Trading. *Journal of Finance* 60:2955-2985.

Jovanovic, B., and A. J. Menkveld. (2011). Middlemen in Limit-Order Markets. *Working Paper.*

Keim, D., and A. Madhavan. (1995). Anatomy of the trading process Empirical evidence on the behavior of institutional traders. *Journal of Financial Economics* 37:371-398.

Keim, D., and A. Madhavan. (1996). Transactions costs and investment style: an inter-exchange analysis of institutional equity trades. *Journal of Financial Economics* 46:265-292.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun. (2011). The Flash Crash: The Impact of High Frequency Trading on an Electronic Market. *Working Paper.*

Madhavan, A., and G. Sofianos. (1998). An Empirical Analysis of NYSE Specialist Trading. *Journal of Financial Economics* 48:189-210.

Martinez, A., and I. Rosu. (2011). High Frequency Trading, News and Volatility. *Working Paper.*

Menkveld, A. J. (2011). High Frequency Trading and the New-Market Makers. *Working Paper.*

Menkveld, A. J., S. J. Koopman, and A. Lucas. (2007). Modeling around-the-clock Price Discovery for Cross-listed Stocks using State Space Models. *Journal of Business & Economic Statistics* 25:213–225.

O'Hara, M. (2003). Liquidity and Price Discovery. *Journal of Finance* 58: 1335-1354.

O'Hara, M, C. Yao, and M. Ye. (2011). What's not there: The Odd-Lot Bias in TAQ Data. *Working Paper*.

Pagnotta, E., and T. Philippon (2011). Competing on Speed. *Working Paper*.

Panayides, M. (2007). Affirmative Obligations and Market Making with Inventory. *Journal of Financial Economics* 86:513-542.

Petersen, M. (2009). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies* 22, 435-480.

Securities and Exchange Commission, 2010, Concept Release on Equity Market Structure, Release No. 34-61358; File No. S7-02-10.

Sofianos, G. (1995). Specialist Gross Trading Revenues at the New York Stock Exchange. *Working Paper* 95-01, New York Stock Exchange.

Thompson, S. (2011). Simple Formulas for Standard Errors that Cluster by both Firm and Time. *Journal of Financial Economics* 99, 1-10.

Zhang, F. (2010). High-Frequency Trading, Stock Volatility, and Price Discovery. *Working Paper*.

Table 1: Descriptive Statistics.

This table reports descriptive statistics that are equal weighted averages across stock-days for 118 stocks traded on NASDAQ for 2008 and 2009. Each stock is in one of three market capitalization categories: large, medium, and small. The closing midquote price is the average bid and ask price at closing. Trading volume is the average dollar trading volume and is also reported by HFT type.

Summary Statistics	Units	Source	Large	Medium	Small	All
Market Capitalization	$ Billion	Compustat	$52.47	$1.82	$0.41	$18.23
Price	$	TAQ	$56.71	$30.03	$17.93	$34.95
Daily Midquote Return Volatility	bps.	TAQ	16.5	25.8	42.9	30.3
Bid-Ask Spread	$	NASDAQ	$0.03	$0.04	$0.09	$0.05
Relative Bid-Ask Spread	bps.	TAQ	5.29	13.32	50.20	15.73
NASDAQ Trading Volume	$ Million	NASDAQ	$179.01	$6.35	$1.11	$62.28
HFT^D **Trading Volume**	$ Million	NASDAQ	$77.06	$2.38	$0.26	$26.96
HFT^S **Trading Volume**	$ Million	NASDAQ	$75.86	$1.18	$0.11	$26.18
HFT^{All} **Trading Volume**	$ Million	NASDAQ	$152.92	$3.55	$0.37	$53.13

Table 2: State Space Model of $\mathbf{HFT^{All}}$ and Prices.

The model is estimated for each stock, each day, using HFT trading variables to decompose the observable price (midquote) $p_{i,t}$ for stock i at time t (in one-second increments) into two components: the unobservable efficient price $m_{i,t}$ and the transitory component $s_{i,t}$:

$$p_{i,t} = m_{i,t} + s_{i,t}$$
$$m_{i,t} = m_{i,t-1} + w_{i,t}$$
$$w_{i,t} = \kappa_i^{All}\widetilde{HFT}_{i,t}^{All} + \mu_{i,t}$$
$$s_{i,t} = \phi s_{i,t-1} + \psi_i^{All} HFT_{i,t}^{All} + \upsilon_{i,t}$$

$HFT_{i,t}^{All}$ is HFT overall order flow; $\widetilde{HFT}_{i,t}^{All}$ is the surprise component of the order flow. Each stock is in one of three market capitalization categories: large, medium, and small. T-statistics are calculated using standard errors double-clustered on stock and day.

Panel A: Permanent Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\kappa^{All}}$	**bps. / $10000**	0.25	4.80	7.99	4.13
(t-stat)		(11.32)	(29.33)	(8.55)	(13.55)
$\boldsymbol{\sigma^2(\widetilde{HFT}^{All})}$	**$10000**	3.35	0.72	0.29	1.53
$\boldsymbol{(\kappa^{All} * \sigma(\widetilde{HFT}^{All}))^2}$	**bps.^2**	1.51	18.86	109.63	39.44
(t-stat)		(8.37)	(29.58)	(53.01)	(49.97)
$\boldsymbol{\sigma^2(w_{i,t})}$	**bps.^2**	22.89	208.01	1214.75	438.85

Panel B: Transitory Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\phi}$		0.48	0.48	0.45	0.47
$\boldsymbol{\psi^{All}}$	**bps. / $10000**	-0.03	-2.21	-4.03	-1.97
(t-stat)		(-2.95)	(-25.95)	(-9.08)	(-13.73)
$\boldsymbol{\sigma^2(HFT^{All})}$	**$10000**	3.37	0.74	0.33	1.55
$\boldsymbol{(\psi^{All} * \sigma(HFT^{All}))^2}$	**bps.^2**	0.38	5.74	38.80	13.58
(t-stat)		(6.04)	(26.40)	(55.57)	(53.50)
$\boldsymbol{\sigma^2(s_{i,t})}$	**bps.^2**	2.88	80.13	746.15	248.86

Table 3: State Space Model of $\mathbf{HFT^D}$, $\mathbf{HFT^S}$ and Prices.

The model is estimated for each stock, each day, using HFT trading variables to decompose the observable price (midquote) $p_{i,t}$ for stock i at time t (in one-second increments) into two components: the unobservable efficient price $m_{i,t}$ and the transitory component $s_{i,t}$:

$$p_{i,t} = m_{i,t} + s_{i,t}$$
$$m_{i,t} = m_{i,t-1} + w_{i,t}$$
$$w_{i,t} = \kappa_i^D \widetilde{HFT}_{i,t}^D + \kappa_i^S \widetilde{HFT}_{i,t}^S + \mu_{i,t}$$
$$s_{i,t} = \phi s_{i,t-1} + \psi_i^D HFT_{i,t}^D + \psi_i^S HFT_{i,t}^S + \upsilon_{i,t}$$

$HFT_{i,t}^D$ and $HFT_{i,t}^S$ are HFT liquidity-demanding and liquidity-supplying order flow; $\widetilde{HFT}_{i,t}^D$ and $\widetilde{HFT}_{i,t}^S$ are the surprise components of those order flows. Each stock is in one of three market capitalization categories: large, medium, and small. T-statistics are calculated using standard errors double-clustered on stock and day.

Panel A: Permanent Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\kappa^D}$	**bps. / $10000**	0.59	9.33	49.57	18.10
(t-stat)		(21.60)	(34.32)	(35.06)	(36.92)
$\boldsymbol{\kappa^S}$	**bps. / $10000**	-0.58	-11.14	-73.33	-25.73
(t-stat)		(-25.04)	(-33.01)	(-35.86)	(-37.56)
$\boldsymbol{\sigma^2(\widetilde{HFT}^D)}$	**$10000**	3.32	0.70	0.27	1.50
$\boldsymbol{\sigma^2(\widetilde{HFT}^S)}$	**$10000**	2.50	0.33	0.14	1.04
$\boldsymbol{(\kappa^D * \sigma(\widetilde{HFT}^D))^2}$	$\boldsymbol{bps.^2}$	2.94	23.53	119.96	44.66
(t-stat)		(15.94)	(31.84)	(45.14)	(43.97)
$\boldsymbol{(\kappa^S * \sigma(\widetilde{HFT}^S))^2}$	$\boldsymbol{bps.^2}$	1.70	14.05	112.94	38.82
(t-stat)		(12.09)	(31.33)	(55.54)	(55.84)
$\boldsymbol{\sigma^2(w_{i,t})}$	$\boldsymbol{bps.^2}$	24.70	217.43	1210.12	441.73

Panel B: Transitory Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\phi}$		0.59	0.49	0.43	0.51
$\boldsymbol{\psi^D}$	**bps. / $10000**	-0.12	-4.03	-17.24	-6.54
(t-stat)		(-9.67)	(-30.81)	(-25.40)	(-28.30)
$\boldsymbol{\psi^S}$	**bps. / $10000**	0.12	4.70	28.69	10.15
(t-stat)		(11.19)	(31.74)	(32.29)	(34.55)
$\boldsymbol{\sigma^2(HFT^D)}$	**$10000**	3.34	0.72	0.30	1.52
$\boldsymbol{\sigma^2(HFT^S)}$	**$10000**	2.51	0.34	0.15	1.05
$\boldsymbol{(\psi^D * \sigma(HFT^D))^2}$	$\boldsymbol{bps.^2}$	0.53	6.58	40.19	14.34
(t-stat)		(8.62)	(26.02)	(49.12)	(47.26)
$\boldsymbol{(\psi^S * \sigma(HFT^S))^2}$	$\boldsymbol{bps.^2}$	0.36	4.47	42.52	14.22
(t-stat)		(7.85)	(26.67)	(55.88)	(56.88)
$\boldsymbol{\sigma^2(s_{i,t})}$	$\boldsymbol{bps.^2}$	5.45	79.65	724.48	243.38

Table 4: HFT Revenues.

This table presents results on HFT trading revenue with and without NASDAQ trading fees and rebates. Revenues are calculated for HFT demand, supply, and the sum of both: HFT^{D}, HFT^{S} and HFT^{All}. Each stock is in one of three market capitalization categories: large, medium, and small. Panel A reports results per stock-day, and Panel B reports HFT revenue per stock and day per $10,000 traded. Panels C & D report the same as in Panels A & B after incorporating NASDAQ fees and rebates.

Panel A: HFT Revenue per stock-day

	Large	**Medium**	**Small**	**All**
$\boldsymbol{HFT^{D}}$	$7,464.64	$428.88	$59.91	$2,681.16
(t-stat)	(6.89)	(5.38)	(3.77)	(7.30)
$\boldsymbol{HFT^{S}}$	$-1,911.30	$-46.43	$0.04	$-660.12
(t-stat)	(-2.19)	(-0.92)	(0.00)	(-2.22)
$\boldsymbol{HFT^{All}}$	$5,553.33	$382.45	$59.95	$2,021.04
(t-stat)	(4.03)	(5.01)	(2.94)	(4.32)

Panel B: HFT Revenue per stock-day per $10,000 traded

	Large	**Medium**	**Small**	**All**
$\boldsymbol{HFT^{D}}$	$0.98	$2.10	$3.85	$2.27
(t-stat)	(7.71)	(8.48)	(4.89)	(8.25)
$\boldsymbol{HFT^{S}}$	$-0.27	$-0.34	$-0.23	$-0.29
(t-stat)	(-2.90)	(-1.15)	(0.75)	(-0.06)
$\boldsymbol{HFT^{All}}$	$0.44	$1.45	$2.55	$1.45
(t-stat)	(5.38)	(4.73)	(3.51)	(5.25)

Panel C: HFT Revenue per stock-day after fees

	Large	**Medium**	**Small**	**All**
$\boldsymbol{HFT^{D}}$	$2,433.43	$144.52	$16.18	$874.54
(t-stat)	(2.27)	(1.83)	(1.02)	(2.41)
$\boldsymbol{HFT^{S}}$	$4,209.15	$148.91	$21.62	$1,476.56
(t-stat)	(4.76)	(2.93)	(1.43)	(4.90)
$\boldsymbol{HFT^{All}}$	$6,642.58	$293.44	$37.81	$2,351.11
(t-stat)	(4.70)	(3.88)	(1.85)	(4.99)

Panel D: HFT Revenue per stock-day after fees per $10,000 traded

	Large	**Medium**	**Small**	**All**
$\boldsymbol{HFT^{D}}$	$0.02	$0.53	$0.97	$0.50
(t-stat)	(0.17)	(1.14)	(0.70)	(1.33)
$\boldsymbol{HFT^{S}}$	$0.64	$1.14	$2.49	$1.40
(t-stat)	(6.77)	(3.93)	(3.93)	(6.08)
$\boldsymbol{HFT^{All}}$	$0.48	$1.06	$1.94	$1.14
(t-stat)	(5.00)	(4.51)	(3.38)	(5.02)

Table 5: Descriptive Statistics on High-Permanent Volatility Days.

This table reports summary statistics variables for high-permanent-volatility ($\sigma^2(\mathrm{w}_{i,t})$) days from the state space model in Table 4. High-permanent-volatility days are categorized for each stock when $\sigma^2(\mathrm{w}_{i,t})$ is in the 90th percentile for that stock. Each stock is in one of three market capitalization categories: large, medium, and small. Differences between high-permanent-volatility days and other days are statistically significant at the 1% level using standard errors double-clustered on stock and day for all variables in the table.

Summary Statistics	**Units**	**Source**	**Large**	**Medium**	**Small**	**All**
Daily Midquote Return Volatility	bps.	TAQ	30.99	47.05	72.50	49.96
Bid-Ask Spread	$	NASDAQ	0.04	0.07	0.14	0.08
Relative Bid-Ask Spread	bps.	TAQ	9.14	27.89	82.87	27.81
NASDAQ Trading Volume	$ Million	NASDAQ	$271.00	$8.77	$1.30	$94.70
$\boldsymbol{HFT^D}$ **Trading Volume**	$ Million	NASDAQ	$121.74	$3.31	$0.33	$42.19
$\boldsymbol{HFT^S}$ **Trading Volume**	$ Million	NASDAQ	$119.58	$1.49	$0.15	$40.79
$\boldsymbol{HFT^{All}}$ **Trading Volume**	$ Million	NASDAQ	$241.33	$4.80	$0.48	$82.99

Table 6: State Space Model of $\mathbf{HFT^{All}}$ and Prices on High-Permanent-Volatility Days.
This table reports the estimates for the state space model for high-permanent-volatility ($\sigma^2(w_{i,t})$) days. High-permanent-volatility days are categorized for each stock when $\sigma^2(w_{i,t})$ is in the 90th percentile for that stock. The model is estimated for each stock, each day, using HFT trading variables to decompose the observable price (midquote) $p_{i,t}$ for stock i at time t (in one-second increments) into two components: the unobservable efficient price $m_{i,t}$ and the transitory component $s_{i,t}$:

$$p_{i,t} = m_{i,t} + s_{i,t}$$
$$m_{i,t} = m_{i,t-1} + w_{i,t}$$
$$w_{i,t} = \kappa_i^{All}\widetilde{HFT}_{i,t}^{All} + \mu_{i,t}$$
$$s_{i,t} = \phi s_{i,t-1} + \psi_i^{All} HFT_{i,t}^{All} + \upsilon_{i,t}$$

$HFT_{i,t}^{All}$ is HFT overall order flow; $\widetilde{HFT}_{i,t}^{All}$ is the surprise component of the order flow. Each stock is in one of three market capitalization categories: large, medium, and small. T-statistics are calculated using standard errors double-clustered on stock and day for **differences** between high-permanent-volatility days and other days.

Panel A: Permanent Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\kappa^{All}}$	**bps. / $10000**	0.61	10.48	5.85	5.62
(t-stat) for diff. between hi and other days		(1.92)	(7.43)	(-0.55)	(1.26)
$\boldsymbol{\sigma^2(\widetilde{HFT}^{All})}$	**$10000**	2.72	0.61	0.20	1.24
$\boldsymbol{(\kappa^{All} * \sigma(\widetilde{HFT}^{All}))^2}$	$\boldsymbol{bps.^2}$	10.62	70.26	243.69	100.02
(t-stat) for diff. between hi and other days		(5.85)	(15.71)	(24.03)	(18.60)
$\boldsymbol{\sigma^2(w_{i,t})}$	$\boldsymbol{bps.^2}$	95.96	784.31	4256.00	1559.21

Panel B: Transitory Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\phi}$		0.49	0.40	0.37	0.47
$\boldsymbol{\psi^{All}}$	**bps. / $10000**	-0.15	-5.67	-10.60	-1.97
(t-stat) for diff. between hi and other days		(-1.26)	(-8.74)	(-3.78)	(-5.97)
$\boldsymbol{\sigma^2(HFT^{All})}$	**$10000**	2.74	0.63	0.23	1.26
$\boldsymbol{(\psi^{All} * \sigma(HFT^{All}))^2}$	$\boldsymbol{bps.^2}$	3.12	22.28	81.98	33.01
(t-stat) for diff. between hi and other days		(4.99)	(14.67)	(21.16)	(17.43)
$\boldsymbol{\sigma^2(s_{i,t})}$	$\boldsymbol{bps.^2}$	10.03	128.81	496.23	194.54

Table 7: State Space Model of $\mathbf{HFT^D}$, $\mathbf{HFT^S}$ and Prices on High-Permanent-Volatility Days.

This table reports the estimate for the state space model for high-permanent-volatility ($\sigma^2(w_{i,t})$) days. High-permanent-volatility days are categorized for each stock when $\sigma^2(w_{i,t})$ is in the 90th percentile for that stock. The model is estimated for each stock, each day, using HFT trading variables to decompose the observable price (midquote) $p_{i,t}$ for stock i at time t (in one-second increments) into two components: the unobservable efficient price $m_{i,t}$ and the transitory component $s_{i,t}$:

$$p_{i,t}= m_{i,t}+ s_{i,t}$$
$$m_{i,t}= m_{i,t-1}+ w_{i,t}$$
$$w_{i,t}= \kappa_i^D \widetilde{HFT}_{i,t}^D + \kappa_i^S \widetilde{HFT}_{i,t}^S + \mu_{i,t}$$
$$s_{i,t}= \phi s_{i,t-1} + \psi_i^D HFT_{i,t}^D + \psi_i^S HFT_{i,t}^S + \upsilon_{i,t}$$

$HFT_{i,t}^D$ and $HFT_{i,t}^S$ are HFT liquidity and supplying order flow; $\widetilde{HFT}_{i,t}^D$ and $\widetilde{HFT}_{i,t}^S$ are the surprise components of those order flows. Each stock is in one of three market capitalization categories: large, medium, and small. T-statistics are calculated using standard errors double-clustered on stock and day for **differences** between high-permanent-volatility days and other days.

Panel A: Permanent Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\kappa^D}$	**bps. / \$10000**	1.62	23.00	108.31	40.48
(t-stat) for diff. between hi and other days		(4.73)	(12.75)	(12.92)	(11.98)
$\boldsymbol{\kappa^S}$	**bps. / \$10000**	-1.59	-29.15	-165.16	-59.34
(t-stat) for diff. between hi and other days		(-6.10)	(-10.83)	(-13.52)	(-11.25)
$\boldsymbol{\sigma^2(\widetilde{HFT}^D)}$	**\$10000**	2.78	0.59	0.19	1.24
$\boldsymbol{\sigma^2(\widetilde{HFT}^S)}$	**\$10000**	1.96	0.26	0.11	0.82
$\boldsymbol{(\kappa^D * \sigma(\widetilde{HFT}^D))^2}$	$\boldsymbol{bps.^2}$	15.18	79.76	270.10	112.80
(t-stat) for diff. between hi and other days		(6.30)	(12.19)	(20.86)	(14.89)
$\boldsymbol{(\kappa^S * \sigma(\widetilde{HFT}^S))^2}$	$\boldsymbol{bps.^2}$	9.72	53.46	257.03	97.77
(t-stat) for diff. between hi and other days		(6.30)	(12.19)	(20.86)	(14.89)
$\boldsymbol{\sigma^2(w_{i,t})}$	$\boldsymbol{bps.^2}$	103.77	804.10	4071.99	1515.94

Panel B: Transitory Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\phi}$		0.61	0.41	0.36	0.51
$\boldsymbol{\psi^D}$	**bps. / \$10000**	-0.50	-11.00	-44.43	-6.54
(t-stat) for diff. between hi and other days		(-3.57)	(-12.95)	(-12.94)	(-12.45)
$\boldsymbol{\psi^S}$	**bps. / \$10000**	0.49	12.86	57.40	10.15
(t-stat) for diff. between hi and other days		(4.13)	(10.27)	(9.73)	(10.18)
$\boldsymbol{\sigma^2(HFT^D)}$	**\$10000**	2.80	0.61	0.21	1.27
$\boldsymbol{\sigma^2(HFT^S)}$	**\$10000**	1.97	0.27	0.12	0.83
$\boldsymbol{(\psi^{Init} * \sigma(HFT^D))^2}$	$\boldsymbol{bps.^2}$	3.54	24.89	88.91	36.13
(t-stat) for diff. between hi and other days		(5.53)	(14.81)	(21.63)	(20.35)
$\boldsymbol{(\psi^{Pass} * \sigma(HFT^S))^2}$	$\boldsymbol{bps.^2}$	2.36	16.62	84.58	31.53
(t-stat) for diff. between hi and other days		(4.83)	(10.78)	(17.45)	(14.27)
$\boldsymbol{\sigma^2(s_{i,t})}$	$\boldsymbol{bps.^2}$	17.16	157.29	593.83	235.89

Table 8: HFT and Subsequent Returns Macroeconomic News Announcement.
This table presents results on HFT trading and subsequent returns around macroeconomic announcements. We report the coefficients from a regression of returns from time t+2 to time t+10 on HFT liquidity demand, liquidity supply, and overall: HFT^{D}, HFT^{S} and HFT^{All} from time t-1 to time t+1 after a macroeconomic announcement becomes publicly available. Time t is the second in which a macroeconomic news announcement is publicly available. $HFT^{D,S,All}_{i,t-1,t+1}$ is the HFT dollar volume difference between buying and selling (Buy – Sell), scaled by 10,000 and $Ret_{i,t+2,t+10}$ is the return in basis points from two seconds after the macroeconomic announcement to ten seconds afterwards.

$$Ret_{i,t+2,t+10} = \alpha + \beta\ HFT^{D,S,All}_{i,t-1,t+1} + \varepsilon_{i,t}$$

Each stock is in one of three market capitalization categories: large, medium, and small. Panel A reports the HFT^{D} results, Panel B the HFT^{S} results, and Panel C the HFT^{All} results. T-statistics are calculated using standard errors clustered by day.

Panel A				
	Large	**Medium**	**Small**	**All**
$\boldsymbol{HFT^{D}_{t-1,t+1}}$	0.08	1.06	1.35	0.08
(t-stat)	(2.03)	(2.26)	(1.99)	(2.05)
Panel B				
$\boldsymbol{HFT^{S}_{t-1,t+1}}$	-0.14	0.23	-4.30	-0.14
(t-stat)	(-4.30)	(0.24)	(-1.36)	(-4.33)
Panel C				
$\boldsymbol{HFT^{All}_{t-1,t+1}}$	0.04	1.00	1.15	0.05
(t-stat)	(1.27)	(2.27)	(1.85)	(1.363)

Table 9: Limit Order Book Imbalance and Subsequent HFT.

This table presents results on HFT trading and lagged limit order book imbalance (LOBI). We report the mean coefficient from a set of OLS regressions conducted for each stock on each trading day. LOBI is defined as:

$$LOBI_{i,t} = (Size_{i,t}^{Offer} - Size_{i,t}^{Bid})/(Size_{i,t}^{Offer} + Size_{i,t}^{Bid});$$

where *Size* is the dollar volume of orders available at the NBBO. *LOBI* is scaled by 10,000. Panel A regresses the return in period *t*+1, Ret_{t+1} on LOBI in period *t*:

$$Ret_{\ i,t+1} = \alpha + \beta\ LOBI_{i,t} + \varepsilon_{i,t},$$

Panels B, C, and D report the regressions with the dependent variable as HFT Demand, Supply, and All, respectively:

$$HFT_{i,t+1}^{D,S,All} = \alpha + \beta\ LOBI_{i,t} + \varepsilon_{i,t}$$

where $HFT_{i,t+1}^{D,S,All}$ is the HFT dollar volume difference between buying and selling (Buy – Sell), scaled by 10,000 in period *t*+1. T-statistics are calculated using standard errors double-clustered on stock and day. Each stock is in one of three market capitalization categories: large, medium, and small.

Panel A: Ret_{t+1}				
	Large	**Medium**	**Small**	**All**
$LOBI_t$	-0.01	-0.01	0	-0.01
(t-stat)	(-16.36)	(-4.35)	(0.01)	(-1.19)
Panel B: HFT_{t+1}^{D}				
$LOBI_t$	-0.19	-0.21	-0.13	-0.18
(t-stat)	(-15.17)	(-7.55)	(-2.05)	(-8.49)
Panel C: HFT_{t+1}^{S}				
$LOBI_t$	0.07	0.05	0.06	0.06
(t-stat)	(5.90)	(1.86)	(1.18)	(3.33)
Panel D: HFT_{t+1}^{All}				
$LOBI_t$	-0.06	-0.11	-0.03	-0.07
(t-stat)	(-12.36)	(-11.81)	(-0.63)	(-4.20)

Figure 1: Correlation of HFT and Subsequent Returns.

This figure plots the correlation between HFT^{D}, HFT^{S}, and HFT^{All} and returns contemporaneously and up to ten seconds into the future in one-second increments.

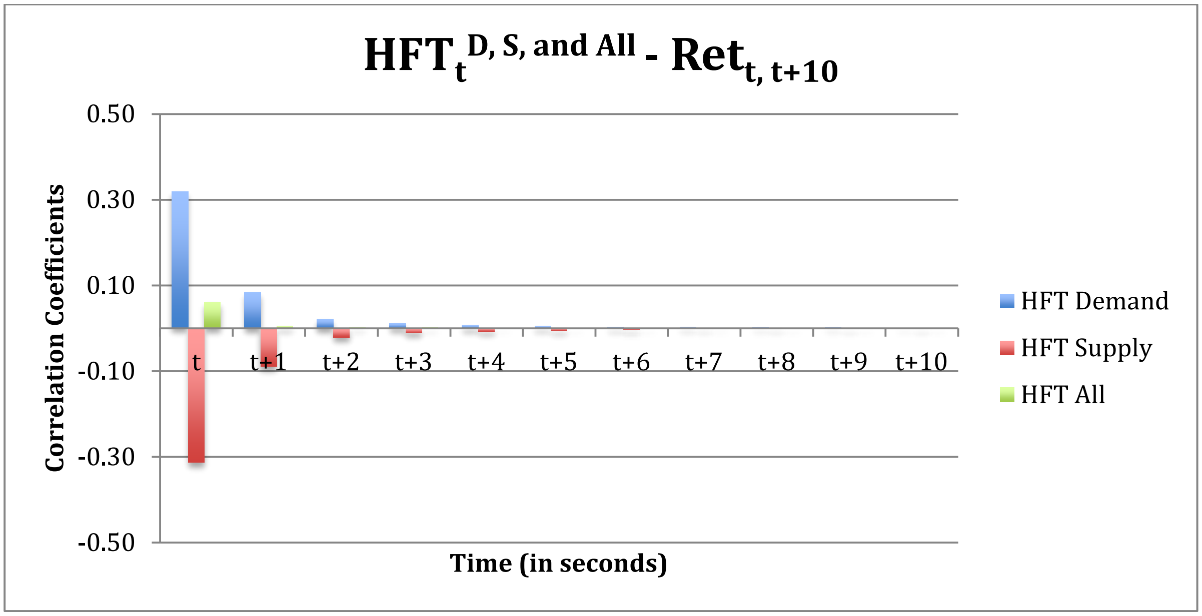

Figure 2: Correlation of Returns and Subsequent HFT.

This figure plots the correlation between returns and HFT^{D}, HFT^{S}, and HFT^{All} contemporaneously and up to ten seconds into the future in one-second increments.

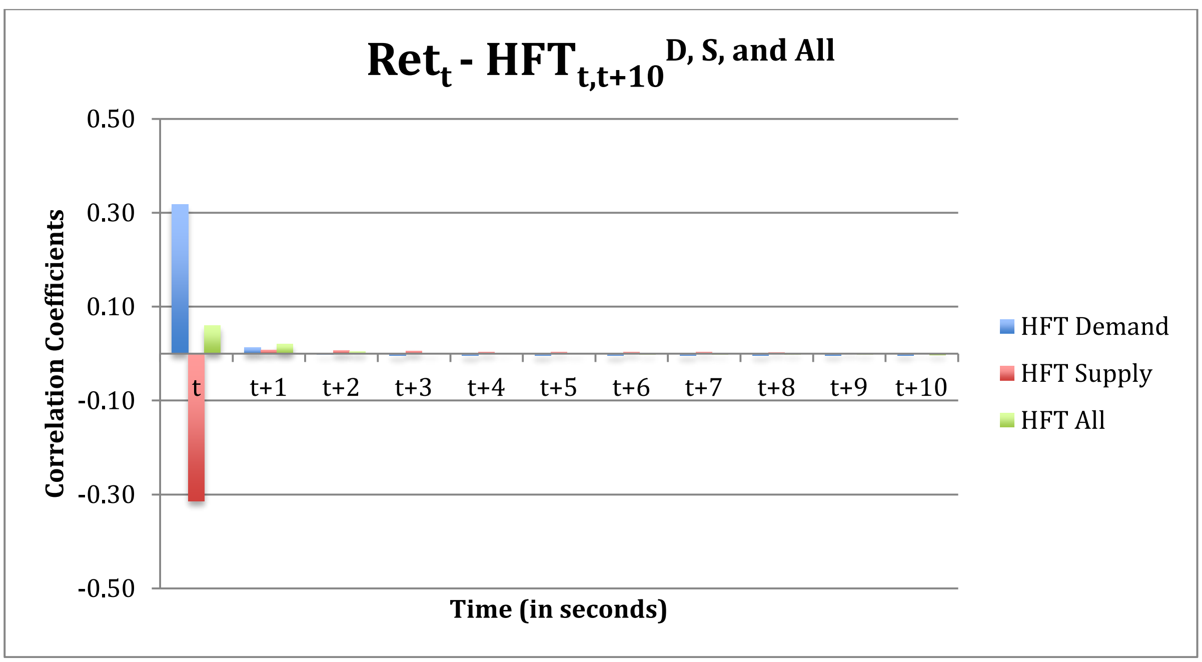

Figure 3: HFT Trading and Portfolio Returns for Positive Macro Announcements. This figure plots the value-weighted sample portfolio return, and HFT^{D}, HFT^{S}, and HFT^{All}. Time is in seconds, and at time t = 0 macroeconomic news is made publicly available. Positive announcements are those above the average analyst forecast.

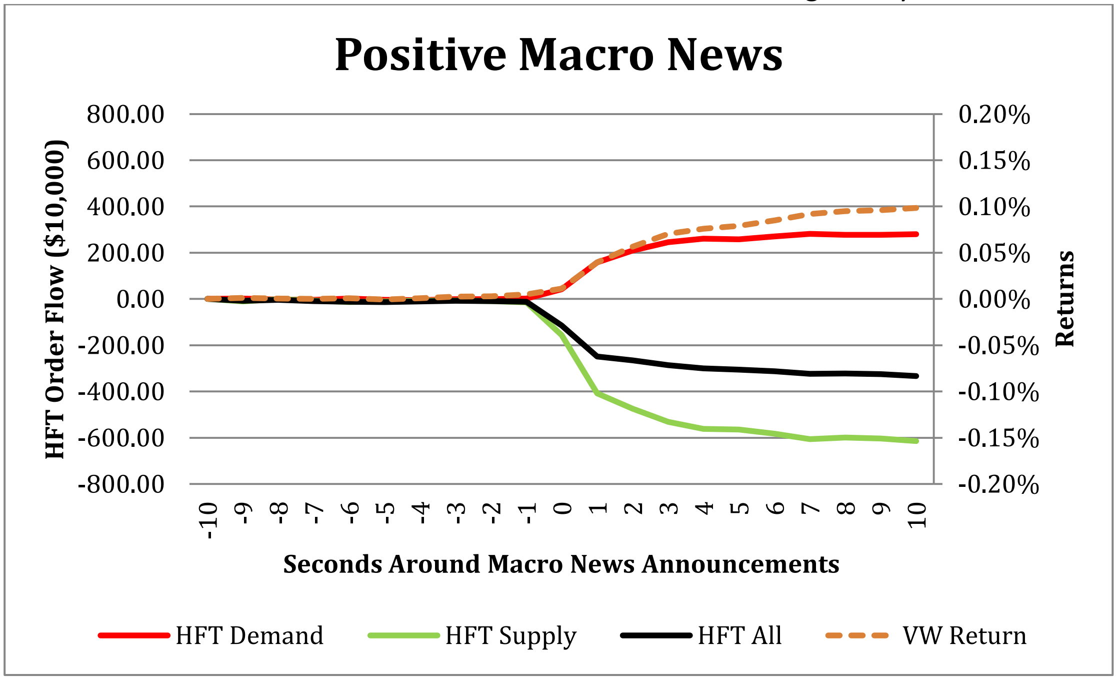

Figure 4: HFT Trading and Portfolio Returns for Negative Macro Announcements. This figure plots the value-weighted sample portfolio return, and HFT^{D}, HFT^{S}, and HFT^{All}. Time is in seconds, and at time t = 0 macroeconomic news is made publicly available. Negative announcements are those below the average analyst forecast.

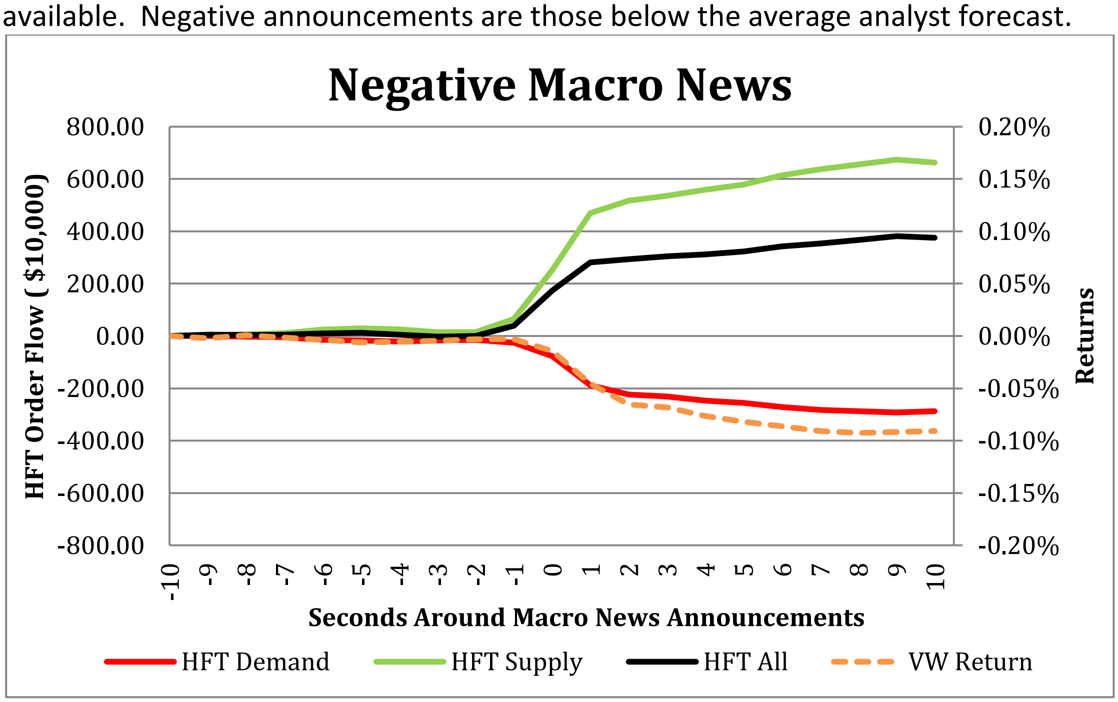

Figure 5: Correlation of Market-wide HFT and Subsequent Market Returns.
This figure plots the correlation between HFT^D, HFT^S, and HFT^{All} aggregated across all stocks in the sample and the value-weighted portfolio return, contemporaneously through ten seconds into the future, in one-second increments.

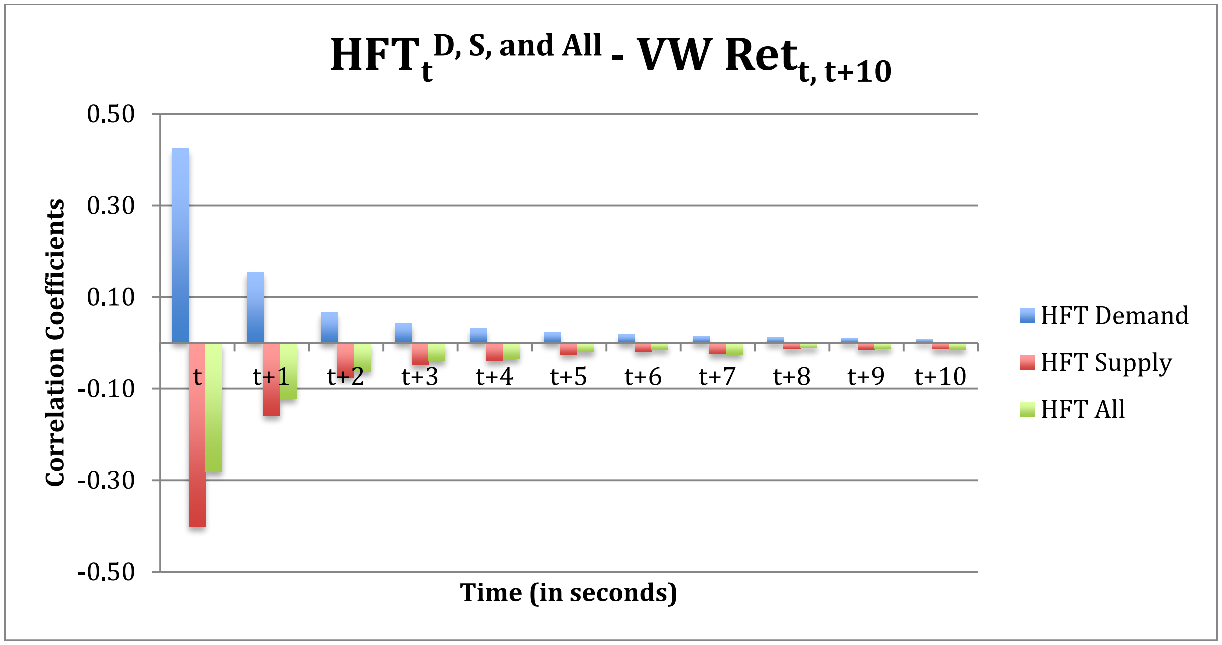

Figure 6: Correlation of Markets Returns and Subsequent Market-wide HFT.
This figure plots the correlation between the value-weighted portfolio return and HFT^D, HFT^S, and HFT^{All} aggregated across all stocks in the sample, contemporaneously through ten seconds into the future, in one-second increments.