High Frequency Trading and Price Discovery*

October 10th, 2011

Terrence Hendershott
(University of California at Berkeley)

Ryan Riordan
(Karlsruhe Institute of Technology)

We examine the role of high-frequency traders (HFT) in price discovery. Overall HFT play a positive role in 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 passive non-marketable orders are adversely selected in terms of the permanent and transitory components as these trades are in the opposite direction as permanent price changes and in the same direction as transitory pricing errors. HFT marketable orders' informational advantage is sufficient to overcome the bid-ask spread and trading fees to generate positive trading revenues. Non-marketable limit orders also result in positive revenues as the costs associated with adverse selection are smaller than the bid-ask spread and liquidity rebates. HFT predicts price changes over short horizons measured in the tens of seconds.

* 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 Álvaro Cartea, Richard Gorelick, Charles Jones, Albert Menkveld, Adam Nunes, Stephen Sapp, and Cameron Smith for helpful comments. All errors are our own.

Historically, financial markets incorporated a group of intermediaries to provide immediacy to outside investors. These intermediaries were often given special status and located on the trading floor of exchanges. Virtually all stock exchanges becoming fully automated (e.g., see Jain (2005)) increased markets' trading capacity and enabled intermediaries to expand their use of technology. This resulted in reduced roles for traditional human market makers and led to the rise of a new class of intermediary, typically referred to as high frequency trader (HFT; HFT is also used for high frequency trading).

Like traditional intermediaries HFTs 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 not available to others. Without such privileges, there is no clear basis for imposing the traditional obligations of market makers (e.g., see Panayides (2007)) on HFT. The substantial, largely negative media coverage of HFT[1] and the so called "flash crash" on May 6th, 2010 raise significant interest and concerns about the role HFT play in the stability and price efficiency of financial markets. 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.

We find that overall HFT play a positive 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 orders and is true on average days and on the most volatile days. In contrast, HFT passive non-marketable orders are adversely selected in terms of the permanent and transitory components as these trades are in the opposite direction as permanent price changes and in the same direction as transitory pricing errors. 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 smaller

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

than the bid-ask spread and liquidity rebates. HFT predicts price changes over short horizons of less than 30 seconds.

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. While each trade pays a narrow bid-ask spread the overall order faces substantial transaction costs. This causes noise in prices due to price pressure arising from liquidity demand by long-term investors. If HFT trades against this transitory volatility it can be viewed as supplying liquidity and reducing long-term investors trading costs. If HFT trades in the direction of transitory volatility (pricing errors) it can be viewed as increasing the costs to those investors. HFT trading in the direction of pricing errors could also arise from attempts to manipulate prices.

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 and passive side of each trade was an 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 link changes in both to HFT. The permanent component is normally interpreted as information and transitory component as transitory volatility or noise.

Overall, HFT's impact is similar to that of other intermediaries and speculators. Speculators can improve price efficiency by getting more information into prices and trading against pricing errors. Manipulative strategies and predatory trading could decrease price efficiency. We find that overall HFT increases price discovery and 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, while our analysis provides some evidence, the counterfactual of how other market participants would behave in the absence of HFT is not known.

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. Section 5 uses a state space model to decompose prices into their permanent/efficient component and transitory/noise component and examines the role of HFT trading in each component. Section 6 relates HFT's role in price discovery to HFT profitability. Section 7 discusses the implications of our findings in general and with respect to social welfare. Section 8 concludes.

2. Literature

This paper fits within the expanding literature on algorithmic trading and HFT.[2] Biais and Woolley (2011) provide background and survey related research. Brogaard (2010) uses the same data as in this paper to examine a number of topics related to HFT, volatility, and trading. O'Hara, Yao, and Ye (2011) also use the same Nasdaq data to examine the role of less than 100 share trades, which are not reported in the public transaction data, in the permanent price discovery process. Our paper differs in that we focus on aggressive and passive HFT trading and the roles both play in the permanent and transitory parts of price discovery. Our results on the permanent price are consistent with Brogaard (2010) and O'Hara, Yao, and Ye (2011).

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.[3] 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. Menkveld (2011) also analyzes the HFT firm's role in the price discovery process along with its profitability and risk taking.

---

[2] Zhang (2010) measures HFT using trading volume relative to institutional portfolio changes in quarterly 13f filings. This measure captures trading frequencies higher than that of long term investors, but identifying the more recent developments in HFT as defined by the SEC (2010) is more difficult.

[3] See Easley, Lopez De Prado, and O'Hara (2010) for analysis of May 6, 2010.

This paper also relates to algorithmic trading, of which HFT is a subset. Hendershott, Jones, and Menkveld (2011) show that algorithmic trading (AT) causally improves liquidity and makes quotes more informative. Chaboud, Chiquoine, Hjalmarsson, and Vega (2009) relate AT to volatility and find little relation. Hendershott and Riordan (2009) find that both AT demanding liquidity and AT supplying liquidity makes prices more efficient. Hasbrouck and Saar (2010) study low-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.

3. Data, Institutional Details, and Descriptive Statistics

The data used in this study are also utilized by Brogaard (2010) and are provided by NASDAQ to academics under a non-disclosure agreement. They consist of trade data for a stratified sample of 120 randomly selected stocks listed on NASDAQ and the NYSE (New York Stock Exchange). 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 which 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.

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

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 the active and passive 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) takes and another HFT supplies the liquidity for a trade. HN trades indicate that an HFT takes and a nHFT supplies liquidity, the reverse is true for NH trades. The remainder of the paper denotes HFT initiating (HH or HN) trades as HFT^{Init} and HFT supplied trades (HH or NH) as HFT^{Pass}. Total HFT trading activity ($HFT^{Init} + HFT^{Pass}$) is labeled as HFT^{All}.

The NASDAQ HFT dataset is supplemented with the National Best Bid Best Offer (NBBO) from TAQ. The NBBO measures the best prices prevailing across all markets to focus on market-wide price discovery. When combining the two data sets two small-cap firms are dropped because their symbols do not appear in TAQ at the beginning of the sample period: BZ and MAKO. Market capitalization data is year-end 2009 data retrieved from Compustat. We focus on continuous trading during normal trading hours by removing opening and closing crosses and trading before 9:30 or after 16:00.

Table 1 reports the descriptive statistics overall and by size category. The average market capitalization of sample firms is $18.2 billion. The range across size categories is high with an average of $52.47 billion in large and $410 million in small. We report average closing prices and volatility of returns. As is typical prices are highest and return volatility is lowest in large stocks with the reverse holding for small stocks.

Insert Table 1 here

We report time-weighted bid-ask spreads in dollars and as a percentage of the prevailing quote midpoint using the TAQ NBBO data sampled at one minute frequencies. On average spreads increase in both dollar and percentage terms from large to small stocks. Relative spreads in small stocks are roughly 10 times higher than for large stocks. Spreads likely play an important role in HFT profitability and behavior, e.g., whether to trade passively or actively. However, spreads calculated based on visible liquidity may overestimate the effective spreads actually paid or received by HFT.

The average per stock NASDAQ daily trading volume is $62.28 million. Trading volume is highest in large stocks at $179 million traded per stock and day and lowest in small stocks with roughly $1.01 million traded per stock and day. The reported trading volume reflects trades occurring during continuous trading on NASDAQ. HFT^{Init} is responsible for roughly 43% of trading volume in large stocks and 23% in small stocks. HFT^{Pass} makes up 41% of trading volume in large and only 10% in small stocks. These numbers confirm the conjecture that HFT is concentrated in large liquid stocks and less in small less 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 order flow (net trading): buy volume minus 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. See Table A3 in the Internet Appendix for the results of augmented Dickey-Fuller (ADF) test for each stock-day. The results of the ADF test suggest that HFT inventory variables may be non-stationary and provide support for the use of order flow rather than inventory levels in the statistical analysis of their trading behavior.

4. Lead-Lag Correlations of Returns and HFT Trading

To better understand the relationship between HFT variables and price changes Table 2 examines the lagged, contemporaneous, and leading correlations of NBBO midquote returns and the HFT order flow variables at ten-second horizons. Correlations are calculated for one lead and lag.

Insert Table 2 here

Consistent with previous studies returns are negatively auto-correlated with a first-order coefficient of correlation of -0.06. The result is increasing across across size categories. The negative return autocorrelation provides evidence that prices contain a transitory component at 10-second frequencies.

HFT order flow is positively auto-correlated overall and all size categories. The contemporaneous correlation between HFT^{Init} and HFT^{Pass} is negative reflecting the fact that HFT^{Pass} often supplies liquidity to HFT^{Init}. This is also evidence that HFT^{Init} and HFT^{Pass} arise from different trading strategies or from different components of strategies being executed actively versus passively, e.g., liquidity-demanding trades initiate positions that may be closed-out via passive limit orders.

The contemporaneous correlation of HFT and returns provide evidence of the relation of their trading to prices at different horizons. HFT^{Init} is positively contemporaneously correlated with returns with a coefficient of 0.22. The positive and large correlation between HFT^{Init} and Rtn_t and the small correlation of HFT_t^{Init} with future returns (Rtn_{t+10}) indicates that HFT^{Init} information is short-lived. These results suggest that when HFT demand liquidity their information is incorporated into prices in less than 1-minute. Figure 1 plots the correlation between HFT order flow and contemporaneous and future returns at 10-second frequencies over a 1-minute period.

Insert Figure 1 here

The figure shows that the correlation between HFT_t^{Init} and Rtn_{t+20} is essentially zero overall and across size categories.

When HFT supply liquidity the relationship is reversed. At the 10-second horizon the correlation between HFT_t^{Pass} and Rtn_t is large and negative with a correlation coefficient of -0.19. The relationship between HFT_t^{Pass} and returns in the subsequent 10-seconds is -0.02, consistent with HFT supplying liquidity being adversely selected. As with initiated HFT trades, the correlation dies off quickly.

Figure 1 shows that the correlation between HFT_t^{Pass} and Rtn_{t+20} is close to zero overall and across size categories. The correlation dies off slightly slower than for HFT_t^{Init}.

The relationship between HFT^{All} and returns is positive contemporaneously and at the 10-second horizon. Overall, the evidence is consistent with HFT playing a role in price discovery at very short horizons. The final plot in Figure 1 shows the correlations for HFT_t^{All} and returns. The figure shows that the correlation between HFT_t^{All} and Rtn_{t+20} is close to zero overall and across size categories.

5. State Space Model of HFT and Prices

The results of the correlation analysis suggest that HFT^{Init}, HFT^{Pass} and HFT^{All} have distinct relationships with prices. To better understand the relationship between the HFT variables, permanent price changes, and transitory price changes we estimate a state space model. 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:

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

The permanent process characterizes information arrivals where $w_{i,t}$ represents the permanent price increments. To capture the overall impact of HFT and the individual impacts of HFT^{Init} and HFT^{Pass}, we formulate and estimate two models. One model incorporates only HFT^{All} while a second includes both HFT^{Init} and HFT^{Pass} variables. Following Hendershott and Menkveld (2010) 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:

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

where $\widetilde{HFT}_{i,t}^{Init}$ and $\widetilde{HFT}_{i,t}^{Pass}$ are the surprise innovations in the corresponding variables calculated analogously to the previous model. A lag length of 60 (five minutes) was used as determined by standard techniques and to be comparable to the 1-minute version of the SSM in the Internet Appendix. 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:

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

The inclusion of $HFT_{i,t}^{Init}$, $HFT_{i,t}^{Pass}$ and $HFT_{i,t}^{All}$ enables measurement of the aggregate and disaggregate roles HFT play in the transitory component of prices. As is standard (and in Hendershott and Menkveld (2010)), the identification assumption is that conditional on the trading variables the innovations in the permanent and transitory components are uncorrelated: $\mathrm{Cov}(\mu_t, \upsilon_t) = 0$.

5.1 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 two state space models enable inference on the overall roles of HFT in prices and the differential role of active and passive HFT. Before presenting the estimation results we discuss how the HFT strategies presented in the SEC concept release impact the estimated state space model coefficients

The predictions on HFT^{Init} and HFT^{Pass} for the permanent component of prices are relatively clear. When utilizing arbitrage and directional strategies we expect HFT^{Init} to be positively correlated with future prices changes. Based on passive market making strategies we generally expect HFT^{Pass} to be negatively correlated with future 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 possible increasing, e.g., arbitrage, or decreasing pricing errors, e.g., manipulation. Whether or not the arbitrage and direction strategies tend to be implemented passively or activity is not specified by the SEC. Finally, the individual strategies are not observable in the data. Therefore, the estimation results will provide evidence on which strategies have the largest effects and do not necessarily demonstrate whether specific strategies are present or not.

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. The costs are directly observable for initiated trades 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 a limit order to sell below the fundamental value (see Amihud and Mendelson (1980), Ho and Stoll (1981), and others).[4] HFT applying price pressure either passively or actively 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^{Init} is straightforward as the decision to trade lies solely with the trade initiator. A positive κ^{Init} is interpreted as HFT^{Init} positively contributing to the discovery of the efficient price. In the transitory equation a positive (negative) ψ^{Init} is interpreted as increasing (decreasing) the noise in prices. Order anticipation and manipulation strategies should result in a positive ψ^{Init}. Identifying pricing errors and profiting from trading against them would yield a negative ψ^{Init}.

The Passive 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^{Pass} would provide evidence that HFT are adversely selected. A positive ψ^{Pass} indicates that HFT is 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 his inventory

[4] 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).

position. This risk management would lead to a positive ψ^{Pass}. Positive ψ^{Pass} could also result from 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 hit them. This is the opposite of momentum ignition whereby the initiating trade is responsible for causing the pricing errors. Section 7 discusses this further.

5.2 State Space Model Estimation

To estimate the state space model for each of the 2,340 10-second time intervals in a trading day for each stock we use the NBBO midquote price, the HFT initiated order flow (dollar buying volume minus selling volume), the HFT initiated passive order flow, and overall HFT order flow (sum of initiated and passive order flows). The state space model is estimated on a stock-day-by-stock-day basis using maximum likelihood via the Kalman filter. There are a number of advantages to using the state space model in our setting. First, the state space model explicitly separates short-term transitory effects from long-term permanent effects. Second, maximum likelihood estimation is asymptotically unbiased and efficient. Third, the state space model offers a structural analysis that helps identify effects that would otherwise be unobserved. After estimation, the Kalman filter offers an in-sample decomposition of the price time series into the efficient and transitory components. The decomposition is available at any point in the sample period using past and current prices.

Our sample contains 118 stocks on 510 trading days. To estimate the state space model we require more than 5 minutes (60 10-second time intervals) where HFT^{Init}, HFT^{Pass} and HFT^{All} are non-zero. We remove observations for all stocks for that day where this is not the case for any stock. After estimating the state space model we end up with 505 days for which we have enough observations. These stock days are used in all analyses for the remainder of the paper. We winsorize estimates of κ and ψ 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 4 reports the results of the HFT^{All} state space model estimation for each size category and overall. Overall we see that HFT^{All} is positively correlated with efficient price changes and negatively correlated with pricing errors. It seems that HFT are able 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 6.61 overall κ coefficient implies that $10,000 of positive HFT order flow (buy volume minus sell volume) is associated with a 6.61 basis point increase in the efficient price. The positive coefficient is consistent with the findings of Brogaard (2010) and O'Hara, Yao, and Ye (2001). The aggregate role of overall HFT in the permanent price process is small: 17.37 basis points in the 10-second permanent price variance of 87.96 basis points. The negative ψ coefficients show that HFT is generally trading in the opposite direction of the pricing errors. This result is statistically significant overall and in the medium and small stocks. The pricing errors are persistent with an AR(1) coefficient between 0.4 and 0.6.

Insert 3 here

In Table 4 we report the results of the disaggregated model of HFT. We include HFT^{Init} and HFT^{Pass} trading variables to better understand the different impacts of both and to provide insight into the trading strategies employed. The two key findings in Panel A on the permanent price component are: (1) HFT initiated trades are correlated with changes in the unobserved permanent price component; (2) HFT passive trades are adversely selected as they are negatively correlated with changes in the permanent price component. The first finding follows from κ^{Init} being positive overall and in each size category. Such relations are typically associated with informed trading. The negative coefficients on κ^{Pass} show that HFT passive trading occurs in the direction opposite to permanent price movements. This relationship exists in models of uninformed liquidity supply where suppliers earn the spread but lose to informed traders.

Insert Table 4 here

As discussed in Section 5.1, the expected relationships between the transitory component of prices and HFT, presented in Panel B of Table 4, are less clear. The state space model estimation results show that ψ^{Init} is negatively related to pricing errors. The negative relation between aggressive HFT trading and the transitory component holds across size categories. This indicates that when HFT initiate trades they trade in the opposite direction to the transitory component of prices, consistent with their trading reducing transitory volatility. 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^{Pass} are positive indicating that HFT^{Pass} trading is associated with a larger transitory component of prices. This is true overall and increasing across size categories. The coefficient is largest for small stocks and smallest for large stocks. One interpretation of the positive ψ^{Pass} is that HFT are adversely selected due to being uninformed about the transitory component price. Passive HFT trading could also be fully aware of the transitory component in prices and be intentionally using price pressure for risk management. Finally, HFT could be attempting to passively manipulate prices in order to trade actively against other market participants. Section 7 discusses these possibilities in more detail.

Aggressive HFT trading is associated with more information being incorporated into prices and smaller pricing errors. It is unclear whether or not the aggressive HFT order submitters know which role any individual trade plays. HFT strategies typically focus on identifying predictability. Whether that predictability arises from the permanent or transitory component is less important. A final finding based on the state space model results is that the permanent component of prices is larger than the transitory component. We find that in the disaggregate model the variance of the permanent component is 335.7 $bps.^2$ and the transitory component is 90.11 $bps.^2$. In relative terms, permanent prices changes are roughly 3.7 times greater than transitory changes.

5.3 State Space Model on High and Low Permanent Volatility Days

The SEC (2010, p.48) expresses 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 permanent volatility 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 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 both 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 information days. Overall total trading volume increases by $32.42 million and by $15.23, $14.61 and $29.86 million for HFT^{Init}, HFT^{Pass} and HFT^{All}. As a percentage of total trading volume HFT^{Init} and HFT^{Pass} increase their participation somewhat. The fact that HFT^{Pass} increases their participation on high-permanent volatility days reveals that HFT continue to supply liquidity in times of market stress. This was a concern highlighted in the SEC concept release.

Table 6 reports the state space model estimates on high-permanent volatility days for the aggregate model. As in Table 3, Panel A reports results for the permanent price component and Panel B for the transitory price component. As in Table 5, statistical inference is conducted on the difference between high-permanent volatility days and other days.

Insert Table 6 here

Comparing Tables 3 and 6 shows that the coefficients in the state space model on high-permanent volatility days all have the same signs and are generally of larger magnitudes than on all days. The difference between high-permanent volatility days and other days are statically significant for most coefficients.

Table 7 presents the results of the disaggregate model's estimates structured as in Table 4. Similar to the aggregate model results we find that the coefficients have the same signs and are larger in magnitude on high-permanent volatility days. The coefficients on HFT^{Init} and HFT^{Pass} for both the permanent and transitory components are roughly two to three times greater than the estimates in Table 6. These show that HFT's role in price discovery is qualitatively similar on high-permanent volatility days, which we interpret as high market stress times.

Insert Table 7 here

6. HFT Revenues

The state space model characterizes the role of HFT trading in the price process. HFT^{Init} gain by trading in the direction of permanent price changes and against transitory changes. HFT^{Pass} lose due to adverse selection and trading in the direction of pricing errors. These gains and losses are before taking into account trading fees and the bid-ask spread. Passive trading earns the spread that aggressive 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, active, and passive HFT. Given that HFT is short-term speculation, it must be profitable or 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 active and passive trading. NASDAQ fees and rebates are taken from the NASDAQ Equity Trader Archive on NasdaqTrader.com. In 2008 and 2009 we identify 6 fee and rebate changes affecting the top volume bracket.[5] Fees for active trades range from \$0.0025 to \$0.00295 per share and rebates for passive trades 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. Distortions in brokers routing decisions based on displayed prices versus prices net of fees may be problematic.

We estimate HFT revenues following Sofianos (1995) and Menkveld (2011). Both analyze primarily passive trading. We decompose total HFT revenue into two components, revenue attributable to HFT^{Init} trading activity and revenue associated with HFT^{Pass} trading activity. We assume that for each stock and each day in our sample, HFT start and end the day without inventories. HFT^{Init} trading revenue for an individual stock for one day is calculated as (each of the *N* transactions within each stock day is subscribed by *n*):

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

[5] 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.

where $INV\_HFT_N^{Init}$ is the daily closing inventory in shares (as opposed to dollars in Table 2) and P_T is the closing quote midpoint. The first term captures cash-flows throughout the day and the second term values the terminal inventory at the closing midquote.[6] $\bar{\pi}^{*Pass}$ is calculated analogously.

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

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

$$\bar{\pi}^{*All} = \bar{\pi}^{*Init} + \bar{\pi}^{*Pass}$$

Table 9 presents the HFT revenue results overall and for initiated and passive trading with and without Nasdaq fees. Panel A provides the average revenue before fees per stock day overall and across size categories. Panel B presents the average revenue after fees per stock day overall and across size categories.

Insert Table 8 here

HFT^{All} earns positive revenues overall and in each size category. In Panel B on a per stock day basis aggressive HFT trading after fees earns \$2,351 overall and \$6,643, \$293 and \$38 in large, medium, and small stocks, respectively. The overall, large, and medium revenue results are all significant at the 1% level using standard errors double clustered on stock and day. HFT earns over 100 times more in large stocks than in small stocks. For one HFT firm Menkveld (2011) also finds significantly higher revenues in larger stocks. Both initiated and passive HFT have positive revenues. Initiated trading's informational advantage is sufficient to overcome the bid-ask spread and fees. Passive trading's informational disadvantage is overcome by revenues from the bid-ask spread and fees.

[6] 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 similiar results.

Given the magnitude of HFT trading volume in Table 1, Panel B of Table 9 suggests that the revenues per dollar traded are low. For large stocks the average initiated HFT trading volume is $77 million per day. Dividing the average day profits of $2,433 by this volume number yields revenues of $0.03 per $10,000 traded. Repeating that calculation for other size categories and passive trading shows that revenue per dollar traded is larger for smaller stocks and is roughly to two to three times higher for passive trading than initiated trading.

As with the analysis of the state space model coefficients, Panels C and D repeat the revenue breakdown on high-permanent volatility days. Initiated revenues are higher on high-permanent volatility days, but the differences are not statistically significant. On high-permanent volatility days passive revenues are higher for medium stocks, but lower for large and small stocks.

7. Discussion

Overall HFT has a beneficial role in the price discovery process in terms of more information getting into prices faster and smaller pricing errors. More informative stock prices can lead to better resource allocation in the economy. However, HFT's information is short-lived at less than 30 seconds. If this information would become public without HFT, then the potential welfare gains may be small. Furthermore, Jovanovic and Menkveld (2011) show how HFT trading on soon-to-be public information could impose adverse selection costs on other investors, leading to less trade and lower welfare. However, as Hasbrouck (1991, p. 190) writes "the distinction between public and private information is more clearly visible in formal models than in practice."

The fact that HFT predicts price movements for only tens of seconds does not demonstrate that the information would inevitably become public. It could also 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 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).

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.

The negative association of overall HFT trading with pricing errors fails to support HFT generally engaging in manipulation. However, passive HFT trading is positively associated with pricing errors. This could be due to adverse selection in the transitory component, risk management, or manipulation. The SEC (2010, p. 53) suggests one passive manipulation strategy: "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."[7] If the limit order is executed before being cancelled, it could result in HFT passive trading 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 passive manipulation could not succeed. While we think risk management is a more plausible explanation for the positive relation between passive HFT trading 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.

[7] 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.

8. Conclusion

We examine the role of high-frequency traders (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 passive 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 that will occur in less than 30 seconds.

One important concern about HFT is their role in market stability.[8] Our results provide no evidence that HFT contribute to market instability in prices. To the contrary HFT overall trades 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 (2008-9).

Our results are one step towards better understanding how HFT affect market structure and performance. Our results are for a single market for a subset of HFT. Better data for both HFT and long-term investors may enable more general conclusions. Studies examining HFT around news announcements, firm's earnings, and other events could provide further identification and understanding. The cross-stock, cross-market, and cross-asset behavior of HFT are also important areas of future research.

HFT are a type of intermediary. When thinking about the role HFT plays in markets it is natural to try to compare the new market structure to the previous 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

[8] 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.

central question is whether there were benefits from the old more highly regulated intermediation sector 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.

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

Brogaard, J. A. (2010). High Frequency Trading and its Impact on Market Quality. *Working Paper.*

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

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. (2010). The Microstructure of the Flash Crash. *Working Paper*.

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. 2010. Price Pressures. *Working Paper*.

Hendershott, T., and R. Riordan. 2009. Algorithmic Trading and Information. *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.*

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.

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, C. Yao, and M. Ye. (2011). What's not there: The odd-lot bias in TAQ data. *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 in 2008 and 2009. Each stock is in one of three market capitalization categories: large, medium, and small. The closing midquote 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^{Init} **Trading Volume**	$ Million	NASDAQ	$77.06	$2.38	$0.26	$26.96
HFT^{Pass} **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: HFT Lead-Lag Correlation Matrix. This table reports the correlation of lagged HFT trading variables and returns. The HFT trading variables are order flow: buy volume minus sell volume. A superscript of Init, Pass, and All corresponds to initiated, passive, and all HFT trading. Correlations are calculated by stock day for 118 stocks for 2008 and 2009. Each stock is in one of three market capitalization categories: large, medium, and small. Correlations are calculated at 10-second frequencies each stock day and then averaged across stock days. Numbers in bold are statistically significantly different from zero at the 1% level using standard errors double clustered on stock and day.

All	$\mathbf{Rtn}_{t-10}$	HFT_{t-10}^{Init}	HFT_{t-10}^{Pass}	HFT_{t-10}^{All}	$\mathbf{Rtn}_{t}$	HFT_{t}^{Init}	HFT_{t}^{Pass}	HFT_{t}^{All}	$\mathbf{Rtn}_{t+10}$	HFT_{t+10}^{Init}	HFT_{t+10}^{Pass}	HFT_{t+10}^{All}
$\mathbf{Rtn}_{t}$	-0.06				1.00				-0.06			
HFT_{t}^{Init}	-0.01	0.03			0.22	1.00			0.02	0.03		
HFT_{t}^{Pass}	0.01	0.01	0.03		-0.19	-0.24	1.00		-0.02	0.00	0.03	
HFT_{t}^{All}	0.00	0.03	0.01	0.04	0.09	0.78	0.36	1.00	0.01	0.02	0.02	0.04
Large	$\mathbf{Rtn}_{t-10}$	HFT_{t-10}^{Init}	HFT_{t-10}^{Pass}	HFT_{t-10}^{All}	$\mathbf{Rtn}_{t}$	HFT_{t}^{Init}	HFT_{t}^{Pass}	HFT_{t}^{All}	$\mathbf{Rtn}_{t+10}$	HFT_{t+10}^{Init}	HFT_{t+10}^{Pass}	HFT_{t+10}^{All}
$\mathbf{Rtn}_{t}$	-0.01				1.00				-0.01			
HFT_{t}^{Init}	-0.03	0.03			0.33	1.00			0.02	0.03		
HFT_{t}^{Pass}	0.02	0.01	0.03		-0.31	-0.36	1.00		-0.03	0.00	0.03	
HFT_{t}^{All}	-0.01	0.03	0.02	0.05	0.10	0.73	0.34	1.00	0.00	0.02	0.03	0.05
Medium	$\mathbf{Rtn}_{t-10}$	HFT_{t-10}^{Init}	HFT_{t-10}^{Pass}	HFT_{t-10}^{All}	$\mathbf{Rtn}_{t}$	HFT_{t}^{Init}	HFT_{t}^{Pass}	HFT_{t}^{All}	$\mathbf{Rtn}_{t+10}$	HFT_{t+10}^{Init}	HFT_{t+10}^{Pass}	HFT_{t+10}^{All}
$\mathbf{Rtn}_{t}$	-0.06				1.00				-0.06			
HFT_{t}^{Init}	0.00	0.03			0.21	1.00			0.03	0.03		
HFT_{t}^{Pass}	0.01	0.01	0.03		-0.15	-0.22	1.00		-0.02	-0.01	0.03	
HFT_{t}^{All}	0.01	0.03	0.01	0.04	0.12	0.83	0.29	1.00	0.02	0.02	0.02	0.04
Small	$\mathbf{Rtn}_{t-10}$	HFT_{t-10}^{Init}	HFT_{t-10}^{Pass}	HFT_{t-10}^{All}	$\mathbf{Rtn}_{t}$	HFT_{t}^{Init}	HFT_{t}^{Pass}	HFT_{t}^{All}	$\mathbf{Rtn}_{t+10}$	HFT_{t+10}^{Init}	HFT_{t+10}^{Pass}	HFT_{t+10}^{All}
$\mathbf{Rtn}_{t}$	-0.10				1.00				-0.10			
HFT_{t}^{Init}	0.01	0.03			0.11	1.00			0.02	0.03		
HFT_{t}^{Pass}	0.00	0.00	0.03		-0.11	-0.14	1.00		-0.02	-0.01	0.03	
HFT_{t}^{All}	0.01	0.02	0.01	0.03	0.04	0.77	0.44	1.00	0.01	0.02	0.02	0.03

Table 3: 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 10 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**	0.22	8.77	10.80	6.61
(t-stat)		(1.29)	(4.76)	(5.59)	(1.30)
$\boldsymbol{\sigma^2(\widetilde{HFT}^{All})}$	**$10000**	4.57	0.37	0.09	1.67
$\boldsymbol{(\kappa^{All} * \sigma(\widetilde{HFT}^{All}))^2}$	**bps.^2**	1.95	14.51	35.80	17.37
(t-stat)		(5.26)	(10.76)	(14.45)	(10.81)
$\boldsymbol{\sigma^2(w_{i,t})}$	**bps.^2**	34.86	70.65	159.10	87.96

Panel B: Transitory Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\phi}$		0.49	0.58	0.52	0.53
$\boldsymbol{\psi^{All}}$	**bps. / $10000**	-0.02	-1.87	-4.04	-1.98
(t-stat)		(-0.55)	(-6.60)	(-3.69)	(-4.88)
$\boldsymbol{\sigma^2(HFT^{All})}$	**$10000**	4.74	0.38	0.09	1.73
$\boldsymbol{(\psi^{All} * \sigma(HFT^{All}))^2}$	**bps.^2**	0.25	1.85	3.98	2.02
(t-stat)		(5.66)	(8.14)	(12.36)	(10.35)
$\boldsymbol{\sigma^2(s_{i,t})}$	**bps.^2**	6.18	27.26	77.04	36.69

Table 4: State Space Model of $\mathbf{HFT^{Init}}$, $\mathbf{HFT^{Pass}}$ 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 10 second increments) into two components: the unobservable efficient price $m_{i,t}$ and the transitory component $s_{i,t}$:

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

$HFT_{i,t}^{Init}$ and $HFT_{i,t}^{Pass}$ are HFT initiated and passive order flow; $\widetilde{HFT}_{i,t}^{Init}$ and $\widetilde{HFT}_{i,t}^{Pass}$ 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
$\boldsymbol{\kappa^{Init}}$	**bps. / $10000**	1.04	14.02	104.53	39.29
(t-stat)		(4.80)	(8.65)	(10.47)	(7.34)
$\boldsymbol{\kappa^{Pass}}$	**bps. / $10000**	-1.87	-18.10	-177.89	-64.96
(t-stat)		(-4.90)	(-7.44)	(-9.83)	(-6.92)
$\boldsymbol{\sigma^2(\widetilde{HFT}^{Init})}$	**$10000**	4.73	0.36	0.08	1.74
$\boldsymbol{\sigma^2(\widetilde{HFT}^{Pass})}$	**$10000**	3.37	0.18	0.04	1.20
$\boldsymbol{(\kappa^{Init} * \sigma(\widetilde{HFT}^{Init}))^2}$	$\boldsymbol{bps.^2}$	4.23	13.26	26.38	14.52
(t-stat)		(11.16)	(12.71)	(14.12)	(12.96)
$\boldsymbol{(\kappa^{Pas} * \sigma(\widetilde{HFT}^{Pass}))^2}$	$\boldsymbol{bps.^2}$	3.10	8.01	26.48	12.41
(t-stat)		(12.41)	(9.38)	(12.03)	(10.23)
$\boldsymbol{\sigma^2(w_{i,t})}$	$\boldsymbol{bps.^2}$	36.18	76.32	175.50	94.70

Panel B: Transitory Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\phi}$		0.28	0.33	0.29	0.30
$\boldsymbol{\psi^{Init}}$	**bps. / $10000**	-0.17	-2.03	-17.73	-6.55
(t-stat)		(-3.18)	(-6.68)	(-9.12)	(-6.77)
$\boldsymbol{\psi^{Pass}}$	**bps. / $10000**	0.41	2.25	25.27	9.17
(t-stat)		(4.18)	(6.49)	(7.97)	(6.24)
$\boldsymbol{\sigma^2(HFT^{Init})}$	**$10000**	4.88	0.37	0.08	1.79
$\boldsymbol{\sigma^2(HFT^{Pass})}$	**$10000**	3.49	0.18	0.04	1.25
$\boldsymbol{(\psi^{Init} * \sigma(HFT^{Init}))^2}$	$\boldsymbol{bps.^2}$	0.43	1.33	3.02	1.58
(t-stat)		(12.70)	(9.08)	(11.48)	(11.09)
$\boldsymbol{(\psi^{Pass} * \sigma(HFT^{Pass}))^2}$	$\boldsymbol{bps.^2}$	0.34	1.04	2.99	1.44
(t-stat)		(14.84)	(8.91)	(11.47)	(10.28)
$\boldsymbol{\sigma^2(s_{i,t})}$	$\boldsymbol{bps.^2}$	5.66	17.18	48.30	23.50

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 a the 1% level using standard errors double clustered on stock and day for.

	Units	Source	Large	Medium	Small	All
Daily Midquote Return Volatility	bps.	TAQ	30.99	47.05	72.50	49.96
Bid-Ask Spread	$	TAQ	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
$\boldsymbol{HFT^{Init}}$ **Trading Volume**	$ Million	NASDAQ	$121.74	$3.31	$0.33	$42.19
$\boldsymbol{HFT^{Pass}}$ **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 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 10 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**	0.10	23.94	16.89	13.71
(t-stat) for diff. between hi and other days		(-0.16)	(4.87)	(0.61)	(2.06)
$\boldsymbol{\sigma^2(\widetilde{HFT}^{All})}$	**$10000**	5.04	0.42	0.09	1.85
$\boldsymbol{(\kappa^{All} * \sigma(\widetilde{HFT}^{All}))^2}$	$\boldsymbol{bps.^2}$	10.91	71.42	130.95	70.83
(t-stat) for diff. between hi and other days		(5.18)	(9.47)	(12.83)	(10.68)
$\boldsymbol{\sigma^2(w_{i,t})}$	$\boldsymbol{bps.^2}$	146.64	299.45	564.91	335.70

Panel B: Transitory Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\phi}$		0.43	0.51	0.45	0.46
$\boldsymbol{\psi^{All}}$	**bps. / $10000**	-0.41	-8.82	-11.06	-6.76
(t-stat) for diff. between hi and other days		(-1.16)	(-5.76)	(-2.31)	(-4.29)
$\boldsymbol{\sigma^2(HFT^{All})}$	**$10000**	5.25	0.44	0.10	1.92
$\boldsymbol{(\psi^{All} * \sigma(HFT^{All}))^2}$	$\boldsymbol{bps.^2}$	1.39	11.36	16.70	9.80
(t-stat) for diff. between hi and other days		(4.32)	(9.12)	(12.19)	(10.52)
$\boldsymbol{\sigma^2(s_{i,t})}$	$\boldsymbol{bps.^2}$	23.03	90.71	157.45	90.11

Table 7: State Space Model of $\mathbf{HFT^{Init}}$, $\mathbf{HFT^{Pass}}$ 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 10 second increments) into two components: the unobservable efficient price $m_{i,t}$ and the transitory component $s_{i,t}$:

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

$HFT_{i,t}^{Init}$ and $HFT_{i,t}^{Pass}$ are HFT initiated and passive order flow; $\widetilde{HFT}_{i,t}^{Init}$ and $\widetilde{HFT}_{i,t}^{Pass}$ 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
$\boldsymbol{\kappa^{Init}}$	**bps. / \$10000**	5.79	35.91	197.11	77.94
(t-stat) for diff. between hi and other days		(-2.79)	(-7.06)	(-9.01)	(-7.07)
$\boldsymbol{\kappa^{Pass}}$	**bps. / \$10000**	-13.41	-59.99	-398.69	-153.97
(t-stat) for diff. between hi and other days		(-3.47)	(-5.41)	(-9.07)	(-6.89)
$\boldsymbol{\sigma^2(\widetilde{HFT}^{Init})}$	**\$10000**	5.07	0.41	0.08	1.89
(t-stat) for diff. between hi and other days		(-1.60)	(-2.29)	(-1.30)	(-1.60)
$\boldsymbol{\sigma^2(\widetilde{HFT}^{Pass})}$	**\$10000**	3.30	0.18	0.04	1.20
(t-stat) for diff. between hi and other days		(-0.45)	(-0.75)	(-2.56)	(-0.10)
$\boldsymbol{(\kappa^{Init} * \sigma(\widetilde{HFT}^{Init}))^2}$	$\boldsymbol{bps.^2}$	20.05	56.03	89.77	54.75
(t-stat) for diff. between hi and other days		(-8.98)	(-11.48)	(-14.03)	(-13.87)
$\boldsymbol{(\kappa^{Pas} * \sigma(\widetilde{HFT}^{Pass}))^2}$	$\boldsymbol{bps.^2}$	16.52	42.44	98.68	51.87
(t-stat) for diff. between hi and other days		(-8.65)	(-7.70)	(-13.13)	(-10.73)
$\boldsymbol{\sigma^2(w_{i,t})}$	$\boldsymbol{bps.^2}$	140.13	311.59	608.98	349.38

Panel B: Transitory Price Component

	Units	Large	Medium	Small	All
$\boldsymbol{\phi}$		0.26	0.27	0.24	0.26
$\boldsymbol{\psi^{Init}}$	**bps. / $10000**	-1.95	-10.60	-50.22	-20.51
(t-stat) for diff. between hi and other days		(-3.66)	(-6.55)	(-9.75)	(-7.87)
$\boldsymbol{\psi^{Pass}}$	**bps. / $10000**	4.31	11.23	79.25	30.93
(t-stat) for diff. between hi and other days		(-4.24)	(-5.41)	(-7.13)	(-6.47)
$\boldsymbol{\sigma^2(HFT^{Init})}$	**$10000**	5.24	0.42	0.09	1.95
(t-stat) for diff. between hi and other days		(-1.66)	(-2.29)	(-1.43)	(-1.64)
$\boldsymbol{\sigma^2(HFT^{Pass})}$	**$10000**	3.44	0.20	0.05	1.25
(t-stat) for diff. between hi and other days		(-0.28)	(-0.92)	(-2.83)	(-0.05)
$(\boldsymbol{\psi^{Init} * \sigma(HFT^{Init})})^2$	$\boldsymbol{bps.^2}$	2.09	7.79	11.79	7.15
(t-stat) for diff. between hi and other days		(-5.93)	(-10.09)	(-10.89)	(-11.71)
$(\boldsymbol{\psi^{Pass} * \sigma(HFT^{Pass})})^2$	$\boldsymbol{bps.^2}$	1.66	5.79	11.06	6.10
(t-stat) for diff. between hi and other days		(-6.00)	(-7.86)	(-11.52)	(-10.31)
$\boldsymbol{\sigma^2}(\mathbf{s_{i,t}})$	$\boldsymbol{bps.^2}$	23.07	61.30	100.03	60.88

Table 8: HFT Revenues.

This table presents results on HFT trading revenue with and without NASDAQ trading fees and rebates. Revenues are calculated for initiated, passive, and all HFT trading: HFT^{Init}, HFT^{Pass} and HFT^{All}. Each stock is in one of three market capitalization categories: large, medium, and small. Panel A reports results per stock day. Panel B reports HFT revenue per stock and day after NASDAQ fees and rebates. Panels C and D report HFT revenues per stock day and with fees and rebates on high permanent-volatility days. *T*-statistics are calculated using standard errors double clustered on stock and day. In Panels C and D the *t*-statistics are based on the **difference** between high permanent-volatility and other days**.**

Panel A: HFT Revenue per stock day

	Large	**Medium**	**Small**	**All**
$\boldsymbol{HFT^{Init}}$	\$7,464.64	\$428.88	\$59.91	\$2,681.16
(t-stat)	(6.89)	(5.38)	(3.77)	(7.30)
$\boldsymbol{HFT^{Pass}}$	\$-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 after fees

	Large	**Medium**	**Small**	**All**
$\boldsymbol{HFT^{Init}}$	\$2,433.43	\$144.52	\$16.18	\$874.54
(t-stat)	(2.27)	(1.83)	(1.02)	(2.41)
$\boldsymbol{HFT^{Pass}}$	\$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 C: HFT Revenue per high-permanent volatility stock day

	Large	**Medium**	**Small**	**All**
$\boldsymbol{HFT^{Init}}$	\$13,090.92	\$1,228.87	\$190.82	\$4,883.63
(t-stat)	(1.13)	(1.43)	(1.27)	(1.28)
$\boldsymbol{HFT^{Pass}}$	\$(3,706.79)	\$155.53	\$(16.78)	\$(1,201.14)
(t-stat)	(-0.53)	(0.56)	(-0.17)	(-0.47)
$\boldsymbol{HFT^{All}}$	\$9,384.13	\$1,384.41	\$174.04	\$3,682.48
(t-stat)	(0.61)	(2.03)	(1.07)	(0.76)

Panel D: HFT Revenue per high-permanent volatility days after fees

	Large	**Medium**	**Small**	**All**
$\boldsymbol{HFT^{Init}}$	\$6,688.54	\$890.09	\$140.32	\$2,597.46
(t-stat)	(0.88)	(1.35)	(1.21)	(1.03)
$\boldsymbol{HFT^{Pass}}$	\$3,960.65	\$380.33	\$9.13	\$1,464.54
(t-stat)	(-0.07)	(0.64)	(-0.12)	(-0.01)
$\boldsymbol{HFT^{All}}$	\$10,649.20	\$1,270.43	\$149.45	\$4,062.01
(t-stat)	(0.63)	(1.99)	(1.05)	(0.78)

Figure 1: Correlation of HFT^{Init} and Returns.
This figure plots the correlation between returns and HFT^{Init}, HFT^{Pass}, and HFT^{All} contemporaneously and 50 seconds into the future in 10 second increments. We report the correlation for large, medium, small stocks and all stocks in our sample.