Available online at www.sciencedirect.com

ScienceDirect

Journal of Financial Markets 16 (2013) 680–711

Journal of FINANCIAL MARKETS

www.elsevier.com/locate/finmar

Very fast money: High-frequency trading on the NASDAQ☆

Allen Carrion*

Lehigh University, Perella Department of Finance, 621 Taylor Street, Bethlehem, PA 18015, USA

Available online 22 June 2013

Abstract

This paper provides evidence regarding high-frequency trader (HFT) trading performance, trading costs, and effects on market efficiency using a sample of NASDAQ trades and quotes that directly identifies HFT participation. I find that HFTs engage in successful intra-day market timing, spreads are wider when HFTs provide liquidity and tighter when HFTs take liquidity, and prices incorporate information from order flow and market-wide returns more efficiently on days when HFT participation is high.

JEL classification: G1; G2

Keywords: High-frequency trading; Trading performance; Intraday return predictability; VWAP; Trading costs; Adverse selection; Market efficiency

1. Introduction

High-frequency trading has become a pervasive feature of the equity markets in a relatively short period of time. Estimates of high-frequency trading activity levels vary, but are large. Consistent with this notion, an identified group of high-frequency traders (HFTs) participates in

☆This paper is based on a chapter of my dissertation at the University of Utah. I wish to thank my committee members Hendrik Bessembinder (co-chair), Michael Cooper (co-chair), Michael Halling, Rachel Hayes, and Kumar Venkataraman for their guidance and generosity with their time. Others who have provided insightful feedback include Shmuel Baruch, Jonathan Brogaard, Patrick Kane, Madhuparna Kolay, Gideon Saar (the editor), Nitish Sinha, Jeff Smith, Chris Stanton, and Laura Tuttle. I have also received helpful comments from seminar participants at the University of Utah, Lehigh University, the FDIC, and the SEC. I thank NASDAQ and Frank Hatheway for supplying the HFT data.

*Tel.: +610 758 6370.

E-mail address: amc312@lehigh.edu

1386-4181/$ - see front matter
http://dx.doi.org/10.1016/j.finmar.2013.06.005

68.3% of the of dollar trading volume in the sample I study in this paper. The developments in market structure (such as decimalization, REG NMS, and automated electronic limit order books) that have created the circumstances for HFTs to flourish are relatively recent. Our understanding of the impact of high-frequency trading on market quality is in its infancy, partly due to its sudden emergence and, until very recently, the lack of high-quality data. There are widely differing views among market participants, regulators, and the financial media on whether HFTs are beneficial, neutral, or detrimental. The disagreements regarding their impact on market quality partly stem from a lack of consensus on the nature of their trading practices. A common view is that HFTs have taken over the market-making function. Under this scenario, they generally benefit the market by increasing competition to provide liquidity, but there are still concerns that they lack the affirmative obligations that bound traditional market-makers and could cause disruptions by exiting the market at their discretion. HFTs are also thought to engage in high-frequency arbitrage, which may have the beneficial effect of making prices more efficient. The alternate perspective is that the liquidity they provide is unreliable, and is outweighed by disruptive practices they are alleged to employ such as order spoofing, predatory trading, herding, or overloading market infrastructure with excessive messages.

I provide evidence on these issues by examining HFT trading and market quality impacts in a sample of NASDAQ trades and quotes that identifies HFT participation. This is the same dataset used in Brogaard (2012) and Brogaard, Hendershott, and Riordan (forthcoming) [BHR (2013) hereafter], but I primarily focus on different research questions and where there is overlap, different empirical strategies are employed that yield additional insights. The first question I address is what are the sources of HFT profitability? I investigate their market timing performance. This is important because it helps characterize their strategies to give insights into their motives for trading, which likely impacts market quality, and also provides evidence on intraday return predictability. My second research question is what trading costs do HFTs face when executing their strategies? This provides additional insights into the sources of their profitability, as well as their decisions on when to supply and demand liquidity. Examining the permanent price impacts of HFT trades also tests theoretical predictions that they impose high adverse selection costs on other traders when demanding liquidity and avoid being adversely selected when providing liquidity. Finally, what impact do HFTs have on market quality? I address this question from the perspective of market efficiency. If HFTs act primarily as liquidity providers or arbitrageurs, we might expect their activity to make prices more efficient, while some of the disruptive strategies they are thought to employ could have the opposite effect.

My main findings are as follows. HFT trading performance as measured in a Volume-Weighted Average Price (VWAP) analysis reflects successful market timing, and this performance is surprisingly strong at longer horizons than might be expected. Trading costs are unconditionally very low in this sample, but spreads are wider on trades where HFTs provide liquidity and tighter on trades where HFTs take liquidity, suggesting that HFTs provide liquidity when it is scarce and consume liquidity when plentiful. I investigate theories that HFTs impose higher adverse selection costs on slower traders and face less adverse selection themselves, and find mixed results that are only significant for specific subsamples and trade types. Prices are more efficient on days when HFTs are more active in a given stock, in the sense that it takes less time for stock prices to incorporate information from order flow and market index returns. This result is driven by HFT liquidity-demanding trades.

These findings should be interpreted with caution. As discussed in more detail below, the sample does not identify the activity of all HFTs, and contains only NASDAQ continuous trading activity in the sample stocks. The sample stocks are traded in multiple venues, and are presumably traded by the sample HFTs in other venues. Also, the NASDAQ exchange is organized as an electronic limit order book with price and time priority, partial pre-trade transparency,[1] post-trade transparency, anonymity, and a maker-taker fee model. It is not clear that any conclusions drawn in this sample will necessarily generalize to markets that are organized differently. These concerns are somewhat mitigated by the facts that the sample contains an economically large amount of trading activity, both in absolute terms and as a share of volume in the sample firms, and the identified HFT firms account for a large share of the observed volume. In addition, although I fail to find evidence of any detrimental effects of HFTs, I can only observe their collective activity and my analysis focuses on their trading and effects aggregated over a variety of market conditions. It is possible that individual HFTs follow disruptive strategies that are hidden by this level of aggregation, or that HFTs collectively have negative impacts in certain market conditions. Nevertheless, this paper should advance our understanding of HFT trading behavior and market quality impacts.

The rest of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 describes the data. Section 4 analyzes HFT trading performance. Section 5 studies trading costs and how they vary with HFT participation. Section 6 presents price efficiency tests. Section 7 concludes.

2. Literature review

In addition to the views of market participants and regulators, there are theoretical reasons to suspect that HFTs may affect market quality. The major sources of trading frictions in the classic market microstructure models are information asymmetry, inventory risk, and order processing costs. HFTs are likely to differ from the intermediaries they have replaced in all of these dimensions. As pointed out in Biais, Foucault, and Moinas (2011) and Jovanovic and Menkveld (2012), the speed advantage of HFTs could allow them to react more quickly to public news than other traders, which would reduce the adverse selection costs they face when providing liquidity while making limit orders riskier for slower traders. Similarly, Stoll (2000) argues that speed differentials play a role in informational frictions, and that increasing the speed parity among traders could reduce spreads under certain conditions. Inventory costs may also play a greater role than in the past. High-frequency traders generally seek to end the day flat. In models such as Garman (1976) and Ho and Stoll (1981), inventory adjustment motives affect liquidity, and recent evidence is supportive (Naik and Yadav, 2003; Panayides, 2007; Comerton-Ford, Hendershott, Jones, Moulton, and Seasholes, 2010). Several studies have shown evidence of market-maker inventory adjustment taking place relatively slowly,[2] and if HFTs manage inventory more aggressively, we might expect the effects on liquidity to increase. Order processing costs should be reduced for HFTs because of their large trading volumes. Rebates for adding liquidity are tiered by volume, and their fixed costs will be spread over more transactions. While the classic microstructure literature has implications for the effects of HFTs, there has also been a recent growth in HFT-specific theoretical literature. Jovanovic and Menkveld (2012) develop a model where the information asymmetry effects can generate either beneficial or

[1]While non-displayed orders are allowed on the NASDAQ, market participants can only observe the displayed limit order book.

[2]Hasbrouck and Sofianos (1993) [HS (1993) hereafter] find cases where inventory takes long periods to revert to apparent target levels. In a more recent sample, Hendershott and Menkveld (2012) reported inventory half-lives of 0.55–2.11 days.

negative impacts, and derive the conditions where each outcome is in effect. Cvitanic and Kirilenko (2010), Biais, Foucault, and Moinas (2011), and Jarrow and Protter (2012) present theoretical models where HFTs can play disruptive roles. The mechanisms include order sniping in Cvitanic and Kirilenko (2010), overinvestment, adverse selection, and the crowding out of slower traders in Biais, Foucault, and Moinas (2011), and a type of herding behavior in Jarrow and Protter (2012).

Despite the emerging theoretical literature and ongoing policy debates concerning HFTs, there has been little empirical research on the behaviors and impacts of HFTs in equity markets until very recently. The empirical studies include Kirilenko, Kyle, Samadi, and Tuzun (2011), Brogaard (2012), Jovanovic and Menkveld (2012), BHR (2013), Hagströmer and Nordén (this issue), Hasbrouck and Saar (this issue), and Menkveld (this issue). Of these, Kirilenko, Kyle, Samadi, and Tuzun (2011) focus on a single extreme event (the 2010 Flash Crash[3]), and Jovanovic and Menkveld (2012) and Menkveld (this issue) study a single HFT, leaving four papers that address the collective behaviors and effects of HFTs across a range of market conditions. Brogaard (2012) studies the same NASDAQ sample with HFT participation identified by the exchange used in my paper, and finds that HFTs provide a large share of the liquidity in the market and dampen volatility, and provides stylized facts on the determinants of their trading behavior. Also using the same dataset, BHR (2013) find that HFTs are an important part of the price discovery process and document stylized facts regarding their trading profits and the determinants of their trading behavior. Hagströmer and Nordén (this issue) study a sample from NASDAQ-OMX Sweden that allows them to track individual HFTs. They find that HFTs follow diverse strategies and identify one group of HFTs that act as market-makers and another group that trades opportunistically. They exploit exogenous changes in HFT activity driven by tick size changes to show that the activity of market-maker HFTs mitigates short-term volatility. Hasbrouck and Saar (this issue) also study recent NASDAQ data and use the intensity of inferred dynamic limit order strategies, which they call strategic runs, to identify periods when HFTs are active in a stock. They find that high-frequency trading “reduces quoted spreads and the total price impact of trades, increases depth in the limit order book, and lowers short-term volatility.”

There is a related thread of empirical studies on algorithmic trading (AT). High-frequency trading is generally considered a subset of AT, but is very different from other types of AT. Hasbrouck and Saar (this issue) explain the distinction clearly. They divide AT into agency algorithms and proprietary algorithms. Agency algorithms are “used by buy-side institutions (and the brokers who serve them) to minimize the cost of executing trades in the process of implementing changes in their investment portfolios.” These can be thought of as engaging in activities such as splitting large orders or alternating between providing and taking liquidity, with the goal of meeting a longer term trading need while minimizing its price impact. Proprietary algorithms encompass the subset of algorithms employed by HFTs. In contrast to the typical users of agency algorithms, HFTs trade their own capital, turnover positions rapidly, have the technology and infrastructure to trade at very high speeds [2–3 milliseconds according to Hasbrouck and Saar (this issue)], and are reluctant to hold inventory overnight. The AT literature does not study high-frequency trading directly, but often touches on related issues or includes HFT trades in AT samples. The empirical AT studies that address market quality issues include Chaboud, Chiquoine, Hjalmarsson, and Vega (2011), Hendershott, Jones, and Menkveld (2011),

[3]The Flash Crash is the popular name for an event that occurred on May 6, 2010, where within a half-hour period, the major U.S. equity indexes dropped more than 5% and quickly reversed most of the losses. Volatility in some ETFs and individual stocks was even greater (Kirilenko, Kyle, Samadi, and Tuzun, 2011).

and Hendershott and Riordan (forthcoming).[4] Chaboud, Chiquoine, Hjalmarsson, and Vega (2011) study AT in foreign exchange markets and find that AT trades contribute less to price discovery than human trades in two of the three currencies in their sample, AT limit orders seem to be placed strategically to face less adverse selection costs, AT reduces liquidity provision before the Nonfarm Payroll (NFP) report and increases it afterwards, and there is some evidence that AT lowers volatility. Hendershott, Jones, and Menkveld (2011) examine market quality measures on the NYSE and find that AT improves liquidity for large capitalization stocks, makes quotes more informative, and reduces the adverse selection costs of trades. They utilize an infrastructure improvement in 2003 to establish causality. Hendershott and Riordan (forthcoming) examine AT in the DAX stocks on the Deutsche Boerse's Xetra platform. They find that algorithmic traders are more likely to demand liquidity when it is cheap and supply when it is expensive, and are faster to react to index returns than human traders. Chaboud Chiquoine, Hjalmarsson, and Vega (2011) and Hendershott and Riordan (forthcoming) study data that identifies algorithmic trader participation, while Hendershott, Jones, and Menkveld (2011) use message traffic as a proxy for AT activity.

3. Data

3.1. Overview

The NASDAQ dataset consists of trades and quotes for a sample of 120 stocks. The stock sample was chosen by Terrence Hendershott and Ryan Riordan. The sample is stratified by market capitalization[5] and is evenly split by NASDAQ and NYSE listing. Table 1 lists the means of selected characteristics and trading volume statistics for the full sample and subsamples by listing exchange and market capitalization.[6] The NASDAQ's share of trading is much higher for the NASDAQ listed stocks than the NYSE listed stocks. This could suggest that the nature of HFT activity on the NASDAQ may be different for the two groups of stocks, a possibility that I allow for in the tests that follow. The sample period covers all of 2008 and 2009 and one week in 2010.[7] The trade sample consists of all trades executed on the exchange in continuous trading, excluding crosses and NASDAQ TRF-reported trades.[8] Trades are time stamped to the millisecond and signed to indicate whether they were initiated by a buyer or seller. The trade signs are high quality, and are based on records of fee and rebate payments used by the exchange. NASDAQ Inside Quotes (BBOs) are provided for subsamples of the data. These subsamples cover the first full trading week in each quarter, the week of September 15–19, 2008 (the week of the Lehman Brothers collapse), and the week of February 22–26, 2010. The BBO data are time stamped to the millisecond and do not have the synchronization problems common in alternate sources. Quotes before 9:30 am, after 4:00 pm, with non-positive or missing ask prices, and with crossed markets are filtered out. The only filter applied to the full trade sample is the removal of trades before 9:30 am and after 4:00 pm. A subsample used for trading cost analysis also requires a quote before and after each trade. Additional filters are applied for some analyses, and specifics are provided in the relevant sections.

[4]There is also a somewhat large AT literature that studies trading algorithms and trading costs for users of algorithms.

[5]With 40 large, 40 medium, and 40 small stocks.

[6]A similar table with individual stock data is provided in the Internet Appendix.

[7]There is one day, October 10, 2008, missing from the dataset.

[8]The FINRA/NASDAQ TRF (Trade Reporting Facility) is a system that reports trades executed in dark trading venues to the Consolidated Tape.

Table 1
Sample stock characteristics and trading volume summary statistics.

Group	N	Market cap (billions)	Price	Avg. dollar trading volume NASDAQ	CT	NASDAQ Share
Full sample	120	17.828	39.59	69.365	201.259	31.5%
NYSE-listed stocks	60	18.347	29.96	32.295	154.916	17.6%
NASDAQ-listed stocks	60	17.309	49.23	106.436	247.602	45.5%
Large cap stocks	40	51.284	66.66	200.086	579.044	32.2%
Mid cap stocks	40	1.796	33.08	6.879	21.264	31.8%
Small cap stocks	40	0.403	19.04	1.130	3.468	30.6%

Sample was selected for NASDAQ by Terrence Hendershott and Ryan Riordan. Sample period is January 2008 – December 2009 and February 22, 2010 – February 26, 2010. Listing venue, price, and market capitalization are from CRSP as of February 26, 2010. Dollar trading volumes are from TAQ for trades between 9:30 am and 4:00 pm and are first averaged over all days in the sample for each stock, then averaged across all stocks in each group.

A unique feature of this dataset is that NASDAQ has manually identified 26 high-frequency trading firms and flagged their activity. Specifically, trades contain a field with the following codes: HH, HN, NH, or NN. H identifies a HFT and N identifies a non-HFT. The first term in the pair classifies the liquidity *taker*, and the second term classifies the liquidity *supplier*. For example, a trade labeled HN would mean an HFT took liquidity from a non-HFT.

The identities of the HFTs are not provided. The selection process was partly manual and somewhat subjective. The principles are described in BHR (2013) as follows: “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.” BHR (2013) and Hasbrouck and Saar (this issue) note that the selection process excludes certain types of firms that engage in high-frequency trading, such as large integrated sell-side firms with proprietary high-frequency trading desks, or smaller HFTs that route trades through direct access brokers. BHR (2013) describe the sample as being composed of independent proprietary trading firms. While the data only identify a subset of HFTs, these firms participate in a large share of the trading volume in the sample (see Section 3.2). The misclassification of some HFTs as non-HFTs biases many of the tests I conduct against finding significant results.

I also obtain supplemental data from CRSP and TAQ. I use CRSP data for the sample stock descriptive statistics only. For several tests I employ midpoint returns, and I consider it preferable to use an NBBO midpoint constructed from the TAQ CQ tape instead of the NASDAQ midpoint. The NBBO includes price data from other market centers, and is available on dates when NASDAQ Inside Quotes are not provided. In addition to the larger sample size available with NBBO quotes, my main considerations in choosing a quote source for a particular application are that my TAQ quotes are only time stamped to the second while NASDAQ quotes are time stamped to the millisecond, and whether I am primarily interested in prices across all markets or liquidity on the exchange where the sample trades occur. I also use TAQ to obtain SPY midpoints to construct a proxy for the market return, and I use trade data from the CT to assess NASDAQ’s volume shares in sample stocks.[9]

[9]SPY is the ticker symbol for an ETF that tracks the S&P 500.

3.2. Descriptive statistics

Table 2 presents trade summary statistics. The second column reports values for the full sample. The full sample covers 509 days and contains 550,118,372 trades for approximately 106 billion shares and a total dollar volume of $3.9 trillion. The daily average share volume in the sample is 208 million shares and the dollar volume is $7.7 billion. There is substantial variation in the daily trading activity. On the 10th percentile day, there is $4.4 billion traded, while on the 90th percentile day, $11.9 billion is traded. The trade size is of particular interest because there is a common perception that trade sizes are much smaller than in the past. They are in fact small in this sample: the average size is 192 shares, the median is 100 shares, and the 90th percentile is 400 shares. The third column reports values for the subsample where matching NASDAQ pre-trade and post-trade quotes are available. This subsample contains 61,272,712 trades for 11.6 billion shares and $444 billion dollars. By comparing the two columns, we can informally assess whether the quote subsample is reasonably representative. The days with quotes have somewhat

Table 2
Trade summary statistics.

Descriptive Statistics	Full Sample	Matched w/quotes
Days in sample	509	49
Number of trades	550,118,372	61,272,712
Total share volume (millions)	105,772	11,642
Total dollar volume (millions)	3,919,037	443,996
Trade size		
Mean	192	190
Std Dev	449	447
10th %ile	50	58
Median	100	100
90th %ile	400	398
Num of trades/day		
Mean	1,080,783	1,250,464
Std Dev	393,491	570,385
10th %ile	691,279	634,906
Median	1,009,167	1,091,299
90th %ile	1,575,009	2,263,314
Daily share volume (millions)		
Mean	208	238
Std Dev	73	97
10th %ile	130	132
Median	197	209
90th %ile	298	396
Daily dollar volume (millions)		
Mean	7,699	9,061
Std Dev	3,147	4,158
10th %ile	4,439	4,537
Median	6,892	7,740
90th %ile	11,912	15,076

Trade and Inside Quote (BBO) data provided by NASDAQ. Trade sample period is January 2008–December 2009 and February 22, 2010–February 26, 2010. Trades are missing on October 10, 2008. Quote sample period is the first full week of each quarter during 2008 and 2009, September 15, 2008–September 19, 2008 (the week of the Lehman Brothers collapse), and February 22, 2010–February 26, 2010. Only trades between 9:30 am and 4:00 pm are used.

more trading activity, but in general appear similar. The subsample covers roughly 10% of the trading days in the full sample, and the aggregate trades, share volume, and dollar volume are around 11% of the full sample values. The daily mean share volume and dollar volume in the subsample are 14% and 18% higher, respectively, than the full sample means. The trade size distributions are very close.

Next, I examine the extent of HFT activity as a share of total dollar trading volume. I construct three measures of HFT participation that differ in how each trade is classified as a HFT or a non-HFT trade. The first counts trades where an HFT participates on either side of a trade (*HFT_ALL*), the second only uses trades where an HFT is the liquidity demander (*HFT_DEMAND*), and the third only uses trades where a HFT is the liquidity supplier (*HFT_SUPPLY*). Trades where HFTs are on both sides are counted in all three measures. The denominator is all trading volume in the NASDAQ sample. The formulas for the three measures are:

$$HFT\_ALL = (HH + HN + NH)/(HH + HN + NH + NN) \quad (1)$$

$$HFT\_DEMAND = (HH + HN)/(HH + HN + NH + NN) \quad (2)$$

$$HFT\_SUPPLY = (HH + NH)/(HH + HN + NH + NN), \quad (3)$$

where each right-hand side term is the dollar volume for the specified counterparty combination.

Table 3 presents summary statistics on the HFT participation variables. Panel A shows that HFTs participate in 68.3% of all dollar trading volume, demand liquidity in 42.2%, and supply liquidity in 41.2% across the full sample. Panel B reports summary statistics for daily participation shares, with trades pooled across all stocks on each sample day. The mean participation shares are similar and little time variation is evident, with standard deviations ranging from 2.4% to 3.6%. These levels are strikingly high and are of a similar order of magnitude to those reported by Brogaard (2012). Panel C summarizes stock-day participation shares, with sample statistics equally-weighted across all stock-days. This removes the extra weight implicitly given to stock-days with more trades in the previous panels. Here we see much lower mean participation levels (48.3%, 32.5%, and 23.2% for *HFT_ALL*, *HFT_DEMAND*, and *HFT_SUPPLY*, respectively), suggesting that HFTs participate more heavily in stock-days with more trading activity. More variability is also evident, with standard deviations from 15.4% to 20.5%. A natural question is whether the variability in HFT participation across stock-days is determined by temporary market conditions or by persistent stock characteristics. To gain some insight into this issue, Panel D summarizes the mean daily participation shares for each stock. This analysis shows that there is substantial variation in long-run mean HFT participation across stocks. For example, the 90th percentile stock has a mean daily *HFT_ALL* share of 72.6%, while the 10th percentile stock has a share of 25.1%. This is consistent with an analysis of HFT participation by stock-day in Brogaard (2012), who finds that some persistent stock characteristics such as market capitalization and market-to-book are determinants of HFT activity.

For some of the tests I conduct later, it is necessary to identify days with high HFT intensity. In light of the observations above, a stock-specific measure that controls for the normal level of HFT activity in that stock is desirable. For each of the three types of HFT participation, I construct indicator variables that take a value of 1 for each stock-day where the dollar volume participation share is in the highest tercile for that stock across all sample days and 0 otherwise. The choice of terciles is somewhat arbitrary, but seems to be a reasonable tradeoff between sample size and extremity. Panel E of Table 3 reports summary statistics on the differences between the mean HFT dollar volume participation shares on days when the indicator variable is one (high participation

Table 3
HFT dollar volume participation shares.

HFT participation definition	*HFT_ALL*	*HFT_DEMAND*	*HFT_SUPPLY*
Panel A: Full sample pooled			
Participation share	68.3%	42.2%	41.2%
Panel B: daily pooling			
N	509	509	509
Mean	68.5%	42.7%	41.1%
Std Dev	2.8%	3.6%	2.4%
10th %ile	65.2%	37.9%	38.2%
Median	68.3%	42.7%	41.1%
90th %ile	72.3%	47.8%	44.1%
Panel C: Stock-day pooling			
N	61,014	61,014	61,014
Mean	48.3%	32.5%	23.2%
Std Dev	20.5%	15.4%	16.8%
10th %ile	19.9%	10.9%	5.5%
Median	49.2%	33.2%	17.9%
90th %ile	75.0%	52.6%	50.5%
Panel D: Stock pooling			
N	120	120	120
Mean	48.3%	32.5%	23.2%
Std Dev	17.7%	11.9%	15.0%
10th %ile	25.1%	15.4%	10.2%
Median	46.4%	34.2%	15.7%
90th %ile	72.6%	47.1%	49.4%
Panel E: Within-stock variation			
N	120	120	120
Mean	16.6%	16.1%	12.5%
Std Dev	5.4%	4.1%	3.6%
10th %ile	9.4%	10.9%	8.2%
Median	16.7%	16.1%	12.0%
90th %ile	23.4%	21.7%	16.9%

HFT participation shares are measured as dollar volume of sample trades with HFT participation divided by total dollar volume of sample trades. Three versions of participation shares are calculated, differing in whether HFT participation is defined as trades where an HFT participates in either side (*HFT_ALL*), the liquidity-demanding side (*HFT_DEMAND*), or the liquidity-supplying side (*HFT_SUPPLY*). Trades where an HFT participates in both sides are used in all three measures. Full Sample Pooled statistics are calculated using trading volume aggregated across the full sample. Daily Pooling statistics use daily participation shares for volume aggregated across all stocks on each sample day, summarized across days. Stock-day Pooling statistics use daily participation shares calculated for each stock, summarized across stock-days. Stock Pooling statistics use daily participation shares calculated for each stock, summarized across stocks. Within-Stock Variation statistics are calculated by first averaging volume for each stock across high HFT participation days and normal HFT participation days separately, taking the differences between the two measures for each stock, then summarizing across stocks. High HFT participation days are days when HFT participation is in its top tercile by stock, other days are classified as Normal HFT participation days. Trade data provided by NASDAQ. Trade sample period is January 2008–December 2009 and February 22, 2010–February 26, 2010. Trades are missing on October 10, 2008. Only trades between 9:30 am and 4:00 pm are used.

days) and other days (normal participation days). In the Internet Appendix, I verify that these variables are well-behaved and suitable for use in the tests that follow.

4. Trading performance

It is important to understand the source of HFT profits for two main reasons. First, it gives insight into their motives for trading and likely impacts on market quality. It is possible that HFTs are simply earning the spread when providing liquidity, and demanding liquidity when necessary to rebalance. If HFTs are instead profiting from market timing, then their trades would be likely to make prices more efficient, yet the apparent liquidity they provide could be overstated. Under that scenario, the liquidity provided by a HFT in the sample trades was only available to counterparties trading against their price forecasts, and was not offered because they perceived the spread to be an adequate incentive to provide liquidity. Second, this analysis provides evidence on the intraday predictability of stock prices. Analogous to the search for signs of longer horizon predictability in the asset manager performance literature, HFT performance and behavior is a natural setting to search for signs of short-term predictability.

4.1. VWAP analysis

In this section, I investigate HFT market timing skills using Volume-Weighted Average Price (VWAP) analysis.[10] This method measures trading performance by comparing the average price obtained on a set of trades of interest to a benchmark based on the average price of an alternate set of trades. It is well-suited to this application because it does not require all of a market participant's trades to generate a valid measure, it does not rely on assumptions about when a trade is reversed or a position is marked, and it does not use benchmarks that are designed from a long -term investor perspective. I calculate the VWAP for various sets of trades as:

$$\text{VWAP} = (1/Vol_{TOTAL}) \sum_{i=1}^{N} (Vol_i)(P_i), \tag{4}$$

where i indexes trades, N is the total number of trades in the set, Vol_i is the number of shares in the ith trade, P_i is the price of the ith trade, and Vol_{TOTAL} is the total number of shares traded in the set of trades. For the main analysis, I calculate the VWAPs of HFT buys, HFT sells, and all sample trades for every stock-day in the sample.[11] I refer to these measures as HFT Buy VWAP, HFT Sell VWAP, and Market VWAP. I also provide a subsample analysis where I calculate these measures separately for HFT liquidity-demanding and liquidity-providing trades and a decomposition analysis where five-minute intervals are used in addition to full trading days.

HFT market timing performance can be assessed by comparing the three VWAP measures. The Market VWAP is traditionally employed as a no-skill benchmark. Berkowitz, Logue, and Noser (1988) refer to this measure as "the price a 'naïve' trader can expect to obtain." If HFTs possess market timing skills, then HFT Buy VWAP will be lower than Market VWAP, HFT Sell VWAP will be higher than Market VWAP, and HFT Buy VWAP will be lower than HFT Sell

[10] VWAP analysis was introduced by Berkowitz, Logue, and Noser (1988) as a measure of broker and money manager trading performance. My implementation is related to the floor trader performance measure from Manaster and Mann (1996) and the traded spread from Stoll (2000).

[11] Sample trades include all continuous market trades executed on NASDAQ between 9:30 am and 4:00 pm, excluding crosses. Trades between two HFTs (type HH trades) are excluded from HFT VWAP measures but included in Market VWAP.

VWAP. A possible objection to the characterization of performance based on these measures as market timing skill is that it will capture liquidity provision as well as what we typically think of as market timing. This is true for the initial analysis, but I address this concern in further tests, first in this section by repeating the analysis for subsamples of liquidity-demanding and supplying trades, and later in Section 5 where the spreads on HFT trades are examined.

The results of this analysis are presented in Table 4. All VWAP differences are signed so a positive number indicates positive trading performance (i.e., Market VWAP—HFT Buy VWAP will be positive if HFTs buy below the Market VWAP). The VWAP and difference calculations are performed daily for each stock and summarized across stock-days, sample days, and stocks. Differences are normalized by dividing by Market VWAP to facilitate aggregation across stocks. Panel A reports summary statistics averaged across all stock-days. The normalized mean of HFT Sell VWAP — HFT Buy VWAP (which I refer to as HFT Sell — Buy VWAP for brevity hereafter) is 6.5 bps. Positive skewness is evident, as the median difference is only 2.3 bps. Buys and sells contribute about equally (3.3 bps below Market VWAP and 3.2 bps above Market VWAP, respectively). This performance is stronger for liquidity providing trades. For liquidity providing HFT trades, the HFT Sell - Buy VWAP is 12.8 bps, while it is only 2.3 bps for liquidity-demanding trades.

In Panel B of Table 4, the VWAP differences for each stock are averaged over each sample day, and the resulting daily values are then averaged to produce a time series of daily measures. The standard deviation and skewness observed in Panel A decrease, which is not surprising because I am essentially creating an equal-weighted portfolio of all the sample stocks every day. The consistency of the HFT liquidity providing trade performance over time becomes apparent. On the 10th percentile day, the mean HFT Sell—Buy VWAP across all stocks is 2.7 bps. In Panel C, the differences are averaged over all sample days for each stock, and then summarized across stocks. These results demonstrate another dimension of HFT liquidity provision performance consistency. In the 10th percentile stock, the HFT Sell—Buy VWAP is 1.9 bps. All of the mean differences indicate positive performance and all are significantly different from 0 at the 5% level or higher, with the exception of the HFT Sell—Market VWAP in Panel C.

These results differ somewhat between NYSE and NASDAQ listed stocks. The most notable difference is that the mean HFT Sell — Buy VWAPs for liquidity-demanding trades are much higher for NYSE stocks and not significantly different from 0 for NASDAQ stocks in any of the three weighting schemes. The other HFT Sell — Buy VWAP differences are qualitatively similar, but point estimates indicate that HFT liquidity-supplying trades in NASDAQ stocks outperform liquidity-supplying trades in NYSE stocks. These results are tabulated and discussed in more detail in the Internet Appendix.

4.2. HFT trading profits

In the preceding analysis, the VWAP differences are presented as measures of observed HFT trading performance relative to benchmarks. It is also possible to use the VWAP differences to estimate HFT trading profits gross of rebates and fees.[12] This requires an assumption regarding the imbalance between observed buy volume and sell volume in a day. It is commonly assumed that HFTs generally end the day flat or close to it, but their daily trading imbalances in the sample often add up to substantial apparent positions at the end of the day. It is not clear to what extent these apparent positions are offset by trades not in this dataset (i.e., not executed in continuous

[12]See BHR (2013) for a discussion of the relationship between trading profits, fees, expenses, and economic profits.

Table 4
HFT VWAP difference summary statistics.

HFT trade type	All			Demand			Supply		
VWAP difference	mkt—buy	sell—mkt	sell—buy	mkt—buy	sell—mkt	sell—buy	mkt—buy	sell—mkt	sell—buy
Panel A: Stock-day weighting									
N	60,692	60,716	60,585	60,260	60,360	59,966	60,084	60,108	59,524
Mean	0.033	0.032	0.065	0.011	0.010	0.023	0.064	0.065	0.128
Std Dev	0.484	0.431	0.560	0.539	0.516	0.659	0.639	0.611	0.813
T	3.26	14.42	22.30	3.57	3.57	6.96	14.43	12.68	25.87
10th %ile	−0.177	−0.179	−0.244	−0.267	−0.268	−0.356	−0.298	−0.292	−0.317
Median	0.011	0.011	0.023	0.002	0.001	0.004	0.023	0.025	0.050
90th %ile	0.269	0.264	0.424	0.305	0.301	0.434	0.469	0.467	0.668
Panel B: Day weighting									
N	509	509	509	509	509	509	509	509	509
Mean	0.033	0.032	0.065	0.011	0.010	0.023	0.064	0.065	0.128
Std Dev	0.056	0.050	0.066	0.071	0.063	0.075	0.100	0.115	0.111
T	13.24	14.38	22.27	3.56	3.55	6.93	14.45	12.71	25.94
10th %ile	−0.017	−0.019	−0.004	−0.053	−0.056	−0.059	−0.018	−0.024	0.027
Median	0.028	0.027	0.058	0.008	0.005	0.020	0.049	0.055	0.108
90th %ile	0.087	0.090	0.147	0.077	0.078	0.109	0.159	0.157	0.257
Panel C: Stock weighting									
N	120	120	120	120	120	120	120	120	120
Mean	0.034	0.032	0.066	0.011	0.009	0.021	0.066	0.067	0.135
Std Dev	0.046	0.033	0.069	0.041	0.047	0.070	0.077	0.062	0.137
T	8.06	10.47	10.50	3.02	1.98	3.35	9.45	11.76	10.80
10th %ile	0.000	0.001	0.003	−0.024	−0.024	−0.031	0.007	0.010	0.019
Median	0.023	0.020	0.052	0.006	0.006	0.013	0.051	0.056	0.111
90th %ile	0.077	0.078	0.139	0.055	0.058	0.095	0.138	0.131	0.258

Differences between VWAP on HFT trades of various categories and same-stock, same-day market VWAP, and differences between VWAP on HFT sells and buys. VWAP differences are scaled by market VWAP and reported as percentages. VWAP differences are signed so that positive numbers indicate positive HFT performance. VWAP differences are first calculated separately for each stock-day and then the stock-day values are summarized with different weightings. Panel A reports summary statistics weighted equally over all stock-days, Panel B reports summary statistics equally weighted by day, and Panel C reports summary statistics equally weighted by stock. *T*-statistics test the null that the mean is 0, and in Panel A use standard errors clustered by day. Trade data provided by NASDAQ. Trade sample period is January 2008–December 2009 and February 22, 2010–February 26, 2010. Trades are missing on October 10, 2008. Only trades between 9:30 am and 4:00 pm are used.

trading on the NASDAQ), and to what extent they are actual overnight positions. For the purpose of estimating HFT profits, I assume the imbalances are offset during the day in unobserved trades (presumably in other trading venues or in the crosses) that are executed at the VWAP of similar HFT trades in the data.[13] Therefore, the estimates should be interpreted as profits to roundtrip strategies where at least one leg is executed on the NASDAQ. I estimate HFT trading profits for every stock-day as follows:

$$\text{HFT Trading Profit} = (\text{HFT Sell}-\text{Buy VWAP}) \times \text{Max}(Vol_{HFT,Buy}, Vol_{HFT,Sell}), \tag{5}$$

where $Vol_{HFT,Buy}$ is the total number of shares bought by HFTs and $Vol_{HFT,Sell}$ is the total number of shares sold by HFTs. I perform this calculation for all HFT trades (with non-HFT counterparties) and for HFT liquidity-demanding and supplying trades separately.

I estimate that the sample HFTs earn trading profits of $2623.84 per stock-day. Considering liquidity-demanding trades and liquidity-supplying trades separately, HFTs lose $691.54 when demanding liquidity and earn $3292.61 when supplying liquidity.[14] Note that the liquidity-supplying and demanding profits do not add up to the total. This is not an error and is due to the effect of liquidity-demanding trades that are offset by liquidity-supplying trades. These are treated as imbalances when analyzing liquidity-demanding and supplying trades separately, but the offsetting effects are recognized when all trades are considered together. These estimates should be interpreted as hypothetical profits under the assumption that the imbalance-offsetting trades were of the same type while ignoring any actual offsetting trades from the other type, rather than a decomposition of the estimated overall profits into liquidity supply and demand components. This approach allows for liquidity-demanding and supplying profit estimates that reflect the price of liquidity in observed HFT trades.

Menkveld (this issue) and BHR (2013) also provide estimates of HFT profits. Menkveld finds that the HFT he analyzes is profitable and earns the bulk of its gross profits from spreads on liquidity-providing trades, but he studies a very different sample so the actual values are not comparable. BHR (2013) studies the same sample and reports trading profits for market capitalization subsamples which, when converted to full sample profits for comparison, would translate to profits of $1990.10 per stock-day for all HFT trades, profits of $2636.35 for liquidity-demanding trades, and losses of $646.26 for liquidity-supplying trades.[15] BHR's all-trade profit estimates are somewhat lower than mine and the contrast between our liquidity-demanding and supplying results is striking. I have identified two methodology choices that explain most of these differences. First, they include trades between HFTs, while I exclude them. This choice is sufficient to make HFT liquidity-demanding trades profitable. An inspection of the HFT Sell — Buy VWAP differences on these trades confirms that when two HFTs trade with each other in this sample, the liquidity-demanding trader tends to outperform the liquidity-supplying trader. Second, on days with an HFT trading imbalance, BHR mark the imbalance at the closing quote

[13]For example, when estimating the profits to all HFT trades, if there is a buy imbalance in the data, the unobserved offsetting sell trades are assumed to execute at the HFT Sell VWAP measured for all HFT sell trades on the same stock-day. Similarly, when estimating the profits to liquidity-demanding HFT trades, a buy imbalance is assumed to be offset by unobserved sell trades that are executed at the HFT Sell VWAP measured for liquidity-demanding HFT sells only. More details are provided in the Internet Appendix.

[14]The mean loss on liquidity-demanding trades is not inconsistent with the positive mean HFT Sell—Buy VWAP of 2.3 bps for these trades reported in Table 4 (using stock-day weighting). This is a result of multiplying the HFT Sell—Buy VWAP by HFT roundtrip trading volume in (5) to estimate the HFT dollar profits for each stock-day, and reflects worse HFT performance in these trades on high volume stock-days.

[15]This conversion assumes a balanced panel, which seems approximately true. In their earlier 2012 working paper that did report full sample profits, this approach came to within roughly 1%.

midpoint, while I mark it at the VWAP of similar HFT trades as described above. The advantage of my method is that it imputes prices on the unobserved HFT trades that adjust for the trading skill and price of liquidity in similar observed HFT trades. Both choices are required to find losses from HFT liquidity-providing trades. I provide a stepwise reconciliation of my results with BHR's and a more detailed description of the two methodologies in the Internet Appendix.[16]

4.3. Market timing decomposition analysis

At what horizon do HFTs have market timing ability? We might expect this ability to be concentrated at the shortest horizons based on their investments in very low-latency technology and various assertions in the media. This relates to questions about the nature of intraday return predictability, whether HFTs are willing to risk their capital on expected price changes that take longer to play out, and also on their potential effects on price formation. I investigate this issue by decomposing the trading performance measures computed above into two components. First, the HFT Sell — Buy VWAP differences summarized in Table 4 are decomposed into shorter term and longer term components. I replace the price on each HFT trade with the Market VWAP for the five-minute interval in which the trade occurred, and recalculate daily HFT Buy and Sell VWAPS for each stock-day using these transformed prices. I then recalculate the HFT Sell — Buy VWAP difference using the adjusted VWAPs, and refer to this as HFT intraday market timing performance. This procedure removes the effects of HFT market timing within five-minute intervals, and leaves only the effect of their choices of how much to buy or sell in a given five-minute interval. I also measure the difference between the total HFT Sell — Buy VWAP and the HFT intraday market timing performance on each stock-day, and call this the HFT short-term timing performance. This can be interpreted as a measure of HFT's ability to time the market within five-minute intervals. If HFT market timing performance is only due to their short-term timing ability, then their intraday market timing performance should be close to zero and the entire HFT Sell — Buy VWAP difference should be allocated to the short-term timing measure, and vice versa.

I demonstrate the decomposition approach with an example. Consider two HFTs (HFT1 and HFT2) who trade the same stock over two five-minute intervals. The market VWAP is 100.00 in the first interval and 102.00 in the second, but within each interval prices higher and lower than the VWAP are available. HFT1 buys 100 shares in the first interval for $99.00, and sells them in the same interval for $101.00, then buys 100 shares for $101.00 in the second interval, and sells them in the same interval for $103.00. HFT2 buys 200 shares in the first interval for $100.00, and sells them in the second interval for $102.00. Both traders would have HFT Buy VWAPs of $100.00, HFT Sell VWAPs of $102.00, and HFT Sell - Buy VWAP differences of $2.00. However, HFT1's performance was driven by trades at briefly available prices and HFT2's performance was driven by buying when prices were low for a sustained period and selling when prices were higher for a sustained period. Replacing the trade prices with the market VWAPs for the intervals in which they traded, HFT intraday market timing performances of $0.00 are calculated for HFT1 and $2.00 for HFT2. Similarly, the HFT short-term timing performances are

[16] I also provide a separate reconciliation between profit estimates for the large cap subsample obtained from my methodology and those reported for the same subsample by BHR. The main implications of my analysis hold in the large cap subsample. The reconciliation is much closer than for the full sample and confirms that the methodology choices identified above explain the substantive differences. These reconciliations also show the robustness of my reported estimates.

$2.00 for HFT1 and $0.00 for HFT2. When I apply this decomposition to the data, if the sample HFTs trade more like HFT1, then more of the HFT Sell - Buy VWAP difference will be attributed to HFT short-term timing performance. If they trade more like HFT2, then more of the HFT Sell — Buy VWAP difference will be attributed to HFT intraday market timing performance.

This decomposition is similar in spirit to the procedure introduced by Hasbrouck and Sofianos (1993) [HS (1993) hereafter] and applied by Menkveld (this issue) and many others, but has important differences. HS (1993) decompose profits into a component attributable to the bid-ask spread and a positioning component.[17,18] This is implemented by replacing the trade prices with pre-trade midpoints and recalculating the total profits. It may superficially appear that the spread component in their approach would be entirely allocated to my short-term timing measure and the positioning component would be very similar to my intraday timing measure, but that is not the case. This would be true if the pre-trade midpoint on every HFT trade was equal to the market VWAP in that five-minute interval. Otherwise, part of the spread component on trades not reversed within same interval could be picked up in the intraday timing measure. The amount of the spread component allocated to the intraday timing measure will depend on multiple factors and is not possible to quantify precisely in this sample due to the incomplete quote coverage, but the liquidity-demanding and supplying trade subsamples can be used to estimate lower and upper bounds on this effect. For liquidity-demanding trades, the intraday timing measure will capture a HS (1993) positioning-like component less some part of the spread component; for liquidity-supplying trades, it will pick up this positioning-like component plus some part of the spread component.[19] The HS (1993) method is not well-suited to this dataset because there is only quote data for a fraction of the sample period and the incomplete set of the sample HFTs' trades precludes reliable inventory calculations.

The results of the decomposition analysis are presented in Table 5. Panel A reports results for stock-day weighting. For each HFT trade type group, the first column reports summary statistics on HFT intraday market timing performance. For all trades, 5.0 bps of the 6.5 bps overall HFT Sell - Buy VWAP is attributable to intraday market timing performance. For liquidity-demanding trades, their intraday market timing performance of 2.8 bps is actually higher than their 2.3 bps HFT Sell - Buy VWAP. For liquidity-supplying trades, 8.1 bps of their 12.8 bps HFT Sell - Buy VWAP is attributable to intraday market timing performance. All of the intraday market timing performance estimates are highly statistically significant. HFT intraday market timing performance also seems to inherit much of the positive skewness from the overall HFT Sell - Buy VWAP difference. For each HFT trade type group, the second column reports summary statistics on HFT short-term timing performance. HFT short-term timing performance for all trades and liquidity-supplying trades is positive, while it is negative for liquidity-demanding trades. All of the short-term timing performance estimates are highly statistically significant. HFT short-term timing performance is also positively skewed, but less so than intraday market timing performance. Panels B and C present the decomposition summarized with day and

[17]There are two other differences to note. HS (1993) work in dollar profits, while my HFT Sell — Buy VWAP measures and components are essentially dollar profits per hundred dollars of HFT roundtrip trading volume. Also, their procedure assumes that the sample contains all of a market participant's trades and handles imbalances as inventories carried across days, while I adjust for incomplete trade data as described above.

[18]HS (1993) refer to the positioning component as "gross quote midpoint trading profits," while Menkveld (this issue) and others use the term "positioning profit." The positioning component is further decomposed by time horizon using spectral analysis.

[19]Even after converting to dollar amounts and ignoring imbalances, this would not be exactly equal to the HS (1993) positioning component because the positioning performance within the 5-minute intervals would be allocated to my short-term timing measure.

Table 5
HFT VWAP difference decomposition.

HFT Trade Type	All		Demand		Supply	
Performance Measure	Intraday market timing performance	Short-term timing performance	Intraday market timing performance	Short-term timing performance	Intraday market timing performance	Short-term timing performance
Panel A: Stock-day weighting						
Mean	0.050	0.015	0.028	−0.005	0.081	0.046
Std Dev	0.558	0.207	0.638	0.146	0.783	0.235
T	17.55	16.94	8.81	−7.38	18.12	33.19
10th %ile	−0.251	−0.031	−0.338	−0.059	−0.358	−0.027
Median	0.016	0.007	0.006	−0.002	0.026	0.020
90th %ile	0.396	0.062	0.431	0.047	0.587	0.136
Panel B: Day weighting						
Mean	0.050	0.015	0.028	−0.005	0.081	0.046
Std Dev	0.064	0.020	0.072	0.015	0.101	0.033
T	17.5	16.87	8.77	−7.38	18.15	31.87
10th %ile	−0.014	−0.001	−0.049	−0.020	−0.010	0.020
Median	0.045	0.012	0.027	−0.005	0.065	0.039
90th %ile	0.125	0.030	0.107	0.009	0.182	0.082
Panel C: Stock weighting						
Mean	0.050	0.016	0.027	−0.006	0.086	0.049
Std Dev	0.060	0.028	0.063	0.018	0.091	0.057
T	9.25	6.32	4.65	−3.41	10.35	9.44
10th %ile	−0.006	0.002	−0.024	−0.016	0.005	0.013
Median	0.040	0.010	0.015	−0.002	0.066	0.032
90th %ile	0.116	0.029	0.099	0.008	0.181	0.108

Decomposition of the differences between VWAP on HFT same-stock, same-day sell and buy trades of various categories into an intraday market timing performance component and a short-term timing component. VWAP differences and components are scaled by market VWAP and reported as percentages. VWAP differences and components are signed so that positive numbers indicate positive HFT performance. The intraday market timing performance is calculated by first replacing the price on each HFT trade with the market VWAP for the same five-minute interval, computing VWAPS for HFT sells and buys using the modified prices for every stock-day, and then taking the difference between the adjusted sell and buy VWAPs. Timing performance is the HFT Sell-Buy VWAP difference calculated using actual trade prices less the intraday market timing performance. *T*-statistics test the null that the mean is 0, and in Panel A use standard errors clustered by day. Trade data is provided by NASDAQ. Trade sample period is January 2008–December 2009 and February 22, 2010–February 26, 2010. Trades are missing on October 10, 2008. Only trades between 9:30 am and 4:00 pm are used.

stock-day weighting, and show very similar results.[20,21] It is noteworthy that HFT short-term timing performance on liquidity-demanding trades is negative. This suggests that, without HFTs' intraday market timing skills, which are negatively affected by spread effects on these trades but

[20]I also perform this analysis in NYSE and NASDAQ subsamples. The results are qualitatively similar. The signs and significance are unchanged. The point estimates indicate that more of the HFT intraday market timing performance comes from liquidity-demanding trades for NASDAQ stocks and more from liquidity-supplying trades for NYSE stocks.

[21]I also conduct an analysis where I rank all 5-minute intervals within each stock-day into 13 groups by market VWAP and examine how HFT activity differs from low-to-high price intervals. The main results are (1) HFTs buy more in the low price periods and sell more in the high price periods, consistent with the Table 5 analysis, and (2) this behavior is relatively continuous across price levels rather than concentrated at the extremes.

still positive and significant, their short-term timing ability in these trades is not sufficient to overcome the bid-ask spread. It is also useful to compare this to a similar analysis of the single HFT in Menkveld (this issue). Menkveld finds that the HFT is only profitable from spreads and in positions that last five seconds or less and consistently loses on positions held longer than a minute. If this was true in my sample, we would expect to see negative intraday timing performance measures for liquidity-demanding trades. The difference is possibly a reflection of differences in market structure or the Hagströmer and Nordén (this issue) finding of heterogeneity in HFT strategies. It is also interesting to note that Menkveld is able to examine the HFT's positions across markets and finds that net long or short positions last "seconds, minutes, and even hours," which is consistent with the possibility of an HFT earning significant intraday market timing profits.

Overall, these results are striking. HFTs would retain most of their market timing ability if they transacted at the market VWAPs for the five-minute intervals in which their trades occur. HFT intraday market timing performance is greater than short-term timing performance for all HFT trade type groups. This even holds for their liquidity-demanding trades, where the intraday market timing performance is expected to be negatively affected by spread effects.

These results suggest that HFTs possess intraday market timing skills, buying when prices are temporarily low and selling when prices are temporarily high. This is consistent with the existence of economically significant predictability in intraday prices. These timing skills are not driven by very short-term signals and are not a result of trading at fleeting prices. Finally, HFT liquidity-providing trades outperform their liquidity-demanding trades. This raises the question of why they engage in so many liquidity-demanding trades. As discussed in Section 3.2, half or more of HFT dollar trading volume (depending on the measurement approach) is liquidity-demanding. It is likely that both types of trades are used together in integrated trading strategies. It is also possible that some of the liquidity-demanding trades are motivated by inventory rebalancing or other risk considerations instead of profits, and that some HFT liquidity-demanding trades are motivated by more time-sensitive information than their liquidity-supplying trades. Note that the results in this section are obtained from the aggregate activity of the sample HFTs, and may mask heterogeneity in the trading strategies and profitability across individual HFTs.

5. Trading costs

In this section I compare the trading costs between trades with HFT participation and those without. The primary metrics I use are effective spreads, permanent price impacts, and realized spreads. All spreads are measured as percentages of the midpoint price prior to the trade, and I report half spreads to reflect one-way rather than roundtrip costs. For this analysis, I use the subsample of trades where both pre- and post-trade quotes are available.

Effective spreads measure the difference between a trade's execution price and the pre-trade midpoint. Effective spreads compensate liquidity providers for adverse selection costs when trading with informed traders [as in Glosten and Milgrom (1985)] and are expected to contain an additional component that covers inventory risk, order processing costs, and market-maker rents. This second component is termed "real friction" by Stoll (2000). An established empirical decomposition method separates the effective spreads into the permanent price impact (adverse selection component) and realized spread (real friction). See Huang and Stoll (1996) and Bessembinder and Kaufman (1997a,b) for a discussion of this methodology and examples of its implementation. The following formulas are used on every trade where quotes are available:

$$\text{Effective Spread} = 100Q(P–M_0)/M_0 \tag{6}$$

$$\text{Permanent Price Impact} = 100Q(M_T–M_0)/M_0 \tag{7}$$

$$\text{Realized Spread} = 100Q(P–M_T)/M_0 = \text{Effective Spread} - \text{Permanent Price Impact}, \tag{8}$$

where Q is a trade sign indicator variable equal to 1 for buys and -1 for sells, P is the trade price, M_0 is the pre-trade quote midpoint, and M_T is midpoint T minutes after the trade. Reported decompositions are computed with T set to 1 minute, and untabulated robustness tests use 15 seconds, 30 seconds, 5 minute, and 30 minute. The last midpoint of the regular trading hours is used when trades are within T minutes of the close. Aside from the traditional interpretations of this decomposition, there are additional reasons it is of particular interest when combined with the HFT identification. If HFTs systematically profit from compensation in the spread, we should observe high realized spreads on their liquidity-providing trades. Otherwise, if HFTs profit from these trades it must be through some other mechanism, such as rebates or superior exit timing (i.e., beating the 1-minute benchmark used in the decomposition). If the realized spreads on these trades are much higher than on those where others provide liquidity, this suggests that HFTs have skill in choosing when to offer liquidity to the market. When taking liquidity, if HFTs are trading on information we should observe high permanent price impacts, while if they are simply re-balancing we should not.

Means and medians of the spread and permanent price impacts are reported in Table 6. These are tabulated for the full sample and for all counterparty type combinations in the data. For the full sample, the mean effective spreads are 2.7 bps, the mean permanent price impacts are 3.9 bps, and the realized spreads are −0.9 bps. These trading cost measures are strikingly low compared to historical estimates. For example, Bessembinder (2003) finds mean effective spreads of 28.9 bps and realized spreads of 17.2 bps in his post-decimalization NASDAQ sample.[22] Many other studies have noted reductions in trading costs over time (e.g., Chordia, Roll, and Subrahmanyam, 2008, 2011; Angel, Harris, and Spatt, 2011), so the low costs in this sample are not entirely unexpected. It is surprising, however, that mean realized spreads are negative for the full sample and all counterparty combinations, and medians are negative in the full sample and negative or zero for all counterparty categories.[23] This means that effective spreads do not fully compensate the liquidity provider for adverse selection costs. It does not necessarily mean that liquidity providers lose money to informed traders on average, because the absolute values are small and at least partially offset by liquidity rebates. It is also possible that some liquidity providers are able to beat the 1-minute post-trade benchmarks built into these measures, which is suggested by the market timing analysis in Section 4.[24] Nevertheless, it does mean that the compensation for liquidity provision is very low based on these widely-used measures. I offer two possible explanations. First, it is possible that increased competition between liquidity providers has driven compensation for liquidity provision down to a level close to the liquidity rebate. Second, it is possible that order submission strategies have evolved in such a way that a large proportion of the trades are between traders seeking liquidity with varying degrees of patience or price sensitivity (Hasbrouck and Saar, 2009), rather than trades between

[22]Originally reported as roundtrip spreads, converted to half spreads here to facilitate comparison with my results.

[23]This observation holds for realized spreads based on 15-second, 30-second, 5-minute, and 30-minute decompositions.

[24]Using the realized spread to estimate liquidity supplier profits also ignores that if a roundtrip can be completed with two liquidity-providing trades, then the effective spread will be earned twice while the price impact is paid only once.

Table 6
Mean and median spread and permanent price impact summary.

			1-minute decomposition	
Category	*N*	Effective spread	Perm price impact	Realized spread
Panel A: Mean spreads and permanent price impacts				
All	61,272,712	0.027	0.036	−0.009
HH	11,631,186	0.023	0.035	−0.012
HN	14,837,559	0.021	0.034	−0.013
NH	19,581,587	0.028	0.032	−0.004
NN	15,222,380	0.035	0.042	−0.007
Panel B: Median spreads and permanent price impacts				
All	61,272,712	0.022	0.025	−0.002
HH	11,631,186	0.023	0.029	−0.016
HN	14,837,559	0.018	0.024	−0.010
NH	19,581,587	0.024	0.023	0.000
NN	15,222,380	0.023	0.024	0.000

All spreads and permanent price impacts are measured as a percent of the pre-trade midpoint. Spreads are reported as 1-way or half-spreads. Trades signs are provided by NASDAQ based on payments to liquidity providers. Uses trade subsample where both a pre-trade and post-trade midpoint are available. The first letter in each trade category label refers to the liquidity taker and the second refers to the liquidity provider. H signifies an HFT; *N* signifies a non-HFT.

an impatient liquidity demander and a liquidity provider largely motivated by the compensation in the spread.

To compare trading costs in trades with HFT participation to those without, I regress these measures of trading costs on indicator variables that capture whether a HFT participated in a trade. I utilize stock-day-half-hour fixed effects to control for stock characteristics and market conditions within half-hour intervals, and include control variables for various trade characteristics. Variations of the following specification are used:

$$\begin{aligned} SPREAD_{itn} = \alpha_{it} &+ \beta_1 HFT + \beta_2 (HFT \times MEDIUM) + \beta_3 (HFT \times LARGE) \\ &+ \beta_4 (HFT \times BUY) + \beta_5 MEDIUM + \beta_6 LARGE + \beta_7 BUY + \varepsilon, \end{aligned} \tag{9}$$

where i indexes stocks, t indexes day-half-hour intervals, and n indexes trades. *SPREAD* is an effective spread, permanent price impact or realized spread. *HFT* is an indicator variable equal to 1 if a trade had HFT participation and 0 otherwise. Different versions of the model define HFT participation by trade side (liquidity-demanding or liquidity-supplying). The interaction terms allow the effects of HFT participation to vary with trade characteristics. *MEDIUM* and *LARGE* are trade size indicator variables. *MEDIUM* trades are defined as at least 500 but less than 1000 shares, and *LARGE* trades are 1000 shares or more. BUY indicates that the buyer took liquidity in the trade. The coefficient on *HFT* is of primary interest and can be interpreted as the difference in the spreads or price impacts between trades with HFT participation (of the specified type) and those without after controlling for the other explanatory variables. In specifications including all interaction terms, the HFT indicator variable captures the trading cost differences for trades of less than 500 shares or sell trades within each stock-day-half-hour, and the coefficients on the HFT indicator variables interacted with characteristics are additional differences on HFT trades above that for small HFT trades or sells. I estimate the regressions using the fixed-effects estimator and cluster the standard errors within day-half-hour intervals following Arellano (1987) and Gormley and Matsa (2013).[25] The fixed-effects estimator (or within-group transformation) implements the fixed effects by demeaning the dependent and independent variables within each group and estimates the model on the transformed data with pooled OLS. It is possible to compute clustered standard errors directly from (9) on the untransformed data in principle, but the fixed-effects estimator is less computationally intensive when many indicator variables are required.

Table 7 reports the results of the regressions. The HFT participation indicator is defined based on the liquidity-demanding side of the trade in Panel A and the liquidity-supplying side of the trade in Panel B. The dependent variable is the effective spread in Models 1–4, the permanent price impact in Models 5–8, and the realized spread in Models 9–12. The first pair of models in each group (Models 1–2, 5–6, 9–10) omit the interaction terms, constraining the effect of HFT participation to be constant across trades, while the second pair (Models 3–4, 7–8, 11–12) allows the effect of HFT participation to vary with trade size or trade direction. The first model in each group includes only the HFT participation indicator and stock-day-half-hour fixed effects, while the other models include trade characteristic controls. The coefficient estimates on Models 1–2 in Panels A and B show that effective spreads are 0.7 bps tighter on trades where an HFT demands liquidity and 0.3 bps wider on trades where an HFT supplies liquidity. The controls add very little explanatory power beyond the fixed effects. These results suggest that HFTs provide liquidity when it is scarce and consume liquidity when it is plentiful. The interaction terms show that this effect is smaller for medium and large trades, and is absent for large liquidity-supplying trades. All estimates on the HFT indicator variables are significant at the 1% level for the effective spread regressions. Models 5–8 in Panels A and B show the results for permanent price impacts, which are generally less statistically significant and arguably economically insignificant. The exception is that there is evidence suggesting that HFTs face lower adverse selection costs than non-HFTs when supplying liquidity in larger trades. The coefficient estimates on Models 9–10 in Panels A and B show that realized spreads are 0.8 bps tighter than on similar trades when HFTs demand liquidity and 0.4 bps wider for trades when HFTs supply liquidity. This confirms that the results observed in the effective spread regressions are not overwhelmed by adverse selection effects, and appear to be strengthened by HFTs' informational advantage when demanding liquidity and their ability to avoid supplying liquidity to informed traders. The size interaction regressions show the same pattern observed for effective spreads when HFTs demand liquidity, and show that their ability to avoid adverse selection on larger trades results in higher realized spreads for trades above 500 shares when supplying liquidity. There is no significant asymmetry in this effect for buys and sells. All estimates on the HFT indicator variables are significant at the 1% level in the constrained realized spread regressions for both HFT demand and supply. In the interaction regressions, the HFT demand indicator variable is significant at the 1% level for small trades and sells, is marginally significant when interacted with *MEDIUM*, and is insignificant elsewhere. The HFT supply indicator variable is significant at the 1% level for small trades and sells, when interacted with *MEDIUM* and *LARGE*, and insignificant when interacted with *BUY*.[26]

[25] Arellano (1987) introduces a method for clustering standard errors by the fixed-effects groups that utilizes the fixed-effects estimator and a cluster adaptation of the White (1980) estimator. Gormley and Matsa (2013) point out the correct degrees of freedom adjustment when extending this approach to handle fixed-effects nested within clusters.

[26] I also perform this analysis in NYSE and NASDAQ subsamples. The results are qualitatively similar for effective spreads and realized spreads, but differ for permanent price impacts. The permanent price impact results are provided in the Internet Appendix.

Table 7
Regression estimates of spreads and permanent price impacts on HFT participation variables and controls.

Panel A: HFT demand participation

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable	Effective spread				Permanent price impact				Realized spread			
HFT	−0.007	−0.007	−0.007	−0.007	0.001	0.001	0.001	0.001	−0.008	−0.008	−0.008	−0.008
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.066)	(0.037)	(0.028)	(0.045)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
HFT × *MEDIUM*			0.001				0.000				0.001	
			(<0.001)				(0.497)				(0.095)	
HFT × *LARGE*			0.001				0.000				0.001	
			(<0.001)				(0.614)				(0.370)	
HFT × *BUY*				0.0005				−0.001				0.001
				(<0.001)				(0.525)				(0.293)
MEDIUM		0.0005	0.0002	0.0005		0.005	0.005	0.005		−0.004	−0.005	−0.004
		(<0.001)	(0.069)	(<0.001)		(<0.001)	(<0.001)	(<0.001)		(<0.001)	(<0.001)	(<0.001)
LARGE		0.001	0.001	0.001		0.008	0.008	0.008		−0.007	−0.008	−0.007
		(<0.001)	(0.026)	(<0.001)		(<0.001)	(<0.001)	(<0.001)		(<0.001)	(<0.001)	(<0.001)
BUY		−0.0003		−0.001		0.003		0.003		−0.003		−0.004
		(<0.001)		(<0.001)		(0.501)		(0.466)		(0.452)		(0.394)
R^2	27.60%	27.60%	27.60%	27.60%	2.44%	2.44%	2.44%	2.44%	1.37%	1.38%	1.38%	1.38%

Panel B: HFT supply participation

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable	Effective spread				Permanent price impact				Realized spread			
HFT	0.003	0.003	0.004	0.003	−0.001	−0.001	0.000	−0.002	0.004	0.004	0.004	0.005
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.011)	(0.041)	(0.297)	(0.014)	(<0.001)	(<0.001)	(<0.001)	(<0.001)

HFT × *MEDIUM*			−0.002				−0.004				0.002	
			(<0.001)				(<0.001)				(<0.001)	
HFT × *LARGE*			−0.004				−0.008				0.004	
			(<0.001)				(<0.001)				(<0.001)	
HFT × *BUY*				0.0004				0.002				−0.002
				(<0.001)				(0.065)				(0.120)
MEDIUM		0.001	0.002	0.001		0.004	0.006	0.004		−0.003	−0.004	−0.003
		(<0.001)	(<0.001)	(<0.001)		(<0.001)	(<0.001)	(<0.001)		(<0.001)	(<0.001)	(<0.001)
LARGE		0.003	0.004	0.003		0.008	0.011	0.008		−0.006	−0.007	−0.006
		(<0.001)	(<0.001)	(<0.001)		(<0.001)	(<0.001)	(<0.001)		(<0.001)	(<0.001)	(<0.001)
BUY		−0.0004		−0.001		0.003		0.002		−0.003		−0.002
		(<0.001)		(<0.001)		(0.499)		(0.678)		(0.445)		(0.570)
R^2	27.41%	27.41%	27.41%	27.41%	2.44%	2.44%	2.44%	2.44%	1.36%	1.37%	1.36%	1.37%

The regression model is:

$$SPREAD_{itn} = \alpha_{it} + \beta_1 HFT + \beta_2(HFT \times MEDIUM) + \beta_3(HFT \times LARGE) + \beta_4(HFT \times BUY) + \beta_5 MEDIUM + \beta_6 LARGE + \beta_7 BUY + \varepsilon,$$

where i indexes stocks, t indexes day-half hours, and n indexes trades. *SPREAD* is an effective spread, permanent price impact, or realized spread. The regression is estimated with stock-day-half hour fixed-effects. *HFT* is an indicator variable that takes a value of 1 if an HFT participated in the trade and 0 otherwise. In Panel A *HFT* is defined using the liquidity-demanding side of the trade; in Panel B *HFT* is defined using the liquidity-supplying side of the trade. *MEDIUM* and *LARGE* are indicator variables that capture trade size. *MEDIUM* indicates a trade size of (500 shares, 1,000 shares), and *LARGE* indicates a trade size≥1,000 shares. *BUY* is a dummy indicating that the trade was buyer-initiated. *P*-values are reported in parentheses and are based on standard errors clustered within half-hour intervals. All regressions use 61,272,712 trade observations.

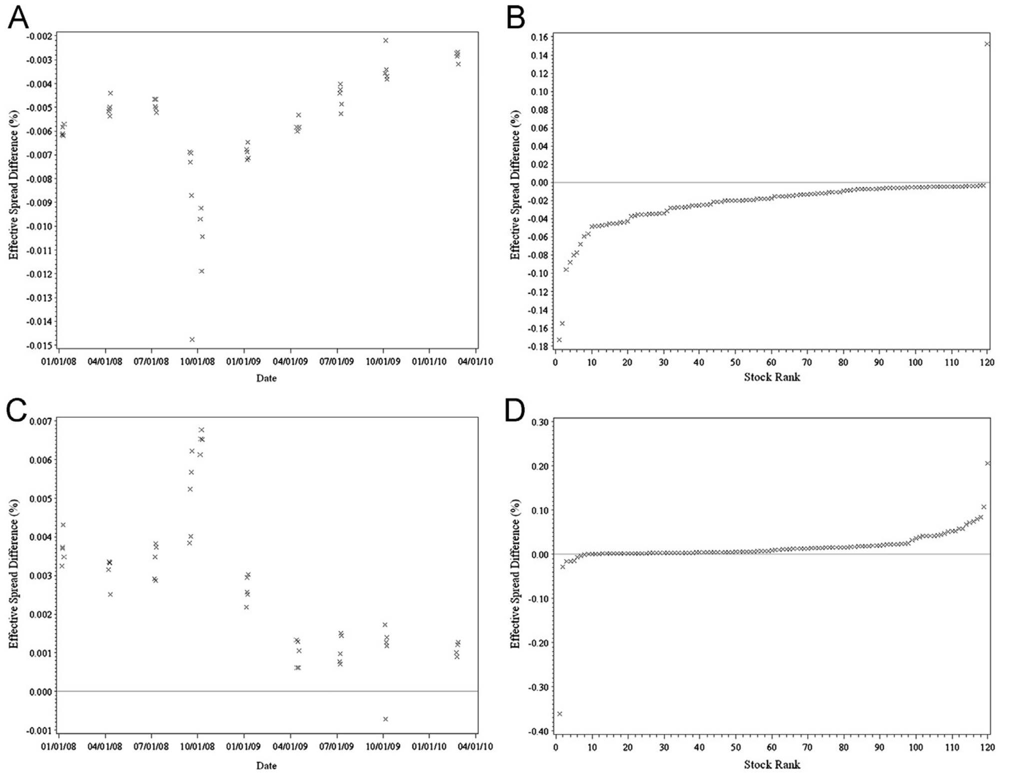

Fig. 1. Time and cross-sectional variation in impact of HFT participation on effective spreads. Panel A: HFT_DEMAND vs. Time, Panel B: HFT_DEMAND vs. Stock Ranking, Panel C: HFT_SUPPLY vs. Time and Panel D: HFT_SUPPLY vs. Stock Ranking. The graphs show coefficients from effective spreads regressed on HFT participation indicator variables and controls estimated one day at a time or one stock at a time. The regression model is:

$$Effective\ \ Spread_{itn} = \alpha_{it} + \beta \times HFT + \text{controls},$$

where i indexes stocks, t indexes day-half hours, n indexes trades, and HFT is an indicator variable that takes a value of 1 when an HFT participates in a trade. HFT is alternately set to *HFT_DEMAND* or *HFT_SUPPLY*, which use participation on the liquidity-demanding or supplying side of the trade respectively. In Panels A and B, HFT participation is defined using the liquidity-demanding side of the trade. In Panels C and D, HFT participation is defined using the liquidity-providing side of the trade. When regression is estimated one stock at a time, stocks are sorted in order of estimated loading on HFT.

Fig. 1 shows the time series and cross-sectional variation in the effect of HFT participation on effective spreads. The plots show the coefficients on the HFT indicator variable in Eq. (9), without interaction terms and using all controls, estimated one day at a time or one stock at a time. From Panels A and C, we see the coefficients do vary over time. For example, in Panel A we see the lowest value for *HFT_DEMAND* is −1.5 bps, and the highest is −0.2 bps. Panels B and D show how the coefficients vary by stock, with stocks sorted by their coefficient estimates. These show relatively little variation except in the tails, and by inspection the tails tend to hold small stocks.[27]

[27]The exception is that the right tail in Panel B of Fig. 1 (higher effective spreads when HFT demand liquidity) is composed primarily of large caps. The one visually obvious outlier is a small cap, however.

It is also of interest to compare the permanent price impact regression estimates with the price discovery analysis in BHR (2013). Understanding the adverse selection in HFT trades is particularly important because one of the detrimental impacts of HFTs predicted by the theoretical literature is that they impose high adverse selection costs on non-HFTs when demanding liquidity, and one of the predicted benefits of HFTs is that they may quote tighter spreads based on their ability to avoid being adversely selected by slower traders [see Biais, Foucault, and Moinas (2011) and Jovanovic and Menkveld (2012) discussed in Section 2]. BHR (2013) find that when demanding liquidity, HFT trades bring information into the market and impose adverse selection on liquidity suppliers more than non-HFTs, and when supplying liquidity themselves are adversely selected, but less so than non-HFTs. The regression models employed in this paper provide an alternate perspective on these questions. These models test whether the permanent price impacts of trades with HFT participation are significantly different from permanent price impacts of other trades, after controlling for the other factors described above. In comparison, BHR uses unexpected HFT and non-HFT trading only, aggregates trading activity at the one-second level, and measures price impact with changes in an efficient price obtained from a state-space model rather than raw quote midpoints. The results in Table 7 from the constrained models show that trades where HFTs demand liquidity do have 0.1 bps higher permanent price impacts than similar trades where they do not, and trades where HFTs supply liquidity have 0.1 bps lower permanent price impacts than similar trades. However, these magnitudes are economically small and the statistical significance is relatively weak considering the number of observations. These results also differ across trade types and stock subsamples. From Model 7 in Panel B we see that the permanent price impacts on small trades where HFTs supply liquidity are not significantly different from those on similar trades where non-HFTs supply liquidity, and small trades are most prevalent in this market. In the Internet Appendix we see that the result of HFTs imposing higher adverse selection costs when demanding liquidity does not hold for the 60 NYSE-listed stocks (in the constrained models, or for small trades, sell trades, and buy trades in the interaction models). On balance, the evidence on the theoretical predictions regarding adverse selection in HFT trades seems mixed. The results are methodology-dependent and weak or absent for small trades and for a large subsample of stocks.

Overall, the results from the regressions confirm the initial observations from the summary statistics. HFT participation explains statistically significant differences in trading cost measures, and HFTs execute their trades at better prices than non-HFTs and have some ability to avoid adverse selection costs on larger trades when supplying liquidity. These results must be interpreted with some caution, however. I cannot assign causality to HFT participation for the differences in trading costs reported. First, it is possible that causality runs in the opposite direction. It is likely that HFTs condition their trading behavior on expected trading costs. Second, even if HFTs do not participate in a given trade, their presence in the market could still affect the cost of that trade through competition or adverse selection.

6. Market efficiency

Pricing efficiency is widely considered to be an important dimension of market quality. Fama (1970) describes an efficient market as one where “security prices at any time ‘fully reflect’ all available information.” Chordia, Roll, and Subrahmanyam (2008) [CRS (2008) hereafter] note that the empirical literature has shown that intraday inefficiencies can exist in markets that are efficient at longer horizons, because it takes investors time to process and react to information.

They further state that "the determinants of this short term predictability deserve a thorough investigation by finance scholars."

It is an open question whether high-frequency trading makes prices more efficient. Theory provides little direct guidance. There is no consensus on how to describe HFT behavior, so it is not clear whether they should be modeled as discretionary market makers, arbitrageurs, predators, or some combination. The HFT-specific models such as Jarrow and Protter (2012) and Jovanovic and Menkveld (2012) describe a variety of mechanisms that could make prices more or less efficient. Empirically, BHR (2013) find that HFTs are an important part of price discovery process and that their abnormal trading is in the direction of removing pricing errors, but this is not equivalent to showing that their activity results in more efficient prices and they do not perform direct efficiency tests on the time series of prices. Also, Brogaard (2012), Hagströmer and Nordén (this issue), and Hasbrouck and Saar (this issue) find some evidence that HFT reduces volatility, which is often informally considered an inverse measure of efficiency. However, total volatility is composed of fundamental volatility and excess volatility. While reducing excess volatility makes prices more efficient, these studies only deal with total volatility. Finally, Hasbrouck and Saar (this issue) find that HFT increases liquidity and CRS (2008) find that liquidity is associated with greater market efficiency.

In this section, I will further investigate this question by comparing the results of direct tests of price efficiency during days with high HFT activity to normal days. A common type of efficiency test measures whether prices are efficient with respect to a specific information set, and I use lagged order imbalances and market returns in this role.[28]

6.1. Order imbalance tests

First, I apply tests inspired by Chordia, Roll, and Subrahmanyam (2005) [CRS (2005) hereafter] and CRS (2008) to examine the incorporation of information from lagged order flows.[29] These tests exploit the concept that efficient prices will follow a random walk, and ex-ante conditioning information will not have explanatory power for future returns. CRS (2005) show that order flow imbalances in individual stocks from one period can predict returns in the next period over some short horizons. CRS (2008) show that the predictive value of lagged order flow imbalance increases on days when liquidity is low, and present a test specification that I adapt to test the effects of HFT on efficiency. The basic form of the model is:

$$R_t = \alpha + \beta_1 OIB_{t-1} + \beta_2 (OIB_{t-1} \mathrm{x} HFT) + \beta_3 MKT_t + \varepsilon, \tag{10}$$

where R_t is the midpoint return calculated from TAQ midpoints, $OIB_{t\text{-}1}$ is the lagged order imbalance, *HFT* is an indicator variable that identifies high-HFT participation days, and *MKT* is the SPY S&P 500 ETF midpoint return calculated from TAQ. I use midpoint returns instead of trade returns because predictability in transaction prices due to bid-ask bounce is not generally considered evidence of informational inefficiency. The HFT indicator in my model replaces the illiquid day indicator variable in CRS (2008), and is defined and discussed in Section 3.2. In different versions of this test, HFT participation is alternately calculated using *HFT_ALL*, *HFT_DEMAND*, or *HFT_SUPPLY*. Following CRS (2008), I use five-minute intervals to measure returns and order imbalances. I also use one-minute intervals because CRS (2008) show

[28] I also use variance ratio analysis in a third unreported set of tests and find qualitatively similar results. I consider the order imbalance and price delay approaches to be more interesting because they provide insight into the specific types of information affecting price formation. The variance ratio results are available upon request.

[29] This efficiency test is also used in Chung and Hrazdil (2010a,b).

that the five-minute horizon predictability has diminished over time, and because HFT effects may be more pronounced at shorter horizons. *OIB* is defined as (Buy Dollar Volume – Sell Dollar Volume)/ Total Dollar Volume. *OIB* is measured over the same interval length as returns. MKT_t is included to reduce the correlation in the residuals across stocks. The regression is estimated one stock at a time, and the time series coefficients are averaged across stocks in a reverse Fama and MacBeth (1973) procedure. *T*-statistics are corrected for correlation in the regression residuals across stocks using the method in CRS (2008). This method adjusts the measured standard errors upwards by $[1+(N-1)\rho]^{1/2}$, where N is the number of individual regressions and ρ is the mean pair-wise correlation across the residuals. If the relationship found in CRS (2005, 2008) holds in this sample, then β_1 will be positive. If the market is more efficient when HFT activity is high, then the sum of β_1 and β_2 will be lower in absolute value than β_1, regardless of whether the CRS finding of a positive β_1 holds.

The results from the order flow imbalance tests are shown in Table 8. The mean coefficients on lagged order imbalance are positive and significant in all models except for the specification using five-minute returns with *HFT_SUPPLY* participation, where it is still positive and marginally significant. The number of stocks with positive and significant coefficients on lagged order imbalance in the individual regressions is higher than the number of stocks with negative significant coefficients, and often much higher. This is consistent with the findings in CRS (2008) for most of their models and subsamples. For both five-minute and one-minute returns, the explanatory power of lagged order imbalance is reduced on high HFT days when HFT participation is defined using *HFT_ALL* or *HFT_DEMAND*. As an illustration, consider the five-minute returns with *HFT_ALL* participation. The mean coefficient on lagged order imbalance is 0.0960. The mean coefficient on lagged order imbalance interacted with the HFT indicator is −0.0704. This means that on high-*HFT_ALL* days, the predicted effect of lagged order imbalance is.0255 (0.0960−0.0704), compared to 0.0960 on normal *HFT_ALL* days, and the *t*-statistic of −2.69 on the coefficient on the interaction term is the test against the null that the difference in lagged *OIB* effect between the high and normal days is 0. There are 37 (out of 120) individual stock interaction coefficients that are significantly negative, while only two are significantly positive. The results for all specifications using *HFT_ALL* or *HFT_DEMAND* participation are qualitatively similar, but are stronger for *HFT_DEMAND* participation and with five-minute returns. For example, the predictive power of order flow is reduced by roughly 20% on high-*HFT_DEMAND* days for one-minute intervals, but it is almost completely removed for five-minute intervals. In both specifications using *HFT_SUPPLY* participation, the mean interaction terms are not significantly different from 0 and there is no strong pattern in individual coefficients.[30]

6.2. Price delay

I employ a second set of efficiency tests using the price delay measures from Hou and Moskowitz (2005). While the CRS tests measure the incorporation of information in past order flow, price delay measures the incorporation of information from market index returns. There are at least two reasons to suspect HFT may affect the incorporation of index return information into individual stock prices.

[30]I also perform this analysis in the NYSE and NASDAQ subsamples and using an alternate intraday definition of HFT participation. The NYSE and NASDAQ subsample results are very similar. The one minute HFT_ALL result becomes marginally significant due to the smaller sample size, and all other results hold. The results with the intraday HFT participation measure are robust for five-minute intervals and insignificant for one-minute intervals. The intraday HFT participation results are in the Internet Appendix.

Table 8
Order imbalance efficiency regressions with HFT participation indicator variable interactions.

HFT participation definition	Variable	Coefficient	t-statistic		Num pos sig	Num neg sig
			Raw	Adjusted		
Panel A: 5-minute returns						
HFT_ALL	Intercept	0.0066	3.34	1.38	9	2
	MKT	825.0872	36.52	15.07	120	0
	$OIB\$_{t-1}$	0.0960	6.62	2.73	59	12
	$OIB\$_{t-1} \times HFT$	−0.0704	−6.51	−2.69	2	37
HFT_DEMAND	Intercept	0.0066	3.33	1.37	9	2
	MKT	825.0813	36.52	15.06	120	0
	$OIB\$_{t-1}$	0.1098	7.42	3.06	59	10
	$OIB\$_{t-1} \times HFT$	−0.1175	−9.73	−4.01	4	57
HFT_SUPPLY	Intercept	0.0066	3.36	1.39	9	2
	MKT	825.1111	36.52	15.07	120	0
	$OIB\$_{t-1}$	0.0651	4.38	1.81	47	24
	$OIB\$_{t-1} \times HFT$	0.0249	2.52	1.04	17	6
Panel B: 1-minute returns						
HFT_ALL	Intercept	0.0004	0.91	0.39	6	2
	MKT	691.5641	31.78	13.60	120	0
	$OIB\$_{t-1}$	0.1047	14.48	6.20	112	1
	$OIB\$_{t-1} \times HFT$	−0.0193	−5.55	−2.37	7	53
HFT_DEMAND	Intercept	0.0004	0.92	0.39	7	2
	MKT	691.5636	31.78	13.60	120	0
	$OIB\$_{t-1}$	0.1066	14.71	6.29	114	1
	$OIB\$_{t-1} \times HFT$	−0.0244	−6.75	−2.89	8	64
HFT_SUPPLY	Intercept	0.0004	0.95	0.41	6	2
	MKT	691.5649	31.78	13.60	120	0
	$OIB\$_{t-1}$	0.0992	13.58	5.81	111	3
	$OIB\$_{t-1} \times HFT$	−0.0024	−0.79	−0.34	19	23

Regressions of 5-minute and 1-minute returns on contemporaneous market returns, lagged order imbalances, and lagged order imbalances interacted with a high-HFT participation day indicator variable. Returns are from TAQ and are calculated using the last midpoint in each interval. SPY ETF returns are used as the market proxy. $OIB\$_{t-1}$ is the dollar value of buyer-initiated trades less the dollar value of seller-initiated trades divided by the total dollar volume during interval $t-1$. A stock is defined as having a high-HFT participation day when its participation share is in its highest tercile for that stock over the entire sample. Participation share is defined as HFT dollar volume divided by the stock's total dollar volume. Three versions of participation shares are calculated differing in whether HFT participation is defined as trades where an HFT participates in either side (*HFT_ALL*), the liquidity-demanding side (*HFT_DEMAND*), or the liquidity-supplying side (*HFT_SUPPLY*). Trades where an HFT participates in both sides are used in all three measures. The regressions are estimated separated for each stock, and cross sectional means of coefficients across all stocks are reported. *T*-statistics test the null that the mean is 0. Adjusted *t*-statistics are corrected for cross-correlation in the residuals. The numbers of positive significant and negative significant coefficients in the individual stock regressions are reported, with significance defined as a *t*-statistic greater than 2 in absolute value. The sample contains 120 stocks. All coefficients are multiplied by 1,000.

First, index returns are a plausible input variable to HFT strategies and index arbitrage is frequently mentioned in informal descriptions of suspected HFT behavior. Second, Jovanovic and Menkveld (2012) find that HFT activity is positively correlated to the explanatory power of the market index for a stock's returns. They attribute this effect to increased HFT activity when hard information has more value, but causality could run the other way as well.

Hou and Moskowitz (2005) refine procedures used earlier by Brennan, Jegadeesh, and Swaminathan (1993) and Mech (1993). While Hou and Moskowitz (2005) use price delay based on weekly data as a stock characteristic in asset pricing tests, I calculate price delay with one-minute and five-minute midpoint returns and employ it as an efficiency measure. To measure price delay, I first estimate the regressions:

$$R_t = \alpha + \beta_1 MKT_t + \delta_1 MKT_{t-1} + \delta_2 MKT_{t-2} \ldots + \delta_6 MKT_{t-6} + \varepsilon \tag{11}$$

$$R_t = \alpha + \beta_1 MKT_t + \varepsilon, \tag{12}$$

where returns are defined as in (10), and six lags of MKT are used. As in the order flow imbalance tests, I use both five-minute and one-minute intervals. These regressions are estimated one stock at a time, separately for high-HFT participation days and normal days. I refer to (11) as the unrestricted model and (12) as the restricted model. Then for each stock I calculate the following price delay measures, separately on high-HFT participation days and normal days:

$$D1 = 1 – (R^2{}_{rest}/R^2{}_{unrest}) \tag{13}$$

$$D2 = (\delta_1 + 2\delta_2 + 3\delta_3 \ldots + 6\delta_6)/(\beta_1 + \delta_1 + 2\delta_2 + 3\delta_3 \ldots + 6\delta_6) \tag{14}$$

$$D3 = (T(\delta_1) + 2T(\delta_2) + \ldots + 6T(\delta_6))/(T(\beta_1) + T(\delta_1) + 2T(\delta_2) \ldots + 6T(\delta_6)), \tag{15}$$

where R^2_{rest} is the R^2 from (12), R^2_{unrest} is R^2 from (11), $T(.)$ is the t-statistic on the coefficient in (11), and other terms are as defined in (11). $D1$ is based on the procedure in Mech (1993) and can be interpreted as the additional explanatory power from the lagged returns as a proportion of the total explanatory power of the unrestricted regression. Coefficient ratios similar to $D2$ and $D3$ were used in Brennan, Jegadeesh, and Swaminathan (1993), but the weightings were introduced by Hou and Moskowitz (2005). $D2$ gives more weight to coefficients on more distant lags of the market return. This serves to increase the measure of price delay when higher coefficients on distant returns indicate that more of the explanatory power in (11) is being driven by returns farther in the past. $D3$ is similar to $D2$, still giving more weight to more distant lags, but adjusts the value of each lagged coefficient based on its estimation precision. Therefore, a large estimated coefficient at a long lag would increase $D2$, but may only slightly increase $D3$ if the standard error of the coefficient estimate is high. Higher values of price delay reflect slower adjustment. For each stock, I calculate each price delay measure separately for high-HFT days and normal days, which are defined as in the order imbalance tests above. Then within each stock I subtract price delay measures on high-HFT days from those on normal days, and average these differences across stocks.

If stock prices incorporate market-wide information more efficiently on days when HFT activity is high, then price delay should be lower on these days and the mean differences will be negative. When testing whether the mean differences are significantly different from zero, it is not clear whether the differences can be considered independent observations. The inputs to the price delay measures are coefficients estimated from regressions on returns in the same sample period. For each stock, two regressions are run in separate subsamples, with different days entering the subsamples for each stock based on stock-specific HFT activity. If there are market-wide mechanisms that cause simultaneous price delays across multiple stocks, then there will be some cross-sectional dependence in the differences because there is correlation between high-HFT participation days across stocks. To correct for this, I use the same standard error adjustment as in the order imbalance tests, calculated from residuals on the unrestricted regression estimated over the full sample (i.e., not divided by the HFT participation category). This is a conservative approach because the sample-splitting procedure should reduce the dependence relative to that in the order imbalance tests.

The results from the price delay tests are shown in Table 9. Price delay is lower on high *HFT_ALL* days and high *HFT_DEMAND* days in all specifications; the results using *HFT_SUPPLY* are insignificant. For *HFT_ALL* and *HFT_SUPPLY*, the price delay difference point estimates are all negative and significant using raw *t*-statistics. Using the conservative *t*-statistics adjustment described above, the five-minute differences become insignificant. One-minute differences are significant at the 10% level for *D*1 with *HFT_ALL* and are insignificant with *HFT_DEMAND*. For *D*2, the differences remain statistically significant at the 5% level in both specifications. For *D*3, the difference is significant at the 10% level using *HFT_ALL* and significant at the 5% level using *HFT_DEMAND*. For *HFT_SUPPLY*, *D*2 and *D*3 differences are significantly negative at five-minute horizons before the adjustment, and none are significant after. Differences are insignificant in other *HFT_SUPPLY* specifications before the adjustment and the point estimates are of mixed signs.[31]

6.3. Interpretation

Overall, these results suggest that prices are more efficient when HFT activity is high. Prices tend to reflect more of the information in past order flows and past market returns on high-HFT activity days, and the effect is stronger when they are demanding liquidity. Based on the evidence presented here, I cannot conclude that HFT activity causes market efficiency increases, only that there is a positive association. However, if HFTs possess comparative advantages in profitably exploiting pricing inefficiencies, it seems unlikely that HFTs choose to trade more and demand liquidity more when the market is more efficient.

The fact that the improvements in measured efficiency are observed primarily when *HFT_DEMAND* is high is relevant to a claim made in CRS (2008). They conjecture that the short-term predictive power of order imbalances is due to the limited ability of market makers to absorb the imbalances without causing price pressure. They argue that liquidity improves efficiency in this setting because arbitrage traders are more likely to trade on this predictability when liquidity is high, and they do so by submitting market orders or marketable limit orders. My results are consistent with a version of this story where HFTs play the role of arbitrageur, and inconsistent with a version where HFTs are enhancing efficiency by improving liquidity.

Price delay reductions on high-HFT days are generally larger at one-minute horizons than at five-minute horizons. This is different from what was observed in the lagged order flow tests. This could be interpreted as supporting the conjecture in Jovanovic and Menkveld (2012) that HFTs trade more aggressively on hard information, including index returns. Their concept of hard information focuses on how quickly incoming information can be used to update quotes.

[31] I also perform this analysis in the NYSE and NASDAQ subsamples. The subsample results are very similar. There is some minor loss of test power due to the smaller sample size. The NYSE subsample has a slightly larger reduction in price delay on high HFT_DEMAND days than the NASDAQ subsample at five-minute horizons, and all other results hold.

Table 9
Comparisons of price delay measures across high and normal HFT participation regimes.

					t-Statistic	
HFT participation definition	PD measure	High	Normal	Diff	Raw	Adjusted
Panel A: Five-minute returns						
HFT_ALL	*D*1	0.030	0.038	−0.008	−3.04	−1.22
	*D*2	0.262	0.375	−0.113	−4.57	−1.83
	*D*3	0.264	0.381	−0.117	−4.46	−1.79
HFT_DEMAND	*D*1	0.030	0.040	−0.010	−3.89	−1.56
	*D*2	0.250	0.382	−0.131	−4.69	−1.88
	*D*3	0.251	0.386	−0.135	−4.54	−1.82
HFT_SUPPLY	*D*1	0.037	0.033	0.003	1.49	0.60
	*D*2	0.301	0.349	−0.047	−2.12	−0.85
	*D*3	0.302	0.354	−0.051	−2.14	−0.86
Panel B: One-minute returns						
HFT_ALL	*D*1	0.068	0.084	−0.016	−4.31	−1.83
	*D*2	0.312	0.445	−0.133	−5.14	−2.18
	*D*3	0.311	0.448	−0.136	−4.71	−2.00
HFT_DEMAND	*D*1	0.070	0.086	−0.016	−2.70	−1.15
	*D*2	0.283	0.468	−0.186	−6.29	−2.67
	*D*3	0.278	0.473	−0.195	−6.01	−2.55
HFT_SUPPLY	*D*1	0.077	0.077	0.000	−0.10	−0.04
	*D*2	0.403	0.389	0.014	0.65	0.28
	*D*3	0.407	0.389	0.018	0.72	0.31

Price Delay measures use regressions of a stock's return on contemporaneous and lagged market returns (unrestricted regression) compared to regressions on contemporaneous returns only (restricted regression) to measure the speed with which market information is incorporated into the stock's price. D1 is derived from R^2 from the restricted and unrestricted regressions. D2 uses ratios of lagged coefficients to all coefficients and gives more weight to longer lags. D3 is similar to D2 but uses t-statistics instead of coefficients, down-weighting less precise estimates. Higher values indicate greater delays. Six lags of market returns are used. Returns are from TAQ and are calculated using the last midpoint in each interval. SPY ETF returns are used as the market proxy. Price Delays are calculated from five-minute returns in Panel A and 1-minute returns in Panel B. A stock is defined as having a high-HFT participation day when its participation share is in its highest tercile for that stock over the entire sample. Participation share is defined as HFT dollar volume divided by the stock's total dollar volume. Three versions of participation shares are calculated, differing in whether HFT participation is defined as trades where an HFT participates in either side (*HFT_ALL*), the liquidity-demanding side (*HFT_DEMAND*), or the liquidity-supplying side (*HFT_SUPPLY*). Trades where an HFT participates in both sides are used in all three measures. Price Delay differences are calculated separately for each stock, and cross sectional means across all stocks are reported. *T*-statistics test the null that the mean is 0. Adjusted *t*-statistics are corrected for cross-correlation in the residuals. The sample contains 120 stocks.

They show evidence of this occurring very rapidly after index futures price changes, while it would seem that multiple trades must be accumulated before an order flow imbalance becomes meaningful.

7. Conclusion

In this paper, I analyze HFT trading performance, trading costs, and effects on market efficiency using a sample of NASDAQ trades and quotes with HFT participation explicitly identified. HFTs seem to possess intraday market timing ability, and this result is not driven solely by very short-term signals or trading at fleeting prices. The magnitude of their market timing performance suggests that there is economically significant predictability in intraday prices. Trading costs are low in this market, but spreads are wider on trades where HFTs provide liquidity and tighter on trades where HFTs take liquidity. This suggests that HFTs provide liquidity when it is scarce and consume liquidity when it is plentiful. Prices incorporate information from order flow and market-wide returns more efficiently on days when HFT participation is high. This effect is driven by HFT demand-side participation, implying that HFTs improve price efficiency when demanding liquidity.

This new evidence can potentially provide guidance to theoretical researchers seeking to model HFT behavior and market quality impacts. For example, the relatively low spreads earned on their liquidity providing trades, their market timing performance, and the large share of their trades that demand liquidity together suggest that one may not want to model HFTs as uniformly following market-making strategies. The HFT intraday market timing results suggest that models where HFTs solely profit from very short-term activities such as trading at fleetingly available prices may be incomplete.

It is worth reiterating that my data are limited to NASDAQ continuous trading and my focus is on the collective trading and market quality impacts of the sample HFTs aggregated over a variety of market conditions. These issues and other limitations of this study are discussed in more detail above. Conclusions drawn in this setting may not generalize to other environments, and continued study of these issues is clearly warranted. In particular, HFT trading strategies and impacts on market quality in extreme market conditions are important topics for future research.

References

Angel, J., Harris, L., Spatt, C., 2011. Equity trading in the 21st century. Quarterly Journal of Finance 1, 1–53.

Arellano, M., 1987. Computing robust standard errors for within-groups estimators. Oxford Bulletin of Economics and Statistics 49, 431–434.

Berkowitz, S.A., Logue, D.E., Noser, E.A., 1988. The total cost of transactions on the NYSE. Journal of Finance 43, 97–112.

Bessembinder, H., 2003. Trade execution costs and market quality after decimalization. Journal of Financial and Quantitative Analysis 38, 747–778.

Bessembinder, H., Kaufman, H.M., 1997a. A comparison of trade execution costs for NYSE and NASDAQ-listed stocks. Journal of Financial and Quantitative Analysis 32, 287–310.

Bessembinder, H., Kaufman, H.M., 1997b. A cross-exchange comparison of execution costs and information flow for NYSE-listed stocks. Journal of Financial Economics 46, 293–319.

Biais, B., Foucault, T., Moinas, S., 2011. Equilibrium algorithmic trading. Working paper, University of Toulouse.

Brennan, M.J., Jegadeesh, N., Swaminathan, B., 1993. Investment analysis and the adjustment of stock prices to common information. Review of Financial Studies 6, 799–824.

Brogaard, J., 2012. Essays on high frequency trading. Northwestern University dissertation.

Brogaard, J., Hendershott, T., Riordan, R., 2013. High frequency trading and price discovery. Working paper, University of California at Berkeley.

Chaboud, A., Chiquoine, B., Hjalmarsson, E., Vega, C., 2011. Rise of the machines: Algorithmic trading in the foreign exchange market. Working paper, Board of Governors of the Federal Reserve System.

Chordia, T., Roll, R., Subrahmanyam, A., 2005. Evidence on the speed of convergence to market efficiency. Journal of Financial Economics 76, 271–292.

Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and market efficiency. Journal of Financial Economics 87, 249–268.

Chordia, T., Roll, R., Subrahmanyam, A., 2011. Recent trends in trading activity and market quality. Journal of Financial Economics 101, 243–263.

Chung, D., Hrazdil, K., 2010a. Liquidity and market efficiency: a large sample study. Journal of Banking and Finance 34, 2346–2357.
Chung, D., Hrazdil, K., 2010b. Liquidity and market efficiency: analysis of NASDAQ firms. Global Finance Journal 21, 262–274.
Comerton-Forde, C., Hendershott, T., Jones, C.M., Moulton, P.C., Seasholes, M.S., 2010. Time variation in liquidity: the role of market-maker inventories and revenues. Journal of Finance 65, 295–331.
Cvitanic, J., Kirilenko, A., 2010. High frequency traders and asset prices. Working paper, California Institute of Technology and MIT Sloan School.
Fama, E.F., 1970. Efficient capital markets: a review of theory and empirical work. Journal of Finance 25, 383–417.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81, 607–636.
Garman, M.B., 1976. Market microstructure. Journal of Financial Economics 3, 257–275.
Glosten, L., Milgrom, P., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of Financial Economics 14, 71–100.
Gormley, T.A., Matsa, D.A., 2013. Common errors: How to (and not to) control for unobserved heterogeneity. Working paper, University of Pennsylvania and Northwestern University.
Hagströmer, B., Nordén, L., 2013. The diversity of high frequency traders. Journal of Financial Markets, http://dx.doi.org/10.1016/j.finmar.2013.05.009, thiss issue..
Hasbrouck, J., Saar, G., 2009. Technology and liquidity provision: the blurring of traditional definitions. Journal of Financial Markets 12, 143–172.
Hasbrouck, J., Saar, G., 2013. Low-latency trading. Journal of Financial Markets, http://dx.doi.org/10.1016/j.finmar.2013.05.003, this issue..
Hasbrouck, J., Sofianos, G., 1993. The trades of market makers: an empirical analysis of NYSE specialists. Journal of Finance 48, 1565–1593.
Hendershott, T., Menkveld, A.J., 2012. Price pressures. Working paper, University of California at Berkeley and VU University Amsterdam.
Hendershott, T., Riordan, R., Algorithmic trading and the market for liquidity. Journal of Financial and Quantitative Analysis, forthcoming.
Hendershott, T., Jones, C.M., Menkveld, A.J., 2011. Does algorithmic trading improve liquidity? Journal of Finance 66, 1–33.
Ho, T., Stoll, H.R., 1981. Optimal dealer pricing under transactions and return uncertainty. Journal of Financial Economics 9, 47–73.
Hou, K., Moskowitz, T.J., 2005. Market frictions, price delay, and the cross-section of expected returns. Review of Financial Studies 18, 981–1020.
Huang, R.D., Stoll, H.R., 1996. Dealer versus auction markets: a paired comparison of execution costs on NASDAQ and the NYSE. Journal of Financial Economics 41, 313–357.
Jarrow, R.A., Protter, P., 2012. A dysfunctional role of high frequency trading in electronic markets. International Journal of Theoretical and Applied Finance, 15.
Jovanovic, B., Menkveld, A.J., 2012. Middlemen in limit order markets. Working paper, New York University.
Kirilenko, A., Kyle, A., Samadi, M., Tuzun, T., 2011. The flash crash: The impact of high frequency trading on an electronic market. Working paper, CFTC and University of Maryland.
Manaster, S., Mann, S.C., 1996. Life in the pits: competitive market making and inventory control. Review of Financial Studies 9, 953–975.
Mech, T.S., 1993. Portfolio return autocorrelation. Journal of Financial Economics 34, 307–344.
Menkveld, A.J., High frequency trading and the new-market makers. Journal of Financial Markets, http://dx.doi.org/10.1016/j.finmar.2013.06.006, this issue..
Naik, N.Y., Yadav, P.K., 2003. Do dealer firms manage inventory on a stock-by-stock or a portfolio basis? Journal of Financial Economics 69, 325–353.
Panayides, M.A., 2007. Affirmative obligations and market making with inventory. Journal of Financial Economics 86, 513–542.
Stoll, H.R., 2000. Friction. Journal of Finance 55, 1478–1514.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.