BANK OF ENGLAND

Working Paper No. 469
High-frequency trading behaviour and its impact on market quality: evidence from the UK equity market

Evangelos Benos and Satchit Sagade

December 2012

BANK OF ENGLAND

Working Paper No. 469

High-frequency trading behaviour and its impact on market quality: evidence from the UK equity market

Evangelos Benos[1] and Satchit Sagade[2]

Abstract

We analyse the intraday behaviour of high-frequency traders (HFTs) and its impact on aspects of market quality such as liquidity, price discovery and excess volatility. For that, we use a unique transactions data set for four UK stocks, over the period of a randomly selected week. Our data identifies the counterparties to each transaction, enabling us to track the trading behaviour of individual HFTs. We first find that HFTs differ significantly from each other in terms of liquidity provision: while some HFTs mostly consume liquidity (ie trade more 'aggressively') by primarily executing trades via market orders, others mostly supply liquidity (ie trade more 'passively') by primarily executing trades via limit orders. To examine how trading behaviour is related to these patterns of liquidity provision, we split the HFTs in two groups, according to their trade aggressiveness, and examine the behaviour and impact of each group separately. We find that the 'passive' HFTs follow a trading strategy consistent with market making and as such their trades have alternating signs and are independent of recent (ten-second) price changes. By contrast, 'aggressive' HFTs exhibit persistence in the direction of their trades and trade in line with the recent (ten-second) price trend. We then explore the relationship between HFT activity and market quality. We find that both higher price volatility and lower spreads cause HFT activity to increase. We suggest a number of reasons as to why this might be so. Finally, we use a tick time specification to examine the impact of HFT activity on price discovery (ie information-based volatility) and noise (ie excess volatility). We find that while HFTs have a higher information-to-noise contribution ratio than non-HFTs, there are instances where this is accompanied by a large absolute noise contribution.

Key words: High-frequency trading, liquidity, price discovery, volatility.

JEL classification: G10.

(1) Bank of England. Email: evangelos.benos@bankofengland.co.uk.

(2) ICMA Centre, Henley Business School, University of Reading. Email: s.sagade@icmacentre.ac.uk.

The views expressed are those of the authors and do not necessarily reflect those of the Bank of England or the Monetary Policy Committee, Financial Policy Committee or the Financial Services Authority (FSA). We are grateful to Ana Fernandes and Peter Mason of the FSA for being so patient and helpful with our many questions about the SABRE database. We would also like to thank David Barr, David England, Simon Hargreaves, Stefan Hunt, Marius Jurgilas, Tim Rowe, Edwin Schooling-Latter, Anne Wetherilt, Filip Žikeš, an anonymous referee and seminar participants at the Bank of England, the Norges Bank and the UK Financial Services Authority for useful comments and suggestions. This paper was finalised on 23 November 2012.

The Bank of England's working paper series is externally refereed.

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/publications/Pages/workingpapers/default.aspx

Publications Group, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 Fax +44 (0)20 7601 3298 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

Summary

This paper studies the behaviour of high-frequency traders (HFTs) in the UK equity market and analyses its impact on aspects of market quality such as liquidity, price discovery and excess volatility. Although there is no precise definition of an "HFT", the term is commonly used to describe firms that use computers to trade at high speeds and who also tend to end the day flat, ie carry small or no overnight positions.

HFT activity has increased steadily over the recent years in the US, the UK and continental European equity markets and, following a number of market mishaps (which seem to have been triggered by flawed computer trading algorithms), high-frequency trading has also caught the attention of regulators. However, the empirical evidence on the behaviour and impact of HFTs has so far been relatively limited and inconclusive. Thus, the Bank of England has a natural interest in better understanding HFT behaviour and how it might impact the quality of UK equity markets. In particular, a key question is whether and how HFT activity impacts price efficiency and liquidity.

This paper uses a sample from a data set of transaction reports, maintained by the Financial Services Authority, to attempt to give a first answer to these questions. The data identifies the counterparties to each transaction, which enables us to identify HFTs and study their behaviour.

We first find that HFTs exhibit substantial variability in their trading strategies. For instance, while some HFTs trade primarily passively (by posting orders that rest on the order book of the exchange so that others can trade against them), others trade primarily aggressively (by trading against resting orders of passive traders). In other words, some HFTs mostly supply liquidity and others mostly consume it. For this reason and in order to examine how trading behaviour is related to these patterns of liquidity provision, we split the HFTs in two groups, according to their trade aggressiveness, and examine the behaviour and impact of each group separately.

The "passive" HFTs tend to alternate their positions over the short run (ie their buys tend to be followed by sells and their sells by buys) and their positions also tend to be insensitive to recent price changes. Conversely, "aggressive" HFTs do not alternate their positions, and tend to trade in the direction of the recent price trend (ie they buy when the price rises and sell when it drops).

We next examine whether and how price volatility and the prevailing bid-ask spread influence HFT activity. We find that both "passive" and "aggressive" HFTs trade relatively more when prices are more volatile and when the spread is narrow. We suggest a number of reasons why this might be so.

Finally, we examine the impact of HFT activity on volatility. We note that volatility can be either "good" (when price changes reflect the arrival of new information about fundamentals) or "excessive" (when price changes do not reflect any information about fundamentals). In the latter case it is also referred to as "noise". Clearly, markets are more informationally efficient when there is more "good" volatility and less "noise". We therefore examine the contribution of HFTs to both "good" volatility and "noise". For that, we use an econometric framework that

takes into account the exact time sequencing of HFT trades and price changes and, as such, allows us to isolate and estimate the causal effect of HFT activity on price volatility.

Our results show that there are instances where HFTs contribute (in absolute terms) a large amount of both “good” and “excessive” volatility; more so than the average, non-HFT, trader. This is possible if some of their trades carry a large amount of information while other trades are uninformative. We hypothesise that this may be because HFTs aim to end each day with relatively flat positions: if an HFT must, at some point during the day, only trade in order to adjust their inventory, these trades will have no information content and will likely create noise. For the stocks we analyse, HFTs are more informationally efficient than non-HFTs as their relative contribution of “good” to “excessive” volatility is on average 30% higher than that of non-HFTs. Owing however to the small number of stocks in our sample, we cannot confidently generalise these findings in the entire cross-section of stocks.

Given the instances of large contributions of both “good” and “excessive” volatility by HFTs, it is not immediately clear what the welfare implications of HFT activity are. If improvements in price efficiency at some times come part and parcel with additional noise at other times, then whether HFT activity is socially beneficial or not, will ultimately depend on how the marginal utility of information compares with the marginal disutility of noise, ie on how much additional noise we are willing to tolerate at some times for the benefit of more informed trading at other times. It will also depend on the balance between any beneficial impact HFTs may have on markets during “normal” market conditions and the effect of HFT activity under more “stressed” market conditions. Finally, the welfare implications of HFT activity will also depend on the propensity of errors in the operation of their algorithmic trading to cause harmful disturbances of the type experienced in the “Flash Crash” of 6 May 2010. However, these issues are beyond the scope of this paper.

1 Introduction

This paper utilises unique transactions data from the UK equity market to study the intraday behaviour of High-Frequency Traders (HFTs) and understand their impact on market quality. Although not precisely defined, the term "HFT" is generally used to describe any trading firm that makes extensive use of computer technology in its trading process with the aim of executing a large number of transactions within short time intervals. Furthermore, the firm normally ends the day with a relatively flat position.[1] Following the May 6, 2010 "Flash Crash" in the United States and the potential role that HFTs may have played in the crash, the subject has caught the attention of academics and policy makers worldwide.[2] The debate has mainly centred around the impact of HFT on various aspects of market quality such as liquidity, price discovery and excess volatility. To the extent that HFT impacts these aspects of market quality it has direct implications for financial stability. Indeed, regulators on both sides of the Atlantic have implemented or contemplate rules aimed at either directly curbing HFT activity or limiting its potential impact.[3] Our paper contributes to the ongoing debate on the impact of HFT activity on market quality and also sheds some light on the trading behaviour of HFTs.

For our analysis we use transaction-level data which identifies the counterparties to every trade and thus enables us to track individual HFTs. While our data does not include the trades of HFTs who are not regulated in the European Economic Area, it does include the trades of some of the largest HFTs. To classify firms as HFTs, we use various information sources including the companies' websites and media reports.[4] In our sample, "HFTs" participate in about 27% of all trading volume.

We start by analysing the impact of HFT activity on liquidity. Liquidity is an important aspect of market quality and liquidity provision has frequently been cited in arguments made both in favour and against HFTs.[5] We find that HFTs display significant variability in the level of liquidity provision: while some HFTs predominantly supply liquidity, others primarily consume it. For this reason and owing to confidentiality restrictions, we split the sample of HFTs in two groups based on their trade aggressiveness[6] and study the behaviour and trading impact of each group separately. The "passive" group consists of all HFTs with a below-median aggressiveness ratio and the "aggressive" group consists of all HFTs with an above-median aggressiveness ratio.

Having split the set of HFTs in our sample in these two groups, we then attempt to understand the second-by-second behaviour associated with the different patterns of liquidity provision of each group. We find that the "passive" HFTs exhibit a behaviour broadly consistent with that of a market maker: these HFTs alternate between buys and sells and their trading is neutral to recent (i.e. 10-second) price changes. On the contrary, the "aggressive" HFTs trade in a trending fashion (their buys are followed by buys and their sells by sells) and they also trade with the recent price trend (i.e. they buy when prices rise and sell when they drop). On account of this behaviour, and given the need to balance their positions, they display a stronger long-term sensitivity to inventory levels.

We next look for causal links between HFT activity and two measures of market quality, namely short-term price volatility and the bid-ask spread. The fundamental problem that one encounters in such cases is endogeneity. We exploit the autoregressive behaviour of both spread and volatility and use lagged values of these variables as instruments for current values in order to assess their impact on HFT activity. The results show that both groups of HFTs trade proportionally more when volatility is higher and spreads are narrower. This is consistent with a number of explanations: If higher volatility is suggestive of a higher rate of information arrival, HFTs may be trading proportionally more by exploiting their speed advantage.

[1] For a detailed overview of HFT characteristics see AFM(2010).

[2] See CFTC-SEC (2010) for the official report on the "Flash Crash". See Haldane (2011) and Shapiro (2012) for a regulatory perspective of HFT.

[3] For instance, as part of the ongoing revision of the Markets in Financial Instruments Directive in Europe, regulators contemplate a minimum resting period of half a second for all limit orders while the SEC recently approved proposals to tighten the market-wide circuit-breaking thresholds on a pilot basis.

[4] Companies that are classified as HFTs this way include market makers, other proprietary trading companies and hedge funds. Companies that use algorithms exclusively for optimising order execution are not labeled as HFTs in our sample.

[5] See for example IOSCO (2011)

[6] Trade aggressiveness is the ratio of trading volume (or number of trades) executed via market orders or marketable limit orders over the total trading volume (or total number of trades) in which an HFT participated.

Lower spreads on the other hand may be causing both types of HFTs to trade more for different reasons: "Aggressive" HFTs may find it cheaper to trade when the spread is small since they primarily use market orders. "Passive" HFTs may again be exploiting their speed advantage if a small spread is indicative of higher competition for liquidity provision which results in a faster market (in the sense that quotes need to be updated quickly to be competitive).

An alternative solution to the endogeneity problem is to work in tick time.[7] In tick time, one can sequence transactions and quote updates and, using an appropriate econometric framework, it is possible to isolate the causal link between trades and price changes. We use the standard VAR framework of the Hasbrouck (1991a, 1991b) model of trades and price changes and augment it appropriately to test our hypotheses of interest. Using this framework, we first estimate the price impact (via impulse response functions) of the two groups of HFTs and compare it with that of the rest of the traders. We find that the "aggressive" HFTs have a larger price impact than the "passive" ones (1.2 bps versus 0.5 bps). However this larger price impact by the "aggressive" HFTs is proportional to their aggressive trading volume, suggesting that "aggressive" HFTs' trades do not individually have a larger price impact than those of the "passive" ones. The rest of the traders have a higher average price impact (at 2.6 bps), again, due to their higher trading volume. However, on a "per-share-traded" basis, HFTs have a much larger price impact than the other traders: whereas HFTs initiate about 12% of all aggressive volume, they account for about 40% of the total price impact. If one interprets the price impact as information contribution, then these results suggest that while all other traders contribute more to price discovery in absolute terms (on account of the larger trade volumes that they initiate), the HFTs in our sample contribute more to price discovery on a per-share-traded basis.

A challenge in a high-frequency trading environment is making sense of short-term price volatility. In theory, every trade that has a price impact increases short-term volatility. However, given that the concern is not about volatility per se but rather about whether volatility reflects information or not, we do not wish to potentially brand HFTs as "volatility increasing" if they quickly and accurately respond to the arrival of new information about a security and through their trades cause prices to quickly adjust to their new equilibrium values. For this reason, we follow Hasbrouck (1991b, 1993) and impose some additional structure to the VAR framework: we first assume that the observed price consists of an efficient (non-stationary) and a noisy (stationary) price component. Changes in the efficient price component last longer (i.e. 50 ticks) and as such are assumed to reflect information about fundamentals. Thus, the efficient price innovation equals, at each tick, the overall, long-term price impact. On the contrary, any price changes that tend to quickly reverse and disappear are attributed to noise. After decomposing the price innovation at every tick into the efficient price innovation and the residual noise, we examine what fraction of the variance of each component the two groups of HFTs and all other traders are responsible for.

Given that we do this analysis for each stock separately and that we only have a few stocks in our sample, we cannot draw general conclusions about the impact of HFT on the stock cross-section. However, we can demonstrate that there are instances where HFTs collectively contribute around 50% of both the variance of the efficient price innovation and of the noise variance. This suggests that given their trading volume participation of about 27%, HFTs (particularly the "aggressive" ones) can significantly amplify both price discovery and noise. Overall, HFTs have higher ratios of information-to-noise contribution than all other traders, with the difference being statistically significant for some stocks.

The instances where HFTs contribute large amounts of both information and noise to the prices suggest that while some HFT trades have a large information content, others are entirely uninformative. We believe that this is likely a result of the types of strategies that HFTs follow. For example, if informed trading causes an HFT to build up a position during the day, she will have to unload this position by the end of the day. To the extent that there are no profitable ways to do so, the HFT may be forced to make a series of "uninformed" trades to simply bring her position back to her target inventory levels. This could mean that inherent in the typical HFT strategy of carrying no positions overnight is the creation of noise. Alternatively, HFTs (especially the "aggressive" ones) may be employing trading strategies that rapidly react to news arriving at the market. This reaction however may either be disproportionate or the HFTs may also react to signals that are in reality uninformative. In those cases HFTs will end up creating excess

[7] In the context of an empirical model this means that time is not updated in fixed intervals but, rather, whenever there is a transaction or a quote update.

volatility.

Our paper contributes to a small but fast growing literature on HFT and Algorithmic Trading (AT). While some papers focus on the actual behaviour of HFTs[8] most studies have attempted to assess the impact of HFT on liquidity, price discovery and volatility.

Regarding the impact of HFT activity on liquidity, the evidence so far is mixed. For instance, while Hasbrouck and Saar (2011) and Hendershott et al (2011) find that HFT/AT activity reduces quoted and effective spreads, Hendershott and Moulton (2011) show that following the introduction of NYSE-hybrid in 2006, spreads went up as liquidity providers sought compensation against the risk of being adversely selected by informed traders. Also, Brogaard (2011) documents that HFTs provide less order book depth than non-HFTs.

There seems to be more of a consensus in the literature on the effect of HFT on price discovery. For instance, Hendershott and Moulton (2011) also find that, after the introduction of NYSE-hybrid, order execution times fell thus causing prices to adjust faster to new information. Hendershott et al (2011) find that algorithmic trading enhances the informativeness of quotes on NYSE, while Hendershott and Riordan (2011) find that for DAX-traded equities, algorithmic trades contribute more to price discovery than human trades. Finally, using Nasdaq data, Brogaard (2011) also finds that HFTs contribute to the price discovery process.

Some papers have also examined the impact of HFT on overall volatility. However, most of these papers make no distinction between the informative and uninformative components of volatility. It is then perhaps for this reason that the results of these papers are less conclusive. For instance, Hasbrouck and Saar (2011) find that HFT decreases short-term volatility while Zhang (2010) argues that HFTs increase volatility. A paper that does distinguish between information-based and excess volatility is that of Boehmer et al (2012). Using data from a cross-section of 39 exchanges, over a nine-year period, these authors find that while algorithmic traders improve informational efficiency, they also contribute to excess volatility. However, the authors do not directly observe HFT activity but instead infer it from the ratio of trading volume to message traffic. The fact that most studies suffer from the same problem, may be another reason for the discrepancies in the literature. For instance, Zhang (2010) indirectly estimates HFT activity from institutional investor holdings over quarterly intervals while Hasbrouck and Saar (2011) estimate low-latency activity by the number of sequences of linked messages ("strategic runs") over 10-minute intervals. By using a unique data set where the identities of the counterparties are known, our paper overcomes this difficulty.

2 Data: Properties and Filters

We use transactions data, time stamped to the second, for four FTSE 100 stocks (labeled as W, X, W and Z) over a randomly selected one-week period, within the past two years.[9] The transactions data is obtained from the SABRE and Bloomberg databases. The SABRE database is maintained by the UK Financial Services Authority (FSA) and contains information on transaction prices, sizes, time, location and importantly, counterparty identity. It also identifies the buyer and the seller in each transaction as well as whether the counterparties execute a given transaction in a principal or agent capacity. The SABRE database captures the trading activity of all firms directly regulated by the FSA as well as that of firms that trade through a broker (brokers are regulated and as such must report their clients' transactions). This implies that firms who are not subject to FSA regulation and who do not trade through a broker are not subject to reporting requirements and their trades are not captured in SABRE. However, SABRE does include the trades of some of the largest HFTs.[10] To analyse the behaviour of HFTs in the universe of all trades we complement SABRE with Bloomberg, which contains all transactions executed on the various exchanges as well as all quote updates at the top of the order books.

For our analysis, we apply a number of filters on these two data sets. Since we are interested in the effects of HFT on aspects of market quality, and given that HFT activity is concentrated on venues with

[8] e.g. Kirilenko et al (2011) examine the behaviour of HFTs before and during the May 6, 2010 "Flash Crash" and Menkveld (2012) studies the behaviour of a large market-making HFT on Chi-X and Euronext.

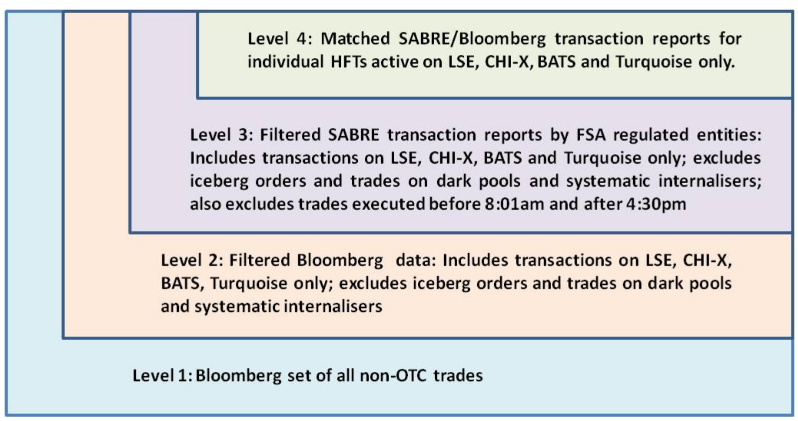
[9] The names of the stocks as well as the exact dates are not revealed for confidentiality reasons.

[10] The SABRE database has only been used before in the context of HFT in Brogaard et al (2012). These authors analyse the effect of HFT on the execution costs of institutional investors.

electronic, public limit order books, we only retain trades that have exclusively been executed on the order book of one of the four largest (in terms of trading volume) exchanges or Multilateral Trading Facilities (MTFs) in the UK: The London Stock Exchange (LSE), Chi-X, BATS and Turquoise. We thus exclude any Over-The-Counter (OTC) transactions reported on these exchanges as well as any transactions executed on Systematic Internalisers (SIs) and Dark Pools (DPs). Dark trades are executed at the midpoint of the prevailing best bid and offer and therefore have no market impact. We also drop transactions that result from fully or partially hidden limit orders (also known as "iceberg orders"). These orders do not affect the prevailing quote because they are invisible and thus do not allow market participants to condition their behaviour on them. Finally, we drop any transactions that are time stamped before 8:01am or after 4:30pm so as to exclude trades that are associated with the opening and closing auctions at any of the four exchanges/MTFs. In all cases, we make sure that the SABRE and Bloomberg samples are always comparable by applying the same filters.[11]

To analyse HFT behaviour, we isolate the identities and trades of individual HFTs in the SABRE database and match their SABRE reports with their Bloomberg trade reports. The matched data set enables us to track individual HFT trades in the Bloomberg universe of all trades. The matching of these transaction reports is done by nearest time stamp, trade size and trade price. Figure 1 illustrates the relationship between the various trade data sets we work with. The "Level 3" filtered SABRE data set contains the transaction reports of all FSA-regulated entities while the "Level 4" matched SABRE/Bloomberg data set is used to analyse individual HFT behaviour. After applying all the data filters, the transactions we retain represent, for each stock and each of the four exchanges/MTFs, more than 70% of their total non-OTC trading volume.[12] These same filters reduce the number of unique SABRE transaction reports from 849,626 to 716,746.

Figure 1: **Illustration of the structure of our data. The data described in the smaller rectangles is a subset of that described in the larger ones.**



The stocks we use in our analysis are all from different industries and exhibit substantial variability in trading volumes and market capitalisations. The first two columns of Table 1 give ranges in which the average daily trading volumes for each of the stocks lie, during our sample time, as well as the quartile rank that these average volumes attain when compared with the weekly trading volumes of the same stocks over the rest of the year. It is evident that not only are the trading volumes cross-sectionally diverse but also that the week we have randomly selected is a high-volume week for two of the stocks and a low/average-volume week for the other two. Similarly, the last two columns of Table 1 give ranges for the market capitalisations of the stocks, along with their quartile size rank in the FTSE 100 index. These numbers show that, although all of our stocks are among the 100 largest, they do exhibit substantial size variability both in absolute terms and relative to the FTSE 100. The smallest stocks have a market capitalisation of less than £20bn while the

[11] See Appendix 1 for a detailed description of how we filter the raw SABRE data.

[12] In terms of the illustration in Figure 1, these are the trading volumes captured in the "Level 2" data set as a fraction of the trading volumes in the "Level 1" data set.

largest stock has a size between £80bn and £160bn. Overall, the stocks in our sample exhibit a reasonable degree of cross-sectional variability in volume and size in both absolute and relative terms.

Table 1: **Stock characteristics**

Stock	*Daily Trading Volume (millions of shares)*	*Quartile Rank (in the year)*	*Market Capitalisation (£bn)*	*Quartile Rank (in the FTSE 100)*
W	0-50	1st	0-20	3rd
X	250-300	3rd	20-80	4th
Y	0-50	4th	0-20	2nd
Z	100-150	2nd	80-160	4th

Notes: This Table shows ranges of daily trading volume (in millions of shares) and market capitalization (in £ billions) to which the stocks in our sample belong. We also report the rankings of the volume and market capitalisation figures as compared respectively with the rest of the trading days in the same year and the rest of the FTSE 100 stocks in the same week. Source: Bloomberg.

3 HFT classification and Summary Statistics

In this section we use the filtered SABRE data set (i.e. the "Level 3" and "Level 4" data sets in Figure 1) to describe some general characteristics of HFTs such as their volume and trade participation, the extent to which they trade aggressively via market orders or passively via limit orders, as well as their trading speeds. Knowledge of the identities of the counterparties in our sample allows us to precisely identify HFTs. To do that, we use press reports and the companies' websites to see which of them are described as HFTs or define themselves as such. For instance, if a company defines itself as a "High-Frequency Trader (HFTs)" (or alternatively a "low-latency trader" or an "electronic market maker") then we classify it as an HFT.

We start by measuring the volume and trade participation of the HFTs and also the extent to which they consume or supply liquidity. A trader consumes liquidity by posting market orders (or marketable limit orders), which get executed against standing limit orders and supplies liquidity by submitting limit orders which rest on the order book.

To do that, we first classify each transaction as buyer- or seller-initiated using the Lee-Ready (1991) algorithm.[13] We then compare this classification with the buy-sell indicator of the SABRE trade report and classify a trade as aggressive or passive as follows:

1 If the Lee-Ready classification matches the buy-sell indicator in SABRE as reported by trader i, then this is an aggressive trade and trader i is consuming liquidity.

2 If the Lee-Ready classification does not match the buy-sell indicator in SABRE as reported by trader i, then this is a passive trade and trader i is supplying liquidity.

The point is that whoever initiates the trade is the party consuming liquidity. So, if the Lee-Ready algorithm classifies a given trade as buyer-initiated and the transaction is reported by trader i who identifies herself as the buyer, then trader i placed the aggressive order. If trader i identifies herself as a seller, then she is not the party who initiated the trade and therefore she must have placed the passive order.

After classifying each transaction j of HFT i as active or passive, we calculate her volume- and trade-weighted aggressiveness ratios, i.e. the ratio of volume (number of trades) traded aggressively, over her total trading volume (total trades). These ratios show whether, on average, an HFT mostly supplies or consumes

[13] See Appendix 2 for a detailed description of the algorithm and how we implement it.

liquidity. The volume and trade-weighted aggressiveness ratios are thus defined as:

$$VWAR_i = \frac{\sum_{j \in AT_i} VOL_{ij}}{\sum_j VOL_{ij}} \tag{1}$$

$$TWAR_i = \frac{\sum_j I_{[j \in AT_i]}}{J_i} \tag{2}$$

where VOL_{ij} is the volume of trade j executed by HFT i, AT_i is the set of all aggressive trades by HFT i and J_i is total number of transactions in which HFT i participates.

Table 2 shows the HFT participation and aggressiveness statistics. Panel A looks at trading volumes (measured in millions of shares traded) and Panel B looks at numbers of trades. HFTs participate in about 27% of the total traded volume and 26% of all executed transactions. This means that the HFTs in our sample are a counterparty to roughly one out of every four shares traded. The third column of the Table shows the HFTs' aggressiveness ratio in terms of volume and number of trades. The aggressiveness ratio of HFTs is about 47% in volume terms. This suggests that on average, HFTs supply slightly more liquidity than they consume. However, this number masks the fact that individual HFT behaviour can vary substantially. To highlight this point, we split our HFT sample into a "passive" and an "aggressive" group with the "passive" group including all HFTs with a below-median aggressiveness ratio and the "aggressive" group including all HFTs with an above-median aggressiveness ratio. The two groups have distinctly different aggressiveness ratios which likely reflects the variety of trading strategies that are employed by HFTs. The HFTs in the "passive" group trade two thirds of their shares (and also execute two thirds of their trades) by posting limit orders and trade the remaining third of shares by market orders, thus supplying a lot more liquidity than they consume. The "aggressive" group on the other hand trades slightly more than half of its shares (and executes about 60% of its trades) via market orders, thus consuming more liquidity than it supplies. The grouping we use (for confidentiality reasons) still masks some of the variation in aggressiveness ratios. For instance, the maximum aggressiveness ratio in our sample is above 70% in volume terms.

Table 2: **HFT participation and aggressiveness**

	Panel A: Volume			
HFT Type	*Volume Participation*	*% of Volume Participation*	*Aggr. Ratio (Vol-weighted)*	*% of Aggr. Volume Participation*
All HFTs	539.1	27.1%	46.9%	12.71%
"Passive" HFTs	214.4	10.8%	32.7%	3.53%
"Aggressive" HFTs	324.6	16.3%	56.3%	9.18%
	Panel B: Trades			
HFT Type	*Trade Participation*	*% of Trade Participation*	*Aggr. Ratio (Trade-weighted)*	*% of Aggr. Trade Participation*
All HFTs	109,390	26.0%	45.2%	11.8%
"Passive" HFTs	58,681	14.0%	31.2%	4.4%
"Aggressive" HFTs	50,709	12.0%	61.3%	7.4%

Notes: Panel A shows absolute volume participation (in millions of shares), relative volume participation as a fraction of total trading volume, volume-weighted aggressiveness ratios as defined in equation (1) and the fraction of aggressive volume participation which is the product of relative volume participation with the volume-weighted aggressiveness ratio. Panel B shows absolute and relative participation by number of trades, trade-weighted aggressiveness ratios as defined in equation (2) and the fraction of aggressive trade participation which is the product of relative trade participation with the trade-weighted aggressiveness ratio. All numbers are aggregated over stocks and HFTs. We classify as HFTs all firms in the SABRE data set for which press reports and their own websites point to such activity. The "passive" group consists of all HFTs with a below-median aggressiveness ratio and the "aggressive" group of all HFTs with an above-median ratio.

These numbers suggest that HFT behaviour is very diverse: for example, the low aggressiveness ratios of the passive group are consistent with high-frequency market-making whereas the aggressive group of HFTs could be employing some other high-frequency strategy such as statistical arbitrage or event trading. This means that in order to examine the effect of HFT activity on market quality, one needs to do that separately for the different types of HFT strategies. Thus, later in the paper we examine separately the impact of the two groups of HFTs on various aspects of market quality.

Table 3 shows statistics on trade sizes, end-of-day of positions (as a fraction of trading volumes) and trading speeds. The average trade size for the HFTs in our sample is 4,731 shares with "aggressive" HFTs' trades being twice as large as "passive" HFTs' ones. To see if the HFTs that we identify conform with the widespread view that they carry small overnight positions, we also calculate for each HFT, the volume-weighted ratio of their absolute end-of-day positions over their total trading volume. Thus, the End-of-Day Position over Volume (EDPV) ratio for HFT j is:

$$EDPV_j = \frac{\sum_{i=1}^{4}\Big[\big(\sum_{t=1}^{5}|POS_{jit}|\big)\big(\sum_{t=1}^{5}VOL_{jit}\big)\Big]}{\sum_{i=1}^{4}\sum_{t=1}^{5}VOL_{jit}} \tag{3}$$

where POS_{jit} is the end-of-day t position of HFT j on stock i and VOL_{jit} is her daily trading volume on the same stock. For this we assume that traders start each day with a zero position. The numbers in Table 3 suggest that indeed, HFTs on average end the day flat as the average ratio, across all HFTs, is 5.6%. "Aggressive" HFTs appear to carry smaller overnight positions than "passive" ones; this could be because "passive" HFTs engage in market-making strategies which require a stock inventory. Of the two groups of HFTs that we analyse, the "passive" one is faster than the "aggressive" group (2.88 vs. 2.49 trades per minute) and also has higher maximum speeds (7.95 vs. 6.70 trades per minute). The average trading speed of the "passive" group is comparable to the 2.74 trades per minute that Menkveld (2012) reports for the large market-making HFT that he studies.[14]

Table 3: **High-Frequency Trader summary statistics**

HFT Type	*Average Trade Size*	*Avg. End-of-day Position over Volume*	*Average Trade Speed (Trades/Min)*	*Maximum Trade Speed (Trades/Min)*
All HFTs	4,731	5.6%	2.69	7.95
"Aggressive" HFTs	6,362	1.7%	2.49	6.70
"Passive" HFTs	3,100	9.5%	2.88	7.95

Notes: The first column shows the average trade size of of the firms in our sample that describe themselves as "high-frequency traders" as well as of the two HFT groups ("passive" and "aggressive") that we analyse. The second column shows their simple average, across HFTs, of the end-of-day position as a fraction of trading volume, defined in equation (3). The last two columns show the average and maximum trading speeds (in trades per minute) across stocks and for all venues. The HFT group includes all firms in our sample for which press reports and their own websites point to such activity. Source: SABRE.

4 HFT positions and their determinants

We next analyse the short-term trading behaviour of the "passive" and "aggressive" HFTs.[15] We do this for two reasons: first, in order to identify any trading patterns that might cause one group of HFTs

[14] Menkveld (2012) reports that the HFT trades 1,397 times per stock per day. Assuming an 8.5 hour trading day, this translates to 2.74 trades per stock, per minute.

[15] The analysis in this section is based on the matched SABRE/Bloomberg data set (the "Level 4" data set in Figure 1).

to supply liquidity and the other to consume it. Second, in order to associate HFTs' short-term behaviour with their impact on short-term volatility and price discovery.

Thus, we attempt to identify the factors that influence the second-by-second changes in HFTs' positions. For each individual HFT in our sample, we estimate an empirical model of change in holdings similar to the one in Kirilenko et al (2011):

$$\Delta y_{ijt} = \alpha + \phi \Delta y_{ijt-1} + \delta y_{ijt-1} + \sum_{k=1}^{10} \beta_k R_{it-k} + \sum_{m=1}^{4} d_m I_{[m=i]} + u_{ijt} \tag{4}$$

where the index i denotes stocks, j denotes HFTs and t denotes a second-long time interval. y_{ijt} is the level of HFT j holdings, in number of shares, in stock i at time t, R_{it-k} is the return of the same stock over the time interval $t-k$ and $I_{[m=i]}$ is the stock i dummy variable. We calculate returns using the quote midpoint at the beginning and end of each second, so as to avoid the bid-ask bounce bias. The coefficients ϕ and δ can be interpreted as the strength of short- and long-term mean reversion respectively. We estimate this model separately for each group of HFTs by a pooled regression.

Table 4 shows the results of this estimation.[16] A striking difference between the two groups of HFTs is their opposite sensitivity to recent position changes (the ϕ parameter). Whereas the "passive" group reverses its position in the following second, the "aggressive" trades in the same direction. Both groups target an inventory level over a longer time period as the negative coefficients on their lagged inventory levels (the δ parameter) indicate. This suggests, that over a longer period, the positions of the HFTs, in both groups, are mean-reverting with an average inventory half-life of about 2.7 and 1.6 hours respectively.[17] The "passive" group has a longer inventory half-life (the δ coefficient is smaller in absolute terms), possibly because the second-by-second reversal of positions allows the HFTs in this group to maintain a relatively constant inventory in the short term so that smaller adjustments are necessary over the longer term.

Finally, the coefficients on past returns are also different across the two groups of HFTs. Following a price change, the "passive" group trades, in the next 10 seconds, in a way that is neutral to the price trend as the small and statistically insignificant sum of the lagged return coefficients (which equals -1.97) suggests. On the contrary, over the same period, the "aggressive" group seems to be trading strongly with the price trend: a 0.01% increase in the stock price in each of the previous 10 seconds (leading to a cumulative increase of 0.1%), causes the HFTs in that group to purchase, on average, an additional 138 shares.[18]

Overall, the "passive" HFTs display characteristics consistent with market-making activity: apart from supplying more liquidity than they consume, they typically reverse their positions from one second to the next and their trading is price-neutral. On the other hand, the "aggressive" HFTs appear to follow a different trading strategy: Although they, too, target an inventory level (albeit over a longer time period), their trading is trending with respect to both position and price changes. A price change over the previous 10 seconds leads to a same-direction trade which in turn leads to an additional same-direction trade in the next second.

5 Analysis of HFT activity in clock time

We next explore the relationship between HFT activity and two dimensions of market quality, namely spread and volatility. Volatility is defined here in the usual sense and we do not attempt to break it down into information-related volatility and noise. Apart from examining if, and to what extent, HFT activity

[16] We estimate the model coefficients using fixed effects. There are however two potential problems with the above specification that need to be addressed. The first has to do with the dynamic bias created by the lagged dependent variable in a panel structure and the second with the potential bias created by the presence of a lagged dependent variable along with autocorrelation in the error terms. The first is less of a concern because the bias tends to zero as the time dimension in a panel increases (see Nickel (1981)), which is the case in our model. To address the second, we estimated the model both without the lagged dependent variable and with additional lags of the dependent variable, thus removing low order serial correlation. The point estimates of both the lagged dependent variable and of the other regressors appeared to be robust across specifications.

[17] The half-life of an HFT's inventory is the time it takes for a deviation from a target level to be reduced by half. It is calculated as $\log(0.5)/\hat{\delta}$.

[18] These results are robust to randomly dropping particular stocks or days from the data.

Table 4: **Position regressions**

Dependent Variable: Δy_{ijt}	**"Passive" HFTs**	**"Aggressive" HFTs**
Independent Variables:		
Δy_{ijt-1}	**-0.0185**	**0.0267**
y_{ijt-1}	$\mathbf{-7.1 \times 10^{-5}}$	$\mathbf{-12.5 \times 10^{-5}}$
R_{it-1}	**-36.349**	**55.723**
R_{it-2}	**7.104**	**29.198**
R_{it-3}	9.807	**21.900**
R_{it-4}	**9.236**	**18.884**
R_{it-5}	9.087	**19.066**
R_{it-6}	-4.801	**8.150**
R_{it-7}	0.014	-0.266
R_{it-8}	9.043	3.498
R_{it-9}	-1.535	-4.397
R_{it-10}	**-3.577**	**-13.302**
Sum ($\sum_{k=1}^{10} \beta_k$)	-1.970	**138.453**
Obs.	610,600	610,600
Inventory Half-life (hours)	2.7	1.6

Notes: Estimation results of the position regression (4) for the "passive" and "aggressive" groups of HFTs that we analyse, over all four stocks and for the entire week in our sample. The "passive" HFTs are those with a below-median aggressiveness ratio and the "aggressive" ones are those with an above-median aggressiveness ratio. y_{ijt} is the stock i, time t level of holdings (in 10,000s of shares) by HFT j and R_{it} is the stock i, time t return. The dependent variable is the contemporaneous change in holdings, Δy_{ijt}. Coefficients that are significant at 5% are in bold. Half-life is calculated as $\log(\mathbf{0.5})/\hat{\delta}$ where $\hat{\delta}$ is the coefficient on the lagged inventory level $\mathbf{y_{it-1}}$.

is correlated with spread and volatility, the main goal is to uncover any causal links between HFT activity and these two parameters of market quality. This is no easy feat when working in clock time because in any single time interval there are multiple transactions and quote updates so that it is not possible to sequence the events and distinguish the cause from the effect. To overcome this problem, we use, in our multivariate analysis, instrumental variables. Unfortunately, the nature of the instruments is such that we can only make causal inferences about the effect of market quality on HFT activity and not the other way around. The impact of HFT activity on market quality is therefore something that we address in the next section where we use a tick-time specification.

5.1 Bivariate Analysis

We start by looking at the unconditional relationship between HFT activity and stock price bid-ask spread and volatility. In Figure 2 we plot the excess volume participation of each HFT group, for every 5-minute interval, against all excess spread and volatility percentiles. Excess HFT volume participation is defined as the difference between the actual, 5-minute, HFT volume participation and the HFTs', same-stock, weekly average participation. Similarly, excess spread/volatility is the difference between the 5-minute average spread/volatility of a stock and its weekly average. The participation ratios that we plot are volume-weighted averages across stocks and HFTs. The plots show unconditional relations but HFT volume participation is likely driven by a multitude of factors. Nevertheless, we state the main hypotheses about the possible causal links between HFT activity and spread and volatility and then more formally test some of these hypotheses.

The spread plot in Figure 2a shows a negative relationship between HFT activity and spread size for both groups of HFTs. This is more pronounced at the highest spread percentiles where higher spreads are associated with 3% lower levels of HFT activity relative to the weekly average. Regarding the "passive" HFTs, there are at least two potential explanations why we might observe this pattern. The first is that the spread increases because the "passive" HFTs decrease their trading activity. Since "passive" HFTs predominantly trade by posting limit orders, their quotes can only narrow the spread (or leave it unchanged). Thus, if "passive" HFTs trade less this likely means that they also quote prices less frequently (or further away from the top of the order book) which would cause spreads to increase. A second explanation is that causality runs the other way around and a smaller spread causes the "passive" HFTs to trade proportionally more. Since a tight spread indicates that there is more competition among liquidity providers, it could be that a high-frequency "passive" trader has an advantage in this competitive environment, because she is able to update faster her quotes, and ends up trading proportionally more.[19] To summarise, the two non-exclusive hypotheses are:

- H_{A1}: The "passive" HFTs' trading causes the spread to narrow.
- H_{A2}: A narrower spread causes the "passive" HFTs' limit orders to be executed relatively more frequently.

The "aggressive" HFTs also reduce their trading activity when the spread is wider. Since aggressive trade execution can only widen the spread or keep it unchanged, we interpret this as evidence that "aggressive" HFTs avoid trading when it is expensive to do so, i.e. when the spread is wider. Thus, the hypothesis here is that:

- H_{B1}: A wider spread causes "aggressive" HFTs to trade proportionally less, presumably because trading is expensive.

Overall, the spread plots are consistent with evidence in the empirical literature that HFT either contributes to the narrowing of the spread (as in Hasbrouck and Saar (2011) and Hendershott et al (2011)) or is driven by it (as in Hendershott and Riordan (2011)).

Figure 2b shows the unconditional relationship between HFT activity and volatility. Volatility is measured as the sum of the quote-to-quote squared returns of the quote midpoint. We use the quote midpoint to calculate volatility so as to avoid the bid-ask bounce bias. Both the "passive" and "aggressive" HFTs' activity is increasing with volatility up to a certain point and then, for the highest volatility percentiles, it is decreasing. However, "passive" HFTs appear to be more sensitive to the level of volatility than "aggressive" HFTs.

There are several alternative explanations consistent with this inverted U-shaped relationship between HFT activity and volatility: One is that "passive" HFT trading activity somehow causes volatility to increase.[20] Alternatively, the increase in the "passive" HFTs' activity with volatility could be because volatility reflects a more erratic and/or concentrated arrival of new information. This may in turn cause market-making HFTs to post relatively more limit orders in order to accommodate the increased order flow on behalf of investors who wish to trade on the arriving information. This explanation is consistent with the well-known positive relationship between volatility and trading volume.[21] The decrease in the "passive" HFTs' relative activity for the highest volatility percentiles could be because the "passive" HFTs deem the market too risky at that stage and decide to scale down their activity, at least in relative terms. The reduced activity by the "aggressive" HFTs at the highest volatility percentiles could be for the same reason. To summarise, the alternative hypotheses for both groups of HFTs are:

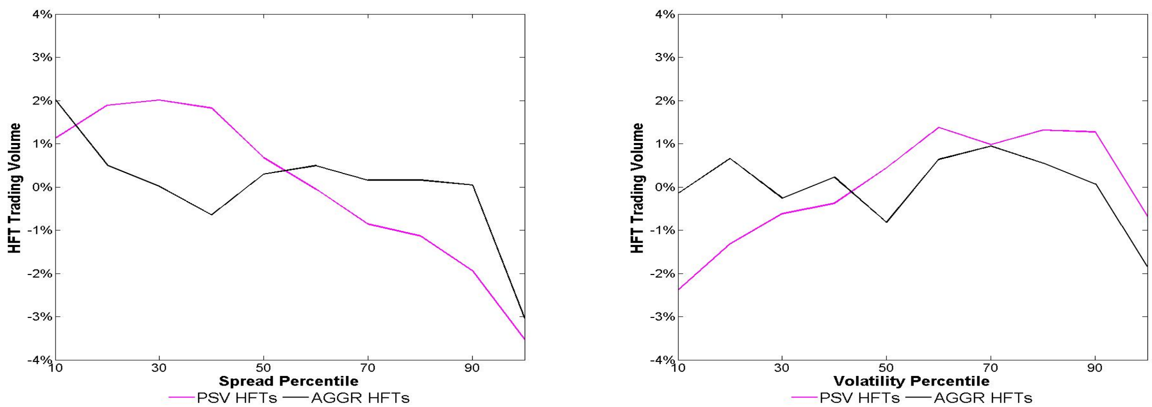
- H_{AB3}: HFT trading affects volatility.
- H_{AB4}: Volatility is influencing the trading activity of the "passive" HFTs (and less so of the "aggressive" HFTs).

[19] Unfortunately we do not have the quote data needed to verify these hypotheses.

[20] It would be difficult however to explain the inverted U-shaped pattern of the "passive" HFTs' trading activity. If "passive" HFTs influence volatility, the inverted U pattern implies that the same level of "passive" HFT activity can influence volatility in different ways.

[21] See for example Schwert (1989).

Figure 2: **Average excess HFT volume participation for each spread (Figure 2a) and volatility (Figure 2b) percentile, over 5-minute intervals. The purple lines show the excess volume participation of the "passive" HFT group and the black lines show the excess volume participation of the "aggressive" HFT group. For each 5-minute interval and each stock we calculate the difference between the average 5-minute spread/volatility of that stock and its weekly average. We then place these differences in a decile bucket. Over the same 5-minute interval we also record the excess volume participation of each HFT in that stock. Excess volume participation is the difference between the actual, HFT volume participation over the 5-minute interval and the HFTs' same-stock weekly average volume participation. We average excess volume participation across stocks and HFTs using volume weights. Source: SABRE and Bloomberg**



(a) *Spread Q Plot* (b) *Volatility Q Plot*

In the multivariate analysis that follows we directly test some of the above hypotheses on the relationship between HFT activity, volatility and spreads.

5.2 Multivariate Analysis

The multivariate analysis tests for contemporaneous causal effects of spread and volatility on HFT activity. The fundamental problem that one encounters in such tests is endogeneity. When working in clock time, even a short time interval will normally contain multiple trades and quote updates and for this reason it is difficult to establish the direction of causality between HFT activity and measures of market quality. One potential solution is to use appropriate instrumental variables, assuming these are available. Given the autoregressive behaviour of both spread and volatility, we use lagged values of these variables as instruments for current values. Unfortunately, we cannot instrument HFT activity in a similar fashion as it exhibits almost zero serial correlation. This implies that the clock-time analysis only allows us to assess the impact of spread and volatility on HFT behaviour.

Therefore, in terms of the notation previously introduced, we are testing hypotheses H_{A2}, H_{B1} and H_{AB4}. Our (simple) empirical specification is:

$$HFT_{it} = bSPREAD_{it} + cVOL_{it} + u_{it}, \quad u_{it} \sim IID(0, \sigma^2) \tag{5}$$

where i indexes a stock and t indexes 15-second intervals. All variables are standardised to have a zero daily mean and unit standard deviation which means there is no need for an intercept in our specification. HFT_{it} is the excess HFT trading volume (in %) on stock i at time interval t i.e., it is the time t percentage of HFT volume participation in excess of the daily average. $SPREAD_{it}$ is the excess prevailing spread and VOL_{it} is the excess volatility calculated as the sum of squared returns over time interval t. As before, these returns are calculated using the prevailing quote mid-point so as to avoid capturing any jumps in prices caused by the bid-ask bounce. By standardising the variables around their daily means we are also effectively controlling for stock-specific fixed effects.

An obvious concern is that the $SPREAD$ and VOL variables may be endogenous: it is easy to imagine HFT activity affecting both contemporaneous spreads and price volatility. For this reason, we instrument these variables using lagged spread and volatility. Both of these variables are strongly autoregressive rendering their lagged values good predictors of their current values. Furthermore, there is no conceivable way in which current HFT activity could influence lagged spreads and volatility. We therefore estimate model (5) via two-stage least squares (2SLS).

Table 5 shows the results of this estimation. Although spread is statistically significant (in the same direction) for both groups of HFTs, it is not economically significant: a one standard deviation increase in the spread relative to its daily mean causes HFT activity to drop by only about 0.05 to 0.06 standard deviations. Nevertheless, for the "passive" HFTs hypothesis H_{A2} is confirmed in the data: a narrower spread results in proportionally more trades for HFTs. This could be because the higher speed with which the "passive" HFTs in our sample post quotes and trade gives them a comparative advantage when the provision of liquidity is competitive. Alternatively, it could be because "passive" HFTs do more inventory rebalancing when spreads are tight. The "aggressive" HFTs' negative sensitivity to spread is likely because it is much cheaper to trade via market orders when the spread is narrow (which supports hypothesis H_{B1}).

The effect of volatility on HFT activity is larger: A one standard deviation increase in volatility causes "passive" HFT activity to increase by 0.2 standard deviations and "aggressive" HFT activity to increase by 0.13 standard deviations. The positive reaction of "passive" HFTs to volatility is likely due to mechanical reasons: controlling for the spread, a standing limit order is more likely to be executed when the fundamental price of the stock (as captured by the quote midpoint) is more volatile. The "aggressive" group's sensitivity to volatility may be because they engage in trading strategies where the profit opportunities are greatest when volatility is high. Overall, these results demonstrate that changes in short-term (i.e. 15-second) volatility and spread have an impact on contemporaneous HFT behaviour.

Table 5: **Multivariate analysis of HFT activity, spread and volatility**

Dependent Variable: HFT_{it}	**"Passive" (HFTs)**	**"Aggressive" (HFTs)**
Independent Variables:		
$SPREAD_{it}$	-0.046	-0.064
	(0.01)	(0.04)
VOL_{it}	0.206	0.126
	(0.02)	(0.08)

Notes: Estimation results of the HFT activity model (5) for the "passive" and "aggressive" groups of HFTs that we analyse, over all four stocks and for the entire week in our sample. The "passive" HFTs are those with a below-median aggressiveness ratio and the "aggressive" ones are those with an above-median aggressiveness ratio. HFT_{it} is the aggregate activity of each group of HFTs on stock i at time interval t. $SPREAD_{it}$ and VOL_{it} are the spread and volatility respectively of the same stock and at the same time interval. All variables are standardised. Significance is established using the Newey-West (1987) standard errors. p-values are in parentheses.

6 HFT activity and market quality in tick time

In the last section of the paper we use the Hasbrouck (1991a, 1991b and 1993) VAR framework to assess, in tick time, the impact of the two groups of HFTs on two dimensions of market quality: price discovery and volatility. These two concepts are related to each other as price discovery is essentially price volatility that occurs in response to information about fundamentals. However, trading can also lead to excess volatility (or "noise") that is unrelated to information about fundamentals. An informationally efficient market is characterised by high levels of volatility associated with price discovery and low levels of excess volatility. We are therefore interested in measuring the impact of HFTs on both price discovery and excess volatility.

We carry out our analysis using transaction reports exclusively from LSE. We do this because we want to capture the tick-by-tick relationship between HFT activity and prices. Since UK exchanges and MTFs are not directly linked to one another, prices and spreads are not simultaneously synchronised across venues. This means that unless one concentrates on a single order book, any estimates of the relationship between HFT activity and market conditions, in tick time, will be noisy. We therefore do our analysis using LSE data since this is the largest venue by trading volume.

We start by estimating the "long-term" price impact of the "aggressive" (i.e. liquidity consuming) trades of the two groups of HFTs and of the rest of the traders. The price impact is a measure of the information content of the trades of each group and, as such, will give us an idea of the relative contribution of each group to price discovery. However, it does not allow us to assess the overall, absolute contribution of each trader category to price discovery and noise. For this reason, we use the Hasbrouck (1991b and 1993) framework and estimate the contributions of the two groups of HFTs and of the rest of the traders to the efficient and noisy components of stock price innovations. This allows us to assess how much each trader category contributes to overall price discovery and noise.

Potential biases due to the partial SABRE coverage

As we mentioned in Section 2, although SABRE includes several large HFTs, it does not include all HFTs active in the UK equity market. This means that when drawing comparisons between the two groups of HFTs and all other traders (labeled "Others"), our results are likely to be biased because the "Others" category will include any HFTs not covered by SABRE. To the extent however that these HFTs have a similar impact to the ones we analyse, this bias will understate the differences between HFTs and non-HFTs. In other words, in the absence of the bias, any differences between the HFTs and the rest of non-HFT traders, would have been even more pronounced.

6.1 The VAR model and HFT price impact

We measure the information content of a trade by calculating its permanent price impact via an impulse response function. For that, we estimate, in tick time, a Hasbrouck (1991a)-inspired VAR model of transactions and returns. We estimate the model separately for each stock and for the entire time span of our sample. Using 10 lags, the VAR specification is:

$$
\begin{aligned}
r_t &= \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=0}^{10} \beta_i q_{t-i}^{\text{“Pass.” HFTs}} + \sum_{i=0}^{10} \gamma_i q_{t-i}^{\text{“Aggr.” HFTs}} + \sum_{i=0}^{10} \delta_i q_{t-i}^{\text{Others}} + \epsilon_{1t} \qquad (6)\\
q_t^{\text{“Pass.” HFTs}} &= \sum_{i=1}^{10} \mu_i r_{t-i} + \sum_{i=1}^{10} \nu_i q_{t-i}^{\text{“Pass.” HFTs}} + \sum_{i=1}^{10} \xi_i q_{t-i}^{\text{“Aggr.” HFTs}} + \sum_{i=1}^{10} \zeta_i q_{t-i}^{\text{Others}} + \epsilon_{2t}\\
q_t^{\text{“Aggr.” HFTs}} &= \sum_{i=1}^{10} \kappa_i r_{t-i} + \sum_{i=1}^{10} \lambda_i q_{t-i}^{\text{“Pass.” HFTs}} + \sum_{i=1}^{10} \pi_i q_{t-i}^{\text{“Aggr.” HFTs}} + \sum_{i=1}^{10} \rho_i q_{t-i}^{\text{Others}} + \epsilon_{3t}\\
q_t^{\text{Others}} &= \sum_{i=1}^{10} \phi_i r_{t-i} + \sum_{i=1}^{10} \chi_i q_{t-i}^{\text{“Pass.” HFTs}} + \sum_{i=1}^{10} \psi_i q_{t-i}^{\text{“Aggr.” HFTs}} + \sum_{i=1}^{10} \omega_i q_{t-i}^{\text{Others}} + \epsilon_{4t}
\end{aligned}
$$

where t gets updated whenever there is either a transaction or a quote update, r_t is the mid-quote return calculated from one event to the next and q_t^j takes the values $-1, 0, +1$ depending whether a member of trader category j sells aggressively, does not trade at all, or buys aggressively. By updating time this way, we always have a quote update when there is a transaction but not the other way around. That is, there can be a quote update without a contemporaneous transaction.

Apart from the return equation, the model features a transaction equation for each of the two groups of HFTs and one for all other traders. In this sense, our approach is similar to that of Barclay et al (2003), Chaboud et al (2011) and Hendershott and Riordan (2011). Transaction variables can contemporaneously influence the return variable but not vice versa. This implies that causality runs contemporaneously only from trades to quote revisions. This is true in an electronic order book since a market order of sufficient size can exceed the depth of the best available quote and thereby contemporaneously change the prevailing best quote.[22] Also, the model does not allow trades of the various groups to contemporaneously influence each other.[23] This is because we only examine transactions resulting from orders submitted to a single order book. A contemporaneous relationship between transactions is possible only if traders can divert their orders in multiple venues all of which are fully integrated.[24]

We estimate the VAR model for the entire week appropriately adjusting for end-of-day effects.[25] The error terms satisfy: $E[\epsilon_{it}\epsilon_{is}] = E[\epsilon_{it}\epsilon_{js}] = 0$ for $i \neq j$ and $s < t$. Although the inclusion of contemporaneous terms in the return equation implies that $E[\epsilon_{1t}\epsilon_{2t}] = E[\epsilon_{1t}\epsilon_{3t}] = E[\epsilon_{1t}\epsilon_{4t}] = 0$, the error terms of the transaction equations are contemporaneously correlated with one another because, at each tick, only one of the three transaction variables can be non-zero.[26] This means that $E[\epsilon\epsilon'] = \Sigma$, where ϵ is the error term vector and Σ has non-diagonal elements.

After estimating this model, we invert the lag polynomials and apply a Choleski decomposition on matrix Σ in order to obtain a vector moving average (VMA) form with orthogonal errors:[27]

$$\begin{bmatrix} r_t \\ q_t^{\text{“Pass.” HFTs}} \\ q_t^{\text{“Aggr.” HFTs}} \\ q_t^{\text{Others}} \end{bmatrix} = \begin{bmatrix} a(L) & b(L) & c(L) & d(L) \\ e(L) & f(L) & g(L) & h(L) \\ i(L) & j(L) & k(L) & l(L) \\ m(L) & n(L) & o(L) & p(L) \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \\ e_{4t} \end{bmatrix} \tag{7}$$

where each of $a(L)$ to $p(L)$ is an infinite order lag polynomial of the form: $a(L) = \sum_{i=0}^{\infty} a_i L^i$ and $E[ee'] = I$.

In this setup, the 50-period impulse response functions of the “passive” HFT group, the “aggressive” HFT group and the other traders are given by $\sum_{i=0}^{50} \hat{b}_i$, $\sum_{i=0}^{50} \hat{c}_i$ and $\sum_{i=0}^{50} \hat{d}_i$ respectively.[28] These sums are equal to the average aggregate price change, over 50 lags, following a single trade by each type of trader.[29]

Since the variance of the transformed error terms in the VMA model is unity, the impulse response functions capture both the impact on prices of a given trade as well as the actual number of trades executed.

In Figure 3 we plot the average (across stocks) impulse response functions of the three groups of traders. The confidence bands are at 95% and are calculated for each stock using bootstrap. The “aggressive” HFTs in our sample have a larger long-term price impact than “passive” HFTs (1.2 bps versus 0.5 bps). This implies that their collective trades contribute more to price discovery than those of the “passive” HFTs. However this larger contribution to information by the “aggressive” HFTs is roughly proportional to their larger aggressively executed trading volume[30], suggesting that “aggressive” HFTs’ individual trades are not more informative than those of the “passive” HFTs’. The rest of the traders have a much larger price impact due to their larger share of aggressively traded volume. On a per-share-traded basis however, “aggressive” and “passive” HFTs have a much larger price impact than the rest of the traders: Whereas HFTs initiate

[22] Although Hasbrouck (1991a) also assumes that transactions contemporaneously influence quotes, this happens in his paper for a different reason: The quotes are set by specialists who condition them on the incoming order flow. This creates a contemporaneous causal link from transactions to quotes.

[23] Nevertheless, our results are not sensitive to this assumption. As a robustness check we also estimate the VAR model assuming different causal orderings between the trades of the various groups and find no material differences.

[24] See for example Barclay et al (2003) who examine the choice of routing orders via ECNs versus Nasdaq market makers.

[25] The first 10 observations of each day are used as independent variables for the 11th observation but not as dependent variables themselves.

[26] Whenever there is a quote update, all three transaction variables are zero.

[27] If $A(L)x_t = \epsilon_t$ is the original VAR model and $x_t = B(L)\epsilon_t$ is the moving average form with non-orthogonal errors, then the moving average form with orthogonal errors is $x_t = C(L)e_t = B(L)Q^{-1}e_t$ where Q satisfies: $Q^{-1}Q^{-1'} = \Sigma$ and also $C(0) = B(0)Q^{-1}$ is upper-triangular.

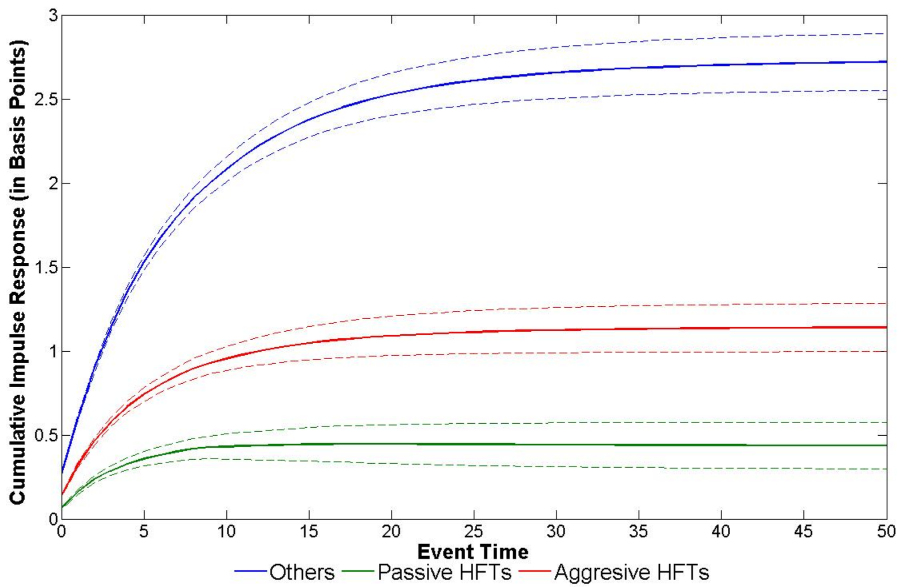
[28] It is worth noting that the impulse response function does not only capture the direct effect on prices of a trade by a given group but also the indirect effect of the trades of the other groups which occur in response to the original trade.

[29] For robustness, we also estimate impulse response functions over 100 lags. These results are similar to the numbers we report.

[30] 9.18% vs. 3.53%. See Table 2, Panel A

12.7% of all trading volume, their collective price impact is about 40% of the total price impact by all trader categories.

Figure 3: **Average, across stocks, 50-step impulse response functions measured in bps for the "passive" HFTs (green line), the "aggressive" HFTs (red line) and all other traders (blue line). The "passive" HFTs are those with a below-median aggressiveness ratio and the "aggressive" ones are those with an above-median aggressiveness ratio. We calculate the impulse response functions using LSE trades and quote updates. We first estimate the VAR model (6) for each stock and for the entire week-long time horizon of our sample. Then we convert it into the vector moving average form (7). The IRFs equal the sums of the estimated coefficients of the first 50 lags in the VMA model: $\sum_{i=0}^{50} \hat{b}_i$, $\sum_{i=0}^{50} \hat{c}_i$ and $\sum_{i=0}^{50} \hat{d}_i$. We do 1,000 bootstrap iterations to calculate 95% confidence bands for the impulse response functions of each stock.**



6.2 HFT contribution to information

We next utilise the variance decomposition method suggested by Hasbrouck (1991b) to assess the overall fraction of information and noise that the two HFT groups contribute through their trades. A trader can theoretically contribute both to information and to noise: think, for example, of an HFT, half of whose trades are informed with the rest being used to calibrate her inventory levels. We also examine how that contribution compares with the information and noise contributions of all other traders and to the information or noise arriving independently of trades.

The Hasbrouck (1991b) framework is suitable for such an exercise as it adds some structure to the price process by assuming that the observed price has a non-transitory (random walk) component that reflects the efficient price, and a transitory (stationary) component that reflects the residual noise. Thus, the observed price (i.e. the prevailing quote midpoint) is given by:

$$p_t = p_t^* + s_t \tag{8}$$

where the efficient price p_t^* is a random walk:

$$p_t^* = p_{t-1}^* + v_t, \quad v_t \sim iid(0, \sigma_v^2) \tag{9}$$

and the residual noise is assumed to have no long-term impact on prices, i.e. $\lim_{h\to\infty} E_t[s_{t+h}] = 0$. In other words, the efficient price and the noise are both defined in a statistical sense. Assuming that the efficient price follows a random walk, any stationary deviations from a random walk behaviour are classified as "noise". Given the above, the observed price innovation is then equal to:

$$r_t = p_t - p_{t-1} = \Delta p_t^* + \Delta s_t \tag{10}$$

where Δp_t^* is the permanent price effect due to information and Δs_t is the transitory effect due to noise. Using the VAR framework in equations (6) and (7), the permanent price effect is given by:

$$\Delta p_t^* = (\sum_{i=0}^{\infty} a_i)e_{1t} + (\sum_{i=0}^{\infty} b_i)e_{2t} + (\sum_{i=0}^{\infty} c_i)e_{3t} + (\sum_{i=0}^{\infty} d_i)e_{4t} \tag{11}$$

Since the VMA model errors are spherical with unit variance, the variance of the permanent price impact is:

$$\sigma_v^2 = \Big(\sum_{i=0}^{\infty} a_i\Big)^2 + \Big(\sum_{i=0}^{\infty} b_i\Big)^2 + \Big(\sum_{i=0}^{\infty} c_i\Big)^2 + \Big(\sum_{i=0}^{\infty} d_i\Big)^2 \tag{12}$$

The first term in the above sum represents the contribution to the efficient price innovation from public information and the rest of the terms represent the contributions from private information by the "passive" and "aggressive" HFTs and all other traders respectively.[31]

In Table 6 we report these contributions as a fraction of the total variance, i.e. we report the ratios:

$$\frac{\Big(\sum_{i=0}^{50} \hat{a}_i\Big)^2}{\hat{\sigma}_v^2}, \frac{\Big(\sum_{i=0}^{50} \hat{b}_i\Big)^2}{\hat{\sigma}_v^2}, \frac{\Big(\sum_{i=0}^{50} \hat{c}_i\Big)^2}{\hat{\sigma}_v^2}, \frac{\Big(\sum_{i=0}^{50} \hat{d}_i\Big)^2}{\hat{\sigma}_v^2} \tag{13}$$

where $\hat{\sigma}_v^2$ has also been estimated using 50 lags. The variance decomposition results show that across the four stocks of our sample "aggressive" HFTs contribute on average about 30% of the total information versus 20% for "passive" HFTs and about 33% for all other traders. The difference between "aggressive" and "passive" HFTs is primarily driven by stocks X and Y where the "aggressive" group makes a significantly larger (in the statistical sense) contribution to information than the "passive" group.[32] Public, non-trade related information accounts for about 17% of the efficient price variance. As expected, public information makes the largest relative contribution for stocks X and Z, the stocks with the largest market capitalisation and trading volumes (see Table 1): large companies generate more news as they are more frequently on the spotlight and they are followed more closely by the financial press.

Overall, these results are consistent with the ordering suggested by the impulse response functions and show that there are instances (e.g. stocks W and Z) where HFTs collectively contribute a large amount of information which is, in the statistical sense, significantly larger than that of the rest of the traders. They also hint to the fact that HFTs make a substantial contribution to price discovery in relative terms: while the HFTs in our sample participate in about 27% of the total traded volume and initiate about 13% of it, they collectively contribute more than half of the total private information that is impounded in prices via trades. However, to assess their overall informational efficiency, one needs to compare their contribution to price discovery with their contribution to noise. We do this next.

[31] The term "private information" is used here to describe the ability of traders to quickly and accurately process public information and subsequently make informed trades. In the same fashion, "public information" is any information in the public domain that requires little or no further processing to inform a trade (e.g. an earnings announcement). As such, "public information" is reflected in prices primarily via quote updates.

[32] "Passive" HFTs contribute less information to stock X as their overall aggressiveness in that stock is low (exact numbers not reported). This in turn is likely because "passive" HFTs in stock X execute a larger fraction of their standing limit orders, owing to the stock's large trading volume.

Table 6: **Efficient price variance decomposition results for trading done on the LSE**

Stock	**Private Information**			**Public Information**	**Differences**	
	"Passive" HFTs	**"Aggressive" HFTs**	**Others**		**"Aggressive" minus "Passive" HFTs**	**All HFTs minus Others**
W	31.62%	30.21%	30.88%	7.29%	-1.41%	**30.95%**
	(4.08%)	(2.81%)	(2.18%)	(0.57%)	(6.54%)	(5.55%)
X	3.30%	23.84%	34.15%	38.72%	**20.54%**	-7.02%
	(2.39%)	(2.78%)	(1.89%)	(2.44%)	(3.79%)	(5.03%)
Y	17.96%	35.44%	40.04%	6.55%	**17.48%**	13.37%
	(6.04%)	(3.22%)	(3.61%)	(0.64%)	(8.77%)	(8.17%)
Z	27.49%	29.13%	29.10%	14.27%	1.64%	**27.52%**
	(2.74%)	(2.57%)	(1.75%)	(0.99%)	(4.66%)	(4.01%)
Average	20.09%	29.66%	33.54%	16.71%	9.56%	16.21%

Notes: We first estimate the VAR model (6) for each stock and for the entire week-long time horizon of our sample. Then we convert it into the vector moving average form (7) and report the efficient price variance contribution ratios in (13). These ratios show the fraction of the efficient price change that is due to private and public information. The fraction of price change due to private information is further broken down between "passive" HFTs, "aggressive" HFTs and all other traders. The "passive" HFTs are those with a below-median aggressiveness ratio and the "aggressive" ones are those with an above-median aggressiveness ratio. We also report the difference in variance contributions between "aggressive" and "passive" HFTs and between all HFTs and all other traders. Statistically significant (at 5%) differences are in bold. Standard deviations are in parentheses and are calculated by 1,000 bootstrap iterations.

6.3 HFT contribution to noise

Here we assess the contribution of HFTs to noise, i.e. the transitory (and stationary) component of the efficient price. The goal is to measure the contributions of each type of trader to the variance of noise, σ_s. The structure imposed by equations (8), (9), (10) and (11) is still valid and the only additional assumption is that noise is a moving average (MA) process of the residuals of the VMA form, as in Hasbrouck (1993):

$$s_t = \sum_{i=0}^{\infty} \tilde{\alpha}_i e_{1t-i} + \sum_{i=0}^{\infty} \tilde{\beta}_i e_{2t-i} + \sum_{i=0}^{\infty} \tilde{\gamma}_i e_{3t-i} + \sum_{i=0}^{\infty} \tilde{\delta}_i e_{4t-i} \tag{14}$$

Although Hasbrouck (1993) uses actual transaction prices to calculate the returns r_t and to estimate σ_s, here we calculate returns using the quote midpoint. We do this because we do not want the bid-ask bounce to affect our noise estimates. Instead, we want to focus on the effect of HFT on the efficiency of the mid-quote.[33]

From the expression of returns r_t in the VMA model and equations (10), (11) and (14) it then follows that the noise process parameters are equal to:

$$\tilde{\alpha}_j = -\sum_{k=j+1}^{\infty} a_k, \quad \tilde{\beta}_j = -\sum_{k=j+1}^{\infty} b_k, \quad \tilde{\gamma}_j = -\sum_{k=j+1}^{\infty} c_k, \quad \tilde{\delta}_j = -\sum_{k=j+1}^{\infty} d_k \tag{15}$$

The variance of the noise is then given by:

$$\sigma_s^2 = \sum_{i=0}^{\infty} \left(\tilde{\alpha}_i^2 + \tilde{\beta}_i^2 + \tilde{\gamma}_i^2 + \tilde{\delta}_i^2 \right) \tag{16}$$

Because of the way this model is identified, this estimate is a lower bound for σ_s and therefore the results should be interpreted as such.[34] The estimated contributions to this variance (using 50 lags) by the various groups of traders are:

$$\frac{\sum_{i=0}^{50} \hat{\tilde{\alpha}}_i^2}{\hat{\sigma}_s^2}, \quad \frac{\sum_{i=0}^{50} \hat{\tilde{\beta}}_i^2}{\hat{\sigma}_s^2} \quad \frac{\sum_{i=0}^{50} \hat{\tilde{\gamma}}_i^2}{\hat{\sigma}_s^2}, \quad \frac{\sum_{i=0}^{50} \hat{\tilde{\delta}}_i^2}{\hat{\sigma}_s^2} \tag{17}$$

where $\hat{\sigma}_s^2$ is also estimated using 50 lags. The first ratio captures the contribution to noise by sources other than trading, which we term "public noise".[35] The rest of these ratios capture the contributions to noise by the "aggressive" group of HFTs, the "passive" group of HFTs and the rest of the traders respectively.

Table 7 shows the contributions to the noisy price innovations of each trader group. The first thing to notice is that "aggressive" and "passive" HFTs collectively contribute roughly 45% of the noise variance across all four stocks, although they initiate about 13% of the total trading volume in our sample. This is comparable to their overall contribution to price discovery of about 50% (see Table 6). Furthermore, there are instances where HFTs contribute significantly more noise than the rest of the traders. This is the case for stocks W and Z which are the stocks for which HFTs were also found to be contributing significantly more information. For stock X, HFTs contribute less noise than the rest of the traders but this is driven by the very low participation (and consequently low noise contribution) by the "passive" HFTs. Similar to their information contribution, the "aggressive" group contributes significantly more to noise than the "passive" one for stocks X and Y. It thus appears that the more HFTs trade aggressively the more they contribute to both price discovery and excess volatility.

It is not immediately clear why HFTs are large contributors to noise. However, several explanations come to mind: One is that HFTs end the day with relatively flat positions: If informed trading causes

[33] However, as a robustness check we also calculate the contributions to noise of each trader group using transaction prices. These calculations show a proportionally larger noise contribution by the two groups of HFTs and a smaller noise contribution by the rest of the traders. Therefore, the results we report are conservative as far as the HFT contribution to noise is concerned. The transaction price results are available upon request.

[34] See Hasbrouck (1993) for a discussion of the alternative identification strategies and their properties.

[35] This includes changes to the prevailing quote that may happen, absent a trade, as a response to news arriving at the market. The quote change will constitute noise to the extent it is disproportionate to the news.

Table 7: **Noisy price variance decomposition results for trading done on the LSE**

Stock	**Private Noise**			**Public Noise**	**Differences**	
	“Passive” HFTs	**“Aggressive” HFTs**	**Others**		**“Aggressive” minus “Passive” HFTs**	**All HFTs minus Others**
W	32.89%	28.92%	34.96%	3.24%	-3.97%	**26.84%**
	(6.61%)	(4.77%)	(4.61%)	(0.45%)	(10.43%)	(9.43%)
X	1.75%	18.95%	35.06%	44.23%	**17.19%**	**-14.36%**
	(1.76%)	(4.12%)	(2.43%)	(3.26%)	(4.83%)	(6.76%)
Y	10.06%	39.34%	48.96%	1.65%	**29.29**%	0.44%
	(4.41%)	(4.80%)	(4.63%)	(0.23%)	(7.90%)	(9.53%)
Z	23.95%	25.86%	37.31%	12.89%	1.91%	**12.49**%
	(4.56%)	(4.32%)	(4.03%)	(1.52%)	(7.29%)	(7.24%)
Average	17.16%	28.27%	39.07%	15.50%	11.11%	6.35%

Notes: We first estimate the VAR model (6) for each stock and for the entire week-long time horizon of our sample. We then convert the model into the vector moving average form (7) and report the noisy price variance contribution ratios in (17). These ratios show the fraction of the noisy price that is due to private and public information. The fraction of noise due to private investor trading is further broken down between “passive” HFTs, “aggressive” HFTs and all other traders. The “passive” HFTs are those with a below-median aggressiveness ratio and the “aggressive” ones are those with an above-median aggressiveness ratio. We also report the difference in variance contributions between “aggressive” and “passive” HFTs and between all HFTs and all other traders. Statistically significant (at 5%) differences are in bold. Standard deviations are in parentheses and are calculated by 1,000 bootstrap iterations.

an HFT to build up a position during the day, the HFT will have to unload this position by the end of the day. To the extent that there is no profitable way to do so, the HFT may be forced to make a series of "uninformed" trades to simply bring her position back to her target inventory levels. This could mean that inherent in the typical HFT strategy of carrying no positions overnight is the creation of noise. Alternatively, HFTs (especially the "aggressive" ones) may be employing trading strategies that rapidly process any signals that HFTs receive and trade accordingly. It could then be that owing to the speed at which they react, HFTs may be making errors in interpreting these signals or that the signals themselves are uninformative. Unfortunately, we cannot empirically disentangle these hypotheses.

Given the increased contribution of HFTs to both information and noise, it is important to make an overall assessment of their informational efficiency. In Table 8 we compare the overall informational efficiency of the various trader groups, by looking at the relative contributions of information and noise (i.e. the Information-to-Noise ratio) for each stock. It turns out that for all stocks HFTs are, as a group, about 30% informationally more efficient than the rest of the traders: The ratio of information-to-noise contributions is 1.13 for all HFTs, and 0.86 for all other traders, with the differences in this ratio being statistically significant for stocks Y and Z.

These results suggest that there are instances where HFTs' trades are, overall, informationally more efficient, in the sense that they contribute more information than noise. The previous tables suggest however that this is happening because HFTs amplify (or scale up) both the beneficial and detrimental components of price volatility: The overall improvement in price discovery that HFTs bring about comes part and parcel with a large amount of excess volatility.

The small cross-section of our sample means that we cannot generalise these results. However, our paper does establish that there are instances where HFT contributes significantly to both price discovery and noise. It is unclear what are the overall welfare implications of high-frequency trading in these instances. This will likely depend on the marginal utility and disutility of information and noise respectively. In other words, on how much additional noise we are willing to tolerate at some times for the benefit of more informed trading at others.

7 Summary and Conclusions

We utilise high-frequency transactions data from the UK equity markets to study the intraday HFT behaviour and its impact on market quality. Our sample covers four FTSE 100 stocks and one week of trading. While the small cross-section of our sample does not allow us to draw conclusions about the overall impact of HFT on market quality, it does allow us to demonstrate the potential impact of HFT at certain instances. The stocks in the sample are relatively diverse in terms of market capitalisation and traded volumes and the particular, randomly selected, week is not uncommon when compared with the other weeks of the same year. The fact that we observe trader identities allows us to study the behaviour of particular types of HFTs and thus disaggregate HFT behaviour.

We find that HFTs exhibit variability in their trading strategies which translates into differences in terms of trading speed, liquidity provision, price discovery and impact on short-term, excess volatility. More specifically, we analyse the behaviour of two groups of HFTs which differ in terms of their liquidity provision: one (labeled "passive") consists of those HFTs that mostly supply liquidity while the other (labeled "aggressive") consists of HFTs who mostly consume it. In terms of trading behaviour, we find that the "passive" HFTs alternate their trading direction from one second to the next and that their trading is price-neutral. On the contrary, the "aggressive" HFTs' position changes are trending and they tend to trade with the previous 10-second price trend. Given the need to maintain a balanced overall position, "aggressive" HFTs' trading is more sensitive in the long run to their inventory levels.

We find that both groups of HFTs increase their relative trading activity when volatility is higher and when the spread is tight. We articulate several explanations as to why this may be so. In terms of their impact on market quality, we show that there are instances where HFTs contribute significantly more to both price discovery and to noise than the rest of the traders. Furthermore, there are instances where "aggressive" HFTs are leading the "passive" ones in both information and noise contribution. In all of these

Table 8: **Ratios of information to noise contribution for trading done on the LSE**

Stock	**Information/Noise Ratio**				**Differences**	
	"Passive" HFTs	**"Aggressive" HFTs**	**All HFTs**	**Others**	**"Aggressive" minus "Passive" HFTs**	**All HFTs minus Others**
W	0.96	1.04	1.00	0.88	0.08	0.12
	(0.09)	(0.07)	(0.04)	(0.05)	(0.14)	(0.08)
X	1.88	1.26	1.31	0.97	-0.62	0.34
	(1.42)	(0.15)	(0.19)	(0.05)	(1.48)	(0.23)
Y	1.79	0.90	1.08	0.82	-0.89	**0.26**
	(0.41)	(0.08)	(0.07)	(0.07)	(0.47)	(0.13)
Z	1.15	1.13	1.14	0.78	-0.02	**0.36**
	(0.10)	(0.08)	(0.05)	(0.02)	(0.15)	(0.07)
Average	1.44	1.08	1.13	0.86	-0.36	0.27

Notes: For each stock and each trader group, the Table shows the ratio of information to noise contributions. It also shows the differences in this ratio between "aggressive" and "passive" HFTs and between all HFTs and all other traders. The "passive" HFTs are those with a below-median aggressiveness ratio and the "aggressive" ones are those with an above-median aggressiveness ratio. Standard errors are in parentheses. Statistically significant (at 5%) differences are in bold. Significance is established by 1,000 bootstrap iterations.

cases, HFTs' trading is, overall, informationally more efficient than that of the rest of the traders in the sense that they have a higher ratio of information to noise contribution. Nevertheless, the overall welfare implications of HFT are unclear; these will depend on how the marginal benefit of information at some times compares with the marginal cost of excess volatility at other times, including in periods of market stress. They will also depend on the propensity for errors in the operation of HFTs' algorithmic trading to cause harmful disturbances of the type experienced in the "Flash Crash" of 6 May 2010. However, these issues are beyond the scope of this paper.

Appendix 1: Filters applied on the SABRE data set

This appendix describes the procedure we follow in order to filter the raw SABRE transactions data. We first drop all transactions for which the reported venue in the SABRE reporting form is "XOFF". These are off-exchange reported OTC transactions. We then drop any transactions that were not reported on one of the four exchanges: LSE ("XLON"), Chi-X ("CHI-X"), BATS ("BATE") and Turquoise ("TRQX"). We do this in order to eliminate trades that were conducted in quote-driven markets (such as X-Plus), systematic internalisers (such as Nomura) and dark pools (such as Sigma-X). We next drop all transactions for which the reported venue is one of the above four exchanges (e.g. "XLON") but for which the BIC (Business Identification Code) of the clearing house is not reported as a counterparty in the "counterparty1" field. These are on-exchange reported but off-the-order book executed OTC transactions. In particular, we drop all transactions which do not have one of the following clearing house identifiers in the SABRE form "counterparty1" field: "LCHLGB2EXXX" and "LCHLGB2LXXX" for LCH, "EMCFNL2AXXX", "EMCFNL21XXX", "EMCFNL22XXX" and "EMCFNL2A" for EMCF, "ECCPGB22XXX", "ECCPGB2LGLC" and "ECCPGB2LXXX" for EuroCCP and "CLRXCHZZXXX" for SIX X Clear. We however retain transactions where the "counterparty1" field has erroneously been filled with the name of an exchange instead of the BIC of the clearing house. These transactions have been executed on an order book. The values in the "counterparty1" field for which the transactions are retained are: "XLON:LSE", "XLON:" "TRQX:null", "CHIX:CHI-X" and "BATE".

The next step is to remove from the data set those transactions that were executed on the dark pools of BATS, Chi-X and Turquoise as well as any transactions on the limit order book that are the result of "iceberg orders"[36]. Since SABRE does not have identifiers for these types of trades, we rely on Bloomberg to exclude them from our sample. In particular, we match the SABRE transaction reports with their Bloomberg counterparts (where the matching is based on time, price and trade size) and drop all SABRE transactions whose matches are classified by Bloomberg as "dark" or "iceberg". The Table below lists the exact Bloomberg condition codes for which a trade is dropped. For LSE, we only retain those transactions that were generated by the system through automatic execution (Bloomberg condition code "AT"). For this exchange we are not able to remove trades resulting from iceberg orders, as Bloomberg condition codes do not distinguish such trades.

Exchange	*Dark order Condition Code*	*Iceberg Order Condition Code*
BATS UK	BD	DB, DS, DT
Chi-X	CD	DT
Turquoise UK	DD	DV

[36] Iceberg orders are limit orders that rest on the order book but which are partially or fully hidden.

Appendix 2: Description and Implementation of the Lee-Ready algorithm

To classify trades as buyer or seller-initiated, we apply the Lee-Ready (1991) algorithm. The algorithm classifies a trade as buyer- (seller-) initiated if it is executed above (below) the midpoint of the bid and ask quotes. The rationale is that a trade executed close to the bid (ask) price is more likely to be seller- (buyer-) initiated. For the trades that lie on the bid-ask midpoint, the "tick test" is used for their classification. This procedure classifies a midpoint trade as buyer-initiated if it is executed at a higher price than the previous trade (i.e., if it is an "uptick") and as seller-initiated if it is executed at a lower price ("downtick"). If the previous trade is executed at the same price, the tick test looks at the next most recent trade until it reaches a change in the trade price ("zero uptick" or "zero downtick").

To implement the Lee-Ready algorithm, we go through the following steps:

1 *Top of order book reconstruction*: We first reconstruct the top of the order books for each of the four exchanges that we consider: LSE, BATS, Chi-X and Turquoise. Since the quote data from Bloomberg is reported in chronological order, the bid and ask quotes at a given time are set equal to the most recently reported quotes. The table below illustrates this process.

	Quote	*Top of book*
ASK	5	–
BID	3	3-5
BID	2	2-5
ASK	4	2-4
ASK	6	2-6

2 *Matching trades with quotes*: The next step is to match transactions with same-second and same-venue top of the book quotes. This is not straightforward because for each second there are usually multiple top of the book quotes.[37] To overcome this, we compare, for each time stamp, the transaction price with a spread equal to the average of the actual spreads within the same time stamp.[38] If the transaction price equals the midpoint of the "average spread", we apply the tick test and compare the transaction price with the most recent different transaction price.

Finally, unlike in the US, UK (and European) exchanges are not directly connected to each other, meaning that there is no single consolidated order book for any given stock. Instead, exchanges are indirectly connected through firms that have multiple exchange memberships. For this reason, we recreate the tops of all four exchange order books and match each transaction with the prevailing quote of the venue in which it took place.

[37]For both SABRE and Bloomberg, time granularity is at one second. In the presence of high-frequency traders who can post multiple quotes in a second, this level of granularity does not allow one to associate each trade and quote with a unique time stamp.

[38]If for example, the top of the book quotes in the above table were reported in the same second, the average bid and ask for that second would be 2.25 and 5 respectively.

References

[1] Authority for the Financial Markets (AFM), 2010, "High frequency trading: The application of advanced trading technology in the European marketplace", Report of the Dutch Authority for the Financial Markets, November, available at:
http://www.afm.nl/layouts/afm/default.aspx /media/files/rapport/2010/hft-report-engels.ashx

[2] Barclay M, Hendershott T and T McCornick, 2003, "Competition among Trading Venues: Information and Trading on Electronic Communication Networks", *Journal of Finance*, 58, 2637-2665

[3] Boehmer E, Fong K and J Wu, 2012, "International Evidence on Algorithmic Trading", *working paper*

[4] Brogaard J, 2011, "High Frequency Trading and its Implication on Market Quality", *working paper*

[5] Brogaard J, Hendershott T, Hunt S, Latza T, Pedace L and C Ysusi, 2012, "High Frerquency Trading and the Execution Costs of Institutional Investors", *working paper*

[6] CFTC-SEC, 2010, "Findings Regarding the Market Events of May 6, 2010", Report of the staffs of the CFTC and SEC to the joint advisory committee on emerging regulatory issues, September, available at: http://www.sec.gov/news/studies/2010/marketevents-report.pdf

[7] Chaboud A, Chiquoine B, Hjalmarsson E and C Vega, 2011, "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market", *FRB International Finance Discussion Paper No. 980*

[8] Haldane A, 2011, "The Race to Zero", Speech, International Economic Association Sixteenth World Congress, Beijing, July 8; available at:
http://www.bankofengland.co.uk/publications/Documents/speeches/2011/speech509.pdf

[9] Hasbrouck J, 1991a, "Measuring the Information Content of Stock Trades", *Journal of Finance*, 46, 179-207

[10] Hasbrouck J, 1991b, "The Summary Informativeness of Stock Trades: An Econometric Analysis", *Review of Financial Studies*, 4, 571-595

[11] Hasbrouck J, 1993, "Assessing the Quality of a Security Market: A New Approcah to Transaction-Cost Measurement", *Review of Financial Studies*, 6, 191-212

[12] Hasbrouck J and G Saar, 2011, "Low-Latency Trading", *working paper*

[13] Hendershott T, Jones C and A Menkveld, 2011, "Does Algorithmic Trading Improve Liquidity?", *Journal of Finance*, 66, 1-33

[14] Hendershott T and P Moulton, 2011, "Automation, Speed and Stock Market Quality: The NYSE's Hybrid", *Journal of Financial Markets*, 14, 568-604

[15] Hendershott T and R Riordan, 2011, "Algorithmic Trading and Information", *NET Institute Working Paper No 09-08*

[16] IOSCO, 2011, "Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency", Consultation Report, available at:
http://www.iosco.org/library/pubdocs/pdf/ioscopd354.pdf

[17] Kirilenko A., Kyle A., Samadi M. and T. Tuzun, 2011, "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market", *working paper*

[18] Lee M. C. and M. Ready, 1991, "Inferring Trade Direction from Intraday Data", *Journal of Finance*, 46, 733-746

[19] Menkveld A, 2012, "High Frequency Trading and the *New-Market* Makers", *working paper*

[20] Nickel S., 1981, "Biases in Dynamic Models with Fixed Effects", *Econometrica*, 49, 1417-1426

[21] Shapiro M, 2012, "Introductory Remarks at SEC's Market Technology Roundtable", Washington DC, October 2; available at: http://www.sec.gov/news/speech/2012/spch100212mls.htm

[22] Schwert G. W., 1989, "Why Does Stock Market Volatility Change Over Time?", *Journal of Finance*, 44, 1115-1153.

[23] Zhang F., 2010, "The Effect of High-Frequency Trading on Stock Volatility and Price Discovery", *working paper*