

Price discovery and the cross-section of high-frequency trading

Evangelos Benos*

Satchit Sagade†

Abstract

We quantify the price discovery contribution of high-frequency traders (HFTs) in the United Kingdom equity market and examine how it varies in their cross-section. For this, we group individual HFTs according to their liquidity taking/making activity. HFTs contribute about 14% of all trade-induced information, with aggressive HFTs accounting for two-thirds of this contribution. This suggests that HFTs who pursue strategies that require use of aggressive trades are the most informed, as opposed to passive HFTs who more likely act as market makers. However, information shares decline with the amount of aggressive volume, suggesting that HFTs' news trading strategies are not scalable.

JEL classification: G10

Keywords: High-Frequency Trading, Price Discovery

*Bank of England, Threadneedle Street, London, EC2R 8AH, UK evangelos.benos@bankofengland.co.uk.

†Department of Finance and Research Center SAFE, Goethe University, Frankfurt, +49 69 798 30085, sagade@safe.uni-frankfurt.de. Corresponding author.

This is a revised version of an earlier paper titled: “High-Frequency Trading Behaviour and Its Impact on Market Quality: Evidence from the UK Equity Market.” This version utilizes a significantly expanded dataset. We would like to thank Ana Fernandes and Peter Mason of the Financial Conduct Authority for being so helpful with our many questions about the ZEN database. We would also like to thank Alfonso Dufour, Simon Hargreaves, Stefan Hunt, Tim Rowe, Edwin Schooling-Latter, Erik Theissen, Michalis Vasios, Christian Westheide, Filip Žikès, an anonymous referee and seminar participants at the Bank of England, the Copenhagen Business School, the Norges Bank, the UK Financial Conduct Authority, and the ICMA Centre for useful comments and suggestions. Satchit Sagade gratefully acknowledges research support from the Research Center SAFE, funded by the State of Hessen initiative for research LOEWE. The views expressed in this paper do not necessarily reflect those of the Bank of England or the Financial Conduct Authority.

1 Introduction

Over the past few years, “High-Frequency Trading” (HFT) activity has become prevalent in a number of securities and derivatives markets, accounting for a significant fraction of the trading volume.^{1,2} As a result, and following the May 6, 2010 “Flash Crash” in the United States, HFT has caught the attention of academics and policy makers worldwide. The debate has mainly centered around the impact of HFT on various aspects of market quality, such as liquidity and price discovery.³ Indeed, regulators on both sides of the Atlantic have implemented or are contemplating rules aimed at mitigating any potentially detrimental effects of HFT activity on market quality.⁴

This paper utilizes proprietary transactions data from the equity market in the United Kingdom (U.K.), over a four-month period, to assess and quantify the impact of various types of HFT firms on price discovery and also compare and contrast their information contribution against each other and against all other market participants. Thus, in this paper we extend previous work on the subject (e.g., Brogaard, Hendershott, and Riordan (2014a) and references therein) by examining how price discovery varies in the cross-section of HFT firms.

The data we use is uniquely suited for this exercise as it allows us to identify market participants and track the activity of individual HFT firms. Additionally, our data captures the activity of some of the largest stand-alone HFTs, in one of the world’s major stock markets. Although we observe activity at the firm level, for confidentiality reasons we group HFT firms according to their overall liquidity taking/making behavior into

¹We use the acronym “HFT” to denote both high-frequency trading activity and high-frequency traders.

²For a detailed overview of HFT characteristics, see AFM (2010), SEC (2010), and Jones (2013).

³See CFTC-SEC (2010) for the official report on the Flash Crash. See Haldane (2011) and Shapiro (2012) for a regulatory perspective of HFT.

⁴For instance, as part of the revision of the Markets in Financial Instruments Directive (MiFID) in Europe, regulators will require HFT firms to provide a detailed description of the algorithms being used and to test-run these algorithms extensively before rolling them out. MiFID will also limit the ratio of submitted orders to executed trades, which is bound to affect HFT activity. (See Articles 17.2 and 48.6-9 of the MiFID document available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0065>.) In the U.S., the Securities and Exchange Commission (SEC) is contemplating “anti-disruptive” trading rules as a means of inhibiting excess volatility that might be caused by HFT. See the relevant section of a June 5, 2014 speech by SEC Commissioner Mary Jo White, available at: <https://www.sec.gov/News/Speech/Detail/Speech/1370542004312>.

“passive,” “neutral,” and “aggressive” groups and study the information contribution of each group.

We start our analysis by examining the inventory dynamics of the above groups, across the four largest U.K. lit venues. We do this in order to see how HFT aggressiveness (our grouping criterion) is associated with HFTs’ trading patterns. We find that over longer horizons (measured in hours), the inventories of all groups are mean-reverting and exhibit very similar patterns of serial correlation. However, over shorter horizons (measured in seconds), the various groups manage their inventories in markedly different ways. In particular, aggressive HFTs appear to be insensitive to recent (i.e., 10 second) price changes, consistent with news trading. Neutral HFTs are trend chasers (i.e., they trade in the direction of recent price changes), which is suggestive of momentum strategies. Finally, passive HFTs trade in the opposite direction of only the previous second price change, which is consistent with a market making strategy.⁵ At the same time, aggressive and passive HFTs are more sensitive to their accumulated inventory level, whereas neutral HFTs are less so and as a result accumulate larger intra-day positions. The analysis of HFT inventory dynamics reveals substantial heterogeneity in trading patterns. This raises the question of how HFT firms with such different trading patterns contribute to price discovery and motivates the rest of the paper.

To examine how information contribution varies in the cross-section of HFT firms, we look at the price impact of the liquidity-taking trades of the HFT firms in each group. We focus our price discovery analysis on a single order book and for this reason we use data on trades exclusively executed on the London Stock Exchange (LSE), the largest U.K. lit venue by trading volume. Our LSE data captures a larger share of overall market activity than the data used previously to study HFT. This is because lit volume in the U.K. is less fragmented than in the U.S., where most other HFT studies use data from, and as such, the LSE attracts a larger share of lit volume than any of the U.S. exchanges individually does. For example, during our sample period and depending on the particular stock, the LSE accounted for 55%-70% of the total U.K. lit volume.

⁵A price increase (decrease) is likely to hit an ask (bid) quote and thus trigger a sell (buy) by a HFT present at the top of the book.

For our analysis, we first estimate a calendar time specification where we regress price changes on contemporaneous and lagged group-level HFT order flow. We find that the order flow of aggressive HFTs is in the direction of future price changes whereas future prices are either insensitive or move in the opposite direction of the order flow of the other groups. This is consistent with the order flow of aggressive HFTs having a larger information content than that of the other groups.

Next, we use the VAR framework introduced by Hasbrouck (1991a, 1991b) to examine, in tick time, the information contribution of the various groups. In tick time, one can sequence transactions and quote updates and, using an appropriate econometric specification, it is possible to more accurately isolate a (Granger) causal link between trades and price changes. Thus, we avoid the potential endogeneity bias of clock time specifications with contemporaneous variables. In the Hasbrouck VAR framework, a trade in the direction of a permanent price change is associated with price discovery whereas a trade in the direction of a transitory price change is associated with noise. We therefore assess how much the aggressive trades of each group and of the other market participants contribute to the variance of permanent price changes and compare these contributions with each other.

We find that HFTs collectively contribute about 6% of the total information through their aggressive trades, the rest of the market participants contribute about 37%, while quote updates account for the remaining 57.3%. Furthermore, HFT information contribution varies substantially across groups, with aggressive, neutral, and passive HFTs contributing respectively 4.1%, 1.3%, and 0.7% of the total information. This HFT information contribution amounts to about 14% of all trade-induced information contribution.

The larger information contribution by aggressive HFTs is consistent with them following news trading or statistical arbitrage strategies (both of which rely heavily on aggressive trades). The limited amount of information that the passive group contributes is consistent with market making activity, which involves posting limit orders and only using aggressive trades sporadically to make quick inventory adjustments. It should be noted however that these results do not necessarily imply direct causation from HFTs'

trades to permanent price changes. HFTs could simply be anticipating quote updates or trades by other informed investors rather than generating the information themselves. In the context of the Hasbrouck model we cannot distinguish between these alternative explanations and even if an HFT's trades simply predict - rather than cause - permanent price changes, this still counts as information contribution.⁶

When scaling the shares of information contribution by the shares of the aggressively executed volume of each group, we find that aggressive HFTs have a lower information-to-volume ratio than neutral HFTs, which in turn have a lower ratio than passive HFTs (0.81 vs. 1.80 vs. 4.96). We interpret this as evidence of the lack of scalability in the types of strategies that at least the most aggressive HFTs pursue. Intuitively, as the set of signals to which HFTs choose to respond via aggressive trades expands, the average amount and/or quality of information that can be extracted and traded upon declines. This can either be because the signals become noisier, or because they become more complex and thus harder to process, or because the HFTs are capacity-constrained and are forced to allocate less computing power per signal, as their number increases. Alternatively, it could be because of sheer competition among HFTs.⁷ These findings also cast doubt on the notion that passively trading, market-making, HFTs only use aggressive orders to make inventory adjustments. Rather, they point to the presence of multiple trading desks and strategies even within these predominantly passively trading firms.

Our paper contributes to a fast-growing literature on HFT and algorithmic trading. The impact of the various HFT strategies on price efficiency is unclear. For example, in Martinez and Rosu (2013) and Foucault, Hombert, and Rosu (2015), HFTs react faster to news and therefore cause prices to become more efficient. On the other hand, in Jarrow and Protter (2012), HFTs react simultaneously to common signals, which creates an excessive price impact and moves prices away from fundamentals. Our paper is closest in scope to Brogaard, Hendershott, and Riordan (2014a), who use NASDAQ transactions

⁶We thank an anonymous referee for highlighting this point.

⁷For example, Baron et al. (2016) show that HFTs compete fiercely in the E-mini market to the extent that new entrants are often forced to exit the market.

to establish that, overall, HFTs contribute to price efficiency by trading in the direction of permanent price changes and in the opposite direction of pricing errors. We use a similar framework as they do, but focus instead on how HFT price discovery varies in the cross-section of HFT firms. Additionally, we provide an estimate of the absolute amount of information impounded onto prices via aggressive HFT trades. Related work also includes Hirschey (2013), who uses the same NASDAQ data as Brogaard, Hendershott, and Riordan (2014a) and finds that HFT firms can predict the order flow of non-HFT ones. This may suggest that HFTs could be free-riding on the information acquisition efforts of other market participants. In our paper, we do not attempt to explain and test the reasons why some types of HFTs contribute more information than others. Our paper is also related to Benos et al. (2015), who document that HFT activity is more correlated than that of a control group of firms and that instances of correlated trading are associated with permanent price impacts, suggesting that they are information-driven.

Other related papers include Hendershott and Riordan (2011), who find that for DAX-traded equities, algorithmic trades (not just HFT ones) contribute more to price discovery than human trades, Hendershott et al. (2011), who find that algorithmic trading enhances the informativeness of quotes on the NYSE, and Chaboud et al. (2014), who show that algorithmic traders improve the efficiency of the foreign exchange market as captured by the frequency of triangular arbitrage opportunities and the autocorrelation of high-frequency returns. Also, Hendershott and Moulton (2011) look at price efficiency before and after the introduction of the NYSE-hybrid market. They find that with the NYSE-hybrid market in place, order execution times fell, thus causing prices to adjust faster to new information. However, these authors do not focus specifically on HFT activity. Other HFT studies that employ proprietary data are Hagströmer and Nordén (2013), who give a thorough description of the activity of the various HFTs in the Swedish stock market and Brogaard et al. (2014b) who study the impact of HFT on the execution costs of institutional investors. Finally, our paper is related to recent studies of HFT competition and industry structure. These include Baron et al. (2016), who show that the HFT industry has a “winner-takes-all” structure, whereby most profits accrue to

a few incumbent firms and Boehmer, Li, and Saar (2015), who document competition among HFTs in a variety of common strategies.

The rest of the paper is organized as follows: In Section 2, we describe the data and present summary statistics on the behavior of the HFT firms in our sample. In Section 3, we analyze the inventory dynamics of each group and in Section 4 we examine the cross-section of information contribution. We conclude in Section 5.

2 Data and summary statistics

2.1 Data properties and filters

Our main dataset consists of transaction reports on 92 FTSE 100 stocks,⁸ over the four-month period from September 1, to December 31, 2012, executed on the four largest (by volume) U.K. venues: LSE, Chi-X, BATS, and Turquoise. The data are obtained from the ZEN database, which is maintained by the U.K. Financial Conduct Authority (FCA).⁹ The trader-submitted reports contain information on the execution price, trade size, time stamp (accurate to the second), location, and, importantly, submitter identity. The reports also include a buyer/seller flag and information on whether the submitter executes a given transaction in a principal or agent capacity. The ZEN database captures the trading activity of all firms directly regulated by the FCA, as well as that of firms that trade through a broker (brokers are regulated and as such must report their clients' transactions).

Knowledge of the identities of the transaction report submitters allows us to precisely identify individual HFT firms. We classify a firm in our sample as a HFT one based on discussions with FCA supervisors.¹⁰ We thus identify 28 firms as HFTs; however, we do our analysis using the ten largest firms, which account for more than 95% of the

⁸These are the stocks that always remained in the FTSE 100 Index during our sample period.

⁹The FCA is the financial regulatory body in the U.K tasked with supervising financial markets, exchanges, and firms.

¹⁰Classification as an HFT firm is based on a number of criteria, such as the frequency of orders and trades, the ratio of orders to executed trades, the amount of overnight positions held, the lifetime of limit orders, whether a firm uses proprietary capital, as well as whether it utilizes latency-reducing technologies. To be classified as an HFT, a firm would have to satisfy multiple of these criteria.

total volume of the 28 firms. For confidentiality purposes, we do not report statistics on individual firms; instead we allocate the HFT firms into three groups according to their overall liquidity making/taking activity (i.e., their aggressiveness), and report statistics for each group separately.

One limitation of the data is that we do not observe the trades of any HFTs that are direct members of the various exchanges but are not FCA-regulated. Nevertheless, almost all large proprietary HFT firms operate in the U.K. as local legal entities and as such are captured in our data. We also cannot identify the activity of HFT desks of larger institutions with multiple trading desks (such as investment banks) operating in the same market. This is because the trades of these institutions are all reported under a single name. Furthermore, although we do observe the behavior of HFTs that trade through a broker, it is not easy to identify these trades as the various brokers use internal codes to identify their client HFTs. Nevertheless, the majority of stand-alone HFTs trade through their own accounts as sponsored access would likely slow them down. As such, these HFTs are captured in our data.

We complement the ZEN data with Bloomberg data on trades and quotes from the LSE. The Bloomberg data have three useful properties. First, the data contain the LSE-executed trades of all market participants regardless of whether they are FCA regulated or not. This allows us to accurately estimate the volume participation of the HFT firms in our sample. Second, Bloomberg transaction reports with the same second-level time stamp are chronologically ordered. Thus, by matching the ZEN and Bloomberg trade reports for the HFTs, we can precisely identify the exact sequence of HFTs' trades and also how they are ordered relative to all other trades. Finally, the Bloomberg data include all quote updates at the top of the LSE order book. Thus, by matching the HFTs' ZEN trade reports with Bloomberg quotes we can identify the direction of each trade executed on the LSE by the HFTs using the Lee and Ready (1991) algorithm. In the Appendix, we explain in more detail exactly how we apply the Lee-Ready algorithm to our data. Given that the Bloomberg data contain no information on trader identity, the matching itself is done by time stamp, trade size, and trade price. The matching accuracy is high as we

are able to match more than 99% of ZEN HFT reports with their equivalent Bloomberg reports. Over our sample period, LSE daily volumes accounted for 55% to 70% of the total U.K. lit volume for FTSE 100 stocks.¹¹

For our analysis, we apply a number of additional filters on the ZEN and Bloomberg data sets. Since we are interested in the impact of HFT activity on lit markets, we exclude any over-the-counter (OTC) transactions, as well as any transactions executed on dark pools. We also only retain matched reports that are time stamped after 8:01am and before 4:30pm so as to exclude trades that are associated with the opening and closing auctions. After applying these filters, we are left with 83% of the total volume executed by all HFTs in our sample.¹²

2.2 Variable definitions

Using the ZEN transactional data, we first construct a number of variables in order to describe overall HFT activity. Let Vlm_{istk} denote the signed volume of transaction k in stock s on day t by HFT firm i . Therefore, the daily volume on that stock traded by the HFT is given by:

$$Vlm_{ist} = \sum_k |Vlm_{istk}| \quad (1)$$

and the total value traded on that stock-day is:

$$Value_{ist} = \sum_k |Vlm_{istk}| P_{istk}, \quad (2)$$

where P_{istk} is the execution price of trade k . Assuming that HFTs start the day with a flat position, their end-of-day position for stock s on day t is:

$$EoDPOS_{ist} = \sum_k Vlm_{istk}. \quad (3)$$

¹¹Fidessa Fragmentation Index available at <http://fragmentation.fidessa.com/fragulator/>.

¹²ZEN OTC/dark trade reports appear to be less accurately time stamped than order book trade reports. For instance, for a subset of the HFT firms in our sample, we could only match 50% of the ZEN reports with a same-second Bloomberg trade report.

Given that HFTs generally end the day with a relatively flat position, it is also interesting to see what their maximum absolute intra-day position is for any given stock. This can be calculated as:

$$IDMax_{ist} = \max_k \left| \sum_k Vlm_{istk} \right|, \quad (4)$$

where again, k denotes trades. Using the above variables, we can also calculate a HFT's trading revenue. For firm i , stock s , and day t , the trading revenue is:

$$Rev_{ist} = \sum_k -Vlm_{istk}P_{istk} + EoDPOS_{ist}P_{st}, \quad (5)$$

where P_{st} is the closing price on day t for stock s , the assumption being that any outstanding positions at the end of the day are liquidated at the closing price.

Finally, using LSE quotes, we classify LSE-executed trades by HFTs as either liquidity making or liquidity taking. These are trades that are executed via limit and market orders respectively and are classified as such using the Lee and Ready (1991) algorithm. In the Appendix, we explain in detail how this is done. After classifying each transaction k of firm i as either aggressive or passive, we calculate the firm's volume-weighted aggressiveness ratio for stock s , on day t , i.e., the ratio of volume traded aggressively divided by its total trading volume:

$$AggrRatio_{ist} = \frac{Vlm_{ist}^{take}}{Vlm_{ist}^{make} + Vlm_{ist}^{take}}, \quad (6)$$

where the superscripts indicate whether the trading volume has been aggressively (“take”) or passively (“make”) executed. Thus, the aggressiveness ratio shows the extent to which a HFT supplies or consumes liquidity. In order to highlight the diversity of HFT strategies and examine separately their impact on price discovery, we split the HFTs in our sample into three groups based on their median aggressiveness ratio over stock-days: the “aggressive” ones (with $median(AggrRatio) > 0.67$) the “neutral” ones (with $0.33 < median(AggrRatio) \leq 0.67$), and the “passive” ones (with $median(AggrRatio) \leq$

0.33).¹³

2.3 Summary statistics

Table 1 shows the summary statistics over HFT-stock-days for the variables described above. These statistics have been calculated both for all HFTs grouped together and separately for each of the three groups (aggressive, neutral, and passive). We also report volume-weighted averages and standard deviations for some of the variables so as to more accurately describe aggregate HFT activity. In conjunction with Table 1, the various plots of Figure 1 show the actual frequency distributions, over HFT-stock-days, of some of the variables.

The average HFT firm trades about 360,000 shares and £2 million in a given stock, on a given day, across the four U.K. lit venues. This amounts to about 3% of the daily volume for that stock, or up to a maximum of 30% for all the HFTs in our sample collectively.¹⁴ However, the large differences between average and median volumes and values suggest that there is substantial heterogeneity across HFTs, with few large HFTs skewing the statistics. The group statistics indicate that these large HFTs tend to be aggressive traders: Indeed, the median volumes and values of the aggressive HFTs are twice as large as those of the neutral and passive HFTs.

The HFTs in our sample trade on average 439 times per stock-day. As with traded volumes and values, this number is skewed by a few HFTs since the mean number of trades is substantially larger than the median of 224 trades. The frequency distribution on the number of trades in Figure 1(b) shows that HFT activity is variable: there are stock-days where some HFTs barely trade and others when they trade heavily. In fact, HFTs trade fewer than 20 times in about 12% of the stock-days and importantly, this is also the mode of the frequency distribution. Additionally, it is in less than 3% of all

¹³We use fixed breakpoints instead of terciles in order to classify HFTs so as to ensure that the classification accurately reflects the activity of each HFT. If we used terciles, a hypothetical HFT with an aggressiveness ratio of above 50% could in principle be classified as passive.

¹⁴To the extent that HFTs trade with each other, adding up their volume participation will result in double-counting. This number is therefore an upper bound of HFT total participation and is attained if the HFTs in our sample do not trade with each other. Unfortunately, the data does not allow us to assess the extent to which HFTs trade with each other.

stock-days that a HFT trades more than 2,500 times per stock-day. The most frequently trading firms are the aggressive HFTs followed by the passive and neutral ones.

Over our sample period, the average HFT generates a trading revenue of about £760 per stock, per day, before rebates and fees, across all four lit venues. However, the median HFT generates only £20 per stock-day while the 5th and 95th percentiles are in the order of thousands of pounds. This is also true for each group and suggests that there is some time variation in the generated profits. This, in turn, highlights the fact that the business model of HFTs is largely based on diversification over many trades: although on a given stock-day, HFTs may experience substantial gains or losses, they generate on average a modest trading surplus.

HFTs also end the day with relatively flat positions (10% of the daily volume traded on the limit order book) with the aggressive and passive HFTs ending the day with even smaller relative positions at 6% and 7% of their daily volume respectively. When one also takes into account trades executed by HFTs off the four main limit order books, the relative end-of-day position becomes even smaller at about 5% for all HFTs. This suggests that HFTs use to some extent the OTC market in order to make inventory adjustments. Additionally, the maximum intra-day positions reached by HFTs are not substantial, suggesting that HFTs tightly manage their inventories throughout the day. This is also evident from the number of zero crossings (i.e., instances when their inventory shifts between values that are above and below of the daily average): on average, HFTs experience about 15 zero crossings on a stock-day basis but again, this number is skewed by a few HFTs in the neutral category with a much larger number of crossings. Thus, the median HFT experiences only six zero crossings on a given stock-day.

On a stock-day basis, HFTs are more frequently passive than they are aggressive, with an average ratio of 46%. In value terms however, they are mostly aggressive with a value-weighted average of 52%.¹⁵ Nevertheless, these numbers mask the substantial heterogeneity of average aggressiveness across HFTs. This is evident by comparing the average ratios of each of the three groups: the aggressive HFTs take liquidity 82% of the

¹⁵This is in contrast to Hagströmer and Nordén (2013), who find that on NASDAQ-OMX Stockholm exchange, liquidity supplying HFTs account for about 57%-59% of the value traded.

time, whereas the passive HFTs do so only 11% of the time. Figure 2 shows the average aggressiveness ratio of each of the three groups and for all HFTs over time. It is evident that these ratios are relatively stable, suggesting that HFTs do not significantly alter their liquidity taking/making behavior over time.

Overall, the summary statistics suggest that HFT behavior is very diverse: for example, the low aggressiveness ratios of the passive firms are consistent with high-frequency market-making, whereas the aggressive group of HFTs could be employing some other high-frequency strategy that relies more heavily on a speedy order execution, such as statistical arbitrage or event trading. Thus, in what follows, we examine separately the activity of each group, along with its impact on price discovery.

3 HFT inventory dynamics

We start by looking at the long- and short-term dynamics of HFT inventories. Our goal is to see whether the different aggressiveness ratios of the groups also translate into markedly different inventory patterns. This is motivated by the intuition that different patterns of liquidity taking/making activity should be associated with markedly different strategies. For example, SEC (2010) suggests that the primary source of profits for market-making HFT strategies is the bid-ask spread and the liquidity rebates offered by the various venues. This implies that market-making HFTs might be sensitive to their inventory levels and therefore quickly reverse a position in order to reduce their exposure to unanticipated price changes. SEC (2010) also describes two additional types of HFT strategies: arbitrage strategies try to exploit pricing differences between related assets (e.g., an ETF and its components) while directional strategies involve taking an unhedged position based on anticipated price changes (e.g., exploiting differences between observed and fundamental values). HFTs employing either of these strategies might have a lower short-term sensitivity to inventory levels due to the need to hold on to a position until prices move in the desired direction. In addition, a directional strategy exploiting momentum in prices would involve trading in the direction of the short-term price trend.

3.1 HFT long-term inventory dynamics

Although the end-of-day positions reported in Table 1 are not large and do not vary much between the three groups, they are still not zero, suggesting that perhaps some longer-term dynamics might be at play. To examine this, we look at lagged autocorrelations of inventories sampled at various frequencies.¹⁶ For this, we first construct long-term inventory positions for each HFT and stock by carrying over any previous day outstanding positions. We then calculate autocorrelations of inventories by sampling the data at various frequencies and finally average the estimated autocorrelations for all the HFTs in each of the three groups. In Figures 3 and 4, we plot autocorrelation functions constructed in this manner for data sampled on daily and two-hour frequencies respectively.¹⁷ It is evident from Figures 1 and 2 that over the longer run the HFT inventories of all groups are mean-reverting. In particular, Figure 4 shows that any serial dependencies disappear after three lags (i.e., six hours).¹⁸

3.2 HFT short-term inventory dynamics

We estimate a model of second-by-second inventory changes for the HFTs in our sample, to see whether there are broad differences in trading behavior across the three groups over the short term. For this, we estimate an empirical model of change in holdings, similar to the one in Kirilenko et al. (2014), by aggregating the holdings of the HFT firms across all of the four lit venues in our sample (LSE, Chi-X, BATS, and Turquoise). Aggregating in this manner is necessary as HFTs have been shown to control inventory by simultaneously trading in multiple venues (Menkveld, 2011). Thus, for each HFT,

¹⁶We thank an anonymous referee for this suggestion.

¹⁷We can calculate autocorrelations of higher than daily frequency using order book data only. This is because OTC and dark pool trades are not accurately time stamped or not time stamped at all. Daily autocorrelations including OTC and dark trades are almost identical to the order book ones reported here and for this reason are omitted. They are available upon request.

¹⁸Autocorrelation functions calculated in the same manner at different frequencies are consistent with these patterns and are omitted for brevity. Also, partial autocorrelation functions show significance only at the first lag for all intra-day sampling frequencies, suggesting that any short-term conditional dependencies are fully captured by the first inventory lag. Finally, in a series of augmented Dickey-Fuller tests for each HFT-stock, we reject the null hypothesis of a unit root for all groups and all sampling frequencies (ranging from five minutes to two hours) in more than 97% of the cases. These results are also available upon request.

stock and day, we estimate the following model of position changes:

$$\Delta y_t = \phi \Delta y_{t-1} + \delta y_{t-1} + \sum_{k=1}^{10} \beta_k R_{t-k} + u_t, \quad (7)$$

where t denotes a second-long time interval. To ensure statistical significance, we only use for the estimation HFT-stock-days with at least 50 non-zero observations. y_t is the level of the HFT inventory in number of shares, in a given stock and on a given day at time t , assuming the HFT starts the day with a zero inventory¹⁹. R_t is the return of a given stock over the time interval t . We calculate returns using the last quote midpoints of each second, so as to minimize the bid-ask bounce bias. All variables are demeaned and standardized using the same HFT-stock-day means and standard deviations. We do this so as to render the estimated coefficients comparable across HFT-stock-days. The coefficients ϕ and δ can be interpreted as the strength of short- and long-term mean reversion respectively and are intended to capture the dynamics of HFT inventory management. The coefficients on the lagged returns on the other hand are intended to capture the dynamics of HFT strategies.

After this first set of estimations, we test whether the simple and weighted averages (over HFT-stock-days) of the estimated coefficients are zero, separately for the aggressive, neutral, and passive groups. Thus, the tests we do for each group are:

$$\bar{\phi} \equiv \sum_{i,s,d} w^{isd} \hat{\phi}^{isd} = 0, \quad \bar{\delta} \equiv \sum_{i,s,d} w^{isd} \hat{\delta}^{isd} = 0, \quad \bar{\beta}_k \equiv \sum_{i,s,d} w^{isd} \hat{\beta}_k^{isd} = 0, \quad (8)$$

where i denotes HFT firms, s denotes stocks and d denotes days. For these tests, we cluster the test statistics by stocks and HFT firms. The weights are $w^{isd} = 1/(i \times s \times d)$ for the simple averages and $w^{isd} = Val^{isd} / \sum_{i,s,d} Val^{isd}$ for the weighted ones and Val^{isd} is the value (in GBP) traded by firm i , in stock s , on day d .²⁰

¹⁹The results are identical when we use instead positions constructed by carrying over any previous day positions. This is because the key variable is the change in inventory and also because all variables are standardized.

²⁰We also 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.

Table 2 shows the results of these tests of means. The table shows that the aggressive and neutral groups trade in the same direction from one second to the next, whereas the passive group tends to alternate between buys and sells. The table also shows that the HFTs (especially the aggressive and passive ones) are sensitive to their inventory levels ($\bar{\delta} < 0$). In other words, over a longer period, the positions of the HFTs are mean-reverting. This is consistent with the previously reported inventory autocorrelations and also with the SEC (2010) concept release suggesting that HFTs do not take large directional positions and instead aim to end the day with as flat a position as possible. Nevertheless, the three groups of HFTs differ in their sensitivities to inventory levels, with the neutral group being the least sensitive in a statistical sense. This would imply that neutral HFTs inventory levels are more persistent over time and more likely to reach high values intra-day, which is confirmed in the summary statistics in Table 1: the volume-weighted maximum intra-day position of the neutral group is about three times as large (at 27%) as that of the passive and aggressive groups (10%).

As far as HFTs' sensitivity to recent (i.e., 10-second) returns is concerned, there are also stark differences between the groups. Aggressive firms are insensitive to recent price changes, which is consistent with trading strategies that involve reacting quickly to new information. Neutral firms trade in the direction of price changes, buying when prices rise and selling when they drop. This relationship is primarily driven by the most recent price changes and it decays monotonically over the course of a 10-second period. This suggests that neutral firms are trend chasers which could be the result of them following some kind of momentum strategy.

Finally, passive firms trade in the opposite direction of recent price changes with this effect being entirely driven by the most recent (i.e., one-second) price change. This is consistent with market making behavior: as prices rise (or fall), passive firms' ask (bid) quotes are likely to be hit, triggering a sell (or buy) transaction. It is also consistent with passive HFTs' trades being adversely selected due to stale quotes. It is worth mentioning that these patterns are similar to the price/HFT flow correlations reported by Brogaard, Hendershott, and Riordan (2014a) in Figures 1-3, despite the fact that our specification

and the underlying data set are entirely different. This points to consistent patterns of HFT behavior across markets and over time.

Overall, these results highlight the diversity of HFT activity in the U.K. equity market. They point to different inventory control mechanisms and different trading strategies that result in the various groups having different degrees of sensitivity to short-term price changes. They also suggest that the different HFT strategies may have a different impact on price discovery.

4 HFT activity and price discovery

In this section, we assess the informativeness of the trading activity of the various groups of HFTs. That is, we assess and compare the informativeness of the liquidity-taking trades of the groups in our sample, as well as that of the other market participants. Our goal is to see whether and how the information content of trades varies with the type of HFT firm and how this content compares to that of other traders. As mentioned above, HFTs may either make a positive contribution to price discovery [e.g., by reacting faster to news as in Martinez and Rosu (2013)] or a negative one [e.g., by simultaneously reacting to common signals and thus creating price pressure, as in Jarrow and Protter (2012)]. It is therefore plausible that HFTs contribution to price discovery varies with the trading strategies employed.

To examine this, we first assess the information contribution of each group by estimating a standard price impact regression in calendar time. However, calendar time price impact regressions have some limitations. For example, the estimated coefficients could be biased due to endogeneity arising from the use of contemporaneous order flow and returns.

For this reason, we also study HFT information contribution using the Hasbrouck (1991a, 1991b) VAR framework. This well-established methodology has several advantages. First, it operates in tick time and utilizes the exact sequence of trades. This means that it does not suffer from endogeneity issues. Second, it provides a way of

quantifying and comparing the contribution of a given trade to fundamental volatility. Price discovery is essentially price volatility that occurs in response to information about fundamentals. This information is incorporated into prices via trades and/or quote updates. Thus, by quantifying the contribution of trades to the volatility of permanent price changes, the Hasbrouck VAR framework allows us to quantify the contribution of trades to information.

We estimate both our price impact calendar time regressions and the Hasbrouck VAR using transaction reports exclusively for the LSE, the largest lit venue in the U.K. by traded volume. We do this because we want to capture the tick-by-tick relationship between HFT activity and prices on a single order book. As mentioned earlier, the LSE captured between 55% and 70% of the total U.K. lit volume over our sample period.

In doing so, it is worth noting that although we can distinguish the activity of stand-alone HFTs, we cannot distinguish any HFT activity of particular trading desks within larger institutions. This means that when drawing comparisons between the various groups of HFTs and all other traders (labeled “Others”), our results are likely to be biased because the “Others” category may include HFTs not covered by ZEN. To the extent however that the unobserved, desk-level, HFT activity is similar to that of stand-alone firms, this bias will understate the differences between HFTs and non-HFTs. In other words, in the absence of the bias, any differences between HFTs and non-HFTs would likely be even more pronounced. Similarly, we cannot assess the degree to which HFTs (especially the passive ones) contribute to price formation via quotes since we do not observe individual quoting behavior.

4.1 Clock-time price impact regressions

We start our analysis of the information contribution of the various HFT groups by estimating a standard price impact specification in calendar time, similar to the one used in Chordia and Subrahmanyam (2004), albeit at a higher frequency. More specifically,

we estimate the following model for each stock in our sample:

$$R_t = a + b_1 OF_t^{Others} + b_2 OF_{t-1}^{Others} + c_1 OF_t^{Aggressive} + c_2 OF_{t-1}^{Aggressive} \quad (9)$$

$$+ d_1 OF_t^{Neutral} + d_2 OF_{t-1}^{Neutral} + e_1 OF_t^{Passive} + e_2 OF_{t-1}^{Passive} + u_t,$$

where t denotes either 10, or 30-second time intervals and OF_t^i is the order flow of each group of market participants aggregated over time interval t . For instance, the order flow of the aggressive group is defined as: $OF_t^{Aggressive} = \sum_{i \in Aggr.HFT} OF_{i,t}$ where $OF_{i,t} = Vlm_{i,t}^{take}(buy) - Vlm_{i,t}^{take}(sell)$. The other groups are the neutral and passive HFTs, as well as all other market participants (labeled ‘‘Others’’). All variables are normalized by their own standard deviations at the stock level. For the estimation, we drop any stocks with negligible HFT activity. In particular, we drop those stocks for which there are fewer than 10 aggressive trades by each group over the entire sample period. This reduces the number of stocks to 84. After estimating the model, we test if the means of the estimated coefficients are statistically different from zero. We do this for all the stocks in our sample as well as each stock size quartile separately. The average coefficients and the associated t -statistics are shown in Table 3. Panels A and B of Table 3 show the results obtained when aggregating activity over 10-second and 30-second intervals, respectively.²¹

The results in Table 3 show that order flow of the various groups and of all other market participants is strongly contemporaneously correlated with returns. This is to be expected and consistent with the literature. The coefficients on lagged order flow however, show some important differences between the various groups. In particular, aggressive HFT firms appear to be trading in the direction of future price changes to a much larger degree than neutral and passive firms. In fact, the order flow of the passive firms is associated with price reversals, as the negative average point estimates of their lagged order flow suggest. These effects persist regardless of the time frequency used for the aggregation of the variables or the size quartile to which the stocks belong. These results are consistent with the liquidity-taking trades of aggressive firms contributing a larger

²¹We also estimate model (9) using excess returns as a dependent variable, but the results are similar.

amount of information and these firms being likely the ones primarily responsible for the positive impact of HFT activity on price discovery documented in Brogaard, Hendershott, and Riordan (2014a). On the contrary, neutral and passive HFTs' order flow is not in the direction of future price changes and in several instances is in the opposite direction. This trading activity is less consistent with information contribution. For passive HFTs, this may well be the case if these firms engage primarily in market making strategies. In any case, the above results cannot be definitive due to potential simultaneity or reverse causality present in the contemporaneous effects. For this reason, we next estimate a tick-time specification.

4.2 Hasbrouck VAR regressions and HFT information contribution

To measure the information contribution of HFTs' trades, we next 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:

$$\begin{aligned}
r_t &= \sum_{i=1}^{10} \alpha_i r_{t-i} + \sum_{i=0}^{10} \beta_i q_{t-i}^{\text{Psv HFTs}} + \sum_{i=0}^{10} \gamma_i q_{t-i}^{\text{Neu HFTs}} + \sum_{i=0}^{10} \delta_i q_{t-i}^{\text{Aggr HFTs}} + \sum_{i=0}^{10} \zeta_i q_{t-i}^{\text{Others}} + \epsilon_{1t} \\
q_t^{\text{Psv HFTs}} &= \sum_{i=1}^{10} \eta_i r_{t-i} + \sum_{i=1}^{10} \theta_i q_{t-i}^{\text{Psv HFTs}} + \sum_{i=1}^{10} \kappa_i q_{t-i}^{\text{Neu HFTs}} + \sum_{i=1}^{10} \lambda_i q_{t-i}^{\text{Aggr HFTs}} + \sum_{i=1}^{10} \mu_i q_{t-i}^{\text{Others}} + \epsilon_{2t} \\
q_t^{\text{Neu HFTs}} &= \sum_{i=1}^{10} \nu_i r_{t-i} + \sum_{i=1}^{10} \xi_i q_{t-i}^{\text{Psv HFTs}} + \sum_{i=1}^{10} \pi_i q_{t-i}^{\text{Neu HFTs}} + \sum_{i=1}^{10} \rho_i q_{t-i}^{\text{Aggr HFTs}} + \sum_{i=1}^{10} \sigma_i q_{t-i}^{\text{Others}} + \epsilon_{3t} \\
q_t^{\text{Aggr HFTs}} &= \sum_{i=1}^{10} \tau_i r_{t-i} + \sum_{i=1}^{10} \upsilon_i q_{t-i}^{\text{Psv HFTs}} + \sum_{i=1}^{10} \phi_i q_{t-i}^{\text{Neu HFTs}} + \sum_{i=1}^{10} \chi_i q_{t-i}^{\text{Aggr HFTs}} + \sum_{i=1}^{10} \psi_i q_{t-i}^{\text{Others}} + \epsilon_{4t} \\
q_t^{\text{Others}} &= \sum_{i=1}^{10} \omega_{1i} r_{t-i} + \sum_{i=1}^{10} \omega_{2i} q_{t-i}^{\text{Psv HFTs}} + \sum_{i=1}^{10} \omega_{3i} q_{t-i}^{\text{Neu HFTs}} + \sum_{i=1}^{10} \omega_{4i} q_{t-i}^{\text{Aggr HFTs}} + \sum_{i=1}^{10} \omega_{5i} q_{t-i}^{\text{Others}} + \epsilon_{5t},
\end{aligned} \tag{10}$$

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 on 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 mid-quote return update when there is a transaction but not the other way around. That is, there can be a mid-quote return update without a contemporaneous transaction.

Apart from the return equation, the model features a transaction equation for each of the three groups and one for all other traders. In this sense, our approach is similar to that of Barclay et al. (2003) 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. Also, the model does not allow trades of the various groups to contemporaneously influence each other.^{22,23}

This specification explicitly takes into account potential interactions and feedback effects between the activity of the various groups and of the other traders. Although HFTs do not observe each others' (and other market participants') identities, our specification is motivated by the fact that HFTs are known to engage in "order flow" (or "order anticipation") trading strategies whereby they attempt to identify and exploit patterns in the trading volume and order flow of a given stock.

We estimate the VAR model for the entire sample period, appropriately adjusting for end-of-day effects.²⁴ We apply the same filter as in the calendar time price impact regressions and drop those stocks for which there are fewer than 10 aggressive trades by each group over the entire sample period. This reduces the number of stocks to 84. After estimating this model, we invert the lag polynomials and apply a Choleski decomposition

²²A contemporaneous relationship between transactions is meaningful only if one aggregates the trading activity of various participants in fixed calendar time intervals. When operating in tick time, two aggressive trades cannot, by definition, coincide.

²³The 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 $E[\epsilon_{1t}\epsilon_{2t}] = E[\epsilon_{1t}\epsilon_{3t}] = E[\epsilon_{1t}\epsilon_{4t}] = E[\epsilon_{1t}\epsilon_{5t}] = 0$, the error terms of the transaction equations are contemporaneously correlated with one another because, at each tick, only one of the four transaction variables can be non-zero. This means that $E[\epsilon\epsilon'] = \Sigma$, where ϵ is the error term vector and Σ has non-diagonal elements.

²⁴The first 10 observations of each day are used as independent variables for the 11th observation but not as dependent variables themselves.

on the error covariance matrix Σ in order to obtain a vector moving average (VMA) form with orthogonal errors.

We next utilize the variance decomposition method of Hasbrouck (1991b) to assess the information contribution of the trades of the three groups and also to see how this compares with the information contribution of the rest of the traders (labelled “Others”). The Hasbrouck (1991b) framework allows us to do this since 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 assumed to equal:

$$p_t = p_t^* + s_t, \quad (11)$$

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

$$p_t^* = p_{t-1}^* + v_t, \quad v_t \sim iid(0, \sigma_v^2) \quad (12)$$

and the residual noise is assumed to have no long-term impact on prices, i.e., $\lim_{h \rightarrow \infty} E_t[s_{t+h}] = 0$. In other words, assuming that the efficient price follows a random walk, any stationary deviations from a random walk behavior are classified as “noise.” The observed price innovation is then equal to:

$$r_t = p_t - p_{t-1} = \Delta p_t^* + \Delta s_t, \quad (13)$$

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 Equation (10), along with its VMA representation, the efficient price innovation is given by:

$$\Delta p_t^* = \left(\sum_{i=0}^{\infty} a_i \right) e_{1t} + \left(\sum_{i=0}^{\infty} b_i \right) e_{2t} + \left(\sum_{i=0}^{\infty} c_i \right) e_{3t} + \left(\sum_{i=0}^{\infty} d_i \right) e_{4t} + \left(\sum_{i=0}^{\infty} f_i \right) e_{5t}, \quad (14)$$

where e_{1t} , e_{2t} , etc. are the orthogonal errors of the VMA representation and the sums

are over the coefficients of each of the infinite order lag polynomials $a(L)$ to $f(L)$ of the VMA representation of r_t . Since the VMA model errors are spherical with unit variance, the variance of the permanent price impact is:

$$\sigma_v^2 = \left(\sum_{i=0}^{\infty} a_i \right)^2 + \left(\sum_{i=0}^{\infty} b_i \right)^2 + \left(\sum_{i=0}^{\infty} c_i \right)^2 + \left(\sum_{i=0}^{\infty} d_i \right)^2 + \left(\sum_{i=0}^{\infty} f_i \right)^2. \quad (15)$$

The first term in Equation (15) represents the contribution to the efficient price innovation from public information (i.e., quote updates) and the rest of the terms represent the contributions from private information by the passive, neutral, and aggressive HFTs and all other traders, respectively.²⁵

Thus, the relative contributions to information by the various groups of traders are:

$$\frac{\left(\sum_{i=0}^{50} \hat{a}_i \right)^2}{\hat{\sigma}_v^2}, \frac{\left(\sum_{i=0}^{50} \hat{b}_i \right)^2}{\hat{\sigma}_v^2}, \frac{\left(\sum_{i=0}^{50} \hat{c}_i \right)^2}{\hat{\sigma}_v^2}, \frac{\left(\sum_{i=0}^{50} \hat{d}_i \right)^2}{\hat{\sigma}_v^2}, \frac{\left(\sum_{i=0}^{50} \hat{f}_i \right)^2}{\hat{\sigma}_v^2}, \quad (16)$$

where $\hat{\sigma}_v^2$ has also been estimated using 50 lags.

Table 4 shows the estimates of the information contributions for the HFT groups that we have identified, the rest of the traders, and of quote updates. Table 5 shows the differences, along with the associated t -statistics, of the information ratios of the groups and the rest of the traders. In both cases, we report estimates for the entire sample of stocks, as well as for each stock size quartile.

Table 4 shows that the stand-alone HFTs in our sample collectively account for about 6% of the total information contribution in all of the 84 stocks. This is broken down into a 4.1% contribution for aggressive HFTs, a 1.3% contribution for neutral HFTs and a 0.7% contribution for passive ones. The rest of the traders contribute 37% of the information and quote updates contribute the remaining 57.3%. Overall, HFTs contribute 14% of the total information impounded into prices exclusively via aggressive trades while the rest of the traders collectively contribute 86%. Thus, the stand-alone HFTs in our sample make

²⁵The term “private information” is used here to describe any piece of information that is incorporated into prices through trades. In the same fashion, “public information” is any piece of information that is reflected in prices via quote updates. The fact however that quotes are provided by (potentially informed) traders, suggests that the labels “private” and “public” are probably not very accurate. We nevertheless use them as they are somewhat standard in the microstructure literature.

a significant contribution to price discovery. This also suggests that HFTs are able to exploit information signals that are distinct from those that slower market participants trade on. This is consistent with the findings of Brogaard, Hendershott, and Riordan (2014a) and Chaboud et al. (2014), who show that HFTs make a positive contribution to price discovery.

Furthermore, there are substantial differences between the various groups that we examine. HFTs in the aggressive group have a larger total information contribution via aggressive trades than neutral HFTs, who in turn have a larger contribution than passive ones. Of the 14% total HFT trade-induced information contribution, about 10% is accounted for by aggressive HFTs, 3% by neutral ones, and the remaining 1% by passive HFTs. This is presumably driven by the amount of aggressive trading that each group does and reflects the type of strategies that different HFTs employ. Aggressive HFTs are more likely to try to capitalize on the immediacy of market orders by reacting quickly to news and information: thus, their larger information contribution. On the other hand, passive HFTs are more likely to follow market making strategies and be less concerned about trading aggressively and quickly on new information. These shares do not vary significantly across stock size quartiles. Table 5 shows that the differences in information shares among HFTs, other traders, and returns are almost always statistically significant.

In Table 6 we divide the fractions of trade-related information contribution of each group with their respective fractions of total aggressive volume. The general pattern that emerges is that the more aggressive HFTs are, the less information they contribute on a per-trade basis. We think that this is indicative of the finite investment opportunities associated with HFT. In other words, as HFTs attempt to exploit more information signals, either these signals become noisier or they become more difficult to process. This, in turn, could be either because of constraints in HFT computational capacity or because of sheer competition among HFTs. Thus, the informativeness of the average aggressive trade declines. This confirms anecdotal evidence of relatively limited capital being dedicated to HFT activity because the strategies that HFTs employ, although highly profitable, are not scalable. Overall, these results give us additional insights about

the importance of HFT activity for price efficiency and also how this varies in the cross-section of HFTs and with the strategies that they employ.

5 Summary and future research

We utilize transactions data from the U.K. equity market to quantify HFTs' contribution to price discovery and see how this varies in the HFT cross-section. The fact that we observe trader identities allows us to group HFTs, based on their liquidity taking/making behavior, into aggressive, neutral, and passive ones and to examine how price discovery contribution varies across these three groups. Our sample covers 92 FTSE 100 stocks over a four-month period.

We first show that HFTs exhibit variability in their trading strategies by documenting differences in liquidity provision, end-of-day and maximum intra-day positions, trading revenues, etc. The variability in strategies also translates into different sensitivities of HFT position changes to inventory levels and to recent price changes. Whereas all HFTs have mean-reverting inventories over longer horizons, aggressive HFTs tend to trade in a contrarian fashion (i.e., in the opposite direction of recent price changes), neutral HFTs are trend chasers (i.e., they trade in the same direction as recent price changes) and passive HFTs exhibit behavior consistent with market making.

We then examine HFT contribution to price discovery and how this varies across groups. We start by estimating a calendar time price impact specification that shows that the order flow of the aggressive HFTs is in the direction of future price changes. This is not the case for either the neutral or the passive HFTs; in fact, the order flow of the passive group is in the opposite direction of future price changes. This is consistent with the aggressive HFTs being the main contributors to price discovery.

We complement the price impact regressions with a variance decomposition analysis that utilizes the framework of Hasbrouck (1991a, 1991b) and we show that HFTs collectively contribute about 14% of the total trade-related information (or price discovery). Of this, about 10% is accounted for by aggressive HFTs, 3% by neutral ones, and the

remaining 1% by passive HFTs. We interpret this as evidence that the aggressive HFTs are the ones who primarily employ strategies that aim to quickly and accurately respond to and trade on new information as it arrives at the market place. The passive HFTs' limited information contribution is consistent with the notion that these HFTs primarily act as market makers. These results are relatively stable with stock market capitalization.

Finally, we show that when scaled with the fraction of aggressively executed volume, these information shares decline as aggressive volume increases. This suggests that HFTs' trading strategies (at least those that rely on aggressive trading) are not scalable. If anything, it appears that as HFTs expand the set of signals on which they choose to trade upon, the average quality of the signal extraction process declines. This could be either because the signals themselves become noisier and/or because HFTs' signal processing capability is limited.

The literature on HFT is expanding quickly, particularly as new and better data gradually becomes available. A natural extension of our current work would be to examine how HFT-induced price discovery materializes across trading venues. In other words, to see if and to what extent HFT firms respond to signals received from the order book of one exchange, to trade in another.

Appendix

A. Description and implementation of the Lee and Ready (1991) algorithm

To classify trades as buyer or seller-initiated, we apply the Lee and Ready (1991) algorithm in the Bloomberg universe of all trades and quotes. 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. The “tick test” classifies a midpoint trade as buyer-initiated if it is executed at a higher price than the previous trade (an “uptick”) and as seller-initiated if it is executed at a lower price (a “downtick”). If the previous trade is also executed at the same price and on the quote midpoint, the tick test looks at the next most recent trade until it reaches a change in the trade price (“zero uptick” or “zero downtick”).

More precisely, to implement the Lee-Ready algorithm, we go through the following steps:

1. *Top of order book reconstruction:* We reconstruct the top of the LSE order book over the entire sample period. Since the quote data from Bloomberg are reported in chronological order, the bid and ask quotes at a given time are set equal to the most recently reported quotes.
2. *Implementing the Lee-Ready algorithm:* The next step is to implement the Lee-Ready algorithm in the Bloomberg universe by matching transactions with the prevailing top of the book quotes. Thus, each transaction price is matched with the most recent, best bid and ask quotes. The fact that Bloomberg trades and quotes are chronologically ordered allows us to do an exact time matching. We then classify trades as buyer- or seller-initiated by comparing the transaction price with the prevailing quote. If the transaction price happens to equal the midpoint

of the prevailing spread, we then apply the tick test as described above.

3. *Classifying the HFT trades:* To identify the instances where a HFT is the initiator of a trade, we compare the Lee-Ready classification with the buy-sell indicator of the ZEN trade report submitted by the HFT: if the Lee-Ready algorithm classifies a given trade as buyer-initiated and the reporting HFT i identifies herself as the buyer in the matched ZEN report, then this HFT placed the aggressive order. If the HFT identifies herself as a seller, then she is not the party who initiated the trade and therefore she placed the limit order. In short, the identification procedure is as follows:

1. If the Lee-Ready classification matches the buy-sell indicator in ZEN as reported by HFT i , then this is an aggressive trade and HFT i is consuming liquidity.
2. If the Lee-Ready classification does not match the buy-sell indicator in ZEN as reported by HFT i , then this is a passive trade and HFT i is supplying liquidity.

References

- 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>.
- Barclay M., Hendershott T. and McCornick T., 2003, Competition among Trading Venues: Information and Trading on Electronic Communication Networks, *Journal of Finance* 58(6), 2637-2665.
- Baron M., Brogaard J., Hagströmer B. and Kirilenko A., 2016, Risk and Return in High Frequency Trading, working paper.
- Benos E., Brugler J., Hjalmarsson E. and Zikes F., 2015, Interactions Among High Frequency Traders, Bank of England working paper No. 523.
- Biais B., Foucault T. and Moinas S., 2013, Equilibrium Fast Trading, working paper.
- Boehmer E., Li D. and Saar G., 2015, Correlated High-Frequency Trading, working paper.
- Brogaard J., Hendershott T. and Ryordan R., 2014a, High frequency trading and price discovery, *Review of Financial Studies* 27(8), 2267-2306.
- Brogaard J., Hendershott T., Hunt S., Latza T., Pedace L. and Ysusi C., 2014b, High Frerquency Trading and the Execution Costs of Institutional Investors, *Financial Review* 49(2), 345-369.
- Carrion A., 2013, Very fast money: high-frequency trading on the NASDAQ, *Journal of Financial Markets* 16(4), 680-711.
- 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>.
- Chaboud A., Chiquoine B., Hjalmarsson E. and Vega C., 2014, Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market, *Journal of Finance* 69(5), 2045-2084.
- Chordia T. and Subrahmanyam A., 2004, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics* 72(3), 485-518.

- Foucault T., Hombert J. and Rosu I., 2015, News Trading and Speed, *Journal of Finance* 71(1), 335-382.
- Glosten L. and Milgrom P., 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14(1), 71-100.
- 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>.
- Hagströmer B. and Nordén L., 2013, The diversity of high-frequency traders, *Journal of Financial Markets* 16(4), 741-770.
- Hasbrouck J., 1991a, Measuring the Information Content of Stock Trades, *Journal of Finance* 46(1), 179-207.
- Hasbrouck J., 1991b, The Summary Informativeness of Stock Trades: An Econometric Analysis, *Review of Financial Studies* 4(3), 571-595.
- Hendershott T. and Moulton P., 2011, Automation, Speed and Stock Market Quality: The NYSE's Hybrid, *Journal of Financial Markets* 14(4), 568-604.
- Hendershott T. and Riordan R., 2011, Algorithmic Trading and Information, NET Institute Working Paper No 09-08.
- Hirshey N., 2013, Do HFTs anticipate buying and selling pressure?, working paper.
- Jarrow R. and Protter P., 2012, A Dysfunctional Role of High Frequency Trading in Electronic Markets, *International Journal of Theoretical and Applied Finance* 15(3).
- Jones C., 2013, What do we know about high-frequency trading?, Columbia Business School Research Paper No. 13-11.
- Kirilenko A., Kyle A., Samadi M. and Tuzun T., 2014, The Flash Crash: The Impact of High Frequency Trading on an Electronic Market, working paper.
- Lee M. C. and Ready M., 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance* 46(2), 733-746.
- Martinez V. and Rosu I., 2013, High Frequency Traders, News and Volatility, working paper.
- Menkveld A. (2013), High frequency trading and the *New Market* makers, *Journal of Financial Markets* 16(4), 712-740.

Securities and Exchange Commission, 2010, Concept Release on Equity Market Structure.

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>.

Table 1: HFT summary statistics. This table reports weighted and un-weighted summary statistics over HFT-stock-days collectively for all HFTs in our sample, as well as separately for each of the three HFT groups. The weighting is done using each HFT's traded volume. The data include transactions for the 92 FTSE 100 stocks in our sample, executed on four U.K. lit venues (LSE, CHI-X, BATS, and Turquoise) with the exception of the aggressiveness ratios, which have been calculated using LSE-executed trades only. The passive, neutral and aggressive groups consist of all HFTs with $median(AggrRatio) \leq 0.33$, $0.33 < median(AggrRatio) \leq 0.67$, and $0.67 < median(AggrRatio)$ respectively. The aggressiveness ratio is defined in equation (6). *Trading revenue* is defined in equation (5). *Zero crossings* is the number of times that a HFT's inventory becomes zero on a given stock-day. *ID Max Position* is the HFT's maximum absolute inventory level as a fraction of its traded volume for the same day. *EoD Position* is the ratio of the HFT's end-of-day inventory over its traded volume for the same day. We report statistics for end-of-day positions using trades executed on the limit order book only and also statistics using both limit order book and OTC trades. The sample period is from September 1, 2012, to December 31, 2012. HFT-Stock-days with no activity are excluded.

	All HFTs										Passive HFTs				
	Mean	Std. Dev.	Median	95th perc.	5th perc.	Mean	Std. Dev.	Median	95th perc.	5th perc.	Mean	Std. Dev.	Median	95th perc.	5th perc.
<i>Value (GBP)</i>	1,956,040	7,929,132	659,221	7,681,122	905	1,288,426	2,201,031	512,547	5,636,657	83					
<i>Volume</i>	364,375	1,428,026	85,539	1,285,990	2,000	262,672	821,812	63,092	1,032,083	717					
<i>Number of Trades</i>	439	640	224	1,673	5	473	780	245	1,821	2					
<i>Volume Participation (%)</i>	3.1%	2.9%	2.3%	8.6%	0.1%	2.5%	2.4%	2.1%	7.1%	0.0%					
<i>Average Trade Size</i>	712	1,252	365	2,460	101	677	1,348	341	2,404	87					
<i>Average Trade Size (in GBP)</i>	3,757	8,675	3,142	7,739	90	2,662	3,482	2,083	7,395	29					
<i>Trading Revenue (GBP)</i>	762	29,649	20	5,180	-4,280	2,382	55,353	41	26,426	-16,358					
<i>Zero Crossings</i>	15	28	6	60	1	12	20	5	43	1					
<i>Aggr. Ratio</i>	46.0%	40.0%	40.7%	100.0%	0.0%	11.3%	20.4%	2.4%	58.2%	0.0%					
<i>Vlm-w Aggr. Ratio</i>	52.1%	36.9%	-	-	-	18.8%	20.7%	-	-	-					
<i>ID Max Position (% of Volume)</i>	26%	27%	15%	100%	2%	33%	36%	13.5%	100%	3%					
<i>Vlm-w ID Max Position (% of Volume)</i>	13%	15%	-	-	-	10%	14%	-	-	-					
<i>EoD Position (% of Volume)</i>	19%	28%	7%	100%	0%	28%	37%	7%	100%	0%					
<i>Vlm-w EoD Position (% of Volume)</i>	10%	15%	-	-	-	7%	13%	-	-	-					
<i>Vlm-w EoD Position (% of Volume, all trades)</i>	5%	12%	-	-	-	5%	12%	-	-	-					
<i>HFT presence (% of stock-days)</i>	68.9%	-	-	-	-	49.1%	-	-	-	-					

Table 1 continued

	<i>Neutral HFTs</i>					<i>Aggressive HFTs</i>				
	Mean	Std. Dev.	Median	95th perc.	5th perc.	Mean	Std. Dev.	Median	95th perc.	5th perc.
<i>Value (GBP)</i>	1,930,682	12,878,734	503,407	6,938,065	20,409	2,359,987	3,718,006	958,958	9,634,538	94,381
<i>Volume</i>	218,009	824,630	64,710	807,295	2,000	551,980	2,020,152	125,404	2,071,428	8,282
<i>Number of Trades</i>	363	488	180	1,502	9	481	638	255	1,743	34
<i>Volume Participation (%)</i>	2.2%	2.4%	1.6%	7.2%	0.1%	4.2%	3.2%	3.7%	10.1%	0.6%
<i>Average Trade Size</i>	615	1,284	296	2,209	108	823	1,134	481	2,709	129
<i>Average Trade Size (in GBP)</i>	4,189	14,433	3,142	8,214	1,022	4,189	1,814	3,842	7,578	1,849
<i>Trading Revenue (GBP)</i>	188	5,017	9	3,119	-2,705	59	2,532	39	2,580	-2,366
<i>Zero Crossings</i>	19	33	5	101	1	14	31	6	28	1
<i>Aggr. Ratio</i>	43.4%	29.5%	39.7%	100.0%	0.0%	81.9%	24.8%	97.0%	100.0%	29.1%
<i>Vlm-w Aggr. Ratio</i>	58.3%	29.5%	-	-	-	74.4%	25.9%	-	-	-
<i>ID Max Position (% of Volume)</i>	25%	27%	15%	89%	1%	21%	17%	16%	56%	4%
<i>Vlm-w ID Max Position (% of Volume)</i>	27%	23%	-	-	-	10%	8%	-	-	-
<i>EoD Position (% of Volume)</i>	18%	27%	4%	89%	0%	13%	16%	8%	47%	1%
<i>Vlm-w EoD Position (% of Volume)</i>	23%	24%	-	-	-	6%	7%	-	-	-
<i>Vlm-w EoD Position (% of Volume, all trades)</i>	8%	18%	-	-	-	4%	6%	-	-	-
<i>HFT presence (% of stock-days)</i>	58.4%	-	-	-	-	99.1%	-	-	-	-

Table 2: Inventory control regression results for the aggressive, neutral, and passive groups. We first estimate model (7) for each HFT-stock-day with at least 50 non-zero innovations in the dependent variable. We then report simple (Panel A) and weighted (Panel B) averages of the estimated coefficients over HFT-stock-days and for each group, as well as t -statistics of the hypotheses that the means equal zero. The t -statistics are double-clustered by stock and HFT firm. The weighting in Panel B is done using the value traded by each HFT on each stock-day. All the variables in the HFT-stock-day regressions are centered and standardized using the same-day means and standard deviations of the variables. y_t is the time t level of inventory by a HFT on a given stock-day and R_t is the return in second t of a given stock. All coefficients have been multiplied by 10^3 . The data include transactions executed on four U.K. lit venues (LSE, CHI-X, BATS, and Turquoise), over the time period from September 1, 2012, to December 31, 2012.

Dep. Var: Δy_t	Panel A: Equally weighted			Panel B: Equally weighted		
	Aggressive HFTs	Neutral HFTs	Passive HFTs	Aggressive HFTs	Neutral HFTs	Passive HFTs
	Coefficients	t -statistics	Coefficients	t -statistics	Coefficients	t -statistics
Δy_{t-1}	9.29	21.5	11.65	1.7	-0.80	-0.2
y_{t-1}	-7.35	-4.3	-15.07	-1.0	-8.03	-4.8
R_{t-1}	-0.62	-0.2	5.23	7.4	-48.65	-4.7
R_{t-2}	-2.43	-0.8	2.90	3.6	2.46	7.3
R_{t-3}	-2.01	-0.8	2.24	3.0	1.99	7.2
R_{t-4}	-1.46	-0.5	1.59	1.6	1.19	5.2
R_{t-5}	-1.55	-0.6	1.34	1.2	0.75	6.6
R_{t-6}	-1.54	-0.7	1.03	1.1	1.06	9.8
R_{t-7}	-1.45	-0.7	0.98	1.1	0.33	4.9
R_{t-8}	-1.47	-0.8	0.91	0.9	0.64	6.1
R_{t-9}	-1.30	-0.8	0.69	0.8	0.21	1.9
R_{t-10}	-1.19	-0.8	0.66	0.7	0.37	3.7
$\sum_{k=1}^{10} \beta_k$	-15.02	-0.7	17.56	2.0	-39.66	-4.2

Table 2 continued

Dep. Var: Δy_t	Aggressive HFTs			Panel B: Value weighted Neutral HFTs			Passive HFTs		
	Coefficients	<i>t</i> -statistics	<i>t</i> -statistics	Coefficients	<i>t</i> -statistics	<i>t</i> -statistics	Coefficients	<i>t</i> -statistics	<i>t</i> -statistics
Δy_{t-1}	13.72	4.5	23.53	23.53	2.5	2.5	-8.17	-3.5	-3.5
y_{t-1}	-8.65	-2.5	-16.42	-16.42	-0.7	-0.7	-3.90	-10.0	-10.0
R_{t-1}	6.14	2.7	4.91	4.91	5.4	5.4	-58.75	-5.1	-5.1
R_{t-2}	1.37	3.2	3.78	3.78	62.1	62.1	4.65	6.6	6.6
R_{t-3}	0.43	0.6	2.51	2.51	3.6	3.6	3.40	6.4	6.4
R_{t-4}	0.49	0.3	3.04	3.04	1.9	1.9	2.91	7.2	7.2
R_{t-5}	-0.16	-0.1	1.35	1.35	1.3	1.3	2.14	6.3	6.3
R_{t-6}	-0.32	-0.3	1.09	1.09	1.2	1.2	2.21	8.0	8.0
R_{t-7}	-0.47	-0.4	0.56	0.56	0.5	0.5	1.39	8.0	8.0
R_{t-8}	-0.76	-0.8	0.29	0.29	0.2	0.2	1.54	8.2	8.2
R_{t-9}	-0.61	-0.9	0.32	0.32	0.3	0.3	0.83	5.9	5.9
R_{t-10}	-0.52	-0.8	1.20	1.20	3.1	3.1	0.50	9.4	9.4
$\sum_{k=1}^{10} \beta_k$	5.60	0.8	19.06	19.06	2.9	2.9	-39.17	-4.6	-4.6

Table 3: Clock-time price impact regression results. For each stock in our sample, we estimate model (9) and then test whether average coefficients (across stocks) are equal to zero. The model is estimated by aggregating activity over 10-second (Panel A) and 30-second intervals (Panel B). We report averages over all stocks, as well as over each stock size quartile. For the estimation we include only those stocks with a minimum of 10 aggressive trades by each group over the entire sample period. The resulting data include all transactions from 84 FTSE 100 stocks, executed on the LSE, over the time period from September 1, 2012, to December 31, 2012.

Panel A: 10-second interval						
	<i>All Stocks</i>		<i>1st Qrt</i>		<i>2nd Qrt</i>	
	<i>Coefficients</i>	<i>t-statistics</i>	<i>Coefficients</i>	<i>t-statistics</i>	<i>Coefficients</i>	<i>t-statistics</i>
OF_t^{Others}	0.235	26.48	0.234	12.56	0.255	14.71
OF_{t-1}^{Others}	0.013	10.41	0.010	3.72	0.013	4.81
$OF_t^{Aggressive}$	0.072	17.60	0.069	7.84	0.071	8.05
$OF_{t-1}^{Aggressive}$	0.003	5.55	0.003	3.01	0.004	3.65
$OF_t^{Neutral}$	0.046	9.99	0.066	6.26	0.043	4.71
$OF_{t-1}^{Neutral}$	0.000	0.65	0.000	-0.13	0.000	-0.36
$OF_t^{Passive}$	0.060	4.45	0.074	3.09	0.075	2.47
$OF_{t-1}^{Passive}$	-0.001	-1.07	-0.001	-0.34	-0.001	-0.28
Avg. R^2	0.10		0.10		0.12	
	<i>3rd Qrt</i>		<i>4th Qrt</i>			
	<i>Coefficients</i>	<i>t-statistics</i>	<i>Coefficients</i>	<i>t-statistics</i>		
OF_t^{Others}	0.250	13.48	0.199	13.66		
OF_{t-1}^{Others}	0.010	6.32	0.018	7.75		
$OF_t^{Aggressive}$	0.064	10.70	0.085	9.88		
$OF_{t-1}^{Aggressive}$	0.002	1.91	0.005	2.85		
$OF_t^{Neutral}$	0.042	5.48	0.032	4.06		
$OF_{t-1}^{Neutral}$	0.001	0.64	0.001	1.36		
$OF_t^{Passive}$	0.070	1.94	0.018	3.47		
$OF_{t-1}^{Passive}$	-0.005	-1.40	0.001	1.43		
Average R^2	0.12		0.07			

Table 3 continued

Panel B: 30-second intervals						
	<i>All Stocks</i>		<i>1st Qrt</i>		<i>2nd Qrt</i>	
	<i>Coefficients</i>	<i>t-statistics</i>	<i>Coefficients</i>	<i>t-statistics</i>	<i>Coefficients</i>	<i>t-statistics</i>
OF_t^{Others}	0.269	22.70	0.249	11.19	0.289	12.22
OF_{t-1}^{Others}	0.009	9.57	0.002	3.37	0.009	5.72
$OF_t^{Aggressive}$	0.076	13.83	0.063	6.28	0.077	6.85
$OF_{t-1}^{Aggressive}$	0.001	3.97	-0.001	-1.81	0.003	3.27
$OF_t^{Neutral}$	0.052	8.93	0.070	5.67	0.050	4.33
$OF_{t-1}^{Neutral}$	-0.002	-0.46	-0.004	-0.10	-0.002	-0.30
$OF_t^{Passive}$	0.065	4.53	0.082	2.97	0.085	2.45
$OF_{t-1}^{Passive}$	-0.002	-1.20	-0.003	-0.22	-0.004	-0.25
Average R^2	0.13		0.11		0.15	
	<i>3rd Qrt</i>		<i>4th Qrt</i>			
	<i>Coefficients</i>	<i>t-statistics</i>	<i>Coefficients</i>	<i>t-statistics</i>		
OF_t^{Others}	0.291	11.64	0.244	10.75		
OF_{t-1}^{Others}	0.006	4.75	0.017	8.50		
$OF_t^{Aggressive}$	0.070	7.85	0.095	7.41		
$OF_{t-1}^{Aggressive}$	0.002	1.58	0.002	1.94		
$OF_t^{Neutral}$	0.048	4.30	0.038	3.53		
$OF_{t-1}^{Neutral}$	-0.001	-0.50	-0.002	-0.61		
$OF_t^{Passive}$	0.068	2.11	0.023	2.90		
$OF_{t-1}^{Passive}$	-0.002	-4.27	0.001	0.99		
Average R^2	0.15		0.10			

Table 4: Hasbrouck information shares. We estimate the VAR model (10) for each stock and for the entire sample period and convert the model into a vector moving average form. We then report the average - across stocks - efficient price variance contribution ratios in (16). The averages are calculated over all stocks as well as over each stock size quartile. The null hypothesis being tested for the Information contributions is that they are equal to zero. t -statistics are in parentheses. The data include all transactions from 84 FTSE 100 stocks, executed on the LSE, over the time period from September 1, 2012, to December 31, 2012.

<i>Returns</i>	<i>Others</i>	<i>Passive</i>	<i>Neutral</i>	<i>Aggressive</i>	<i>Returns</i>	<i>Others</i>	<i>Passive</i>	<i>Neutral</i>	<i>Aggressive</i>
	<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>		<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>
<i>All Stocks</i>									
57.3%	36.6%	0.7%	1.3%	4.1%	62.6%	30.7%	0.5%	2.0%	4.2%
(44.37)	(30.19)	(3.98)	(8.73)	(21.51)	(29.08)	(18.16)	(4.03)	(4.81)	(11.01)
<i>1st Qtr of Stocks</i>									
<i>2nd Qtr of Stocks</i>									
53.9%	39.1%	1.2%	1.6%	4.3%	56.8%	37.6%	1.0%	0.9%	3.7%
(20.5)	(15.85)	(2.46)	(5.12)	(12.82)	(17.19)	(11.51)	(1.92)	(5.34)	(10.65)
<i>3rd Qtr of Stocks</i>									
<i>4th Qtr of Stocks</i>									
55.9%	39.1%	0.2%	0.8%	4.0%					
(38.32)	(31.88)	(3.91)	(3.97)	(8.6)					

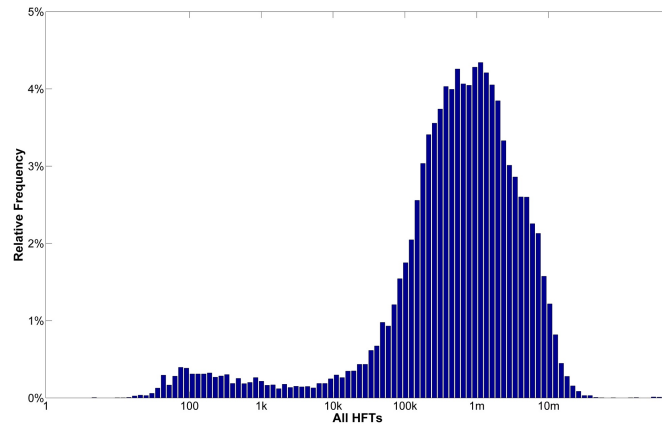
Table 5: Differences in Information ratios among returns, HFT groups, and all other traders. We estimate the VAR model (10) for each stock and for the entire sample period and convert the model into a vector moving average form. We then report the average -across stocks - differences of the Information ratios. The averages are calculated over all stocks as well as over each stock size quartile. The null hypotheses being tested are that the Information ratios are equal across groups. t -statistics are in parentheses. The data include all transactions on the 84 FTSE 100 stocks of our sample executed on the LSE over the time period from September 1, 2012, to December 31, 2012.

<i>Others - Passive HFTs</i>	<i>Others - Neutral HFTs</i>	<i>Others - Aggressive HFTs</i>	<i>Returns- Passive HFTs</i>	<i>Returns- Neutral HFTs</i>	<i>Returns- Aggressive HFTs</i>	<i>Aggressive HFTs - Passive HFTs</i>	<i>Aggressive HFTs - Neutral HFTs</i>	<i>Neutral HFTs - Passive HFTs</i>
<i>All stocks</i>								
35.8% (27.57)	35.2% (29.54)	32.5% (28.75)	56.5% (45.23)	55.9% (41.45)	53.2% (37.61)	3.3% (11.36)	2.7% (13.25)	0.6% (2.49)
<i>1st Qtr Stocks</i>								
30.2% (17.48)	28.7% (18.47)	26.5% (17.73)	62.1% (29.04)	60.6% (24.88)	58.3% (23.81)	3.7% (9.19)	2.2% (5.62)	1.5% (3.66)
<i>2nd Qtr Stocks</i>								
37.9% (13.83)	37.5% (16.13)	34.7% (15.23)	52.7% (21.21)	52.3% (18.57)	49.5% (17.32)	3.2% (4.75)	2.8% (6.99)	0.4% (0.65)
<i>3rd Qtr Stocks</i>								
36.6% (10.28)	36.6% (11.32)	33.8% (11.11)	55.8% (17.8)	55.8% (16.64)	53.0% (14.91)	2.8% (3.74)	2.8% (8.01)	-0.1% (-0.11)
<i>4th Qtr Stocks</i>								
38.8% (31.41)	38.3% (29.92)	35.1% (31.19)	55.6% (38.05)	55.0% (37.36)	51.9% (28.86)	3.8% (8.16)	3.2% (6.15)	0.6% (3.15)

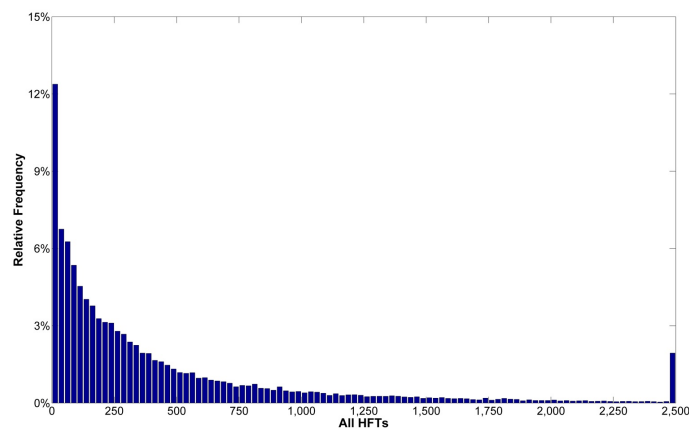
Table 6: Fractions of trade-related information contribution, fractions of aggressive volume, and the ratios of the two for each group and all other traders for all stocks and each stock size quartile. The data include all transactions on the 84 FTSE 100 stocks of our sample executed on the LSE over the time period from September 1, 2012, to December 31, 2012.

	<i>Others</i>	<i>Passive</i>	<i>Neutral</i>	<i>Aggressive</i>	<i>Others</i>	<i>Passive</i>	<i>Neutral</i>	<i>Aggressive</i>
	<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>	<i>HFTs</i>
	<i>All stocks</i>							
<i>Information</i>	85.6%	1.7%	3.1%	9.6%	82.1%	1.3%	5.3%	11.3%
<i>Aggr. Vlm</i>	86%	0.3%	1.7%	11.8%	85.3%	0.4%	2.0%	12.4%
<i>Ratio</i>	0.99	4.96	1.80	0.81	0.96	3.59	2.68	0.92
	<i>2nd Qtr of Stocks</i>							
<i>Information</i>	84.7%	2.6%	1.9%	3.8%	86.9%	2.3%	1.2%	3.4%
<i>Aggr. Vlm</i>	87.6%	0.4%	1.6%	10.4%	88.0%	0.3%	0.9%	10.8%
<i>Ratio</i>	0.97	6.01	1.22	0.37	0.99	8.59	1.27	0.31
	<i>4th Qtr of Stocks</i>							
<i>Information</i>	88.5%	0.6%	1.9%	9.1%				
<i>Aggr. Vlm</i>	88.3%	0.1%	0.8%	10.8%				
<i>Ratio</i>	1.00	4.56	2.44	0.84				

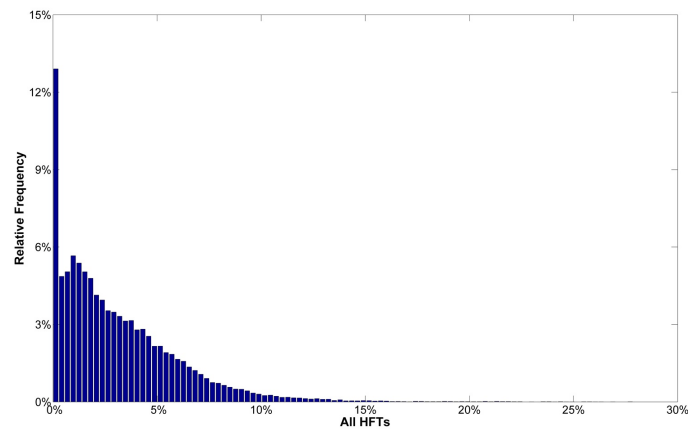
Figure 1: Frequency distributions of HFT activity variables. We calculate the HFT activity variables for each HFT-stock-day and report the resulting frequency distributions. The data include transactions for the 92 FTSE 100 stocks in our sample, executed on four U.K. lit venues (LSE, CHI-X, BATS, and Turquoise) over the time period from September 1, 2012, to December 31, 2012. HFT-stock-days with no activity are excluded.



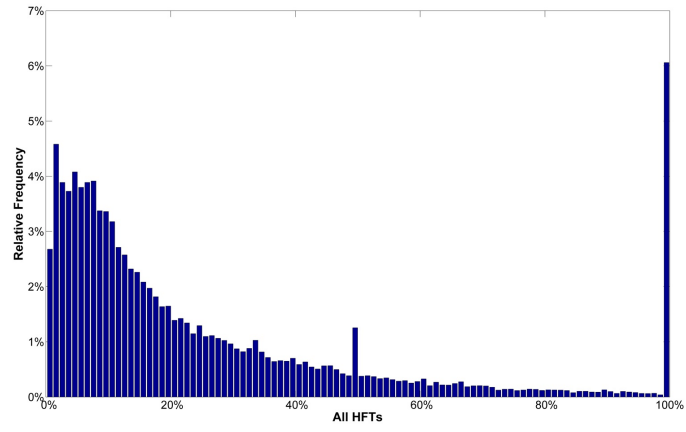
(a) *Value Traded (GBP)- log scale*



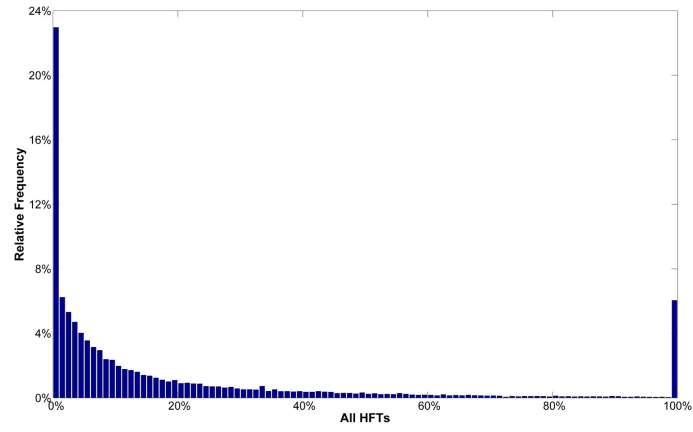
(b) *Number of Trades*



(c) *Volume Participation (%)*



(d) *Maximum Intra-day Position (% of Volume)*



(e) *End-of-day Position (% of Volume)*

Figure 2: Average aggressiveness ratios of the various groups over time. The aggressiveness ratios are calculated for each HFT-stock-day as defined in equation (6) and are based on LSE trades and quotes. The passive, neutral and aggressive groups consist of all HFTs with $median(AggrRatio) \leq 0.33$, $0.33 < median(AggrRatio) \leq 0.66$ and $0.66 < median(AggrRatio)$ respectively. The aggressiveness ratios are then averaged over the HFT firms of each group. The data include all transactions on the 92 FTSE 100 stocks of our sample executed on the LSE over the time period from September 1, 2012, to December 31, 2012. HFT-stock-days with no activity are excluded

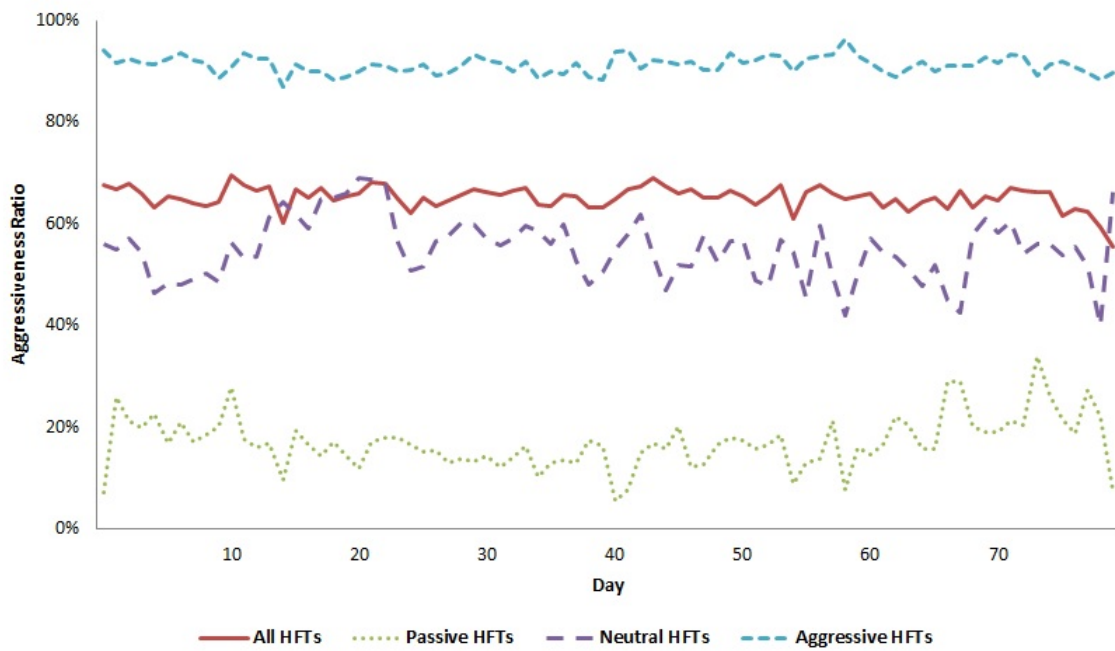
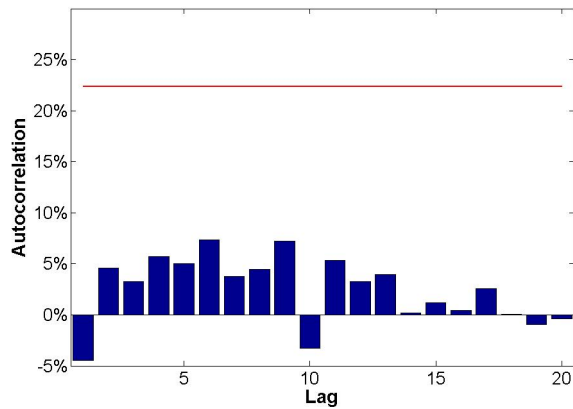
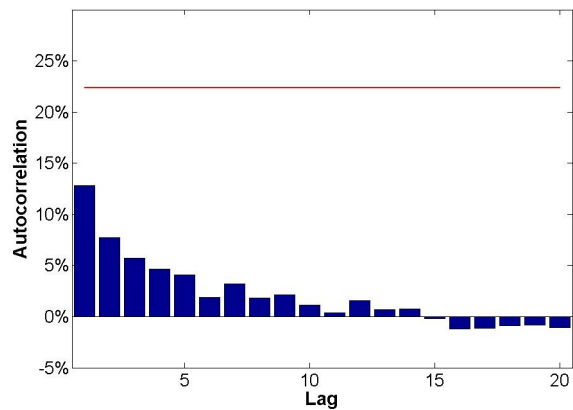


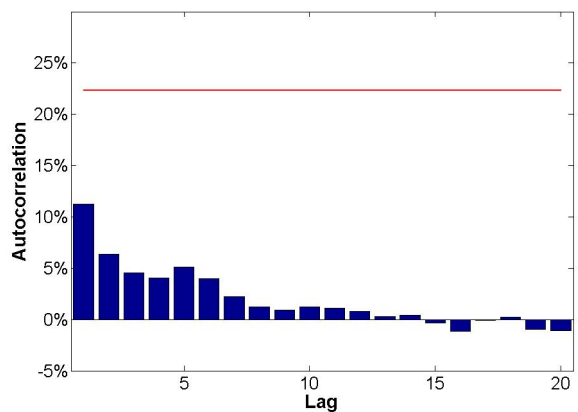
Figure 3: Autocorrelations of HFT inventory positions, sampled every day. For each HFT-stock, we calculate the inventory autocorrelations up to 20 lags and then average these autocorrelations over each of the three groups. The inventories are calculated by carrying over the previous day outstanding positions. The horizontal line denotes significance at 5%. The passive, neutral and aggressive groups consist of all HFTs with $median(AggrRatio) \leq 0.33$, $0.33 < median(AggrRatio) \leq 0.66$ and $0.66 < median(AggrRatio)$ respectively. The aggressiveness ratios are then averaged over the HFT firms of each group. The data include transactions for the 92 FTSE 100 stocks in our sample, executed on four U.K. lit venues (LSE, CHI-X, BATS, and Turquoise) over the time period from September 1, 2012, to December 31, 2012.



(a) *Aggressive HFTs*

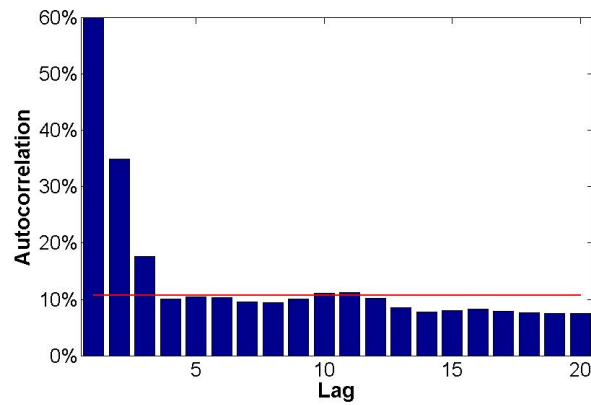


(b) *Neutral HFTs*

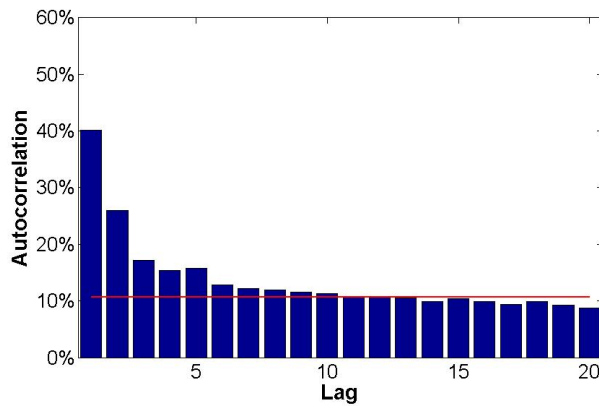


(c) *Passive HFTs*

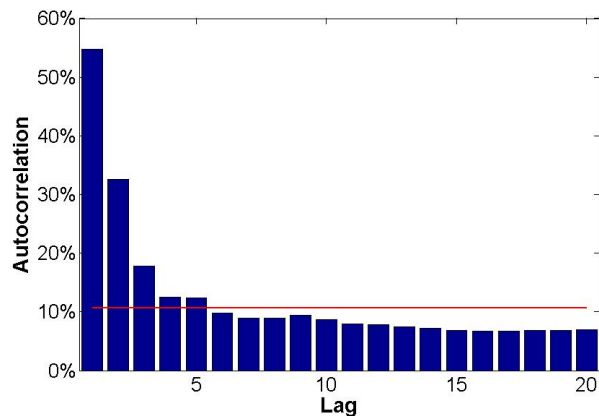
Figure 4: Autocorrelations of HFT inventory positions, sampled every two hours. For each HFT-stock, we calculate the inventory autocorrelations up to 20 lags and then average these autocorrelations over each of the three groups. The inventories are calculated by carrying over the previous day outstanding positions. The horizontal line denotes significance at 5%. The passive, neutral and aggressive groups consist of all HFTs with $median(AggrRatio) \leq 0.33$, $0.33 < median(AggrRatio) \leq 0.66$ and $0.66 < median(AggrRatio)$ respectively. The aggressiveness ratios are then averaged over the HFT firms of each group. The data include transactions for the 92 FTSE 100 stocks in our sample, executed on four U.K. lit venues (LSE, CHI-X, BATS, and Turquoise) over the time period from September 1, 2012, to December 31, 2012.



(a) *Aggressive HFTs*



(b) *Neutral HFTs*



(c) *Passive HFTs*