BANK OF ENGLAND

Working Paper No. 523

Interactions among high-frequency traders

Evangelos Benos, James Brugler, Erik Hjalmarsson and Filip Zikes

February 2015

BANK OF ENGLAND

Working Paper No. 523

Interactions among high-frequency traders

Evangelos Benos,[1] James Brugler,[2] Erik Hjalmarsson[3] and Filip Zikes[4]

Abstract

Using unique transactions data for individual high-frequency trading (HFT) firms in the UK equity market, we examine if the trading activity of individual HFT firms is contemporaneously and dynamically correlated with each other, and what impact this has on price efficiency. We find that HFT order flow exhibits significantly higher commonality than the order flow of a control group of investment banks, both within and across stocks. However, intraday HFT order flow commonality is associated with a permanent price impact, suggesting that commonality in HFT activity is information-based and so does not generally contribute to undue price pressure and price dislocations.

Key words: High-frequency trading, correlated trading strategies, price discovery.

JEL classification: G10, G12, G14.

(1) Bank of England. Email: evangelos.benos@bankofengland.co.uk
(2) University of Cambridge, Faculty of Economics. Email: jab244@cam.ac.uk
(3) University of Gothenburg, Department of Economics. Email: erik.hjalmarsson@economics.gu.se
(4) Bank of England. Email: filip.zikes@bankofengland.co.uk.

The views expressed in this paper are those of the authors and not necessarily those of the Bank of England or any of its committees. We are grateful to Satchit Sagade for his help in cleaning and processing the data. We would also like to thank Bjőrn Hagstrőmer, Edwin Schooling Latter, Nick Vause, Matthew Willison, Graham Young, an anonymous referee, seminar participants at the UK Financial Conduct Authority and participants of the 2015 Copenhagen Business School workshop on high-frequency trading for helpful comments and suggestions. This paper was finalised on 26 January 2015.

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/research/Pages/workingpapers/default.aspx

Publications Team, 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

High-frequency trading (HFT), where automated computer traders interact at lightning-fast speed with electronic trading platforms, has become an important feature of many modern financial markets. The rapid growth, and increased prominence, of these ultrafast traders have given rise to concerns regarding their impact on market quality and market stability. These concerns have been fuelled by instances of severe and short-lived market crashes such as the 6 May 2010 'Flash Crash' in the US markets. One concern about HFT is that owing to the high rate at which HFT firms submit orders and execute trades, the algorithms they use could interact with each other in unpredictable ways and, in particular, in ways that could momentarily cause price pressure and price dislocations in financial markets.

Using unique transactional data that allows us to identify the activity of HFT firms present in the UK equity market, we examine if their activity is indeed correlated and what this means for market quality. We focus our analysis on the ten largest HFT firms, which account for the bulk of the stand-alone HFT firm activity in our sample. In doing so we compare their activity with that of the ten largest investment banks present in our sample.

We estimate a dynamic regression model of order flow, by HFT firms and investment banks, in individual stocks as well as across different stocks. Order flow is defined as the net aggressive buying volume over a given time interval. In other words, it is the difference between the number of shares bought and sold via orders that are executed immediately at the best available price. The estimation is done using data sampled at a ten-second frequency in order to capture any short-lived interactions across HFT firms.

We find that HFT order flow is more correlated over time than that of the investment banks, both within and across stocks. This means that HFT firms tend more than their peer investment banks to buy or sell aggressively the same stock at the same time. Also, a typical HFT firm tends to simultaneously aggressively buy and sell multiple stocks at the same time to a larger extent than a typical investment bank.

What does that mean for market quality? A key element of a well-functioning market is price efficiency; this characterises the extent to which asset prices reflect fundamental values. Dislocations of market prices are clear violations of price efficiency as they happen in the absence of any news about fundamental values.

To assess the impact of correlated trading by HFT firms on price efficiency, we first construct a metric that captures the extent of correlated trading within a day by HFT firms and investment banks. We then run regressions of stock returns on contemporaneous and lagged order flow by HFT firms and investment banks. If order flow has a longer-lasting (i.e. 'permanent') price impact, then this is indicative of informed trading; for if the trade had no information content, its price impact would be temporary as the induced price change would not be justified by any changes in fundamentals and market participants would force the price back to its original value. The key question is then if our metric of correlated trading is associated with a permanent or temporary price impact.

We find that instances of correlated trading by HFT firms are associated with a permanent price impact whereas correlated trading by investment banks is associated with only a temporary price impact. We interpret this as evidence that HFT correlated trading is information-based; in other words, HFT firms appear to be reacting simultaneously and quickly to new information as it arrives at the market place, which makes prices more efficient. This suggests that correlated trading by HFT firms does not appear to contribute to undue price pressure and price dislocations on a systematic basis in the UK equity market. Of course, this does not mean that HFT activity may never cause or exacerbate any price dislocations either in the equity or other markets. To assess that, additional research with more data, covering periods of market stress, would be necessary.

1 Introduction

High-frequency trading, where automated computer traders interact at lightning-fast speed with electronic trading platforms, has become an important feature of many modern markets. The rapid growth, and increased prominence, of these ultra fast traders have given rise to concerns regarding their impact on market quality and market stability. Over the past few years, numerous empirical studies have analyzed the market impact of high-frequency trading (HFT), as well as algorithmic trading (AT) more generally.[1,2] With some recent exceptions, most of these studies have analyzed aggregate measures of HFT and AT in various markets.[3] In the current paper, we aim to shed light on the ways in which individual HFTs interact with each other and assess the impact of this interaction on price efficiency.

Automated high-frequency trading is made possible by the direct interaction between electronic trading platforms and pre-programmed computers. Although this lends HFTs a huge speed advantage over "human" traders—computers are simply much faster at receiving, processing and reacting to new information—the pre-programmed systematic nature of high-frequency trading might also limit the diversity of the strategies that HFTs implement. This notion is given empirical support by Chaboud, Chiquoine, Hjalmarsson, and Vega (2014), who document evidence consistent with computer-based strategies being more correlated than those of human traders in the foreign exchange market. Possible correlation of HFT strategies is often viewed as a source of concern, as it could potentially have destabilizing effects on the market (Haldane, 2011, and White, 2014). The "Quant Meltdown" in August 2007, when many long-short equity funds pursuing similar strategies suffered major losses and quickly unwound their strategies amid great

[1] Algorithmic trading refers to any automated trading where computers directly interact with electronic trading platforms; high-frequency trading is therefore a subset of algorithmic trading. Given the focus of the current paper, in the subsequent discussion we mostly refer to high-frequency trading, although many of the arguments apply to both AT and HFT.

[2] HFT will be used to denote both high-frequency *trader* and high-frequency *trading*; AT will be used in an analogous manner. In our data, we can identify the trading activity of individual high-frequency trading (HFT) firms. We will therefore refer to both HFTs and HFT firms, where the latter formulation is used to emphasize this unit of observation.

[3] See, for instance, Hendershott, Jones, and Menkveld (2011), Hendershott and Riordan (2013), Brogaard, Hendershott, and Riordan (2014), and Chaboud, Chiquoine, Hjalmarsson, and Vega (2014). Benos and Sagade (2015) analyze the activity of various subgroups of HFTs distinguished according to their liquidity making/taking behaviour.

market turmoil, is a prime example of the possible negative impact of highly correlated strategies among a large segment of market participants; Khandani and Lo (2011) provide an in-depth analysis of these events.[4]

The implications of correlation among HFTs' trading strategies is not unambiguous, however, and depends on the underlying reasons behind it. If the strategy correlation is a result of many HFTs focusing on the same arbitrage opportunities, this may help improve price efficiency as implied by the models of Kondor (2009) and Oehmke (2009) in the context of "convergence trades". This positive effect from competition is not a foregone conclusion, however. Stein (2009) and Kozhan and Wah Tham (2012) both argue that increased competition for arbitrage opportunities could cause a crowding effect, which might result in prices being pushed away from fundamentals.

Alternatively, HFT activity could be correlated because HFTs trade on common signals. Again, the effect on prices is ambiguous. In the model of Martinez and Rosu (2013), correlated trading by HFTs makes prices more efficient, whereas in that of Jarrow and Protter (2012), HFTs' simultaneity in trading causes prices to "overshoot", creating excess volatility. Additionally, HFTs might also create deviations in prices from fundamentals if they follow simple trading rules like the positive-feedback traders in DeLong, Shleifer, Summers, and Waldmann (1990), or the chartists in Froot, Scharfstein, and Stein (1992).

In the current paper, we use data from the UK equity market to analyze cross-sectional aspects of high-frequency trading and assess their impact on market quality. In particular, we analyze both the interactions between different HFT firms in a given stock, as well as the trading patterns of a given HFT firm trading across several stocks. That is, we analyze both the *within-stock/across-firm* trading correlation, and the *within-firm/across-stock* trading correlation. The purpose of these analyses is to form a better understanding of: (i) the extent to which a given HFT firm tends to trade in a similar manner and direction to its high-frequency competitors, and (ii) the extent to which a given HFT firm tends to follow similar, or correlated, strategies over many stocks.

Both of these analyses speak towards the greater question of whether HFTs might be

[4]The quantitative strategies referred to in this episode where not necessarily implemented through computer-based trading.

a source of concern from the perspective of market stability. A greater correlation across HFT firms suggests that HFTs act more as a uniform group with a greater potential for (possibly adverse) market impacts, as discussed above. A greater correlation in the trading patterns across multiple stocks for a given HFT might suggest a more market-wide impact of the trades of a given HFT.[5]

We use data on transactions for the 20 largest stocks (by market capitalization) of the FTSE 100 index, executed on the electronic limit order book of the London Stock Exchange (LSE). These data are accessed through the ZEN database, maintained by the UK Financial Conduct Authority (FCA), and the sample spans four months, from September 1st through December 31st, 2012. The data explicitly identifies the submitter of each trade report along with other detailed information such as volume, execution price and time stamp. We focus our analysis on trading in 10 individual HFT firms, which together represent more than 98 percent of the total HFT volume in our sample. By focusing on a limited number of large firms, which are behind the vast majority of high-frequency trading, we are able to conduct a detailed analysis of the interactions between HFT firms. In addition, we also use trade data for the 10 largest investment banks (IBs) active in our sample. These serve as a reference group against which to compare the results for the pure high-frequency firms. IBs clearly engage in a wide variety of trading activities, including trades by their own proprietary trading desks as well as customer order-driven trades. Whereas the proprietary trading activities might involve high-frequency strategies, the overall activities of investment banks are quite distinct from that of pure HFT firms. We therefore view IBs as a relevant comparison group, proxying for the behaviour of informed traders in the market.

To analyze correlations, and possible causations, between the activities of individual HFTs in a given stock, we use a high-frequency vector autoregression (VAR). In particular, for each of the 20 stocks in our sample, we formulate a VAR with trading activity in all 10 HFTs and all 10 IBs as dependent variables. The VAR also controls for time trends, the bid-ask spread, stock return volatility, trading volume, and past

[5]That is not to say that individual HFTs trading in a given asset cannot under certain circumstances adversely affect the entire market, as may have happened during the May 6, 2010 "flash crash"(e.g., Menkveld, 2013, and Kirilenko, Kyle, Samadi, and Tuzun, 2014). However, the key contract traded during the flash crash was the E-mini S&P 500 futures contract. It seems less likely that trading in an individual stock, as we study here, would cause similar disturbance.

returns. Trading activity is measured either as (i) order flow (buyer-initiated volume minus seller-initiated volume), or (ii) total transacted volume. The VAR is estimated by pooling data from all stocks, yielding a set of interpretable results.

Within this VAR model, we evaluate whether trading by one HFT firm is affected by the trading of the other HFT firms, through Granger causality tests. In particular, if the trading activity by a given firm is *positively* Granger caused by the activity of the other HFT firms, we view this as consistent with correlation in the strategies of HFT firms. Granger causality is essentially a predictive property, and this concept of strategy correlation is somewhat different from the usual static correlation that one might think of. However, thinking of trade correlations in this dynamic setting has several advantages. First, Granger causality, especially in high-frequency settings, may be indicative of actual (contemporaneous) causality, which is arguably more interesting to analyze than non-causal correlations. Second, the predictive nature of the tests might actually be at least as relevant as any contemporaneous correlations or causations. That is, from a market stability perspective it seems of key interest to understand whether trades in a given stock by a given firm eventually triggers more trades by other firms in that stock.

The main empirical results show that there is indeed a positive order flow correlation among HFTs. In particular, HFTs' order flows are positively temporally dependent across different HFTs, whereas those of IBs are negatively dependent. In other words, aggressive buying by an HFT Granger-causes additional aggressive buying by other HFTs, whereas aggressive *buying* by an IB Granger-causes aggressive *selling* by the other IBs, and vice versa for aggressive selling.

Similarly, we use a VAR specification to test whether a given HFT firm's trading strategies tend to be similar across different stocks. In particular, we estimate a VAR system with each equation capturing the trade activity of the HFT firm in a given stock. The data are now pooled for all HFTs, and a pooled panel VAR is again estimated. An analogous panel VAR is also estimated for IBs. The same controls as before are included in the regressions. Similarly to the case of within-stock/across-firm trading, we find that individual HFTs' strategies tend to be considerably more correlated across stocks than those of the IBs, both in terms of order flow and volume traded. That is, a trade by a given HFT is likely to generate additional trades by the same HFT, and in the same direction, across a number of stocks, to a significantly larger extent than a trade by an IB.

One interpretation of these results is that HFT algorithms may have a degree of commonality embedded in their design, which could potentially give rise to price pressure and excess volatility as in Jarrow and Protter (2012). An alternative interpretation is that HFTs use strategies that are uniformly more efficient in receiving, processing, and trading on information as it arrives at the marketplace, as in Martinez and Rosu (2013). In this case, the observed commonality is the result of HFT firms trading on common sources of information.

To test these two hypotheses, we construct a high-frequency metric of HFT and IB order flow correlation and use it as an explanatory variable in a price impact regression. The key finding is that HFT correlation is associated with a permanent price impact, whereas IB correlation is associated with price reversals, over a 5-minute period. This is consistent with HFT commonality being the result of informed trading and thus contributing to price discovery.

Our study adds to the growing empirical literature on high-frequency trading specifically, and algorithmic trading generally. In relation to previous work, we contribute to the understanding of the correlation of HFT strategies across different firms, and also the extent to which individual HFT firms tend to follow similar strategies across many different stocks. Most previous studies have been restricted to using aggregate measures of HFT or AT participation and have primarily focused on the speed aspects of computer-based trading, and less on the "cross-sectional" aspects.[6]

The rest of the paper is organized as follows. Section 2 describes the data that we use and presents some summary statistics and preliminary motivating analysis. Section 3 introduces the main empirical framework and presents the results on interactions across HFT firms. Section 4 studies whether these correlation patterns appear to have any impact on market quality and Section 5 concludes. Some supplemental results are

[6]Benos and Sagade (2015), Hagströmer and Nordén (2013), and Hagströmer, Nordén, and Zhang (2014) also make explicit use of the ability to follow individual HFT firms. Their focus is, however, quite different from ours, and mostly on classifying and distinguishing HFTs along market-maker and market-taker lines and assessing the aggregate impact of HFTs on market quality. Brogaard, Hagströmer, Nordén, and Riordan (2014) study the importance of co-location across HFT firms.

presented in the Appendix.

2 Data, Summary Statistics, and Preliminary Analysis

2.1 The ZEN Database

Our data consist of reports for trades executed on the electronic order book of the LSE, for the 20 largest stocks (by market capitalization) of the FTSE 100 index, over the four months from September 1st to December 31st 2012, a period spanning 80 business days.[7]

The transactions data are obtained from the proprietary ZEN database, which is maintained by the UK Financial Conduct Authority (FCA) and consists of trader-submitted transaction reports that contain information on execution price, trade size, time stamp to the nearest second, location and, importantly, submitter identity. The reports also indicate if the submitter is the buyer or the seller in each transaction, as well as whether a given transaction is executed in a principal or agent capacity. Although ZEN also includes reports of trades that are executed in the OTC space, we restrict our analysis to trades executed on the LSE, the largest UK venue by trading volume. The LSE accounted for between 55 to 70 percent of the total ("lit") volume for the FTSE 100 shares during our sample period.[8]

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 must report their clients' transactions. Firms who are not subject to FCA regulation, and who do not trade through a broker, are not subject to reporting requirements and their reports are not included in ZEN. For our purposes, this implies that we do not observe the trades of HFTs that are direct members of the various UK exchanges, but who are not FCA-regulated. This includes the foreign branches of HFT firms that also have a UK branch; i.e., the activity of the UK branch is captured in ZEN, but the activity of the foreign branch is not. Informal conversations with market regulators suggest that

[7] Our data ends on December 31st 2012, although the last trading day we use in our sample is December 21st 2012. We drop the (two) trading days between Christmas and New Years, since these are days with extremely low volume of trade.

[8] In comparison, NASDAQ, from which many studies on HFTs draw their data never exceeded 25 percent of the total S&P 500 volume over the same period (see the Fidessa Fragmentation Index available at `http://fragmentation.fidessa.com/fragulator/`).

most firms choose to trade on the LSE via their local branches, and we therefore do not expect this to affect coverage in a substantial way. We also cannot identify the activity of individual HFT desks of larger institutions—with multiple trading desks operating in the same market—since all trades from such an institution are reported under a single name. Similarly, it is not feasible to identify the trades of individual HFTs that trade through a broker.

For these reasons, we focus our analysis on stand-alone HFTs that are known to be trading on a proprietary basis. We classify trading firms as HFTs based on discussions with FCA supervisors and from this group we select the 10 largest firms, which account for about 98% of the total trading volume of all such identified HFTs. For confidentiality reasons, we cannot list the names of these 10 HFTs, but they include some of the largest stand-alone HFTs. We also use reports on proprietary trades submitted by the 10 largest investment banks (IBs) in order to compare and contrast the trading activity of the IBs with that of HFTs.[9] As mentioned above, the IB transaction reports originate from all trading activities within these institutions, and as such are a reasonable proxy of overall market activity.

For the remainder of the paper, we will refer to both HFTs and IBs as (trading) firms.

We focus our analysis on the 20 largest of the FTSE 100 stocks, as measured by market capitalization at the beginning of the sample period, for two primary reasons. First, since we aim to study the high-frequency actions and impact of HFTs, it is crucial to use data for stocks that are in fact actively and frequently traded. Second, we need to keep the total number of stocks in the sample to a reasonable number in order to conduct the within-firm/across-stocks VAR analysis, where the dimension of the model grows with the number of stocks as detailed in Section 3.2. This sample restriction is also similar to those used in other studies, such as Hagströmer and Nordén (2013) and Hendershott and Riordan (2013), who use the 30 largest stocks traded on NASDAQ-OMX Stockholm and Deutsche Boerse, respectively.

Finally, we use quote data from the LSE, obtained via Bloomberg, in order to reconstruct the top of the order book every second and to match the ZEN trade reports with

[9]The ZEN data contains a flag that allows us to distinguish between proprietary and agency trades.

the prevailing best bid and ask prices at the time of a given transaction. This allows us to classify trades as either buyer- or seller-initiated, using the usual classification scheme of Lee and Ready (1991). That is, trades that are executed at prices closer to the prevailing bid (ask) are classified as seller- (buyer-) initiated. Trades executed at the quote midprice are classified based on a tick rule: uptick (downtick) trades are classified as buyer- (seller-) initiated. Finally, we also use Bloomberg transaction data to calculate aggregate volume and market wide order-flow for each stock.

2.2 Variable Definitions

From the ZEN data and the matched Bloomberg quote data, we create a number of variables that we use in the analysis. As mentioned above, each trade report is time-stamped to the second and we are therefore able to create time series of trade variables observed at a second-by-second frequency. Our measure of trading volume used in the empirical analysis is the number of shares bought or sold within a given time interval, by a given HFT or IB, in a given stock (in the summary statistics, we also present some figures for the transacted value and the number of trades). In particular, for each firm i (HFT or IB), in stock s at time t, we calculate $Vlm_{i,s,t}$, representing the sum of the number of shares bought and sold during period t. In the raw data, t represents seconds, although, as detailed in the empirical sections, we will use coarser sampling intervals in the actual analysis.

Based on our trade classification scheme, we also measure the "aggressive" and "passive" volume of each firm, for each stock. The "aggressive" volume is the part of the trading volume where the firm acts as the initiator of the trade (i.e., where the firm acts as the market-"taker"), and the "passive" volume is the part of the trading volume where the firm provides the quote hit by another trader (i.e., where the firm acts as the market-"maker"). These volumes will also be referred to as the take- and make-volumes, denoted by $Vlm^{take}_{i,s,t}$ and $Vlm^{make}_{i,s,t}$, respectively. The sum of the aggressive and passive volumes, of course, add up to the total trading volume. The aggressiveness ratio is defined as the fraction of take-volume relative to total volume,

$$AggrRatio_{i,s,t} = \frac{Vlm_{i,s,t}^{take}}{Vlm_{i,s,t}^{take} + Vlm_{i,s,t}^{make}} = \frac{Vlm_{i,s,t}^{take}}{Vlm_{i,s,t}}. \quad (1)$$

Order flow is defined as the difference between aggressive buy-volume and aggressive sell-volume, with the direction of trade viewed from the perspective of the trade initiator (aggressor). The order flow of firm i in stock s is thus given by

$$OF_{i,s,t} = Vlm_{i,s,t}^{take}\left(Buy\right) - Vlm_{i,s,t}^{take}\left(Sell\right), \quad (2)$$

where $Vlm_{i,s,t}^{take}\left(Buy\right)$ and $Vlm_{i,s,t}^{take}\left(Sell\right)$ represent the aggressive buy and sell volumes, respectively.

Aggregate measures of volume and order flow, across HFTs or IBs, are obtained by summing up the variables across all HFTs (IBs). That is,

$$Vlm_{s,t}^{HFT} = \sum_{i \in HFT} Vlm_{i,s,t}, \quad (3)$$

and

$$OF_{s,t}^{HFT} = \sum_{i \in HFT} OF_{i,s,t}. \quad (4)$$

$Vlm_{s,t}^{IB}$ and $OF_{s,t}^{IB}$ are defined analogously, as are aggregates across other variables. The "residual" market-wide volume and order flow, for a given stock, are defined as the sum of the respective variables across all market participants that we observe in Bloomberg, except for the 10 HFTs and 10 IBs.

2.3 Summary Statistics

We start by briefly summarizing some of the characteristics of the HFTs in our sample. For the purpose of comparison, and since they are subsequently used as a reference group, we also show summary statistics for IBs. In particular, in Table 1 we show summary statistics for the daily values of volume (number of shares) and value (in pounds) traded, market share, number of trades and aggressiveness ratio. Separate statistics are shown for HFTs and IBs. The first column in each section (HFT or IB) shows the mean across all firm-stock-days. For instance, the first row shows the average number of shares traded across all HFTs and across all stock-day observations. The following three columns in each section show the corresponding standard deviation and the 5th and 95th percentile of the data pooled across all HFTs and across all stock-day observations. The final column in each section shows the standard deviation of the firm-averaged observations. That is, for each HFT (IB), the average values across all stock-days are calculated, and the standard deviation of these 10 firm-level averages are reported in the final columns (labeled Std($\overline{HFT_i}$) and Std($\overline{IB_i}$), respectively). This final statistic provides an idea of the dispersion or variation in activity across HFTs (IBs).

Table 1 shows that, on average, an HFT firm trades about 300,000 shares and 1.4 million pounds per-stock per-day in the 20 largest FTSE 100 stocks. There is great variation around this average, however, as seen by the standard deviation and the 5th and 95th percentiles of the distribution. There is also considerable variation across firm averages, as seen in the column labeled Std($\overline{HFT_i}$). IBs generally trade more heavily than HFTs, trading on average about 700,000 shares and 3.2 million pounds per-stock per-day. This is expected since IBs are larger organizations with multiple trading desks that simultaneously execute a variety of strategies. Similar to HFTs, there is also great variation in IB trading, although the across-firm variation is relatively smaller than for HFTs (reported in the column labeled Std($\overline{IB_i}$)).

These activity levels give rise to average market shares of about 2.4% for each HFT and 5.5% for each IB. In aggregate, these figures indicate a market share of approximately 25% for the 10 HFTs used in our study. This is very similar to the 25-30% market share of the group of pure HFT firms identified by Hagströmer and Nordén (2013) on NASDAQ-OMX Stockholm.[10]

The aggressiveness ratio statistics reveal a diversity of strategies across HFTs. On average, HFTs trade approximately equal amounts aggressively and passively (average aggressiveness ratio is equal to 0.48), but the pooled standard deviation across all firm-

[10]Substantially higher HFT market shares for U.S. markets are reported in, for instance, Carrion (2013), with HFT market shares of upwards of 70%. As discussed in Hagströmer and Nordén (2013), the lower figures for NASDAQ-OMX Stockholm partly reflects the fact that the activity of hybrid firms, which engage in both purely proprietary high-frequency trading as well as other client-driven activities, are not included when calculating the HFT market share. Similarly, we only count the market share of a set of purely proprietary HFTs.

stock-days is equal to 0.38 and the across firm standard deviation is equal to 0.29. There is thus great variation in aggressiveness, both across stock-days as well as across different HFT firms. IBs have a somewhat lower average aggressiveness ratio (0.41), and there is substantially less variation, especially across different IBs.

2.4 Do HFTs Trade in the Same or Different Stocks?

One of the main questions of this paper is whether an HFT's activity in a given stock is correlated with the trading activity of other HFTs in that same stock. In the next section, we perform a high-frequency VAR analysis to answer this question, but here we first provide some simple illustrative results, based on daily data. The purpose here is not to perform the same analysis on a daily frequency as we do below on a high intra-daily frequency, but rather to document some interesting daily covariations between the activity of different HFTs. As such, we focus on somewhat different measures of trading activity in the daily analysis compared to the high-frequency analysis. It should also be stressed that whereas the high-frequency analysis aims at providing some causal results and interpretations, at least in a Granger sense, the daily analysis must be viewed as strictly non-causal, capturing only (partial) correlations or associations.

In the daily analysis, we consider the covariation across HFTs in the following two trade activity measures: (i) the relative capital allocation ($RCA_{i,s,t}$), defined as the value traded by HFT i, in stock s, on day t, divided by the total value traded across all stocks by HFT i, on day t,[11] and (ii) the aggressiveness ratio ($AggrRatio_{i,s,t}$), as defined in equation (1). Both of these measures capture "relative" aspects of the behaviour of HFTs. The relative capital allocation measures how active an HFT is in a given stock, compared to other stocks. The aggressiveness ratio captures how aggressive (in a liquidity-taking sense) an HFT is in a given stock, relative to its overall trading activity in that stock. Loosely speaking, one might also think of the relative capital allocation as a decision on whether to trade a given stock, and the aggressiveness ratio as the type of trading conditional on being active in a given stock. These measures of trading activity are, at least to some extent, best viewed as a form of (daily) averages, rather than "spot"

[11] For each HFT, the relative capital allocations are calculated based on the total traded volume of that HFT in all one hundred FTSE 100 shares. The average allocation seen in Table 2 is therefore not equal to five percent.

observations, and are generally not well defined at high frequencies where, at a given instant in time, an HFT might be trading only in a single stock as maker or taker. In the high-frequency analysis of the next section, we use actual traded volume and order flow as measures of trading activity, since they are well defined over any sampling interval.

Letting $y_{i,s,t}$ denote either $RCA_{i,s,t}$ or $AggrRatio_{i,s,t}$, we estimate the following regression to analyze daily covariation in HFT behaviour,

$$\begin{aligned} y_{i,s,t} &= \alpha_i + \beta_{RCA} RCA_{-i,s,t} + \beta_{AggrRatio} AggrRatio_{-i,s,t} \\ &\quad + \beta_{MrktCap} MrktCap_{s,t} + \beta_{RV} RV_{s,t} + \beta_{Spread} Spread_{s,t} + u_{i,s,t}. \end{aligned} \tag{5}$$

Here, $RCA_{-i,s,t}$ is the relative volume participation of all other HFTs in stock s on day t, defined as the sum of the volumes traded by all other HFTs (i.e., except HFT i) over the total trading volume of the stock for that day.[12] $AggrRatio_{-i,s,t}$ is the aggregate aggressiveness ratio for all other HFTs in stock s on day t, constructed as the sum of total aggressive volume across all other HFTs, divided by the total trading volume of those HFTs. $MrktCap_{s,t}$ is the market capitalization of the stock (in £billion), $RV_{s,t}$ is the daily realized volatility based on intra-day 5-minute log-returns and $Spread_{s,t}$ is the depth-weighted daily average of all the intra-day quoted spreads expressed as a percentage of the quote midpoint. The main variables of interest in this regression are RCA_{-i} and $AggrRatio_{-i}$, which captures the activity of other HFTs. The other three variables, namely market capitalization, volatility, and bid-ask spread, are known to correlate (at least weakly) with the activity of HFTs and are used as control variables in the analysis (e.g., Hendershott and Riordan, 2013, and Benos and Sagade, 2015). For completeness, we estimate the model with and without these three control variables and report both sets of results.

[12]This definition of $RCA_{-i,s,t}$ is not completely analogous to the definition of the relative capital allocation for an individual firm ($RCA_{i,s,t}$). The latter captures the firm's relative distribution of capital to a given stock, whereas the former captures the fraction of traded volume in a given stock that is due to HFTs other than HFT i. However, we believe this definition of "other" HFT activity is more relevant than the complete analogue of $RCA_{i,s,t}$, which would capture the aggregate relative capital allocation of all other HFTs to stock i. In a given stock, it seems more likely that a given HFT reacts to the overall fraction of trading due to other HFTs (i.e., the current definition of $RCA_{-i,s,t}$), rather than to whether other HFTs have allocated more or less capital to that stock relative to other stocks. We still use the notation $RCA_{-i,s,t}$ to denote the current definition, since this notation makes clear that the left-hand side variable $RCA_{i,s,t}$ and the right-hand side variable $RCA_{-i,s,t}$ are both intended to capture the relative trading activity of HFTs.

An analogous regression for IB trading is also estimated, in order to facilitate comparison with the HFTs. In this case, $RCA_{i,s,t}$ and $AggrRatio_{i,s,t}$ represent the relative capital allocation and aggressiveness ratio, respectively, of IB i, in stock s, on day t. Similarly, $RCA_{-i,s,t}$ and $AggrRatio_{-i,s,t}$ now represent the relative volume participation and aggregate aggressiveness ratio, respectively, for all other IBs in stock s on day t.

Table 2 shows summary statistics for the variables used in the regression. As seen, there are no remarkable differences between HFTs and IBs. Both types of firms allocate relatively uniformly their trading activity across stocks and also trade by using both aggressive and passive orders.

Using the sample of the 20 largest FTSE 100 shares, equation (5) is estimated separately for HFTs and IBs, and in each case pooled across all stocks and HFTs (IBs); i.e., pooled across i and s. An HFT (IB) -specific intercept, α_i, is included in the pooled regressions, capturing any time/stock-invariant HFT (IB) characteristics. The slope coefficients are all kept constant across stocks and HFTs (IBs). Since there are stock-days during which there is no HFT (IB) activity, the model with $y_{i,s,t} = RCA_{i,s,t}$ is estimated using a pooled Tobit specification that accounts for the mass of zeros in the dependent variable. For $y_{i,s,t} = AggrRatio_{i,s,t}$, we do a pooled least squares estimation dropping stock-days for which there is no HFT (IB) trading volume, since on those days the aggressiveness ratio is not defined. We also performed a least squares estimation with a Heckman correction, to account for any potential bias induced by dropping the zero activity days. However, the results from this estimation were almost identical to the plain least squares estimates, and are not presented here.

The results of these estimations are shown in Table 3, with the first four columns showing the HFT and IB estimates for relative capital allocation, with and without control variables, and the last four columns showing the corresponding estimates for the aggressiveness ratio. Table 3 reveals several interesting results. First, HFTs' relative capital allocations are negatively correlated with the aggressiveness of other HFTs. In other words, HFTs appear to avoid trading in stocks where other HFTs are trading more aggressively (columns 1 and 2). However, conditional on trading in a given stock, a given HFT tends to be more aggressive if other HFTs are present (columns 5 and 6). These results are consistent with some specialization across stocks among HFT firms, but also with the possibility that when interacting with other HFTs, a given HFT will become more aggressive because of adverse selection concerns; in other words, if aggressive HFTs can react faster to information than passive ones can cancel outstanding orders, the latter are at risk of having their orders executed at disadvantageous, stale prices. This is also in line with the idea that HFTs jointly trade aggressively in response to common signals, as in Martinez and Rosu (2013).

Without controlling for additional factors, there is evidence that the relative capital allocation of HFTs across stocks is positively correlated (column 1). However, once the control variables are included, this effect disappears and the coefficient turns negative, although it is not statistically significant (column 2). This suggests that while HFTs might prefer to trade in, say, larger stocks, once one controls for this fact, there is no evidence that HFTs allocate more capital to stocks to which other HFTs also allocate more capital. If anything, the opposite appears to be the case, given the negative point estimate seen in the regression with control variables included.

The results for the IBs are somewhat different. IBs tend to allocate more capital to stocks where other IBs are active and aggressive, suggesting, perhaps unsurprisingly, less specialization across stocks among IBs (columns 3 and 4 of Table 3). In addition, IBs tend to be more aggressive in stocks where other IBs are more aggressive (columns 7 and 8). This is somewhat distinct from the HFTs, which appear to be more aggressive in the mere presence of other HFTs.

As documented elsewhere in the literature, HFTs trade more heavily in larger stocks and when volatility is higher and the spread is narrower. Table 3 also suggests that IB activity is very similar to that of HFTs as far as its correlation with the above variables is concerned. Perhaps somewhat surprisingly, HFTs (and IBs) appear to trade more aggressively in stocks with *lower* volatility, which is the opposite of the effect for relative capital allocation. However, it is difficult to give any strong interpretation of this result given the lack of causal identification. Likewise, IBs tend to trade more aggressively in stocks with higher spread, which again is difficult to interpret in a non-causal setting.

Overall, the regression results suggest that there may be significant interdependence between the trading patterns of individual HFTs. In particular, there is some evidence that HFTs tend to avoid each other, but when they do trade in the same stock they appear to be more aggressive than otherwise. Without identifying any causal directions, one should be careful not to push the interpretation of these results too far. The current daily analysis does not take into account the simultaneity in the trading decisions of the individual HFTs (or IBs), and the aggregation of the data to a daily frequency clearly confounds the high-frequency nature of these decisions. However, the patterns seen in these daily aggregates must, by construction, originate from the intra-daily high-frequency trading process, and therefore serve as some motivation for pursuing a high-frequency analysis.

3 Dynamic Correlations in HFT Activity

We now turn to the main part of our analysis, where we attempt to pin down the extent of intra-day correlation in HFT strategies across different HFT firms and also the extent of intra-day correlation in HFT strategies across different stocks. We address both of these questions through the use of (reduced-form) high-frequency vector autoregressions (VARs), which capture the "dynamic" correlations in HFT activity both within-stocks/across-firms and within-firm/across-stocks. That is, we are interested in determining the extent to which current trading by some HFT firm might cause subsequent trading by other HFT firms and also the extent to which current trading by an HFT firm in one stock might cause subsequent trading by the same HFT firm in another stock. As measures of trading activity, we use either total traded volume or order flow.

As before, we use the activity of IBs as a benchmark. In other words, we want to see how much more correlated is the activity of HFTs, within and across stocks, compared to that of the IBs. We use the IBs as a benchmark because of the wide variety of trading strategies that these institutions simultaneously employ across their trading desks. Such strategies may include market making, optimal client-order execution, and high- and low-frequency proprietary trading. IBs should thus be a reasonable proxy of overall market activity.

The high-frequency VAR setup addresses many of the shortcomings of the simple daily correlations presented above. In particular, by explicitly using high-frequency intra-daily data, one avoids the dilution and potential biasing of effects that may result from the daily aggregation. The conditional nature of VARs is also well suited to further explore how HFTs react in response to the actions of other HFTs. Furthermore, the high frequency of observation allows for a much better temporal ordering of events, which in turn provides ways of identifying a more causal chain of events. Formally, we perform a type of Granger causality test. Granger causality is essentially a predictive property and not necessarily "causation" in the typical (contemporaneous) sense used in economics and finance. However, at high frequencies, where the frequency of observation might be similar to the actual frequency of "events" (e.g., trading decisions), evidence of Granger causality might also correspond more closely to the usual notion of causation.

In line with these considerations, the VAR analysis is performed at a 10-second frequency; i.e., using data sampled every 10 seconds. Using 30 lags, the VAR models capture the dynamic dependence in trading over a five-minute period. While we have access to second-by-second trade data, the use of such a high frequency in the VAR modeling would inevitably lead to a much shorter period over which one captures dynamic trade dependencies (i.e., with a 1-second sampling frequency, 30 lags span only half a minute). We therefore view the 10-second sampling choice as a reasonable trade-off between capturing the high-frequency decision making process and allowing for a reasonable period of time over which to model the dynamic evolution of the trade process (i.e., the temporal span of the lags). In addition, if one chooses too high a sampling frequency, most of the trade variables entering the VAR will be zero almost all of the time, which may also lead to bias in the results (similar concerns are expressed by Chaboud, Chiquoine, Hjalmarsson, and Vega, 2014).

Figures 1 and 2 provide further justification for the 10-second sampling frequency. The left-hand panels in the figures show the distributions of the number of trades, by any HFT (or IB) in a given stock, for those periods where *at least one* trade occurred in that stock. Results for 1-second and 10-second time periods are shown in the top left corner, and results for 1- and 5-minute periods are shown in the bottom left corner. That is, conditional on there being at least one trade in a given stock in a given time period, the graphs show the relative frequencies of the number of trades in those periods, averaged across all 20 stocks. Thus, for example, during those 1-second periods for which there was at least one trade, we see in the top left-hand graph in Figure 1 that in approximately 95 percent of these periods, there was, in fact, only one HFT trade. A similar result holds for IBs, as seen in the corresponding graph in Figure 2. In the 10-second periods, nearly 80 percent of those periods with trading activity contain only a single trade (in a given stock), and there is very rarely more than two trades in any 10 second period. The right-hand-side panels in Figures 1 and 2 show analogous distributions for the number of unique traders (i.e., the number of unique HFT or IB firms) trading in a given period. The distributions here look similar to those seen for the number of trades, but they are even more concentrated at unity. In particular, only about 15 (20) percent of the 10-second periods with at least one trade have two unique HFTs (IBs), and hardly any 10-second periods have more than two unique traders. Taken together, these results suggest that by aggregating data to the 10-second frequency, one seldom loses much information on individual trades, since most 10-second intervals only contain at most one trade by any HFT or IB firm in a given stock. Similarly, one does not induce much additional "simultaneity" into the observed trading process by sampling at the 10-second rather than the 1-second frequency, since very few 10-second intervals contain trades by more than one unique trader.

3.1 Within-Stock/Across-Firms HFT Activity

We start by analyzing the correlation of trading activity in a given stock across HFT firms. Let $HFT_{i,s,t}$ be the trading activity of HFT firm i at time t in stock s, and analogously, let $IB_{i,s,t}$ be the trading activity of IB i at time t in stock s. As mentioned above, trading activity is measured either by order flow or total volume, both based on the number of shares traded as discussed in the data section, and sampled at 10-second intervals. Further, define $\mathbf{HFT}_t^s$ as the vector of stacked trading activity in stock s at time t for all $i = 1, ..., 10$ HFTs and define $\mathbf{IB}_t^s$ as the corresponding vector of IB trading activity. That is,

$$\mathbf{HFT}_t^s = \begin{bmatrix} HFT_{1,s,t} \\ \vdots \\ HFT_{10,s,t} \end{bmatrix} \text{ and } \mathbf{IB}_t^s = \begin{bmatrix} IB_{1,s,t} \\ \vdots \\ IB_{10,s,t} \end{bmatrix}.$$

Let $\mathbf{Y}_t^s \equiv (\mathbf{HFT}_t^{s\prime}, \mathbf{IB}_t^{s\prime})'$ denote the stacked trading activity by both HFT and IB firms, and formulate the following high-frequency VAR for stock s,

$$\mathbf{Y}_t^s = \mu^s + \sum_{k=1}^{30} A_k \mathbf{Y}_{t-k}^s + \Lambda \mathbf{X}_{t-1}^s + \Psi \mathbf{G}_t + \epsilon_t^s. \tag{6}$$

The dependent variable, $\mathbf{Y}_t^s \equiv (\mathbf{HFT}_t^{s\prime}, \mathbf{IB}_t^{s\prime})'$, is thus a 20×1 vector of 10-second trading activity in the 10 HFT and 10 IB firms, and A_k, $k = 1, ..., 30$, are 20×20 lag matrix coefficients. Since the data are sampled every 10 seconds, the 30 lags included in the VAR cover the previous five minutes of trading. $\mathbf{X}_{t-1}^s$ consists of lagged control variables not modeled in the VAR. In particular, $\mathbf{X}_{t-1}^s$ includes the cumulative return on stock s during the five minutes prior to the t^{th} observation, the realized volatility of the 10-second returns in stock s during the five minutes prior to the t^{th} observation, and the mean inside spread in stock s during the five minutes prior to the t^{th} observation.[13] In addition, $\mathbf{X}_{t-1}^s$ includes the total market-wide trading activity, captured by the market-wide order flow and the market-wide volume in stock s during the five minutes prior to the t^{th} observation; that is, both market-wide order flow and volume are included as control variables irrespective of whether the dependent variables in the VAR represent firm-specific order flow or volume. $\mathbf{G}_t$ includes deterministic functions of time. In particular, $\mathbf{G}_t$ represents linear and quadratic functions of the time of day (measured by the intra-daily observation number, ranging from 1 to 3060), and linear and quadratic functions of the daily observation number (ranging from $1 - 80$).

The VAR is estimated by pooling data across the sample of the 20 largest FTSE 100 stocks, allowing for stock-specific intercepts (μ^s). All other coefficients are pooled across stocks. There are 80 days in the sample, and 3060 intra-daily intervals each day, resulting in $244,800$ observations per stock, and $4,896,000$ observations in the pooled regression for all 20 stocks.[14] Before inclusion in the VAR, all variables are standardized by their stock-specific standard deviations. This standardization should make the pooling assumption less restrictive and immediately makes the parameter magnitudes correspond

[13]The variables in $\mathbf{X}_{t-1}^s$ are all measured up until one period prior to the current observation; hence the subscript $t - 1$. For instance, the past five-minute returns on stock s are defined as the five-minute returns ending at time $t - 1$.

[14]The first and last five minutes of each trading day are discarded in order to avoid any beginning or end of day effects.

to standard deviation effects.

In this framework, we are interested in testing the following hypotheses: (i) To what extent does trading by an HFT firm in a given stock cause subsequent trading activity by other HFTs in the same stock? (ii) Do we observe similar causation between HFTs and IBs, viewing these two types of traders as distinct groups? We attempt to test these hypotheses within the above VAR model by mapping the general questions into specific coefficient restrictions. In order to facilitate the testing of these hypotheses, it is useful to write the VAR in a format where $\mathbf{Y}_t^s = (\mathbf{HFT}_t^{s\prime}, \mathbf{IB}_t^{s\prime})'$ is written out explicitly. That is, partitioning the coefficient matrices, we can write equation (6) as,

$$\begin{bmatrix} \mathbf{HFT}_t^s \\ \mathbf{IB}_t^s \end{bmatrix} = \mu^s + \sum_{k=1}^{30} \begin{bmatrix} A_{11,k} & A_{12,k} \\ A_{21,k} & A_{22,k} \end{bmatrix} \begin{bmatrix} \mathbf{HFT}_{t-k}^s \\ \mathbf{IB}_{t-k}^s \end{bmatrix} + \Lambda \mathbf{X}_{t-1}^s + \Psi \mathbf{G}_t + \epsilon_t^s. \quad (7)$$

The parameter sub-matrices $(A_{11,k}, A_{12,k}, A_{21,k}, A_{22,k})$ now group the coefficients for the HFTs and IBs. $A_{11,k}$ $(A_{22,k})$ correspond to lag-correlations among HFTs (IBs). The sub-matrix $A_{12,k}$ $(A_{21,k})$ captures the effects of past trading by IBs (HFTs) on the current trading of HFTs (IBs).

To test whether lagged trading in other HFTs affect a given HFT's current trading, we evaluate the null hypothesis that the sum of the off-diagonal coefficients in $A_{11,k}$ across all k lags is equal to zero. Similarly, we test whether past trading by IBs (HFTs) causes current trading of HFTs (IBs) by evaluating the null hypothesis that the sum of all the coefficients across all lags in $A_{12,k}$ $(A_{21,k})$ is equal to zero. In both cases, the null of no causation is rejected if the sum is statistically significant from zero. The sum of the coefficients on the lags of a given variable is proportional to the long-run impact of that variable, and the test can essentially be viewed as a form of long-run Granger causality test.[15] Importantly, to the extent that the relationship is significant, the sign of the sum of coefficients also indicates the direction of the (long-run) relationship; i.e., whether current trading causes more or less trading in the future.

[15]Testing whether, say, $A_{12,k} = 0$ for all k, represents a normal Granger causality test of whether trading by IBs Granger causes trading by HFTs. Testing whether the off-diagonal elements in $A_{11,k}$, for all k, are all equal to zero does not constitute a proper Granger causality test since in this case one is not testing a block-exogeneity hypothesis. Additional coefficient restrictions would need to be imposed in order to formally test for Granger causality.

Table 4 provides the full list of hypotheses that we evaluate, along with the formal coefficient restrictions corresponding to each hypothesis. Results are shown for trading activity measured either as order flow or as total trading volume. In each case, the total sum of all the coefficients are given, along with the value of the Wald test for the null hypothesis that the sum is equal to zero and the corresponding p-value. The test statistics and p-values are obtained through bootstrapping. In particular, they are calculated using a non-parametric block bootstrap at the daily level. By sampling with replacement from the 80 trading days in our sample and including all observations within each day, we preserve any unspecified intra-day error correlation across stocks and firms. As such, these bootstrapped test statistics and p-values are robust to arbitrary intra-day error correlation and we rely on these for our inference.[16]

Starting with the results for order flow, the first row of Table 4 shows strong statistical evidence that current trading in a given stock by a given HFT firm is affected by the past trading in that stock by other HFT firms. In particular, the order flow results suggest that, on average, the current trading of an HFT will tend to be in the same direction as that of the past trades of other HFTs (the sum of the order flow coefficients is positive). In contrast, the second row of Table 4 indicates that the current trading direction of a given IB will tend to be in the opposite direction of past trades by other banks (the sum of the order flow coefficients is negative and strongly significant).

Rows three and four of Table 4 show that there is no strong evidence of the current trading direction of HFTs being affected by the past trade direction of IBs, and vice versa. Specifically, the response of HFTs to IBs is significant at the 5% level, but not at the 1% level, and the response of IBs to HFTs is not statistically significant at any conventional level. The latter result also suggests that HFTs are not anticipating the orders of IBs.[17] The final row of Table 4 provides a formal test of whether the observed difference between HFTs and IBs is also statistically significant. In particular, it confirms that HFTs are significantly more positively correlated than IBs.

The results for volume, which are shown in the last three columns of Table 4, provide

[16]The results of our hypotheses tests using standard test statistics (not reported) are very similar to those using the bootstrapped ones.

[17]Using data from NASDAQ, Hirschey (2013) finds that HFTs anticipate the orders of the general population of non-HFTs. The discrepancy between his and our results could be because he is comparing HFTs with all non-HFTs rather than with just investment banks.

some additional information on the dynamic interaction among HFTs and IBs. Total trading volume is not associated with a given direction of trade, and provides a measure of overall trading activity rather than trading direction. It is therefore not surprising that the results for HFTs and IBs now go in the same direction. In particular, past trading volume by other HFTs (IBs) predict a larger current trading volume for a given HFT (IB), as seen in the first and second rows, respectively. However, the sum of the coefficients for the own lag effect of HFTs (first row) is markedly smaller in magnitude than the own effect for IBs (0.35 versus 1.23), and the difference is statistically significant as seen in row five. As seen in rows three and four, there is also strong evidence that past trading volume by IBs (HFTs) leads to increased volume of HFTs (IBs) as a group.

Overall, the results in Table 4 suggest that both HFTs and IBs tend to increase their trading activity in response to past trading activity of other HFTs or IBs (the results for total trading volume). However, whereas HFTs tend to trade in the same direction as past trades by other HFTs, IBs tend to trade in the reverse direction of past IB trades (the order flow results). Neither HFTs nor IBs appear to be strongly influenced in their trading direction by the trade direction of past IB or HFT trades; that is, trade direction within the own group (HFT or IB) appears to matter the most. There is thus some evidence that HFTs, as a group, are more prone to "dynamically" correlate in their trading direction. In terms of the magnitude of the effects, the sums of the own group order flow coefficients are similar for HFTs and IBs, but of opposite signs (0.22 and -0.35). However, the own group volume coefficients for HFTs is markedly smaller in magnitude than the own effect for IBs (0.35 versus 1.23).

As a form of diagnostic data and model check, the covariance matrices for the residuals from the fitted VAR models are presented in the Appendix. These covariance matrices exhibit a near-diagonal nature, which further validates the choice of sampling frequency. That is, aggregation of the data to a 10-second frequency does not appear to have induced much contemporaneous correlation, corroborating the conclusions drawn from Figures 1 and 2.

3.2 Within-Firm/Across-Stocks HFT Activity

We next turn to the question of whether the trading activity for a given HFT firm is correlated across different stocks. That is, do individual HFTs follow strategies which tend to result in similar trading activities across different stocks? In order to address this question we formulate a VAR similar to the previous one, with the focus on uncovering dynamic correlations across stocks for a given HFT. As previously, $HFT_{i,s,t}$ ($IB_{i,s,t}$) denotes the trading activity of HFT (IB) firm i at time t in stock s, with trading activity measured either by order flow or total volume. But, instead of stacking the trading activity of all firms in a given stock, we now stack the trading activity in all stocks for a given firm. That is, we define,

$$\mathbf{HFT}_t^i = \begin{bmatrix} HFT_{i,1,t} \\ \vdots \\ HFT_{i,20,t} \end{bmatrix} \text{ and } \mathbf{IB}_t^i = \begin{bmatrix} IB_{i,1,t} \\ \vdots \\ IB_{i,20,t} \end{bmatrix}.$$

For each HFT i, we formulate the following VAR,

$$\mathbf{HFT}_t^i = \mu^{i,HFT} + \sum_{k=1}^{30} B_k \mathbf{HFT}_{t-k}^i + \Lambda^{HFT}\mathbf{X}_{t-1} + \Psi^{HFT}\mathbf{G}_t + \epsilon_t^{i,HFT}. \tag{8}$$

The VARs are pooled across HFT firms, allowing for firm-specific intercepts $\mu^{i,HFT}$ and common coefficients B_k, Λ^{HFT} and Ψ^{HFT}. The same VAR is also estimated using IB trading activity, $\mathbf{IB}_t^i$, and again the estimation is done by pooling across all IBs,

$$\mathbf{IB}_t^i = \mu^{i,IB} + \sum_{k=1}^{30} C_k \mathbf{IB}_{t-k}^i + \Lambda^{IB}\mathbf{X}_{t-1} + \Psi^{IB}\mathbf{G}_t + \epsilon_t^{i,IB}. \tag{9}$$

That is, we estimate separate VARs for HFTs and IBs, such that we obtain parameter estimates for both types of firms. The control variables included in $\mathbf{X}_{t-1}$ are the same as those in the within-stocks/across-firms VAR discussed in the previous subsection,[18]

[18] In particular, $\mathbf{X}_{t-1}$ now stacks, for stocks $s = 1, 2, ..., 20$, the following stock-specific control variables: the cumulative return on stock s during the five minutes prior to the t^{th} observation, the realized volatility of the 10-second returns in stock s during the five minutes prior to the t^{th} observation, the mean inside spread in stock s during the five minutes prior to the t^{th} observation, and the market-wide order flow and the market-wide volume in stock s during the five minutes prior to the t^{th} observation. As before, both market-wide order flow and volume are included as control variables irrespective of whether the

and we also use the same standardization of the data prior to estimation. As before, we compute our p-values and test statistics using a non-parametric block bootstrap at the daily level, since these are robust to unspecified intra-day error correlations.

The parameters obtained from the VARs in equations (8) and (9) represent the degree of dynamic correlation of trading for a firm of a given type (HFT or IB) across all 20 stocks in our sample. Table 5 shows the results based on equations (8) and (9). The first two rows in the table show that there is strong evidence of dynamic correlation across stocks for both HFTs and IBs. The results highlight, however, that the correlation for HFTs is considerably stronger than for IBs, as seen from both the coefficient estimates and the test statistics. Both IBs and HFTs thus appear to pursue trading strategies that result in dynamically clustered trading patterns across stocks. However, this effect appears to be substantially stronger for HFTs.[19]

In the Appendix, the covariance matrices for the residuals from the within-firm/across-stocks VAR models are also reported. As in the within-stock/across-firms case, these covariance matrices are almost diagonal, again suggesting that the 10-second sampling frequency does not lead to any issues with contemporaneous correlation.

4 Price impact of correlated HFTs

Given the evidence on dynamically correlated trading activity among HFTs, we end our analysis with a look at the actual impact of correlated trading on stock prices. The potential impact of such behaviour on market prices has been a concern among authorities (e.g., Haldane, 2011). Simultaneous HFT activity in the same stock, and in the same direction, could potentially have an excessively large price impact, causing prices to temporarily deviate from fundamentals. Therefore, in this section, we directly examine if instances of highly correlated trading within stocks have any predictive power for contemporaneous and future returns, and whether the impact of correlated trading by HFTs is any different from that of correlated trading by IBs. We restrict our attention to correlated trading within stocks, since this leads to a natural analysis of whether such

dependent variables in the VAR represent firm-specific order flow or volume.

[19]The strong correlation in trading across stocks appears consistent with HFTs' pursuing index arbitrage strategies. That is, strategies where one attempts to profit from miss-pricing between a traded index (or index futures product) and the underlying basket of stocks that makes up the index.

correlations have an impact on the price process of the given stock.

To capture the extent of correlated trading by HFTs and IBs within stocks, we construct a metric similar to the one used by Lakonishok, Shleifer, and Vishny (1992) to measure herding among institutional investors. In particular, for each stock s and time interval t we calculate

$$CorrTrading_{s,t}^{HFT} = N\left(Buy\right)_{s,t}^{HFT} - \frac{N\left(Buy\right)_{s,t}^{HFT} + N\left(Sell\right)_{s,t}^{HFT}}{2}, \tag{10}$$

where $N\left(Buy\right)_{s,t}^{HFT}$ is the number of aggressive HFT buyers and $N\left(Sell\right)_{s,t}^{HFT}$ is the number of aggressive HFT sellers in stock s in time period t. In a given stock, over a given time interval, an HFT is classified as an aggressive buyer (seller) if its total aggressive buy volume is greater (smaller) than its total aggressive sell volume in that stock during that time interval. That is, if the majority of the HFT's "take-" volume is on the buy (sell) side, it is classified as an aggressive buyer (seller). An HFT that performs no aggressive trading—or if its aggressive buy and sell volumes are identical—in a given stock in a given time interval adds neither to the number of aggressive buyers nor sellers in that time period.

The metric defined in equation (10) effectively calculates the number of excess aggressive buyers or sellers at any given point in time, relative to a situation where HFTs randomly buy and sell with equal probability, independently of one another. When all 10 HFTs in our sample aggressively buy, this metric takes a value of +5, whereas when all 10 HFTs aggressively sell at the same time, the metric takes the value of −5. When aggressive HFTs are equally split between buyers and sellers, or if no HFTs are trading aggressively at all, the metric equals zero. An analogous metric is also constructed for IBs, denoted by $CorrTrading_{s,t}^{IB}$.

The correlation metrics, $CorrTrading_{s,t}^{HFT}$ and $CorrTrading_{s,t}^{IB}$, are calculated for all stocks in the sample of the 20 largest FTSE 100 shares, using minute-by-minute data. The slower one-minute sampling frequency (compared to the 10-second frequency in the VAR analysis), is motivated by the need to sample coarsely enough for there to be sufficiently many observations where numerous HFTs (and/or IBs) trade during the same time interval. That is, the higher the sampling frequency, the more likely it is that just one, or very few, HFT(s) trade in a given time interval, rendering the above correlation metric less useful. At the same time, as in the VAR analysis, the sampling frequency also needs to be high enough to capture the relevant time horizons over which HFTs operate. As a robustness check, we also present results for data sampled at the 5-minute frequency.

The 1-minute and 5-minute sampling intervals are motivated by the bottom graphs in Figures 1 and 2. As discussed previously, for a given sampling frequency, Figures 1 and 2 show the average number of trades in a given stock (left-hand panels), and the average number of unique traders in that stock (right-hand panels), in periods during which there was at least one trade. As seen in the bottom panels, with sampling frequencies of either 1 minute or 5 minutes, there is a fairly wide range of both the likely number of trades as well as the number of unique traders. This suggests that, at these frequencies, one can reasonably expect to capture the contemporaneous correlation of HFTs.

To measure the contemporaneous and lagged price impact associated with correlated trading, we regress 1-minute returns on contemporaneous and lagged order flow, the correlated trading metrics and their lags, as well as the interaction of the two. Thus, our specification takes the form,

$$\begin{aligned}
R_{s,t} \;=\; & \alpha_s + \sum_{i=0}^{5} \beta^{HFT}_{OF,i} OF^{HFT}_{s,t-i} + \sum_{i=0}^{5} \beta^{IB}_{OF,i} OF^{IB}_{s,t-i} + \sum_{i=1}^{5} \beta^{Res}_{OF,i} OF^{Res}_{s,t-i} \\
& + \sum_{i=0}^{5} \beta^{HFT}_{Corr,i} CorrTrading^{HFT}_{s,t-i} + \sum_{i=0}^{5} \beta^{IB}_{Corr,i} CorrTrading^{IB}_{s,t-i} \\
& + \sum_{i=0}^{5} \beta^{HFT}_{OF\times Corr,i} \left(OF^{HFT}_{s,t-i} \times CorrTrading^{HFT}_{s,t-i} \right) \\
& + \sum_{i=0}^{5} \beta^{IB}_{OF\times Corr,i} \left(OF^{IB}_{s,t-i} \times CorrTrading^{IB}_{s,t-i} \right) + u_{s,t}. \qquad (11)
\end{aligned}$$

Here $R_{s,t}$ is the 1-minute return of stock s in period t, and $OF^{HFT}_{s,t}, OF^{IB}_{s,t}$, and $OF^{Res}_{s,t}$ are the order flows from HFTs, IBs and the remainder of the market (the "residual" order flow). $CorrTrading^{HFT}_{s,t}$ and $CorrTrading^{IB}_{s,t}$ are the correlation metrics for HFTs and IBs defined in equation (10). The model is estimated by least squares, pooling the data across stocks while allowing for stock-specific intercepts α_s, and including five lags of all variables. To achieve comparability across stocks, we normalize the order flow variables at the stock level. Specifically, we consider two different normalizations. The first, labeled "own normalization", divides HFT, IB and residual order flow by their own standard deviations at the stock level. Under this normalization, the regression coefficients for these variables may thus be interpreted as the price impact for a typical trade by HFTs, IBs and the rest of the traders, respectively. Alternatively, we normalize all order flow variables by the standard deviation of the market-wide order flow, labeled "market normalization". In this case, the regression coefficients measure the impact of a typical market-wide trade. The returns on the left-hand side of the regressions are always standardized by their own standard deviation at the stock level.[20].

Table 6 reports the regression results. For brevity, we only report the sum of the coefficients for the five lags, and the associated (robust) t-statistics. In column 1, we first run a simple regression of 1-minute returns on contemporaneous and lagged market-wide order flow; the market-wide order flow is denoted by $OF_{s,t}^{Mkt}$ in the table, and is defined as $OF_{s,t}^{Mkt} \equiv OF_{s,t}^{HFT} + OF_{s,t}^{IB} + OF_{s,t}^{Res}$.[21] Consistent with previous findings in the literature, the contemporaneous coefficient is positive and highly statistically significant. The sum of the coefficients for the lagged order flow is negative and also significant, implying that at least a part of the contemporaneous price impact tends to be subsequently reversed.

We next allow the HFT, IB and residual order flows to enter separately into the regression. The results are reported in columns 2 and 5 for the own and market normalization, respectively. We find qualitatively similar results to the regression with market-wide order flow. That is, positive contemporaneous correlation between order flow and returns and negative correlation between past order flow and returns, uniformly across HFTs, IBs and the rest. When normalizing by its own standard deviation, the HFTs' price impact and reversal coefficients (i.e., the sum of the lag coefficients) are substantially smaller in magnitude, but the differences largely disappear when normalizing all order flow variables by the market-wide order flow standard deviation. This is consistent with the typical trade of an HFT being smaller in size than that of IBs. Overall, we find only

[20]The correlation metrics, $CorrTrading_{s,t}^{HFT}$ and $CorrTrading_{s,t}^{IB}$, are not scaled prior to estimation since they are already in a standardized format, taking on values between +5 and -5

[21]This regression can viewed as a restricted version of equation (11), where one imposes the restrictions $\beta_{OF,i}^{HFT} = \beta_{OF,i}^{IB} = \beta_{OF,i}^{Res}$ for $i = 0, ...5$, and all other coefficients are restricted to equal zero.

small differences in the price impacts and reversals between HFTs and IBs.

We next add our metrics of correlated trading to the regressions. The estimation results are reported in columns 3 and 6 of Table 6. We find similar estimates across the two different normalizations, but significantly different results between HFTs and IBs. While the price impact coefficients for HFTs' and IBs' correlated trading are both positive, significant, and of similar magnitude, the impact of HFTs' correlated trading are not subsequently reversed, unlike for IBs. This suggests that HFTs' correlated trading is informed, leading to a permanent price impact. Finally, we also include the interaction terms in the regression, and the results are reported in columns 4 and 7. The interaction terms appear mostly insignificant, with only some weak evidence of statistical significance for the contemporaneous interaction between HFT order flow and trade correlation, and their inclusion has virtually no impact on the coefficients of the other regressors. The actual coefficients for the interaction terms are also very small in absolute value, further signalling that these interactions are not of economic significance.

As a robustness check, we also run the same regressions as above, but using data sampled every five minutes. That is, 5-minute returns are now regressed on the order flow and trade correlation variables. However, in order to keep the temporal span of the lags identical to the 1-minute specification, only one lag is now included. Otherwise, the two specifications are identical, and the results are presented in Table 7. The results presented in Table 7 strongly echo those seen in Table 6. The statistical significance of some of the estimates based on the 5-minute data is somewhat weaker than in the 1-minute case, but otherwise the results are consistent across the two sampling frequencies. Importantly, there is no evidence that HFTs' correlated trading leads to price reversals, although the reversals for IBs are no longer statistically significant.

Overall, these results suggest that HFTs' correlated trading is likely the result of HFTs trading on the same, "correct" , information. In contrast, the correlated trading of IBs is associated with price reversals, suggesting that the correlation in IB strategies is less informationally driven.

5 Conclusion

Using a unique data set on the transactions of individual high-frequency traders (HFTs), we examine the interactions between different HFTs and the impact of such interactions on price discovery. Our main results show that for trading in a given stock, HFT firm order flows are positively correlated at high-frequencies. In contrast, when performing the same analysis on our control sample of investment banks, we find that their order flows are negatively correlated. Put differently, aggressive (market-"taking") volume by an HFT will tend to lead to more aggressive volume, in the same direction of trade, by other HFTs over the next few minutes. For banks the opposite holds, and a bank's aggressive volume will tend to lead to aggressive volume in the opposite direction by other banks. As far as activity across different stocks is concerned, HFTs also tend to trade in the same direction across different stocks to a significantly larger extent than banks.

Given the apparent tendency to commonality in trading activity and trading direction among HFTs, we further examine whether periods of high HFT correlation are associated with price impacts that are subsequently reversed. Such reversals might be interpreted as evidence of high trade correlations leading to short-term price dislocations and excess volatility. However, we find that instances of correlated trading among HFTs are associated with a permanent price impact, whereas instances of correlated bank trading are, in fact, associated with future price reversals. We view this as evidence that the commonality of order flows in the cross-section of HFTs is the result of HFTs' trades being informed, and as such have the same sign at approximately the same time. In other words, HFTs appear to be collectively buying and selling at the "right" time. The results are also in agreement with the conclusions of Chaboud, Chiquoine, Hjalmarsson, and Vega (2014), who find evidence of commonality among the trading strategies of algorithmic trades in the foreign exchange market, but who also find no evidence that such commonality appears to be creating price pressures and excess volatility that would be detrimental to market quality.

A final caveat is in order. The time period we examine is one of relative calm in the UK equity market. This means that additional research on the behaviour of HFTs, particularly during times of severe stress in equity and other markets, would be necessary in order to fully understand their role and impact on price efficiency.

Appendix: Error Covariance Matrices from the VARs

Tables A1 and A2 show the error covariance matrices from the within-stock/across-firms VARs represented by equation (6), using either order flow or total volume as dependent variables, respectively, and sampled at the 10-second frequency. Since the data are standardized prior to estimation, the diagonal elements in the covariance matrices (i.e., the variances) are close to unity, and the off-diagonal elements are thus close to correlations. As is seen, when using order flow as the measure of trade activity (Table A1), the off-diagonal elements are all close to zero, and never greater than 0.1 in absolute value. This suggests that at the current sampling frequency of 10-second intervals, there is little concern that the analysis is confounded by strong *contemporaneous* interactions. Table A2 shows the corresponding error covariance matrix when trading activity is measured by total volume. In this case, some of the off-diagonal elements take on slightly larger values than in the order flow case, although they are still all below 0.25 in absolute value and most of them are still below 0.1. Thus, although the contemporaneous error structure is not quite as clean as in the order flow case, the 10-second sampling frequency does not seem to induce any large contemporaneous correlations in the case of volume either.

Tables A3 and A4 show the error covariance matrices from the within-firm/across-stocks VARs in equations (8) and (9), using order flow as the measure of trade activity. In this specification, a separate VAR is estimated for HFTs and IBs, respectively, and Tables A3 and A4 show the corresponding error covariance matrices. In both cases, the error structure is almost perfectly diagonal. This is also true when using total volume rather than order flow as a dependent variable, although these results are omitted in the interest of space.

Table A1: Covariance matrix of the error terms from the within-stock/across-firms VAR specification in equation (6), with order flow as dependent variable. HFT_i (IB_i) represents the residuals for the ith HFT (IB) firm in the VAR specification. The results are based on 10-second data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012.

	HFT_1	HFT_2	HFT_3	HFT_4	HFT_5	HFT_6	HFT_7	HFT_8	HFT_9	HFT_{10}	IB_1	IB_2	IB_3	IB_4	IB_5	IB_6	IB_7	IB_8	IB_9	IB_{10}
HFT_1	**0.85**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HFT_2	0.00	**1.00**	0.02	-0.02	0.08	0.00	0.02	0.07	0.01	0.01	0.01	0.03	0.02	0.02	0.00	0.01	0.03	0.07	0.09	0.06
HFT_3	0.00	0.02	**0.99**	0.00	0.01	0.00	0.03	0.01	0.00	0.01	-0.01	0.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.01	0.00
HFT_4	0.00	-0.02	0.00	**0.75**	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	-0.02	-0.01	-0.03	-0.03
HFT_5	0.00	0.08	0.01	0.00	**0.90**	0.00	0.01	0.03	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.02	0.02	0.04	0.02
HFT_6	0.00	0.00	0.00	0.00	0.00	**0.94**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HFT_7	0.00	0.02	0.03	0.00	0.01	0.00	**0.84**	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00
HFT_8	0.00	0.07	0.01	-0.01	0.03	0.00	0.00	**0.99**	0.00	0.00	0.00	0.01	-0.01	-0.01	-0.01	-0.01	0.01	0.01	0.03	0.01
HFT_9	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HFT_{10}	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.00	0.00	**0.94**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
IB_1	0.00	0.01	-0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	**0.99**	0.00	-0.01	0.00	0.00	-0.01	0.00	0.00	0.01	0.00
IB_2	0.00	0.03	0.00	-0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	**0.99**	0.00	0.00	0.00	-0.01	0.00	0.01	0.02	-0.01
IB_3	0.00	0.02	-0.01	0.00	0.01	0.00	0.00	-0.01	0.00	0.00	-0.01	0.00	**0.99**	0.00	-0.01	0.00	0.00	0.01	0.01	-0.01
IB_4	0.00	0.02	0.00	0.00	0.01	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	**0.98**	0.00	0.00	0.00	0.00	0.01	0.00
IB_5	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	-0.01	0.00	**0.98**	-0.01	0.00	0.00	0.00	-0.02
IB_6	0.00	0.01	0.00	0.00	0.01	0.00	0.00	-0.01	0.00	0.00	-0.01	-0.01	0.00	0.00	-0.01	**0.97**	0.01	0.00	0.01	0.00
IB_7	0.00	0.03	0.00	-0.02	0.02	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	**0.99**	0.01	0.01	0.00
IB_8	0.00	0.07	0.00	-0.01	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	**0.99**	0.05	0.01
IB_9	0.00	0.09	0.01	-0.03	0.04	0.00	0.01	0.03	0.00	0.01	0.01	0.02	0.01	0.01	0.00	0.01	0.01	0.05	**1.00**	0.04
IB_{10}	0.00	0.06	0.00	-0.03	0.02	0.00	0.00	0.01	0.00	0.01	0.00	-0.01	-0.01	0.00	-0.02	0.00	0.00	0.01	0.04	**0.99**

Table A2: Covariance matrix of the error terms from the within-stock/across-firms VAR specification in equation (6), with total volume as dependent variable. HFT_i (IB_i) represents the residuals for the ith HFT (IB) firm in the VAR specification. The results are based on 10-second data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012.

	HFT_1	HFT_2	HFT_3	HFT_4	HFT_5	HFT_6	HFT_7	HFT_8	HFT_9	HFT_{10}	IB_1	IB_2	IB_3	IB_4	IB_5	IB_6	IB_7	IB_8	IB_9	IB_{10}
HFT_1	**0.90**	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.02
HFT_2	0.01	**0.98**	0.11	0.04	0.11	0.02	0.06	0.13	0.04	0.04	0.12	0.19	0.14	0.11	0.10	0.13	0.14	0.23	0.17	0.16
HFT_3	0.01	0.11	**0.94**	0.01	0.07	0.01	0.07	0.12	0.02	0.04	0.08	0.11	0.07	0.08	0.05	0.08	0.08	0.16	0.10	0.10
HFT_4	0.00	0.04	0.01	**0.74**	0.01	0.00	0.00	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.03	0.02	0.06	0.04
HFT_5	0.00	0.11	0.07	0.01	**0.88**	0.01	0.03	0.09	0.02	0.02	0.05	0.08	0.06	0.07	0.03	0.06	0.06	0.10	0.09	0.09
HFT_6	0.00	0.02	0.01	0.00	0.01	**0.93**	0.00	0.02	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.01	0.02	0.01	0.01
HFT_7	0.00	0.06	0.07	0.00	0.03	0.00	**0.83**	0.04	0.01	0.02	0.03	0.04	0.02	0.03	0.01	0.03	0.03	0.06	0.05	0.03
HFT_8	0.01	0.13	0.12	0.03	0.09	0.02	0.04	**0.98**	0.03	0.04	0.08	0.13	0.09	0.09	0.05	0.09	0.09	0.17	0.12	0.13
HFT_9	0.00	0.04	0.02	0.01	0.02	0.00	0.01	0.03	**0.99**	0.01	0.02	0.03	0.02	0.02	0.02	0.02	0.02	0.04	0.04	0.03
HFT_{10}	0.00	0.04	0.04	0.01	0.02	0.01	0.02	0.04	0.01	**0.91**	0.02	0.04	0.03	0.02	0.01	0.03	0.03	0.06	0.04	0.04
IB_1	0.00	0.12	0.08	0.01	0.05	0.00	0.03	0.08	0.02	0.02	**0.97**	0.08	0.08	0.06	0.04	0.08	0.08	0.12	0.09	0.10
IB_2	0.00	0.19	0.11	0.01	0.08	0.01	0.04	0.13	0.03	0.04	0.08	**0.97**	0.10	0.08	0.05	0.10	0.09	0.17	0.16	0.15
IB_3	0.00	0.14	0.07	0.01	0.06	0.00	0.02	0.09	0.02	0.03	0.08	0.10	**0.97**	0.06	0.04	0.06	0.07	0.12	0.11	0.10
IB_4	0.01	0.11	0.08	0.01	0.07	0.01	0.03	0.09	0.02	0.02	0.06	0.08	0.06	**0.96**	0.05	0.06	0.07	0.12	0.09	0.09
IB_5	0.00	0.10	0.05	0.00	0.03	0.00	0.01	0.05	0.02	0.01	0.04	0.05	0.04	0.05	**0.97**	0.04	0.03	0.08	0.08	0.07
IB_6	0.01	0.13	0.08	0.01	0.06	0.01	0.03	0.09	0.02	0.03	0.08	0.10	0.06	0.06	0.04	**0.95**	0.08	0.13	0.10	0.10
IB_7	0.00	0.14	0.08	0.03	0.06	0.01	0.03	0.09	0.02	0.03	0.08	0.09	0.07	0.07	0.03	0.08	**0.97**	0.13	0.11	0.11
IB_8	0.01	0.23	0.16	0.02	0.10	0.02	0.06	0.17	0.04	0.06	0.12	0.17	0.12	0.12	0.08	0.13	0.13	**0.96**	0.20	0.18
IB_9	0.01	0.17	0.10	0.06	0.09	0.01	0.05	0.12	0.04	0.04	0.09	0.16	0.11	0.09	0.08	0.10	0.11	0.20	**0.98**	0.15
IB_{10}	0.02	0.16	0.10	0.04	0.09	0.01	0.03	0.13	0.03	0.04	0.10	0.15	0.10	0.09	0.07	0.10	0.11	0.18	0.15	**0.97**

Table A3: Covariance matrix of the error terms from the within-firm/across-stocks VAR specification in equation (8), for HFTs, using order flow as dependent variable. S_i represents the residuals for the ith stock in the VAR specification. The results are based on 10-second data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
S_1	**0.99**	0.03	0.01	0.02	0.01	0.01	0.01	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.02
S_2	0.03	**0.99**	0.02	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.06	0.02	0.02	0.01
S_3	0.01	0.02	**0.90**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S_4	0.02	0.03	0.01	**0.99**	0.01	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
S_5	0.01	0.02	0.01	0.01	**0.90**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.01
S_6	0.01	0.02	0.01	0.01	0.01	**0.90**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S_7	0.01	0.02	0.01	0.02	0.01	0.01	**1.00**	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01
S_8	0.03	0.03	0.01	0.02	0.01	0.01	0.02	**0.99**	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.03
S_9	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	**0.89**	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01
S_{10}	0.01	0.02	0.01	0.02	0.01	0.01	0.02	0.01	0.01	**0.80**	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01
S_{11}	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	**0.99**	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01
S_{12}	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**0.99**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S_{13}	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**0.89**	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S_{14}	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01	**0.40**	0.00	0.01	0.01	0.01	0.01	0.01
S_{15}	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.00	**0.99**	0.01	0.01	0.01	0.01	0.03
S_{16}	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**0.90**	0.01	0.01	0.01	0.01
S_{17}	0.02	0.06	0.01	0.01	0.00	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**1.00**	0.01	0.01	0.01
S_{18}	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**1.00**	0.01	0.01
S_{19}	0.01	0.02	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**1.00**	0.01
S_{20}	0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	**0.90**

Table A4: Covariance matrix of the error terms from the within-firm/across-stocks VAR specification in equation (9), for IBs, using order flow as dependent variable. S_i represents the residuals for the ith stock in the VAR specification. The results are based on 10-second data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
S_1	**0.99**	0.03	0.01	0.02	0.01	0.01	0.01	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.02
S_2	0.03	**0.99**	0.02	0.03	0.02	0.02	0.02	0.03	0.02	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.06	0.02	0.02	0.01
S_3	0.01	0.02	**0.90**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S_4	0.02	0.03	0.01	**0.99**	0.01	0.01	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02
S_5	0.01	0.02	0.01	0.01	**0.90**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.01	0.01
S_6	0.01	0.02	0.01	0.01	0.01	**0.90**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S_7	0.01	0.02	0.01	0.02	0.01	0.01	**1.00**	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01
S_8	0.03	0.03	0.01	0.02	0.01	0.01	0.02	**0.99**	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.03
S_9	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	**0.89**	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01
S_{10}	0.01	0.02	0.01	0.02	0.01	0.01	0.02	0.01	0.01	**0.80**	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01
S_{11}	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	**0.99**	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01
S_{12}	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**0.99**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S_{13}	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**0.89**	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S_{14}	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01	**0.40**	0.00	0.01	0.01	0.01	0.01	0.01
S_{15}	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.00	**0.99**	0.01	0.01	0.01	0.01	0.03
S_{16}	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**0.90**	0.01	0.01	0.01	0.01
S_{17}	0.02	0.06	0.01	0.01	0.00	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**1.00**	0.01	0.01	0.01
S_{18}	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**1.00**	0.01	0.01
S_{19}	0.01	0.02	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	**1.00**	0.01
S_{20}	0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	**0.90**

References

Benos E. and S. Sagade, 2015, Price Discovery and the Cross-section of high-frequency trading, working paper.

Brogaard J., Hagströmer B., Nordén L. and R. Riordan, 2014, Trading Fast and Slow: Colocation and Market Quality, working paper.

Brogaard J., Hendershott T. and R. Riordan, 2014, High frequency trading and price discovery, *Review of Financial Studies* 27, 2267-2306.

Carrion, A. (2013), Very Fast Money: High-Frequency Trading on the NASDAQ, *Journal of Financial Markets* 16, 680-711.

Chaboud A., Chiquoine B., Hjalmarsson E. and C. Vega, 2014, Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market, *Journal of Finance* 69, 2045-2084.

DeLong, B., Shleifer, A., Summers, L. and R.J. Waldmann, 1990, Positive Feedback Investment Strategies and Destabilizing Rational Speculation, *Journal of Finance* 45, 379-395.

Froot K., Scharfstein D. and J. Stein, 1992, Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation, *Journal of Finance* 47, 1461-1484.

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 L. Nordén, 2013, The diversity of high-frequency traders, *Journal of Financial Markets* 16, 741-770.

Hagströmer B., Nordén L., and D. Zhang, 2014, The aggressiveness of high-frequency traders, *The Financial Review*, forthcoming.

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

Hendershott T. and R. Riordan, 2013, Algorithmic Trading and the Market for Liquidity, *Journal of Financial and Quantitative Analysis* 48, 1001-1024.

Hirschey N., 2013, Do High-Frequency Traders Anticipate Buying and Selling Pressure?, working paper.

Jarrow R. and P. Protter, 2012, A Dysfunctional Role of High Frequency Trading in Electronic Markets, *International Journal of Theoretical and Applied Finance* 15.

Khandani A. and A. W. Lo, 2011, What Happened To The Quants In August 2007?: Evidence from Factors and Transactions Data, *Journal of Financial Markets* 14, 1-46.

Kirilenko A., Kyle A., Samadi M. and T. Tuzun, 2014, The Flash Crash: The Impact of High Frequency Trading on an Electronic Market, working paper.

Kondor P., 2009, Risk in Dynamic Arbitrage: Price Effects of Convergence Trading, *Journal of Finance* 64, 631-655.

Kozhan R. and W. Wah Tham, 2012, Execution Risk in High-Frequency Arbitrage, *Management Science* 58, 2131-2149.

Lakonishok J., Shleifer A. and R. Vishny, 1992, The Impact of Institutional Trading on Stock Prices, *Journal of Financial Economics* 32, 23-43.

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

Martinez V. and I. Rosu, 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, 712-740.

Oehmke M., 2009, Gradual Arbitrage, working paper, Columbia University.

Stein J., 2009, Presidential Address: Sophisticated Investors and Market Efficiency, *Journal of Finance* 64, 1571-1548.

White M.J., 2014, Enhancing Our Equity Market Structure, speech, Sandler O'Neill & Partners, L.P. Global Exchange and Brokerage Conference New York, N.Y. available at: `http://www.sec.gov/News/Speech/Detail/Speech/1370542004312`.

Table 1: Summary statistics on HFT and IB activity, based on data pooled across firm-stock-days. The first two columns in each panel indicate the mean and standard deviation across all firm-stock-days, and columns three and four indicate the 5th and 95th percentile of all firm-stock-day observations. The final column in each panel reports the standard deviation of the firm-specific means; i.e., the standard deviation of the 10 firm-level averages for each variable. The results are based on aggregated daily data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012.

	High-Frequency Traders					Investment Banks				
	Mean	Std(pooled)	5%	95%	Std ($\overline{HFT}_i$)	Mean	Std(pooled)	5%	95%	Std ($\overline{IB}_i$)
Volume traded	322,734	1,591,999	0	1,329,051	524,703	753,463	1,866,751	10,873	3,476,878	439,942
Value traded ('000s)	1,367	2,521	0	5840	608	3202	3818	166	9725	695
Market share	2.40%	3.70%	0	10.60%	3.0%	5.50%	4.50%	0%	14.30%	3.1%
No. of trades	201	341	0	814	225	427	481	30	1232	239
AggrRatio	0.48	0.38	0	1.00	0.29	0.41	0.20	0.14	0.80	0.10

Table 2: Summary statistics of the stock-day level variables. The statistics are calculated over stocks, days and firms in each of the two groups (HFTs and IBs). The relative capital allocation (RCA) by a given firm, on a given stock-day, is the value traded on that stock and day over the total volume traded across all stocks on the same day, by the same firm. $MrktCap$ is the market capitalization on a given stock-day, expressed in £billions. The aggressiveness ratio ($AggrRatio$) is the sum of total aggressive volume on a given stock-day by the firms of each group divided by their total traded volume on the same stock-day. The realized volatility (RV) is the realized average daily volatility based on 5-minute log returns. Finally, *Spread* is the depth-weighted daily average of the intra-day spreads expressed as a percentage of the quote midpoint. The results are based on aggregated daily data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012.

	RCA		*AggrRatio*		*MrktCap*	*RV*	*Spread*
	HFT	IB	HFT	IB	*(£bn)*	*(%)*	*(%)*
Mean	0.034	0.032	0.478	0.415	40.38	1.656	0.060
Std	0.044	0.028	0.378	0.198	32.44	1.457	0.019
Min	0	0	0	0	2.60	0.391	0.038
Max	0.169	0.112	1	1	122.1	9.272	0.160
N	16,000	16,000	12,224	15,963	16,000	16,000	16,000

Table 3: Daily regression results. The table shows the estimation results of equation (5). The dependent variable is the relative capital allocation (*RCA*) of the HFT or IB in a given stock and on a given day, or the aggressiveness ratio (*AggrRatio*) defined in equation (1). The model with relative capital allocation as the dependent variable is estimated using a pooled Tobit specification, and the model with the aggressiveness ratio as the dependent variable is estimated using pooled least squares. The model is estimated with and without the control variables (*MrktCap*, *RV*, and *Spread*). RCA_{-i} is the relative volume participation of all other HFTs in a given stock and $AggrRatio_{-i}$ is the aggregate aggressiveness ratio for all other HFTs in a given stock. *MrktCap*, *RV* and *Spread* are the market capitalization, the daily realized volatility and the depth-weighted average intra-day spread of stock i on day t, respectively. The results are based on aggregated daily data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012. t-statistics based on robust standard errors are in parentheses; * and ** denote significance at 5% and 1% respectively.

	RCA				*AggrRatio*			
	HFT	HFT	IB	IB	HFT	HFT	IB	IB
RCA_{-i}	0.0689^{**}	-0.0096	0.0384^{**}	0.0499^{**}	0.2787^{**}	0.1163^{**}	0.0262^{*}	0.0163
	(12.39)	(−1.62)	(22.87)	(31.83)	(9.48)	(3.58)	(2.54)	(1.46)
$AggrRatio_{-i}$	-0.0190^{**}	-0.0361^{**}	0.0238^{**}	0.0637^{**}	0.0592^{**}	0.0187	0.1695^{**}	0.1298^{**}
	(−6.32)	(−12.24)	(7.05)	(19.70)	(3.73)	(1.15)	(7.92)	(5.68)
MrktCap		0.0002^{**}		0.0001^{**}		0.0001		-0.0002^{**}
		(16.06)		(19.68)		(0.97)		(−3.86)
RV		0.0063^{**}		0.0070^{**}		-0.0200^{**}		-0.0078^{**}
		(16.27)		(38.26)		(−10.88)		(−7.28)
Spread		-0.7924^{**}		-0.5127^{**}		-0.4484^{**}		0.4482^{**}
		(−29.83)		(−42.37)		(−2.85)		(4.80)
N	16,000	16,000	16,000	16,000	12,224	12,224	15,963	15,963

Table 4: Within-Stock/Across-Firms HFT activity. The table reports point estimates and test statistics of hypotheses regarding HFT activity within stocks and across HFT firms. The hypotheses restrictions are based on the corresponding VAR model in equation (6). The trading activity variables are either the HFT/IB order flow or traded volume. The results are based on 10-second data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012. * and ** denote significance at 5% and 1%, respectively.

		Order Flow			Total Volume		
Hypothesis to be tested	Coefficient Restriction	Coefficient outcome	Wald statistic	p-value	Coefficient outcome	Wald statistic	p-value
Are HFTs correlated within stocks?	$\sum_k \sum_i \sum_{j \neq i} (A_{11,k})_{i,j} = 0$	0.2227^{**}	39.51	0.0000	0.3567^{**}	34.84	0.0000
Are banks correlated within stocks?	$\sum_k \sum_i \sum_{j \neq i} (A_{22,k})_{i,j} = 0$	-0.3517^{**}	41.55	0.0004	1.2346^{**}	27.28	0.0000
Do HFTs respond to banks?	$\sum_k \sum_i \sum_{j \neq i} (A_{12,k})_{i,j} = 0$	-0.1003^{*}	5.19	0.0230	0.6057^{**}	30.94	0.0000
Do banks respond to HFTs?	$\sum_k \sum_i \sum_{j \neq i} (A_{21,k})_{i,j} = 0$	0.0871	2.21	0.1370	0.6649^{**}	51.32	0.0000
Are HFTs more correlated than banks?	$\sum_k \sum_i \sum_{j \neq i} (A_{11,k})_{i,j} - \sum_k \sum_i \sum_{j \neq i} (A_{22,k})_{i,j} = 0$	0.5744^{**}	69.39	0.0000	-0.8779^{**}	13.22	0.0000

Table 5: Within-Firm/Across-Stocks HFT activity. The table reports point estimates and test statistics of hypotheses regarding individual HFT activity across stocks. The hypotheses restrictions are based on the corresponding VAR models in equations (8) and (9). The trading activity variables are either the HFT/IB order flow or traded volume. The results are based on 10-second data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012. * and ** denote significance at 5% and 1% respectively.

		Order Flow			Total Volume		
Hypothesis to be tested	Coefficient Restriction	Coefficient outcome	Wald statistic	p-value	Coefficient outcome	Wald statistic	p-value
Is HFT trading correlated across stocks?	$\sum_k \sum_i \sum_{j \neq i} (B_k)_{i,j}$	3.9195**	103.85	0.0000	5.8664**	1030.30	0.0000
Is bank trading correlated across stocks?	$\sum_k \sum_i \sum_{j \neq i} (C_k)_{i,j}$	0.7494**	26.18	0.0000	2.5783**	212.29	0.0000
Are the parameters of HFTs and banks equal across stocks?	$\Lambda^{HFT} = \Lambda^{IB}$, $\Psi^{HFT} = \Psi^{IB}$, $B_k = C_k \quad \forall k$		60446	0.0000		55431	0.0000

Table 6: Results for correlated trading, using 1-minute data. The table shows regressions of 1-minute returns on contemporaneous and lagged order flow, correlated trading metrics and interactions of the two. The order flow variables are calculated as the total aggressive buy volume minus the total aggressive sell volume by the HFTs, IBs and all other market participants, respectively. The measure of correlated trading $CorrTrading_t^{HFT}$ ($CorrTrading_t^{IB}$), is defined in equation (10). The results are based on 1-minute data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012. Robust t-statistics are in parentheses. * and ** denote significance at 5% and 1%, respectively.

		Own normalization			Market normalization		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
OF_t^{MKT}	0.438** (34.60)						
$OF_{t-1,t-5}^{MKT}$	−0.095** (−10.10)						
OF_t^{HFT}		0.224** (19.80)	0.151** (13.90)	0.151** (14.00)	0.407** (30.70)	0.279** (18.50)	0.279** (18.60)
$OF_{t-1,t-5}^{HFT}$		−0.041** (−7.95)	−0.046** (−7.85)	−0.046** (−7.90)	−0.078** (−8.84)	−0.087** (−8.93)	−0.087** (−8.95)
OF_t^{IB}		0.369** (23.70)	0.305** (20.70)	0.305** (20.90)	0.432** (28.80)	0.359** (24.50)	0.359** (24.70)
$OF_{t-1,t-5}^{IB}$		−0.081** (−8.15)	−0.066** (−6.70)	−0.066** (−6.71)	−0.099** (−9.44)	−0.083** (−7.54)	−0.083** (−7.52)
OF_t^{RES}		0.376** (35.20)	0.342** (32.40)	0.342** (32.40)	0.489** (30.60)	0.445** (28.10)	0.445** (28.00)
$OF_{t-1,t-5}^{RES}$		−0.086** (−9.94)	−0.075** (−9.22)	−0.075** (−9.26)	−0.113** (−9.73)	−0.099** (−9.01)	−0.099** (−9.04)
$CorrTrading_t^{HFT}$			0.252** (12.50)	0.252** (12.40)		0.244** (11.80)	0.244** (11.80)
$CorrTrading_{t-1,t-5}^{HFT}$			0.018* (2.24)	0.019* (2.25)		0.020** (2.59)	0.020** (2.60)
$CorrTrading_t^{IB}$			0.201** (17.00)	0.201** (17.00)		0.201** (16.90)	0.201** (16.90)
$CorrTrading_{t-1,t-5}^{IB}$			−0.068** (−6.65)	−0.068** (−6.64)		−0.065** (−6.34)	−0.064** (−6.34)
$OF_t^{HFT} \times CorrTrading_t^{HFT}$				−0.006 (−1.83)			−0.010 (−1.57)
$OF_{t-1,t-5}^{HFT} \times CorrTrading_{t-1,t-5}^{HFT}$				0.000 (−0.12)			0.001 (0.12)
$OF_t^{IB} \times CorrTrading_t^{IB}$				−0.005 (−1.45)			−0.004 (−1.55)
$OF_{t-1,t-5}^{IB} \times CorrTrading_{t-1,t-5}^{IB}$				0.000 (0.11)			−0.001 (−0.28)
R^2	0.191	0.191	0.214	0.215	0.194	0.217	0.217
N	719,900	719,900	719,900	719,900	719,900	719,900	719,900

Table 7: Results for correlated trading, using 5-minute data. The table shows regressions of 5-minute returns on contemporaneous and lagged order flow, correlated trading metrics and interactions of the two. The order flow variables are calculated as the total aggressive buy volume minus the total aggressive sell volume by the HFTs, IBs and all other market participants, respectively. The measure of correlated trading $CorrTrading_t^{HFT}$ ($CorrTrading_t^{IB}$), is defined in equation (10). The results are based on 5-minute data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012. Robust t-statistics are in parentheses. * and ** denote significance at 5% and 1%, respectively.

		Own normalization			Market normalization		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
OF_t^{MKT}	0.362** (23.38)						
OF_{t-1}^{MKT}	−0.066** (−10.97)						
OF_t^{HFT}		0.084** (5.53)	0.053** (3.04)	0.053** (3.04)	0.176** (7.18)	0.119** (3.94)	0.119** (3.94)
OF_{t-1}^{HFT}		−0.017** (−3.82)	−0.022** (−4.15)	−0.022** (−4.19)	−0.038** (−5.25)	−0.047** (−5.56)	−0.048** (−5.59)
OF_t^{IB}		0.383** (20.95)	0.352** (20.92)	0.352** (21.15)	0.425** (20.57)	0.390** (20.79)	0.390** (21.00)
OF_{t-1}^{IB}		−0.075** (−8.57)	−0.074** (−8.13)	−0.074** (−8.19)	−0.086** (−10.19)	−0.086** (−9.55)	−0.086** (−9.61)
OF_t^{RES}		0.274** (21.82)	0.260** (20.87)	0.260** (20.87)	0.358** (19.51)	0.340** (18.77)	0.340** (18.74)
OF_{t-1}^{RES}		−0.056** (−8.50)	−0.053** (−8.34)	−0.053** (−8.40)	−0.076** (−8.73)	−0.073** (−8.58)	−0.073** (−8.62)
$CorrTrading_t^{HFT}$			0.098** (6.65)	0.098** (6.65)		0.092** (6.03)	0.092** (6.03)
$CorrTrading_{t-1}^{HFT}$			0.011 (1.87)	0.011 (1.88)		0.012* (2.27)	0.012* (2.27)
$CorrTrading_t^{IB}$			0.086** (9.73)	0.086** (9.76)		0.087** (10.04)	0.087** (10.05)
$CorrTrading_{t-1}^{IB}$			−0.006 (−1.39)	−0.006 (−1.37)		−0.004 (−0.92)	−0.004 (−0.91)
$OF_t^{HFT} \times CorrTrading_t^{HFT}$				−0.007 (−1.69)			−0.014 (−1.75)
$OF_{t-1}^{HFT} \times CorrTrading_{t-1}^{HFT}$				−0.005 (−1.20)			−0.006 (−0.89)
$OF_t^{IB} \times CorrTrading_t^{IB}$				−0.006 (−0.94)			−0.005 (−0.71)
$OF_{t-1}^{IB} \times CorrTrading_{t-1}^{IB}$				0.001 (0.38)			0.000 (0.07)
R^2	0.130	0.140	0.148	0.148	0.144	0.150	0.150
N	143,980	143,980	143,980	143,980	143,980	143,980	143,980

Figure 1: Frequency distributions of the number of HFT trades and the number of HFT firms present within various time intervals. All the distributions are conditional on there being at least one HFT trade in a given interval. The time intervals range from one second to five minutes. The results are based on data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012.

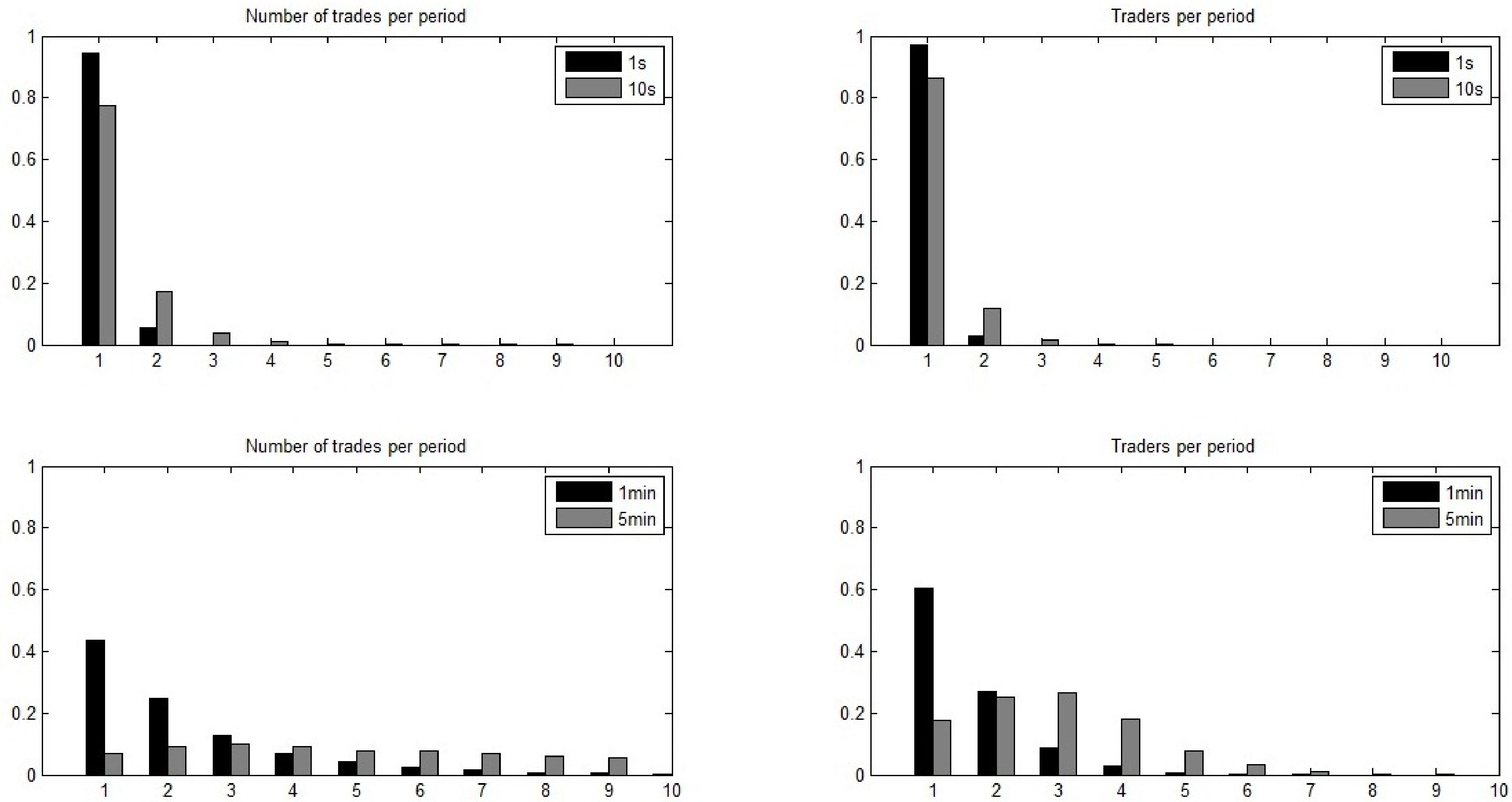

Figure 2: Frequency distributions of the number of IB trades and the number of IB firms present within various time intervals. All the distributions are conditional on there being at least one IB trade in a given interval. The time intervals range from one second to five minutes. The results are based on data using all LSE-executed transactions in the 20 largest, by market capitalization, FTSE 100 shares, covering the sample period between September 1st and December 31st 2012.