UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 680

Interactions among High-Frequency Traders

Evangelos Benos, James Brugler, Erik Hjalmarsson, and Filip Zikes

December 2016

EFMD
EQUIS
ACCREDITED

Interactions among High-Frequency Traders

Evangelos Benos   James Brugler   Erik Hjalmarsson

Filip Zikes[*]

December 12, 2016

Abstract

Using unique transactions data for individual high-frequency trading (HFT) firms in the U.K. equity market, we examine the extent to which the trading activity of individual HFT firms is correlated with each other and the impact on price efficiency. We find that HFT order flow, net positions, and total volume exhibit significantly higher commonality than those of a comparison group of investment banks. 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.

Keywords: High-Frequency Trading; Correlated Trading Strategies; Price Discovery.

JEL Classification: G10, G12, G14.

I   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 stability. Recent events, such as the "flash crashes" in U.S. equity markets on May 6, 2010 and U.S. Treasury markets on October 15, 2014, have highlighted such worries. 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] The current paper aims to shed light on the ways in which individual HFTs interact with each other and assess the effect of this interaction on price efficiency.

The main purpose of our analysis is to better understand the extent to which a given HFT firm tends to trade in a similar manner and direction as its high-frequency competitors. This speaks towards the greater question of whether HFTs might be 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. Whether such correlations among HFTs played an important role in recent flash crashes is not clear, but is certainly a relevant concern. A clear example—albeit from outside the domain of HFT—of the possible negative impact of highly correlated strategies among

---

[1]Algorithmic trading refers to any automated trading where computers directly interact with electronic trading platforms; HFT is therefore a subset of AT. Given the focus of the current paper, in the subsequent discussion we mostly refer to HFT, 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).

a large segment of market participants is provided by the "Quant Meltdown" in August 2007. During this episode, many long-short equity funds pursuing similar strategies suffered major losses and quickly unwound their strategies amid great market turmoil (Khandani and Lo, 2011).

Our data document transactions for the stocks in the U.K. 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 U.K. Financial Conduct Authority (FCA); our sample spans four months, from September 1 through December 31, 2012. The data explicitly identify the submitter of each trade report along with other detailed information such as volume, execution price, and time stamp. We focus 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. IBs clearly engage in a wide variety of trading activities. Although these activities might also involve high-frequency strategies, the overall activities of investment banks are clearly quite distinct from those of pure HFT firms. We therefore view IBs as a relevant comparison group, proxying for the behavior 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 stock in our sample, we formulate a VAR with trading activity in all 10 HFTs and all 10 IBs as dependent variables.[4] Trading activity is measured either as (i) order flow (buyer-initiated volume minus seller-initiated volume), (ii) total transacted volume, or (iii) change-in-inventory (i.e., change in net position). The VAR is formulated in "trade" time (or "tick" time), such that

---

[4]Trading activity by other market participants, not included in these two groups of HFTs and IBs, is modelled in a final separate equation in the VAR. The VAR thus forms a complete system of all trading activity, represented by the HFTs, IBs, and the "residual" market.

the time index only changes when there is a trade event, and is estimated by pooling data from all stocks, yielding a set of interpretable results. The tick-time formulation avoids any temporal aggregation of the data and arguably provides the cleanest way of estimating the relationship between a given trading activity and subsequent trades.

The main empirical results from the VAR can be summarized as follows. In a lead-lag (Granger causality) sense, HFT trading activity tends to be strongly positively related across firms, both for directional and non-directional measures of activity (i.e., both order flow and total volume). In particular, aggressive buying (selling) by an HFT is associated with subsequent additional aggressive buying (selling) by other HFTs. Similarly, changes in inventory for HFTs are also positively related, such that accumulation (reduction) of inventory in a given stock by a given HFT tends to be followed by an accumulation (reduction) of inventory in that same stock by other HFTs. For IBs, we find little evidence of such lead-lag relationships for either order flow or total volume. Changes in inventory for IBs, however, are strongly *negatively* related across IB firms, and also negatively related to changes in HFT inventory, suggesting that IBs tend to absorb inventory from each other as well as from HFTs.

The VAR results thus suggest that HFTs do exhibit commonality in their trading behavior, especially relative to what is observed for IBs. One possible interpretation of this result 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 the model of Jarrow and Protter (2012). An alternative interpretation is that HFTs use strategies that are uniformly more efficient in receiving, processing, and trading on information when 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

4

regression. The key finding is that HFT correlation is associated with a permanent price impact, whereas IB correlation tends to be associated with price reversals. This is consistent with HFT commonality being the result of informed trading and thus contributing to price discovery, along the lines of Martinez and Rosu (2013). Specifically, our analysis suggests that the times when HFTs exhibit commonality in their behavior, are in fact times when they each possesses some (correlated) "private" information and act as informed traders. Correlation in trading activity among HFTs might therefore, at least partly, be driven by correlations in their private information signals. This result expands upon previous findings that HFTs, on average, tend to act as informed traders and trade in the direction of permanent price changes (e.g., Carrion, 2013, and Brogaard, Hendershott, and Riordan, 2014).

The rest of the paper is organized as follows. Section II provides a brief literature review and Section III describes the data and some summary statistics. Section IV introduces the VAR specification and presents the results on interactions across HFT firms. Section V studies whether these correlation patterns appear to have any impact on market quality, and Section VI concludes.

II   Related Literature

Automated HFT 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 HFT 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 implications of correlation among HFTs' trading strategies is not unambiguous, however, and depends on the underlying reasons behind it. If the 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).

Overall, 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 its potential impact on price discovery. Most previous studies have been restricted to using aggregate measures of HFT or AT participation and have focused more on the speed aspects of computer-based trading, and less on the "cross-sectional" aspects.[5] A concurrent study by Boehmer, Li, and Saar (2016)

---

[5]Benos and Sagade (2016), 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 (2016) study the importance of co-location across HFT firms, and Brogaard, Garriott, and Pomeranets (2014) analyze entry and competition among HFT firms.

also analyzes correlations across HFTs, although their focus is very distinct from ours. Their main finding is that increased correlation among HFT strategies is associated with lower stock volatility and that this effect likely stems from more efficient market making on behalf of HFTs. Their overall conclusions are thus in line with ours, namely, that there is a fair degree of correlation among HFTs but that this correlation appears beneficial rather than detrimental to the market. Anand and Venkataraman (2016) study correlations among (high-frequency) market makers on the Toronto Stock Exchange and find significant positive correlation in the liquidity provision across different market makers. Interestingly, the correlation among market makers tends to be higher when volatility is lower, alleviating some regulatory concerns that liquidity is withdrawn *en masse* in stressful times.

III Data and Summary Statistics

A The ZEN Database

Our data consist of reports for trades executed on the electronic order book of the LSE, for all stocks in the FTSE 100 index, over the four months from September 1 to December 31, 2012, a period spanning 80 business days. The transactions data are obtained from the proprietary ZEN database.[6] This database is maintained by the U.K. Financial Conduct Authority (FCA) and consists of trader-submitted transaction reports, which 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 seller in each transaction, as well as whether a given transaction is executed in a principal or agent capacity. We

---

[6]Our data ends on December 31, 2012, although the last trading day we use in our sample is December 21, 2012. We drop the (two) trading days between Christmas and New Years, as these days have an extremely low volume of trade. We focus on stocks that remained in the FTSE 100 index throughout our sample period, and we omit shares with multiple classes trading simultaneously on the London Stock Exchange (e.g., Royal Dutch Shell A-class and B-class shares) due to issues in matching trades between the ZEN and Bloomberg databases for these securities (see discussion below on matching the two databases). This leaves a total of 92 stocks in our sample, which for simplicity we refer to as the FTSE 100 sample.

restrict our analysis to trades executed on the LSE, which accounted for between 55 and 70 percent of the total ("lit") volume for the FTSE 100 shares during our sample period.[7]

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 that are not subject to FCA regulation, and that 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 U.K. exchanges, but that are not FCA-regulated. This group includes the foreign branches of HFT firms that also have a U.K. branch; i.e., the activity of the U.K. branch is captured in ZEN, but the activity of the foreign branch is not. Informal conversations with market regulators suggest that 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—because 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. Although the exact details are confidential, the FCA scheme for identifying HFT firms is based on a number of criteria such as order-submission and trade frequencies, the ratio of orders to

---

[7]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/`).

executed trades, the amount of overnight positions held, the duration of limit orders, the use of proprietary capital, as well as the utilization of latency-reducing technologies. To be classified as an HFT, a firm would have to satisfy several of these criteria. These criteria are also consistent with other schemes used to identify HFTs, such as those in Baron, Brogaard, and Kirilenko (2014), Kirilenko, Kyle, Samadi, and Tuzun (2016), and Korajczyk and Murphy (2016). The resulting data set of HFT activity is very similar to that used by Benos and Sagade (2016).

We also use reports on proprietary trades submitted by the 10 largest investment banks (IBs) to compare and contrast the trading activity of the IBs with that of HFTs.[8] For the remainder of the paper, we refer to both HFTs and IBs as (trading) firms.

Finally, we use quote data from the LSE, obtained via Bloomberg, to reconstruct the top of the order book and to match the ZEN trade reports with 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. We also use Bloomberg transaction data to calculate the total aggregate (market-wide) volume and order flow for each stock. The details of the matching procedure between the Bloomberg and the Zen databases are described in Appendix A.

Importantly, as is detailed in the Appendix, we can be confident that the actual *order* of trades and quotes in our merged data set is accurate. Thus, although our transaction data are time stamped only to the nearest second, we are able to create a complete chronological *ordering* of trades and quote updates. In the subsequent VAR analysis, we make explicit use of this fact, as we estimate the model in trade

---

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

time, rather than calendar time.

B   Variable Definitions

We create a number of variables from the matched ZEN and Bloomberg data. Our measure of trading volume used in the empirical analysis is the number of shares bought or sold within a given time interval (or in a single trade), by a given HFT or IB, in a given stock. 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 summary statistics, we also present the transacted value (in British pounds, GBP) and the number of trades.

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 in which the firm acts as the initiator of the trade (i.e., the firm acts as the market-"taker"), and the "passive" volume is the part of the trading volume in which the firm provides the quote hit by another trader (i.e., the firm acts as the market-"maker"). These volumes will also be referred to as the take- and make-volumes, denoted by $Vlm_{i,s,t}^{take}$ and $Vlm_{i,s,t}^{make}$, respectively. The sum of the aggressive and passive volumes, of course, add up to the total trading volume of each firm.

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

$$OF_{i,s,t} = Vlm_{i,s,t}^{take}\left(Buy\right) - Vlm_{i,s,t}^{take}\left(Sell\right),\tag{1}$$

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.

Finally, the (change in the) net position is defined as the difference between

overall buy volume and overall sell volume,

$$NP_{i,s,t} = Vlm_{i,s,t}\left(Buy\right) - Vlm_{i,s,t}\left(Sell\right). \tag{2}$$

That is, the net position measures the direction of trade, irrespective of whether trading is conducted through make or take orders.

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

$$Vlm_{s,t}^{HFT} = \sum_{i \in HFT} Vlm_{i,s,t}, \tag{3}$$

$$OF_{s,t}^{HFT} = \sum_{i \in HFT} OF_{i,s,t}, \tag{4}$$

and

$$NP_{s,t}^{HFT} = \sum_{i \in HFT} NP_{i,s,t}. \tag{5}$$

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

C   Summary Statistics

We start by briefly summarizing some of the characteristics of the HFT firms in our sample, along with the corresponding statistics for the IB firms. Summary statistics are also shown for all "Other" firms that are not part of the 10 HFTs and 10 IBs used in our main analysis. The "Other" category thus includes market participants such as traditional asset managers, hedge funds, and retail investors.

Table 1 shows summary statistics for daily firm-stock characteristics, including the daily volume (number of shares) and value (in GBP) traded, the number of trades, trade size, the absolute change in net position over the day (measured in

11

GBP), the ratio of net-position change to daily volume (based on GBP values), and the number of times that the estimated inventory crosses zero during the day.[9] Separate statistics are shown for HFTs, IBs, and "Other" firms. The first column in each section (HFT, IB, or Other) shows the mean across all firm-stock-days. For instance, the first row in the first column shows the average number of shares traded across all HFTs and across all stock-day observations. The second column shows the corresponding standard deviation across all firm-stock-days.

[**INSERT TABLE 1 HERE**]

Table 1 shows that, on average, an HFT firm trades about $188,000$ shares and $840,000$ pounds per-stock per-day in the FTSE 100 stocks on the London Stock Exchange. These values are distributed over approximately 145 trades per stock during the day. There is great variation around these averages, however, as seen by the standard deviations. IBs generally trade a bit more heavily than HFTs, trading on average about $289,000$ shares and 1.3 million pounds per-stock per-day, distributed over 215 trades. This is expected as IBs are larger organizations with multiple trading desks that simultaneously execute a variety of strategies. "Other" firms trade considerably less frequently than the large HFTs and IBs in our sample, averaging about 47 trades per day in a given stock. However, when these firms trade, they tend to trade much larger amounts than the HFTs and IBs (41,000 pounds versus around 5,000 pounds for HFTs and IBs).

The final three rows in Table 1 show daily statistics for the (absolute) change in net position over the day (in GBP), the average ratio of this change to the overall volume (in GBP), and the number of times that the inventory of the firm crosses zero during the day.[10] All three measures capture aspects of the notion that HFTs

[9]The inventory calculation are based on the assumption that each firm starts the day with zero inventory for each stock.

[10]When calculating the averages and standard deviations of the daily ratio of position change

take positions over short periods, are reluctant to build up inventory, and do not follow longer-term directional strategies. As is seen, the ratio of the change in inventory to overall traded volume is 16 percent for HFTs, 35 percent for IBs, and 43 percent for Others. The inventory of an HFT crosses zero about seven times per day, whereas for IBs and Others, the corresponding figures are around two and one times per day, respectively.[11]

IV   Interactions among HFTs

We now attempt to pin down the extent of correlation, or dependency, in HFT strategies across different HFT firms. We address this question through the use of trade-time vector autoregressions (VARs), which capture the dependency in trading activity of HFTs within a given stock. That is, we are interested in determining the extent to which current trading by some HFT firm might lead to, or be associated with, subsequent trading by other HFT firms. We run these regressions in trade time (or tick time), where time is updated after each transaction in a given security, rather than after a fixed chronological window, or calendar time.[12] Trade time is arguably a better representation of how HFTs analyze information and formulate strategies compared with clock time (Easley, de Prado, and O'Hara, 2012).

Importantly, the trade time formulation allows for a complete ordering of events. To the extent that no trades occur exactly at the same time, the formula-

to volume, the daily observations are weighted by the number of trades on that day (for the given firm type, i.e., HFT, IB, or Other). This ensures that days where there is more activity, and where the daily ratio of position change to volume is better measured, are given more weight. That is, the (daily) ratio of position change to volume is in itself best viewed as an average across a number of trades and is therefore better measured when there is more trading activity.

[11]Note that the summary statistics reported here differ from those in Benos and Sagade (2016), who use a similar data set sourced from the ZEN database, because the latter report statistics based on trades from four trading venues (the LSE, BATS, Chi-X, and Turquoise) rather than just the LSE. Much of the analysis in Benos and Sagade (2016) is, however, restricted to LSE-only data.

[12]Trade, or event, time model formulations are also used in, for instance, Benos and Sagade (2016) and Brogaard, Henderschott, and Riordan (2014).

tion therefore captures the impact of a given trade on the immediately following trades. Or, put alternatively and without claims of a causal effect, the trade time formulation allows for capturing the immediate, or what one might term the "contemporaneous," association between trading decisions. In particular, the VAR will capture both "correlations" in trading decisions among HFTs, where the trades of several HFTs trading on a similar signal arrive in sequence, as well as "causal" relationships, where the trades of one HFT may trigger the trades of other HFTs.

The VAR specification is used to explore how HFTs react in response to the actions of other HFTs, as well as IBs and the overall market, as explained in more detail below. Formally, we perform a type of Granger causality tests, which, in line with the discussion above, capture both contemporaneous correlations as well as actual causality in the sense that the actions of one trading firm cause a subsequent action by another firm. In the below discussion, we simply interpret the results in terms of lead-lag relationships but with no claim that the effects are truly causal.

As measures of trading activity, we use three related, but distinct, variables: order flow, total volume traded, and change-in-inventory (or equivalently, change-in-net-position during that trade). The unit of all three trade activity measures is number of shares traded.

A A Panel VAR of Stock Trading

Let $HFT_{i,s,t}$ be the trading activity of HFT firm i at trade event t in stock s, and, analogously, let $IB_{i,s,t}$ be the trading activity of IB i at trade event t in stock s. As mentioned above, trading activity is measured either by order flow, total volume, or change-in-inventory, and t is measured in trade time. Below, we sometimes simply refer to t as time. Further, define $\mathbf{HFT}_t^s$ as the vector of stacked trading activity in stock s at event t for all $i = 1, ..., 10$ HFTs and define $\mathbf{IB}_t^s$ as the corresponding

vector of IB trading activity. That is,

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

Also define M_t^s as the residual trading activity in stock s during time t, i.e., the activity of the entire market less the activity of HFTs and IBs. Let $\mathbf{Y}_t^s \equiv (\mathbf{HFT}_t^{s\prime}, \mathbf{IB}_t^{s\prime}, M_t^s)'$ denote the stacked trading activity by both HFT and IB firms, as well as the residual market activity, and formulate the following trade-time VAR for stock s,

$$
\mathbf{Y}_t^s = \mu^s + \sum_{k=1}^{10} A_k \mathbf{Y}_{t-k}^s + \Lambda \mathbf{X}_{t-1}^s + \Psi \mathbf{G}_t + \varepsilon_t^s. \tag{6}
$$

The dependent variable, $\mathbf{Y}_t^s \equiv (\mathbf{HFT}_t^{s\prime}, \mathbf{IB}_t^{s\prime}, M_t^s)'$, is thus a 21×1 vector of trading activity in the 10 HFT firms, the 10 IB firms, and the residual market, during time t. The VAR therefore forms a complete system of all trading activity, represented by the HFTs, IBs, and the "residual" market.[13] μ^s is a 21×1 vector of stock-specific intercepts and A_k, $k = 1, ..., 10$ are 21×21 lag matrix coefficients. We include 10 lags in the VAR, corresponding to the 10 previous trades in that stock.[14] $\mathbf{X}_{t-1}^s$ consists of lagged control variables not modelled in the VAR. In particular, $\mathbf{X}_{t-1}^s$ includes the cumulative return on stock s during the 10 trades prior to the tth ob-

---

[13]For change-in-inventory, the residual market activity is actually a linear combination of the components of $\mathbf{HFT}_t^s$ and $\mathbf{IB}_t^s$, as, by construction, the net inventory change of the entire market $(\sum_i HFT_{i,s,t} + \sum_i IB_{i,s,t} + M_t^s)$ must equal zero for each stock s during each trade t. The residual market, M_t^s, thus can not be included in the VAR when trading activity is measured by changes in inventory. The dependent variable, $\mathbf{Y}_t^s$, therefore reduces to a 20×1 vector in this case, with corresponding adjustments of the coefficient dimensions in the VAR.

[14]As a robustness check, the VAR model was also estimated with 20 lags. The coefficients and test results for the 10-lag VAR are almost identical to those from the 20-lag VAR, indicating that the coefficients for lags 11-20 are mostly indistinguishable from zero. This is also further confirmed by the plots in Figures 1 and 2, which graph coefficients across lags. As is seen, in most cases, by lag 10 the coefficients are very close to zero. In the interest of space, the results for the 20-lag VAR are not reported.

servation, the realized volatility during the 10 trades prior to the tth observation,[15] and the average spread and depth at the best bid and offer in stock s during the 10 trades prior to the tth observation.[16] $\mathbf{G}_t$ includes deterministic functions of time. In particular, $\mathbf{G}_t$ represents linear and quadratic functions of the daily observation number (ranging from 1 to 80) and intra-day dummy variables for each distinct half-hour period within the trading day (i.e., 8:00AM-8:30AM, 8:31AM-9:00AM, and so forth).

The VAR is estimated by pooling data across the full sample of FTSE 100 stocks, allowing for stock-specific intercepts in each equation (μ^s). All other coefficients are pooled across stocks. In total, there are 25,230,628 observations (trades) in the pooled regression, stretching across the 80-day sample period between September 1 and December 31, 2012. Data are sampled during the normal trading hours between 8:00AM and 4:30PM, although activity in the first and last five minutes of each trading day is discarded in order to avoid any beginning or end of day effects. Standard errors and parameter covariance matrices are computed using a non-parametric block bootstrap at the daily level. This method (described in detail in Appendix B) produces consistent estimates of standard errors that are robust to heteroskedasticity as well as any error dependency within each trading day. In particular, the bootstrap approach is robust to cross-sectional dependence across stocks in the panel VAR.

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 lead to (Granger cause) subsequent trading activity by other HFTs in the same stock? (ii) To what extent does trading by an HFT firm in a given stock lead to subsequent trading activity by other market participants in the same stock? (iii) Do we observe similar relationships within and between HFTs and IBs, viewing these two types of traders

---

[15]Realized volatility is defined as the sum of squared mid-quote returns.

[16]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 returns on stock s are defined as the returns over the $t - 10^{th}$ period to the $t - 1^{th}$ period.

as distinct groups? We attempt to test these hypotheses within the above VAR model by mapping the general questions into specific coefficient restrictions. To facilitate the testing of these hypotheses, it is useful to write the VAR in a format where $\mathbf{Y}_t^s$ is written out explicitly. That is, partitioning the coefficient matrices, we can write equation (6) as

$$
\begin{bmatrix}
\mathbf{HFT}_t^s \\
\mathbf{IB}_t^s \\
M_t^s
\end{bmatrix}
= \mu^s + \sum_{k=1}^{10}
\begin{bmatrix}
A_{11,k} & A_{12,k} & A_{13,k} \\
A_{21,k} & A_{22,k} & A_{23,k} \\
A_{31,k} & A_{32,k} & A_{33,k}
\end{bmatrix}
\begin{bmatrix}
\mathbf{HFT}_{t-k}^s \\
\mathbf{IB}_{t-k}^s \\
M_{t-k}^s
\end{bmatrix}
+ \Lambda \mathbf{X}_{t-1}^s + \Psi \mathbf{G}_t + \epsilon_t^s.
$$

$$(7)$$

In equation (7), the parameter sub-matrices $(A_{ij,k})$, $i, j = 1, ..., 3$, now group the coefficients for the HFTs, IBs, and the residual market. $A_{11,k}$ $(A_{22,k})$ correspond to lag-dependencies 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), while the sub-matrix $A_{31,k}$ $(A_{32,k})$ corresponds to the effect of HFTs (IBs) on residual market trading activity. $A_{13,k}$ $(A_{23,k})$ corresponds to the lag-effects of residual market activity on HFTs (IBs).

To test whether lagged trading in other HFTs affects (Granger causes) 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 affects current trading of HFTs by evaluating the null hypothesis that the sum of all the coefficients across all lags in $A_{12,k}$ is equal to zero. In both cases, the null of no Granger causation is rejected if the sum is statistically significantly different from zero. Analogous tests are used to evaluate how a given IB's trading responds to lagged trading by other IBs and lagged HFT trading. 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. Importantly, to the extent that the relationship is significant, the sign of the sum also indicates the direction of the

17

(long-run) relationship; i.e., whether current trading leads to more or less trading in the future.

For the order flow and total volume specifications, we can also test how lagged trading of HFTs or IBs affects trading by the remainder of market participants. Specifically, we can test if increased trading activity of HFTs (IBs) leads to increased trading activity by the remaining firms in the market by testing the null that the sum of the elements of $A_{31,k}$ ($A_{32,k}$) are equal to zero. For completeness, we also test whether increased market activity affects HFTs (IBs) by considering the sum of the elements of $A_{13,k}$ ($A_{23,k}$). Again, the null of no causation is rejected if the sum of these parameters is statistically significantly different from zero.[17]

B   Empirical Results

Table 2 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 as order flow, total trading volume, and change-in-inventory. In each case, the total sum of all the coefficients is given, along with the bootstrapped p-value corresponding to the Wald test of the null hypothesis that the sum is equal to zero, which might be interpreted as a null hypothesis of no (long-run) Granger causality. As mentioned above, the p-values are obtained through a bootstrap procedure, which controls for heteroskedasticity and cross-sectional dependence between stocks (see Appendix B for details).

[INSERT TABLE 2 HERE]

Starting with the results for order flow, the first row of Table 2 shows strong statistical evidence that current trading in a given stock by a given HFT firm is

---

[17]As explained in footnote 13, it is not possible to include residual market activity in the specification with changes in net positions.

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 2 indicates that the current trading direction of a given IB is little affected by the past trades of other IBs. The estimated effect is not significant at the 5 percent level (the p-value is 0.08) and very small in magnitude. Consistent with these findings, row 3 shows that the null hypothesis that the lag effects are identical for HFTs and IBs is strongly rejected.

Rows four to six of Table 2 show that the impact of past HFT order flow on current IB trading is almost identical to the analogous impact of past IB order flow on current HFT activity. In addition, as seen in rows 7 through 9 of Table 2, current trading by the remainder of the market (residual trading) reacts strongly to previous order flows of both HFTs and IBs, although somewhat less to past HFT flows than past IB flows. The final three rows in Table 2 show that neither IBs nor HFTs react much to previous trading by the rest of the market; the coefficients are statistically significant but very small in absolute magnitude.

For order flow, the lead-lag relationship between IBs and HFTs, viewed as two trader groups, is fairly symmetric, with IBs and HFTs each responding similarly to the other group's past trading. Thus, HFTs do not *lead* the trading of IBs to any greater extent than IBs lead the trading of HFTs. Past HFT and IB order flows also tend to lead the rest of the market in a similar way, with, in fact, IBs having a somewhat stronger effect. Hirschey (2016) and Tong (2015) both argue that HFTs anticipate the orders of other investors, whereas van Kervel and Menkveld (2015) find evidence to the contrary. Our findings mostly concern the relative aspects of HFTs and IBs, suggesting that in terms of lead-lag relationships with each other and with the rest of the market, HFTs and IBs are quite similar.

Figures 1 and 2 graphically display some of the relationships emerging from the VAR model, on a lag-by-lag basis. In particular, Figure 1 shows the total

response of HFTs (IBs) to the trading activity of other HFTs (IBs). Figure 2 shows the corresponding responses of HFTs to IB trading, and vice versa. That is, Figures 1 and 2 show the coefficients reported in Table 2 broken down by each lag.[18]

[INSERT FIGURES 1 and 2 HERE]

The graphs tell essentially the same story as the coefficients and test results reported in Table 2. However, breaking down the effects lag-by-lag provides a better idea of how the lead-lag relationships evolve over time. As evident from Graph A in both Figures 1 and 2, which show the order flow results, the majority of the effects are concentrated in the first few lags. Higher-order lag coefficients are typically close to zero and/or not statistically significant (the vertical bars around each lag coefficient indicate 95 percent confidence intervals).

The lag coefficients reported in Figures 1 and 2 also have simple economic interpretations. In particular, each reported coefficient represents the total effect on current trading by all HFT or IB firms from a 100-share trade by each firm in the lagged period. For instance, in the left-hand-side chart of Graph A in Figure 1, the first lag coefficient is around 50. This implies that if each HFT traded 100 shares in the previous period, the current aggregate HFT trading increases by 50 shares, ignoring any effects coming from own past trading by a given HFT. Analogous interpretations apply to the other graphs in Figures 1 and 2.

The results for volume, which are shown in the middle two columns of Table 2, are broadly in line with those obtained using order flow. Total trading volume is not associated with a given direction of trade, and these regressions thus provide a measure of how overall trading activity, rather than trading direction, is related for HFTs, IBs, and the rest of the market. Past trading volume by other HFTs predicts a larger current trading volume for a given HFT (row 1) and for a given

---

[18]The coefficient estimates shown in Table 2 are thus obtained by summing up the coefficients across all lags in the corresponding graph in Figure 1 or 2 (and dividing by 100).

IB (row 5). Past trading volume by IBs does *not* predict a larger current volume for other IBs (row 2), but past IB trading is predictive of future HFT trading (row 4). In contrast to the order flow results, past HFT trading volume does not have a significant impact on the current trading volume by the remainder of the market (row 7), whereas past IB volume is still significant (row 8). Formally, however, we cannot reject that these effects on the remainder of the market are the same for HFTs and IBs (row 9). Graph B in Figures 1 and 2 show the volume results broken down by each lag. In comparison to the order flow results, shown in Graph A in the figures, there is a tendency for the volume effects to be less concentrated to the first few lags. In the case of the response of IB trading to previous trading by other IBs (seen in Figure 1), there is also evidence of an initial negative effect, which is subsequently reversed.

The results for change-in-inventory, or net position, are shown in the final two columns of Table 2 and provide some additional information regarding the interactions among HFTs and IBs. Because change-in-inventory captures both aggressive and passive trading, these regressions highlight the degree to which firms are actually trading with each other (i.e., taking opposite positions over a series of trades). For HFTs, we find that changes in inventory are positively related over time, or, in other words, HFTs tend to accumulate or reduce inventory in a given stock at the same time (row 1). In contrast, for IBs we find that changes in inventory are negatively related and that these firms therefore tend to absorb inventory from each other (row 2). We also find that past HFT inventory accumulation (reduction) is associated with reduction (accumulation) in IB inventory, providing further evidence that HFTs do not appear to front-run IBs (row 5). Graph C in Figures 1 and 2 shows the lag-by-lag results. These highlight, in particular, the strong negative lag effect for IBs (Figure 1), which persists over many lags. As mentioned previously (see footnote 13), the residual market-wide change-in-inventory is a linear combination of the change-in-inventory of the HFTs and IBs. As such, we cannot include the M_t^s variable in the panel-VAR for this measure of trading activity.

21

These change-in-inventory results might also help explain, or further elaborate on, some recent findings by Korajczyk and Murphy (2016) and van Kervel and Menkveld (2015). The essential finding in both of these studies is that when large traders begin a sequence of trades (i.e., a split-up of a large buy or sell order), HFTs initially act as liquidity providers by trading in the opposite direction of the large trade. However, after a while (around 15 minutes in Korajczyk and Murphy and 2 hours in van Kervel and Menkveld), the HFTs learn of the trade sequence and instead start trading in the same direction as the large trader. This switch in trade direction by HFTs leads to substantially higher trading costs during this part of the trade. If HFTs all tend to trade in the same direction, it suggests that it might be hard to find a (market making) HFT to accommodate your trade if your trade is in the "wrong" direction. That is, liquidity would either be plentiful, as all HFTs are willing to trade with you, or dry up, as they all want to trade in the same direction as you. This could explain the rather drastic increase in execution costs as HFTs switch direction a bit into a large order. IBs, on the other hand, have less of a systematic direction as a group.

In summary, the VAR results suggest that the lead-lag dependencies in trading activity between HFT firms appear to be considerably stronger and more significant than for IB firms. This is true when activity is measured either by order flow or overall volume. When looking at changes in inventory, we find that HFTs tend to be positively related, whereas IBs are strongly negatively related. HFTs thus have a tendency to act coherently as a group, jointly building up or decreasing their overall position in a stock. IBs on the other hand, appear to trade more with each other, such that a decrease in net position for some firm is associated with a subsequent increase by another firm. These results are also consistent with those of Chaboud, Chiquoine, Hjalmarsson, and Vega (2014), who find that HFTs (or ATs more generally) tend to trade relatively less with each other in the foreign exchange market.

V Price Impact of Correlated HFTs

Given the evidence on correlated trading activity among HFTs, we continue the 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.

To capture the extent of correlated trading by HFTs and IBs, 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

$$CorrTrading_{s,t}^{HFT} = N\left(Buy\right)_{s,t}^{HFT} - \frac{N\left(Buy\right)_{s,t}^{HFT} + N\left(Sell\right)_{s,t}^{HFT}}{2}, \qquad (8)$$

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 (8) effectively calculates the number of excess aggressive buyers or sellers at any given 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 FTSE 100 shares using minute-by-minute data. The one-minute sampling frequency 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.[19] At the same time, the sampling frequency still 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 five-minute frequency.

To measure the contemporaneous and lagged price impact associated with correlated trading, we regress one-minute returns on contemporaneous and lagged order flows, the correlated trading metrics and their lags, as well as interactions of the two. Because both order flows and the correlated trading metrics can take on both positive and negative values, a negative order flow and a negative trade correlation would result in a positive interaction term. To avoid this canceling out of signs, the order flows are instead interacted with the absolute values of the

---

[19]Clearly, one could not conduct this analysis with tick-by-tick observations, as in the VAR model, since then one would end up with just one trade in a given time period.

trade-correlation metrics. Thus, our full specification takes the form

$$
\begin{aligned}
R_{s,t} \; = \; & \alpha_s + \sum_{i=0}^{5} \beta_{OF,i}^{HFT} OF_{s,t-i}^{HFT} + \sum_{i=0}^{5} \beta_{OF,i}^{IB} OF_{s,t-i}^{IB} + \sum_{i=1}^{5} \beta_{OF,i}^{Res} OF_{s,t-i}^{Res} \\
& + \sum_{i=0}^{5} \beta_{Corr,i}^{HFT} CorrTrading_{s,t-i}^{HFT} + \sum_{i=0}^{5} \beta_{Corr,i}^{IB} CorrTrading_{s,t-i}^{IB} \\
& + \sum_{i=0}^{5} \beta_{|Corr|,i}^{HFT} |CorrTrading|_{s,t-i}^{HFT} + \sum_{i=0}^{5} \beta_{|Corr|,i}^{IB} |CorrTrading|_{s,t-i}^{IB} \\
& + \sum_{i=0}^{5} \beta_{OF\times|Corr|,i}^{HFT} \left(OF_{s,t-i}^{HFT} \times |CorrTrading|_{s,t-i}^{HFT} \right) \\
& + \sum_{i=0}^{5} \beta_{OF\times|Corr|,i}^{IB} \left(OF_{s,t-i}^{IB} \times |CorrTrading|_{s,t-i}^{IB} \right) + u_{s,t}. \qquad (9)
\end{aligned}
$$

Here $R_{s,t}$ is the one-minute return of stock s in period t, and $OF_{s,t}^{HFT}, OF_{s,t}^{IB}$, and $OF_{s,t}^{Res}$ are the order flows from HFTs, IBs, and the remainder of the market (the "residual" order flow). $CorrTrading_{s,t}^{HFT}$ and $CorrTrading_{s,t}^{IB}$ are the correlation metrics for HFTs and IBs defined in equation (8). To ensure that the interaction terms do indeed capture the interacting effects between order flows and absolute trade correlations, the absolute trade-correlation metrics also enter into the regression separately.

The main coefficients of interest in equation (9) are those in front of the HFT and IB trade-correlation metrics. In particular, we are interested in whether correlated trading among HFTs (or IBs) is associated with an "extra" price impact, over and above the price impact of order flow, and whether that additional price impact is subsequently reversed or not. That is, keeping HFT order flow constant, does shifting the degree of trade correlation among HFTs alter the overall price impact? The coefficient on the HFT trade-correlation metric, controlling for order flow, answers this question.[20]

---

[20]The coefficients on the *interactions* between order flows and the (absolute) trade-correlation metrics measures whether this "extra" price impact becomes more or less pronounced in periods when order flow is large.

The model is estimated by least squares, pooling the data across all 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 by the standard deviation of the total order flow for that stock (i.e., the sum of the HFT, IB, and Residual order flows). The returns on the left-hand side of the regressions are standardized by their own standard deviation at the stock level.[21] Prior to being interacted, the order flows and the absolute correlated trading metrics are de-meaned (at the stock level) such that the main coefficients in all regressions are reported at the sample mean and thus are comparable across the specifications with and without the interaction terms. That is, the total effect of HFT (IB) order flow, evaluated at the sample mean of (absolute) correlated trading, is therefore simply given by the coefficient on the HFT (IB) order flow, enabling a direct comparison of the order flow coefficients in the specifications with and without interaction terms. As stock-specific intercepts (i.e., fixed effects) are included in the regressions, this demeaning does not in any way alter the regression specifications, but merely allows for an easier interpretation of the coefficients.

Summary statistics of the (non-standardized) returns, order flow, and trade-correlation variables are presented in Table 3, along with the correlation matrix for these variables. The correlations between order flows and the trade-correlation metrics are around 0.25 for both HFTs and IBs. Thus, although positively related, the order flows and trade-correlation metrics are clearly distinct activity measures.

[INSERT TABLE 3 HERE]

Table 4 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 one-minute returns on contemporaneous and

---

[21]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.

lagged total order flow; the total order flow is denoted by $OF_{s,t}^{Tot}$ in the table, and is defined as $OF_{s,t}^{Tot} \equiv OF_{s,t}^{HFT} + OF_{s,t}^{IB} + OF_{s,t}^{Res}$.[22] 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 part of the contemporaneous price impact tends to be subsequently reversed.

[INSERT TABLE 4 HERE]

In column 2 of Table 4, HFT, IB, and residual order flows enter separately into the regression. The results are qualitatively the same as in the specification with total order flow. That is, there 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 of the market.

We next add our metrics of correlated trading to the regressions. The estimation results are reported in column 3 of Table 4. The contemporaneous price impact coefficients for HFTs' and IBs' correlated trading are both positive and significant, although the IB coefficient is larger in magnitude. Most importantly, however, the coefficient on the lagged trade-correlation metric for HFTs is positive (and small in magnitude), whereas the coefficient on lagged trade correlation for IBs is negative (and large in magnitude). That is, keeping order flow fixed, the impact of HFTs' correlated trading is not subsequently reversed, unlike for IBs. Put differently, the results show that correlated HFT trading mitigates the reversal effect of lagged order flow, whereas correlated IB trading exacerbates the reversal effect. These observed differences in the point estimates for HFTs and IBs are also statistically significant, as evident from the formal Wald tests reported toward the bottom of the table. The regression results thus suggest that HFTs' correlated

---

[22]This regression can viewed as a restricted version of equation (9), 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.

trading is informed, leading to a permanent price impact.

Finally, the interactions between the order flows and the absolute values of the correlated trading metrics are included in the regression. The results are reported in column 4 of Table 4. The contemporaneous interaction terms are negative and statistically significant for both HFTs and IBs, indicating that a shift in trade correlation has a larger effect when order flow is closer to its mean.[23] The lagged interactions are negative for HFTs and positive for IBs, although the estimated coefficients are fairly small in magnitude. The coefficients for the (non-interacted) trade-correlation metrics remain virtually identical after including the interaction terms, and inclusion of the interactions does not alter the main conclusions.

To get a sense of the economic magnitude of the estimated effects, recall first that the returns on the left-hand side of the regression are standardized to have a unit standard deviation. The normalized HFT order flow has a standard deviation of around 0.5,[24] and a one-standard-deviation HFT order flow shock is thus associated with a 0.2-standard-deviation shock to returns ($\beta_{OF,0}^{HFT} \times 0.5 \approx 0.4 \times 0.5$), keeping all else constant. A unit shift in the HFT trade-correlation metric would similarly lead to a 0.2-standard-deviation move in returns ($\beta_{Corr,0}^{HFT} \approx 0.2$). Most interestingly, perhaps, the final specification in column 4 of Table 4 shows that the effect of correlated trading might "cancel out" the reversal effect of past order flow, such that the overall effect on returns of past order imbalances and correlated trading is positive, highlighting the likely informed nature of correlated HFT trading.[25]

---

[23]For a given value of HFT order flow, the impact of a unit increase in HFT trade correlation is given by $\beta_{Corr,0}^{HFT} + \beta_{|Corr|,0}^{HFT} + \beta_{OF \times |Corr|,0}^{HFT} \times OF_{s,t}^{HFT} \approx 0.2 - 0.07 \times OF_{s,t}^{HFT}$, where $OF_{s,t}^{HFT}$ is measured in deviations from the mean. The total effect evaluated at the mean of HFT order flow is thus simply equal to 0.2. If HFT order flow is above the mean, the total impact clearly decreases. The same reasoning would apply to a negative shift in trade correlation, provided that order flow in that case is also assumed to be below its mean.

[24]The HFT, IB, and residual order flows are normalized by the standard deviation of the *total* order flow for each stock. Each of these normalized order flows will therefore have a standard deviation less than unity.

[25]Keeping all else constant, the estimated total price impact of a unit shock to lagged correlated HFT trading and a 1-standard-deviation (≈ 0.5) shock to lagged HFT order flow, is given by $-0.023 OF_{s,t-1}^{HFT} + 0.026 CorrTrading_{s,t-1}^{HFT} + 0.010 |CorrTrading|_{s,t-1}^{HFT} - 0.036(OF_{s,t-1}^{HFT} \times$

As a robustness check, we also estimate the same regressions as above using data sampled every five minutes. That is, five-minute returns are now regressed on the order flow and trade-correlation variables constructed over five-minute intervals. However, in order to keep the temporal span of the lags identical to the one-minute specification, only one lag is now included. Otherwise, the two specifications are identical, and the results are presented in Table 5. The results in Table 5 strongly echo those seen in Table 4. The statistical significance of some of the estimates based on the five-minute data is somewhat weaker than in the one-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.

[INSERT TABLE 5 HERE]

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. Previous studies, including Carrion (2013) and Brogaard, Hendershott, and Riordan (2014), have also documented that HFTs tend to contribute to price efficiency by trading (aggressively) in the direction of permanent price changes and in the opposite direction of transitory price changes. Such findings are consistent with HFTs acting as informed traders (e.g., Kyle, 1985). Our results add to these previous findings by showing that periods when the trading activity of HFTs is correlated tend to be periods when HFTs possess private information, i.e., act as informed traders. The correlation in trading activity would thus appear to be the result of correlations in "private" information. The findings here also contribute to the view that the private information held by HFTs appears to be relevant over horizons that stretch for at least

$|CorrTrading|_{s,t-1}^{HFT}) = -0.023 \times 0.5 + 0.026 + 0.01 - 0.036 \times 0.5 = 0.0065.$

a few minutes, and not just over shorter intervals of a few seconds.[26]

VI Conclusion

Using a unique data set of 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 firms' trading activities are positively related at high frequencies. This is true both for overall trading volume, as well as for directional measures of trading such as order flow and changes in net position. In contrast, when performing the same analysis for a group of investment banks, we find that order flow is much more weakly related across the banks, whereas changes in net positions are, in fact, strongly negatively related. The results for net positions, in particular, highlight that HFT firms have a tendency to all trade in the same direction at the time, whereas investment banks instead tend to trade more disparately and absorb each others' changes in inventory.

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

---

[26]Brogaard, Hendershott, and Riordan (2014) document that HFTs appear able to predict price movements over the next few seconds, whereas Carrion (2013) find that HFTs also appear to have the ability to time the market over longer (greater than five minutes) intra-daily intervals. In conjunction with our results, this suggests that HFTs likely possess the ability to time the market at a range of different intra-daily horizons.

to be collectively buying and selling at the "right" time, and correlations in their trading activity appear to, at least partly, be driven by correlations in their private information signals.

In summary, our study finds strong support for the notion that the strategies of HFT firms tend to be correlated with each other. However, our results also suggest that such correlations are not destabilizing for the market, but rather reflect HFT firms trading on the same (correct) information.

Appendix A. Matching the ZEN and Bloomberg data

The Bloomberg data set is time stamped to the nearest second and contains both trade and quote information. In addition to the one-second time stamp, these data also contain a variable indicating the chronological order of all events of either kind (trades or quote changes). We can therefore exactly match the trade and quote information within the Bloomberg data set, creating an exactly *ordered* trade and quote data set.

The trade data in ZEN and Bloomberg are subsequently matched on multiple criteria (execution price, trade size, as well as time to the nearest second) using the fact that we also observe trade information in Bloomberg. By matching on time stamp, as well as trade size and execution price, we are able to almost perfectly match the Bloomberg trades (and, by implication, the Bloomberg quotes) to the ZEN trade information, with an excess of 99% definitive matches for the trades in ZEN. The remaining less than 1 percent of trades that could not be matched are dropped from the analysis.

Since the Bloomberg data provide a correct chronological ordering of both the trades and the quotes, we can also be confident that the actual *order* of trades and quotes in our final merged data set is accurate. Thus, although our transaction data are time stamped only to the nearest second, we are able to create an exact *ordering* of the trades.

Our matching scheme therefore also alleviates most of the concerns raised in the literature on accurately matching trades and quotes in order to classify trade direction (e.g., Easley, Lopez de Prado, and O'Hara, 2012, Chakrabarty, Pascual, and Shkilko, 2014, and Holden and Jacobsen, 2014). This is typically a problem in many data sets because trades and quotes observed at coarse time intervals are either not individually sequenced or not sequenced against each other (i.e., trades versus quotes) within each time interval. Our procedure still suffers from limita-

tions in the Lee and Ready trade signing algorithm, but most studies suggest that this approach works very well provided quotes and trades are correctly matched (e.g., Carrion and Kolay, 2014).

Appendix B. The Bootstrap Procedure

Standard errors and parameter covariance matrices for the VAR in Section A are computed using a non-parametric non-overlapping block bootstrap, where blocks are defined by trading days in our sample. This method re-samples with replacement from the 80 trading days in our data and constructs a bootstrap sample containing all trades from the re-sampled days for each bootstrap run, $b = 1, ..., B$. For each bootstrapped run, b, we construct the data matrices $\left\{ \mathbf{Y}_t^{s,b}, \mathbf{X}_{t-1}^{s,b}, \mathbf{G}_t^b \right\}$ for each stock s and then estimate the pooled parameters $\left\{ A_1^b, ..., A_{10}^b, \Lambda^b, \Psi^b \right\}$ and the stock-specific intercepts $\left\{ \mu^{s,b} \right\}$. From the B bootstrapped parameter estimates, we directly estimate the covariance matrix of the parameters of the VARs. These covariance matrices are used for the hypothesis tests in Table 2. We set B equal to 100 for each regression. This number of replications is chosen for reasons of computational feasibility.

This method of estimating parameter standard errors and covariances has a number of advantages. First and most important, these estimates are robust to both arbitrary heteroskedasticity as well as arbitrary error correlations within a given trading day. In other words, we treat any error terms within a trading day as potentially dependent and compute standard errors that are robust to both these issues (the bootstrap is robust to heteroskedasticity, regardless of the block structure). While the lag structure of the VAR should account for serial correlation in the error terms of each equation, it is still possible that the error terms across stocks are correlated for trades occurring relatively close to one another in time. Second, the non-parametric bootstrap is based on less restrictive assumptions than alternatives such as the parametric bootstrap (Hall, 2005) or usual parametric

VAR standard errors. Third, the nature of exchange trading, whereby continuous trading occurs for only 8.5 hours of each day before ceasing for the next 15.5 hours, imparts a natural block structure for the dependence in our data that we directly model; error terms within blocks (trades on the same days) are plausibly correlated while error terms across blocks (i.e., trades occurring on separate days) are plausibly independent. Lastly, the number of blocks naturally increases with the data, a requirement for consistency (Kreiss and Lahiri, 2012).

References

Anand, A., and K. Venkataraman. "Market Conditions, Fragility, and the Economics of Market Making." *Journal of Financial Economics* 121 (2016), 327-349.

Baron, M., J. Brogaard, and A. Kirilenko. "Risk and Return in High Frequency Trading." Working Paper (2014).

Benos, E., and S. Sagade. "Price Discovery and the Cross-section of High-Frequency Trading." *Journal of Financial Markets* 30 (2016), 54-77.

Boehmer, E., D. Li, and G. Saar. "Correlated High-Frequency Trading." Working Paper (2016).

Brogaard, J., C. Garriott, and A. Pomeranets. "High-Frequency Trading Competition" Working Paper, Bank of Canada (2014).

Brogaard, J., B. Hagströmer, L. Nordén, and R. Riordan. "Trading Fast and Slow: Colocation and Liquidity.", forthcoming, *Review of Financial Studies* (2016)

Brogaard, J., T. Hendershott, and R. Riordan. "High Frequency Trading and Price Discovery." *Review of Financial Studies* 27 (2014), 2267-2306.

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

Carrion, A., and M. Kolay. "Trade Signing in Fast Markets.", Working Paper (2014).

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

Chakrabarty, B., R. Pascual, and A. Shkilko. "Evaluating Trade Classification Algorithms: Bulk Volume Classification versus the Tick Rule and the Lee-Ready Algorithm." Working Paper (2014).

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

Driscoll, J.C., and A.C. Kraay. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *Review of Economics and Statistics* 80 (1998), 549-560.

Easley, D., M.M. Lopéz de Prado, and M. O'Hara. "The Volume Clock: Insights into the High Frequency Paradigm" *Journal of Portfolio Management* 39 (2012), 19-29.

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

Hagströmer, B., and L. Nordén. "The Diversity of High-Frequency Traders." *Journal of Financial Markets* 16 (2013), 741-770.

Hagströmer, B., L. Nordén, and D. Zhang. "The Aggressiveness of High-Frequency Traders." *The Financial Review* 49 (2014), 395-419.

Haldane, A. "The Race to Zero.", Speech, International Economic Association Sixteenth World Congress, Beijing (2011); available at: `https://core.ac.uk/download/pdf/24060887.pdf`.

Hall, A.R. "Generalized Method of Moments.", Oxford University Press, Oxford, U.K. (2005).

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

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

Hirschey, N. "Do High-Frequency Traders Anticipate Buying and Selling Pressure?" Working Paper (2016).

Holden, C.W., and S. Jacobsen. "Liquidity Measurement Problems in Fast Competitive Markets: Expensive and Cheap Solutions." *Journal of Finance* 69 (2014), 1747-1885.

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

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

Kirilenko, A., A. Kyle, M. Samadi, and T. Tuzun. "The Flash Crash: High Frequency Trading in an Electronic Market.", forthcoming, *Journal of Finance* (2016)

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

Korajczyk, R.A., and D. Murphy. "High-Frequency Market Making to Large Institutional Trades." Working Paper (2016).

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

Kreiss, J.P., and S.N. Lahiri. "Bootstrap Methods for Time Series." In *Handbook of Statistics*, Vol. 30, Time Series Analysis: Methods and Applications, T.S. Rao, S.S. Rao, and C.R. Rao, eds., Oxford, U.K.: Elsevier North-Holland (2012).

Kyle, A. "Continuous Auctions and Insider Trading." *Econometrica* 53 (1985), 1315-1336.

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

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

Martinez, V., and I. Rosu. "High Frequency Traders, News and Volatility." Working Paper (2013).

Oehmke, M. "Gradual Arbitrage.", Working Paper, Columbia University (2009).

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

Tong, L. "A Blessing or a Curse? The Impact of High Frequency Trading on Institutional Investors." Working Paper (2015).

van Kervel, V., and A.J. Menkveld. "High-Frequency Trading around Large Institutional Orders." Working Paper (2015).

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

Table 1: Summary Statistics.

The table reports means and standard deviations of various measures of trade activity for HFTs, IBs, and all "Other" firm activity based on data pooled across firm-stock-days. The means and standard deviations are calculated over all firm-stock-days with at least one trade. The sample includes all limit order book trades in FTSE 100 stocks on the LSE from September 1 to December 31, 2012. Specifically, the summary statistics are based on data for the 92 stocks that remained in the FTSE 100 index throughout the sample period and that did not trade with multiple classes simultaneously on the London Stock Exchange.

	HFT		IB		Other	
	Mean	Std.	Mean	Std.	Mean	Std.
Volume (Shares '000)	188.13	1,108.36	288.70	976.94	122.23	1,016.99
Volume (GBP '000)	842.5	1,790.2	1,344.5	2,393.5	482.6	2,543.2
Number of trades per day	145.4	251.0	215.3	323.7	47.1	198.0
Trade size (GBP)	5,455	7,207	5,527	5,297	41,311	651,667
Absolute daily position change (GBP '000)	133.3	349.6	455.6	917.2	198.9	1,269.4
Ratio of absolute position change to volume (GBP/GBP)	0.16	0.23	0.35	0.27	0.43	0.36
Zero-inventory crossings per day	7.28	17.52	2.21	2.90	1.29	24.97

Table 2: VAR Results.

The table reports results from the trade-time panel-VAR model specified in equation (6), using pooled data across all stocks. Coefficient estimates and p-values for hypotheses tests regarding HFT and IB activity are shown. The first two columns list a description of the tested hypothesis and the corresponding formal coefficient restrictions, respectively. Separate results for trading activity measured as order flow, traded volume, and change-in-inventory are shown. All variables are sampled in trade time, with each trade contributing one time period to the sample. The sample includes all limit order book trades in FTSE 100 stocks on the LSE from September 1 to December 31, 2012. Specifically, the analysis uses data for the 92 stocks that remained in the FTSE 100 index throughout the sample period and that did not trade with multiple classes simultaneously on the London Stock Exchange. In total, there are 25,230,628 observations. The p-values are obtained from the bootstrap procedure described in Appendix A and are robust to heteroskedasticity and cross-sectional dependence.

Hypothesis to be tested	Coefficient Restriction	Order flow		Total volume		Change-in-inventory	
		Coefficients	p-value	Coefficients	p-value	Coefficients	p-value
1. Are HFTs correlated within stocks?	$\sum_k \sum_i \sum_{j\neq i} \left(A_{11,k}\right)_{i,j} = 0$	0.7529	0.000	0.3030	0.000	0.2623	0.000
2. Are IBs correlated within stocks?	$\sum_k \sum_i \sum_{j\neq i} \left(A_{22,k}\right)_{i,j} = 0$	0.0506	0.082	0.0192	0.629	-0.8927	0.000
3. Are HFTs more correlated than IBs?	$\sum_k \sum_i \sum_{j\neq i} \left(A_{11,k}\right)_{i,j} - \sum_k \sum_i \sum_{j\neq i} \left(A_{22,k}\right)_{i,j} = 0$	0.7023	0.000	0.2838	0.000	1.1550	0.000
4. Do HFTs respond to IBs?	$\sum_k \sum_i \sum_{j\neq i} \left(A_{12,k}\right)_{i,j} = 0$	0.3059	0.000	0.1886	0.000	-0.1868	0.000
5. Do IBs respond to HFTs?	$\sum_k \sum_i \sum_j \left(A_{21,k}\right)_{i,j} = 0$	0.2848	0.001	0.3843	0.000	-0.4607	0.000
6. Do HFTs respond differently to IBs than IBs do to HFTs?	$\sum_k \sum_i \sum_j \left(A_{12,k}\right)_{i,j} - \sum_k \sum_i \sum_j \left(A_{21,k}\right)_{i,j} = 0$	0.0210	0.802	-0.1956	0.016	0.2740	0.000
7. Does residual trading respond to HFTs?	$\sum_k \sum_i \sum_j \left(A_{31,k}\right)_{i,j} = 0$	2.0014	0.000	0.1153	0.419	-	-
8. Does residual trading respond to IBs?	$\sum_k \sum_i \sum_{j\neq i} \left(A_{32,k}\right)_{i,j} = 0$	2.5581	0.000	0.3478	0.000	-	-
9. Does residual trading respond differently to HFTs than to IBs?	$\sum_k \sum_i \sum_j \left(A_{31,k}\right)_{i,j} - \sum_k \sum_i \sum_j \left(A_{32,k}\right)_{i,j} = 0$	-0.5568	0.020	-0.2325	0.108	-	-
10. Do HFTs respond to residual trading?	$\sum_k \sum_i \sum_j \left(A_{13,k}\right)_{i,j} = 0$	0.0305	0.000	0.0208	0.000	-	-
11. Do IBs respond to residual trading?	$\sum_k \sum_i \sum_{j\neq i} \left(A_{23,k}\right)_{i,j} = 0$	0.0555	0.000	0.0627	0.000	-	-
12. Do HFTs respond differently to residual trading than IBs?	$\sum_k \sum_i \sum_j \left(A_{13,k}\right)_{i,j} - \sum_k \sum_i \sum_j \left(A_{23,k}\right)_{i,j} = 0$	-0.0249	0.000	-0.0419	0.000	-	-

Table 3: Summary Statistics for Returns, Order Flows, and Trade Correlations. The table reports means and standard deviations for returns, HFT, IB, and "Residual" order flows, as well as for the metric of correlated trading for HFTs and IBs. Returns are measured in basis points and order flows in number of shares. The statistics are based on data pooled across firm-stock-days. The lower part of the table reports the correlation matrices for these variables. The statistics are based on data pooled across firm-stock-days, sampled either at the one-minute frequency (left-hand-side columns) or the five-minute frequency (right-hand-side columns). The one-minute and five-minute samples are constructed from all limit order book trades in FTSE 100 stocks on the LSE from September 1 to December 31, 2012. Specifically, the summary statistics are based on data for the 92 stocks that remained in the FTSE 100 index throughout the sample period and that did not trade with multiple classes simultaneously on the London Stock Exchange.

One-minute frequency

	$Returns$	OF^{HFT}	OF^{IB}	OF^{Res}	$CorrTrading^{HFT}$	$CorrTrading^{IB}$
Mean	0.014	39.7	110.8	-139.2	0.000	0.006
Std.	5.50	8,331	12,481	14,570	0.302	0.476
				Correlation matrix		
$Returns$	1.000	0.078	0.130	0.071	0.189	0.293
OF^{HFT}		1.000	0.182	-0.268	0.271	0.070
OF^{IB}			1.000	-0.291	0.057	0.266
OF^{Res}				1.000	-0.025	-0.028
$CorrTrading^{HFT}$					1.000	0.154
$CorrTrading^{IB}$						1.000

Five-minute frequency

	$Returns$	OF^{HFT}	OF^{IB}	OF^{Res}	$CorrTrading^{HFT}$	$CorrTrading^{IB}$
Mean	0.030	91.6	250.2	-312.3	-0.001	0.008
Std.	9.13	12,626	21,736	24,353	0.330	0.558
				Correlation matrix		
$Returns$	1.000	0.048	0.141	0.041	0.110	0.266
OF^{HFT}		1.000	0.193	-0.285	0.203	0.054
OF^{IB}			1.000	-0.323	0.032	0.216
OF^{Res}				1.000	-0.021	-0.032
$CorrTrading^{HFT}$					1.000	0.124
$CorrTrading^{IB}$						1.000

Table 4: Price Impact Regressions using one-minute data.

The table shows regressions of returns on contemporaneous and lagged order flows, correlated trading metrics, absolute correlated trading metrics, and interactions of order flows and the absolute correlated trading metrics. t-statistics are reported in parentheses below the coefficient estimates. The regressions are estimated by least squares, pooling the data across all stocks while allowing for stock-specific intercepts. The order flow variables are normalized, at the stock level, by the standard deviation of the total order flow for that stock, and the returns on the left-hand side of the regressions are normalized by their own standard deviation at the stock level. Prior to being interacted, the order flows and the absolute correlated trading metrics are also de-meaned, such that the main coefficients in all regressions are reported at the sample mean. The results are based on data using all limit order book trades in FTSE 100 stocks on the LSE from September 1 to December 31, 2012. Specifically, the analysis uses data for the 92 stocks that remained in the FTSE 100 index throughout the sample period and that did not trade with multiple classes simultaneously on the London Stock Exchange. In total, there are 3,311,540 one-minute observations. The t-statistics and Wald tests are based on Driscoll and Kraay (1998) standard errors, which are robust to heteroskedasticity, serial correlation, and cross-sectional dependence.

|	(1)	(2)	(3)	(4)
OF_t^{Tot}	0.472			
	(183.8)			
$OF_{t-1,t-5}^{Tot}$	-0.067			
	(-34.72)			
OF_t^{HFT}		0.439	0.351	0.378
		(149.8)	(126.4)	(107.2)
$OF_{t-1,t-5}^{HFT}$		-0.039	-0.040	-0.023
		(-11.71)	(-11.31)	(-5.310)
OF_t^{IB}		0.473	0.384	0.410
		(179.6)	(190.1)	(159.6)
$OF_{t-1,t-5}^{IB}$		-0.075	-0.046	-0.046
		(-37.86)	(-21.03)	(-16.74)
OF_t^{RES}		0.519	0.466	0.457
		(156.9)	(157.1)	(155.9)
$OF_{t-1,t-5}^{RES}$		-0.071	-0.058	-0.058
		(-28.57)	(-23.98)	(-23.94)
$CorrTrading_t^{HFT}$			0.186	0.203
			(45.72)	(47.23)
$CorrTrading_{t-1,t-5}^{HFT}$			0.017	0.026
			(3.024)	(4.389)
$CorrTrading_t^{IB}$			0.310	0.324
			(85.08)	(83.93)
$CorrTrading_{t-1,t-5}^{IB}$			-0.154	-0.155
			(-48.42)	(-48.99)
$|CorrTrading|_t^{HFT}$				0.002
				(0.520)
$|CorrTrading|_{t-1,t-5}^{HFT}$				0.010
				(1.511)
$|CorrTrading|_t^{IB}$				-0.003
				(-1.064)
$|CorrTrading|_{t-1,t-5}^{IB}$				-0.010
				(-2.689)
$OF_t^{HFT} \times |CorrTrading|_t^{HFT}$				-0.071
				(-12.90)
$OF_{t-1,t-5}^{HFT} \times |CorrTrading|_{t-1,t-5}^{HFT}$				-0.036
				(-5.465)
$OF_t^{IB} \times |CorrTrading|_t^{IB}$				-0.046
				(-15.64)
$OF_{t-1,t-5}^{IB} \times |CorrTrading|_{t-1,t-5}^{IB}$				0.002
				(0.625)
Wald test $H_0 : \beta_{Corr,0}^{HFT} = \beta_{Corr,0}^{IB}$ (p-values)			0.000	0.000
Wald test $H_0 : \sum_{i=1}^{5} \beta_{Corr,i}^{HFT} = \sum_{i=1}^{5} \beta_{Corr,i}^{IB}$ (p-values)			0.000	0.000

Table 5: Price Impact Regressions using five-minute data.

The table shows regressions of returns on contemporaneous and lagged order flows, correlated trading metrics, absolute correlated trading metrics, and interactions of order flows and the absolute correlated trading metrics. t-statistics are reported in parentheses below the coefficient estimates. The regressions are estimated by least squares, pooling the data across all stocks while allowing for stock-specific intercepts. The order flow variables are normalized, at the stock level, by the standard deviation of the total order flow for that stock, and the returns on the left-hand side of the regressions are normalized by their own standard deviation at the stock level. Prior to being interacted, the order flows and the absolute correlated trading metrics are also de-meaned, such that the main coefficients in all regressions are reported at the sample mean. The results are based on data using all limit order book trades in FTSE 100 stocks on the LSE from September 1 to December 31, 2012. Specifically, the analysis uses data for the 92 stocks that remained in the FTSE 100 index throughout the sample period and that did not trade with multiple classes simultaneously on the London Stock Exchange. In total, there are 662,308 five-minute observations. The t-statistics and Wald tests are based on Driscoll and Kraay (1998) standard errors, which are robust to heteroskedasticity, serial correlation, and cross-sectional dependence.

	(1)	(2)	(3)	(4)
OF_t^{Tot}	0.428			
	(118.9)			
$OF_{t-1,t-5}^{Tot}$	-0.052			
	(-26.11)			
OF_t^{HFT}		0.311	0.261	0.276
		(39.89)	(29.24)	(28.21)
$OF_{t-1,t-5}^{HFT}$		-0.044	-0.049	-0.053
		(-10.77)	(-11.30)	(-10.39)
OF_t^{IB}		0.462	0.394	0.431
		(96.37)	(86.95)	(79.53)
$OF_{t-1,t-5}^{IB}$		-0.063	-0.057	-0.062
		(-25.01)	(-21.18)	(-18.18)
OF_t^{RES}		0.429	0.389	0.385
		(85.34)	(85.60)	(84.80)
$OF_{t-1,t-5}^{RES}$		-0.056	-0.053	-0.052
		(-19.95)	(-19.58)	(-19.38)
$CorrTrading_t^{HFT}$			0.047	0.054
			(8.647)	(9.849)
$CorrTrading_{t-1,t-5}^{HFT}$			0.010	0.007
			(2.813)	(1.994)
$CorrTrading_t^{IB}$			0.167	0.184
			(52.82)	(57.88)
$CorrTrading_{t-1,t-5}^{IB}$			-0.034	-0.036
			(-15.97)	(-17.22)
$|CorrTrading|_t^{HFT}$				-0.001
				(-0.281)
$|CorrTrading|_{t-1,t-5}^{HFT}$				0.010
				(2.462)
$|CorrTrading|_t^{IB}$				-0.006
				(-1.953)
$|CorrTrading|_{t-1,t-5}^{IB}$				0.002
				(0.929)
$OF_t^{HFT} \times |CorrTrading|_t^{HFT}$				-0.040
				(-3.688)
$OF_{t-1,t-5}^{HFT} \times |CorrTrading|_{t-1,t-5}^{HFT}$				0.014
				(1.853)
$OF_t^{IB} \times |CorrTrading|_t^{IB}$				-0.060
				(-16.57)
$OF_{t-1,t-5}^{IB} \times |CorrTrading|_{t-1,t-5}^{IB}$				0.010
				(3.800)
Wald test $H_0: \beta_{Corr,0}^{HFT} = \beta_{Corr,0}^{IB}$ (p-values)			0.000	0.000
Wald test $H_0: \sum_{i=1}^{5} \beta_{Corr,i}^{HFT} = \sum_{i=1}^{5} \beta_{Corr,i}^{IB}$ (p-values)			0.000	0.000

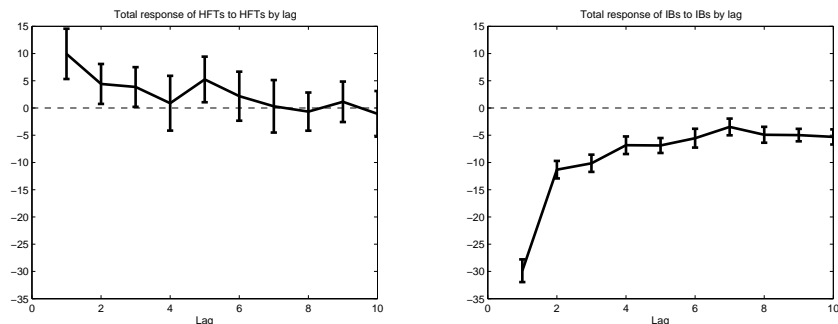Figure 1: Lag-by-lag responses to trades by firms in the same category.

The figure displays the total responses across all traders in either category (HFT or IB) to a 100-share increase in past activity by all other traders in that same category, across lags 1 to 10. Trade activity is measured either as order flow, total volume, or net position. For a given lag, the responses are calculated by first summing the parameters describing the response of firm i to all firms $j \neq i$ in the same category and then summing this quantity across all firms i in the given category (i.e., the responses of HFTs to HFTs are calculated as $\sum_i \sum_{j \neq i} (A_{11,k})_{i,j}$ for each lag $k = 1, ..., 10$). The resulting double sum is scaled by 100 to represent the response to an activity change of that size. The plotted values have a direct relationship with the coefficients reported in Table 2, whereby the sum of the parameters across all lags in each plot are identical to the corresponding coefficients in Table 2 scaled by 100. The vertical bars surrounding each point in the graphs represent 95 percent confidence intervals based on the bootstrapped standard errors.
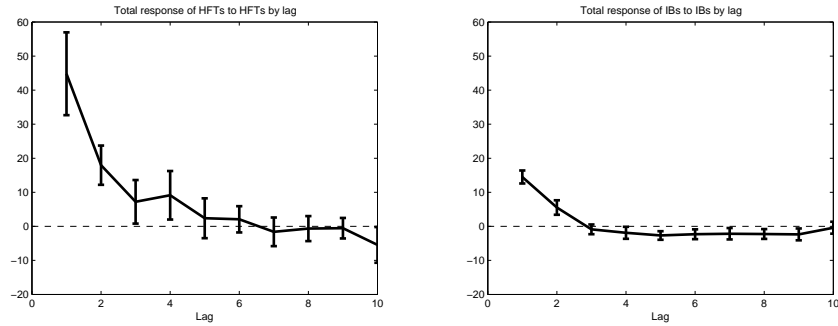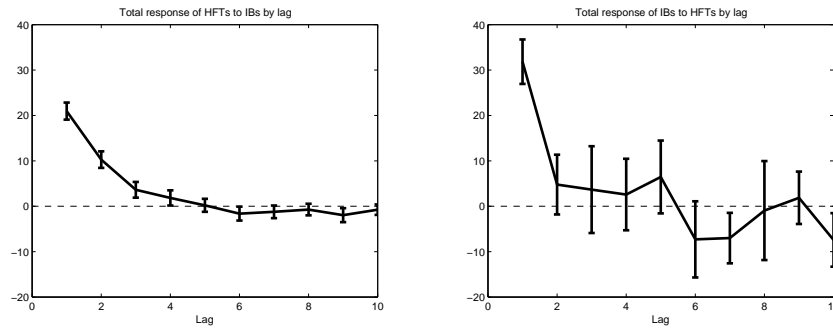

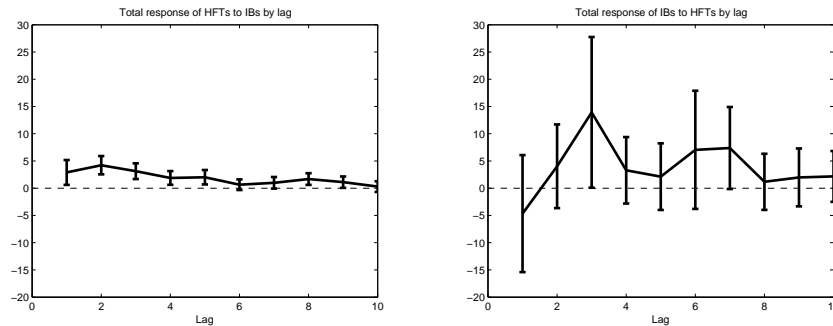(a) *Graph A. Order Flow*
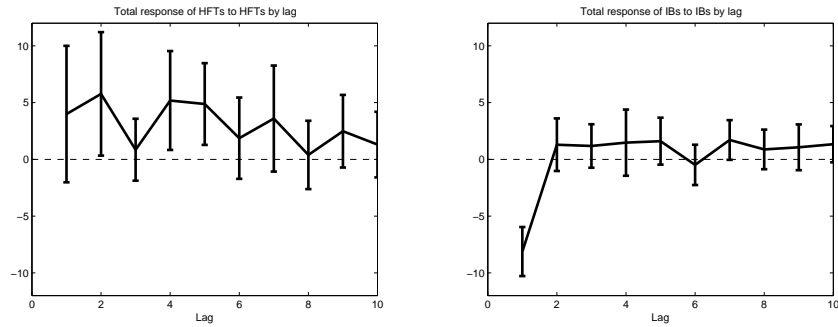
(b) *Graph B. Total Volume*

(c) *Graph C. Net Position*

Figure 2: Lag-by-lag responses to trades by firms in the other category
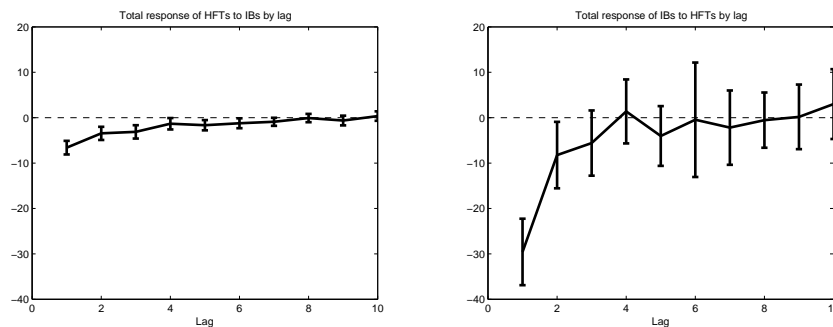
The figure displays the total responses across all traders in either category (HFT or IB) to a 100-share increase in past activity by all traders in the other category, across lags 1 to 10. That is, the graphs show the responses of HFTs to past trading by IBs and vice versa. Trade activity is measured either as order flow, total volume, or net position. The responses are calculated in an analogous manner to Figure 1. As per Figure 1, the plotted values have a direct relationship with the coefficients reported in Table 2, whereby the sum of the parameters across all lags in each plot are identical to the corresponding coefficients in the table scaled by 100. The vertical bars surrounding each point in the graphs represent 95 percent confidence intervals based on the bootstrapped standard errors.

(a) *Graph A. Order Flow*

(b) *Graph B. Total Volume*

(c) *Graph C. Net Position*