Interactions among High-Frequency Traders

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

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 com-
monality 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.

I.   Introduction

High-frequency trading (HFT), where automated computer traders interact at
lightning-fast speed with electronic trading platforms, has become an important
feature of many modern markets. The rapid growth and increased prominence of
these ultra-fast traders have given rise to concerns regarding their impact on mar-
ket 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 Oct. 15, 2014, have high-lighted such worries. Over the past few years, numerous empirical studies have analyzed the market impact of 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 (see, e.g., Hendershott, Jones, and Menkveld (2011), Hendershott and Riordan (2013), Brogaard, Hendershott, and Riordan (2014), and Chaboud, Chiquoine, Hjalmarsson, and Vega (2014)). The current article 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 toward 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 like 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 a large segment of market participants is provided by the "Quant Meltdown" in Aug. 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. Financial Times Stock Exchange (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);[3] our sample spans 4 months, from Sept. 1 through Dec. 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% 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

---

[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 article, 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]For information on the transaction reporting system underlying the Zen database, see https://www.fca.org.uk/markets/transaction-reporting.

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 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 in the following manner. In a lead–lag (Granger causality) sense, HFT trading activity tends to be strongly positively related across firms, for both directional and nondirectional 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 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

---

[4]Trading activity by other market participants not included in these two groups of HFTs and IBs is modeled 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.

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), Brogaard et al. (2014)).

The remainder of the paper is organized as follows: Section II provides a brief literature review, and Section III describes the data and presents 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 preprogrammed 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 preprogrammed, systematic nature of HFT might also limit the diversity of the strategies that HFTs implement. This notion is given empirical support by Chaboud et al. (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 because it could potentially have destabilizing effects on the market (Haldane (2011), 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 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 the model 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) 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 a 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 4 months from Sept. 1 to Dec. 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. 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 restrict our analysis to trades executed on the LSE, which accounted for between 55% and 70% 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

---

[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 (2015) study the importance of co-location across HFT firms, and Brogaard, Garriott, and Pomeranets (2014) analyze entry and competition among HFT firms. Dobrev and Schaumburg (2016) explicitly analyze cross-market linkages in high-frequency trading.

[6]Our data end on Dec. 31, 2012, although the last trading day we use in our sample is Dec. 21, 2012. We drop the 2 trading days between Christmas and New Year's, 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 LSE (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 following discussion on matching the 2 databases). This leaves a total of 92 stocks in our sample, which, for simplicity, we refer to as the FTSE 100 sample.

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

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; that is, 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 executed trades, the amount of overnight positions held, the duration of limit orders, the use of proprietary capital, and 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 (2017), 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 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 mid price 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 Appendix A, 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

---

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

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 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 $\text{VLM}_{i,s,t}$, 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 $\text{VLM}_{i,s,t}^{\text{TAKE}}$ and $\text{VLM}_{i,s,t}^{\text{MAKE}}$, respectively. The sums 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

$$(1) \qquad \text{OF}_{i,s,t} \quad = \quad \text{VLM}_{i,s,t}^{\text{TAKE}}(\text{BUY}) - \text{VLM}_{i,s,t}^{\text{TAKE}}(\text{SELL}),$$

where $\text{VLM}_{i,s,t}^{\text{TAKE}}(\text{BUY})$ and $\text{VLM}_{i,s,t}^{\text{TAKE}}(\text{SELL})$ 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,

$$(2) \qquad \text{NP}_{i,s,t} \quad = \quad \text{VLM}_{i,s,t}(\text{BUY}) - \text{VLM}_{i,s,t}(\text{SELL}).$$

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 the variables across all HFTs (IBs). That is,

$$(3) \qquad \text{VLM}_{s,t}^{\text{HFT}} \quad = \quad \sum_{i \in \text{HFT}} \text{VLM}_{i,s,t},$$

$$(4) \qquad \text{OF}_{s,t}^{\text{HFT}} \quad = \quad \sum_{i \in \text{HFT}} \text{OF}_{i,s,t},$$

and

$$(5) \qquad \text{NP}_{s,t}^{\text{HFT}} \quad = \quad \sum_{i \in \text{HFT}} \text{NP}_{i,s,t}.$$

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 sums 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 GBP), the ratio of net-position change to daily volume (based on GBP values), and the number of times that the estimated inventory crosses 0 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.

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 LSE. 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 because 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,

---

TABLE 1

Summary Statistics

Table 1 reports the means and standard deviations for various measures of trade activity for high frequency traders (HFTs), investment banks (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 1 trade. The sample includes all limit-order book trades in Financial Times Stock Exchange (FTSE) 100 stocks on the London Stock Exchange (LSE) from Sept. 1 to Dec. 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 LSE.

	HFT		IB		OTHER	
Activity Measure	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Volume (shares, thousands)	188.13	1,108.36	288.70	976.94	122.23	1,016.99
Volume (GBP, thousands)	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, thousands)	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

---

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

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 vs. around 5,000 pounds for HFTs and IBs).

The final 3 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 0 during the day.[10] All three measures capture aspects of the notion that HFTs 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% for HFTs, 35% for IBs, and 43% for OTHERs. The inventory of an HFT crosses 0 about 7 times per day, whereas for IBs and OTHERs, the corresponding figures are approximately 2 and 1 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 VARs, which capture the dependency in the 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, Lopéz 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 formulation 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.

---

[10]When calculating the averages and standard deviations of the daily ratio of position change 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 4 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 by, for instance, Benos and Sagade (2016) and Brogaard et al. (2014).

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 in the next subsection. Formally, we perform a type of Granger causality tests, which, in line with the previous discussion, capture both contemporaneous correlations and actual causality in the sense that the actions of one trading firm cause a subsequent action by another firm. In the following 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 the number of shares traded.

A.   A Panel VAR of Stock Trading

Let $\text{HFT}_{i,s,t}$ be the trading activity of HFT firm i at trade event t in stock s, and, analogously, let $\text{IB}_{i,s,t}$ be the trading activity of IB i at trade event t in stock s. As mentioned previously, trading activity is measured by either order flow, total volume, or change in inventory, and t is measured in trade time. In the following discussion, 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 $i = 1, \ldots, 10$ HFTs, and define $\mathbf{IB}_t^s$ as the corresponding vector of IB trading activity. That is,

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

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 $\mathbf{Y}_t^s \equiv \left(\mathbf{HFT}_t^{s\prime}, \mathbf{IB}_t^{s\prime}, M_t^s \right)'$ 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:

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

The dependent variable, $\mathbf{Y}_t^s \equiv \left(\mathbf{HFT}_t^{s\prime}, \mathbf{IB}_t^{s\prime}, M_t^s \right)'$, 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, $k = 1, \ldots, 10$ are 21×21 lag matrix coefficients.

---

[13]For change in inventory, the residual market activity is actually a linear combination of the components of $\mathbf{HFT}_t^s$ and $\mathbf{IB}_t^s$ because, by construction, the net inventory change of the entire market ($\sum_i \text{HFT}_{i,s,t} + \sum_i \text{IB}_{i,s,t} + M_t^s$) must equal 0 for each stock s during each trade t. The residual market, M_t^s, thus cannot 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.

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 modeled 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 observation, 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 intraday dummy variables for each distinct half-hour period within the trading day (i.e., 8:00–8:30AM, 8:31–9:00AM, 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 Sept. 1 and Dec. 31, 2012. Data are sampled during the normal trading hours between 8:00AM and 4:30PM, although activity in the first and last 5 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 nonparametric 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 and 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 as distinct groups? We attempt to test these hypotheses within the previously described 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

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

---

[14] As a robustness check, we also estimate the VAR model 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 0. 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 0. In the interest of space, the results for the 20-lag VAR are not reported.

[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 1 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.

In equation (7), the parameter submatrices $(A_{ij,k})$, $i, j = 1, \ldots, 3$, now group the coefficients for the HFTs, IBs, and the residual market. $A_{11,k}$ $(A_{22,k})$ corresponds to lag dependencies among HFTs (IBs). The submatrix $A_{12,k}$ $(A_{21,k})$ captures the effects of past trading by IBs (HFTs) on the current trading of HFTs (IBs), and the submatrix $A_{31,k}$ $(A_{32,k})$ 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 0. Similarly, we test whether past trading by IBs affects the 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 0. In both cases, the null of no Granger causation is rejected if the sum is statistically significantly different from 0. 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 (long-run) relationship, that is, 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 $A_{31,k}$ $(A_{32,k})$ is equal to 0. 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 0.[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 (in parentheses) corresponding to the Wald test of the null hypothesis that the sum is equal to 0, which might be interpreted as a null hypothesis of no (long-run) Granger causality. As mentioned previously, the p-values are obtained through a bootstrap procedure, which controls for heteroskedasticity and cross-sectional dependence between stocks (see Appendix B for details).

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 affected by the past trading in that stock by other HFT firms. In particular, the

---

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

TABLE 2

VAR Results

Table 2 reports results from the trade-time panel VAR model specified in equation (6), using pooled data across all stocks. Coefficient estimates and p-values (in parentheses) for hypotheses tests regarding high-frequency trader (HFT) and investment bank (IB) activity are shown. The first 2 columns give 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 1 time period to the sample. The sample includes all limit-order book trades in Financial Times Stock Exchange (FTSE) 100 stocks on the London Stock Exchange (LSE) from Sept. 1 to Dec. 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 LSE. 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
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)	—

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 past trades of other IBs have little effect on the current trading direction of a given IB. The estimated effect is not significant at the 5% level (p-value = 0.08) and is 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 4–6 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 7–9 of Table 2, current trading by the remainder of the market (residual trading) reacts strongly to the 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 2 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]

The graphs in Figures 1 and 2 tell essentially the same story as the coefficients and test results reported in Table 2. However, the lag-by-lag breakdown of effects provides a better idea of how the lead–lag relationships evolve over time. As is evident from Graph A in Figures 1 and 2, both of 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 0 and/or not statistically significant (the vertical bars around each lag coefficient indicate 95% 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 a given HFT's own past trading. Analogous interpretations apply to the other graphs in Figures 1 and 2.

The results for volume, which are shown in the 2 middle 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 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 the possibility that these effects on the remainder of the market

---

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

FIGURE 1
Lag-by-Lag Responses to Trades by Firms in the Same Category

Figure 1 displays the total responses across all traders in either category (high-frequency trader (HFT) or investment bank (IB)) to a 100-share increase in past activity by all other traders in that same category across lags 1–10. Trade activity is measured 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 $k = 1, \ldots, 10$). 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 sums 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% confidence intervals based on the bootstrapped standard errors.

Graph A. Order Flow

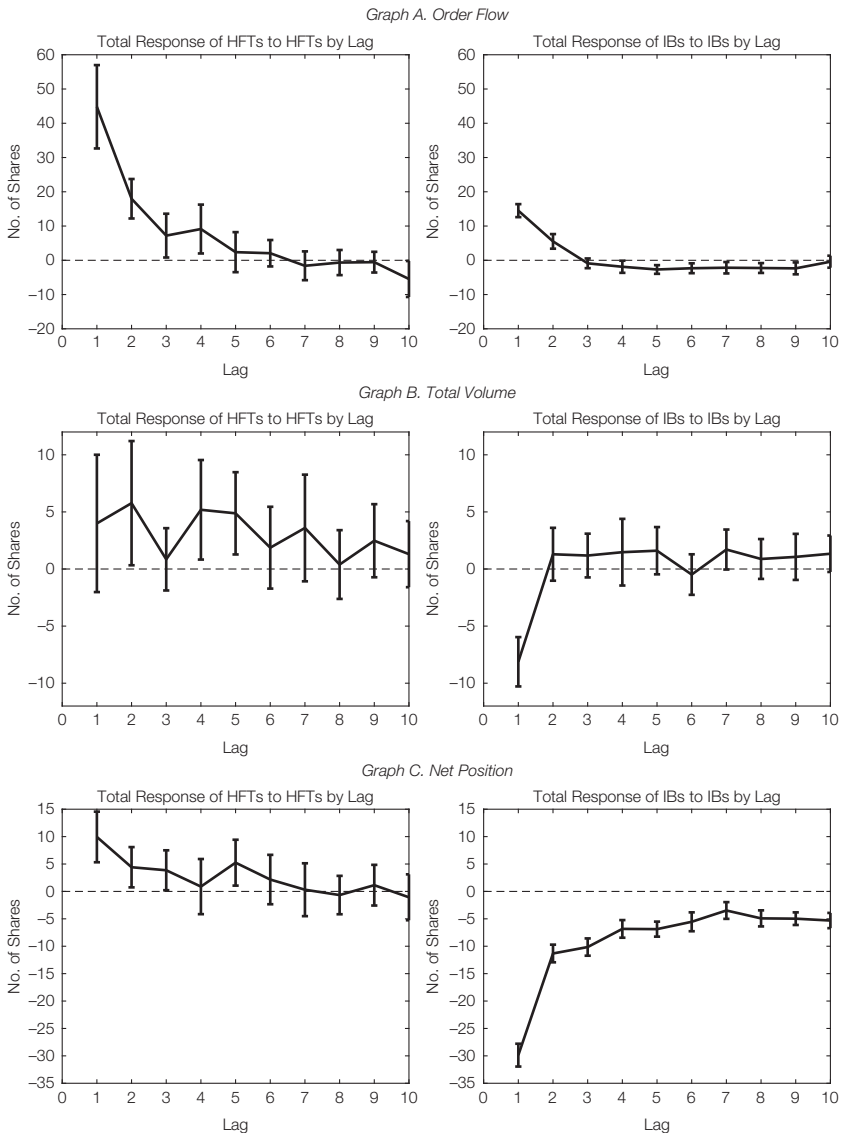Graph B. Total Volume

Graph C. Net Position

FIGURE 2

Lag-by-Lag Responses to Trades by Firms in the Opposite Category

Figure 2 displays the total responses across all traders in either category (high-frequency trader (HFT) or investment bank (IB)) to a 100-share increase in past activity by all traders in the opposite category across lags 1–10. 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 sums 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% confidence intervals based on the bootstrapped standard errors.

Graph A. Order Flow

Graph B. Total Volume

Graph C. Net Position
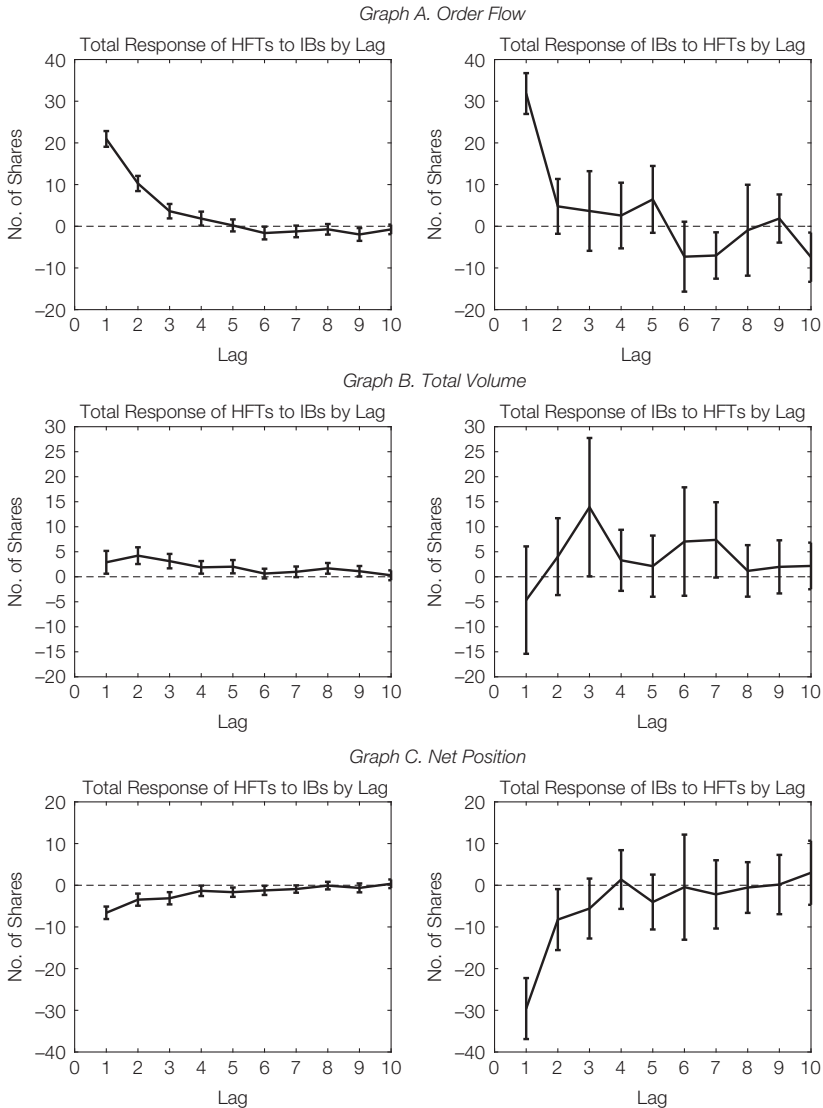
are the same for HFTs and IBs (row 9). Graph B in Figures 1 and 2 shows the volume results broken down by each lag. In comparison to the order flow results, shown in Graph A in the figures, there is a tendency for the volume effects to be less concentrated in the first few lags. In the case of the response 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 2 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; 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 a 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.

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 because all HFTs are willing to trade with you, or it would dry up because 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, conversely, 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 those 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. Conversely, IBs 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 et al. (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 behavior 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

$$(8) \quad \text{CORR\_TRADING}_{s,t}^{\text{HFT}} = N\,(\text{BUY})_{s,t}^{\text{HFT}} - \frac{N\,(\text{BUY})_{s,t}^{\text{HFT}} + N\,(\text{SELL})_{s,t}^{\text{HFT}}}{2},$$

where $N\,(\text{BUY})_{s,t}^{\text{HFT}}$ is the number of aggressive HFT buyers and $N\,(\text{SELL})_{s,t}^{\text{HFT}}$ 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 that has identical aggressive buy and sell volumes) in a given stock in a given time interval does not add to the number of aggressive buyers or 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 a 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 0. An analogous metric is also constructed for IBs, denoted by $\text{CORR\_TRADING}_{s,t}^{\text{IB}}$.

The correlation metrics, $\text{CORR\_TRADING}_{s,t}^{\text{HFT}}$ and $\text{CORR\_TRADING}_{s,t}^{\text{IB}}$, are calculated for all stocks in the sample of FTSE 100 shares using minute-by-minute data. The 1-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 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 5-minute frequency.

To measure the contemporaneous and lagged price impact associated with correlated trading, we regress 1-minute returns on contemporaneous and lagged order flows, the correlated trading metrics and their lags, and 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 trade-correlation metrics. Thus, our full specification takes the form

$$
(9) \quad R_{s,t} = \alpha_s + \sum_{i=0}^{5} \beta_{\text{OF},i}^{\text{HFT}} \text{OF}_{s,t-i}^{\text{HFT}} + \sum_{i=0}^{5} \beta_{\text{OF},i}^{\text{IB}} \text{OF}_{s,t-i}^{\text{IB}} + \sum_{i=1}^{5} \beta_{\text{OF},i}^{\text{RES}} \text{OF}_{s,t-i}^{\text{RES}}
$$

$$
+ \sum_{i=0}^{5} \beta_{\text{CORR},i}^{\text{HFT}} \text{CORR\_TRADING}_{s,t-i}^{\text{HFT}}
$$

$$
+ \sum_{i=0}^{5} \beta_{\text{CORR},i}^{\text{IB}} \text{CORR\_TRADING}_{s,t-i}^{\text{IB}}
$$

$$
+ \sum_{i=0}^{5} \beta_{|\text{CORR}|,i}^{\text{HFT}} |\text{CORR\_TRADING}|_{s,t-i}^{\text{HFT}}
$$

$$
+ \sum_{i=0}^{5} \beta_{|\text{CORR}|,i}^{\text{IB}} |\text{CORR\_TRADING}|_{s,t-i}^{\text{IB}}
$$

$$
+ \sum_{i=0}^{5} \beta_{\text{OF}\times|\text{CORR}|,i}^{\text{HFT}} \left(\text{OF}_{s,t-i}^{\text{HFT}} \times |\text{CORR\_TRADING}|_{s,t-i}^{\text{HFT}} \right)
$$

$$
+ \sum_{i=0}^{5} \beta_{\text{OF}\times|\text{CORR}|,i}^{\text{IB}} \left(\text{OF}_{s,t-i}^{\text{IB}} \times |\text{CORR\_TRADING}|_{s,t-i}^{\text{IB}} \right) + u_{s,t}.
$$

Here, $R_{s,t}$ is the 1-minute return of stock s in period t, and $\text{OF}_{s,t}^{\text{HFT}}, \text{OF}_{s,t}^{\text{IB}}$, and $\text{OF}_{s,t}^{\text{RRS}}$ are the order flows from HFTs, IBs, and the remainder of the market (the "residual" order flow). $\text{CORR\_TRADING}_{s,t}^{\text{HFT}}$ and $\text{CORR\_TRADING}_{s,t}^{\text{IB}}$ 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,

---

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

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]

The model is estimated by least squares, pooling the data across all stocks while allowing for stock-specific intercepts α_s and including 5 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 deviations 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. Because stock-specific intercepts (i.e., fixed effects) are included in the regressions, this de-meaning does not in any way alter the regression specifications but merely allows for an easier interpretation of the coefficients.

Summary statistics for the (nonstandardized) 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 they are positively related, the order flows and trade-correlation metrics are clearly distinct activity measures.

Table 4 reports the regression results. For brevity, we only report the sum of the coefficients for the 5 lags and the associated (robust) t-statistics. In column 1, we first run a simple regression of 1-minute returns on contemporaneous and 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.

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 a positive contemporaneous correlation between order flow and returns and a negative correlation between past order flow and returns uniformly across HFTs, IBs, and the rest of the market.

---

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

[21]The correlation metrics, $CORR\_TRADING_{s,t}^{HFT}$ and $CORR\_TRADING_{s,t}^{IB}$, are not scaled prior to estimation because they are already in a standardized format, taking on values between $+5$ and -5.

[22]This regression can be 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 $i = 0, \ldots, 5$, and all other coefficients are restricted to equal 0.

TABLE 3

Summary Statistics for Returns, Order Flows, and Trade Correlations

Table 3 reports means and standard deviations for returns, high-frequency trader (HFT), investment bank (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 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 1-minute frequency (Panel A) or the 5-minute frequency (Panel B). The 1-minute and 5-minute samples are constructed from all limit-order book trades in Financial Times Stock Exchange (FTSE) 100 stocks on the London Stock Exchange (LSE) from Sept. 1 to Dec. 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 LSE.

| | Panel A. 1-Minute Frequency | | | | | | Panel B. 5-Minute Frequency | | | | | |
Summary Statistics	RETURNS	OF^{HFT}	OF^{IB}	OF^{RES}	$\text{CORR\_TRADING}^{\text{HFT}}$	$\text{CORR\_TRADING}^{\text{IB}}$	RETURNS	OF^{HFT}	OF^{IB}	OF^{RES}	$\text{CORR\_TRADING}^{\text{HFT}}$	$\text{CORR\_TRADING}^{\text{IB}}$
Mean	0.014	39.7	110.8	−139.2	0.000	0.006	0.030	91.6	250.2	−312.3	−0.001	0.008
Std. Dev.	5.50	8,331	12,481	14,570	0.302	0.476	9.13	12,626	21,736	24,353	0.330	0.558
Correlation Matrix												
RETURNS	1.000	0.078	0.130	0.071	0.189	0.293	1.000	0.048	0.141	0.041	0.110	0.266
OF^{HFT}		1.000	0.182	−0.268	0.271	0.070		1.000	0.193	−0.285	0.203	0.054
OF^{IB}			1.000	−0.291	0.057	0.266			1.000	−0.323	0.032	0.216
OF^{RES}				1.000	−0.025	−0.028				1.000	−0.021	−0.032
$\text{CORR\_TRADING}^{\text{HFT}}$					1.000	0.154					1.000	0.124
$\text{CORR\_TRADING}^{\text{IB}}$						1.000						1.000

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 is evident from the formal Wald tests reported toward the bottom of the table. The regression results thus suggest that HFTs' correlated 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

---

[23] For a given value of HFT order flow, the impact of a unit increase in HFT trade correlation is given by $\beta_{\text{CORR},0}^{\text{HFT}} + \beta_{|\text{CORR}|,0}^{\text{HFT}} + \beta_{\text{OF}\times|\text{CORR}|,0}^{\text{HFT}} \times \text{OF}_{s,t}^{\text{HFT}} \approx 0.2 - 0.07 \times \text{OF}_{s,t}^{\text{HFT}}$, where $\text{OF}_{s,t}^{\text{HFT}}$ 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.

TABLE 4

Price-Impact Regressions Using 1-Minute Data

Table 4 reports 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 deviations 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 Financial Times Stock Exchange (FTSE) 100 stocks on the London Stock Exchange (LSE) from Sept. 1 to Dec. 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 LSE. In total, there are 3,311,540 1-minute observations. The t-statistics (reported in parentheses) and Wald tests are based on Driscoll and Kraay (1998) standard errors, which are robust to heteroskedasticity, serial correlation, and cross-sectional dependence.

Right-Hand-Side Variables	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 (149.8)	0.351 (126.4)	0.378 (107.2)
$OF_{t-1,t-5}^{HFT}$		−0.039 (−11.71)	−0.040 (−11.31)	−0.023 (−5.310)
OF_t^{IB}		0.473 (179.6)	0.384 (190.1)	0.410 (159.6)
$OF_{t-1,t-5}^{IB}$		−0.075 (−37.86)	−0.046 (−21.03)	−0.046 (−16.74)
OF_t^{RES}		0.519 (156.9)	0.466 (157.1)	0.457 (155.9)
$OF_{t-1,t-5}^{RES}$		−0.071 (−28.57)	−0.058 (−23.98)	−0.058 (−23.94)
$CORR\_TRADING_t^{HFT}$			0.186 (45.72)	0.203 (47.23)
$CORR\_TRADING_{t-1,t-5}^{HFT}$			0.017 (3.024)	0.026 (4.389)
$CORR\_TRADING_t^{IB}$			0.310 (85.08)	0.324 (83.93)
$CORR\_TRADING_{t-1,t-5}^{IB}$			−0.154 (−48.42)	−0.155 (−48.99)
$\|CORR\_TRADING\|_t^{HFT}$				0.002 (0.520)
$\|CORR\_TRADING\|_{t-1,t-5}^{HFT}$				0.010 (1.511)
$\|CORR\_TRADING\|_t^{IB}$				−0.003 (−1.064)
$\|CORR\_TRADING\|_{t-1,t-5}^{IB}$				−0.010 (−2.689)
$OF_t^{HFT} \times \|CORR\_TRADING\|_t^{HFT}$				−0.071 (−12.90)
$OF_{t-1,t-5}^{HFT} \times \|CORR\_TRADING\|_{t-1,t-5}^{HFT}$				−0.036 (−5.465)
$OF_t^{IB} \times \|CORR\_TRADING\|_t^{IB}$				−0.046 (−15.64)
$OF_{t-1,t-5}^{IB} \times \|CORR\_TRADING\|_{t-1,t-5}^{IB}$				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

the estimated coefficients are fairly small in magnitude. The coefficients for the (noninteracted) 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 1-standard-deviation HFT order flow shock is thus associated with a 0.2-standard-deviation shock to returns ($\beta_{\text{OF},0}^{\text{HFT}} \times 0.5 \approx 0.4 \times 0.5$), 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 ($\beta_{\text{CORR},0}^{\text{HFT}} \approx 0.2$). 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]

As a robustness check, we also estimate the same regressions using data sampled every 5 minutes. That is, 5-minute returns are now regressed on the order flow and trade-correlation variables constructed over 5-minute intervals. However, to keep the temporal span of the lags identical to the 1-minute specification, only 1 lag is now included. Otherwise, the two specifications are identical. The results, shown in Table 5, strongly echo those seen in Table 4. The statistical significance of some of the estimates based on the 5-minute data is somewhat weaker than in the 1-minute case, but otherwise, the results are consistent across the two sampling frequencies. Importantly, there is no evidence that HFTs' correlated trading leads to price reversals.

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 those by Carrion (2013) and Brogaard et al. (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

---

[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 $-0.023\text{OF}_{s,t-1}^{\text{HFT}} + 0.026\text{CORR\_TRADING}_{s,t-1}^{\text{HFT}} + 0.010|\text{CORR\_TRADING}|_{s,t-1}^{\text{HFT}} - 0.036(\text{OF}_{s,t-1}^{\text{HFT}} \times |\text{CORR\_TRADING}|_{s,t-1}^{\text{HFT}}) = -0.023 \times 0.5 + 0.026 + 0.01 - 0.036 \times 0.5 = 0.0065$.

TABLE 5

Price-Impact Regressions Using 5-Minute Data

Table 5 reports 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 deviations 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 Financial Times Stock Exchange (FTSE) 100 stocks on the London Stock Exchange (LSE) from Sept. 1 to Dec. 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 LSE. In total, there are 662,308 5-minute observations. The t-statistics (reported in parentheses) and Wald tests are based on Driscoll and Kraay (1998) standard errors, which are robust to heteroskedasticity, serial correlation, and cross-sectional dependence.

Right-Hand-Side Variables	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 (39.89)	0.261 (29.24)	0.276 (28.21)
$OF_{t-1,t-5}^{HFT}$		−0.044 (−10.77)	−0.049 (−11.30)	−0.053 (−10.39)
OF_t^{IB}		0.462 (96.37)	0.394 (86.95)	0.431 (79.53)
$OF_{t-1,t-5}^{IB}$		−0.063 (−25.01)	−0.057 (−21.18)	−0.062 (−18.18)
OF_t^{RES}		0.429 (85.34)	0.389 (85.60)	0.385 (84.80)
$OF_{t-1,t-5}^{RES}$		−0.056 (−19.95)	−0.053 (−19.58)	−0.052 (−19.38)
$CORR\_TRADING_t^{HFT}$			0.047 (8.647)	0.054 (9.849)
$CORR\_TRADING_{t-1,t-5}^{HFT}$			0.010 (2.813)	0.007 (1.994)
$CORR\_TRADING_t^{IB}$			0.167 (52.82)	0.184 (57.88)
$CORR\_TRADING_{t-1,t-5}^{IB}$			−0.034 (−15.97)	−0.036 (−17.22)
$|CORR\_TRADING|_t^{HFT}$				−0.001 (−0.281)
$|CORR\_TRADING|_{t-1,t-5}^{HFT}$				0.010 (2.462)
$|CORR\_TRADING|_t^{IB}$				−0.006 (−1.953)
$|CORR\_TRADING|_{t-1,t-5}^{IB}$				0.002 (0.929)
$OF_t^{HFT} \times |CORR\_TRADING|_t^{HFT}$				−0.040 (−3.688)
$OF_{t-1,t-5}^{HFT} \times |CORR\_TRADING|_{t-1,t-5}^{HFT}$				0.014 (1.853)
$OF_t^{IB} \times |CORR\_TRADING|_t^{IB}$				−0.060 (−16.57)
$OF_{t-1,t-5}^{IB} \times |CORR\_TRADING|_{t-1,t-5}^{IB}$				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

HFTs appears to be relevant over horizons that stretch for at least a few minutes, not just over shorter intervals of a few seconds.[26]

VI.   Conclusion

Using a unique data set of the transactions of individual 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 and 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 same time, whereas investment banks instead tend to trade more disparately and absorb each other's 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, they have the same sign at approximately the same time. In other words, HFTs appear 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 instead reflect that HFT firms are 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 1-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, and time to the nearest second) using the fact that we also

---

[26]Brogaard et al. (2014) document that HFTs appear to be able to predict price movements over the next few seconds, and Carrion (2013) finds that HFTs also appear to have the ability to time the market over longer (greater than 5 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.

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.

Because 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 to classify trade direction (e.g., Easley et al. (2012), Chakrabarty, Pascual, and Shkilko (2015), 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 are not sequenced against each other (i.e., trades vs. quotes) within each time interval. Our procedure still suffers from the limitations in the Lee and Ready (1991) trade-signing algorithm, but most studies suggest that this approach works very well provided that 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 IV.A are computed using a nonparametric nonoverlapping block bootstrap, where blocks are defined by trading days in our sample. This method resamples with replacement from the 80 trading days in our data and constructs a bootstrap sample containing all trades from the resampled days for each bootstrap run, $b = 1, \ldots, B$. For each bootstrapped run b, we construct the data matrices $\{\mathbf{Y}_t^{s,b}, \mathbf{X}_{t-1}^{s,b}, \mathbf{G}_t^b\}$ for each stock s and then estimate the pooled parameters $\{A_1^b, \ldots, A_{10}^b, \Lambda^b, \Psi^b\}$ and the stock-specific intercepts $\{\mu^{s,b}\}$. 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 $B = 100$ 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 and 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 of these issues (the bootstrap is robust to heteroskedasticity, regardless of the block structure). Although 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 nonparametric 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, whereas error terms across blocks (i.e., trades occurring on separate days) are plausibly independent. Last, 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, University of Washington (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, Cornell University (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." *Review of Financial Studies*, 28 (2015), 3407–3443.

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, University of Utah (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." *Journal of Financial Markets*, 25 (2015), 52–79.

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.

Dobrev, D., and E. Schaumburg. "High-Frequency Cross-Market Trading: Model Free Measurement and Applications." Working Paper, Board of Governors of the Federal Reserve System (2016).

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." *Financial Review*, 49 (2014), 395–419.

Haldane, A. "The Race to Zero." Speech, International Economic Association 16th World Congress, Beijing. Available at: https://core.ac.uk/download/pdf/24060887.pdf (2011).

Hall, A. R. *Generalized Method of Moments*. Oxford, UK: Oxford University Press (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, London Business School (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), 1250022-1–1250022-15.

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." *Journal of Finance*, forthcoming (2017).

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, Northwestern University (2016).

Kozhan, R., and W. W. 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, T. S. Rao, S. S. Rao, and C. R. Rao, eds. Oxford, UK: 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, HEC Paris (2013).

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

Stein, J. "Presidential Address: Sophisticated Investors and Market Efficiency." *Journal of Finance*, 64 (2009), 1517–1548.

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

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

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