High-Frequency Trading around Large Institutional Orders[1]

Vincent van Kervel and Albert J. Menkveld

January 29, 2016

[1]Vincent van Kervel, Pontificia Universidad Católica de Chile, Avenida Vicuña Mackenna 4860, Macul 7820436, Santiago, Chile, vincentvankervel@gmail.com and Tinbergen Institute. Albert J. Menkveld, VU University Amsterdam, FEWEB, De Boelelaan 1105, 1081 HV, Amsterdam, Netherlands, +31 20 598 6130, albertjmenkveld@gmail.com and Tinbergen Institute. We thank Jonathan Brogaard, Björn Hagströmer, Terrence Hendershott, Ron Kaniel, Robert Korajzyck, Charles-Albert Lehalle, Anna Obizhaeva, Neil Pearson, Marc Ponsen, Yazid Sharaiha, George Sofianos, and Avi Wohl for their very useful comments. We are also grateful for comments received during conferences at NBIM and Tel Aviv University and seminars at AFM, CFTC, Goldman Sachs, and the U.S. Securities and Exchange Commission. We thank Bernard Hosman and Sailendra Prasanna Mishra for excellent research assistance. Menkveld gratefully acknowledges NWO for a VIDI grant. The authors further acknowledge support from APG, DNB, NBIM, Swedbank Robur, and QuantValley/FdR Quantitative Management Initiative.

High-Frequency Trading around Large Institutional Orders

Abstract

Liquidity suppliers lean against the wind. We analyze whether high-frequency traders (HFTs) lean against large institutional orders that execute through a series of child orders. The alternative is that HFTs go "with the wind" and trade in the same direction. We find that HFTs initially lean against orders but eventually turn around and go with them for long-lasting orders. This pattern explains why institutional trading cost is 46% lower when HFTs lean against the order (by one standard deviation) but 169% higher when they go with it. Further analysis supports recent theory, suggesting HFTs "back-run" on informed orders.

(for internet appendix click http://goo.gl/mzfUh0)

1 Introduction

Migration to electronic trading created a new type of market participant: high-frequency traders (HFTs). The U.S. Securities and Exchange Commission (SEC) characterized this type as "professional traders acting in proprietary capacity" who use "extraordinarily high-speed and sophisticated computer programs for generating, routing, and executing orders" and end the trading day "in as close to a flat position as possible." HFTs entered securities markets in the late 2000s, appearing first in equity markets. Their market participation, in percentage terms, is typically a couple of deciles (SEC, 2010; Securities and Authority, 2014).

High-frequency trading has triggered a great deal of academic study, particularly after Nasdaq released data that labeled HFTs in trades and quotes. The evidence is by and large favorable for HFTs emphasizing reduced bid–ask spreads and increased price efficiency. The evidence is mixed, however, on how HFTs relate to "excess" volatility, such as in flash crashes. Jones (2014), Biais and Foucault (2014), and the SEC (2014) survey the young and rapidly growing high-frequency trading literature.[1]

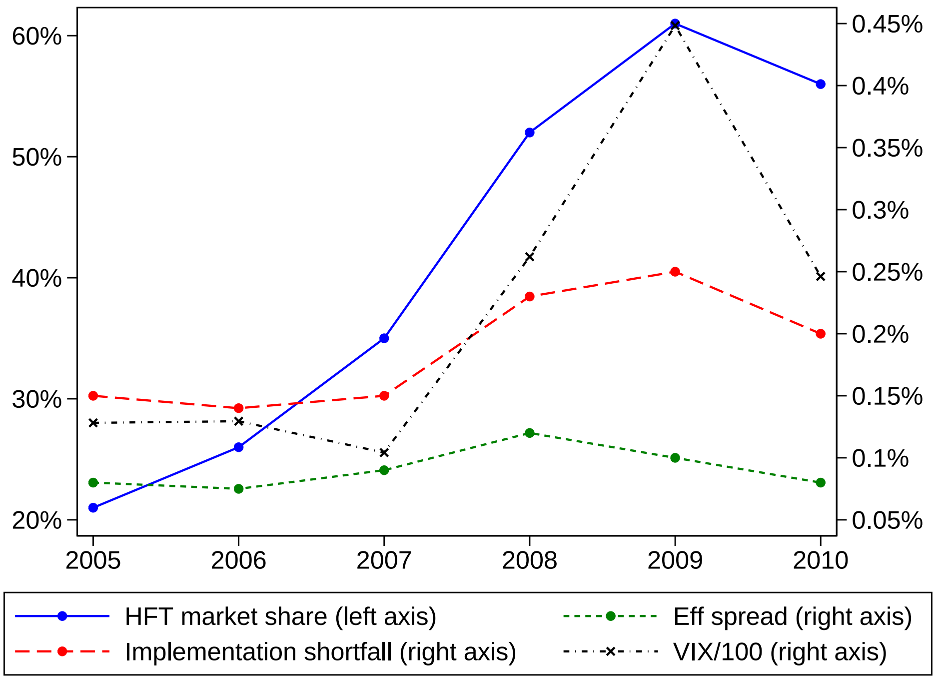
Relatively unexplored is how HFT affects trading by an important group of end users of securities markets: institutional investors. Retail investors benefit from a smaller bid–ask spread, since there generally is enough depth at the best quote to execute their order. Institutional investors, however, need to "work their order" by splitting it up into smaller pieces and feeding them to the market sequentially. They care about "implementation shortfall," that is, the average price at which the entire order executed relative to the price at which it started. In other words, how far did they push the price away from them? Institutional investors care about *cumulative* price impact rather than the half-spread paid on a single child order execution. Some have expressed concern that trading costs have increased and attribute this to HFT presence.[2]

[1] Several empirical studies find that HFT activity reduces bid–ask spreads (Hendershott, Jones, and Menkveld, 2011; Hasbrouck and Saar, 2013; Menkveld, 2013; Brogaard, Hagströmer, Nordèn, and Riordan, 2015; van Kervel, 2015) and improves price efficiency (Boehmer, Fong, and Wu, 2014; Brogaard, Hendershott, and Riordan, 2014). The effect of high-frequency trading activity on short-term volatility and crashes is mixed: Some studies document a negative correlation (Hasbrouck and Saar, 2013; Chaboud, Chiquoine, Hjalmarsson, and Vega, 2014; Hasbrouck, 2015) whereas others document a positive correlation (Gao and Mizrach, 2013; Ye, Yao, and Gai, 2013; Boehmer, Fong, and Wu, 2014; Kirilenko, Kyle, Samadi, and Tuzun, 2014).

[2] See, for example, "Institutional Investors Air HFT Concerns" *Financial Times*, September 12, 2011; "Wealth Fund

Figure 1: Time trend HFT participation and institutional trading costs

This figure plots the overall time trend in HFT equity market participation, institutional trading costs, effective spread, and CBOE Volatility Index (VIX). The VIX is divided by 100 to align with the other variables. Institutional trading cost is measured as implementation shortfall. This is the average execution price on a large order, expressed relative to the price at the start of order execution (multiplied by -1 for sell orders). Institutional trading costs and the effective spread are from Anand, Irvine, Puckett, and Venkataraman (2013, Figure 1), who base it on Abel/Noser data. The HFT data are from "High Frequency Trading: Evolution and the Future," a report published by Capgemini in 2012.



The time trend seems to support institutional investor concern. Figure 1 plots trading costs in U.S. equity markets from 2005 to 2010, when HFT participation grew from about 21% to 56%. The implementation shortfall for institutional orders grew from 15 basis points (bps) to 20 bps, an increase of 33%. If one takes the effective half-spread as a proxy for retail investor costs, one finds that their costs did not change; it was 4 bps at the start and at the end of this period. The 2008–2009 peak in trading costs coincides with the financial crisis and is arguably due to elevated volatility. We should not overinterpret these time trends, but they do seem to warrant further study.

Cautions against Costs Exacted by High-Speed Trading" *The New York Times*, October 20, 2013; and "Berkshire's Munger: High-Frequency Tradings' Basically Evil," *Berkshire Munger*, May 3, 2013. Sofianos and Xiang (2013) analyze Goldman Sachs agency trades and argue that it is hard for HFTs to profitably trade on them. Such analysis demonstrates that clients of large sell-side banks are concerned about being detected by HFTs.

Objective. This paper's objective is to relate trading by HFTs to implementation shortfall. We conjecture that HFT net flow (i.e., buy minus sell volume) over the lifetime of an institutional order correlates with the order's implementation shortfall. In particular, HFTs reduce costs when they lean "against the wind," that is, they trade in opposite direction to the institutional order. They increase costs when they "go with the wind."

The empirical analysis is based on a sample that combines proprietary institutional investor execution data with publicly available HFT trade data (no inference needed). The sample runs from January 1, 2011, until March 31, 2013, and pertains to trading in Swedish stocks. The execution data were provided by four large institutional investors (APG, DNB, NBIM, and Swedbank Robur) and consist of 801,341 child order executions. We construct daily "meta-orders" by grouping into a single order all the child trades of an institution in a particular stock on a particular day. For brevity, we refer to these meta-orders as institutional orders. The final sample contains 5,910 orders, which, on average, contain 135 child order executions. Not surprisingly, we find that these orders are directional, that is, an institution's child order executions on a particular "stock–day" are either almost exclusively buys or almost exclusively sells. Finally, institutional orders are large, on average: \$1.940 million, or 4.0% when expressed as a percentage of the average daily volume.

An important benefit of this particular sample is that HFTs had to reveal their trades on NASDAQ OMX,[3] which was the dominant market, with a two-thirds market share.[4] We select Europe's largest high-frequency trading firms according to *Financial News*: Citadel, Flow Traders, Getco, IAT, IMC, Knight, Optiver, Spire, Susquehanna, and Virtu.[5] Collectively, their participation rate in trades is almost a third in our sample.

Findings. The empirical analysis yields three main findings. First, HFTs (as a group) lean against the wind in the first hour of an institutional order execution but go with the wind for multi-hour executions. The with–wind flow is so strong that HFT net flow over the lifetime of the order is eventually positive for long-lasting institutional buy orders and negative for sell orders. HFTs,

[3] This changed on March 23, 2014, when NASDAQ OMX changed to voluntary reports. Many HFT firms opted to go under the radar and not report their trades. See "Changes to Post Trade Counterparty Visibility in NASDAQ OMX Nordic Blue Chip Shares," *GlobeNewswire*, February 6, 2014.

[4] These numbers are from Fidessa, a trade reporting company (http://tinyurl.com/ozo8ytm).

[5] See "Europe's Top 10 High-Frequency Kingmakers," *Financial News*, October 3, 2011.

therefore, seem to be actively taking positions as opposed to simply mean-reverting "inventory." We stated this result cautiously, since we are aware that HFTs could have entered offsetting positions in alternative markets or highly correlated securities. We consider this concern somewhat unlikely, since perfectly correlated securities are hard to find for stocks and NASDAQ OMX is by far the largest equity exchange for Swedish stocks in our sample. Finally, one could worry that institutional investors and HFTs both respond to certain market conditions, which implies that their mutual correlation is driven by a "third factor." To rule out such an explanation, we create a placebo sample in which none of the institutional investors were active yet market conditions were similar (in terms of volume, volatility, and idiosyncratic and market return). We find against–wind behavior also in the placebo sample, but no with–wind behavior. The with–wind behavior, therefore, seems intimately related to the presence of the institutional order.

Second, the implementation shortfall on institutional orders correlates significantly with HFT net flow, controlling for standard covariates. The average shortfall is 7.4 bps. It is reduced by 3.4 bps for a one standard deviation against–wind flow ($235,000), a reduction of 46%. A one standard deviation with–wind flow ($241,000) increases shortfall by 12.5 bps, an increase of 169%. The magnitude is slightly smaller when implementation shortfall is measured in dollars. In this case, a one standard deviation with–wind flow increases it by $3,024, an increase of 106% relative to the sample average of $2,860. We further note that the implementation shortfall in our sample is the same order of magnitude as that reported in Anand, Irvine, Puckett, and Venkataraman (2013, Table 1), who document an order size-weighted shortfall of 25 bps. The equivalent number for our sample is 14.7 bps.

Third, we analyze how our findings compare to the predictions of various theories. Exploiting cross-sectional variation in institutional orders, we find strongest support for HFTs "back-running" on informed orders. Yang and Zhu (2015) develop an equilibrium model in which intermediaries learn about an investor's informed order in the first trading round to trade along with it in a second round.[6] The investor is fully aware of this risk of detection. In equilibrium, the investor trades off

[6]Boulatov, Bernhardt, and Larionov (2016) propose a model that also generates back-running but considers the exogenous price impact function of Almgren and Chriss (1999); however, they go beyond two periods and analyze Nash strategies in a continuous-time setting.

hiding more of the order in the early round (thus reducing the risk of detection), against a higher total execution cost due to less trading with early-round noise traders. We find support for back-running, since (i) HFTs only exhibit with–wind trading late in the lifetime of an institutional order and (ii) the permanent price impact is larger for orders that they decide to run on, controlling for standard covariates.

These results are inconsistent with predatory trading ("front-running," in a general sense) as proposed by Brunnermeier and Pedersen (2005). Their predatory trading model focuses on an *uninformed* institutional investor who is in need of trading, whereas the institutional investors in our data seem to be informed.

Finally, we find empirical support for HFT market making, as in Grossman and Miller (1988). The against–wind HFT flow observed early in the lifetime of institutional orders and in the placebo sample suggest that HFTs generally trade against price pressures (consistent with what Brogaard, Hendershott, and Riordan (2014) find for U.S. equities). The negative correlation of against–wind HFT flow and investor transaction costs is also consistent with market making. However, we do not find a significant correlation between against–wind flow and the *transitory* component of price impact; market making predicts it to be negative.

A couple of additional findings are worth mentioning. First, HFTs do not seem to detect large long-lasting institutional orders right from the start, since they also lean against the wind initially for these orders. Second, HFT gross trading revenue is \$364, on average, for the stock–days when at least one of the institutions is executing an order. We find that, controlling for standard covariates, a one standard deviation against–wind flow raises their profit by \$795. A one standard deviation with–wind flow raises it by \$1,244. The latter is 100%*(1,244/3,024)=41% of the institution's increase in implementation shortfall for a one standard deviation with–wind flow. Third, for the institutional orders that HFTs detected, we find that, prior to detection, the institutional investor traded quite aggressively, since both the investor's participation in volume and trade size are above the investor's overall average. Fourth, HFTs are not a homogeneous group, in spite of each of them *individually* trading against the wind when they collectively trade against the wind. The same observation holds for their with–wind trading. However, HFTs differ significantly in terms of *how much* they trade against or with the wind. Some lean more toward against–wind trading and others more toward with–wind trading.

Contribution relative to contemporary papers. Our paper contributes to three contemporary papers on trading by institutional investors and HFTs. Korajczyk and Murphy (2014) study how HFTs trade around large orders for a Canadian sample. They assign HFT labels based on trader behavior and infer large orders from aggregate flow by broker–client account. They document against–wind flow initially and with–wind flow later in the course of large order executions. They further find that the implementation shortfall paid on large orders is higher for "stressful trades," that is, trades for which HFT liquidity provision is predicted to be lower.

Our study differs from that of Korajczyk and Murphy (2014) in three ways. First, we identify end user net flow as opposed to client flow per broker. End users often use multiple brokers to execute their orders (Linnainmaa and Saar, 2012). Second, we observe HFT names in the sample and therefore do not need to rely on inference based on behavior. Such inference is necessarily imperfect, since speed, for example, is one important HFT characteristic that is hard to observe in trade and quote data. Third, we complement their results on HFT liquidity provision by focusing on substantial with–wind trading by HFTs. Inspired by theory, we document that such with–wind activity could be interpreted as back-running on informed institutional orders. Cross-sectional analysis supports such conjecture; orders associated with more with–wind HFT activity exhibit greater long-term price impact (and unchanged transitory price impact).

The second contemporary paper is that of Tong (2015), who relates the average implementation shortfall to high-frequency trading intensity for a U.S. equity sample. The author averages across all institutional investors who participate in the Abel/Noser dataset and documents that high shortfall days coincide with days of high HFT intensity, both for HFT "market making" and HFT "directional trades." The benefit of our dataset is that it has *intraday* timestamps (as opposed to daily timestamps) and identifies HFTs by name (as opposed to an exchange-labeled category). This allows us to study *exactly* how HFTs trade during the lifetime of an order. Contrary to Tong (2015), we find that HFT market making lowers institutional cost.[7]

[7] Tong (2015) identifies HFT market making by mean reversion in cumulative net flow, whereas we identify it as

The third related paper is that of Hirschey (2014), who documents that current-second HFT aggressive flow predicts non-HFT aggressive flow in the next 30 seconds. We add to the author's findings in two ways. First, we focus on institutional investor flow, which is a subset of non-HFT flow. The latter also contains flow by other intermediaries, for example, proprietary trading by sell-side banks. Second, our results suggest that HFTs respond to investor flow at low frequencies. They suggest that HFTs should be thought of more broadly than as machines engaged in sub-millisecond speed races. In fact, it is likely that their superior information technology enables them to better generate signals from big data assembled over multiple hours and trade on it profitably.

Some words of caution. Our study faces several limitations and we therefore caution against overinterpretation of the results. First, we do not claim causality. All the results are either unconditional or conditional (regression) associations. We nevertheless believe that the interaction of a set of sophisticated intermediaries with large institutional investors should be of general interest. Moreover, the empirical patterns being more supportive of one theory (back-running) over others helps us understand more deeply the functioning of securities markets.

Second, HFTs seem to engage in back-running, but they might not be the only ones. The data identify HFTs, but not other important intermediaries such as the proprietary trading desks of sell-side banks or hedge funds. In the data, such intermediaries are grouped together with end user flow in the reporting by sell-side banks. In fact, these alternative intermediaries could engage in any of the three hypothesized behaviors: market making, predatory trading, or back-running. There simply is no way for us to tell.

Third, back-running by intermediaries most likely plagued end users long before the rise of HFTs (Harris, 1997). Intermediaries have existed for centuries. One could even argue that back-running was easier in human-intermediated markets because broker–dealers observed their client flow directly, since it had to pass through their hands. They were not allowed to trade on this information but enforcement was extremely difficult (traders could tip each other off). Our contribution is simply to document evidence consistent with back-running, without claiming it is somehow

leaning against an order. Note that predatory trading on institutional orders also implies mean reversion in net position (Brunnermeier and Pedersen, 2005). For this reason, we prefer to identify market making as leaning against an order.

unique to HFTs operating in electronic markets. Indeed, there could very well be *less* of it now. Finally, in a broader sense, cream skimming could also be interpreted as back-running. For example, a New York Stock Exchange specialist would only improve the price if he suspected the order originated from an uninformed investor, thereby raising the price impact for informed investors in equilibrium (Rock, 1990). Another example is dealers or trading locals paying for retail order flow (Easley, Kiefer, and O'Hara, 1996).

Fourth, back-running is not necessarily costly to end users as a group. If the HFT sector is competitive, HFTs could willingly incur a (net) loss when initially trading against the wind. Doing so, they might discover large long-lasting orders on which they can then profitably back-run. This implies a transfer from large long-lasting orders to small institutional orders or to retail orders.

2 Price and Trade Patterns, as Predicted by Theory

In this section, we summarize several priors for trading by HFTs around large institutional orders. Figure 2 depicts the expected price pattern and HFT cumulative net flow in the lifetime of an institutional buy order and shortly thereafter. The three panels correspond to market making (Grossman and Miller, 1988), predatory trading (Brunnermeier and Pedersen, 2005), and back-running (Yang and Zhu, 2015). These predictions are briefly discussed in the next three paragraphs.

Grossman and Miller (1988) predict that market-making HFTs sell to an institutional buy order and close out their position in the long run. The price rises when the institutional investor buys in order to compensate risk-averse market makers for supplying "immediacy." Market makers will offload their position at fundamental values *eventually* but incur a price risk on their inventory in the short run. Selling to the institutional buyer at a (temporarily) elevated price compensates them for such risk. The identifying features are therefore that HFTs sell in the lifetime of the institutional buy order and the price impact is transitory (see Panel (a) of Figure 2).

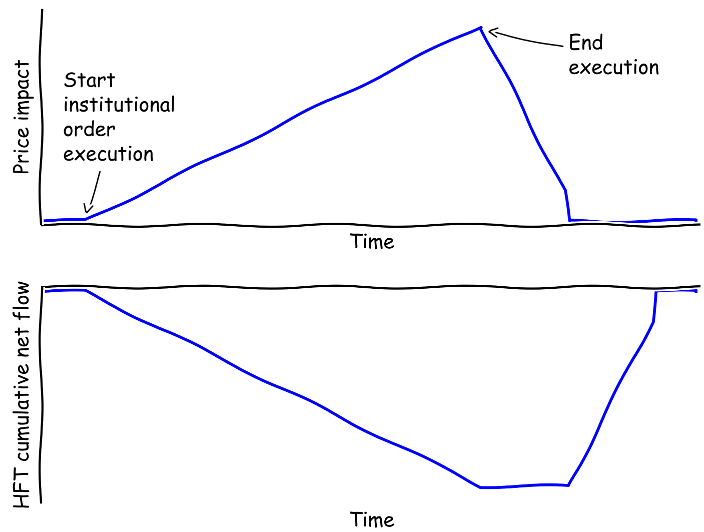
Brunnermeier and Pedersen (2005) model strategic trading by predators who are aware that an institutional investor is forced to trade for liquidity reasons. Their findings imply that predatory HFTs first trade along with an institutional buy order and thus add to price pressure, but eventually turn around and sell to the institutional order. Since HFTs enter positions at a price below the price at which they exit, they earn a positive profit in expectation. The institutional investor suffers, since price pressure is stronger than what it otherwise would have been. Predatory trading is front-running in a general sense. The key features are as follows: HFTs trade along with the institutional order initially and unwind their position at elevated prices and the price impact is transitory (see Panel (b) of Figure 2).

Yang and Zhu (2015) propose back-running as intermediaries learning about the execution of a privately informed order. If HFTs could sniff out those orders, they would trade along with them. This is profitable for them, since these informed orders continue to execute until the privately known fundamental value is reached. If HFTs can unwind their position after completion of the institutional order, they earn positive expected profit. Institutional investors suffer as HFTs effectively take part of their "informational rent." That is, prices are pushed to fundamental values more quickly when HFTs join the trade. The key features are as follows: Buying by HFTs is delayed due to an initial learning period and the price impact is permanent (Panel c).

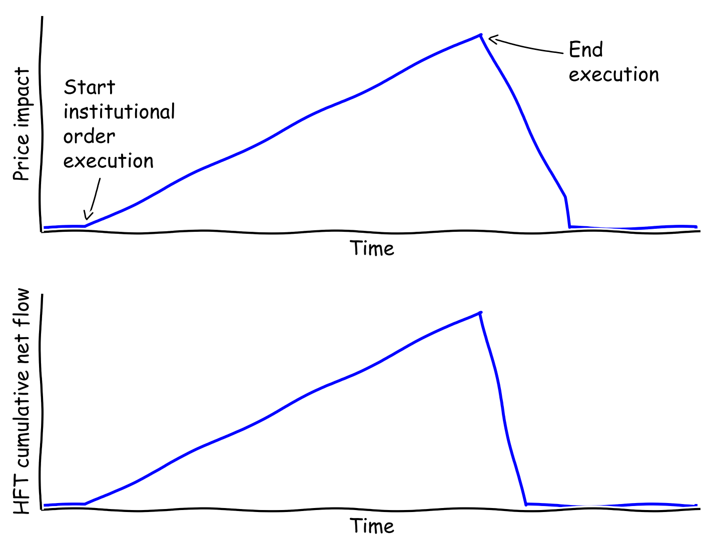
Figure 2: Price and trade patterns around institutional buy orders predicted by theory

This figure schematically depicts three possible price and trade patterns around institutional buy orders, as predicted by theory. Each panel summarizes the predicted price pattern in the lifetime of an institutional order and shortly thereafter (top graph). It further shows how intermediaries (i.e., HFTs) trade by plotting their cumulative net flow in this period (bottom graph). The three panels correspond to market making (Grossman and Miller, 1988), predatory trading (Brunnermeier and Pedersen, 2005), and back-running (Yang and Zhu, 2015), respectively.

(a) Market making (Grossman and Miller, 1988)



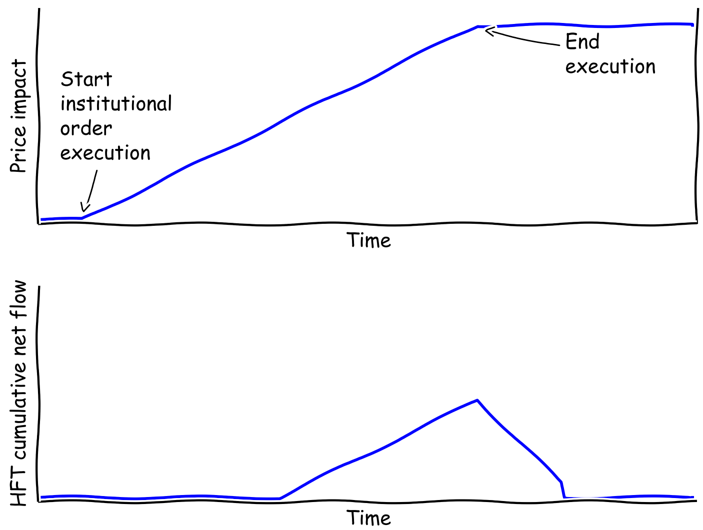
(b) Predatory trading (Brunnermeier and Pedersen, 2005)



Continued on next page.

(Figure 2 continued)

(c) Back-running (Yang and Zhu, 2015)



3 Trade Environment and Data

This section describes the NASDAQ OMX trading environment and presents the public and proprietary datasets. The public dataset contains trades with exchange member identities that are used to identify HFT trades. The proprietary dataset contains the child order execution records of four large institutional investors. These investors cannot be identified in the public data, since they are not exchange members. They use brokers to route their orders to the exchange. Both datasets pertain to trading in the 30 Swedish index stocks from January 1, 2011, through March 31, 2013.

3.1 Trading on NASDAQ OMX

NASDAQ OMX runs mostly like a standard limit-order market to trade their Swedish stocks. The most notable idiosyncratic feature is *ex post* trade transparency on who traded. Trade records that are revealed in real time contain the usual fields, that is, a time stamp in milliseconds, a transaction price, and a transaction quantity. However, at the end of each trading day, NASDAQ OMX also reveals who was trading with whom for each transaction. This identification is carried out at the exchange member level and therefore does not reveal end users. HFTs, banks, and brokers are exchange members, but not institutional investors who trade through banks or brokers. A total of 89 exchange members were active in our sample.

At the time of our sample, NASDAQ OMX faced competition from other regular exchanges and multilateral trading facilities (e.g., dark pools). Its market share for exchange-traded volume was 65%. The most active rival exchange, Chi-X, had a 20% market share. The remaining 15% was shared by five other exchanges. An important caveat of this study is that we do not observe the trades by HFTs on these alternative markets. We revisit this issue in the robustness analysis presented in Section 7.3.

3.2 Public and Proprietary Data

Public data. Two sets of public data are used in this study: equity transactions with member identification and index future returns. Both are obtained from the Thomson Reuters Tick History. As mentioned in the introduction, we use member identities to identify the aggregate net flow of the 10 largest HFTs.

Proprietary data. The proprietary data consist of the child order transactions of four large institutional investors highly active in Swedish index stocks. The data contain detailed NASDAQ OMX execution data, that is, a second time stamp, price, and quantity.

The child order transactions are aggregated at the a stock–day-institution level. For each stock and each day, all child order executions by a single institution are aggregated into an institutional meta-order. The rationale for constructing meta-orders is that an execution desk at the institution receives orders from different portfolio managers and will internally match buy and sell orders. It will therefore worry about obtaining the best execution on *net* flow at the institutional level. We further define meta-orders at a daily frequency, as opposed to lower frequencies, since we are interested in the trading behavior of HFTs. The latter are known to make intraday round trips and prefer to “go home flat.” We refer to institutional meta-orders as institutional orders in the remainder of the manuscript for brevity.

Two filters are applied to establish the sample used in all the analyses. First, institutional orders with low directionality are excluded, since the focus is on orders that built a position in the course of the day (as opposed to intra-day trading strategies). Directionality is based on net flow and defined as the absolute value of the difference between the buy and sell volumes, divided by total volume (all in shares).[8] Orders with directionality below 0.90 are removed from the sample. This filter removes 11.5% of orders. We find that 95% of the remaining orders consist of either purely buys or purely sells. Second, HFT net flow is winsorized at the 1% and 99% levels. This takes care of extreme outliers in this variable.

3.3 Summary Statistics

Table 1 presents various summary statistics. The top panel shows the trading activity of the four institutional investors, the HFTs, and the market at large. We report statistics for all institutional orders combined, as well as separately for institutional buy orders and institutional sell orders.

If an institution trades on a particular stock–day, it trades 140,000 shares, on average. This corresponds to an average order size of $1.940 million. Expressed relative to the average daily volume, the order size is 4.0%. The meta-orders are extremely directional, with an average of 1.00 for both buy and sell orders. These levels are therefore far above the 0.90 threshold we set as a filter.

HFT participation in shares is 22.2% on stock–days when an institution is active. It is 28.5% in terms of the number of trades, which implies that their trade size is slightly smaller than average. In dollar terms, their trade size is $8,625. HFTs strongly mean-revert their positions intradaily, as indicated by the low average HFT directionality of 0.08. Mean reversion is even stronger across days, since the average daily HFT *net* flow is 900 shares out of the 1.70 million shares they trade, on average. This result is in sharp contrast to that for the institutions, who are expected to have

[8] It is inspired by the imbalance measure of Chordia and Subrahmanyam (2004). The precise definition of directionality is |S-B|/(S+B), where B and S are the buy and sell share volume, respectively.

Table 1: Summary statistics

This table presents the mean and standard deviation (in brackets) of various trade variables for the active institution (inst), the group of HFTs (HFTs), and the entire market (all). The sample consists of days when an institution executed an order in a Swedish index stock between January 1, 2011, and March 31, 2013. The leftmost column briefly describes the variable, with the unit of measurement in brackets. Most are self-explanatory, except the following: net flow is buys minus sells, directionality is defined as |buys-sells|/(buys+sells), ADV is average daily volume based on the full sample, implementation shortfall is defined as what was paid to execute a buy order minus what would have been paid had the price on all child transactions been equal to the midquote price at the time the order started executing (equivalent definition for a sell order; the relative implementation shortfall is obtained by dividing by the total order size), and HFTs' gross trading revenue pertains to their trading in the lifetime of an institutional order, where any nonzero position at the end time is valued at the end-of-day price. Weighted averages are obtained by weighting by order size. Swedish kronas are converted to U.S. dollars at the sample average exchange rate.

	All inst orders			Inst buy orders			Inst sell orders		
	Inst	HFTs	All	Inst	HFTs	All	Inst	HFTs	All
Avg volume (10,000 shares)	14	170	767	13	179	807	17	155	698
	(39)	(184)	(877)	(36)	(200)	(974)	(43)	(152)	(679)
Avg net flow (10,000 shares)	1.62	0.09	0.00	12.70	0.10	0.00	-17.09	0.08	0.00
	(41)	(22)	(0)	(35)	(22)	(0)	(43)	(21)	(0)
Avg directionality (based on shares)	1.00	0.08	0.00	1.00	0.07	0.00	1.00	0.08	0.00
	(0.01)	(0.07)	(0.00)	(0.01)	(0.07)	(0.00)	(0.01)	(0.08)	(0.00)
Avg volume (#trades)	139	2,910	10,216	130	3,055	10,810	153	2,665	9,212
	(248)	(2,009)	(6,271)	(235)	(2,178)	(6,690)	(267)	(1,656)	(5,342)
Avg net flow (#trades)	25	-8	0	129	-5	0	-151	-12	0
	(281)	(307)	(0)	(235)	(299)	(0)	(265)	(321)	(0)
Avg directionality (based on #trades)	0.99	0.08	0.00	1.00	0.07	0.00	0.99	0.08	0.00
	(0.01)	(0.07)	(0.00)	(0.01)	(0.07)	(0.00)	(0.01)	(0.07)	(0.00)
Avg volume ($100,000)	20	251	1126	18	258	1164	25	239	1063
	(47)	(194)	(1,034)	(45)	(209)	(1,166)	(51)	(164)	(757)
Avg net flow ($100,000)	1.94	0.76	0.00	17.58	0.85	0.00	-24.47	0.61	0.00
	(51)	(35)	(0)	(44)	(36)	(0)	(51)	(33)	(0)
Avg directionality (based on $100,000)	1.00	0.08	0.00	1.00	0.07	0.00	1.00	0.08	0.00
	(0.01)	(0.07)	(0.00)	(0.01)	(0.07)	(0.00)	(0.01)	(0.08)	(0.00)

Continued on next page.

(Table 1 continued)

	All inst orders			Inst buy orders			Inst sell orders		
	Inst	HFTs	All	Inst	HFTs	All	Inst	HFTs	All
Avg duration (hours)	3.73			4.09			3.11		
	(3.20)			(3.19)			(3.13)		
Avg order size ($1,000)	1,940			1,758			2,447		
	(5129)			(4,442)			(5,128)		
Avg order size relative to ADV (%)	4.0			3.5			5.0		
	(7.6)			(7.2)			(8.2)		
Avg nr child trades	135			124			153		
	(231)			(221)			(246)		
Avg imp shortfall ($)	2,860			2,464			3,511		
	(33,949)			(34,711)			(32,653)		
Avg imp shortfall (bps)	7.4			6.5			8.8		
	(59.6)			(67.5)			(43.8)		
Wgt avg imp shortfall (bps)	14.7			14.7			14.7		
	(72.6)			(82.4)			(59.2)		
Avg gross trading revenue ($)	364			429			258		
	(3,013)			(3,418)			(2,187)		
Number of observations	5,910			3,675			2,235		

longer trading horizons.

Table 1 further reveals various characteristics of the institutional orders. An order generates 135 child trades, on average, in a time span of 3.73 hours. The average implementation shortfall in dollars is $2,860, which corresponds to 14.7 bps (an average dollar shortfall is effectively an order size-weighted relative shortfall). The equally weighted average relative shortfall is only 7.4 bps, which is not surprising, given that larger orders are generally more expensive to execute.

4 HFT Net Flow during the Execution of Institutional Orders

In this section, we investigate whether HFTs "lean against the wind" or "go with the wind" while an institutional order executes through a series of child orders. We also set up a placebo sample to study whether HFT behavior is really due to the presence of an institutional order or is simply the result of market conditions that prevailed on the stock–days when institutions implemented their orders.

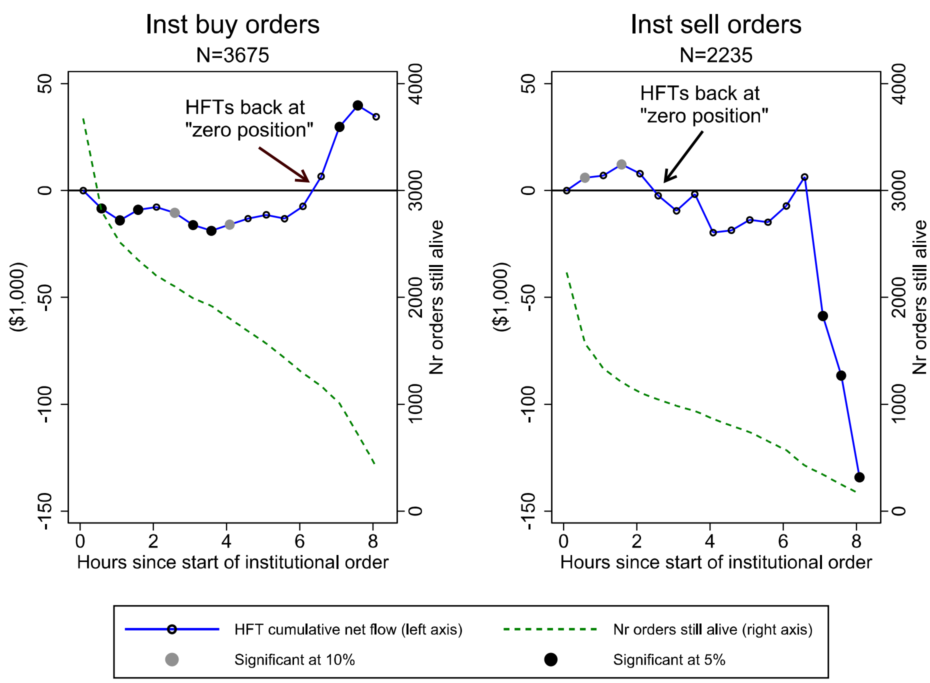
4.1 HFT Net Flow in the Lifetime of an Institutional Order

In this subsection, we plot how HFT net flow develops in the lifetime of an institutional order. We track HFT net flow cumulatively, starting from the first institutional child trade. Snapshots are taken every 30 minutes. At the time of the snapshot, HFT net flow is averaged over all institutional orders that are still alive. This implies that, for the first half hour, for example, the average is taken across short- and long-lasting orders. We consider this a reasonable approach, since HFTs are unlikely to know the order duration *ex ante* (a view supported by the data, as shown later in Table 3). If the average net flow is negative, then HFTs lean against the order (against the wind). If it is positive, then they trade along with it (with the wind). An equivalent analysis is done for institutional sell orders.

Figure 3 contains the results for buy orders (left panel) and sell orders (right panel). We observe that HFTs lean against buy orders in the first six hours of execution, since they go short for approximately $20,000. This result is statistically significant only in the first four hours, as indi-

Figure 3: HFT net flow in the lifetime of an institutional order

This figure plots average HFT net flow from the start of an institutional order to various time points, sampled at 30-minute intervals. The average is taken across all orders that are still active at the end point of the interval. The size of the sample at each time point is indicated by the dashed green line (right axis). Statistical significance is established based on the *t*-value of the mean across stock–institution fixed effects (same as the overall mean), with residuals clustered at the stock–day level.



cated by the solid dots. The right panel shows that HFTs lean against sell orders only in the first two hours. This result is statistically weaker (at only a 10% significance level).

Strikingly, HFTs turn around and go with the order if it lasts more than six hours for buys and more than two hours for sells. The result is statistically significant only if the order execution lasted more than seven hours. After eight hours, HFTs are long $39,900 for buy orders and short $134,000 for sell orders. The with–wind behavior is stronger for sell orders, since HFTs switch earlier and obtain larger with–wind positions. We suspect this is due to the higher execution intensity of sell orders. Indeed, Table 1 reveals that sell orders are 39.2% larger and almost an hour shorter, on average (3.11 hours instead of 4.09 hours).

The dashed green line shows the number of institutional orders still alive at each point in time. For buy orders, we observe significant with–wind flow for the 1,010 orders that are still alive after seven hours. This is about a third of all buy orders. For sell orders, there is significant with–wind flow for 343 orders, which is about a sixth of all sell orders.

In an unreported figure, we plot the institutional net flow over the lifetime of the order (the equivalent of the HFT net flow plot of Figure 3). It reveals that institutional positions build up at an almost linear rate for both buy and sell orders. It further shows that institutional sell orders execute more aggressively, since the slope is about 25% larger in magnitude.

4.2 HFT Net Flow Relative to Placebo Days

A placebo sample is created to identify whether the HFT flow pattern is related to the institutional order or to market conditions that prevailed at the time of the order.

The matching procedure. The placebo sample is constructed by matching each stock–day when one of our institutional investors was active to a similar stock–day for the same stock but when none of our four institutional investors were active. We proceed as follows. The placebo stock–day is selected based on matching four trade variables across two periods:

1. From market open until the first child trade of the order in the "treated" sample
2. During the lifetime of the order, that is, from the first to the last child order execution.

The period from market open until the first execution is added to account for potential endogenous timing by the institution regarding the start of the order (Hendershott, Jones, and Menkveld, 2013). The four trade variables used are the volume, market return, idiosyncratic return (with a beta obtained from Reuters), and realized volatility (based on one-minute midquote returns). A "nearest-neighbor" matching procedure is used. We follow Davies and Kim (2009, p. 183) with one modification: The distance is measured in standard deviation units as opposed to percentages.[9]

Note that the placebo analysis controls for a momentum-based explanation of HFT behavior. If institutional orders have a price impact and if HFTs trade on momentum, then a with–wind pattern occurs naturally. However, if the with–wind HFT pattern is present in the original sample *minus* the placebo sample (i.e., in the differential), then this alternative explanation becomes less likely. Both samples are constructed to have the same price pattern in terms of market return and idiosyncratic return.

Table 2 presents trade statistics for both the (treated) institutional trade sample and the placebo sample. It shows that the match seems reasonable in terms of distance. Its value is small and evenly distributed across all matched variables.

Results of the matched sample analysis. Figure 4 plots the HFT net flow for the treated sample and the placebo sample. The treated sample line is the same as in Figure 3. The only difference is that the dots now denote significance of the HFT net flow tested against the placebo HFT net flow, as opposed to zero. In other words, it pertains to a test on whether the *differential* between the two HFT net flows is zero.

In the placebo sample, HFT net flow mostly leans against the wind. It is negative for buy stock–

[9]The relative distance measure of Davies and Kim (2009) is inappropriate when matching on returns because the distance gets large when returns approach zero (division by (almost) zero). We therefore choose to express distance in standard deviation units. Specifically, for the eight matching variables x^k and for treated observation i, we select the matching observation as follows:

$$\underset{j\in N}{\operatorname{argmin}} \sum_{k=1}^{8} \left(\frac{|x_i^k - x_j^k|}{\sqrt{1/N \sum_j \left(|x_i^k - x_j^k|\right)^2}} \right).$$

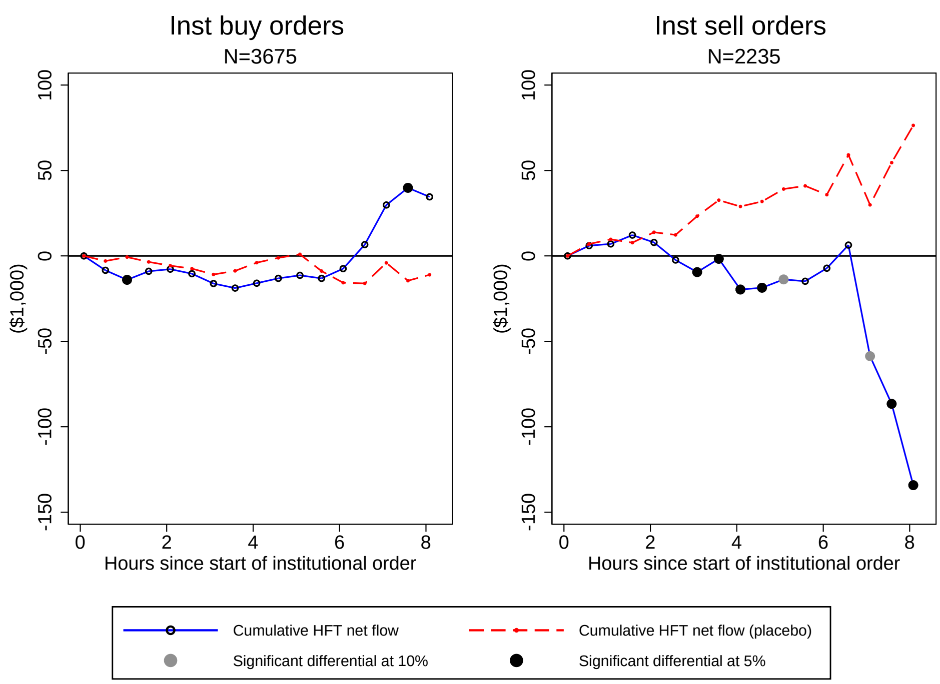
Table 2: Quality of the match with a placebo sample

This table presents the mean and standard deviation of four variables that were used to construct a placebo sample. The main sample consists of stock–days when one of the institutions executed an order. Each such stock–day is matched to a stock–day without an order by any of the four institutional investors. The match is done by a nearest-neighbor algorithm. It uses the average of four trade variables computed for two time intervals: (i) from the market open until the start of the order and (ii) from the start of the order until the end of the order. The four trade variables are the dollar volume, index return, the stock's idiosyncratic return, and realized volatility (based on one-minute midquote returns). The distance is the average distance across all variables in a particular time period. Distance is expressed in standard deviation units.

	Sample stock–days				Placebo stock–days				
	Volume	r^{Index}	r^{Idio}	Volatility	Volume	r^{Index}	r^{Idio}	Volatility	Distance
Panel A: Institutional buy orders									
Mean (open–start)	11.3	3.5	-0.2	6.3	10.5	2.3	0.4	5.3	1.4
St dev (open–start)	20.7	75.8	75.6	15.5	18.3	60.8	64.1	10.8	
Mean (start–end)	17.2	2.6	6.3	10.0	16.9	0.0	6.0	8.2	1.5
St dev (start–end)	25.5	93.8	73.0	23.6	24.3	69.0	58.5	18.1	
Panel B: Institutional sell orders									
Mean (open–start)	10.5	0.5	-2.6	3.9	10.2	0.9	-0.4	3.8	1.2
St dev (open–start)	17.9	46.0	71.0	6.7	16.1	41.2	62.5	5.6	
Mean (start–end)	12.4	-2.1	-9.6	4.5	12.2	-0.8	-7.5	4.2	1.2
St dev (start–end)	20.6	53.4	53.9	15.3	19.2	46.5	44.7	12.9	

Figure 4: HFT net flow on placebo stock–days

This figure plots average HFT net flow for the placebo sample (and the main sample for reference). It echoes Figure 3 for the sample of stock–days when one of the four institutional investors executed an order. The placebo sample consists of stock–days that match these treated stock–days in terms of trading conditions (volume, index return, idiosyncratic return, and volatility) but did not feature trades by any of these four investors. Statistical tests pertain to the differential across treated and placebo stock–days. These are done based on the *t*-value of the mean across stock–institution fixed effects (same as the overall differential mean), with residuals clustered at the stock–day level.



days and positive for sell stock–days. When compared to the placebo sample, the HFT against–wind pattern in the treated sample becomes mostly insignificant, whereas the against–wind pattern becomes statistically significant more often. These findings suggest that the against–wind pattern is due to market conditions, whereas the with–wind pattern seems truly related to the presence of the institutional order.

One interpretation of these placebo results is that HFTs use a market-making strategy in normal market conditions. In the treated sample, institutional orders have a price impact (see Table 2), that is, the average idiosyncratic return is positive for buy orders and negative for sell orders. Given that the placebo sample is matched on this variable, it seems HFTs trade against price changes. They sell when prices go up and buy when prices go down, at least initially. The wedge between the two lines suggests that HFTs switch from market making to speculation when they detect a persistent, directional long-lasting order. They stick to market marking in the placebo sample.

It is important to stress that, while there are no order executions by the four identified institutional investors in the placebo sample, other institutional investors could have executed similar orders. We are not overly worried, since this would bias against us finding anything. In other words, this observation implies not only that the results that we *do* find are really there, but also that they underestimate the true strength of the effect.

4.3 Do HFTs Detect Large Long-Lasting Orders Early?

In this subsection, we repeat the HFT net flow analysis for various sub-samples of the data. Specifically, we split the institutional order sample into a small- and a large-order sample to study whether HFTs have the ability to detect large long-lasting orders early. The results in Table 3 show that this does not seem to be the case. HFTs also lean against large long-lasting orders initially. Panel A shows that HFTs lean significantly against buy orders in the first hour. They lean against such orders also when they are either larger than \$1 million or larger than the median in terms of the percentage of the average daily volume (ADV). If anything, the against–wind pattern is larger in magnitude for such orders and more significant.

Panel A of Table 3 further shows that HFTs also lean against long-lasting buy orders in the

Table 3: Large long-lasting orders

This table presents the means of HFT net flows and their *t*-statistics for institutional buy orders and institutional sell orders. These averages were calculated in the lifetime of the order, from the start until one hour later, from the start until two hours later, and so forth. This table is similar to Figure 3, but adds to it by showing the pattern for various sub-samples based on the characteristics of the institutional order. Statistical significance is established based on the *t*-value of the mean across stock–institution fixed effects (same as the overall mean), with residuals clustered at the stock–day level. The superscripts ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Hours since first child order execution							
	1	2	3	4	5	6	7	8
Panel A: HFT net flow mean in the lifetime of institutional buy orders								
All	-14.0***	-7.7	-16.2**	-16.0*	-11.4	-7.4	29.8**	34.6
t-Stat	(3.9)	(1.5)	(2.4)	(1.9)	(1.1)	(0.6)	(2.1)	(1.4)
N	2,516	2,203	1,992	1,803	1,569	1,290	1,010	418
Size ≤ median %ADV	-9.7*	-9.6	-28.6***	-32.4**	-24.1	-7.9	28.0	32.0
t-Stat	(1.7)	(1.2)	(2.7)	(2.5)	(1.5)	(0.4)	(1.3)	(0.8)
Size > median %ADV	-17.3***	-6.4	-6.9	-3.4	-2.2	-7.1	30.9*	35.9
t-Stat	(3.6)	(1.0)	(0.8)	(0.3)	(0.2)	(0.5)	(1.7)	(1.1)
Size ≤ $1,000,000	-8.9**	-8.9	-24.3***	-30.0***	-26.3**	-23.1	8.5	8.7
t-Stat	(2.0)	(1.4)	(3.0)	(2.9)	(2.1)	(1.6)	(0.5)	(0.3)
Size > $1,000,000	-21.4***	-6.1	-4.5	4.8	9.8	14.1	58.1**	60.7
t-Stat	(3.5)	(0.7)	(0.4)	(0.3)	(0.6)	(0.7)	(2.4)	(1.5)
Duration < 4 hours	-14.3**	1.6	-16.6					
t-Stat	(2.0)	(0.1)	(0.6)					
Duration > 4 hours	-13.9***	-9.8*	-16.1**	-16.0*	-11.4	-7.4	29.8**	34.6
t-Stat	(3.3)	(1.8)	(2.3)	(1.9)	(1.1)	(0.6)	(2.1)	(1.4)
Panel B: HFT net flow mean in the lifetime of institutional sell orders								
All	7.0	7.9	-9.5	-19.7	-13.8	-7.2	-58.7**	-134.2**
t-Stat	(1.3)	(0.9)	(0.8)	(1.4)	(0.7)	(0.3)	(2.1)	(2.4)
N	1,334	1,110	986	863	740	571	343	161
Size ≤ median %ADV	13.8	18.3	-16.1	-0.4	10.3	56.6	-46.4	13.1
t-Stat	(1.4)	(1.2)	(0.7)	(0.0)	(0.3)	(1.0)	(0.7)	(0.1)
Size > median %ADV	4.6	4.5	-7.3	-25.9*	-20.8	-24.9	-61.7*	-176.6***
t-Stat	(0.7)	(0.5)	(0.6)	(1.7)	(1.0)	(0.9)	(1.9)	(2.9)
Size ≤ $1,000,000	6.2	12.6	-7.8	-8.8	-5.4	33.0	10.1	3.8
t-Stat	(0.8)	(1.0)	(0.4)	(0.4)	(0.2)	(0.8)	(0.3)	(0.0)
Size > $1,000,000	7.5	5.4	-10.4	-25.2	-17.6	-25.1	-86.2**	-191.1***
t-Stat	(1.0)	(0.5)	(0.7)	(1.5)	(0.8)	(0.8)	(2.4)	(2.9)
Duration < 4 hours	18.9**	55.4***	41.0					
t-Stat	(2.0)	(2.9)	(1.4)					
Duration > 4 hours	0.7	-5.1	-16.1	-19.7	-13.8	-7.2	-58.7**	-134.2**
t-Stat	(0.1)	(0.6)	(1.4)	(1.4)	(0.7)	(0.3)	(2.1)	(2.4)

first hour. Long-lasting orders are defined as orders with a lifespan of more than four hours (half a trading day). The result is statistically more significant and equal in magnitude when compared to the results for short-lived orders. Panel B reveals that HFTs lean against sell orders, large or long-lasting, but the results are statistically insignificant. Note, however, that for full hours, Figure 3 also shows that HFT net flow is insignificant for sell orders. The figure shows the weak significance of the against–wind pattern only for the half-hour and the one-and-a-half-hour time points.

5 Implementation Shortfall and HFT Gross Trading Revenue

In this section, we investigate the impact of HFT net flow—against–wind or with–wind—on institutional trading cost and HFT gross trading revenue.

5.1 Implementation Shortfall and HFT Net Flow

A standard measure of institutional trading cost is implementation shortfall. It is defined as

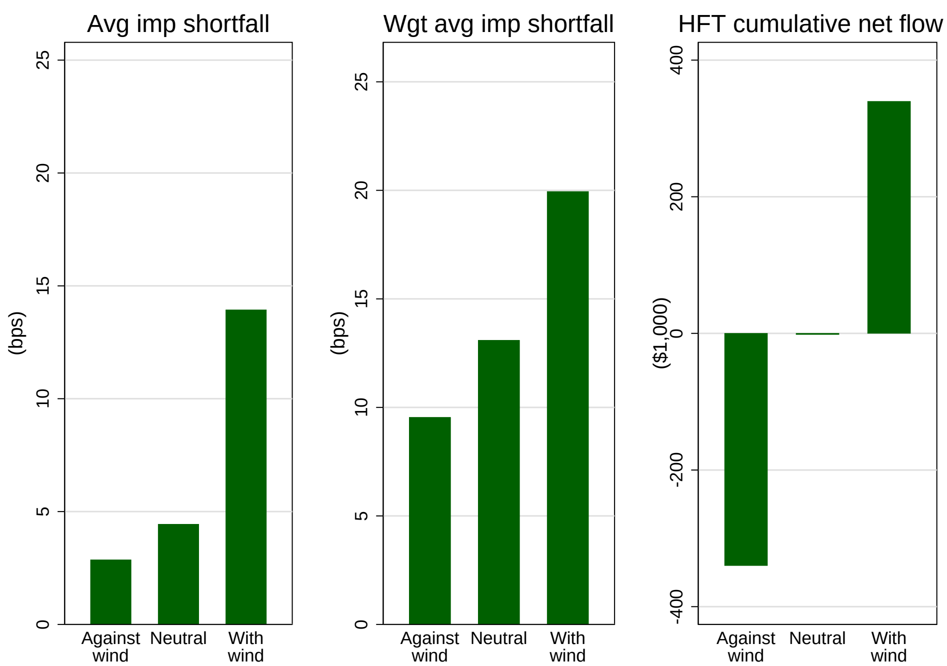
$$ImpShortfall_{ijt} = D_{ijt} \times (\log \overline{P_{ijt}} - \log P_{ijt}^{\text{Start}}), \quad (1)$$

for institutional investor i, stock j, and day t, where D_{ijt} is a buy–sell indicator that equals one for institutional buy orders and -1 for institutional sell orders, $\overline{P_{ijt}}$ is the average trade price on the order, and P_{ijt}^{Start} is the midquote price that prevailed at the start of the order. Implementation shortfall is expected to be positive, on average, since buy orders have a positive price impact and sell orders have a negative price impact that is multiplied by -1 in the definition. Implementation shortfall is defined in relative terms but often expressed in dollar terms by multiplying it with the dollar size of the order.

Before turning to regression analysis, it is useful to inspect whether there is any correlation between the two variables of interest, implementation shortfall and the nature of the HFT net flow. To that end, HFT net flow in the lifetime of an institutional *sell* order is first multiplied by -1. HFT net flows for these orders can then be meaningfully compared to those for buy orders. Negative HFT net flow can then be interpreted as against–wind trading and positive net flow as

Figure 5: Implementation shortfall by HFT net flow terciles

This figure plots the average implementation shortfall, the order size-weighted average implementation shortfall, and the average HFT net flow for the against–wind, neutral, and with–wind HFT net flow terciles. The terciles are created by first multiplying the HFT net flow during institutional sell orders by -1. All HFT net flow observations are then sorted and put into equal-sized bins. The tercile with the lowest values consists of strong against–wind HFT net flows, the middle tercile contains small HFT net flows in terms of size (we label this tercile neutral), and the tercile with the highest values contains strong with–wind HFT net flows.



with–wind trading. These net flows are sorted and binned into equal-sized terciles. The average implementation shortfall is then calculated for each tercile.

Figure 5 illustrates that implementation shortfall is lower when HFTs lean against the order, but higher when they go with the order. Implementation shortfall increases monotonically across the three terciles: It is 3.5 bps for the against–wind tercile, 4.2 for the neutral tercile, and 14.5 basis for the with–wind tercile. The order size-weighted implementation shortfall exhibits a similar pattern. One noteworthy observation is that the increase for with–wind flow is larger in magnitude than the decrease for against–wind flow. This is not simply due to HFT net flow being larger in magnitude for the with–wind bin, as the rightmost panel in the figure shows that they are equal in magnitude; both are about $275,000. Institutional investors seem to suffer more from with–wind HFT net flow than they benefit from against–wind HFT net flow.

For the regression analysis, we separate the two types of HFT net flow by creating the following two variables:

$$AgainstWindHFTCumNetFlow_{ijt} = 1_{\{\mathrm{sgn}(HFTCumNetFlow_{ijt}) \neq \mathrm{sgn}(InstOrder_{ijt})\}} \times \left|HFTCumNetFlow^*_{ijt}\right| \quad (2)$$

and

$$WithWindHFTCumNetFlow_{ijt} = 1_{\{\mathrm{sgn}(HFTCumNetFlow_{ijt}) = \mathrm{sgn}(InstOrder_{ijt})\}} \times \left|HFTCumNetFlow^*_{ijt}\right|, \quad (3)$$

where 1_A is the indicator function, that is, it equals one when A is true and zero otherwise; sgn(A) is the sign function, that is, it is +1 if A is positive, zero if A is zero, and -1 if A is negative; and $HFTNetFlow^*$ is the standardized HFT net flow.[10] These HFT net flow variables appear on the right-hand side of regressions, either in dollar terms or expressed relative to the stock's number of shares outstanding.[11]

The following panel regression is run to verify whether the general pattern of Figure 5 holds

[10]We standardize all right-hand-side variables to make the coefficients more easily interpretable. The sign of almost none of the HFT net flow observations changes, since the overall average HFT net flow is close to zero.

[11]We scale by the number of shares outstanding to make the results comparable across stocks. We borrow this argument from empirical studies that typically use turnover instead of volume as a control variable.

up when standard control variables are added. The model specification is

$$
\begin{aligned}
ImpShortfall_{ijt} &= \alpha_{ij} + \beta_1 AgainstWindHFTCumNetFlow_{ijt} \\
&\quad + \beta_2 WithWindHFTCumNetFlow_{ijt} + \gamma' X_{ijt} + \varepsilon_{ijt},
\end{aligned} \tag{4}
$$

where α_{ij} is shorthand notation for the addition of institution and stock fixed effects, X_{ijt} is a vector with control variables, and ε_{ijt} is a residual that is allowed to exhibit correlation within a stock–day but not across stock–days (standard errors are clustered at the stock–day level). We use two model specifications in the regressions. The first specification expresses implementation shortfall in basis points and HFT net flow relative to the number of shares outstanding. The second specification expresses both variables in dollar terms.

The control variables are standard market condition variables and order-specific variables (e.g., Anand, Irvine, Puckett, and Venkataraman, 2012). In particular, we add the size of the institutional order expressed relative to the average daily volume (ADV), the duration of the order, realized volatility, and stock turnover (share volume divided by the number of shares outstanding).

Table 4 presents the regression results. We observe that against–wind HFT flow is significant and negatively correlated with implementation shortfall, both for the relative shortfall and the dollar shortfall specification. A one standard deviation against–wind HFT flow ($235,000) reduces *relative* implementation shortfall by 3.4 bps, a reduction of 46% relative to the sample average of 7.4 bps. The coefficient of the *dollar* implementation shortfall is insignificant, but the point estimate is of similar magnitude.

With–wind HFT flow is significant and positively correlated with shortfall, both for the relative shortfall and the dollar shortfall. A one standard deviation HFT net flow ($241,000) increases the *relative* implementation shortfall by 12.5 bps, an increase of 169% relative to the sample average. A one standard deviation HFT net flow increases the *dollar* implementation shortfall by $3,024, an increase of 106% relative to the sample average of $2,860. The positive effect of with–wind flow is significantly larger than the negative effect of against–wind flow, as indicated by the *t*-test in the bottom of the table.

Table 4: Implementation shortfall regressed on HFT net flow and control variables

This table presents the panel regression results where implementation shortfall (IS) is the dependent variable. The main explanatory variable is the HFT net flow accumulated over the lifetime of the institutional order. It is expressed either relative to the number of shares outstanding (%) or in dollar terms ($). The values are then standardized and subsequently signed based on whether the direction is the same as the institutional order, in which case it is assigned a positive sign (with–wind), or the opposite, in which case it is assigned a negative sign (against–wind). Various variables are added as controls, all of which are standardized: ADV is the average daily volume based on the full sample. Order size and ADV are measured in shares. Turnover and volatility are measured from the start to the end of the order. Turnover is the daily stock volume divided by the number of shares outstanding and volatility is measured as realized volatility based on one-minute midquote returns. Also reported are the *p*-values of (i) a test of whether the against–wind coefficient equals minus the with–wind coefficient and (ii) a test of whether the model coefficients are indeed equal for institutional buy and sell orders (this table shows the pooled regression results). The regressions include stock and institution fixed effects and standard errors are clustered by stock–date. The *t*-values are in parentheses. The variable units are in brackets and reported right after the variable names. The superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Implementation shortfall (bps)	Implementation shortfall ($1,000)
Against–wind HFT Cum Net Flow (%)	-3.360**	
	(-2.1)	
With–wind HFT Cum Net Flow (%)	12.458***	
	(6.8)	
Against–wind HFT Cum Net Flow ($)		-0.516
		(-1.2)
With–wind HFT Cum Net Flow ($)		3.024***
		(4.8)
Order size relative to ADV(%)	5.439***	5.161***
	(6.1)	(9.5)
Order duration (hours)	-0.996	-0.821*
	(-0.9)	(-1.9)
Stock volatility (%)	-0.642	-0.099
	(-0.9)	(-0.6)
Turnover (%)	-0.185	0.994
	(-0.1)	(1.4)
Observations	5,910	5,910
R-Squared	0.042	0.115
p-value “WW=-AW”	0.001	0.005
p-value “Buys=Sells”	0.205	0.465

The statistically significant control variables have the expected signs. The strongest covariate is order size as a percentage of average daily volume. A one standard deviation increase raises shortfall by 5.4 bps, or \$5,161. In the dollar shortfall models, the duration of the order also significantly affects implementation shortfall. A three-hour-longer order (one standard deviation) reduces shortfall by \$821, a reduction of 29%. All else being equal, spreading the order over a longer horizon reduces execution costs. The coefficients of volatility and turnover are insignificant.

Finally, we test whether the regression coefficients differ between institutional buys and sells, that is, whether we are allowed to pool institutional buy and sell orders in one regression (as we did). The p-values in Table 4 reveal that, indeed, the null hypothesis that all coefficients are equal cannot be rejected. The p-values are 0.205 and 0.465 for the relative and dollar models, respectively.

5.2 HFT Gross Trading Revenue and HFT Net Flow

The previous section showed that HFT net flow is strongly correlated with institutional trading cost. Do HFTs make money off such behavior and, if so, how much? To this end, we repeat the regressions of the previous section but use HFT gross trading revenue (GTR) as a dependent variable to proxy for their gross profit.

We calculate HFT gross trading revenue (GTR) over the lifetime of the order according to Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010). It is a simple accounting exercise, where HFTs start off with no position in the stock and zero cash, they buy and sell the stock along the way (borrowing at zero cost), and at the end time any nonzero position in the stock is converted into cash using the end-of-day stock price. HFTs' gross trading revenue is defined as the cash position that they then have at the end of the period. The variable GTR is measured in dollars but it can also be expressed relative to the total amount HFTs traded in the stock. The measure then indicates how much HFTs make on each dollar they trade. Both the dollar and relative GTR values are used in the regressions (in parallel to the approach for the implementation shortfall regressions).

Table 5 presents the regression results. The relative GTR value does not seem affected by

Table 5: HFT gross trading revenue regressed on HFT net flow and control variables

This table presents the panel regression results where the dependent variable is HFT gross trading revenue in dollar terms or expressed relative to the HFT dollar volume. The main explanatory variable is HFT net flow accumulated over the lifetime of the institutional order. It is expressed either relative to the number of shares outstanding (%) or in dollar terms ($). The values are then standardized and subsequently signed based on whether the direction is the same as that of the institutional order, in which case it is assigned a positive sign (with–wind), or the opposite, in which case it is assigned a negative sign (against–wind). Various variables are added as controls, all of which are standardized: ADV is the average daily volume based on the full sample. Order size and ADV are measured in shares. Turnover and volatility are measured from the start to the end of the order. Turnover is the daily stock volume divided by the number of shares outstanding and volatility is measured as realized volatility based on one-minute midquote returns. Also reported are the p-values of (i) a test of whether the against–wind coefficient equals minus the with–wind coefficient and (ii) a test of whether the model coefficients are indeed equal for institutional buy and sell orders (this table shows the pooled regression results). The regressions include stock and institution fixed effects and standard errors are clustered by stock–date. The t-values are in parentheses. The variable units are in brackets and reported right after the variable names. The superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	HFT gross trading revenue ratio (bps)	HFT gross trading revenue ($1,000)
Against–wind HFT Cum Net Flow (%)	-0.289	
	(-0.6)	
With–wind HFT Cum Net Flow (%)	0.517	
	(1.6)	
Against–wind HFT Cum Net Flow ($)		0.795***
		(2.7)
With–wind HFT Cum Net Flow ($)		1.244***
		(3.3)
Order size relative to ADV(%)	0.024	-0.048
	(0.1)	(-0.5)
Order duration (hours)	-0.081	-0.174
	(-0.2)	(-1.3)
Stock volatility (%)	-0.156	-0.155
	(-1.0)	(-1.1)
Turnover (%)	0.549**	0.487***
	(2.0)	(2.7)
Observations	5,910	5,910
R-Squared	0.001	0.032
p-value “WW=-AW”	0.761	0.000
p-value “Buys=Sells”	0.313	0.117

HFT net flow, but the dollar GTR value is significantly higher the more HFTs engage in either with–wind or against–wind trading. Taken together, it seems that the institutional order gives them additional trading opportunities, that is, they trade more. These trading opportunities, however, are not more profitable in terms of the margin they make on each dollar they trade.

The coefficient is largest for with–wind flow. A one standard deviation HFT with–wind flow raises HFT gross profit by \$1,244, which is an increase of 342% relative to the sample average of \$364. This increase is also large in terms of the additional cost institutional investors incur in such a case, that is, it is $100\% * (\$1,244/\$3,024) = 41\%$ of their cost increase (see Section 5.1). The coefficient for against–wind flow is \$795 per standard deviation. This coefficient is 36% smaller than the with–wind effect, but the difference is not significant (unreported).

We further find that the dollar GTR value is larger during times of more overall volume, all else being equal. The relative GTR does not seem to depend on volume. Again, more overall volume enables HFTs to make more money, but they do not earn more on a per-dollar-traded basis.

Summary. Figure 6 summarizes our investor transaction cost and HFT profit results. It plots the estimated impact of HFT net flow on (i) the implementation shortfall of an institutional buy order and (ii) on HFT gross trading revenue (both in dollars). The estimated impact holds *ceteris paribus* because it is taken from regressions that included standard control variables (see the model specification in (4)). We observe that HFT against–wind flow lowers the institutional trading cost, whereas HFT with–wind flow raises it. The with–wind effect is substantially stronger. HFTs benefit either way, but their trading revenue is slightly higher for with–wind trading. The plots further show that changes in HFT revenue are smaller in magnitude than changes in institutional trading cost.

6 Institutional Investors' Trade Motivation

In this section, we hope to come full circle and understand our findings at a deeper level by relating them to theoretical predictions. Section 2 summarized three types of behaviors HFTs could

Figure 6: Marginal impact of HFT net flow on implementation shortfall and HFT revenue

This figure plots the marginal impact of HFT net flow on the dollar implementation shortfall and HFT gross trading revenue. The marginal impact is estimated in Tables 4 and 5. The impact, therefore, controls for standard covariates, such as order size, duration, volume, and volatility. The domain spans the interval from -1 to +1 standard deviation of the HFT net flow.

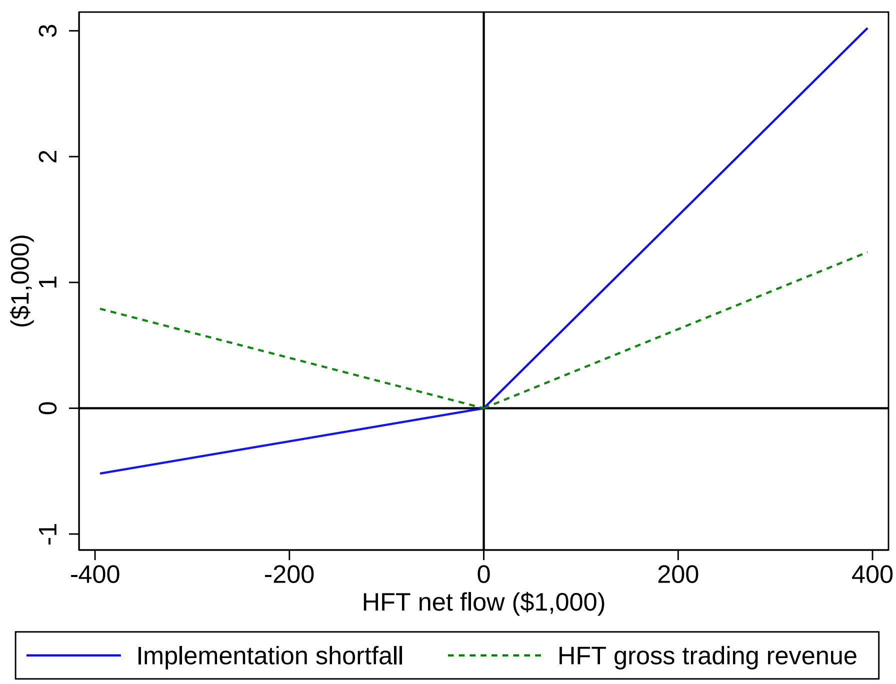


Figure 7: How many names did an investor trade in the days the investor traded?

This figure depicts the empirical distribution of the number of (stock) names an institutional investor traded in the days that the investor traded.

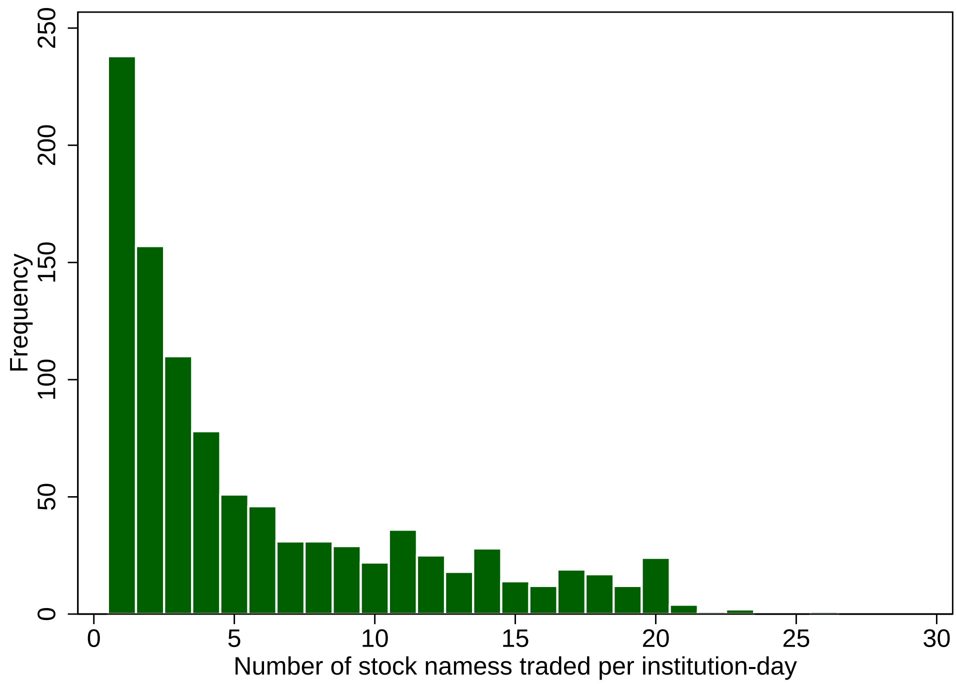


exhibit based on three theoretical studies: market making, predatory trading, and back-running. Against–wind trading is consistent with market making, but is with–wind trading a sign of predatory trading or back-running? To complete the picture, we exploit an important distinguishing feature: Predatory trading hypothesizes that an investor trades to satisfy a pressing liquidity need, whereas back-running assumes that the investor trades to capitalize on a private signal of the fundamental value. This section studies price response to distinguish the two.

Before turning to the price-impact analysis, it is useful to study whether institutional investors simply traded the market portfolio. If this were true, a back-running story based on private information would be less credible (private information would pertain to the state of the economy but would then be a great deal cheaper to trade on in the futures market).

Figure 7 plots the number of (stock) names an institution traded for the days that the institution traded. Trading the market portfolio would entail trading all the 30 constituent names. The histogram reveals that none of the institutions traded all the stocks when they traded—far from it: They most often traded only a single stock. On average, they traded 5.9 names, with a standard deviation of 5.6. We turn to a price impact analysis to study whether such trading could be interpreted as stock picking.

For each order, we compute the overall price impact (OPI) in the lifetime of the order and split it into a transitory and a permanent price impact (TPI and PPI, respectively).[12] The variable OPI is defined as simply the midquote return from the start to the end of the institutional order execution. The OPI value for sell orders (as well as TPI and PPI) is multiplied by -1 to arrive at a price impact number that is positive, on average. The variable PPI is defined as the return from the start of the order until the *next day's* end-of-day price. One simply "waits out" at least one full day to establish what the permanent price impact is.[13] The variable TPI, then, simply is OPI minus PPI. In the entire analysis, we focus on the idiosyncratic return and therefore subtract beta times the market return. We expect investors to learn firm-specific information in case they engage in stock-picking activities.

Before turning to regression analysis, it is useful to inspect whether there is any correlation between the variables of interest OPI, TPI, and PPI, on the one hand, and the nature of HFT net flow, on the other. To that end, we bin by HFT net flow tercile and compute the order size-weighted average of all three variables (similar to the approach in Figure 5). Figure 8 depicts the results. First, note that OPI in each tercile exceeds implementation shortfall (depicted in Figure 5) by at least a third. This result is not surprising, since OPI is the total price impact, whereas IS is the average price impact. Second, the decomposition indicates that the orders in all the bins are primarily informed orders. In the decomposition, PPI makes up the lion's share of OPI, at least 80%. This is the case for all three terciles. Third, in the cross section of orders, with–wind HFT

[12]This approach is inspired by a standard spread decomposition method where one decomposes the effective spread (i.e., price impact) into an adverse selection cost component (permanent) and a realized spread component (transitory) by using future midquotes (Glosten, 1987; Hendershott, Jones, and Menkveld, 2011).

[13]A longer wait could improve accuracy, but at the cost of reduced statistical power, since the fundamental value continues to evolve. We believe the full-day wait strikes a good balance, since longer waits imply qualitatively similar but statistically insignificant results. Finally, one could be worried about prices not having rebounded for multi-day parent orders. In the robustness analysis, we show that the results are robust to such concerns (Section 7.3).

Figure 8: Overall, transitory, and permanent price impacts by HFT net flow terciles

This figure plots the average order size-weighted overall price impact (OPI), transitory price impact (TPI), and permanent price impact (PPI) for against–wind, neutral, and with–wind HFT net flow terciles. The variable OPI is the stock return over the lifetime of the order, which is then decomposed into TPI and PPI. The variable PPI is the return from the start of the order execution until the next day's end-of-day price. The variable TPI is defined simply as OPI minus TPI. All variables are based on idiosyncratic returns, meaning we subtract beta times the market return from the raw return. The terms OPI, TPI, and PPI are multiplied by -1 for institutional sell orders to arrive at positive numbers in expectation. The terciles are created by first multiplying HFT net flow during institutional sell orders by -1. The HFT net flow observations are then sorted and put into equal-sized bins. The tercile with the lowest values consists of strong against–wind HFT net flows, the middle tercile contains small HFT net flows in terms of size (we label this tercile neutral), and the tercile with the highest values contains strong with–wind HFT net flows.

activity seems to center on orders with the largest PPI.

Table 6 presents the regression results and finds that, indeed, the permanent price impact correlates positively with with–wind HFT activity, all else being equal. Standard covariates are added as controls. The regressions of both the relative OPI on relative with–wind HFT and the dollar OPI on the dollar with–wind HFT exhibit a significantly positive coefficient for with–wind HFT. Rerunning these regressions for the two components of OPI reveals PPI is driving this result. Stronger with–wind HFT flow is associated with more informed orders.

Taken together, the findings in this section are consistent with HFTs back-running on informed institutional orders. Not only are the orders in our sample largely informed orders, the with–wind activity is strongest for the most informed among them.

7 Other Findings

This final section presents analyses to answer additional questions not immediately relevant to the paper's main story line. In particular, what could have prompted HFTs to turn to a with–wind strategy? And are HFTs to be thought of as a homogeneous group or is there a decomposition into against–wind types and with–wind types meaningful? The section further presents robustness analysis.

7.1 What Prompts HFTs to Change from Against-Wind to With-Wind Trading?

One of the paper's main findings is that HFTs lean against the wind initially but turn to with–wind trading eventually for long-lasting orders (see Table 3). This finding suggests that HFTs somehow detect the order and change their trading strategy accordingly, which begs the following question: What could have prompted such a change of strategy?

Table 7 characterizes trading in the lifetime of institutional orders for which HFTs traded with–wind. We again select the tercile of orders with the strongest HFT with–wind activity, henceforth

Table 6: Overall, transitory, and permanent price impact regressed on HFT net flow and control variables

This table presents the panel regression results where the dependent variables are the overall price impact (OPI), the transitory price impact (TPI), and the permanent price impact (PPI). The variable OPI is the stock return over the lifetime of the order, which is then decomposed into TPI and PPI. The variable PPI is the return from the start of the order execution until the next day's end-of-day price. The variable TPI is defined simply as OPI minus TPI. All the variables are based on idiosyncratic returns, meaning we subtract beta times the market return from the raw return. The terms OPI, TPI, and PPI are multiplied by -1 for institutional sell orders to arrive at positive numbers in expectation. The main explanatory variable is HFT net flow accumulated over the lifetime of the institutional order. It is expressed either relative to the number of shares outstanding (%) or in dollar terms ($). The values are then standardized and subsequently signed based on whether the direction is the same as that of the institutional order, in which case it is assigned a positive sign (with–wind), or the opposite, in which case it is assigned a negative sign (against–wind). Various variables are added as controls, all of which are standardized: ADV is the average daily volume based on the full sample. Order size and ADV are measured in shares. Turnover and volatility are measured from the start to the end of the order. Turnover is the daily stock volume divided by the number of shares outstanding and volatility is measured as realized volatility based on one-minute midquote returns. Also reported are the *p*-values of (i) a test of whether the against–wind coefficient equals minus the with–wind coefficient and (ii) a test of whether the model coefficients are indeed equal for institutional buy and sell orders (this table shows the pooled regression results). The regressions include stock and institution fixed effects and standard errors are clustered by stock–date. The *t*-values are in parentheses. The variable units are in brackets and reported right after the variable names. The superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	OPI (bps)	TPI(bps)	PPI (bps)	OPI ($)	TPI($)	PPI ($)
Against–wind HFT Cum Net Flow (%)	-2.061	1.791	-3.852			
	(-1.0)	(0.5)	(-1.0)			
With–wind HFT Cum Net Flow (%)	5.504**	-4.115	9.619**			
	(2.0)	(-1.2)	(2.4)			
Against–wind HFT Cum Net Flow ($)				-0.007**	0.000	-0.007
				(-2.0)	(0.1)	(-1.1)
With–wind HFT Cum Net Flow ($)				0.013***	-0.009	0.022***
				(2.9)	(-1.5)	(3.1)
Order size relative to ADV(%)	10.713***	-0.095	10.808***	0.019***	0.001	0.018***
	(8.5)	(-0.0)	(5.0)	(8.4)	(0.4)	(4.5)
Order duration (hours)	0.811	-0.750	1.561	0.003	0.000	0.003
	(0.5)	(-0.3)	(0.5)	(0.9)	(0.0)	(0.5)
Stock volatility (%)	-0.796	1.481	-2.278	-0.002	0.006	-0.008*
	(-0.7)	(0.7)	(-1.2)	(-0.9)	(1.2)	(-1.9)
Turnover (%)	-2.025	-0.125	-1.900	-0.007	-0.003	-0.004
	(-0.7)	(-0.0)	(-0.4)	(-1.4)	(-0.5)	(-0.6)
Observations	5,910	5,910	5,910	5,910	5,910	5,910
R-Squared	0.027	0.001	0.009	0.027	0.001	0.009
p-val "$WW == -1 \times AW$"	0.389	0.681	0.379	0.333	0.403	0.206
p-val "$Buys == Sells$"	0.006	0.016	0.008	0.037	0.030	0.020

Table 7: What prompts HFTs to change from against–wind to with–wind trading?

This table presents the results of an analysis on what could prompt detection, that is, HFTs changing from against–wind to with–wind trading in the lifetime of an institutional order. It reports the average of various trading variables pertaining to the institutional investor, the market, and HFTs. All variables are flow variables, expressed in units per five-minute interval. The averages are calculated for the lifetime of all institutional orders and for the tercile of orders where HFTs trade strongly with the wind. The execution period for this subset of orders is split into a pre-WW and a post-WW sub-period. The separating interval is the last five-minute interval in which the HFT cumulative net flow surpasses zero. The period following this interval, if available, is labeled post-WW; the preceding period is labeled pre-WW. The variable units are in brackets and are reported right after the variable names. For the imbalance variable, buy (sell) indicates the total volume where the buyer (seller) initiated the transaction, that is, the trigger was a market buy (sell) order. The volatility and variance ratios are computed based on midquote returns. The superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Variable	Overall avg	Pre-WW	Post-WW	Diff
Investor	Inst order's participation in share volume (%)	17.0	22.7	6.9	-15.9***
	Inst volume (in 1000 shares per 5 min)	6.7	10.2	5.2	-5.0***
	Number of child trades (per 5 min)	7.4	6.7	4.8	-1.9***
	Size child trades (in shares)	932	1,513	1,074	-439***
Market	Volume (in 1000 shares per 5 min)	44.4	46.3	50.1	3.8
	Order imbalance, |buy-sell|/(buy+sell) (%)	17.1	22.2	16.5	-5.7***
	Volatility (bps per 5 min)	21.1	15.2	43.1	27.9***
	Variance ratio (1-min-ret var/ 60 times 1-sec-ret var)	1.0	1.2	1.0	-0.2***
	Variance ratio (5-min-ret var/ 5 times 1-min-ret var)	1.4	1.5	1.4	-0.1**
HFTs	HFT relative use of market orders (%)	43.9	42.0	46.1	4.1***
	HFT participation in share volume (%)	25.7	27.6	25.8	-1.8***

referred to as WW or detected orders. In particular, the table splits WW orders' execution period into two sub-periods, a final period when HFTs trade with the wind and an earlier period when HFTs (potentially) trade against the wind. We refer to these periods as the post-WW and pre-WW periods, respectively. To separate these periods, the *last* five-minute interval in which the HFT cumulative net flow crosses zero is denoted the separating interval (and added to the pre-WW period).

For orders for which the institutional investor is eventually detected, the institutional investor trades aggressively in the pre-WW period, compared to both the investor's overall average trading and post-WW trading. In the pre-WW period, the investor's total volume is 22.7% relative to the Nasdaq volume, compared to the investor's overall average of 17.0% (in the lifetime of *all* of the investor's orders), and is substantially larger than the investor's participation rate in the post-WW period, 6.9%.[14] The investor's pre-WW trade size is 1,513 shares, which is far above the investor's overall average of 932 shares and significantly lower than the investor's post-WW average, 1,074 shares. In the pre-WW period, the investor is therefore more aggressive in terms of higher market participation with larger trades. The investor's pre-WW trade frequency, however, is lower than the investor's overall average: 6.7 trades instead of 7.4 trades per five minutes. This frequency drops significantly in the post-WW period to 4.8 trades per five minutes. Note that, on all three dimensions, the investor seems to hit the brake after detection.

The general market conditions in the pre-WW period could be characterized as average volume, lower volatility, more price momentum, and a larger order imbalance[15] when compared to both the overall sample and the post-WW period. The larger imbalance and stronger price momentum are likely due to the disproportionately high participation of the institutional investor with one-sided flow. We unfortunately cannot verify whether the imbalance is *directly* the result of the institutional investor's trades, since we cannot sign the investor's trades because they are time stamped to the second, whereas bid and ask quotes show large variations within a second. We therefore cannot identify whether the investor is the initiator or not in any of the trades.

HFTs participate slightly more in the pre-WW period and execute more of their trades through limit orders as opposed to market orders. Their volume participation rate is 27.6% in the pre-WW period, which is higher than the 25.7% overall average and significantly higher than the 25.8% average for the post-WW period. They execute 42.0% of their trades through market orders in the pre-WW period, which is lower than the 43.9% overall average and significantly lower than the 46.1% in the post-WW period. These findings suggest that HFTs are in some kind of rush to build their with–wind positions.

[14]Note that we cannot test for statistical significance in the comparison of the pre-WW sample and the overall sample, since the former is part of the latter.

[15]The order imbalance is computed following standard practice in microstructure. It is expressed as a ratio whose numerator is the absolute value of the sum of signed trade sizes, where the sign depends on who initiated the trade, that is, a plus for market buy orders and a minus for market sell orders. The denominator is the total volume.

7.2 Do All HFTs Follow the Same Strategy?

Disaggregation of HFT cumulative net flow reveals significant differences between the 10 HFTs that make up the total flow. Table 8 shows the average cumulative net flow for each HFT, separately for institutional orders where, *in the aggregate*, HFTs trade against–wind or with–wind. Since we cannot reject the null, that aggregate HFT net flow is equal for both categories of orders ($-335,800 and $331,300, respectively), testing whether they are equal in size for each HFT individually is a meaningful exercise. The joint test that they are equal for each HFT is firmly rejected (Wald test p-value is 0.000). Some HFTs have significantly greater net flow for the AW orders (HFT 2 and HFT 6), while others have greater net flow for the WW orders (HFT 5, HFT 7, and HFT 10). These individual differences are economically large but should be interpreted as HFTs doing either relatively more against–wind trading or relatively more with–wind trading. Note that they all seem to engage in both types of trading. They *all* trade against–wind individually when they trade against–wind in the aggregate and they *all* trade with–wind individually when they trade with–wind in the aggregate.

7.3 Robustness Tests

This subsection presents the robustness analysis. The results are added to the Internet Appendix.

NASDAQ OMX market share. The NASDAQ OMX market share of the exchange-traded volume was 65% at the time of our sample. The HFT cumulative net flow analysis is based on NASDAQ OMX only because this was the sole exchange that reported the trader's identity for each trade (at the end of the trading day). If HFTs trade randomly across exchanges, then the NASDAQ OMX sub-sample is representative and all the results go through. If not, NASDAQ OMX is likely to be the exchange whose majority of trades execute simply because it is the exchange with the largest market share. This line of thought inspired us to conduct the following robustness check: Select only those stock–days when the NASDAQ OMX market share was beyond some threshold and rerun the main analysis. The largest market share we can select while still having sufficient

Table 8: Do all HFTs follow the same strategy?

This table disaggregates HFT cumulative net flow in the lifetime of an institutional order. It distinguishes between institutional orders where HFTs *collectively* trade against the wind (AW) or with the wind (WW). HFT cumulative net flow is signed positive if it had the same sign as the institutional order and is negative otherwise. Since the size of total cumulative net flow for against–wind orders ($-335,800) is not significantly different from that of with–wind orders ($331,300), it is meaningful to test whether the same is true at the disaggregated level, that is, for all HFTs individually. To that end, the rightmost column tests whether the difference between the WW average and (-1) times the AW average is significantly different from zero. The table further reports the *p*-value of the joint test on whether this is true for all HFTs simultaneously. The superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	HFT cumulative net flow (in $1000)		
	Against–wind (AW)	With–wind (WW)	Difference: “WW-(-AW)”[a]
HFT 1	-0.5	4.3	3.8**
HFT 2	-94.4	52.3	-42.1***
HFT 3	-18.2	13.6	-4.6
HFT 4	-2.5	1.6	-0.9
HFT 5	-77.7	92.8	15.1***
HFT 6	-59.8	47.1	-12.7***
HFT 7	-50.0	72.5	22.4***
HFT 8	-1.9	1.7	-0.3
HFT 9	-6.1	8.8	2.8
HFT 10	-24.7	36.6	11.9***
Total	-335.8	331.3	-4.5

[a] The *p*-value of the Wald test that “WW=-AW” for all HFTs is 0.00.

observations for statistical inference is 75%, which reduces the sample from 5,910 orders to 281 orders. We can confirm our main result, where HFTs initially trade against–wind but eventually trade with–wind for long-lasting orders. This result remains statistically significant.

HFTs build position *before* the first child trade. Although we find that HFTs collectively trade against–wind after an investor's first child trade, one could argue that they might have traded with–wind *ahead* of the execution of this first child trade. If the investor polled interest in, for example, dark pools or submitted a non-marketable limit order on an exchange, then this signal could have been picked up by HFTs. It could have prompted them to trade with–wind immediately after receiving such a signal. In this case, the against–wind flow after the first child trade could just be due to the unwinding of a predatory with–wind position. We do not observe investor orders (only their trades), so we cannot directly test this hypothesis. Instead, we calculate HFT cumulative net flow in the hour before an investor's first child trade. We find that HFTs are net sellers ahead of institutional buy orders and net buyers ahead of institutional sell orders. We therefore consider the alternative interpretation of our main result to be highly unlikely (since HFTs must have picked up a signal, if any, at least an hour before an investor's first trade and then started unwinding a speculative position even before the institutional investor executes his or her first trade).

An institutional investor trading the same stock, in the same direction, on consecutive days could be due to a multi-day parent order. The available data do not contain parent order information, only child order executions. We therefore base our analysis on stringing together all child trades for a single institution per stock–day and call them meta-orders. One reasonable alternative approach is to string child trades together across days if an institution trades the same stock in the same direction for two or more consecutive days.[16] If one does so, then the number of meta-orders shrinks from 5,910 to 3,450. Two-thirds of these orders are single-day executions, 17% are two-day executions, and less than 10% execute across three or more days. When repeating our main

[16]We chose not to make this our default approach, since we consider it somewhat more arbitrary. A portfolio manager is unlikely to interact with the trading desk intradaily, but is more likely to be in touch daily. A multi-day meta-order could, therefore, be either a single parent order sent by the portfolio manager once or multiple parent orders where the portfolio manager communicates orders daily after parsing previous-day market outcomes.

analysis for the single-day orders, we again find statistical support for our main results: HFTs lean against–wind initially and with–wind eventually for long-lasting orders (Figure 3); institutional trading costs are lower when HFTs trade against–wind and higher when they trade with–wind (Table 4).

8 Conclusion

This paper is the first to document how trading by HFTs affects the trading cost of end users. End users are known to execute their large orders through a series of small child order transactions. A sample of order executions by four large institutional investors was studied for Swedish stocks in 2011–2013. We find that HFTs initially lean against an order (trade in the direction opposite to it) but, if the order lasts more than a couple of hours, they then turn around and go with the order. HFT gross profit is positive either way. Institutional investors' costs are lower for against–wind HFT net flow but disproportionately larger for with–wind HFT flow.

Against–wind trading by HFTs is generally consistent with classic market making. The novel finding of with–wind high-frequency trading largely supports recent theory on intermediaries back-running on informed orders (Yang and Zhu, 2015). With–wind trading occurs only after several hours, which could be interpreted as the initial learning period. Further analysis on the cross section of institutional orders reveals that, all else being equal, larger permanent price impacts correlate positively with HFT with–wind activity. HFTs seem to run on the most informed orders.

HFT back-running on institutional orders does not necessarily improve market quality. One could argue that prices become more efficient in the short run. HFT trading in the same direction as informed investors makes prices reveal private information more quickly. The worrisome side effect is that, in the long run, prices could become less efficient. Institutional investors could discontinue costly analyst research, since informational rents have to be shared with others in the trading process. Research might no longer be privately profitable. This could become socially costly if informational externalities are large (i.e., information benefits the allocation of capital across entrepreneurs).

We believe the market structure debate should re-center around end user costs. Data are hard to come by, but it should be in the interests of end users and retail and institutional investors to make their trade data available (as was done for this study, e.g.). Alternatively, regulators could demand more data granularity from data centers, much in the spirit of what U.S. regulators did after the 1987 crash. Exchanges were required to identify retail orders in the consolidated equity audit trail data (CAUD). For each trade they completed, brokers had to report whether it was a principal or an agency trade and, if an agency trade, whether it was for a retail investor or for an institutional investor. This would enable more analysis to inform future debates on market quality. The recent SEC initiative to amend rule 613 and create a consolidated audit tape (for regulatory use only) seems like a step in the right direction.[17]

References

Almgren, R., and N. Chriss, 1999, "Value Under Liquidation," *Risk*, pp. 61–63.

Anand, A., P. Irvine, A. Puckett, and K. Venkataraman, 2012, "Performance of Institutional Trading Desks: An Analysis of Persistence in Trading Costs," *Review of Financial Studies*, 25(2), 557–598.

——— , 2013, "Institutional Trading and Stock Resiliency: Evidence From the 2007-2009 Financial Crisis," *Journal of Financial Economics*, 108(3), 773–797.

Biais, B., and T. Foucault, 2014, "HFT and Market Quality," *Bankers, Markets & Investors*, (128), 5–19.

Boehmer, E., K. Fong, and J. Wu, 2014, "International Evidence on Algorithmic Trading," *Working Paper EDHEC Business School*.

Boulatov, A., D. Bernhardt, and I. Larionov, 2016, "Predatory and Defensive Trading in a Dynamic Model of Optimal Execution by Multiple Traders," Manuscript, Higher School of Economics (Moscow).

Brogaard, J., B. Hagströmer, L. Nordèn, and R. Riordan, 2015, "Trading Fast and Slow: Colocation and Market Quality," *Forthcoming Review of Financial Studies*.

Brogaard, J., T. Hendershott, and R. Riordan, 2014, "High Frequency Trading and Price Discovery," *Review of Financial Studies*, 27(8), 2267–2306.

[17] See https://www.sec.gov/divisions/marketreg/rule613-info.htm.

Brunnermeier, M. K., and L. H. Pedersen, 2005, "Predatory Trading," *Journal of Finance*, 60(4), 1825–1863.

Chaboud, A. P., B. Chiquoine, E. Hjalmarsson, and C. Vega, 2014, "Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market," *Journal of Finance*, 69(5), 1540–6261.

Chordia, T., and A. Subrahmanyam, 2004, "Order Imbalance and Individual Stock Returns: Theory and Evidence," *Journal of Financial Economics*, 72(1), 485–518.

Comerton-Forde, C., T. Hendershott, C. M. Jones, P. C. Moulton, and M. S. Seasholes, 2010, "Time Variation in Liquidity: The Role of Market-Maker Inventories and Revenues," *Journal of Finance*, 65(1), 295–331.

Davies, R. J., and S. S. Kim, 2009, "Using Matched Samples to Test for Differences in Trade Execution Costs," *Journal of Financial Markets*, 12(2), 173–202.

Easley, D., N. M. Kiefer, and M. O'Hara, 1996, "Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow," *Journal of Finance*, 51, 811–833.

Gao, C., and B. Mizrach, 2013, "Market Quality Breakdowns in Equities," *Working Paper Rutgers University*.

Glosten, L., 1987, "Components of the Bid Ask Spread and the Statistical Properties of Transaction Prices," *Journal of Finance*, 42, 1293–1307.

Grossman, S. J., and M. H. Miller, 1988, "Liquidity and Market Structure," *The Journal of Finance*, 43(3), 617–633.

Harris, L. E., 1997, "Order Exposure and Parasitic Traders," Manuscript, University of Southern California.

Hasbrouck, J., 2015, "High Frequency Quoting: Short-Term Volatility in Bids and Offers," *Working Paper NYU*.

Hasbrouck, J., and G. Saar, 2013, "Low-latency Trading," *Journal of Financial Markets*, 16(4), 646–679.

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

——— , 2013, "Implementation Shortfall with Transitory Price Effects," in *High Frequency Trading: New Realities for Traders, Markets and Regulators*, ed. by D. Easley, M. Lopez de Prado, and M. O'Hara. Risk Books, London, chap. 9.

Hirschey, N. H., 2014, "Do High-Frequency Traders Anticipate Buying and Selling Pressure?," *Working Paper University of Texas at Austin*.

Jones, C. M., 2014, "What Do We Know About High-Frequency Trading?," *Columbia Business School Research Paper No. 13-11*.

Kirilenko, A. A., A. S. Kyle, M. Samadi, and T. Tuzun, 2014, "The Flash Crash: The Impact of High Frequency Trading on an Electronic Market," *Working Paper CFTC*.

Korajczyk, R., and D. Murphy, 2014, "High Frequency Market Making to Large Institutional Trades," *Working Paper Northwestern University*.

Linnainmaa, J. T., and G. Saar, 2012, "Lack of Anonymity and the Inference from Order Flow," *Review of Financial Studies*, 25(5), 1414–1456.

Menkveld, A. J., 2013, "High Frequency Trading and the New Market Makers," *Journal of Financial Markets*, 16(4), 712–740.

Rock, K., 1990, "The Specialist's Order Book and Price Anomalies," Manuscript, Harvard University.

SEC, 2010, "Concept Release on Equity Market Structure," Release No. 34-61358; File No. S7-02-10.

——— , 2014, "Equity Market Structure Literature Review. Part II: High Frequency Trading," Literature review.

Securities, E., and M. Authority, 2014, "High-Frequency Trading Activity in EU Equity Markets," ESMA Economic Report 2014-1.

Sofianos, G., and J. Xiang, 2013, "Do Algo Executions Leak Information?," in *High-Frequency Trading*, ed. by D. Easley, M. M. López de Prado, and M. O'Hara. Risk Books.

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

van Kervel, V., 2015, "Competition for Order Flow with Fast and Slow Traders," *Review of Financial Studies*, 28(7), 2094–2127.

Yang, L., and H. Zhu, 2015, "Back-Running: Seeking and Hiding Fundamental Information in Order Flows," *Working Paper MIT*.

Ye, M., C. Yao, and J. Gai, 2013, "The Externalities of High Frequency Trading," *Working Paper University of Illinois at Urbana-Champaign*.