

High-Frequency Trading around Large Institutional Orders^{*}

Vincent van Kervel and Albert J. Menkveld

February 21, 2018

^{*}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 Andrew Ainsworth, Bruno Biais (the editor), Jonathan Brogaard, Björn Hagströmer, Terrence Hendershott, Ron Kaniel, Robert Korajzyck, Charles-Albert Lehalle, Anna Obizhaeva, Neil Pearson, Marc Ponsen, Talis Putnins, Yazid Sharaiha, George Sofianos, René Wells, Avi Wohl, Shihao Yu, an anonymous associate editor and two anonymous referees for their very useful comments. We are also grateful for comments received during conferences at Alan Turing Institute, NBIM, and Tel Aviv University and seminars at AFM, CFTC, Université Paris Dauphine, Goldman Sachs, SEC, University of Sydney, UNSW, and UTS. 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. The authors have signed non-disclosure agreements for the proprietary data used in this study.

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 HFTs trading “with the wind,” that is, in the same direction. We find that HFTs initially lean against these orders but eventually change direction and take position in the same direction for the most informed institutional orders. Our empirical findings are consistent with investors trading strategically on their information. When deciding trade intensity, they seem to trade off higher speculative profit against higher risk of detection by HFTs and being preyed on.

JEL code: G10

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 few deciles (SEC, 2010; ESMA, 2014).

High-frequency trading has triggered a great deal of academic study, particularly after NASDAQ released data that labeled HFT trades. 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. Biais and Foucault (2014), O’Hara (2015a), and Menkveld (2016) are examples of survey articles on the young and rapidly growing HFT literature.¹

Relatively unexplored is how HFT relates to 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 a parent order up into small child orders that are fed

¹Several empirical studies find that HFT activity reduces bid-ask spreads (Hendershott, Jones, and Menkveld, 2011; Hasbrouck and Saar, 2013; Menkveld, 2013; Brogaard et al., 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 et al., 2014; Hasbrouck, 2015) whereas others document a positive correlation (Gao and Mizrach, 2013; Ye, Yao, and Gai, 2013; Boehmer, Fong, and Wu, 2014; Kirilenko et al., 2017).

to the market sequentially. They care about “implementation shortfall,” that is, the average price at which the entire order executed relative to the price that prevailed when it started executing. In other words, how far did they push the price away from them while executing? Institutional investors care about *cumulative* price impact rather than the half-spread paid on, for example, a single market order. Some have expressed concern that trading costs have increased and attribute this to HFT presence.²

Objective. Our objective is to document how HFTs trade while an institutional order executes through a series of child trades, and to provide an economic narrative inspired by theoretical studies. Three of them speak directly to our empirical setting. HFTs could be *market makers* who take the other side of the order. They would, for example, be selling in the period when a buy order executes (Grossman and Miller, 1988). Or, they could engage in *predatory trading*. They would buy along with the institutional order initially to then sell when the buy order is almost finished executing, essentially riding the (transitory) price-pressure wave (Brunnermeier and Pedersen, 2005). Finally, they might need time to learn that the buy order is executing, to *back-run* on it later by also buying (Yang and Zhu, 2017).³ Taking such long position relatively late however only makes money if the order is information-motivated (as opposed to an uninformed order seeking liquidity).

²See, for example, “Institutional Investors Air HFT Concerns” *Financial Times*, September 12, 2011; “Wealth Fund 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.

³Boulatov, 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.

These three studies have distinctively different predictions for how HFTs trade as intermediaries while an institutional order executes, and the way that the price responds (see Section I).

The predicted patterns are taken to the data. The empirical analysis is based on a sample that combines proprietary data on order executions by institutional investors with public HFT trade data. The sample runs from January 1, 2011 through March 31, 2013 and covers trading in Swedish index stocks. The order execution data were provided by four large institutional investors (APG, DNB, NBIM, and Swedbank Robur) and include 801,341 child trades. We construct “parent orders,” which we refer to as meta orders, by stringing together child trades by a single institution in a single stock, possibly spanning multiple days (for details see Section II.B on p. 14). The final sample consists of 5,136 such meta orders that, on average, contain 156 child trades and last a bit longer than four hours. Not surprisingly, we find that these orders are directional in the sense that all child trades in a single order are either almost exclusively buys or almost exclusively sells. Finally, institutional orders are large on average: \$2.2 million, or 3.88% when expressed relative to average daily volume.

An important benefit of this particular sample is that HFTs had to reveal their trades on NASDAQ OMX,⁴ which at the time was the dominant market with a market share of two-thirds for the local index stocks.⁵ 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.⁶

⁴This 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.

⁵These numbers are from Fidessa, a trade reporting company (<http://tinyurl.com/ozo8ytm>).

⁶See “Europe’s Top 10 High-Frequency Kingmakers,” *Financial News*, October 3, 2011.

Collectively, their participation rate in NASDAQ OMX trades is almost a third in our sample.

Findings. The empirical analysis yields three main findings. First, the pattern of HFT trading fits none of the three theoretical studies perfectly. We find that HFTs (as a group) appear to lean against the wind when an order starts executing but if it executes more than seven hours (possibly multiple days), they seem to reverse course and trade with the wind. They not only close out their position, but they actually *build* a speculative position in the direction of the order. For buy orders, for example, HFTs sell initially and after seven hours start buying, but lots more than they sold initially. This first result is stated somewhat cautiously, since we are aware that HFTs could have entered offsetting positions in alternative markets or in highly correlated securities. We nevertheless consider this concern unlikely because perfectly correlated securities are hard to find for single stocks, and NASDAQ OMX is by far the largest equity exchange for Swedish stocks.

A more pressing concern is that this first result could be entirely mechanical. Institutional investors *and* HFTs might both respond to certain market conditions, which implies that any correlation in their trading pattern is simply driven by a third factor. To rule out such explanation, we create a placebo sample in which none of the institutional investors were active yet market conditions were similar (in terms of volume, idiosyncratic return, market return, volatility, and (market) order imbalance). We find HFT against-wind trading also in the placebo sample, but not the with-wind trading. The with-wind trading therefore seems to be robustly related to the presence of an institutional order. This finding motivates the remainder of the empirical analysis and the overall message of the paper.

Late HFT with-wind trading is consistent with both predatory trading and back-running. As

alluded to earlier, the two theories can be distinguished by analyzing the information content of the institutional orders. This is where we turn next.

The second main finding is that institutional orders appear mostly information-motivated, in particular the ones with long-lasting executions that HFTs eventually trade along with. A couple of findings support this claim. First, on days when institutions trade, they trade typically only one or a few stock names. On the spectrum from stock picking based on informational signals to uninformed trading of the market portfolio, institutions seems to be closer to the stock-picking end. Analysis of the informational content of their orders based on the extent to which prices rebound after their final child trade, supports the conjectured informed trading. More specifically, the information component dominates the transitory component in their order's overall price impact. Finally, in the cross-section of orders, this information component is larger for orders that HFTs trade with-wind on. This association is economically large as the average permanent price impact is 15 basis points instead of 6 after a one standard deviation increase in with-wind flow.

The third main finding is a set of results that frames the observed with-wind trading as strategic interplay between investors and HFTs. The overarching conjecture is that investors are privately informed and optimally trade on their signal in *full* awareness of HFTs preying on the footprint they leave in the market. More specifically, when deciding how intensively to trade on their signal an investor faces the following trade-off. Trading more intensively raises his speculative profit, but it also raises the risk of being detected by HFTs and preyed on through their with-wind trading. The following findings generate empirical support for this conjecture.

- Investors seem to trade cautiously on their information as this information is revealed only

slowly in prices. Half of their private signal is revealed after three hours, on average. This slow revelation frames our first finding — it could explain why HFTs need *seven* hours to discover an institutional order to trade with the wind on it.

- The ability of HFTs to eventually detect information-motivated executions is demonstrated by them trading most strongly with-wind on the long-lasting orders that are most informed.
- Investors seem aware of actual detection as they decelerate trading after HFTs turn with-wind on their order. Their participation in volume, for example, drops by about 50%.
- Detection is costly to investors, and profitable to HFTs. A one standard deviation with-wind trading by HFTs associates with \$4,480 higher order-execution cost and with \$850 higher HFT gross profit.

Contribution relative to contemporary papers. Our paper contributes to three contemporary papers on trading by institutional investors and HFTs. Korajczyk and Murphy (2016) 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 (2016) 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, they focus on HFT liquidity provision, whereas our focus is on unwind trading by HFTs which is more robust in our sample. We develop an economic perspective and support it with further analysis (e.g., show that institutional orders are information-motivated).

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.⁷

The third related paper is that of Hirschey (2014), who documents that this second’s 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 retail investors and other intermediaries, such as proprietary trading by sell-side banks. Second, our results suggest that HFTs respond to investor flow at low frequencies. We believe that HFTs should be thought of more broadly than as machines engaged in sub-millisecond speed races. In fact, we suspect that their superior information technology enables them to generate better signals from “big data” assembled over multiple hours and trade

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

on it profitably.

Some words of caution. Our study faces several limitations and we therefore caution against over-interpretation 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 (i.e., Yang and Zhu, 2017) over others helps us understand more deeply the functioning of securities markets.

Second, HFTs seem to trade with-wind on institutional orders, 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, the order flow of such intermediaries is grouped together with end-user flow.

Third, with-wind trading by intermediaries most likely plagued investors 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 with-wind trading on institutional orders without claiming that it is somehow 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 with-wind trading. 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 paying for retail order flow (Easley, Kiefer, and O'Hara, 1996).

Fourth, with-wind trading is not necessarily costly to investors 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.

The paper proceeds as follows. Section I offers the theoretical motivation of the study. Section II describes the data. Section III analyzes HFT net flow during the execution of institutional orders, which is then related to predatory trading and back-running in Section IV. Section V offers evidence that with-wind trading is consistent with strategic interplay between HFTs and institutional investors. Section VI offers robustness checks, and Section VII concludes.

I. Price and Trade Patterns, as Predicted by Theory

In this section, we summarize several priors for trading by HFTs around large institutional orders. Figure 1 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, 2017). 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 1).

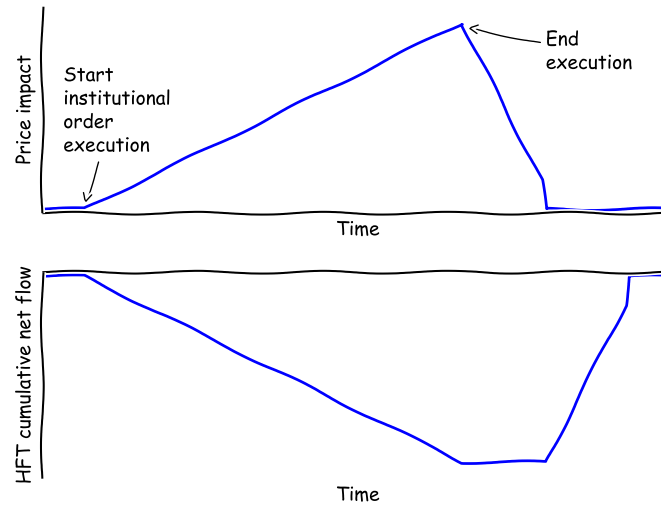
Brunnermeier and Pedersen (2005) model strategic trading by predators who are aware that an institutional investor is forced to trade for liquidity reasons. Their findings imply that predatory HFTs first trade along with an institutional buy order and thus add to price pressure, but eventually turn around 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 1).

Yang and Zhu (2017) 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 (see Panel (c) of Figure 1).

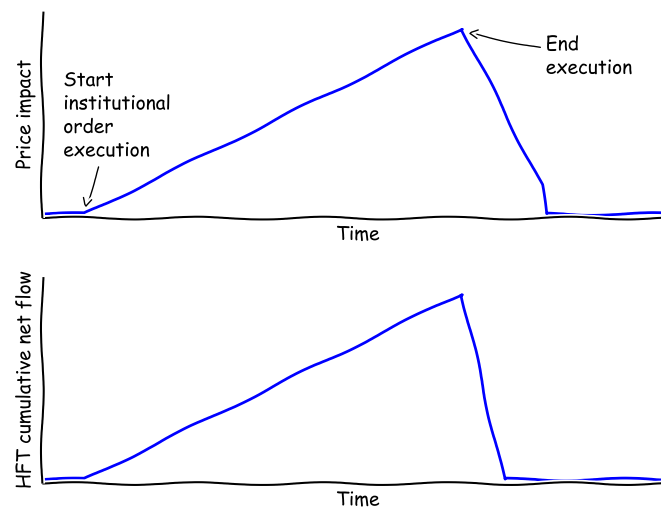
Figure 1 Price and trade patterns around institutional buy orders predicted by theory

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, 2017), respectively.

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



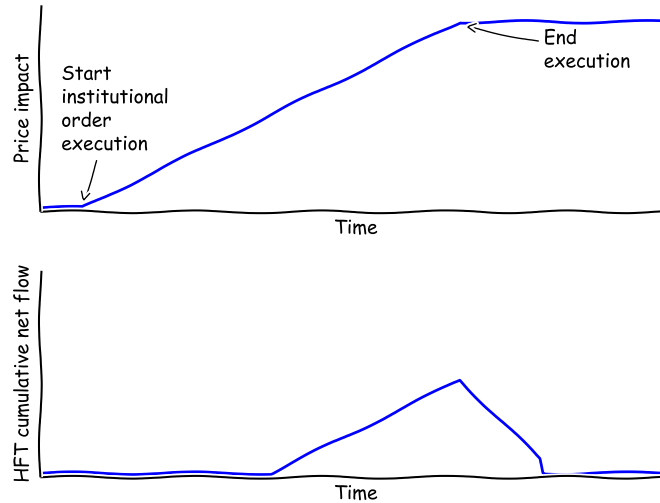
(b) Predatory trading (Brunnermeier and Pedersen, 2005)



Continued on next page.

(Figure 1 continued)

(c) Back-running (Yang and Zhu, 2017)



II. 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.

A. 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 trading venues such as multilateral trading facilities and 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 four 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 VI.

B. 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 timestamp (to the second), price, and quantity.

As parent-order information is missing, the best we can do is infer them from child order executions. We string these child trades together into “meta orders” and create observations by stock-period-institution. This is done as follows. First, for each stock and each day, all child order executions by a single institution are aggregated into stock-day-institution orders. Then, following Korajczyk and Murphy (2016), these orders are further strung together across days if there is a child trade in the last half hour of one day followed by a child trade in the first half hour in the next day. It turns out that 577 out of the 5,136 orders span multiple days. We refer to institutional meta orders as institutional orders in the remainder of the manuscript for brevity.

Filters. 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.

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

C. Summary Statistics

Table I presents summary statistics of the 5,136 stock-period-institution observations. It reports statistics for the entire sample as well as for institutional buy and sell orders separately.

Institutional orders are large and long-lasting. The average order size is \$2.2 million or 3.88% of average daily volume. It consists of 156 child trades that last 4.26 hours, on average. The fraction of the orders executed through market orders (as opposed to limit orders) is 40.6%, which we consider close to the 34.5% documented for ITG Posit data in the U.S. (O'Hara, 2015b). The average implementation shortfall is \$4,294 or 8.27 basis points when expressed relative to order size. Institutional sell orders are somewhat larger on average and execute in a shorter period of time.

In the period it executes, the institutional order is a sizable part of the market. Its net volume for buy or sell orders is \$2.2 million which compares to an overall volume of \$61.2 million in this period (i.e., it is 3.6%). Its net number of trades is 156 out of 5,507 in the market (2.8%). The directionality is at least 0.99 on average. This level is much above the 0.90 threshold that was used to filter the data (see Section II.B).

While the institutional order executes, HFTs are highly active. They trade \$13.1 million which is 21% when compared to overall volume. This volume is executed through, on average, 1,514 net trades. Note that this suggests that the average trade size is \$8,652 (\$13.1 million / 1,514) which is about half of what the same ratio yields for institutional child order size: \$14,100. Directionality in HFT trades is only 8%. The relatively small trade size and low directionality are in line with a well known HFT preference for keeping inventory at a minimum (SEC, 2010).

Table I Summary statistics

This table presents summary statistics for all order executions in Swedish index stocks from January 1, 2011 through March 31, 2013. An observation is defined as the period (consisting potentially of multiple days) during which an institution executed a (meta) order for a particular stock. Observations are therefore indexed by stock-period-institution (SPI). The table presents SPI means and standard errors (in brackets) of various trade variables for the active institution (Inst), the group of HFTs (HFTs), and the entire market (All). The leftmost column describes the variable with the unit of measurement in brackets. Most are self-explanatory, except for the following: directionality is defined as $|\text{buys-sells}|/(\text{buys+sells})$ (note: all buys and sells are based on whether the institution bought or sold, *not* based on which side was the market order), 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 one would have paid had one been able to buy all at the midquote price that prevailed just before the order started executing (sell-order definition is analogous; relative implementation shortfall is obtained by dividing by total order size), and HFTs' gross trading revenue pertains to their trading in the lifetime of the institutional order where any nonzero position at the end time is valued at the end-of-day price. 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 order size (\$100,000)	22 (1)			19 (1)			27 (1)		
Avg order size relative to ADV (%)	3.88 (0.11)			3.41 (0.13)			4.63 (0.18)		
Avg nr child trades	156 (5)			145 (6)			173 (7)		
Avg duration (hours)	4.26 (0.08)			4.66 (0.10)			3.63 (0.11)		
Avg fraction trades through market orders (%)	40.6 (30.4)			40.2 (30.0)			41.3 (30.9)		
Avg imp shortfall (\$)	4,294 (375)			4,366 (496)			4,181 (566)		
Avg imp shortfall (bps)	8.27 (0.74)			8.21 (1.08)			8.37 (0.86)		
Avg HFT gross trading revenue (\$)	285 (82)			261 (123)			323 (85)		
Avg volume (\$100,000)	22 (1)	131 (3)	612 (14)	19 (1)	145 (4)	675 (19)	27 (1)	108 (4)	508 (20)
Avg directionality (based on \$100,000)	1.00 (0.00)	0.08 (0.00)	0.00 (0.00)	1.00 (0.00)	0.07 (0.00)	0.00 (0.00)	1.00 (0.00)	0.08 (0.00)	0.00 (0.00)
Avg volume (#trades)	156 (5)	1,514 (32)	5,507 (116)	145 (6)	1,718 (45)	6,304 (160)	177 (8)	1,178 (43)	4,194 (153)
Avg directionality (based on #trades)	0.99 (0.00)	0.08 (0.00)	0.00 (0.00)	1.00 (0.00)	0.07 (0.00)	0.00 (0.00)	0.99 (0.00)	0.08 (0.00)	0.00 (0.00)
Number of observations	5,136			3,151			1,985		

Inspection of start- and end-times of the institutional orders shows the following daily pattern (see Figure A1 in the online appendix). These times are quite evenly dispersed throughout the day with two exceptions. End-times tend to concentrate at the close with 31% of the orders ending in the last fifteen minutes of the trading day. Start- and end-times peak between 15:45 and 16:00, which is right after U.S. equity markets open. These observations led us to flag these orders in later panel regressions by adding dummies.

The online-appendix contains an additional set of summary statistics which show that HFTs mean-revert their position substantially within the day, but also across days. Table AI in this appendix again reports directionality (i.e., $|B-S|/(B+S)$), but this time for stock-days, stock-weeks, and stock-months. It is 0.084, 0.048, and 0.028, respectively. These numbers suggest that HFTs are close to flat at the end of the day, but not entirely. Part of the position they accumulate within a day, they smooth out across weeks or even months. Not mean-reverting entirely to zero is perhaps not surprising, given that we do not see their trading outside of NASDAQ OMX (see Section II.A). HFTs might also hedge their positions with correlated assets to obtain a flat position *in terms of risk*. The table further shows that the directionality of institutional investors, on the other hand, is much larger across days, weeks and months: 0.99, 0.91, and 0.83, respectively.

III. 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 associated with the presence of the institutional order, or simply relates to the market conditions that prevailed at the time the institution implemented its order.

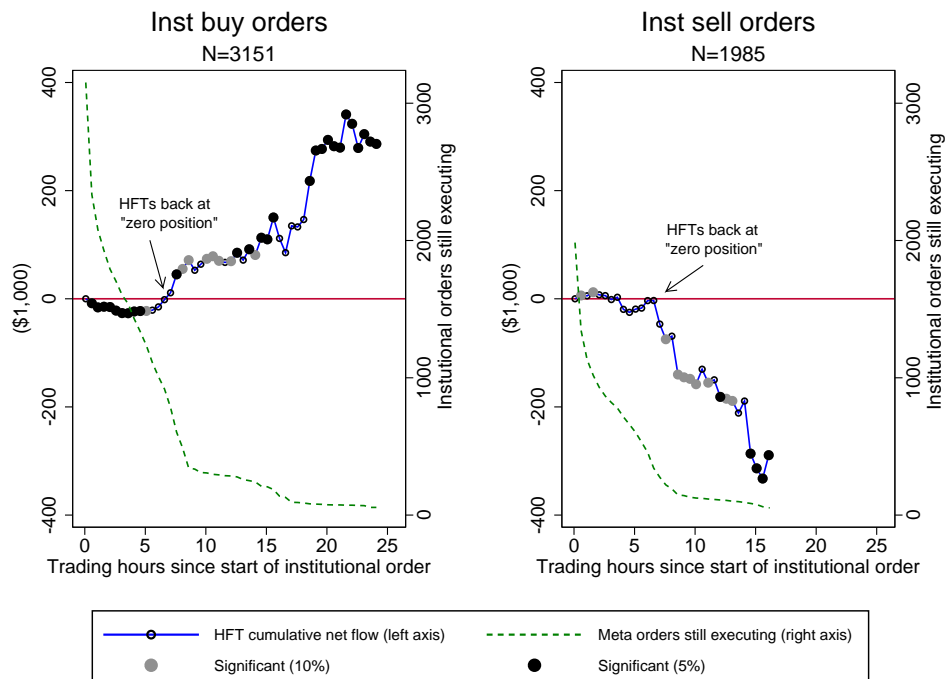
A. *HFT Net Flow in the Lifetime of an Institutional Order*

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 executing (i.e., their final child trade has not arrived yet). 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 III). If the average net flow is negative, then HFTs lean against the order (against wind). If it is positive, then they trade along with it (with wind). An equivalent analysis is done for institutional sell orders.

Figure 2 plots the results for buy orders (left panel) and sell orders (right panel). The plots show that HFTs lean against buy orders in the first seven hours of execution, since they go short

Figure 2 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 in the lifetime of the order. These time points are multiples of 30 minutes. At each point the average is taken across all institutional orders that are still executing at that point (i.e., the final child trade has not arrived yet). The size of the sample at each time point is indicated by the dashed green line (right y-axis). Statistical significance is tested based on the t -value of the mean across stock-institution fixed effects (same as the overall mean), with residuals clustered at the stock-period level.



for approximately \$20,000. This result is statistically significant only in the first five hours, as indicated 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 (only significant at a 10% significance level).

Strikingly, HFTs turn around and go with the order if it lasts more than seven hours for buys and more than two hours for sells. The result turns statistically significant only if the order lasts longer than seven hours. HFT cumulative net flow steadily grows until it plateaus at around \$300,000 with-wind. It reaches this level after about 20 hours for institutional buys and after 15 hours for institutional sells.

Overall it seems that with-wind trading is slightly stronger for sell orders since initial against-wind trading is weaker and the large with-wind position is reached earlier. We suspect this is due to the higher execution intensity of sell orders. Indeed, Table I reveals that sell orders are 38% larger and about an hour shorter, on average. This finding might not be all that surprising as investors might rationally choose to specialize (Van Nieuwerburgh and Veldkamp, 2009). As a result, they are better informed about stocks they own (and consider selling) than stocks they do not own (but consider buying).⁹

The dashed green line indicates the number of institutional orders still executing at each point in time (right y-axis). For buy orders, we observe significant with-wind flow for the 500 orders that are still executing after seven hours. This is about a sixth of all buy orders. For sell orders, there is significant with-wind flow for 200 orders, which is about a tenth of all sell orders.

⁹In Section IV.B we show that larger orders tend to be more informed and there is no *additional* effect from the “meta order is buy” dummy. It therefore is the larger size of sell orders that makes them more informed and therefore exhibit stronger with-wind trading.

The analysis in this section raises new questions, two of which are answered by additional results reported in the online appendix. First, do institutions execute their orders linearly over time, or is there front or back-loading? Figure A2 in the appendix shows that on average investors execute fairly linearly, for both buy and sell orders. Sell orders however execute more intensively, since their slope is about 50% larger in magnitude. This is not surprising since, as noted earlier, sell orders are larger and execute over a shorter horizon. Second, can the HFTs be treated as a single group, or do they individually follow different strategies in response to the institutional orders? Table AII in the appendix shows that if HFTs in the aggregate trade against-wind in the lifetime of an order, each individual HFT trades against wind on average as well. We find a similar result for with-wind trading. Having said that, we do want to mention that there is some heterogeneity across HFTs in terms of how intensively they trade against-wind or with-wind. The overall finding of relative homogeneity for these HFTs is somewhat remarkable given that the HFT literature discusses a variety of HFT strategies (see, e.g., Menkveld, 2016, for a literature survey).

B. 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-period when one of the institutional investors was active to a similar stock-period for the same stock but when none of the four institutional investors were active. We proceed as follows. The placebo stock-period is selected based on matching five trade variables across two periods:

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

The period from market open until the first child trade is added to account for potential endogenous timing by the institution with regards to when to start executing (Hendershott, Jones, and Menkveld, 2013). The five trade variables used are volume, market return, idiosyncratic return¹⁰, realized volatility (based on one-minute midquote returns), and order imbalance (i.e., standard initiator-signed imbalance or otherwise referred to as market-order imbalance). The imbalance measure should pick up HFTs building or managing positions *in general*, and if the large institutional orders would just be run-of-the-mill position building then any effect would disappear when comparing the treated and untreated samples.

A “nearest-neighbor” matching procedure is used to identify the placebo days. We follow Davies and Kim (2009, p. 183) with one modification: The distance is measured in standard deviation units as opposed to percentages.¹¹

Note that the placebo analysis controls for a momentum-based explanation of HFT behavior. If

¹⁰We used a beta estimate from Reuters that is based on monthly data for a five-year rolling window (<http://www.reuters.com/finance/markets/indices>).

¹¹The 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 close-to-zero). We therefore choose to express distance in standard deviation units. Specifically, for the ten matching variables x^k and for treated observation i , we select the matching observation j as follows:

$$\operatorname{argmin}_{j \in J} \sum_{k=1}^{10} \left(\frac{|x_i^k - x_j^k|}{\sqrt{1/J \sum_{l=1}^J |x_i^k - x_l^k|^2}} \right),$$

where j runs over all candidate placebo periods.

Table II Quality of the match with placebo periods

This table presents the mean and standard deviation of the five variables that were used to construct a placebo sample. The main sample consists of the period when an institution executed an order. Each such period is matched with the same period elsewhere in the sample but this time without an order by any of the four institutional investors. The match is done by a nearest-neighbor algorithm. It uses the average of the five trade variables computed for two time intervals: (i) from market open until the start of order execution (i.e., first child trade) and (ii) from the start of execution until the end of it. The five trade variables are dollar volume, index return, the stock's idiosyncratic return, realized volatility (based on one-minute midquote returns), and NASDAQ market order imbalance (buyer minus seller initiated volume). The distance reported in the table is the average distance across all matching variables. Distance is expressed in standard deviation units.

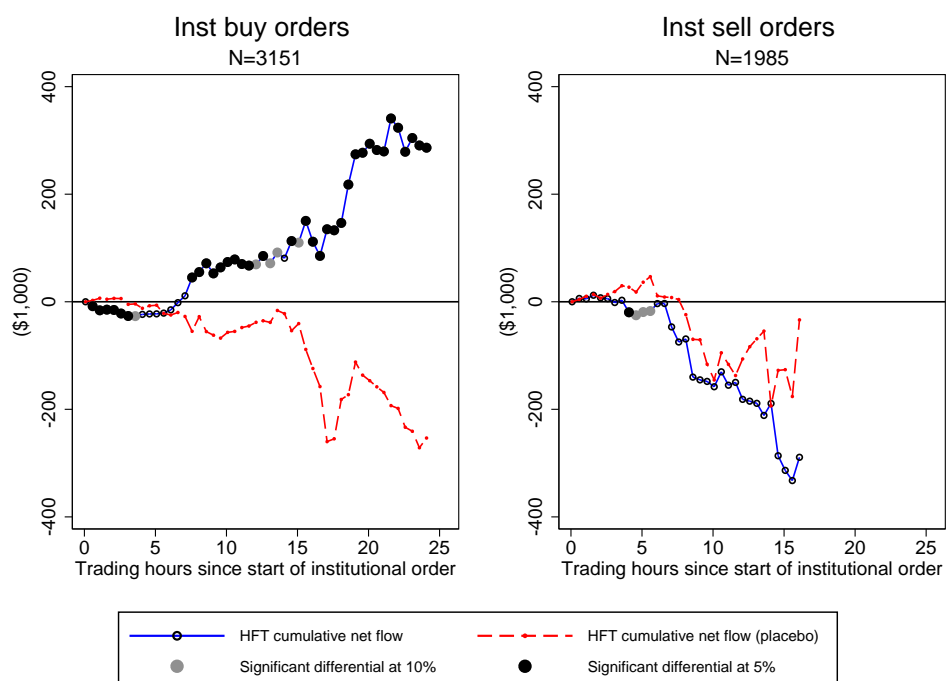
	Sample stock-periods					Placebo stock-periods					Distance
	Volume	r^{Index}	r^{Idio}	Volatility	Imb	Volume	r^{Index}	r^{Idio}	Volatility	Imb	
<i>Panel A: Institutional buy orders</i>											
Mean (open-start)	14.5	0.0	1.0	5.1	0.6	13.4	-0.2	1.6	4.2	0.6	4.2
Std (open-start)	21.3	79.4	79.7	13.5	2.9	18.6	67.6	64.9	6.9	2.4	
Mean (start-end)	29.5	3.0	7.8	6.1	0.5	0.3	1.9	4.4	4.1	1.2	3.2
Std (start-end)	45.6	98.7	91.1	45.0	3.6	0.4	70.1	75.6	26.6	3.9	
<i>Panel B: Institutional sell orders</i>											
Mean (open-start)	13.5	1.7	-0.5	3.2	0.5	13.0	1.7	-0.1	3.0	0.5	3.6
Std (open-start)	20.2	60.7	75.9	9.6	3.2	17.5	52.2	61.6	6.3	2.1	
Mean (start-end)	21.2	-1.5	-10.5	3.7	0.3	0.2	-0.7	-5.7	3.1	0.7	3.7
Std (start-end)	38.3	53.6	65.2	29.6	3.0	0.3	48.8	53.0	19.2	2.9	

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 II presents trade statistics for both the (treated) institutional order 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.

Figure 3 HFT net flow in placebo sample

This figure plots average HFT net flow for the placebo sample (and the main sample for reference). It mirrors Figure 2 but this time includes a placebo sample that consists of trading periods that match the “treated” sample in terms of trading conditions (volume, index return, idiosyncratic return, volatility, and market order imbalance) but this time does not contain trades by any of the four institutional investors. Statistical tests pertain to the differential across treated and placebo stock-periods. 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-period level.



Results of the matched sample analysis. Figure 3 plots HFT net flow for the treated and the placebo sample. The treated sample line is the same as in Figure 2. The only difference is that the dots now denote significance of HFT net flow tested against placebo HFT net flow, as opposed to testing against 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 orders throughout and positive for sell orders in the first eight hours. When compared to the placebo sample, treated buy orders show significantly more against-wind trading in the first five hours and more with-wind trading after eight hours. For sell orders results are weaker, and we only see significantly more with-wind trading between five and seven hours. For sells, there seems to be a with-wind pattern in the placebo sample as well, but this applies to a small sample of orders that are still executing. When pooling buys and sells into pseudo-buy orders,¹² Figure A3 in the online appendix shows that the treated sample has significantly more against-wind flow in the first three hours, and more with-wind flow after seven hours. Taken together, these findings suggest that some of the against-wind pattern is due to market conditions, whereas the with-wind pattern seems truly associated with 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 II), 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

¹²Buy and sell orders can be meaningfully pooled into pseudo-buy orders by multiplying HFT cumulative net flow by minus one for sell orders.

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 mostly stick to market marking in the placebo sample (except for institutional sell orders that last longer than eight hours in which case the placebo pattern becomes erratic as the number of orders still executing is less than 100, see Figure 2).

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, though, 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.

C. Further Analysis of Institutional Order Detection by HFTs

The patterns we documented thus far suggest that HFTs need time to detect large, long-lasting orders in order to trade with-wind on them. Two questions arise naturally at this point. First, the empirical patterns are based on average HFT positions and thus do not rule out that HFTs detect the subset of large, long-lasting orders *right* from the start. Second, even if they do need time to learn about these orders *how* do they learn about them? In other words, are there particular order or market characteristics that somehow reveal their presence? This subsection takes up these questions.

Table III HFT net flow in the lifetime of an institutional order, by subsample

This table documents HFT cumulative net flow in the lifetime of institutional orders where the sample is split based on order size or order duration. For completeness it also contains results for the full sample which are depicted in Figure 2. We only report odd hours for brevity but the complete set of results is available upon request. Each column shows statistics for the orders that are still executing at that point, which explains the decreasing number of observations across columns. 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-period level. The superscripts ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Hours since first child order execution							
	1	3	5	7	9	11	13	15
<i>Panel A: HFT net flow mean in the lifetime of institutional buy orders</i>								
All	-15.9***	-26.5***	-22.5*	11.1	52.8	70.2*	71.5	110.0**
t-stat	(4.2)	(3.4)	(1.8)	(0.6)	(1.3)	(1.7)	(1.6)	(2.0)
N	2,079	1,603	1,236	785	335	293	259	206
Size ≤ median %ADV	-13.8**	-33.4***	-23.3	24.5	145.0*	115.3	36.2	
t-stat	(2.3)	(2.7)	(1.1)	(0.9)	(1.7)	(1.3)	(0.4)	
N	846	634	468	264	63	60	43	
Size > median %ADV	-17.4***	-21.9**	-22.0	4.4	31.4	58.5	78.6	135.7**
t-stat	(3.5)	(2.3)	(1.4)	(0.2)	(0.7)	(1.2)	(1.6)	(2.2)
N	1,233	969	768	521	272	233	216	186
Size ≤ \$1,000,000	-11.3**	-27.5***	-25.5	0.8	51.2	48.7	-9.1	-97.9
t-stat	(2.4)	(2.9)	(1.6)	(0.0)	(0.9)	(0.8)	(0.2)	(1.3)
N	1,185	893	665	387	106	97	79	47
Size > \$1,000,000	-21.9***	-25.2**	-18.9	21.1	53.5	80.8	106.9*	171.5***
t-stat	(3.5)	(2.0)	(1.0)	(0.8)	(1.0)	(1.5)	(1.9)	(2.6)
N	894	710	571	398	229	196	180	159
Duration < 4 hours	-18.0***	-23.6						
t-stat	(2.5)	(0.9)						
N	640	164						
Duration > 4 hours	-15.0***	-26.8***	-22.5*	11.1	52.8	70.2*	71.5	110.0**
t-stat	(3.3)	(3.3)	(1.8)	(0.6)	(1.3)	(1.7)	(1.6)	(2.0)
N	1,439	1,439	1,236	785	335	293	259	206
<i>Panel B: HFT net flow mean in the lifetime of institutional sell orders</i>								
All	5.5	-1.4	-19.1	-46.8	-145.2*	-155.1*	-188.8*	-313.5***
t-stat	(1.0)	(0.1)	(1.0)	(1.5)	(1.9)	(1.7)	(1.7)	(2.5)
N	1,132	820	603	276	142	117	101	77
Size ≤ median %ADV	7.6	-13.7	15.3	49.1				
t-stat	(0.7)	(0.6)	(0.4)	(0.8)				
N	294	197	126	40				
Size > median %ADV	4.8	2.5	-28.2	-63.0*	-184.2**	-191.8**	-245.3**	-358.5***
t-stat	(0.7)	(0.2)	(1.3)	(1.8)	(2.4)	(2.2)	(2.2)	(2.9)
N	838	623	477	236	132	109	93	72
Size ≤ \$1,000,000	2.5	-10.7	-9.1	24.3				
t-stat	(0.3)	(0.6)	(0.3)	(0.5)				
N	426	278	181	59				
Size > \$1,000,000	7.3	3.4	-23.4	-66.1*	-187.0**	-196.5**	-251.9**	-373.6***
t-stat	(1.0)	(0.2)	(1.0)	(1.8)	(2.3)	(2.1)	(2.1)	(2.9)
N	706	542	422	217	127	104	88	69
Duration < 4 hours	14.0	45.7						
t-stat	(1.4)	(1.5)						
N	409	97						
Duration > 4 hours	0.7	-7.7	-19.1	-46.8	-145.2*	-155.1*	-188.8*	-313.5***
t-stat	(0.1)	(0.7)	(1.0)	(1.5)	(1.9)	(1.7)	(1.7)	(2.5)
N	723	723	603	276	142	117	101	77

Detecting large large, long-lasting orders early? We first analyze HFT net flow patterns over time for subsamples based on order size and duration. If the patterns are similar for these subsamples, and in particular around the start of the orders, then indeed HFTs do seem to need time to detect the presence of large, long-lasting institutional orders.

The results in Table III show the following. First, if the sample is split into small and large orders then HFTs trade against-wind initially for both, but only significantly so for buy orders. The sell-order insignificance is not surprising given that there was no such significance in the full sample either. (We report only the odd hours to save space; the even ones show a similar pattern.) These observations hold irrespective of whether the order size is measured relative to a stock's average daily volume (ADV) or measured simply as the order's dollar value. One could even argue that the initial against-wind pattern is stronger for large orders and more significant.

Second, the initial against-wind pattern also holds for both subsamples when they are based on order duration as opposed to size. Long-lasting orders are defined as those that last longer than four hours (half a trading day). The result is again significant only for buy orders. Together, they suggest that HFT cannot distinguish between large or long-lasting orders early in the lifetime.

What order or market covariates explain detection? Given that HFTs need time to learn, what order characteristics or market conditions seem conducive to detection? We perform two regressions to generate some insight. The dependent variable is either signed HFT net flow in the lifetime of the order, or a dummy variable that is equal to one for the tercile of institutional orders for which HFTs traded strongly with-wind. The latter regression is a Probit one. The explanatory

Table IV Determinants of cumulative HFT net flow over the lifetime of the institutional order

The dependent variable is HFT net flow accumulated over the lifetime of the institutional order, signed by the direction of the order. The variable is standardized to have zero mean and unit variance, and it is measured in dollars (first column) or transformed into a dummy variable that is equal to one if an observation belongs to the tercile of highest (with-wind) values (second column). The dummy specification is estimated with Probit, and we report marginal effects evaluated at the means of all variables. Buys and sells are pooled to create pseudo-buys. The explanatory variables are mostly self-explanatory and should become clear with the following notes. Order size and ADV (based on the full sample) are both measured in shares. The dummies “Start at U.S. open” and “End at U.S. open” are equal to one if the order starts or ends, respectively, in the thirty-minute interval after the opening of the U.S. market. Turnover and volatility are measured from the start to the end of the order. Turnover is the stock volume divided by shares outstanding and volatility is realized volatility based on one-minute midquote returns. The market concentration measure HHI is the Herfindahl-Hirschman Index, defined as the sum of daily squared market shares across lit trading venues. All explanatory variables are standardized. Regressions include stock and institution fixed effects and standard errors are clustered by stock-date. *t*-values are in parentheses. Variable units are in brackets. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Signed HFT Net Flow (\$)	With-wind tercile
<i>Order-specific variables</i>		
Order size relative to ADV (%)	0.00 (0.1)	0.01 (1.5)
Order duration (hours)	0.05 (1.5)	0.05*** (3.7)
Trades through market orders (%)	0.01 (0.6)	0.00 (0.2)
Meta order is buy dummy	-0.12** (-2.3)	-0.07** (-2.4)
Start at U.S. open dummy	0.03 (0.8)	-0.06** (-2.2)
End at U.S. open dummy	-0.08* (-1.7)	0.01 (0.3)
Closing auction dummy	-0.07 (-1.5)	0.02 (0.9)
<i>Market-specific variables</i>		
Stock volatility (%)	-0.01 (-0.7)	-0.00 (-0.4)
Turnover (%)	0.08** (2.1)	0.04*** (2.7)
Market concentration HHI (%)	-0.25 (-0.7)	0.09 (0.6)
Observations	5,136	5,136
R-squared	0.012	
Pseudo R-Squared		0.06

variables capture either order or market characteristics. Order characteristics include size, duration, buy or sell, the use of market orders, but also time dummies that identify whether an order starts or ends around the U.S. open, or ends in the closing auction. The market characteristics include volatility, turnover, and a market fragmentation measure. Table IV contains the regression results as well as a detailed description of these variables.

The regression results show that whether HFTs collectively trade with-wind on an order is hard to predict. The explanatory variables are mostly insignificant, except for the following ones. Longer duration orders are more likely to show HFT with-wind trading, which might simply mean that a long and persistent imbalance in the market is harder to attribute to “normal market conditions.” With-wind trading further seems to be more prevalent on high turnover days.¹³ With-wind trading is less likely when institutional orders start at the U.S. open or end at the U.S. close. Either detection might be harder at such informationally intensive periods, or taking a position is more costly at these disproportionately volatile periods. Finally, with-wind trading appears more likely for institutional sell orders (as opposed to buy orders). As mentioned in our discussion of Figure 2, this might be due to the observation that when investors sell they trade more intensively than when they buy. This arguably raises detection risk for sell orders.

The general observation of with-wind trading being hard to predict might not be that surprising if one frames the trading patterns as an interplay between institutional investors and HFTs. The latter benefit from learning about an institutional order executing in order to trade with-wind on it. This might hurt investors who will, in response, do their utmost to camouflage their trades. If they

¹³Further analysis in Section IV.B will show that the institutional orders in our sample seem information-motivated. All else equal, it is likely that an investor trades more intensively on his information at times of high turnover.

do so successfully, it is hard to predict HFT behavior with the standard explanatory variables (i.e., investors might even randomize order-specific variables). Or, to put it more strongly, if one of these variables *was* strongly predicting HFT with-wind trading, then investors have a strong incentive to adjust their strategy. We will further develop this interpretation of investor-HFT interplay in Section V, and perform additional analysis that is motivated by it.

IV. Predatory Trading or Back-running?

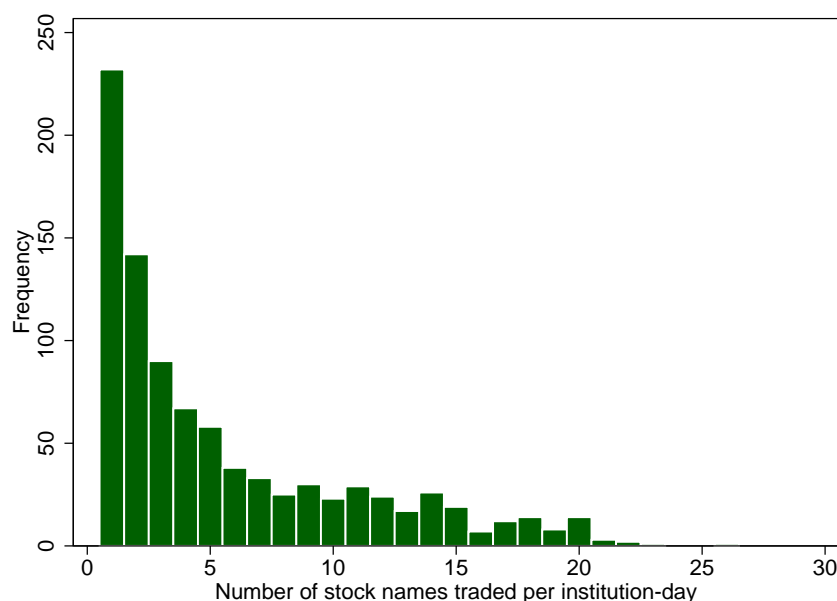
The patterns of how HFTs trade around institutional orders, documented in the previous section, do not simply unequivocally support one of three theoretical patterns summarized in Figure 1: market making, predatory trading, or back-running. Rather, the HFT against-wind pattern is consistent with market making, but is the with-wind pattern a sign of predatory trading or back-running? To answer this question we exploit an important distinguishing feature between the two: Predatory trading hypothesizes that an uninformed investor trades to satisfy a pressing liquidity need, whereas back-running assumes that the investor trades to capitalize on a private (informational) signal. In this section we employ a standard microstructure technique to assess whether at least some institutional orders are information-motivated. And, if so, are these the ones that HFT back-run on?

A. *Plain-Vanilla Trading of the Market Portfolio or Stock Picking?*

But, before turning to a full blown price-impact analysis, some exploratory analysis is useful. If these institutions are uninformed investors trading for liquidity reasons, one expects them to

Figure 4 How many names did an investor trade on days that he traded?

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



trade the market portfolio. If instead, they trade on private signals then one would expect them to *not* trade the market portfolio (as the market is more cheaply traded through index futures), but small sets of stocks, or even a single stock. This broad distinction suggests computing how many stock (names) an institution trades on days that it trades. This is where we turn next.

Figure 4 plots the histogram of how many names an institution trades when it trades. Trading the market portfolio would entail trading all the 30 constituent names. The histogram reveals that none of the institutions ever traded the entire market portfolio — far from it: They most often traded a single stock only. On average, they traded 5.6 names, with a standard deviation of 5.3. We now turn to a price-impact analysis to study whether such trading could be interpreted as “stock picking” based on information.

B. Information in Institutional Orders and Its Correlation with HFT Net Flow

To measure the informational content of institutional orders we extrapolate from the standard microstructure “spread decomposition” approach (Glosten, 1987; Hendershott, Jones, and Menkveld, 2011). This decomposition is used to identify the information in a *single* trade. This is done by simply waiting long enough for the transitory component in the trade’s price impact to disappear, which is often taken to be five minutes. More specifically, the spread decomposition measures the “effective spread” of a (market) buy trade as its immediate price impact: trade price minus the midquote price that prevailed just before the execution, the pre-trade midquote. This trade’s (permanent) “adverse-selection” component is the midquote five minutes later minus the pre-trade midquote. The residual (transitory) component is referred to as the trade’s “realized spread.” Sell trades are pooled with buy trades by multiplying prices by minus one and thus make them “pseudo buys.” Estimates of the effective spread and its two components relies on averaging them across a large sample of trades.

An equivalent price-impact analysis to assess the informational content of institutional orders that comprise *multiple* trades proceeds as follows. The overall price impact (OPI) for buy orders is defined as the transaction price of the final child trade minus the midquote that prevailed just before execution of the first child trade, the pre-order midquote. The decomposition into transitory and permanent price impact (TPI and PPI, respectively) is then done by waiting 24 hours after the last child executed (i.e., the same time on the next trading day) and use the midquote that prevails then.¹⁴ Sell orders are pooled with buy orders after converting them to pseudo buys. The price

¹⁴A wait longer than 24 hours could improve accuracy, but at the cost of reduced statistical power, since the fundamental value continues to evolve. We believe a 24-hour wait strikes a good balance, since longer waits imply

impacts is measured in basis points, and can be converted to dollars by multiplying it by the dollar-size of the institutional order. In all subsequent price-impact analysis we focus on idiosyncratic returns and therefore subtract off from prices the stock's beta times the market return. The reason is that we expect investors to trade on firm-specific information because, if they trade, they trade either a single stock or only a few stocks (see Figure 4). Accordingly, the exact definitions of OPI, PPI, and TPI are:

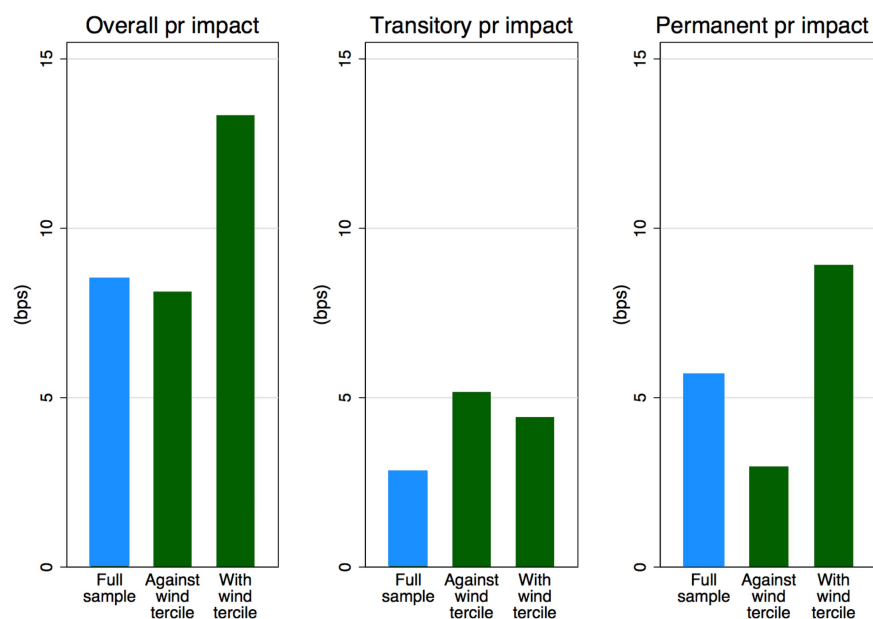
$$\begin{aligned}
 \text{OPI} &= D \left(\frac{p_{\text{end}} - p_{\text{start}}}{p_{\text{start}}} - \beta^M \frac{p_{\text{end}}^M - p_{\text{start}}^M}{p_{\text{start}}^M} \right) \times 10,000, \\
 \text{PPI} &= D \left(\frac{p_{\text{end}+1\text{day}} - p_{\text{start}}}{p_{\text{start}}} - \beta^M \frac{p_{\text{end}+1\text{day}}^M - p_{\text{start}}^M}{p_{\text{start}}^M} \right) \times 10,000, \\
 \text{TPI} &= \text{OPI} - \text{PPI},
 \end{aligned} \tag{1}$$

where D is one for institutional buy orders and minus one for institutional sell orders, and p refers to the share midquote price at start or the end of execution period of the order, while p^M refers to the index level.

Figure 5 illustrates the information content for institutional orders sorted by terciles of HFT cumulative net flow, calculated from order start to end. The figure shows that the information content in institutional orders is large on average, and larger for those where HFT trade with-wind. The overall price impact is 8.5 basis points on average, with 8.1 basis points for the tercile of against-wind orders and 11.4 basis points for the tercile of with-wind orders. The permanent price impact is larger than the transitory impact, 5.7 vs. 2.8, respectively. Interestingly, this wedge is qualitatively similar but statistically insignificant results.

Figure 5 Institutional order's overall, transitory, and permanent price impact, by HFT net flow tercile

This figure plots the average overall price impact (OPI), decomposed into a transitory price impact (TPI) and a permanent price impact (PPI). OPI is the stock return over the lifetime of the order. PPI is the return from the start of an order until 24 hours *after* the end of it. TPI is OPI minus PPI. All variables are based on idiosyncratic returns, meaning we subtract beta times the market return from the raw return. OPI, TPI, and PPI are computed by HFT cumulative net flow terciles. The impacts are plotted separately for the full sample, and the against-wind and with-wind HFT net flow terciles. The HFT net flow is computed over the lifetime of the order. The lowest tercile has strongly negative values (against-wind), the middle one is mostly zero (neutral), and the highest one has strongly positive values (with-wind). Sell order observations are converted to pseudo buys by multiplying where appropriate by minus one. They can then be pooled with buy orders.



larger for with-wind orders, suggesting that, in the cross-section of orders, HFTs trade with-wind on the more informed orders. This finding however might be driven by alternative factors, for example order size, which is why we turn to regression analysis next to control for the effect of various covariates.¹⁵

Table V presents the results of price-impact regressions and shows that the positive correlation between permanent price impact and HFT with-wind trading is robust. The main explanatory variables of interest in these regressions are with-wind and against-wind HFT cumulative net flow. The with-wind variable is defined as HFT cumulative net flow if (in the lifetime of the order) it carries the same sign as the institutional order, and zero otherwise. The against-wind variable is defined analogously, such that total HFT flow equals with-wind flow minus against-wind flow. Standard covariates are added as controls.

The OPI regressions show that HFT with-wind flow's coefficient is positive, both when these variables are expressed in relative terms or in dollar terms. Rerunning these regressions for OPI's two components separately reveals that PPI is driving this result. Stronger with-wind HFT flow is associated with more informed orders in the cross-section. This association is economically large as a one standard deviation increase in with-wind flow implies a 9.28 basis points higher PPI, which is high given that the overall PPI average is 5.69 basis points.

The regressions further show that the consistently significant covariate is order size. Its coefficient is significantly positive for both overall and permanent price impact, and is insignificant

¹⁵Note that the heterogeneity of orders, the against-wind ones seem liquidity-motivated and the with-wind ones information-motivated, seems to call for a next generation of multi-period optimal-execution models that consider such heterogeneity explicitly. An early example of such model is Choi, Larsen, and Seppi (2017).

Table V Institutional order's overall, transitory, and permanent price impact regressed on HFT net flow and control variables

This table presents regression results where the dependent variable is an institutional order's overall price impact (OPI), its transitory price impact (TPI), or its permanent price impact (PPI). Sell order observations are converted to pseudo buys by multiplying where appropriate with minus one. OPI is the stock return over the lifetime of the order. PPI is the return from order start until 24 hours *after* its end. TPI is OPI minus PPI. All variables are based on idiosyncratic returns, meaning we subtract beta times the market return from the raw return. The main explanatory variable is HFT net flow accumulated over the lifetime of the institutional order. It is standardized, and expressed either relative to shares outstanding (%) or in dollar terms (\$). Values in the same (opposite) direction as the institutional order are assigned to with-wind (against-wind) net flow and get a positive sign. The variables are zero otherwise. Order- and market-specific variables are added as controls. Order size and ADV (based on the full sample) are both measured in shares. The dummies "Start at U.S. open" and "End at U.S. open" are equal to one if the order starts or ends, respectively, in the thirty-minute interval after the opening of the U.S. market. Turnover and volatility are measured from order start to end. Turnover is the stock volume divided by shares outstanding and volatility is realized volatility based on one-minute midquote returns. The market concentration measure HHI is the Herfindahl-Hirschman Index, defined as the sum of daily squared market shares across lit trading venues. All explanatory variables are standardized. Regressions include stock and institution fixed effects and standard errors are clustered by stock-date. *t*-values are in parentheses. Variable units are in brackets. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	OPI (bps)	TPI (bps)	PPI (bps)	OPI (\$)	TPI (\$)	PPI (\$)
<i>Order-specific variables</i>						
Against-wind HFT cum net flow (%)	-4.74* (-1.7)	1.78 (0.5)	-6.53 (-1.5)			
With-wind HFT cum net flow (%)	6.50* (1.9)	-2.78 (-0.8)	9.28** (2.1)			
Against-wind HFT cum net flow (\$)				-0.01** (-2.1)	-0.00 (-0.6)	-0.01 (-0.8)
With-wind HFT cum net flow (\$)				0.01** (2.5)	-0.00 (-0.7)	0.02** (2.3)
Order size relative to ADV (%)	11.33*** (5.8)	2.79 (1.3)	8.55*** (3.0)	0.02*** (5.9)	0.00 (1.2)	0.02*** (3.1)
Order duration (hours)	0.99 (0.3)	2.40 (0.8)	-1.42 (-0.3)	0.00 (0.1)	0.00 (0.8)	-0.00 (-0.4)
Trades through market orders (%)	1.70* (1.8)	-0.19 (-0.1)	1.89 (1.0)	0.00 (1.4)	-0.00 (-0.3)	0.00 (1.0)
Meta order is buy dummy	-1.33 (-0.5)	-8.02 (-1.3)	6.69 (1.0)	-0.00 (-0.1)	-0.02 (-1.4)	0.02 (1.2)
Start at U.S. open dummy	1.01 (0.4)	1.05 (0.2)	-0.04 (-0.0)	0.00 (0.1)	0.00 (0.1)	-0.00 (-0.0)
End at U.S. open dummy	-9.65*** (-2.7)	-8.53 (-1.4)	-1.12 (-0.2)	-0.02** (-2.5)	-0.02 (-1.6)	0.00 (0.1)
Closing auction dummy	-3.20 (-0.8)	-5.03 (-1.0)	1.83 (0.3)	-0.00 (-0.3)	-0.01 (-1.5)	0.01 (1.0)
<i>Market-specific variables</i>						
Stock volatility (%)	1.13 (0.6)	-0.77 (-0.4)	1.89 (0.7)	0.00 (0.3)	-0.00 (-0.9)	0.00 (0.8)
Turnover (%)	-1.26 (-0.3)	-0.91 (-0.2)	-0.35 (-0.1)	-0.00 (-0.6)	-0.00 (-0.5)	-0.00 (-0.1)
Market concentration HHI (%)	-0.78 (-0.0)	19.82 (0.5)	-20.60 (-0.5)	-0.01 (-0.3)	0.07 (1.0)	-0.08 (-1.0)
Observations	5,136	5,136	5,136	5,136	5,136	5,136
R-squared	0.030	0.002	0.008	0.030	0.002	0.008

for transitory price impact. This is consistent with institutional investors using their private information strategically, trading larger quantities when having stronger signals. Note further that the with-wind flow's positive association with the information content of orders (PPI) is therefore not simply an artifact of with-wind activity for *larger* orders. The regression analysis shows that when size is controlled for, the positive association remains.

If the institutional order ends in the half hour after the opening of the U.S markets, the OPI is significantly lower. This is likely due to higher overall trading activity in this period and improved liquidity, as the TPI regression suggests that most of the reduction comes from the transitory component.

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, the with-wind activity is strongest for the most informed ones among them. We caution that our results do not reject the predatory trading hypothesis *in general*. Predatory trading requires stressed orders (e.g., forced liquidations) and is therefore more likely to surround orders by highly leveraged institutions (e.g., hedge funds).¹⁶ Such institutions are not part of our sample, which consists of orders by two mutual funds, a sovereign wealth fund, and a pension fund.

¹⁶Barbon et al. (2017) offer evidence suggesting that brokers who execute large positions on behalf of clients at fire sale prices may share this information with other clients, who in turn prey on these large orders. The incentive for the broker, they argue, is to establish a reputation of having valuable information.

V. Putting It All Together: An Economic Interpretation of With-Wind Trading

We documented that HFT trade against-wind early in the lifetime of institutional orders, and turn to with-wind trading later on the most informed ones. This final section develops an economic perspective that could rationalize the with-wind trading by HFTs. It first develops predictions based on economic reasoning and then adds further empirical evidence to verify whether these predictions hold up in the data.

A. *The Interplay between Investors Trading on Information and HFTs*

We conjecture that the with-wind trading pattern is driven by an interplay between two types of agents who both trade strategically:

- HFTs are continuously looking for signals on information-motivated trading by investors in order to prey on it.
- Investors are aware of this and try to hide their private signals, but face a trade-off: Trading at high intensity yields a large speculative position (and thus raises profit) but it also increases the risk of being detected and preyed on (which reduces profit).

This interplay leads us to articulate the following four predictions that will be taken to the data in the next four subsections.

Pred1: Investors trade cautiously on their private information and prices therefore are slow to reveal it.

Pred2: HFTs prey on such information-motivated trading, need time to detect it, but ultimately are able to identify among all long-lasting orders the most informed ones.

Pred3: Investors are aware of HFT detection risk, and respond to detection by decelerating their trading.

Pred4: Investors do indeed experience larger implementation shortfall after being detected, and HFTs generate positive gross trading revenue.

B. Pred1: Speed with Which Investor Information Is Revealed in Prices

The first prediction states that investors trade cautiously on their private information so that prices are slow to reveal it. To measure how successful investors are in this respect, we turn to “unbiasedness regressions.” They have been used before to study the speed of learning (Frenkel, 1977; Hodrick, 1987; Biais, Hillion, and Spatt, 1999). The general specification is:

$$v = [V - E(V|I_0)] = \alpha_t + \beta_t [P_t - E(V|I_0)] + \varepsilon_t, \quad (2)$$

where V is the fundamental value to be learned, $E(V|I_0)$ is its best estimate at the time learning starts, v therefore is the value yet to be discovered at time 0, and P_t is the market price at time t . Performing this regression for various t and plotting the alpha and beta estimates as a function of t yields insight in the speed of learning. Biais, Hillion, and Spatt (1999), for example, use it to study whether the series of indicative prices in the 90 minutes before the Paris Bourse opens reflect information or if they are just noise. In their analysis, v is the close-to-close return, and $[P_t - E(V|I_0)]$ is the close-to-indicative-pre-opening-price-at- t .

Implementing unbiasedness regressions in our sample. We will use these unbiasedness regressions to study how quickly prices reflect possible private information embedded in the institutional investors' orders. In this application we proxy the dependent variable v with the permanent price impact (i.e., PPI in equation (1)). The explanatory variable $(P_t - E(V|I_0))$ is taken to be the midquote at time t minus the midquote that prevailed at the start of the meta-order.¹⁷ The analysis is done for idiosyncratic returns (consistent with the PPI analysis in Section IV.B) and pools sell orders with buy orders by making them pseudo buys (i.e., by changing signs appropriately). The regression that is taken to the data therefore is:

$$PPI_i = \alpha_t + \beta_t (MidQuote_{it} - MidQuote_{i0}) + \varepsilon_{it}, \quad (3)$$

where i indexes all (pseudo) buy orders, t is time relative to the start of order execution. Estimates for α_t and β_t are obtained by fixing t and exploiting the cross-section (i.e., the number of observations equals the number of orders).

Before turning to the empirical results, it is useful to develop the estimators for alpha and beta a bit further to help interpreting the results. It will turn out to be useful to decompose the regressor into a martingale and a pricing error component. Let

$$m_t = E(V|I_t) - E(V|I_0), \quad (4)$$

be the martingale component¹⁸ and rewrite the regressor of equation (3) as

$$MidQuote_{it} - MidQuote_{i0} = m_t + e_t \quad (\text{i.e., } \underline{\text{martingale term plus error term}}), \quad (5)$$

¹⁷We assume that this midquote at time zero is the best estimate of V at that time: $P_0 = E(V|I_0)$.

¹⁸It is a martingale by construction as takes the expectation of a random variable on expanding information sets.

where the error term e_t is implicitly defined as the residual after taking out the martingale component. The estimators for alpha and beta can then be written as

$$\hat{\alpha}_t = v - \hat{\beta}_t (m_t + e_t) \quad \text{and} \quad \hat{\beta}_t = \frac{\text{cov}(v, m_t) + \text{cov}(v, e_t)}{\text{var}(m_t) + \text{var}(e_t) + 2\text{cov}(m_t, e_t)}. \quad (6)$$

These expressions lead to the following insights. First off, if the price at the end of the meta-order execution reflects its information perfectly (as in, e.g, Kyle, 1985), then $\hat{\beta}_T = 1$ and $\hat{\alpha}_T = 0$. Note further that $\hat{\alpha}_0 = v$, as at the start of execution no private information has entered the price yet and the intercept equals the average level of private information (PPI). The regressor is trivially zero at this point and $\hat{\beta}_0$ is therefore not defined (i.e., the regression turns to a trivial one with only an intercept at $t = 0$). Second, in the hypothetical case of no pricing errors then $\hat{\beta}_t = 1$ and $\hat{\alpha}_t$ reveals how much of the information v remains to be learned. (Note that this case is hypothetical because economic models need noise trading or private-value based trading as otherwise one lands in the no-trade theorem (Milgrom and Stokey, 1982).) Third, if there are pricing errors and if they are orthogonal to the information process (driven by m_t and v), then beta is below one and, all else equal, alpha tends to reveal v at a slower rate.¹⁹

Estimates. Figure 6 plots the alpha and beta estimates separately for the tercile of with-wind orders and the tercile of against-wind orders. The with-wind orders are largely information-motivated (see Section IV.B) and the plot should therefore reveal the speed of learning. The against-wind or-

¹⁹Technically, this argument holds trivially if PPI equals v but allows for PPI to be a noisy proxy for v as long as this noise does not correlate with the pricing errors e_t that arise during the lifetime of the order. In other words, we assume that any such error e_t dies out within 24 hours after the last child trade.

ders are largely uninformed (again Section IV.B) and therefore serve as a benchmark. We run the regressions with t varying from zero to four hours. To keep a balanced sample, all these regressions are done for the subsample of orders that last at least four hours.

The estimated alpha and beta patterns lead to a couple of observations. First, the alpha estimates for the with-wind orders start at about 16 basis points which is equal to the average PPI for these orders. The private signal is revealed slowly over time, and after about two hours one-third appears to be revealed. Alpha keeps declining gradually afterwards, but becomes statistically insignificant. Half of the private information seems to be revealed after about three hours. This slow revelation could explain why HFTs need seven hours to detect the institutional order and trade with-wind on it (see Section III). For against-wind orders, the alpha estimates are insignificant likely because these orders are largely uninformed.

Second, the beta estimates are often significantly below one in the early hours but eventually, after three and half hours, they are statistically indistinguishable from one. The other salient feature is that they fall further below one for with-wind orders. These patterns suggest that pricing errors are relatively large in the early hours, and more so for with-wind orders. The stronger pattern for with-wind orders suggests that investors do seem to trade very carefully on their information and let pricing errors grow to obfuscate their signal. Or, alternatively, for against-wind orders more of the long-term price impact is public information, which raises the covariance between m_t and v , thus bringing beta closer to one, all else equal.

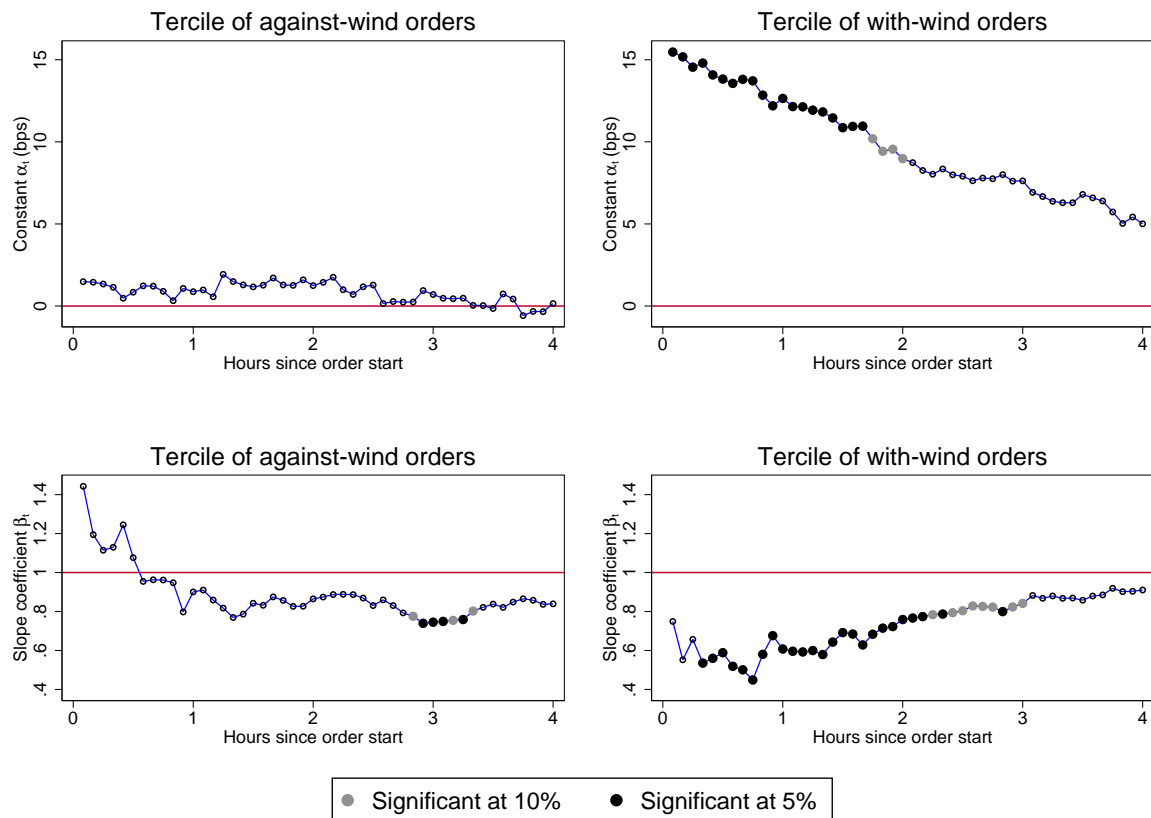
In sum, these empirical results are consistent with the first prediction that investors trade stealthily, so that prices are slow to reveal it.

Figure 6 Unbiasedness regressions to measure speed of information revelation

This figure illustrates the speed of information revelation in the first four hours, for the subset of orders that lasted longer than four hours. It does so separately for the tercile of orders where HFTs traded against-wind and the tercile of orders where they traded with-wind (calculated over the lifetime of the order). The unbiasedness regression that is estimated is:

$$PPI = \alpha_t + \beta_t (P_t - P_0) + \varepsilon_t,$$

where PPI is the long-term price impact of the order (see equation (1)), t measures the time elapsed since the meta order started executing in increments of five minutes, $P_t - P_0$ is the midquote return from the start of order execution until time t . Both PPI and the midquote return are idiosyncratic (i.e., the market return component is removed). The analysis is done on the pooled set of orders where sell orders are converted to pseudo-buy orders. Alphas are tested in terms of being statistically different from zero. Betas are tested relative to one.



C. Pred2: HFT With-Wind Trading and Information in Long-Lasting

Orders

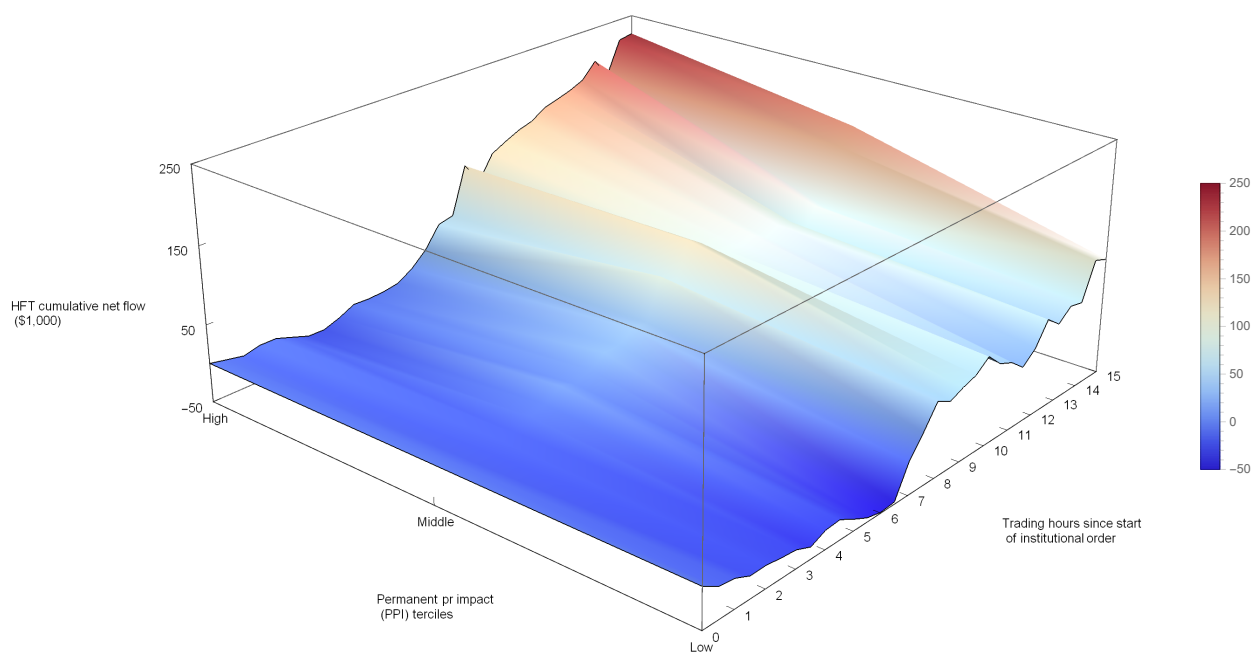
The Section IV.B regressions showed that with-wind trading by HFTs correlates with how informed institutional orders are, and the unbiasedness regressions in Section V.B showed that this information is revealed in prices very slowly. In spite of such slow revelation, HFTs do seem able to detect these orders and trade along with them. But, how skilled are these HFTs? In particular, when do they start trading with-wind for these informed orders? And, does HFT net flow evolve differently over time for informed versus uninformed orders?

Figure 7 illustrates HFT cumulative net flow in the lifetime of institutional orders split into terciles, based on their informational content as measured by PPI. The figure is essentially a 3D version of the signature plot in Figure 2, with information on the additional axis. Sell orders are pooled with buy orders by converting them to pseudo-buy orders.

The figure illustrates that HFTs appear to trade with-wind only on the most informed of long-lasting orders. HFTs cumulative net flow reaches its highest level after about 15 hours for the tercile of high PPI (i.e., most informed) orders. The regression results that accompany this graph are reported below it and reveal that these levels are statistically significant. They turn significantly positive after ten hours of trading. The with-wind levels are substantially lower for the other terciles and, more importantly, they are statistically insignificant. Finally, note that for all three terciles HFTs start out leaning against the wind in the early hours. This against-wind trading is statistically significant, albeit marginally so for the high PPI tercile. Note that after between five and seven hours, all terciles show positive levels which is consistent with the turning point reported

Figure 7 HFT net flow in the lifetime of an institutional order by information content

This figure plots how HFT cumulative net flow develops in the lifetime of institutional orders, where the latter are split in terciles based on their informational content as measured by PPI. It essentially is the 3D version of Figure 2 with informational content on the additional axis. Sell orders are pooled with buy orders by converting them to pseudo-buy orders. The panel below the graph reports coefficient estimates for the odd hours (to conserve space) to reveal what part of it is statistically significant. Each column shows statistics for the orders that are still executing at that point, which explains the decreasing number of observations across columns. The analysis is done through the fifteenth hour of execution to retain at least 50 observations in each tercile.



	Hours since first child order execution							
	(1)	(3)	(5)	(7)	(9)	(11)	(13)	(15)
<i>Low PPI tercile</i>								
Mean	-14.5***	-27.7**	-35.8**	-8.1	42.8	36.1	48.9	98.5
t-stat	(2.7)	(2.5)	(2.0)	(0.3)	(0.8)	(0.6)	(0.7)	(1.3)
N	1091	856	643	377	176	153	133	109
<i>Middle PPI tercile</i>								
Mean	-13.9***	-1.8	17.9	42.9	112.0	90.9	59.8	178.4
t-stat	(2.5)	(0.2)	(0.9)	(1.5)	(1.4)	(1.1)	(0.6)	(1.3)
N	968	688	494	271	110	95	85	65
<i>High PPI tercile</i>								
Mean	-8.7	-18.5*	-2.9	31.7	96.6	151.5***	183.2***	224.4***
t-stat	(1.6)	(1.8)	(0.2)	(1.2)	(1.6)	(2.5)	(2.7)	(2.9)
N	1152	879	702	413	191	162	142	109

in Figure 2.²⁰

The evidence presented here is consistent with the second prediction. There is cross-sectional heterogeneity in the long-lasting orders. Some seem clearly more information-motivated and HFTs only significantly trade with-wind on those ones.

D. Pred3: Trading Before and After HFT Detection

As HFTs seem able to detect the informed orders, are investors aware of such detection risk? Do they decelerate trading after being detected? And, is their trading before detection consistent with trading off a higher speculative profit with heightened detection risk? Such trade-off is admittedly hard to identify in the data. Yet, if their pre-detection trading is somehow more intense as compared to their overall trading and they reduce this intensity after being detected, then such pattern is at least consistent with strategic trading in full awareness of HFT detection risk. Such pattern is predicted in Pred3 and verified empirically in this section.

To study trading before and after detection, the trading period of with-wind orders is split into two sub-periods: before and after detection by HFTs. We identify detection as the time, going backwards from the end of a with-wind order, that HFT cumulative net flow crossed zero. The pre- and post-detection periods are labeled pre-WW and post-WW, respectively. We characterize pre-WW and post-WW trading by generating three sets of trade statistics. The first set focuses on how the institutional investor traded, the second one characterizes overall trading, and the final set zooms in on trading by HFTs. These trade variables are also computed for the overall sample (i.e.,

²⁰We report regression coefficients only for odd hours to conserve space, but they are representative of all hours. The full regression results are reported in Table AIII in the online appendix.

Table VI 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. Child trades are signed using bulk volume classification (Easley, de Prado, and O'Hara, 2016). Specifically, if the one-second return around a child trade is positive, a purchase is labeled market order and a sale is labeled limit order. For the imbalance variable, buy (sell) indicates the total market volume where the buyer (seller) initiated the transaction. Volatility and variance ratios are computed based on midquote returns. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Variable	Overall avg	Pre-WW	Post-WW	Diff	
Investor	Inst volume (1,000 shares per 5 min)	3.39	6.53	3.02	-3.51	***
	Number of child trades (per 5 min)	3.97	5.92	3.31	-2.61	***
	Size child trades (1,000 shares)	1.60	2.77	1.31	-1.46	***
	Investor trades through market orders(%)	40.41	41.48	38.96	-2.52	***
	Investor participation in share volume (%)	24.24	20.17	10.51	-9.66	***
	Std(5-min "Investor participation in share volume")	8.86	14.25	8.84	-5.41	***
Market	Stock volume (1,000 shares per 5 min)	30.40	40.11	38.98	-1.13	
	Order imbalance, buy-sell /(buy+sell) (%)	6.49	14.35	7.91	-6.44	***
	Volatility (bps per 5 minutes)	12.83	14.18	15.20	1.02	***
	Variance ratio (1-min-ret var/ 60 times 1-sec-ret var)	0.87	0.88	0.86	-0.02	***
	Variance ratio (5-min-ret var/ 5 times 1-min-ret var)	0.83	0.84	0.83	-0.01	
HFTs	HFT trades through market orders (%)	39.04	41.03	45.51	4.48	***
	HFT participation in share volume (%)	22.64	26.02	25.66	-0.36	

including against-wind orders) to provide benchmark values.

The results are presented in Table VI and lead to the following observations. First, the investor statistics suggest that he does trade intensively in the pre-WW period. For example, in that period he trades 6,530 shares per five minutes which is about 50% higher than the investor overall average. This higher volume rate is driven both by trading more often and by trading in larger sizes. Pre-WW he trades six instead of four times per five minutes. His trade size is 2,770 shares pre-WW against his overall average of 1,600 shares per trade. This intense trading however does not lead to

an elevated participation in volume.²¹ His participation pre-WW is about 20% compared to 24% overall.²² Finally, the more intense pre-WW trading is further evidenced by the extent to which trades were done by market orders (as opposed to passive orders, i.e., limit orders): 41% instead of 40%.²³ Incidentally, this small variation could explain why this attribute turns insignificant in explaining PPI in Table V.

Second, investors do seem to decelerate strongly after HFTs turn to trade along with them. Comparing their pre- and post-WW trading, one notices that he trades significantly less intensive on all of his trade measures. His volume declines by 54%, trade frequency by 44%, trade size by 53%, his volume participation by 48%, and the fraction of market orders by 5%.

Third, we measure to what extent the investor actively managed execution by benchmarking against the commonly used volume-weighted-average-price (VWAP) algorithm. The perfect implementation of VWAP keeps the volume participation rate flat throughout the execution period. In other words, the standard deviation of this rate sampled, say, every five minutes is expected to be zero. In the overall sample, this standard deviation is about 9% which we do not consider excessive given a mean participation rate of 24% (i.e., the investor seems to often resort to a VWAP strategy). Interestingly, pre-WW this standard deviation increases to 14% to then drop significantly

²¹All volume measures in the paper are single-counted.

²²Note that participation is far above order size divided by ADV as reported in Table I. This wedge is mostly driven by short-duration orders where volume participation can be high, yet size expressed in ADV units is low.

²³This finding of the investor trading at above-average intensity (at least for the first few hours, i.e., pre-detection) on with-wind orders, combined with the observation that these orders are the most informed ones (Section IV), suggests that order size associates positively with PPI. Figure A4 in the online appendix relates the two directly and shows that this is indeed the case.

back to 9% post-WW. We consider this apparent strong deviation from VWAP in the pre-WW period as consistent with the investor optimally trading on a private signal (as is evidenced by the slow revelation result of Section V.B).

Fourth, HFTs trade more aggressively after the detection point. Their relative number of trades through market orders increases significantly from 41% to 46%. Their participation in overall volume does not change significantly when comparing post-WW with pre-WW. They seem to resort to market orders to quickly change their position from against-wind to with-wind.

Incidentally, this analysis of pre-WW and post-WW trading separately could explain why results in this subsection are stronger than the exploratory analysis in Table IV. The latter analysis relates with-wind trading by HFTs to explanatory variables measured over the *entire* lifetime of the order. Pre- and post-WW trading are then lumped together. Strong pre-WW trading by investors, for example, is then moderated by their weak post-WW trading.

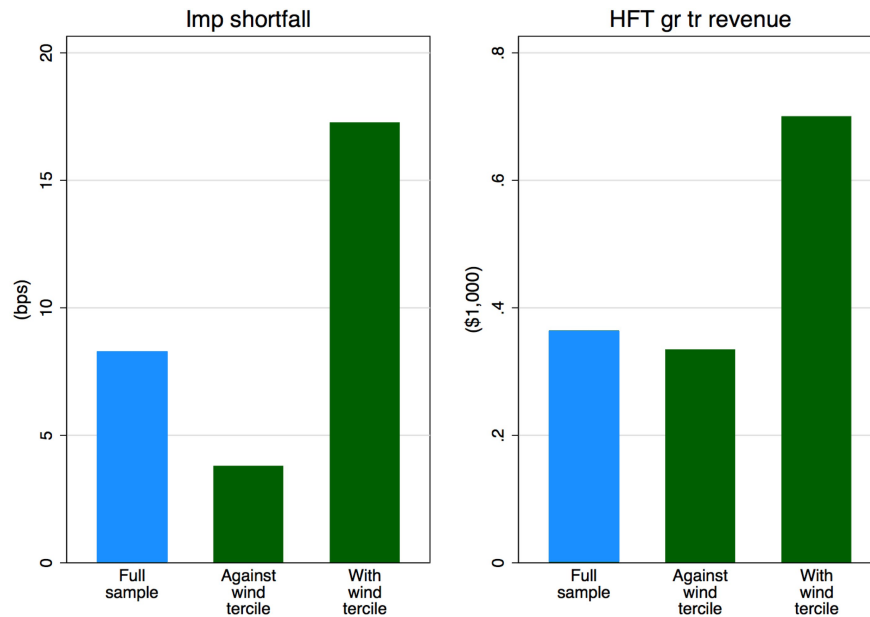
These results provide some empirical support consistent with the third prediction. Investors seem aware of detection risk as they decelerate their trading strongly after HFTs turn to with-trading. They seem to willingly take on this risk, probably to raise their speculative profit, as they do trade more intensively prior to being detected. Pre-detection, they trade more often and in larger size.

E. Pred4: Does With-Wind Trading Correlate with Investor Cost and HFT Profit?

The final prediction states that detection of the investor's informed orders by HFTs is costly to them, and profitable to HFTs. This completes the characterization of the investor-HFT interplay

Figure 8 Implementation shortfall and HFT gross trading revenue by HFT net flow terciles

This figure plots an institution's implementation shortfall and the HFTs' gross trading revenue averages for the full sample, and the against-wind and with-wind HFT net flow terciles. Buy and sell meta orders are pooled after converting sell-order observations to pseudo buys by multiplying with -1 where needed. All HFT net flow observations are then sorted and put into equal-size 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.



as the investor trades off costly detection with profitable speculation. To study this prediction empirically, we compute the (conditional) correlations between HFT with-wind trading and both the implementation shortfall on investor orders and HFT gross trading revenue.

Before turning to regressions to compute conditional correlations, we first plot the bivariate relationship for both variables with the type of HFT trading. Figure 8 stratifies the sample by HFT cumulative net flow tercile (similar to Figure 5) and plots the full-sample average and averages for the against-wind and the with-wind tercile. It shows that implementation shortfall is four basis points below average for against-wind flow and nine basis points above it for with-wind flow.²⁴

²⁴We further note that the implementation shortfall in our sample is the same order of magnitude as that reported

The strongly positive association of with-wind trading and transaction cost is consistent with the conjectured trade-off. We caution that this effect is likely to be overstated whereas the effect of against-wind is understated because of endogeneity. These biases result from reverse relationships. For example, against-wind trading causally reduces price pressures in inventory models, but in the data this effect might be dampened by reverse causality. The reason is that large price pressures make against-wind trading more profitable, which may generate a *positive* correlation between the two. For with-wind trading it works the other way. If the price impact is large then it is more likely to be detected by HFTs, and this reinforces the causal effect from with-wind trading to implementation shortfall.

The right-hand side plots of Figure 8 show that HFTs gross trading revenue is positive overall, and for both the against-wind and the with-wind terciles individually. HFTs seem to generate most revenue when trading with-wind: \$700 per order, which is about twice the overall average.

Table VII complements Figure 8 by presenting *conditional* correlations. It again relates implementation shortfall and HFT gross trading revenue with the type of trading by HFTs, but this time adds control variables. The latter are either variables that further characterize execution of the meta order (e.g., duration of the order) or market characteristics in the lifetime of the order (e.g., volatility). All variables are standardized and therefore results become conditional correlations. The variables of interest — implementation shortfall, HFT gross trading revenue, and HFT against/with-wind trading — are expressed either in dollars or in relative terms for robustness.

in Anand et al. (2013, Table 1), who document an order size-weighted shortfall of 25 basis points. The equivalent number for our sample is 21 basis points.

Table VII Implementation shortfall and HFT revenue regressed on against/with-wind HFT flow and controls

This table presents the panel regression results where implementation shortfall (IS) and HFT gross trading revenue are the dependent variables. The main explanatory variable is HFT net flow accumulated over the lifetime of the institutional order. It is standardized, and expressed either relative to shares outstanding (%) or in dollar terms (\$). Values in the same (opposite) direction as the institutional order are assigned to with-wind (against-wind) net flow and get a positive sign. The variables are zero otherwise. Order- and market-specific variables are added as controls. Variables names are mostly self-explanatory and should become clear with the following notes. Order size and ADV (based on the full sample) are both measured in shares. The dummies “Start at U.S. open” and “End at U.S. open” are equal to one if the order starts or ends, respectively, in the thirty-minute interval after the opening of the U.S. market. Turnover and volatility are measured from the start to the end of the order. Turnover is the stock volume divided by shares outstanding and volatility is realized volatility based on one-minute midquote returns. The market concentration measure HHI is the Herfindahl-Hirschman Index, defined as the sum of daily squared market shares across lit trading venues. All explanatory variables are standardized. Regressions include stock and institution fixed effects and standard errors are clustered by stock-date. *t*-values are in parentheses. Variable units are in brackets. Superscripts ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Imp shortfall (bps)	Imp shortfall (\$1,000)	HFT profit ratio (bps)	HFT gr tr rev (\$1,000)
<i>Order-specific variables</i>				
Against-wind HFT cum net flow (%)	-2.72 (-1.4)		0.38 (0.6)	
With-wind HFT cum net flow (%)	13.02*** (6.3)		1.55*** (2.7)	
Against-wind HFT cum net flow (\$)		-1.28 (-1.3)		0.33 (1.0)
With-wind HFT cum net flow (\$)		4.46*** (3.5)		0.86*** (2.6)
Order size relative to ADV (%)	4.63*** (3.4)	8.50*** (6.9)	-0.19 (-0.5)	-0.00 (-0.0)
Order duration (hours)	-0.75 (-0.3)	-0.42 (-0.2)	0.16 (0.3)	-0.21 (-0.8)
Trades through market orders (%)	2.17*** (3.3)	0.87*** (3.3)	-0.17 (-0.9)	-0.02 (-0.3)
Meta order is buy dummy	3.54* (1.9)	1.90 (1.6)	-0.28 (-0.5)	-0.16 (-0.9)
Start at U.S. open dummy	-3.25** (-2.3)	-0.31 (-0.5)	0.42 (0.6)	0.05 (0.3)
End at U.S. open dummy	-5.93*** (-2.6)	-2.26* (-1.8)	0.57 (1.0)	0.09 (0.5)
Closing auction dummy	-1.82 (-0.7)	-2.97** (-2.2)	-0.03 (-0.0)	-0.45 (-1.4)
<i>Market-specific variables</i>				
Stock volatility (%)	2.63** (2.5)	1.19** (2.0)	0.08 (0.2)	-0.06 (-0.4)
Turnover (%)	4.25 (1.6)	2.63* (1.7)	-0.90 (-1.4)	0.73** (2.3)
Market concentration HHI (%)	-7.94 (-0.4)	2.84 (0.3)	13.10** (2.4)	2.68 (1.2)
Observations	5,136	5,136	5,136	5,136
R-squared	0.062	0.058	0.007	0.005

The results lead to the following additional insights. First, also conditionally with-wind trading is positively correlated with implementation shortfall. The effect is highly statistically significant (i.e., beyond a 1% level). The same holds for the conditional correlation between with-wind trading and HFT gross trading profit. Taken together, these results provide more robust evidence on a positive association between with-wind trading the two variable of interest: transaction cost for investors and gross trading profit for HFTs. This is perhaps the strongest support for the fourth prediction.

Second, the against-wind trading does not significantly correlate with against-wind trading in either the implementation-shortfall or the HFT-revenue regressions. The signs are consistent with Figure 8, but there simply is no statistical significance.

Third, the control variables that are statistically significant in the implementation-shortfall regressions mostly confirm our priors. Transaction costs correlate positively with the relative size of the order, with the extent to which trades were done through market orders (as opposed to limit orders), and with market volatility. These signs are all consistent with the microstructure literature. The surprising results are that implementation shortfall is larger for buy orders and for high-turnover periods. These results are statistically the weakest ones, though, with a confidence level of only 90%.

Fourth, the two variables that turn statistically significant in the HFT-revenue regressions are also intuitive. HFTs earn a higher gross profit when turnover is higher (so that they can layoff inventory easier) and their gross profit margin is higher when the market is more concentrated on the incumbent venue (to probably make up for the higher fees).²⁵

²⁵NASDAQ trading fees were 0.50 basis points for a market order in this period, compared to its main competitor

VI. Robustness

This section presents the robustness analysis. The results are added to the online appendix.

HFT trading around institutional orders in volume clock time (instead of calendar clock time). An alternative way to run the clock to study how HFTs trade around institutional orders is to use volume instead of calendar time. Volume time increments the clock by, for example, an hour when market volume has incremented by $1/8.5$ of a stock's average daily volume (a trading day in Stockholm lasts 8.5 hours). We then track HFT cumulative net flow on this new time grid. To avoid issues of concatenating trades across days, we do this analysis only for the first day of an order's execution period.²⁶ Figure A5 in the online appendix shows that the HFT trading pattern in volume time echoes the pattern in calendar time of the signature plot of Figure 2: against-wind initially and with-wind eventually for long-lasting orders.

The with-wind pattern is somewhat weaker than in the calendar time analysis and this might be due to the following. There is with-wind trading on some orders that are executed in periods of low market volume. These orders by definition are short-lasting in *volume* time. They also have a high participation rate, which makes hiding more difficult, and explains why one sees lots of with-wind flow. However, such orders leave the sample quickly in a volume time analysis. During periods of low volume, the investor may rationally choose to take on more detection risk if his clock ticks in entrant market Chi-X that charged 0.28 basis points.

²⁶That is, some order might stretch across multiple days according to a calendar time concatenation, but not according to volume time concatenation. To compare apples to apples, we decided to do this analysis only based on the first day of each meta order.

natural time. For example, his information advantage dissipates in calendar time when the trading day finishes (and overnight news arrives that makes his private signal redundant), or when a public news announcement comes closer.

Is with-wind trading an end-of-day phenomenon (i.e., HFTs clearing out their inventory)?

To study whether with-wind trading is an end-of-day-phenomenon, we do the following additional analysis. First, we plot the dispersion of order start- and end-times of the with-wind subsample in Figure A6 in the online appendix. It shows that these times are dispersed throughout the day in a way similar to the full sample (Figure A1). Further analysis reveals that the HFT trading pattern is insensitive to whether the order ends in the final trading hour or not. Figure A7 in the online appendix redoes the paper's signature plot after splitting the sample based on whether an order stopped executing before the final trading hour. The general pattern of against-wind first and eventually with-wind for long-lasting orders emerges for both subsamples. These results suggest that our main result is not an end-of-day-phenomenon.

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 subsample 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 orders on 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 observations for statistical inference is 70%, which reduces the sample from 5,136 orders to 863 orders. We can then confirm statistical significance of our main result in Figure A8: HFTs initially trade against-wind but eventually trade with-wind for long-lasting orders.²⁷

Unfortunately we cannot go beyond the 70% cutoff level and incomplete HFT data therefore remains a concern, but we are not overly worried for the following reasons. First, if the NASDAQ HFT position were perfectly offset by opposite positions in non-NASDAQ markets, then the HFT NASDAQ position would be economically meaningless, or just “noise”, and this would work against us when trying to establish empirical patterns. The fact that we find statistically significant patterns therefore suggests that such position is not pure noise. Second, Table AI in the online appendix shows that HFT directionality (on NASDAQ) approaches zero when it is sampled at lower frequencies. HFTs seem to mean-revert at least some of their NASDAQ position across days which they would not need to do if they had perfectly offsetting positions in non-NASDAQ markets. Taken together, these arguments suggest that our HFT cumulative net flow variable is potentially noisy but carries signal as well.

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-

²⁷We also redid the signature plot by stratifying the sample into tercile based on market concentration. Figure A9 in the online appendix again shows that the pattern is robust and, not surprisingly, stronger for the high concentration subsample (as HFT cumulative net flow is measured more precisely there since it is only based on NASDAQ data, i.e., on data of the incumbent market).

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 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 start tracking HFT cumulative net flow one hour before an investor's first child trade. Figure A10 shows that this flow is small and statistically insignificant. We therefore consider this alternative interpretation highly unlikely (since HFTs must have picked up a signal, if any, at least an hour before an investor's first trade).

VII. Conclusion

Our most robust finding is that HFTs trade along with institutional orders, with the wind, if such orders last at least a few hours. Additional analysis suggests that in these cases investors' trading is information-motivated and HFTs try to detect such trading in order to join it. Such HFT with-wind trading is costly to investors and profitable to HFTs.

HFTs appear to need considerable time to detect these informed orders (several hours) which appears at odds with their typical characterization as ultra-high speed traders. We believe HFTs should be more broadly characterized as high-tech traders. Technology enables them to trade at sub-microsecond speed, but also endows them with an ability to parse "big data" to detect informed trading across several hours (such period becomes big data if one includes all message traffic for the security at hand and all related securities).

How do our findings relate to market quality? One might say that HFTs trading with-wind on institutional orders speeds up price discovery. In other words, HFTs trading in the same direction as informed investors could make prices reveal the private information more quickly. It might make prices more efficient in the short run. 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 on end-user costs. Data are hard to come by, but it should be in the interest of end users — retail and institutional investors — to make their trade data available. Alternatively, regulators could demand the reporting of more granular data, 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, for agency trades, whether it was for a retail investor or for an institutional investor. This enabled 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.²⁸

²⁸See <https://www.sec.gov/divisions/marketreg/rule613-info.htm>.

REFERENCES

- Almgren, Robert, and Neil Chriss, 1999, Value Under Liquidation, *Risk* 12, 61–63.
- Anand, Amber, Paul Irvine, Andy Puckett, and Kumar Venkataraman, 2013, Institutional Trading and Stock Resiliency: Evidence From the 2007-2009 Financial Crisis, *Journal of Financial Economics* 108, 773–797.
- Barbon, Andrea, Marco Di Maggio, Francesco Franzoni, and Augustin Landier, 2017, Brokers and Order Flow Leakage: Evidence from Fire Sales, Unpublished Manuscript Harvard Business School.
- Biais, Bruno, and Thierry Foucault, 2014, HFT and Market Quality, *Bankers, Markets & Investors* 5–19.
- Biais, Bruno, Pierre Hillion, and Chester Spatt, 1999, Price Discovery and Learning during the Preopening Period in the Paris Bourse, *Journal of Political Economy* 107, pp. 1218–1248.
- Boehmer, Ekkehart, Kingsley Fong, and Juan Wu, 2014, International Evidence on Algorithmic Trading, Unpublished Manuscript EDHEC Business School.
- Boulatov, Alex, Dan Bernhardt, and Ivan Larionov, 2016, Predatory and Defensive Trading in a Dynamic Model of Optimal Execution by Multiple Traders, Unpublished Manuscript Higher School of Economics.
- Brogaard, Jonathan, Bjorn Hagströmer, Lars Nordén, and Ryan Riordan, 2015, Trading Fast and Slow: Colocation and Market Quality, *Forthcoming Review of Financial Studies* .
- Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2014, High Frequency Trading and Price Discovery, *Review of Financial Studies* 27, 2267–2306.
- Brunnermeier, Markus K., and Lasse Heje Pedersen, 2005, Predatory Trading, *Journal of Finance* 60, 1825–1863.
- Chaboud, Alain P., Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2014, Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market, *Journal of Finance* 69, 1540–6261.
- Choi, Jin Hyuk, Kasper Larsen, and Duane Seppi, 2017, Information and Trading Targets in a Dynamic Market Equilibrium, Unpublished Manuscript Carnegie Mellon University.
- Chordia, Tarun, and Avanidhar Subrahmanyam, 2004, Order Imbalance and Individual Stock Returns: Theory and Evidence, *Journal of Financial Economics* 72, 485–518.
- Davies, Ryan J., and Sang Soo Kim, 2009, Using Matched Samples to Test for Differences in Trade Execution Costs, *Journal of Financial Markets* 12, 173–202.

- Easley, David, Marcos Lopez de Prado, and Maureen O'Hara, 2016, Discerning information from trade data, *Journal of Financial Economics* 120, 269 – 285.
- Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1996, Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow, *Journal of Finance* 51, 811–833.
- ESMA, 2014, High-Frequency Trading Activity in EU Equity Markets, European Securities and Markets Authority Economic Report 2014-1.
- Frenkel, Jacob A., 1977, The Forward Exchange Rate, Expectations, and the Demand for Money: The German Hyperinflation, *American Economic Review* 67, 653–670.
- Gao, Cheng, and Bruce Mizrahi, 2013, Market Quality Breakdowns in Equities, Unpublished Manuscript Rutgers University.
- Glosten, Lawrence R., 1987, Components of the Bid Ask Spread and the Statistical Properties of Transaction Prices, *Journal of Finance* 42, 1293–1307.
- Grossman, Sanford J., and Merton H. Miller, 1988, Liquidity and Market Structure, *The Journal of Finance* 43, 617–633.
- Harris, Lawrence E., 1997, Order Exposure and Parasitic Traders, Unpublished Manuscript University of Southern California.
- Hasbrouck, Joel, 2015, High Frequency Quoting: Short-Term Volatility in Bids and Offers, Unpublished Manuscript New York University.
- Hasbrouck, Joel, and Gideon Saar, 2013, Low-latency Trading, *Journal of Financial Markets* 16, 646–679.
- Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does Algorithmic Trading Improve Liquidity?, *Journal of Finance* 66, 1–33.
- Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2013, Implementation Shortfall with Transitory Price Effects, in David Easley, Marco Lopez de Prado, and Maureen O'Hara, eds., *High Frequency Trading: New Realities for Traders, Markets and Regulators*, chapter 9 (Risk Books, London).
- Hirschey, Nicholas H., 2014, Do High-Frequency Traders Anticipate Buying and Selling Pressure?, Unpublished Manuscript London School of Economics.
- Hodrick, Robert J., 1987, *The Empirical Evidence on the Efficiency of Forward and Futures Foreign Exchange Markets* (Harwood Academic Publishers, Chur, Switzerland).
- Kirilenko, Andrei A., Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2017, The Flash Crash: High-Frequency Trading in an Electronic Market, *The Journal of Finance* 72, 967–998.

- Korajczyk, Robert, and Dermot Murphy, 2016, High Frequency Market Making to Large Institutional Trades, Unpublished Manuscript Northwestern University.
- Kyle, Albert, 1985, Continuous Auctions and Insider Trading, *Econometrica* 53, 1315–1335.
- Linnainmaa, Juhani T., and Gideon Saar, 2012, Lack of Anonymity and the Inference from Order Flow, *Review of Financial Studies* 25, 1414–1456.
- Menkveld, Albert J., 2013, High Frequency Trading and the New Market Makers, *Journal of Financial Markets* 16, 712–740.
- Menkveld, Albert J., 2016, The Economics of High-Frequency Trading: Taking Stock, *Annual Review of Financial Economics* 8, 1–24.
- Milgrom, Paul, and Nancy Stokey, 1982, Information, Trade and Common Knowledge, *Journal of Economic Theory* 26, 17–27.
- O’Hara, Maureen, 2015a, High Frequency Market Microstructure, *Journal of Financial Economics* 116, 257–270.
- O’Hara, Maureen, 2015b, High Frequency Market Microstructure, *Journal of Financial Economics* 116, 257–270.
- Rock, Kevin, 1990, The Specialist’s Order Book and Price Anomalies, Unpublished Manuscript Harvard University.
- SEC, 2010, Concept Release on Equity Market Structure, Release No. 34-61358; File No. S7-02-10.
- Tong, Lin, 2015, A Blessing or a Curse? The Impact of High Frequency Trading on Institutional Investors, Unpublished Manuscript University of Iowa.
- van Kervel, Vincent, 2015, Competition for Order Flow with Fast and Slow Traders, *Review of Financial Studies* 28, 2094–2127.
- Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2009, Information Immobility and the Home Bias Puzzle, *The Journal of Finance* 64, 1187–1215.
- Yang, Liyang, and Haoxiang Zhu, 2017, Back-Running: Seeking and Hiding Fundamental Information in Order Flows, Unpublished Manuscript Massachusetts Institute of Technology.
- Ye, Mao, CHen Yao, and Jiading Gai, 2013, The Externalities of High Frequency Trading, Unpublished Manuscript University of Illinois at Urbana-Champaign.

Internet Appendix to “High-Frequency Trading around Institutional Orders”¹

This online appendix contains the following list of tables and figures references in the main document.

List of Tables

AI	Mean reversion in cumulative net flow: Flat at the end of the day?	3
AII	Do all HFTs follow the same strategy?	4
AIII	HFT cumulative net flow in the lifetime of an institutional order by information content	5

List of Figures

A1	Histogram start- and end-time of institutional orders	6
A2	Cumulative net flow of institutional investors in the lifetime of their orders	7
A3	HFT cumulative net flow in the lifetime of an institutional order, buy and sell orders pooled	8
A4	Permanent price impact (PPI) of institutional orders by order size	9
A5	HFT net flow in the lifetime of an institutional order: volume time instead of calendar time	10
A6	Histogram start- and end-time of with-wind institutional orders	11
A7	HFT cumulative net flow in the lifetime of an institutional order, for orders that end before or in the final trading hour	12
A8	HFT net flow during the lifetime of the institutional order, only for orders on stock-days for which NASDAQ market share exceeds 70%	13

¹van Kervel, Vincent, and Albert J. Menkveld, Internet Appendix to “High-Frequency Trading around Institutional Orders,” Journal of Finance.

A9	HFT cumulative net flow in the lifetime of institutional orders, by tercile of market concentration	14
A10	HFT net flow in the lifetime of an institutional order, this time starting one hour before the first child trade executes.	15

Table AI Mean reversion in cumulative net flow: Flat at the end of the day?

This table reports the ratio of absolute cumulative net flow and volume (i.e., directionality), for both HFTs and for institutional investors. Let B be buy volume (in shares or dollars) and S be sell volume, then this ratio is defined as $|B-S|/(B+S)$. For HFTs the ratio is calculated for stock-days when at least one of the investors in our sample was executing an order, and for *all* stock-days to provide a benchmark. The ratio is further calculated for stock-weeks and stock-months to illustrate to what extent there is mean-reversion across days. For investors, it is calculated for stocks-days, stock-weeks, and stock-months, but only for those where they were executing at least one order.

	Directionality: $ B-S /(B+S)$	
	\$100,000	#Trades
HFT: Stock-days with institutional investors executing orders	0.078	0.078
HFTs with-wind tercile	0.080	0.080
HFTs neutral tercile	0.076	0.076
HFTs against-wind tercile	0.078	0.078
HFT: All stock-days	0.084	0.081
HFT: All stock-weeks	0.048	0.049
HFT: All stock-months	0.028	0.029
Institutional investors: Stock-days	0.990	0.984
Institutional investors: Stock-weeks	0.914	0.914
Institutional investors: Stock-months	0.833	0.830

Table AII 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 (\$-397,400) is not significantly different from that of with-wind orders (\$359,300), 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)		Difference: "WW=(-AW)" ^a	
	Against-wind (AW)	With-wind (WW)		
HFT 1	-67.2	83.3	16.0	***
HFT 2	-1.7	1.5	-0.2	
HFT 3	-44.0	74.5	30.5	***
HFT 4	-1.8	2.8	1.0	
HFT 5	-25.7	39.9	14.2	
HFT 6	-21.3	15.3	-5.9	***
HFT 7	-159.9	73.0	-86.9	
HFT 8	-16.7	20.6	4.0	***
HFT 9	-59.2	48.4	-10.8	
HFT 10	-4.3	9.6	5.3	***
Total	-397.4	359.3	-38.1	

^a The *p*-value of the Wald test that "WW=-AW" for all HFTs jointly is 0.00.

Table AIII HFT cumulative net flow in the lifetime of an institutional order by information content

This table completes the one in Figure 7. The latter one only showed the odd hours.

		Hours since first child order execution																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
<i>Low PPI Tercile</i>																		
Mean	-14.5***	-22.8***	-27.7**	-27.1*	-35.8**	-49.3**	-8.1	44.9	42.8	62.2	36.1	35.3	48.9	59.2	98.5	22.5	58.1	
t-stat	(2.7)	(2.8)	(2.5)	(1.9)	(2.0)	(2.4)	(0.3)	(1.1)	(0.8)	(1.1)	(0.6)	(0.6)	(0.7)	(0.9)	(1.3)	(0.2)	(0.4)	
N	1091	942	856	758	643	525	377	240	176	159	153	147	133	126	109	79	57	
<i>Middle PPI Tercile</i>																		
Mean	-13.9***	-11.3	-1.8	17.1	17.9	31.1	42.9	72.4	112.0	86.5	90.9	88.5	59.8	80.2	178.4	192.8	37.2	
t-stat	(2.5)	(1.3)	(0.2)	(1.2)	(0.9)	(1.2)	(1.5)	(1.4)	(1.4)	(1.0)	(1.1)	(1.0)	(0.6)	(0.7)	(1.3)	(1.5)	(0.3)	
N	968	787	688	597	494	365	271	171	110	96	95	91	85	77	65	41	24	
<i>High PPI Tercile</i>																		
Mean	-8.7	-3.8	-18.5*	-11.8	-2.9	1.7	31.7	63.3	96.6	136.9**	151.5***	169.1***	183.2***	178.4**	224.4***	296.5***	236.9*	
t-stat	(1.6)	(0.5)	(1.8)	(0.9)	(0.2)	(0.1)	(1.2)	(1.5)	(1.6)	(2.3)	(2.5)	(2.7)	(2.7)	(2.3)	(2.9)	(2.8)	(1.7)	
N	1152	994	879	790	702	576	413	269	191	176	162	157	142	131	109	71	46	

Figure A1 Histogram start- and end-time of institutional orders

This histogram depicts the distribution of the start- and end-times of institutional orders by binning them into 15-minute intervals.

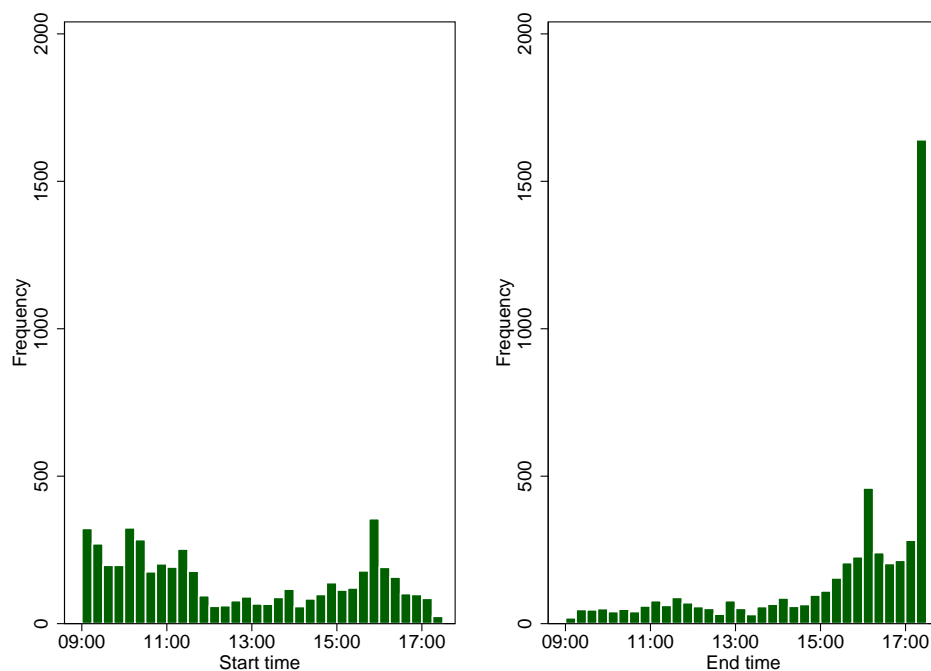


Figure A2 Cumulative net flow of institutional investors in the lifetime of their orders

This figure mirrors Figure 2 in the main text, but this time plots the institution's cumulative net flow instead of HFT cumulative net flow. In other words, it plots cumulative net flow of the institution in the lifetime of its order. The time points are multiples of 30 minutes. At each point the average is taken across all institutional orders that are still executing at that point (i.e., the final child trade has not arrived yet). The size of the sample at each time point is indicated by the dashed green line (right y-axis). Statistical significance is tested based on the t -value of the mean across stock-institution fixed effects (same as the overall mean), with residuals clustered at the stock-period level.

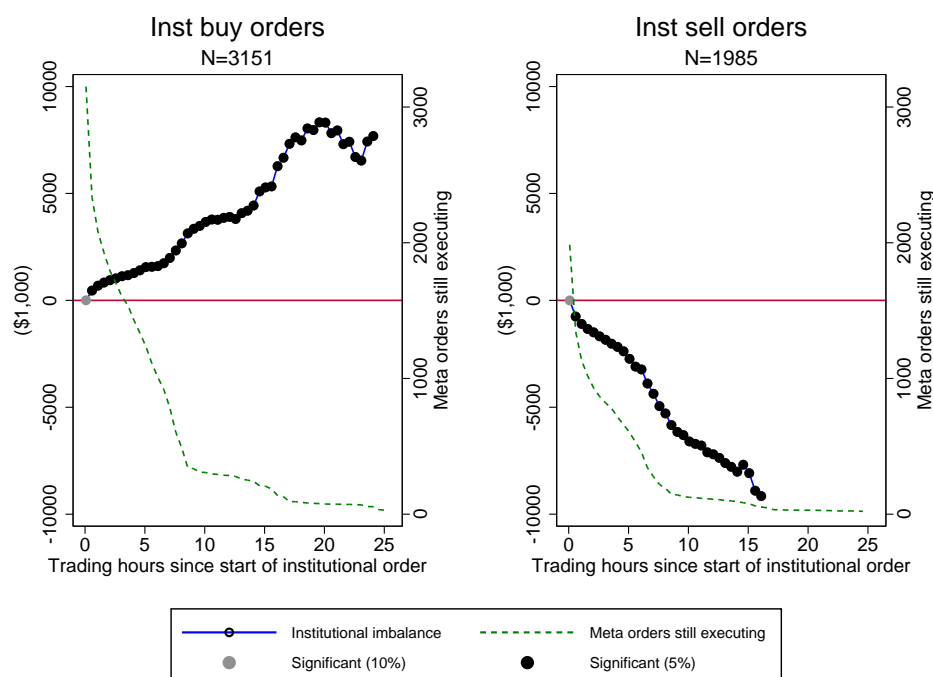


Figure A3 HFT cumulative net flow in the lifetime of an institutional order, buy and sell orders pooled

This figure redoes Figure 2 (top panel) and 3 (bottom panel), but pools buy and sell orders by converting sell orders into pseudo buy orders (by appropriate mirroring).

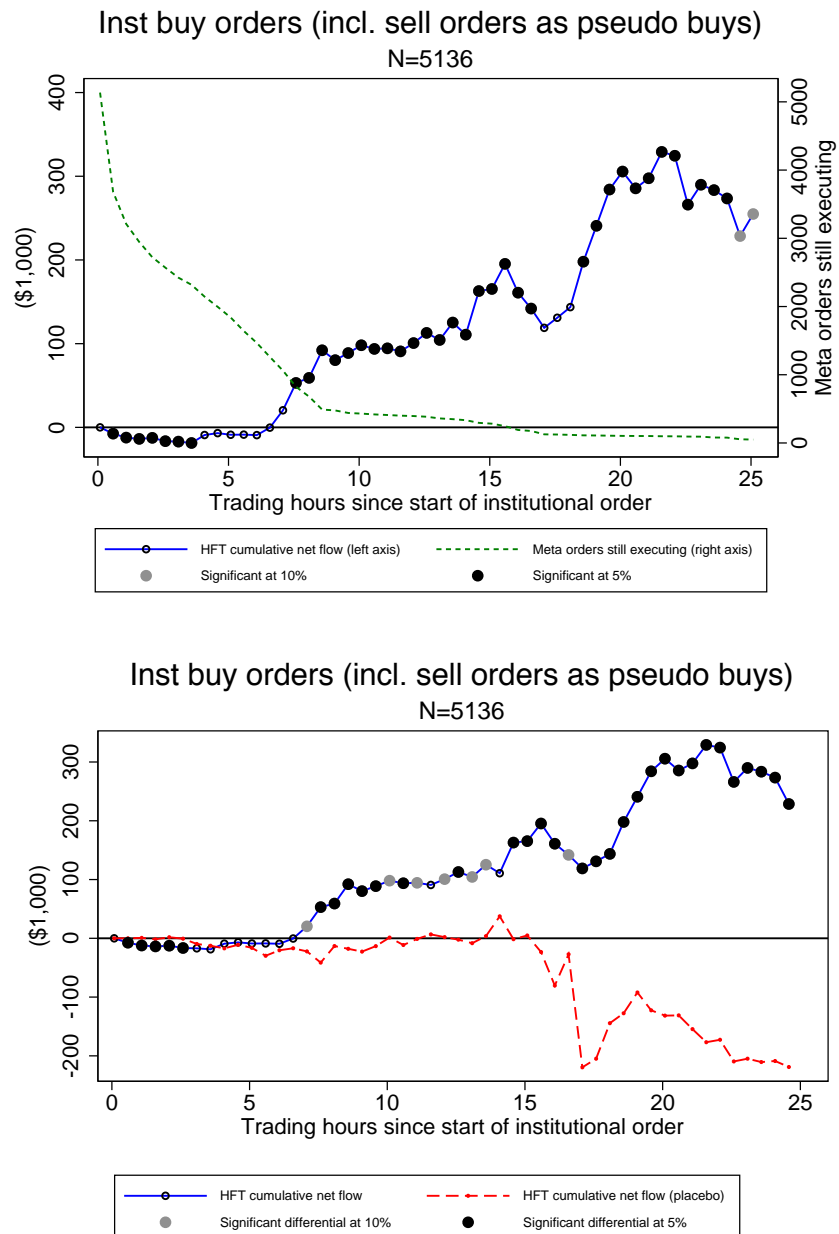


Figure A4 Permanent price impact (PPI) of institutional orders by order size

This figure bins orders by dollar order size or order size expressed relative to average daily volume (ADV). 95% confidence bounds have been added. Units are in brackets.

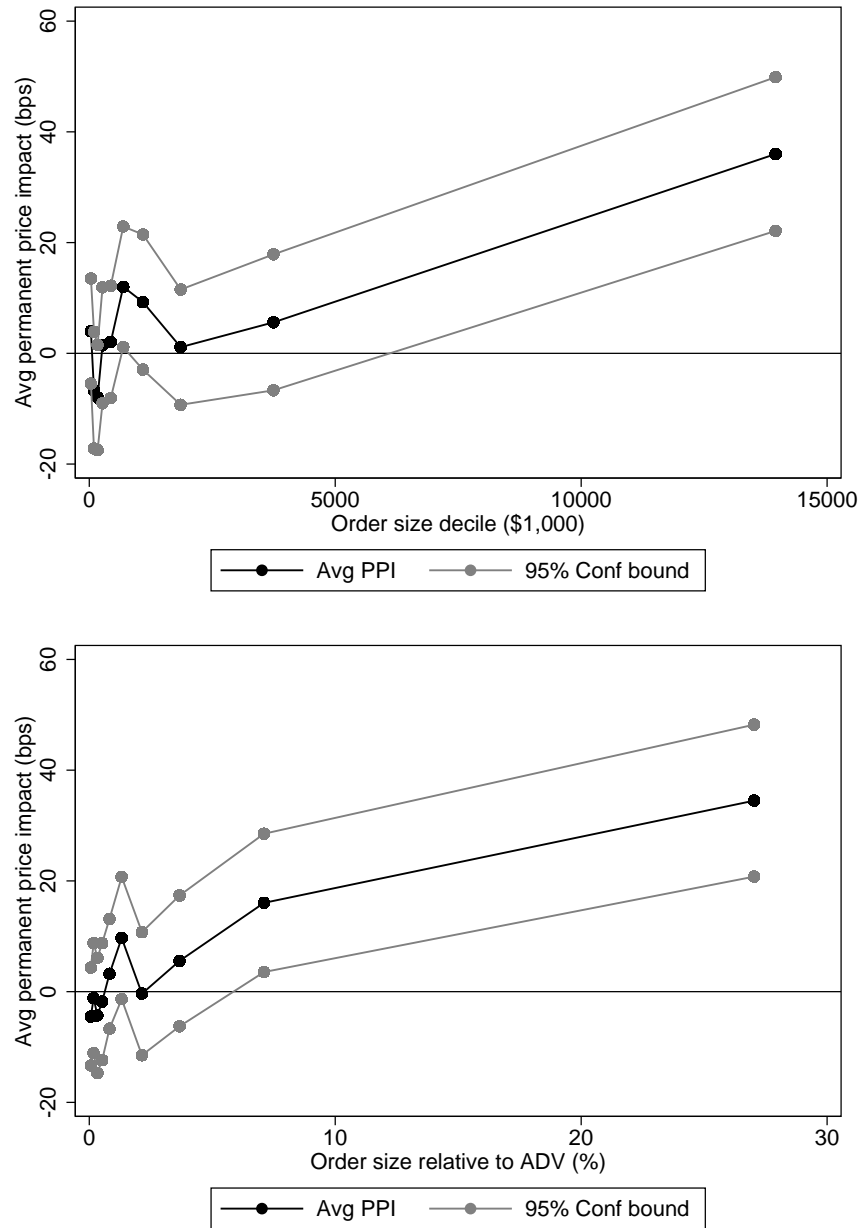


Figure A5 HFT net flow in the lifetime of an institutional order: volume time instead of calendar time

This figure redoes Figure 2 with volume clock time instead of calendar time on the horizontal axis. To create a fair comparison and avoid issues of comparability when concatenating trades across days, we decide to focus only on single-day meta orders combined with the first day only of a multiple-day order. The top panel plots the volume-time pattern, and the bottom panel, for comparison, the calendar-time pattern (similar to Figure 2 but for the subsample only). For the volume-time analysis the time clock is incremented by one hour whenever 1/8.5 of average daily volume is traded on NASDAQ OMX.

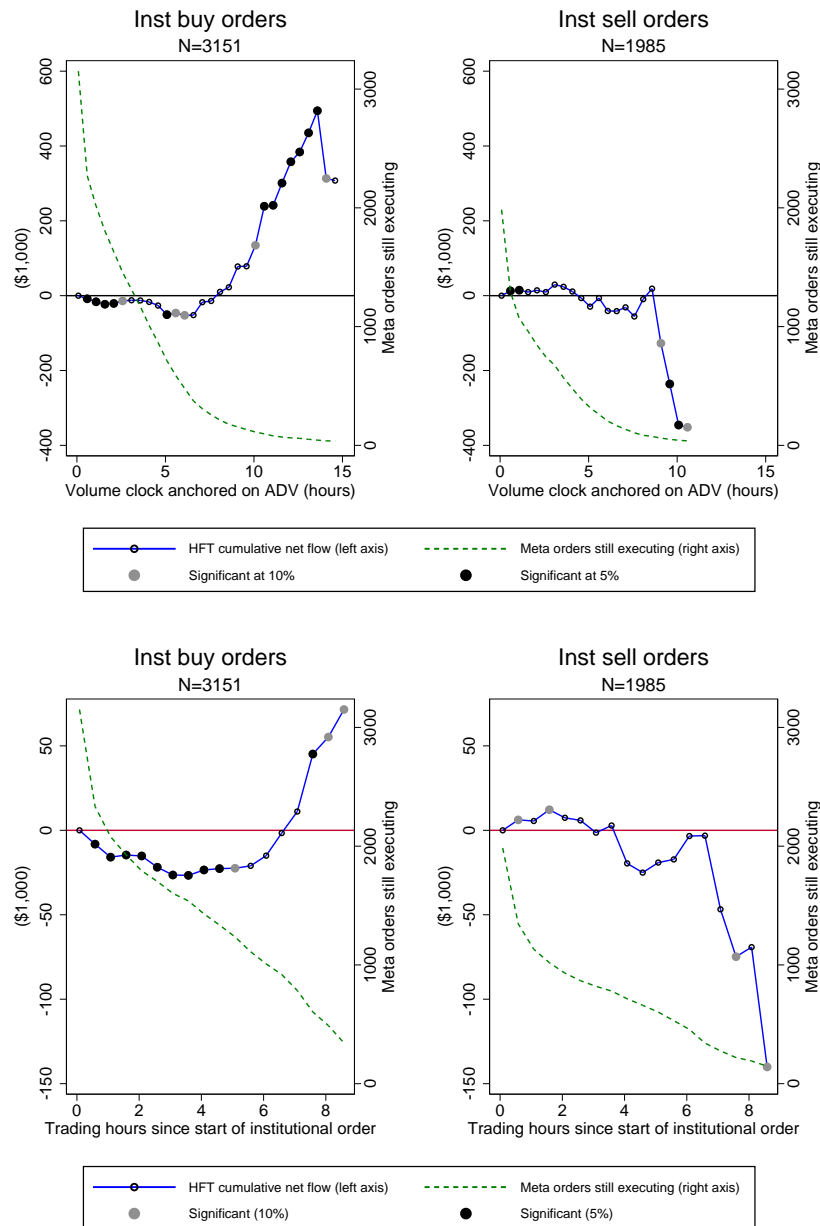


Figure A6 Histogram start- and end-time of with-wind institutional orders

This histograms depict the distribution of the start- and end-times of with-wind institutional orders by binning them in 15-minute intervals. It redoes Figure A1 but now for with-wind orders only, further split into single-day or multiple-day orders.

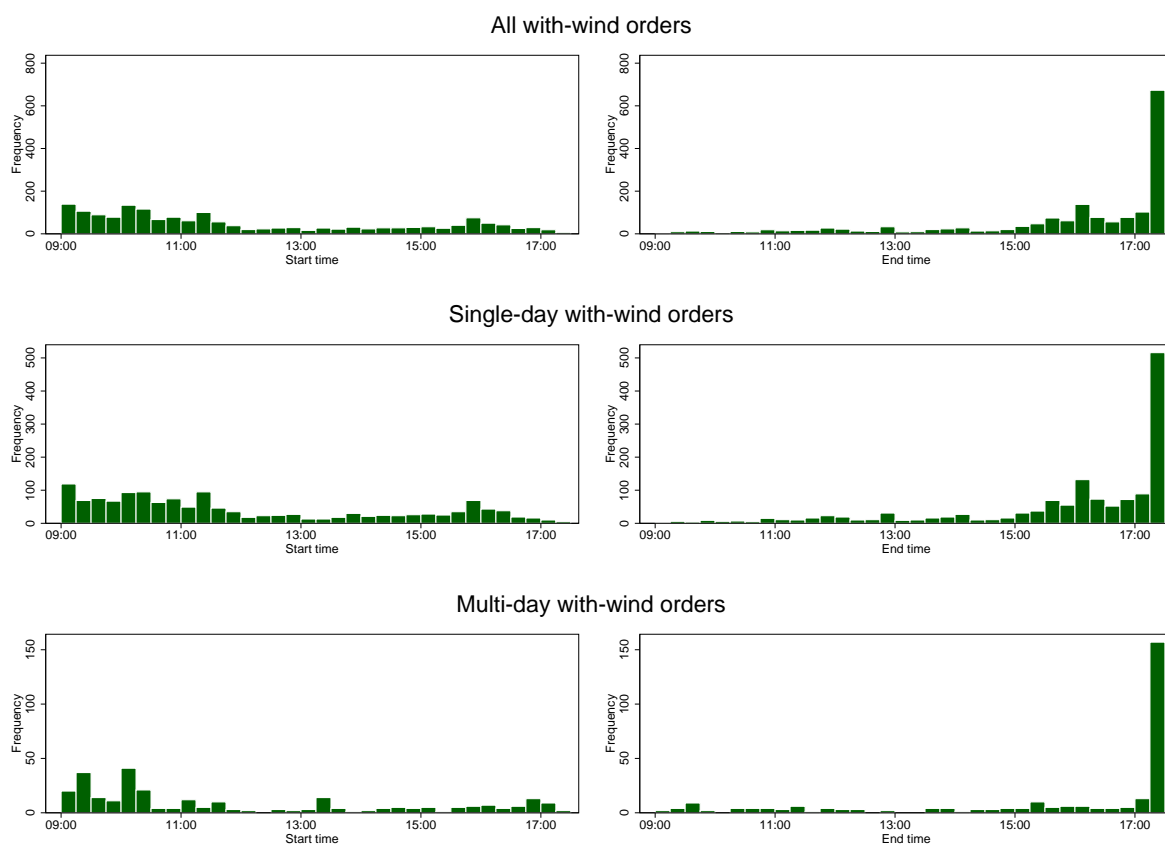


Figure A7 HFT cumulative net flow in the lifetime of an institutional order, for orders that end before or in the final trading hour

This figure redoes Figure 2 for pseudo buys (i.e., buy orders pooled with sell orders by appropriately mirroring the later) split into two subsamples: orders that ended before the final trading hour and orders that ended in the final trading hour.

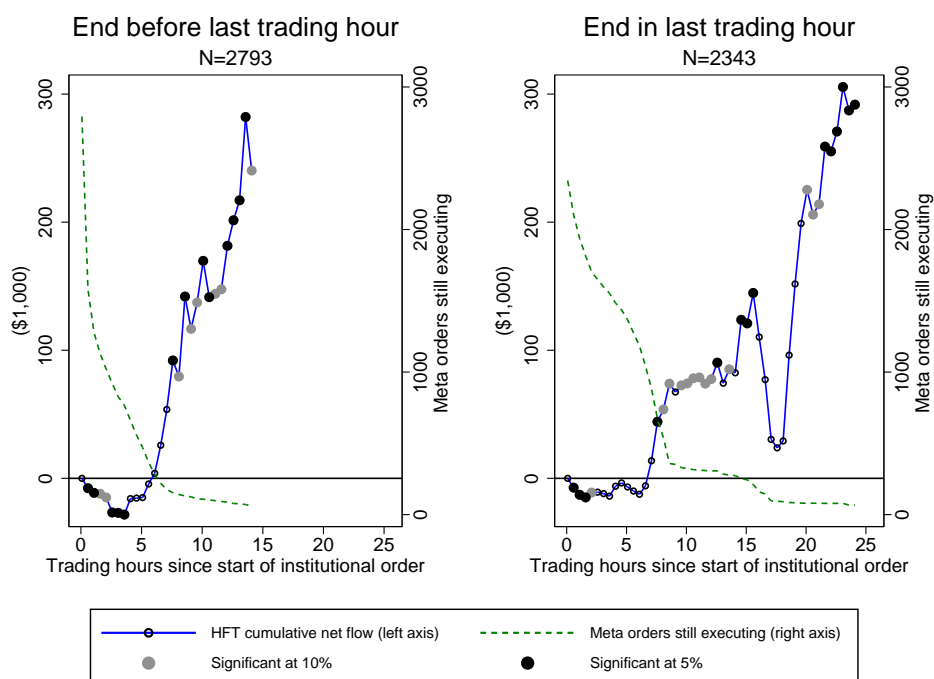


Figure A8 HFT net flow during the lifetime of the institutional order, only for orders on stock-days for which NASDAQ market share exceeds 70%

This figure mirrors Figure 2 in the main text, but is this time based on the subsample of institutional orders where the NASDAQ market share exceeds 70%. The smaller number of observations (863 instead of 5,136) forces us to pool institutional buy and sell orders. We do so by multiplying HFT cumulative net flow by minus one for sell orders. We refer to these orders as pseudo-buy orders.

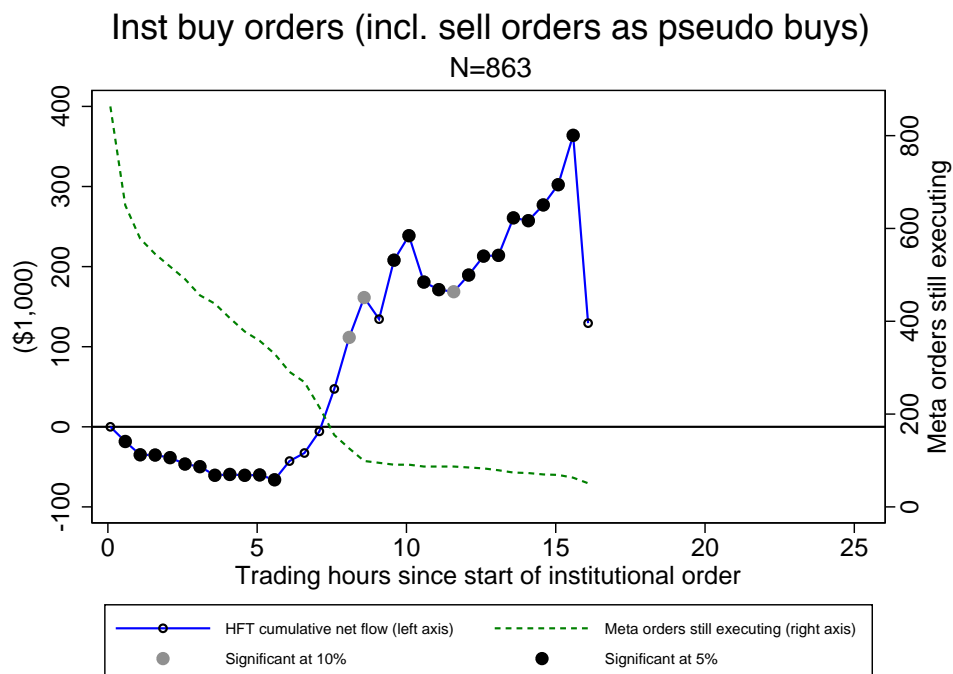


Figure A9 HFT cumulative net flow in the lifetime of institutional orders, by tercile of market concentration

This figure redoes Figure 2 on two subsamples after splitting the sample into terciles based on the market concentration (i.e., the extent to which volume concentrates on one of the lit venues as measured by Herfindahl-Hirschman index). We show results for the high and low tercile. Buys and sells are pooled into pseudo-buy orders for additional power.

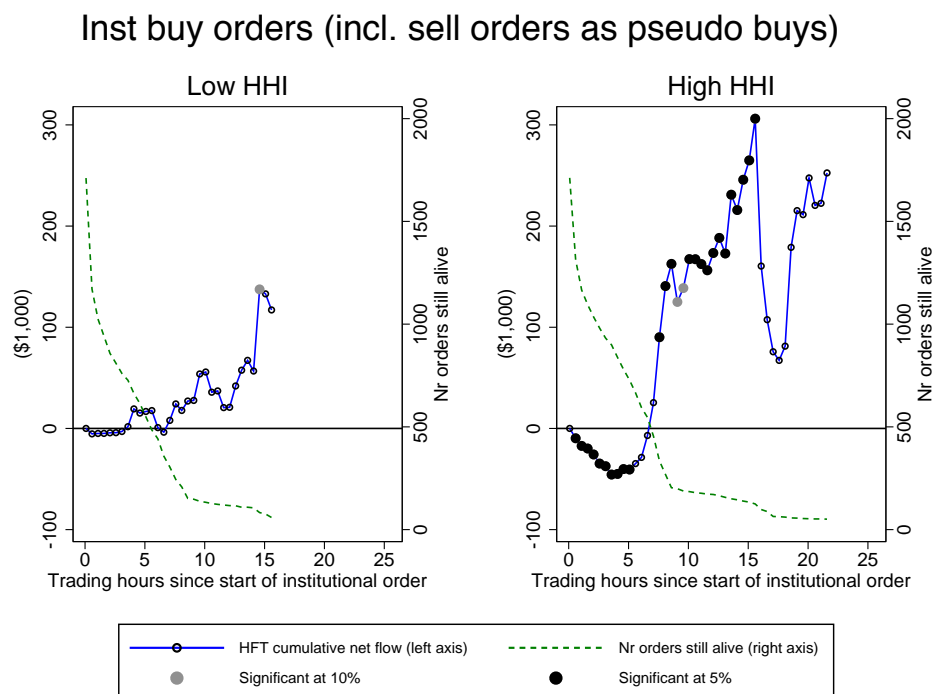


Figure A10 HFT net flow in the lifetime of an institutional order, this time starting one hour before the first child trade executes.

This figure mirrors Figure 2 in the main text, but starts tracking HFT cumulative net flow one hour before the first child trade executes.

