Order Flow Toxicity under the Microscope*

David Abad**

Magdalena Massot

Samarpan Nawn

Roberto Pascual

José Yagüe

This version:
January 2020

* We thank Bidisha Chakrabarty, Ivan Indriawan, Michael Schneider, and the audience at the New Zealand Finance Meeting (Queenstown, 2018), and XXVII Finance Forum (Madrid, 2019) for useful suggestions and comments. This paper is supported by the Spanish DGICYT project ECO2017-86903-P, and the Generalitat Valenciana Grant Prometeo/2017/158. Samarpan Nawn acknowledges the data support received from Financial Research and Trading Laboratory IIM Calcutta (received as a doctoral student at IIM Calcutta). Any errors are entirely our own.

** Magdalena Massot (m.massot@uib.es) is Associate Professor and Roberto Pascual (rpascual@uib.es) is Full Professor in the Department of Business Economics of the University of the Balearic Islands, Spain. Samarpan Nawn (samarpan.nawn@iimu.ac.in) is Assistant Professor in the Department of Finance and Accounting of the Indian Institute of Management Udaipur in Rajasthan, India. David Abad (goliat@ua.es) is Associate Professor in the Department of Business and Accounting of the University of Alicante, Spain. José Yagüe (ppyague@um.es) is Associate Professor in the Department of Business Organization and Finance of the University of Murcia, Spain. Contact author: Roberto Pascual (Phone: +34-971171329; Fax: +34-971172389)

Order Flow Toxicity under the Microscope

Abstract

Which components of the order flow convey information and signal toxicity? We empirically show that the net flow of non-marketable orders conveys more information than the widely used trade-initiator-based order imbalance. The net order flow by HFTs rapidly loses information content with time aggregation, consistent with these traders trading on short-lived valuable signals. Updates of standing limit orders, mostly due to HFT, carry information beyond order submissions, suggesting that HFTs' flickering quotes primarily reflect active risk management rather than manipulative practices. Finally, we find that the HFTs' order flow, both marketable and non-marketable, signals toxicity, while the non-HFTs' order flow does not. We conclude that market authorities should track the HFTs' order flow at or near the best quotes to develop effective leading indicators of order flow toxicity and circuit breaking mechanisms.

Keywords: Order flow, toxicity, order imbalance, high-frequency trading, limit orders, market orders, monitoring.

JEL Classification: G10, G11, G14, G15,

1. Introduction

The order flow is toxic when it adversely selects liquidity providers, who may be unaware they are providing liquidity at a loss (Easley, López de Prado, and O'Hara, 2012). Toxic order flow arises from the presence of informed traders who have advanced signals about fundamentals. In modern high-frequency markets, however, traders can also be informed because they are faster than others in processing public signals (O'Hara, 2015). In this paper, we empirically examine what pieces of the order flow should market authorities, policy makers, professionals, and academics track to infer about underlying information and build advanced indicators of order flow toxicity.

Early models of adverse selection (e.g., Glosten and Milgrom, 1985) presume that fundamental traders endowed with a perishable positive (negative) private signal aggressively buy (sell). As a result, popular empirical tools often attach a pivotal role to the trade initiator or, by extension, to the trade-initiator-based order imbalance (OI), in signaling order flow toxicity (e.g., Hasbrouck, 1991; Easley et al., 1996). Yet, a variety of recent theoretical (e.g., Goettler, Parlour, and Rajan, 2009), empirical (e.g., Anand, Chakravarty, and Martell, 2005), and experimental (e.g., Bloomfield, O'Hara, and Saar, 2005) studies suggest that informed traders might frequently choose to make rather than take liquidity.

In the high-frequency world, O'Hara (2015) claims that sophisticated informed traders rarely cross the spread. Consistently, Brogaard, Hendershott, and Riordan (2019) conclude that price discovery in Canadian markets occurs predominantly through limit orders, the bulk of which is submitted by high-frequency traders (HFTs). Moreover, recent empirical evidence from US markets suggests that initiator-based OIs weakly correlate with toxicity.[1] Overall, the extant literature suggests that to properly capture order flow toxicity in modern high-frequency markets, one needs to account for the net inflow of non-marketable orders.

We use a detail-rich order level database from the National Stock Exchange of India (NSE), a market ruled by algorithmic trading, to study the sources of order flow toxicity. As an overall summary metric, we use the net order flow (NOF), defined as the relative imbalance between the buying and selling pressure over a short time interval. Buying pressure increases with submissions of both market or limit buy orders, cancellations of standing limit orders to sell, and updates of standing limit orders to buy (sell) that increase (decrease) order size. Selling pressure is defined analogously. By decomposing the NOF into pieces attributable to different types of orders (marketable vs. non-marketable), messages (submissions vs. updates), or traders (HFTs vs. non-HFTs), we perform a comprehensive and detailed empirical examination on the information content of the order flow and their ability to signal toxicity.

In our analysis, we pay special attention to HFTs because they currently account for most of the message traffic (e.g., SEC, 2014), and their orders have been shown to convey information (e.g., Brogaard et al., 2019).[2] What is more, the literature suggests that their order flow could be toxic or at least signal order flow toxicity. On the one hand, when HFTs take liquidity they generate adverse-selection costs on slower passive traders (e.g.,

[1] Kim and Stoll (2014) find that OIs do not reflect private information about posterior corporate information events. Collin-Dufresne and Fos (2015) show that initiator-based measures of adverse-selection do not reveal the presence of informed trading. Easley, López de Prado, and O'Hara (2016) find a negative correlation between the OI and low-frequency estimates of the bid-ask spread. Finally, Barardehi, Bernhardt, and Davies (2019) find that periods of higher OI are associated with smaller price impacts.

[2] HFTs process public (hard) information so as to extract valuable signals about the incoming order flow (Hirschey, 2018), firm fundamentals (Chakrabarty, Moulton, and Wang, 2019), or short-term price movements (Brogaard, Hendershott, and Riordan, 2014). HFTs can even free-ride on information acquisition by slower fundamental traders (Yang and Zhu, 2019; van Kervel and Menkveld, 2019), which could actually discourage fundamental research (e.g., Weller, 2018).

Benos and Sagade, 2016). Trades initiated by these so-called "HF-bandits" or "stale-quote snipers" (e.g., Baldauf and Mollner, 2019) are inherently toxic. Accordingly, a higher presence of HF-bandits has been found to correlate with lower liquidity (van Kervel, 2015; Foucault, Kozhan, and Tham, 2017). On the other hand, high-frequency market makers (HF-MMs) take advantage of their superior speed, low monitoring costs, and enhanced information-processing capacity to actively update quotes in response to incoming news or upon detecting informed trading (Jovanovic and Menkveld, 2016). Thus, high rates of cancellations and revisions of their standing limit orders might reflect active risk management. Indeed, HFTs happen to face lower adverse-selection costs (Hoffmann, 2014; Brogaard et al., 2015), and their quotes incorporate information faster (Riordan and Storkenmaier, 2012). So, the HFTs' passive order flow may not be toxic itself, but still signal incoming toxic order flow.

We first examine the dynamic relationship between quote midpoint returns and order flow imbalances for individual stocks. We follow Chordia and Subrahmanyam (2004) but, since order flow imbalances over short intraday bars are more likely to be driven by informative signals, we use intraday rather than daily bars. We find that contemporaneous quote midpoint returns are positively correlated with the NOF. The NOF explanatory power increases as we ignore messages affecting secondary levels of the book. As the length of the time interval increases, quotes become more responsive to changes in the NOF component due to non-marketable limit orders than the component due to market orders (OI).

Next, we examine whether the NOF or its components convey information. In the spirit of Hasbrouck (1991), we use a structural vector-autoregressive (SVAR) model to estimate the permanent impact on the quote midpoint of a shock to the NOF. While some recent studies assess the average informational content of individual orders or trades[3], our interest is on the information content of the NOF computed for alternative intraday time bars. We find that shocks to the NOF have a larger permanent quote midpoint impact than shocks to the OI, the difference increasing with the length of the time bar. Actually, the non-marketable component of the NOF alone has a larger permanent quote midpoint impact than the OI. In line with Brogaard et al. (2019), we show that order-flow imbalance metrics that account for non-marketable limit orders convey more information.

We control by trader type by splitting the NOF into components attributable to HFTs, agency algorithmic traders (AATs), and non-algorithmic traders (NATs). We find that for short time bars, the HFTs' NOF is more informative than the non-HFTs' NOF, and this superior information content is driven by their OI. Consistent with the notion that HFTs trade on extremely short-lived informative signals (e.g., Hirschey, 2018), we find that the information content of the HFTs' NOF weakens with time aggregation while the non-HFTs' NOF remains strongly informative.

We also split the non-marketable NOF into a component due to order submissions and a component due to cancellations and revisions (C&R). We find that, over short time bars, C&R convey information beyond submissions. Our finding aligns with recent studies suggesting that HF-MMs actively manage the risk by refreshing quotes quickly on hard information (Jovanovic and Menkveld, 2016), and do not support the alternative view that HFTs' flickering quotes reflect gaming and fraudulent practices by HFTs (e.g., Eggington, Van Ness, and Van Ness, 2016).

Finally, we address the issue of which components of the NOF can be useful as leading indicators of order flow toxicity. In particular, we examine which NOF components can

[3] See Brogaard et al. (2019), Chakrabarty et al. (2019), and Fleming, Mizrach, and Nguyen (2018).

anticipate short-term liquidity drops, as measured by several high-frequency illiquidity metrics. We find that the absolute NOF is negatively related to illiquidity in the short run. When we look to its absolute non-marketable component alone, the negative relationship persists. As Easley et al. (2016), we also find a weakly negative relationship between the absolute OI and illiquidity. Therefore, neither the NOF nor the OI can be the basis on which to build an effective advanced indicator of adverse selection in modern financial markets.

However, when we control for trader type, we picture changes significantly. We find that the HFTs' NOF signals toxicity, while the non-HFTs' NOF does not. Both the aggressive (OI) and non-aggressive components of the HFTs' absolute NOF are strongly positively related to short-term illiquidity. We also show that rather than impairing the signaling capacity of the HFTs' NOF, the C&R of their standing limit orders contribute to it, reinforcing our conclusion that HFTs' flickering quotes primarily reflect active risk management rather than manipulative practices. Finally, we observe that most of the signaling capacity of the HFTs' NOF is attributable to the order flow that alters the market quotes.

Our study contributes to the extant literature by providing the first detail-rich analysis of the information content and signaling capacity of the order flow in today's high-speed financial markets. Moreover, our findings have relevant practical implications. According to them, academics should track the HFTs' NOF at or near the best quotes to develop effective leading indicators of order flow toxicity. In performing that task, researchers should not rely exclusively on trade data, since the HFTs' non-marketable order flow, even C&R, signals toxicity too. Market authorities could rely on such indicators to design circuit breaker mechanisms that could effectively prevent short-term illiquidity shortfalls.

Our focus on time-aggregated order flow rather than order-by-order data is justified by previous work in the market microstructure field that makes use of trade-initiator-based order imbalances to build toxicity metrics such as the well-known low-frequency PIN (Easley et al., 1996) or the most recent high-frequency VPIN (Easley et al., 2012). The PIN method relies on (daily) OIs, while the VPIN approach uses a probabilistic method called BVC to assign direction to trade volume. Our results suggest that these statistical tools could improve their performance by simply using the proper input. In her analysis of the challenges market microstructure faces in the high-frequency era, O'Hara (2015) calls for new analytical tools for empirical work, since favorite techniques from the microstructure tool kit may no longer work. Our analysis identifies components of the overall NOF that, once isolated, could be the grounds of a new generation of metrics that accurately signal order flow toxicity.

The rest of the paper is organized as follows. In section 2, we give some market background and describe the database. In section 3, we provide methodological details. In section 4, we assess the informativeness of the NOF. In section 5, we control for type of trader and type of message. In section 6, we examine whether the NOF signals order flow toxicity. In section 7, we conclude.

2. Market background, database, and sample

The NSE is the 4th largest exchange in the world in terms of number of trades and the 10th largest exchange in the world in terms of dollar volume.[4] Indian equity markets have a near non-fragmented structure. With only two exchanges where any meaningful trading takes place, the NSE accounts for 80% of the total domestic trading volume.[5] The NSE is a fully electronic order-driven market with no designated market makers. The market

[4] https://www.world-exchanges.org/home/index.php/statistics/annual-statistics
[5] https://www.sebi.gov.in/sebi_data/attachdocs/1463726488005.pdf

opens with a 15-minute pre-opening session. Continuous trading takes place from 9:15 a.m. till 3:30 p.m. The trading system follows price-exposure-time priority. For the benefit of the market participants, the exchange publicly displays on its website real-time information of the top five ask and bid quotes (price and depth).

Like every important modern stock market, NSE is characterized by the prominent presence of algorithmic traders (ATs). Although AT has been allowed since April 2008, it became widespread once the co-location service was introduced in January 2010. Nawn and Banerjee (2019), for example, report that 95% of the order messages and 43% of the trading volume in the 50 largest 2013 NSE-listed stocks comes from AT.

In this study, we use four months, from April to July 2015, of high-frequency data on the fifty constituents of the NSE's benchmark market index, the NIFTY-50, as on April 30, 2015. These stocks are the largest in terms of market capitalization. Together, they account for approximately 60% of the total market value.

The data is provided by the exchange itself. For each trading day, we have an order file and a trade file. The order file contains detailed information on every order message, including submissions of market and limit orders, and cancellations and revisions of standing limit orders. The database identifies orders with special conditions, such as hidden volume (iceberg orders), on-stop, or immediate-or-cancel. Each order is identified with a unique code, meaning that we can track each order's history overtime. The most common type of order by far is that of limit orders, for which we know the limit price, the displayed size, and the hidden size. The trade files provide information on each individual trade, including the size, the price, the code of the orders involved, and the time at which the trade took place. An incoming aggressive order can be executed against several standing limit orders on the opposite side of the book. In such a case, the trade is reported in several entries, one for each passive order executed.

Orders and trades are time-stamped in jiffies (one jiffy is $1/2^{\wedge 16}$th of a second). We use the codes developed by Chakrabarty et al. (2019) to match the order and the trade files, collapse the trades reported in fragments, assign direction to trades, and rebuild the LOB of each NSE-listed stock in our sample at every point in time.

The database includes two exchange market flags that allow us to locate the orders placed by HFTs. On the one hand, the algorithmic trading flag identifies the orders placed by algorithms. On the other hand, the client flag indicates whether an order is for a proprietary or a client account. Based on these two flags, we can classify each order as coming from one of three mutually exclusive and exhaustive groups of traders: algorithmic orders from a proprietary account are attributed to HFTs (e.g., SEC, 2010); algorithmic orders from a client account are attributed to other or "agency" ATs (AATs), and finally orders not placed by algorithms are attributed to non-ATs (NATs). Our identification criterion allows traders to switch their type, moving away from the frequent assumption of an irreversible HFT classification (e.g., Bellia, 2017; Brogaard et al., 2014).

In Table I, we provide sample descriptive statistics. In Panel A, we show that there are economically meaningful differences across the fifty stocks in our sample in terms of market capitalization, activity, liquidity, volatility and price. In Panel B, we provide the contribution of each type of trader to the order flow. Averaged across stocks, HFTs account for 83.5% of the daily message traffic, 48.3% of the daily order submissions, and 86.6% of the daily revisions and cancellations. In contrast, NATs contribution is much lower: 4.7%, 22%, and 3.2%, respectively. In Panel C, we show that the HFTs' message-to-trade ratio is 13.8 (41) times larger than of AATs (NATs), and their cancellation-to-trade ratio is 16.2 (61.5) times larger. These ratios are frequently used in the literature as proxies for HFT (e.g., Chakrabarty, Moulton, and Pascual, 2017), proving the soundness of our method to identify HFT.

[Table I]

3. Methodological details

To perform our analysis, we pre-aggregate volume into time bars of equal length. The shortest bars last one second and the longest half an hour (1800 sec.). Consider first our most general version of the NOF, computed using all messages within each bar. For each bar *b*, we compute the buying pressure for stock *i* as

$$BP_{i,b} = V_{i,b}^{MB} + V_{i,b}^{LB} + V_{i,b}^{CS} \quad [1]$$

where V^{MB} (V^{LB}) represents the accumulated size of all the market or marketable limit (non-marketable limit) orders to buy submitted within bar *b*. V^{CS} is the accumulated size of the standing limit orders to sell cancelled within bar *b*. Revisions of standing limit orders to buy that increase the order size are treated as new non-marketable limit order submissions and therefore added to V^{LB} by the amount of the revision. Similarly, revisions of limit orders to sell that decrease the order size are treated as cancellations and therefore added to V^{CS} by the size of the order.[6] The selling pressure for stock *i* and bar *b* is computed analogously,

$$SP_{i,b} = V_{i,b}^{MS} + V_{i,b}^{LS} + V_{i,b}^{CB} \quad [2]$$

with V^{MS} (V^{LS}) being the accumulated size of all the market or marketable limit (non-marketable limit) orders to sell and V^{CS} being the accumulated size of the standing limit

[6] We ignore order aggressiveness in this version of our metric. As a result, price revisions of standing limit orders have no effect on the NOF. For alternative versions of the NOF we introduce next, price limit revisions that move orders close (farther away) from the best quotes would be treated as order submissions (cancellations) and do will have an effect on the NOF metric.

orders to buy cancelled within bar b. We define the order flow volume for stock i and bar b as the sum of [1] and [2],

$$V_{i,b}^{OF} = BP_{i,b} + SP_{i,b} \tag{3}$$

and the NOF as the relative difference between [1] and [2],

$$NOF_{i,b} = \frac{BP_{i,b} - SP_{i,b}}{V_{i,b}^{OF}} \tag{4}$$

We consider alternative versions of the NOF metric conditioning on order type (m), order aggressiveness (l), and trader type (tr). We use $NOF_{i,b}^{l}(tr, m)$ as our general notation. As for aggressiveness, l = "a" means we consider the whole LOB; otherwise, l = k means that we only use orders equating, hitting, or improving the prevailing k best quotes. Regarding trader types, we consider tr = {a, HFT, AT, NAT}, where "a" in this case means all trader types. Finally, as for order types m = {a, L, S, M}, where "a" means all types of messages, "L" means all messages but market orders and marketable limit orders, "S" means non-marketable limit order submissions, and "M" stands for monitoring, meaning both cancellations and revisions of standing limit orders (e.g., Liu, 2009). To simplify the notation and exposition, non-specified parameters are set to "a". For example, $NOF_{i,b}$ is equivalent to $NOF_{i,b}^{a}(a,a)$, and $NOF_{i,b}(L)$ is equivalent to $NOF_{i,b}^{a}(a,L)$.

Additionally, we define the order imbalance for stock i and bar b as

$$OI_{i,b} = \frac{V_{i,b}^{TB} - V_{i,b}^{TS}}{V_{i,b}^{T}} \tag{5}$$

where V^{TB} (V^{TS}) represents the accumulated buyer- (seller-) initiated volume and V^{T} is the total volume traded,

$$V_{i,b}^{T} = V_{i,b}^{TB} + V_{i,b}^{TS} \quad [6]$$

Notice that V^{TB} in [5] and V^{MB} in [1] are not necessarily equal, since marketable limit orders may not be fully executed.

In Table II, we provide cross-sectional daily average statistics on message traffic. In Panel A, we show that C&R of standing limit orders account for 86.82% of all messages on an average trading session, followed by the submission of non-marketable orders (9.8%), and the submissions of marketable orders (3.38%). In Panel B, we look at the composition of message traffic conditional on the level of the LOB. We distinguish between orders placed or standing at or within the prevailing best quotes, up to the prevailing 5th best quote, and beyond the prevailing 5th best quote. The weight of C&R in total message traffic decreases with aggressiveness. For orders placed or standing at or within the best quotes, C&R represent 56.29% of all messages. For the least aggressive order category, C&R rule the trading game, representing 92.4% of all messages. Finally, in Panel C, we show that HFTs are the main contributors to message traffic, no matter the aggressiveness of the orders involved.

[Table II]

In Table III, we provide cross-sectional average correlations between $NOF_{i,b}^{l}(tr,m)$ and OI for selected time bars (b) in seconds. We use l=1 (Panel A), l=5 (Panel B) and l=a (Panel C). Per l case, we also consider m=a (i.e., all orders), and m=L (i.e., non-marketable limit orders only). Since $NOF_{i,b}^{l}$ comprises $OI_{i,b}$ and the latter is built from trades only, it should not come as a surprise that the correlation between both metrics is positive and significant for all b but decreasing with l. For b=60s, for example, the correlation falls from 0.518 with l=1 to 0.293 with l=a. Yet, average correlations between $NOF_{i,b}^{l}(L)$ and $OI_{i,b}$ are weaker or even negative for larger bars. This negative correlation suggests that larger and positive OIs come with more cancellations on the bid side of the book (perhaps to place aggressive orders instead) and/or with an increase in the submission of new limit orders to sell, triggered by a decrease in their non-execution risk. Apparently, these inflows outweigh new submissions of non-aggressive limit orders to buy or cancellations on the ask side of the book.

[Table III]

4. The informativeness of the NOF

4.1. The NOF-return relationship

We first examine the link between intraday returns and contemporaneous and lagged order flow imbalances. If the NOF were to signal private information, prices should strongly react to shocks to the NOF, and those shocks should have a permanent impact on prices (e.g., Hasbrouck, 1991). We understand this is a necessary condition for the NOF to signal order flow toxicity.

Our analysis is related to Chordia and Subrahmanyam (2004), who show that returns for individual stocks in US markets (in the pre-HFT era) are affected by contemporaneous and lagged OIs. They interpret their findings as consistent with an equilibrium model in which market makers with inventory concerns accommodate serially correlated OIs.[7] In such a context, price changes experience reversals over long horizons as the price pressure caused by OIs is eventually reversed (e.g., Hendershott and Menkveld, 2014).

At the intraday level, however, order-flow imbalances for individual stocks could be induced by either public or private information and therefore have a permanent impact on prices (e.g., Hasbrouck, 1996; O'Hara, Yao, and Ye, 2014). Benos et al. (2017), for

[7] Specifically, in their model order splitting cause positive autocorrelation in OIs, giving rise to price pressure and, then, a positive predictive relation between imbalances and future returns.

example, provide evidence of commonality in trading behavior across HFT firms. Aggressive buying (selling) by a HFT is associated with subsequent aggressive buying (selling) by other HFTs. This correlated trading behavior generates positively auto-correlated OIs over short-term intervals. They show that the HFT order flow correlation is associated with permanent price impacts rather than price reversals, which is consistent with HFT commonality being related to informed trading.

Recently, Johnson and Watson (2018) use Nasdaq ITCH data to show that daily limit order imbalances explain returns, but traditional OIs have more explanatory power.[8] We control for order aggressiveness, account for order size, and drop fleeting orders (Hasbrouck and Saar, 2009) as they also do.[9] However, our analysis differs in many important ways: we look at intraday rather than daily order flow imbalances, distinguish trader types, and consider sources of price pressure other than order submissions, such as cancellations and revisions.

For each time bar, we compute the continuously compound quote midpoint return as

$$r_{i,b} = \ln(q_{i,b}) - \ln(q_{i,b-1}) \qquad [10]$$

where $q_{i,b}$ is the quote midpoint at the end of bar b. We use the return in equation [10], expressed in basis points, as the dependent variable of the following pooled regression model with standard errors clustered by stock,

$$r_{i,b} = \alpha + \sum_{k=0}^{n} \beta_k' IM_{i,b-k} + \sum_{k=1}^{n} \gamma_k r_{i,b-k} + \alpha_O OP_{i,b} + \alpha_C CP_{i,b} + e_{i,b} \qquad [11]$$

[8] Johnson and Watson (2018) conclude that limit order imbalances convey information, but they do not formally prove the information channel and they never rule out the price pressure story of Chordia and Subrahmanyam (2004), whose findings are totally consistent with.

[9] In our approach, fleeting orders cancel themselves out in computing the NOF metric, as they are added when submitted and immediately subtracted when cancelled a few milliseconds later.

where OP (CP) is a dummy that equals one for the first (last) 60 minutes of trading of each trading session and zero otherwise. IM is a vector containing the focal order flow imbalance metric(s). In order to make coefficients comparable across order flow metrics, hereafter we standardize all the variables in IM to have zero mean and unit standard deviation per stock.[10] The optimal number of lags (n) is determined per bar length using the Akaike Information Criterion (AIC) with a maximum of 10 lags.

In Table IV, we provide the estimated β_k in eq. [11] up to the fifth lag for selected time bars, and for the following choices of IM: $NOF_{i,b}^{1}$ (Panel A), $NOF_{i,b}^{5}$ (Panel B), and $NOF_{i,b}$ (Panel C). In general, we find that all the NOF metrics are positively correlated to contemporaneous quote mid-point returns for all b. With 5-second bars, a 1% increase in the standardized $NOF_{i,b}^{1}$ comes with a 0.96% increase in the quote midpoint return in average terms across all stocks in our sample. The model's explanatory power increases with the length of the time bar (b), and decreases with the maximum level of the LOB considered (l), suggesting the NOF becomes noisier when it is aggregated over shorter bars and messages placed beyond the best quotes are taken into account.

[Table IV]

In Table V, we re-estimate equation [11] but with $IM_{i,b} = \left[NOF_{i,b}^{l}(L) \;\; OI_{i,b} \right]$. That is, we decompose the $NOF_{i,b}^{l}$ into two pieces, one due to non-marketable limit orders and the other due to market and marketable limit orders. In Panel A, l=1, and in Panel B, l=5. We find that $NOF_{i,b}^{l}(L)$ is positively correlated with contemporaneous quote midpoint

[10] Results with non-standardized variables are consistent and available at the Internet Appendix of the paper.

returns, even after we account for the OI. Indeed, as we increase the bar size b, returns become slightly more responsive to shocks to $NOF^{l}_{i,b}(L)$ than to shocks to $OI_{i,b}$.

[Table V]

4.2. Permanent quote midpoint impact of the NOF

Our findings in Tables IV and V do not suffice to conclude that the NOF, or any of its components, conveys relevant information for the price discovery process. Chordia and Subrahmanyam (2004) show that positively auto-correlated daily OIs give rise to a positive relation between OI and future returns, but this relation reverses sign when contemporaneous OIs are accounted for. We find similar patterns at the intraday level. In unreported tests, we find OIs to be positively serially correlated, and the strength of that correlation to increase with time aggregation. For example, the first order auto-correlation in the OI series for 60s (300s) bars averaged across stocks is 0.16 (0.23). We also find OIs to be positively correlated with *posterior* returns. In Table V, however, we report negative coefficients for lagged OIs. So, we cannot rule out the possibility that shocks to the NOF cause transitory quote midpoint impacts.

To formally disentangle whether shocks to the NOF cause permanent impacts on prices, suggesting information content, or transitory disruptions, as should be the case under the inventory holding costs paradigm (e.g., Ho and Stoll, 1983), we follow O'Hara et al. (2014) in estimating a structural vector-autoregressive (SVAR) model for quote midpoint changes and order-flow imbalances. In the spirit of Hasbrouck's (1991), we estimate the following bivariate SVAR model with k lags for each stock,

$$\begin{bmatrix} 1 & -\phi_0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_{i,b} \\ NOF^{l}_{i,b} \end{bmatrix} = \Psi(L) \begin{bmatrix} r_{i,b} \\ NOF^{l}_{i,b} \end{bmatrix} + \Lambda G_t + \begin{bmatrix} \varepsilon^{r}_{i,b} \\ \varepsilon^{NOF}_{i,b} \end{bmatrix} \quad [12]$$

where $\Psi(.)$ is a matrix of polynomials of order k in the lag operator L. G_t include deterministic functions of time. In particular, dummies for the first and last trading hour of each session. We exclude observations with lags that date back to the previous day. The optimal number of lags (k) per stock is determined using the AIC criterion. For each stock, we estimate model [12] and obtain the structural cumulative impulse-response function (IRF) to a one-percentage-point increase in the corresponding imbalance metric, our estimate of the permanent quote midpoint impact. Because our order flow imbalance metrics are standardized, the resulting estimated impacts are directly comparable. We provide cross-sectional average estimates of the IRF for different time bar sizes in Table VI. IRFs are reported both in basis points and in relative terms to the standard deviation of the quote midpoint returns.

[Table VI]

Table VI reports positive cross-sectional average permanent quote midpoint impacts across the board. Individually considered, the impacts are also significant for all the stocks in our sample. Even after we control for the standard deviation of the quote midpoint return, we find that the permanent price impact significantly increases with the size of the time bar. For example, a 1% shock to the standardized $NOF_{i,1}^{1}$ (Panel A) has an average permanent quote midpoint impact of 0.75bps (4.7bps) or 0.41 (0.52) times the standard deviation of 1- (60-) second quote midpoint return. Additional tests, not reported in Table VI, show that the permanent quote midpoint impact of a shock to $NOF_{i,b}^{1}$ (Panel A) or $NOF_{i,b}^{5}$ (Panel B) is larger than an equivalent shock to $NOF_{i,b}$ (Panel C) for all b. That is, considering orders placed beyond the five best quotes to compute the NOF adds nothing but noise. For that reason, from this point on we focus exclusively on the order flow affecting up to the five best levels of the LOB, that is, we consider $l \leq 5$.

4.3. Permanent quote midpoint impact of the NOF's components

We examine next the case $IM_{i,b} = \left[NOF_{i,b}^{l}(L) \;\; OI_{i,b} \right]$, that is, we separate the non-marketable order flow from the marketable order flow. We estimate the SVAR model in equation [13], where we presume contemporaneous one-way causality runs from $NOF_{i,b}^{l}(L)$ and $OI_{i,b}$ to returns, and from $OI_{i,b}$ to $NOF_{i,b}^{l}(L)$. This model should render a lower (upper) bound estimate for the IRF of a shock to $NOF_{i,b}^{l}(L)$ ($OI_{i,b}$). By then reversing the later causality assumption, we obtain the corresponding upper (lower) bound. In Table VII, we report cross-sectional averages of the midpoint between the upper bound and lower bound obtained for each IRF for *l*=1.

$$\begin{bmatrix} 1 & -\phi_0^r & -\pi_0^r \\ 0 & 1 & -\pi_0^{NOF} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{i,b} \\ NOF_{i,b}^{l}(L) \\ OI_{i,b} \end{bmatrix} = \Phi(L) \begin{bmatrix} r_{i,b} \\ NOF_{i,b}^{l}(L) \\ OI_{i,b} \end{bmatrix} + \Lambda G_t + \begin{bmatrix} \varepsilon_{i,b}^{r} \\ \varepsilon_{i,b}^{NOF} \\ \varepsilon_{i,b}^{OI} \end{bmatrix} \qquad [13]$$

[Table VII]

Consistent with the pooled regression findings in Table V, Table VII shows that a one-percent-point shock to the standardized $NOF_{i,b}^{1}(L)$ has a larger permanent quote midpoint impact than an equivalent shock to the standardized $OI_{i,b}$ for all *b*. Differences in informational content between the two components increase with *b*. In fact, for *l*=5 (see Table AI in the Appendix), differences are only statistically significant for *b*>5. Therefore, intraday order-flow imbalances computed from non-marketable limit orders have explanatory power on quote midpoint returns beyond that of initiator-based OIs. For relatively large bars, quote midpoints are even more responsive to unexpected changes in the net flow of non-marketable limit orders than to similar unexpected shocks to the OI.

Overall, both marketable and non-marketable order-flow imbalances explain quote midpoint returns over relatively short time intervals, and unexpected shocks to these imbalances convey information and cause a statistically significant permanent impact on the quote midpoint. We therefore conclude that both $NOF^{l}_{i,b}(L)$ and $OI_{i,b}$ convey information. As we show later, this is a necessary but not sufficient condition for them to also signal toxic order flow.

5. NOF and stock returns: conditional tests

5.1. Trader types

The extant literature distinguishes between two types of algorithm traders (ATs): proprietary ATs (a.k.a., HFTs) and agency ATs (e.g., O'Hara, 2015, Menkveld, 2016). Agency AT (AAT) is a service offered by broker-dealers and software developers to buy-side clients to optimize trade execution both across venues and over time. Presumably, the success of the clients' trading strategies does not rely on speed of implementation. In contrast, HFTs are technologically sophisticated investors that use their own algorithms to trade on their own account and capital, implementing a wide variety of trading strategies aimed to profit from the trading process itself at an extraordinary high speed.[11]

The *order entry mode* flag in the NSE database indicates whether a given message is generated, placed, and managed either by an algorithmic terminal or "manually". The *client account* flag distinguishes messages of proprietary accounts from messages of client accounts. Combining these two flags we separate the messages placed by HFTs (AT and proprietary account) from the messages placed by AATs (AT and client

[11] HFTs heterogeneous trading strategies include market making (e.g., Menkveld, 2013), latency arbitrage (e.g., Chaboud et al., 2014), news reaction strategies (e.g., Scholtus, van Dijk, and Frijns, 2014), directional trading (e.g., Brogaard et al., 2014), order-discovery strategies (e.g., van Kervel and Menkveld, 2019), and unethical manipulative strategies such as quote stuffing and spoofing (e.g., Angel and McCabe, 2013).

account). We group the messages not handled by algorithms into the non-AT (hereafter, NAT) category.[12]

We compute the NOF attributable to HFTs, AATs, and NATs, and evaluate their relative informational content using the following SVAR,

$$\begin{bmatrix} 1 & -\phi_{h0}^{r} & -\phi_{a0}^{r} & -\phi_{n0}^{r} \\ 0 & 1 & -\phi_{a0}^{HFT} & -\phi_{n0}^{HFT} \\ 0 & 0 & 1 & -\phi_{n0}^{AAT} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{i,b} \\ IM_{i,b}^{HFT} \\ IM_{i,b}^{AAT} \\ IM_{i,b}^{NAT} \end{bmatrix} = \Phi(L) \begin{bmatrix} r_{i,b} \\ IM_{i,b}^{HFT} \\ IM_{i,b}^{AAT} \\ IM_{i,b}^{NAT} \end{bmatrix} + \Lambda G_t + \begin{bmatrix} \varepsilon_{i,b}^{r} \\ \varepsilon_{i,b}^{HFT} \\ \varepsilon_{i,b}^{AAT} \\ \varepsilon_{i,b}^{NAT} \end{bmatrix} \qquad [14]$$

where $\Phi(L)$ is a vector of lag polynomials of finite order k. As in previous models, the standardized order-flow imbalance metrics ($IM_{i,t}^{j}$) cause the contemporaneous quote-midpoint returns. As in model [13], we alter the causality ordering among the $IM_{i,t}^{j}$ metrics to obtain upper and lower bounds for each IRF. The resulting cross-sectional average accumulated IRFs for l = 1 (l = 5) are reported in Table VIII (Table AII in the Appendix). In Panel A, we provide results for $NOF_{i,b}$. In Panel B, we distinguish between $NOF_{i,b}^{l}(L)$ and $OI_{i,b}$, as in Table VII.

[Table VIII]

In Panel A of Table VIII, we observe that, for all trader types, the $NOF_{i,b}$ conveys information. For b<=5 the quote midpoint impact of a shock to the HFTs' NOF is statistically larger than that of an equivalent shock to the non-HFTs' NOF, both in

[12] Additionally, the database distinguishes between two types of client accounts, custodians and others. According to Chakrabarty et al. (2019), the custodians are members of the exchange that do not conduct their own clearing or settlement. This group primarily involves foreign institutional investors, mutual funds, and financial institutions. The other clients group comprises domestic firms and retail traders that employ their own clearing member. While HFTs only include proprietary ATs, AATs involve custodians (18.9% of messages in our sample) and others (81.1%), and NATs comprise custodians (6.4%), others (54.2%), and proprietary traders (39.4%).

absolute and relative terms. For example, a one-percent-point shock to the standardized $NOF_{i,1}^{1}\left(HFT\right)$ has an average permanent quote midpoint impact of 0.65bps, or 0.35 times the standard deviation of the 1-second quote midpoint return. For AATs (NATs), a similar shock results in an estimated quote midpoint impact of 0.29bps (0.16bps), or 0.40 (0.22) times the standard deviations of the quote midpoint return. All these differences are statistically significant at the 1% level. The HFTs' NOF, however, loses information content with time aggregation. Thus, for *b*>5 the relative quote midpoint impact of a shock to $NOF_{i,b}^{1}\left(HFT\right)$ falls remarkably. Indeed, it even conveys less information than a similar shock to $NOF_{i,b}^{1}\left(AAT\right)$ or $NOF_{i,b}^{1}\left(NAT\right)$. For *b*=300, shocks to $NOF_{i,b}^{1}\left(HFT\right)$ have a significant impact only for 30 stocks in our sample.

From Panel B of Table VIII, we obtain additional insights. We look first at the OI. Consistent with existing evidence for US markets (e.g., Brogaard, et al., 2014, 2017), we find that that trades initiated by HFTs convey information, but only for small bars. For *b*>5, the standardized $OI_{i,b}\left(HFT\right)$ conveys no information. In contrast, for *b*≤5 it conveys more information than the OI of non-HFTs. Our findings therefore support the common understanding that HF-bandits trade on extremely short-lived informative signals (e.g., Hirschey, 2018). Accordingly, the $OI_{i,b}\left(HFT\right)$ aggregated over relatively large time bars should perform poorly as an indicator of toxicity.

In contrast, the permanent quote-midpoint impact of a one-percent-point shock to the $OI_{i,b}\left(AAT\right)$ is strongly positive and significant for all *b*, and larger than for the $OI_{i,b}\left(NAT\right)$ case. This finding might indicate that sophisticated institutional traders at the NSE prefer to handle their information-motivated trades (e.g., Anand et al., 2005) through AT, but it could also reflect that unsophisticated retail traders are adding noise to the $OI_{i,b}\left(NAT\right)$ (e.g., Barber and Odean, 2000).

Regarding $NOF_{i,b}^{1}(L)$, for non-HFTs it conveys information across the board. For HFTs, however, the information content decreases with *b*. For relatively small bar sizes (*b*<=5), differences across trader types are either non-significant or economically negligible. All this evidence leads us to conclude that the higher informational content of $NOF_{i,b\leq 5}^{1}\left(HFT\right)$ reported in Panel A of Table VIII is driven by the marketable orders.

5.2. Message types

Since the advent of HFT, the frequency of quote updates in financial markets has increased dramatically (e.g., Angel, Harris, and Spatt, 2015). Phenomena such as flickering quotes or fleeting orders (e.g., Hasbrouck and Saar, 2009) are not the response to fundamental information arrival (e.g., Hasbrouck, 2018). Rather, they may be the result of HFTs using their low-latency technology to monitor the markets in a near-continuous way to immediately react to market events (e.g., Brolley and Malinova, 2017). Thus, high rates of C&R of standing limit orders may naturally arise as a result of HF-MMs managing their risk of being adversely selected (e.g., Jovanovic and Menkveld, 2016) or undercut (e.g., Baruch and Glosten, 2013), or as a response to order-flow or order-book related signals (e.g., Aït-Sahalia and Saglam, 2017; Dahlström, Hagströmer, and Nordén, 2018). If this is actually the case, shocks to C&R imbalances may convey information and signal toxic order flow. Yet, C&R could also arise from HFT manipulative strategies, such as quote stuffing (e.g., Egginton, Van Ness, and Van Ness, 2016). If high rates of C&R reflect gaming and fraudulent practices by the HFTs, shocks to C&R imbalances may be noisy and uncorrelated to toxic order flow.

By construction, the NOF in equation [4] is unaffected by fleeting orders, since the submission and immediate cancellation of the same order will offset each other out. Therefore, to test whether C&R imbalances convey information, we need to split the $NOF_{i,b}^{l}(L)$ into $NOF_{i,b}^{l}(S)$, due to submissions of limit orders, and $NOF_{i,b}^{l}(M)$, due to C&R of standing limit orders.

In Table IX for l = 1 (Table AIII in the Appendix for l = 5), we provide the estimated permanent quote-midpoint impact of a one-percent-point shock to each of the above standardized NOF components using the SVAR model approach we have described before. For control purposes, we also include the OI in the model specification, meaning that we estimate a 4-equation SVAR model this time. The remaining methodological details are the same as in previous tests.

[Table IX]

We find that shocks to $NOF_{i,b}^{l}(M)$ have a significant positive quote midpoint impact for all b and l, even when we control for contemporaneous aggressive and non-aggressive order submissions. In other words, updates of standing limit orders convey information beyond order submissions. We also observe that with time aggregation quote midpoints become more responsive to unexpected changes in the net flow of both submissions and C&R of non-marketable limit orders. Our findings are therefore in line with the hypothesis that flickering quotes, fleeting orders, and high cancellation- or quotation-to-trade ratios are due to active monitoring of standing limit orders rather than manipulative strategies by HFTs.

6. NOF and order flow toxicity

In this section, we examine which order-flow imbalance metrics can work as advanced indicators of order flow toxicity. Market microstructure theory of adverse selection predicts that toxic order flow should be negatively related to liquidity (e.g., Glosten and Milgrom, 1985). So, we use high-frequency metrics of liquidity computed over regular 1-second and 5-second time bars to evaluate which components of the NOF anticipate short term liquidity withdrawals. In particular, we consider three metrics of illiquidity:

- The time-weighted relative quoted bid-ask spread (RQS): for each best bid and offer (BBO) posted within a given time bar, we divide the quoted bid-ask spread by the quote midpoint, and use the proportion of time it stays in place as its weight.
- The volume-weighted relative effective spread (RES): for each trade within a time bar, we obtain the relative effective spread as two times the relative deviation between the trade price and the prevailing quote midpoint, and use the trade size (in shares) as its weight.
- Amihud's illiquidity ratio (AIR): Amihud (2002) proposes to use the ratio of the absolute stock return to its volume traded within a given interval as an inverse (direct) metric of liquidity or market depth (price impact). This metric has been extensively used in market microstructure and asset pricing empirical research (e.g., Acharya and Pedersen, 2005).[13] As the numerator of the ratio, we use the continuously compound quote midpoint return, computed using the best quotes standing right at the beginning and end of each time bar. As the denominator of the ratio, we use the volume traded in shares within the bar. We multiply the resulting magnitude by one million.

Methodologically, we follow Easley et al. (2016). For each stock, time bar length, illiquidity metric (*ILLIQ*), and standardized order imbalance metric(s) (*IM*) in absolute terms, we estimate the dynamic regression model in equation [15] with Newey and West (1994) heteroscedasticity and autocorrelation consistent (HAC) standard errors. We apply

[13] Goyenko, Holden, and Trzcinka (2009) find that Amihud's metric is a good proxy for price impact.

a Bartlett Kernel to determine the optimal number of lags to consider in analyzing the correlation structure of the residuals.[14]

$$ILLIQ_{i,b} = \alpha_0 + \alpha_1 ILLIQ_{i,b-1} + \beta' \left| IM_{i,b-1} \right| + \delta_1 V_{i,b-1} + \delta_2 \sigma(\Delta q)_{i,b-1} + \lambda_O OP_{i,b} + \lambda_C CP_{i,b} + e_{i,b} \quad [15]$$

All the non-deterministic explanatory variables in the RHS of equation [15] are lagged one period. We opt for this dynamic structure for two main reasons: firstly, we want to minimize reverse causality issues generated by time aggregation, and secondly, we are interested in the utility of the NOF components to build *advanced* indicators of order flow toxicity. If large absolute values of $IM_{t-1}^{i,b}$ signal toxic order flow, we would expect liquidity providers to respond to shocks to $IM_{t-1}^{i,b}$ by demanding higher compensation in exchange for immediacy and/or reducing the available depth near the best quotes, therefore increasing price impact. In other words, we expect β>0 in equation [15].

It is very well known that liquidity experiences regular intraday patterns (e.g., Upson and Van Ness, 2017). We include two dummies in equation [15] to control for regularities in *ILLIQ* during the first (*OP*) and last (*CP*) regular trading hours of the trading session. Liquidity is often found to be inversely related to volatility and positively connected to trading activity (e.g., Stoll, 2000). We control for volume and volatility by including the standard deviation of the quote midpoint changes ($\sigma(\Delta q)_{t-1}^{i,b}$) and the logarithm of the volume in shares ($V_{t-1}^{i,b}$) within each bar *b*. We control that no lag reaches back to the previous day.

In what follows, we provide summary estimates of equation [15] for our sample of 50-NSE listed stocks. In each table, we report cross-sectional average estimates of the

[14] The number of lags equals Int[4(n/100)$^{2/9}$], where n is the number of observations.

coefficient of interest (β). Statistical significance is determined using the aggregated t-statistic of Chordia, Roll and Subrahmanyam (2005).[15] In addition, we report the number of stocks for which β is found to be statistically significant at the 1% and 5% level. Finally, to economize on the presentation of the results, we report only results for l=1 (best quotes only) and for two bar sizes in seconds b = {1, 5}. Results for l = 5 can be found in the Appendix (see Tables AIV to AVI).

In Table X, we look at the aggregated NOF. We consider two choices for $IM_{t-1}^{i,b}$: (a) $NOF_{i,b}^{1}$, and (b) $NOF_{i,b}^{1}(L)$ and $OI_{i,b}$. Despite the NOF is related to contemporaneous price changes (Table IV), and shocks to the NOF have a statistically significant permanent impact on prices (Table VI), the (absolute) NOF turns out to be negatively related to *RQS*, *RES*, and *AIR* in the next time bar. For b=1, this negative relationship is statistically significant for all stocks individually considered. The negative relationship persists for the $NOF_{i,b}^{1}(L)$, meaning that this finding is not driven by the marketable component of the NOF. In line with the evidence provided by Easley et al. (1996) for US derivative markets, we find that the $OI_{i,b}$ is weakly negatively related with immediacy costs. However, it is strongly and positively related with *AIR*. For l=5, we find the same relationships (see Table AIV in the Appendix). Our findings therefore suggest that the aggregated $NOF_{i,b}^{l}$ can hardly be the basis to build effective advance indicators of order flow toxicity.

[Table X]

[15] Chordia et al. (2005) use the cross-sectional average t-statistic corrected by the residual cross-correlation in the individual stock regressions. Assuming constant pairwise residual correlations, Chordia et al. (2005) show that the standard error of the aggregate estimate is inflated by a factor $[1+(N-1)\rho]^{1/2}$, where N is the number of stocks and ρ is the common cross-correlation. We estimate ρ for each particular specification of the regression model [15] as the average of the 1225 unique residual cross-correlations.

In Table XI, we estimate model [15] again, but this time we split the NOF into the components attributable to HFTs, AATs, and NATs. We find that while the negative relationship between $NOF_{i,b}^{1}$ and $ILLIQ_{i,b+1}$ persists for AATs and NATs, we find a strong and positive relationship between the HFTs' NOF and illiquidity, which is consistent with HFTs trades and orders being related to underlying information. Accordingly, results in Table X are driven by the non-HFTs' NOF. In Table AV in the Appendix, we show that for l=5 the positive relationship between the HFTs' NOF and illiquidity persist, but only for b=1, which suggests that order flow toxicity is best captured by the updates in the market quotes caused by the HFTs' order flow.

[Table XI]

In Table XII, we split the $NOF_{i,b}^{1}$ of each trader type into their aggressive ($OI_{i,b}$) and non-aggressive components ($NOF_{i,b}^{1}(L)$). Consistent with Table XI, we find that the strongest connection with toxicity corresponds to the HFTs' order flow, both through marketable and non-marketable orders submissions and updates. While we cannot fully discard that the AATs' and NATs' marketable order flow is also toxic, only the HFTs' non-marketable order flow signals toxicity. Table AVI in the Appendix shows, again, that adding non-marketable limit orders beyond the best quotes (l=5) weakens the positive connection of the $NOF_{i,b}^{l}(L)$ and $ILLIQ_{i,b+1}$.

[Table XII]

Finally, in Table XIII we examine whether the signaling capacity of the HFTs' $NOF_{i,b}^{1}(L)$ stems from the submission of new limit orders or from C&R of limit orders standing in the book. We estimate equation [15] again, but this time we split each trader type's $NOF^1_{i,b}(L)$ into pieces due to order submissions ($NOF^1_{i,b}(S)$), C&R ($NOF^1_{i,b}(M)$), and marketable orders (OI).

[Table XIII]

Consistent with Table XII, we find that the *OI* component of both HFTs' and NATs' (to a lesser extent) NOF are positively correlated with next period's illiquidity. However, only the HFTs' $NOF^1_{i,b}(L)$ signals toxicity. Additionally, we observe now that both $NOF^1_{i,b}(HFT,S)$ and $NOF^1_{i,b}(HFT,M)$ cause $ILLIQ_{i,b+1}$ to increase. This relationship is significant for almost all stocks individually considered. Therefore, rather than impairing the signaling capacity of the HFTs' NOF, HFT's C&R of standing limit orders contribute to it. Once more, our findings support the hypothesis that HFTs' high rates of C&R mostly reflect intense active monitoring of standing limit orders rather than manipulative practices.

Overall, our analysis suggests that the HFTs' NOF signals toxicity, which cannot be said for the non-HFTs' NOF. This signaling capacity, however, is mostly driven by whatever happens close to the best quotes.

7. Conclusions

We have empirically examined which components of the net order flow (NOF) aggregated over short-term time intervals convey information and can be useful as advanced indicators for order flow toxicity. Our NOF metric considers all types of trading messages. It therefore accounts for the information content on the non-marketable order flow (limit orders), both at submission and in posterior updates (C&R). Using granular data for the constituents of the NSE NIFTY-50, we split the NOF into pieces: marketable (OI) vs. non-marketable orders, submissions vs. C&R, aggressively priced vs. nonaggressively priced orders, and orders placed by HFTs vs. non-HFTs. Our analysis focuses on intraday regular short time bars (from 1 to 1800 seconds). Our main findings are as follows:

- Shocks to the non-marketable NOF have a larger permanent price impact than similar shocks to the OI. Thus, our analysis complements and corroborates Brogaard, et al. (2019) recent tick data analysis showing that, in today's markets, price discovery occurs mostly through non-marketable limit orders.
- The informativeness of the HFTs' NOF decreases with time aggregation, which agrees with extant empirical evidence showing that HFTs exploit their speed advantage to trade on extremely short-lived information (e.g., Hirschey, 2018; Easley, O'Hara, and Yang, 2016).
- The NOF component due to C&R, which is mostly driven by HFTs, conveys information. According to our findings, HFTs' flickering quotes and high cancellation-to-trade ratios are more likely the result of a close monitoring of their standing limit orders rather than fraudulent practices (e.g., Jovanovic and Menkveld, 2016).
- The HFTs' NOF signals toxicity, while the non-HFTs' NOF does not. We split the HFTs' NOF into marketable (OI) and non-marketable components to find that high imbalances in any of them precede short-term liquidity drops. Our findings agree with theoretical models such as van Kervel (2015), Biais, Foucault, and Moinas (2015), Foucault et al. (2017), and Menkveld and Zoican (2017), in which trades initiated by the so-called HF-bandits impose adverse selection costs on liquidity providers and, therefore, are inherently toxic. We find, however, that the HFTs' passive order flow also signals toxicity. We attribute the signaling capacity of the HFTs' non-marketable order flow to their enhanced information-processing capacity that allows them to

update quotes at high speed in response to incoming news and market events, or upon detecting informed trading (Hoffmann, 2014; Jovanovic and Menkveld, 2016). As a result, the HFTs' passive order flow anticipates incoming toxic order flow.

- The signaling capacity of the HFTs' NOF mostly concentrates on the best quotes.

We conclude that by isolating the HFTs' NOF at or close to the best quotes, academics could develop efficient leading indicators for order flow toxicity. Market authorities could rely on such indicators to build effective circuit breaker mechanisms directed to protect liquidity suppliers from being adversely selected and prevent toxicity-driven liquidity shortfalls. We also conclude that an effective leading indicator of order flow toxicity should not depend exclusively on the HFTs' initiated trades. Limit order submissions, revisions, and cancellations by the HFTs should be closely monitored too.

We would like to emphasize that our findings do not necessarily imply that HFTs are informed in the sense that they acquire and trade on non-publicly available information. As pointed by Baldauf and Mollner (2019), Hirschey (2018), and van Kervel and Menkveld (2019), among others, HFTs could simply anticipate trades and orders of truly informed investors. Weller (2018) shows the importance of distinguishing between "acquiring information" and "incorporating information into prices" in talking about price discovery. We cannot make that distinction. But, even when HFTs just trade ahead of the truly information-motivated order flow, their NOF still counts as toxic.

References

Acharya, V.V. and L.H. Pedersen, 2005. Asset pricing with liquidity risk. *Journal of Financial Economics* 77 (2), 375-410.

Aït-Sahalia, Y. and M. Saglam, 2017. High frequency market making: implications for liquidity. Available at SSRN: https://ssrn.com/abstract=2908438.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5 (1), 31-56.

Anand, A., S. Chakravarty, and T. Martell, 2005. Empirical evidence on the evolution of liquidity: choice of market versus limit orders by informed and uninformed traders. *Journal of Financial Markets* 8 (3), 288-308.

Angel, J.J., L.E. Harris, and C. S. Spatt, 2015. Equity trading in the 21st century: an update. *The Quarterly Journal of Finance* 5 (1), 1550002.

Angel, J.J. and D. McCabe, 2013. Fairness in financial markets: the case of high frequency trading. *Journal of Business Ethics* 112 (4), 585-595.

Baldauf, M. and J. Mollner, 2019. High-frequency trading and market performance. *Journal of Finance*, forthcoming. http://dx.doi.org/10.2139/ssrn.2674767.

Barardehi, Y.H., D. Bernhardt, and R.J. Davies, 2019. Trade-time measures of liquidity. *The Review of Financial Studies* 32 (1), 126-179.

Barber, B.M. and T. Odean, 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. *The Journal of Finance* 55 (2), 773-806.

Baruch, S. and L. Glosten, 2013. Fleeting orders. *Columbia Business School Research Paper* No.13-43. Available at SSRN: https://ssrn.com/abstract=2278457.

Bellia, M., 2017. High-frequency market making: liquidity provision, adverse selection, and competition. Available at SSRN: https://ssrn.com/abstract=

3074313.

Benos, E., J. Brugler, E. Hjalmarsson, and F. Zikes, 2017. Interactions among high-frequency traders. *Journal of Financial and Quantitative Analysis* 52 (4), 1375-1402.

Benos, E. and S. Sagade, 2016. Price discovery and the cross-section of high-frequency trading. *Journal of Financial Markets* 30, 54-77.

Biais, B., T. Foucault, and S. Moinas, 2015. Equilibrium fast trading. *Journal of Financial Economics* 116 (2), 292-313.

Bloomfield R., M. O'Hara, and G. Saar, 2005. The "make or take" decision in an electronic market: evidence on the evolution of liquidity. *Journal of Financial Economics 75* (1), 165-199.

Brogaard, J., B. Hagströmer, L. Nordén, and R. Riordan, 2015. Trading fast and slow: Colocation and liquidity. *The Review of Financial Studies* 28 (12), 3407-3443.

Brogaard, J., T. Hendershott, and R. Riordan, 2014. High frequency trading and price discovery. *The Review of Financial Studies* 27 (8), 2267-2306.

Brogaard, J., T. Hendershott, and R. Riordan, 2017. High frequency trading and the 2008 short sale ban. *Journal of Financial Economics* 124 (1), 22-42.

Brogaard, J., T. Hendershott, and R. Riordan, 2019. Price discovery without trading: evidence from limit orders. *Journal of Finance* 74 (4), 1621-1658.

Brolley, M. and K. Malinova, 2017. Informed Trading in a low-latency limit order market. Available at SSRN: https://ssrn.com/abstract=3095649.

Chaboud, A.P., B. Chiquoine, E. Hjalmarsson, and C. Vega, 2014. Rise of the machines: algorithmic trading in the foreign exchange market. *The Journal of Finance* 69 (5), 2045-2084.

Chakrabarty, B., T. Hendershott, S. Nawn, and R. Pascual, 2019. Order exposure in high frequency markets. Available at SSRN: https://ssrn.com/abstract= 3074049.

Chakrabarty, B., P. Moulton, and R. Pascual, 2017. Trading systems upgrades and short-sale bans: uncoupling the effects of technology and regulation. *Journal of Empirical Finance* 43, 74-90.

Chakrabarty, B., P. Moulton, and X. Wang, 2019. Attention effects in a high-frequency world. Available at SSRN: https://ssrn.com/abstract=2634621.

Chordia, T., R. Roll, and A. Subrahmanyam, 2005. Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics* 76 (2), 271-292.

Chordia, T. and A. Subrahmanyam, 2004. Order imbalance and individual stock returns: theory and evidence. *Journal of Financial Economics* 72 (3), 485-518.

Collin-Dufresne, P., and V. Fos, 2015. Do prices reveal the presence of informed trading? *The Journal of Finance* 70 (4), 1555-1582.

Dahlström, P., B. Hagströmer, and L. L. Nordén, 2018. Determinants of Limit Order Cancellations. Available at SSRN: https://ssrn.com/abstract=3012831.

Easley, D., N. Kiefer, M. O'Hara, and J.B. Paperman, 1996. Liquidity, information, and infrequently traded stocks. *The Journal of Finance* 51 (4), 1405-1436.

Easley, D., M. López de Prado, and M. O'Hara, 2012. Flow toxicity and liquidity in a high frequency world. *The Review of Financial Studies* 25 (5), 1457-1493.

Easley, D., M. López de Prado, and M. O'Hara, 2016. Discerning information from trade data. *Journal of Financial Economics* 120 (2), 269-285.

Easley, D., M. O'Hara, and L. Yang, 2016. Differential access to price information in financial markets. *Journal of Financial and Quantitative Analysis* 51 (4), 1071-1111.

Egginton, J. F., B. F. Van Ness, and R. A. Van Ness, 2016. Quote stuffing. *Financial Management* 45 (3), 583-608.

Fleming, M.J., Mizrach, B., and Nguyen, G., 2018. The microstructure of a U.S. Treasury ECN: the BrokerTec platform. *Journal of Financial Markets* 40, 2-22.

Foucault, T., R. Kozhan, and W.W. Tham, 2017. Toxic arbitrage. *The Review of Financial Studies* 30 (4), 1053-1094.

Glosten, L. and P. Milgrom, 1985. Bid, ask, and transaction prices in a specialist market with heterogeneously informed agents. *Journal of Financial Economics* 14 (1), 71-100.

Goettler, R.L., C.A. Parlour, and U. Rajan, 2009. Informed traders and limit order markets. *Journal of Financial Economics* 93 (1), 67-87.

Goyenko, R.Y., C.W. Holden, and C.A. Trzcinka, 2009. Do liquidity measures measure liquidity? *Journal of Financial Economics* 92(2), 153-181.

Hasbrouck, J., 1991. Measuring the information content of stock trades. *The Journal of Finance* 46 (1), 179-207.

Hasbrouck, J., 1996. Order characteristics and stock price evolution: an application to program trading. *Journal of Financial Economics* 41 (1), 129-149.

Hasbrouck, J., 2018. High-frequency quoting: short-term volatility in bids and offers. *Journal of Financial and Quantitative Analysis* 53 (2), 613-641.

Hasbrouck, J. and G. Saar, 2009. Technology and liquidity provision: the blurring of traditional definitions. *Journal of Financial Markets* 12 (2), 143-172.

Hendershott, T. and A. J. Menkveld, 2014. Price pressures. *Journal of Financial Economics* 114 (3), 405-423.

Hirschey, N., 2018. Do high-frequency traders anticipate buying and selling pressure? Available at SSRN: https://ssrn.com/abstract=2238516.

Ho, T.S.Y. and Stoll, H.R., 1983. The dynamics of dealer markets under competition. *The Journal of Finance* 38 (4), 1053-1074.

Hoffman, P., 2014. A dynamic limit order market with fast and slow traders. *Journal of Financial Economics* 113 (1), 156-169.

Johnson, B.H. and E.D. Watson, 2018. The explanatory power of order imbalance mesures. *The Quarterly Journal of Finance* 8 (1), 1850003-1-25.

Jovanovic, B., and A.J. Menkveld, 2016. Middlemen in limit order markets. Available at SSRN: https://ssrn.com/abstract=1624329.

Kim, S. and H.R. Stoll, 2014. Are trading imbalances indicative of private information? *Journal of Financial Markets* 20, 151-174.

Liu, W., 2009. Monitoring and limit order submission risk. *Journal of Financial Markets* 12 (1), 107-141.

Menkveld, A.J., 2013. High-frequency trading and the *new-market* makers, *Journal of Financial Markets* 16 (4), 712-740.

Menkveld, A.J., 2016. The economics of high-frequency trading: taking stock, *Annual Review of Financial Economics* 8, 1-24.

Menkveld, A.J. and M.A. Zoican, 2017. Need for speed? Exchange latency and liquidity. *The Review of Financial Studies* 30 (4), 1188-1228.

Nawn, S. and A. Banerjee, 2019. Do proprietary algorithmic traders withdraw liquidity during market stress? *Financial Management* 48 (2), 641-676.

Newey, W.K. and K. West, 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61 (4), 631-653.

O'Hara, M., 2015. High frequency market microstructure. *Journal of Financial Economics* 116 (2), 257-270.

O'Hara, M., C. Yao, and M. Ye, 2014. What's not there: odd lots and market data. *The Journal of Finance* 65 (5), 2199-2236.

Riordan, R. and A. Storkenmaier, 2012. Latency, liquidity and price discovery. *Journal of Financial Markets* 15 (4), 361-492.

Scholtus, M., D. van Dijk, and B. Frijns, 2014. Speed, algorithmic trading, and market quality around macroeconomics news announcements. *Journal of Banking and Finance* 38, 89-105.

SEC, 2010. Concept release on market structure (https://www.se.gov/rules/concept/2010/34-61358fr.pdf)

SEC, 2014. Equity market structure literature review Part II: high frequency trading (https://www.sec.gov/marketstructure/research/hft_lit_review_march_2014.pdf)

Stoll, H.R., 2000. Friction. *The Journal of Finance* 55 (4), 1479-1514.

Upson, J. and R.A. Van Ness, 2017. Algorithmic trading and market liquidity. *Journal of Financial Markets* 32, 28-48.

Van Kervel, V., 2015. Competition for order flow with fast and slow traders. *The Review of Financial Studies* 28 (7), 2049-2127.

Van Kervel, V. and A.J. Menkveld, 2019. High-frequency trading around large institutional orders. *The Journal of Finance* 74 (3), 1091-1137.

Weller, B., 2018. Does algorithmic trading reduce information acquisition? *The Review of Financial Studies* 31 (6), 2184-2226.

Yang, L. and H. Zhu, 2019. Back-running: seeking and hiding fundamental information in order flows. *The Review of Financial Studies*, forthcoming. https://doi.org/10.1093/rfs/hhz070.

TABLE I
Descriptive statistics

This table provides descriptive statistics for our sample, which consists of the 50 largest stocks listed at the NSE in 2015. We use order, trade, and quote data from April to July 2015. In Panel A, we report cross-sectional average statistics on market capitalization, trading activity, volatility and liquidity. Market capitalization is the market value of the company shares in billions of rupees in May 2015. “Volume” is the daily average accumulated traded volume in shares. “Trades” is the daily average accumulated number of trades. Volatility is the daily average high-low ratio (H/L-1). The relative bid-ask spread is averaged weighting by time and reported in basis points. In Panel B, we provide cross-sectional average daily statistics on the order flow composition for all traders together and for three subsets of traders: high-frequency traders (HFTs), agency algorithmic traders (AATs) and non-algorithmic traders (NATs). “Messages” is the sum of all submissions, revisions, and cancellations of orders. “New submissions” is the number or market and limit order submitted. “Cancellations & revisions” is the number of cancellations and revisions either of the limit price or the order size of standing limit orders. “MT/Trades” is the ratio of messages to trades. For the trader types, “MT/Trades” is the ratio of all messages to all trades initiated by the corresponding type of trader. “C&R/Trades” is the equivalent ratio of cancellations plus revisions to trades. Statistical tests compare the equality of the cross-sectional daily means across types of traders.

Panel A: Sample statistics	Mean	Min	Max
Market capitalization (billions of rupees)	1134955.80	195071.30	4871797.20
Volume (/10000)	304.30	5.39	1257.82
Trades	21815.25	3214.55	49271.94
Volatility (x100)	2.61	1.73	3.78
Relative bid-ask spread (bsp)	3.91	1.96	8.57
Depth at the best quotes (sh.) (/10000)	1155.89	52.45	5284.85
Price (Rupees)	949.92	73.40	3895.18

Panel B: Order flow statistics	All	HFTs	AATs	NATs
Messages	1287202.9	1074794.8	152138.5 ***	60269.6 ***
(%)		(83.5)	(11.8)	(4.7)
New submissions	104199.7	50345.9	30942.7 ***	22911.1 ***
(%)		(48.3)	(29.7)	(22.0)
Cancellations & revisions	1183003.2	1024449.0	121195.8 ***	37358.5 ***
(%)		(86.6)	(10.2)	(3.2)

Panel C: Common HFT proxies	All	HFTs	AATs	NATs
MT/Trades	68.67	310.41	22.51 ***	7.40 ***
(std.)	(29.8)	(159.4)	(11.6)	(2.2)
C&R/Trades	63.37	295.68	18.28 ***	4.81 ***
(std.)	(28.7)	(154.0)	(10.4)	(2.1)

***, **, * means statistically different than the HFT's statistic at the 1%, 5%, and 10% level respectively.

TABLE II
Message traffic composition

This table provides cross-sectional daily average statistics on the composition of the message traffic. Our sample consists of the 50 largest stocks listed at the NSE in 2015. We use order, trade, and quote data from April to July 2015. Panel A contains the distribution of messages between: (a) marketable order submissions, (b) non-marketable order submissions, and (c) cancellations and revisions (C&R) of standing limit orders. Panel B provides the same distribution but conditional on the level of the LOB. We distinguish between orders placed or standing (a) at or within the prevailing best quotes, (b) up to the 5th best prevailing quote, and (c) beyond the 5th best prevailing quote. In Panel C, we report the share of each type of trader in all messages placed conditional on the level of the LOB.

Panel A: Message traffic composition

	Submissions (%)		
Statistic	Marketable orders	Non-marketable orders	C&R (%)
Mean	3.38	9.81	86.82
Std. Dev.	1.11	1.88	2.68
Max	6.77	14.43	92.18
Min	1.49	5.84	79.85

Panel B: Message traffic composition and the LOB grid

	Marketable orders	Non-marketable orders	C&R (%)
At or within	24.15	19.56	56.29
Up to level 5	8.32	13.88	77.80
Beyond level 5	0.00	7.59	92.41

Panel C: Trader type and the LOB grid

	Total (%)	HFT (%)	AAT (%)	NAT (%)
At or within	13.48	48.86	35.74	15.41
Up to level 5	37.03	61.09	31.35	7.56
Beyond level 5	62.97	85.37	6.53	8.09

TABLE III
Correlation between NOF and OI

We provide cross-sectional average correlations between the net order flow (NOF) and the trade-initiator-based order imbalance (OI). The NOF and the OI are computed over regular intraday time bars of a fixed duration, from 1 second to 1800 seconds. We consider six versions of the NOF: NOF^l considers all types of orders and messages (submissions, revisions, and cancellations), while $NOF^l(L)$ considers messages involving non-marketable (non-aggressive) limit orders only. The sub-index l, with $l = \{1, 5, a\}$, indicates how many levels of the LOB are taken into account in computing the NOF metrics. In Panel A, l=1, meaning that we consider orders either placed or standing at or within the prevailing best quotes. In Panel B, l=5, meaning that we use orders either placed or standing up to the fifth best ask and bid quotes. In Panel C, l=a, meaning that we consider the whole LOB. The OI is derived exclusively from trades. "Corr." stands for correlation. Our sample consists of the 50 largest stocks listed at the NSE in 2015. We use order, trade, and quote data from April to July 2015. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

Panel A: At or within the prevailing best quotes		
Seconds	Corr(NOF^1,OI)	Corr(NOF^1(L),OI)
1	0.71 ***	0.13 ***
5	0.55 ***	0.12 ***
60	0.52 ***	0.10 ***
300	0.48 ***	-0.04
1800	0.43 ***	-0.19 ***
Panel B: Up to the five prevailing best quotes		
	Corr(NOF^5,OI)	Corr(NOF^5(L),OI)
1	0.54 ***	0.11 ***
5	0.38 ***	0.11 ***
60	0.42 ***	0.12 ***
300	0.41 ***	-0.02
1800	0.36 ***	-0.22 ***
Panel C: All messages		
	Corr(NOF,OI)	Corr(NOF(L),OI)
1	0.41 ***	0.01 **
5	0.25 ***	0.01 ***
60	0.29 ***	-0.07 ***
300	0.28 ***	-0.27 ***
1800	0.23 ***	-0.51 ***

TABLE IV
NOF-return relationship

This table provides the estimated coefficients of a pooled regression model on the relationship between returns and contemporaneous and lagged order flow imbalance over 1, 5, 60, and 300-second time bars. Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015. The dependent variable is the continuously-compound quote-midpoint return (in basis points), computed using the first and last best quotes posted within each time bar. The explanatory variable of interest is the order-flow-based order imbalance or net order flow (NOF). In Panel A, we compute the NOF using all orders that equate, hit, or improve the prevailing best quotes (NOF^1). In Panel B, we compute the NOF using all orders placed or standing within the five best ask and bid quotes of the LOB (NOF^5). Finally, in Panel C, we compute the NOF using the whole LOB. All NOF metrics are standardized per stock. To compute the NOF, we take into account all order submission, revisions, and cancellations. The model also includes lags of the dependent variable and dummies to control for unusual patterns during the initial and final trading hours of each session. The estimated coefficients of these latter variables are not reported. We only report the estimated coefficients of up to five lags of the variables of interest. The optimal number of lags is determined using the AIC criterion. The model is estimated by OLS with standard errors clustered by stock. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

	Time bar length (seconds)			
Coef.	1	5	60	300
Panel A: NOF^1				
NOF(t)	0.5208 ***	0.9612 ***	4.1644 ***	9.1864 ***
NOF(t-1)	0.0482 ***	0.1561 ***	0.0578	-0.6935 ***
NOF(t-2)	0.0118 ***	0.0189 ***	-0.2002 ***	-0.8112 ***
NOF(t-3)	-0.0032 **	-0.0169 ***	-0.1823 ***	-0.5285 ***
NOF(t-4)	-0.0112 ***	-0.0270 ***	-0.1626 ***	-0.4888 ***
NOF(t-5)	-0.0173 ***	-0.0295 ***	-0.1655 ***	-0.4745 ***
Adj-R^2	0.0906	0.1247	0.2652	0.3066
Panel B: NOF^5				
NOF(t)	0.3685 ***	0.7568 ***	4.0977 ***	9.5304 ***
NOF(t-1)	0.0666 ***	0.2177 ***	0.3803 ***	-0.1429
NOF(t-2)	0.0310 ***	0.0757 ***	-0.0883 ***	-0.4966 ***
NOF(t-3)	0.0144 ***	0.0270 ***	-0.0972 ***	-0.3349 ***
NOF(t-4)	0.0053 ***	0.0059 **	-0.1040 ***	-0.3057 ***
NOF(t-5)	-0.0021	-0.0047 **	-0.1051 ***	-0.3079 ***
Adj-R^2	0.0497	0.0838	0.2648	0.3473
Panel C: NOF				
NOF(t)	0.0818 ***	0.2772 ***	2.4528 ***	5.9809 ***
NOF(t-1)	0.0742 ***	0.2151 ***	0.6471 ***	0.7146 ***
NOF(t-2)	0.0407 ***	0.0960 ***	0.1395 ***	-0.0303
NOF(t-3)	0.0244 ***	0.0503 ***	0.0556 ***	-0.1029 **
NOF(t-4)	0.0156 ***	0.0282 ***	0.0215 **	-0.1150 ***
NOF(t-5)	0.0083 ***	0.0150 ***	-0.0191 **	-0.1731 ***
Adj-R^2	0.008	0.0175	0.0994	0.1409

TABLE V
NOF-return relationship by type of order

This table provides the estimated coefficients of a pooled regression model on the relationship between returns and contemporaneous and lagged order flow imbalance over 1, 5, 60, and 300-second time bars. Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015. The dependent variable is the continuously-compound quote-midpoint return (in basis points), computed using the first and last best quotes posted within each time bar. The explanatory variable of interest is the order-flow-based order imbalance or net order flow (NOF) that we decompose into a component due to non-marketable limit orders (NOF(L)) and a component due to market (and marketable limit) orders (OI). In Panel A, we compute the NOF(L) using non-marketable limit orders that equate or improve the prevailing best quotes ($NOF^1(L)$). In Panel B, we compute the NOF(L) using all non-marketable limit orders placed or standing within the five best ask and bid quotes of the LOB ($NOF^5(L)$). In computing the NOF(L), we account for all types of messages (submission, revisions, and cancellations of orders). To compute the OI, we use trades. All NOF and OI metrics are standardized per stock. The model also includes lags of the dependent variable and dummies to control for unusual patterns during the initial and final trading hours of each session. The estimated coefficients of these latter variables are not reported. We only report the coefficients of up to five lags of the variables of interest. The optimal number of lags is determined using the AIC criterion. The model is estimated by OLS with standard errors clustered by stock. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

	Time bar length (seconds)			
Coef.	1	5	60	300
Panel A: $NOF^1(L)$ and OI				
NOF(L)(b)	0.41 ***	0.73 ***	3.08 ***	7.11 ***
NOF(L)(b-1)	0.03 ***	0.11 ***	0.12 ***	-0.40 ***
NOF(L)(b-2)	0.00 **	0.01 ***	-0.13 ***	-0.60 ***
NOF(L)(b-3)	0.00 ***	-0.01 ***	-0.12 ***	-0.40 ***
NOF(L)(b-4)	-0.01 ***	-0.02 ***	-0.12 ***	-0.40 ***
NOF(L)(b-5)	-0.01 ***	-0.02 ***	-0.12 ***	-0.38 ***
OI(b)	0.45 ***	0.73 ***	2.85 ***	6.55 ***
OI(b-1)	0.02 ***	0.06 ***	-0.18 ***	-1.00 ***
OI(b-2)	0.00 ***	-0.02 ***	-0.21 ***	-0.62 ***
OI(b-3)	-0.01 ***	-0.03 ***	-0.18 ***	-0.36 ***
OI(b-4)	-0.02 ***	-0.04 ***	-0.14 ***	-0.29 ***
OI(b-5)	-0.02 ***	-0.04 ***	-0.12 ***	-0.30 ***
Adj-R^2	0.13	0.16	0.29	0.33
Panel B: $NOF^5(L)$ and OI				
NOF(L)(b)	0.25 ***	0.54 ***	3.13 ***	7.63 ***
NOF(L)(b-1)	0.05 ***	0.17 ***	0.34 ***	-0.03
NOF(L)(b-2)	0.02 ***	0.06 ***	-0.04 **	-0.38 ***
NOF(L)(b-3)	0.01 ***	0.02 ***	-0.05 ***	-0.29 ***
NOF(L)(b-4)	0.00 ***	0.01 ***	-0.07 ***	-0.27 ***
NOF(L)(b-5)	0.00	0.00	-0.07 ***	-0.24 ***
OI(b)	0.47 ***	0.75 ***	2.75 ***	6.21 ***
OI(b-1)	0.02 ***	0.07 ***	-0.13 ***	-0.83 ***
OI(b-2)	0.00 *	-0.01 ***	-0.20 ***	-0.52 ***
OI(b-3)	-0.01 ***	-0.03 ***	-0.17 ***	-0.30 ***
OI(b-4)	-0.02 ***	-0.03 ***	-0.13 ***	-0.22 ***
OI(b-5)	-0.02 ***	-0.04 ***	-0.12 ***	-0.22 ***
Adj-R^2	0.10	0.13	0.30	0.37

TABLE VI
Permanent quote-midpoint impact

We report the cross-sectional median permanent quote midpoint impact of a shock to the net order flow (NOF). The NOF takes into account order submissions, revisions, and cancellations. In Panel A, we compute the NOF using the order flow that equates, hits, or improves the prevailing best quotes (NOF^1). In Panel B, we compute the NOF using all messages placing orders or modifying orders already standing within the five best ask and bid quotes of the LOB (NOF^5). Finally, in Panel C, we compute the NOF using all the messages. For each stock, we estimate a bivariate SVAR model for continuously compound quote midpoint returns, expressed in basis points, and the NOF. We impose one-way contemporaneous causality from the NOF to returns. With the estimated model, we obtain the structural IRF of the return to a one-percent-point shock to the NOF. Both returns and NOF are computed over regular time intervals of 1, 5, 60, and 300 seconds. All NOF metrics are standardized per stock. The table reports the cross-sectional median permanent quote midpoint impact both in basis points (IRF) and in relative terms to the standard deviation of the quote midpoint returns (IRF/σ(R)), the corresponding cross-sectional interquartile range, and the number of stocks (out of 50) for which we get a positive and statistically significant impact. Bold format means statistically significant at least at the 5% level. ***, **, * means statistically different than the preceding bar-size's statistic (according to Wilcoxon's rank-sum test). Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015.

	Time bar length (seconds)			
Panel A: Best quotes (NOF^1)	1	5	60	300
Price impact (IRF)	**0.75**	**1.35** ***	**4.70** ***	**9.76** ***
(IQR)	(0.26)	(0.20)	(1.46)	(3.72)
IRF/σ(R)	**0.41**	**0.47** ***	**0.52** ***	**0.54**
(IQR)	(0.05)	(0.07)	(0.11)	(0.15)
Sig. IRF>0 (# stocks)	50	50	50	50
Panel B: Five best quotes (NOF^5)				
Price impact (IRF)	**0.59**	**1.13** ***	**4.64** ***	**10.57** ***
(IQR)	(0.17)	(0.25)	(1.39)	(3.48)
IRF/σ(R)	**0.33**	**0.39** ***	**0.52** ***	**0.57** **
(IQR)	(0.06)	(0.09)	(0.13)	(0.10)
Sig. IRF>0 (# stocks)	50	50	50	50
Panel C: All messages (NOF)				
Price impact (IRF)	**0.26**	**0.62** ***	**3.10** ***	**6.90** ***
(IQR)	(0.17)	(0.28)	(1.25)	(3.20)
IRF/σ(R)	**0.15**	**0.22** ***	**0.35** ***	**0.39**
(IQR)	(0.05)	(0.08)	(0.14)	(0.12)
Sig. IRF>0 (# stocks)	50	50	50	50

TABLE VII
Permanent quote-midpoint impact by type of order

We report the cross-sectional median permanent quote midpoint impact of a shock to the net order flow (NOF). We decompose the NOF into two pieces: NOF(L) considers order submissions, revisions, and cancellations of non-marketable limit orders; OI is the trade-initiator-based order imbalance. In computing the NOF, we use all messages placing non-marketable limit orders or modifying limit orders standing at or within the best ask and bid quotes of the LOB. All NOF and OI metrics are standardized per stock. For each stock, we estimate a SVAR model for the continuously compound midpoint returns (in basis points), NOF(L), and OI. We impose one-way contemporaneous causality running from order flow to returns. Then, we obtain the structural IRF of the return to a one-percent-point shock to each component of the NOF. All variables involved are computed over regular intervals of 1, 5, 60, and 300 seconds. We report the cross-sectional median permanent quote midpoint impact both in basis points (IRF) and in relative terms to the standard deviation of the quote midpoint returns (IRF/σ(R)), the corresponding cross-sectional interquartile range, and the number of stocks (out of 50) for which we get a positive and statistically significant impact. Bold format means statistically significant at least at the 5% level. ***. **, * means a given statistic in the NOF(L) panel is statistically different than the corresponding statistic in the OI panel at the 1%, 5%, and 10% level of statistical significance (according to Wincoxon's ranksum tests). Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015.

	Time bar length (seconds)			
	1	5	60	300
	Shock to NOF(L)			
Price impact (IRF)	**0.51** ***	**0.99** ***	**3.69** ***	**8.34** ***
(IQR)	(0.15)	(0.19)	(1.05)	(3.00)
IRF/σ(R)	**0.29** ***	**0.35** ***	**0.42** ***	**0.46** ***
(IQR)	(0.04)	(0.04)	(0.10)	(0.14)
Sig. IRF>0 (# stocks)	50	50	50	50
	Shock to OI			
Price impact (IRF)	**0.41**	**0.90**	**2.70**	**6.05**
(IQR)	(0.09)	(0.13)	(0.58)	(2.17)
IRF/σ(R)	**0.23**	**0.32**	**0.30**	**0.33**
(IQR)	(0.06)	(0.07)	(0.05)	(0.08)
Sig. IRF>0 (# stocks)	50	50	50	50

TABLE VIII
Quote-midpoint impact and trader type

We report cross-sectional median quote midpoint impact estimates of a shock to the order flow imbalance for three types of traders: proprietary ATs (HFTs), agency ATs (AATs), and non-ATs (NATs). The estimated impact (in basis points) is obtained per stock using a SVAR model for continuously compound quote midpoint returns (*r*) and trader-type-specific measures of order flow imbalance. We impose one-way contemporaneous causality from the order flow to *r*. In Panel A, we use the net order flow (NOF) per trader type (the SVAR model has 4 equations). In Panel B, we decompose the NOF per trader type into NOF(L), based on the non-marketable limit orders, and OI, based on market or marketable limit orders (the SVAR model has 7 equations). In both cases, we compute the NOF using all messages placed at or within the prevailing best quotes of the LOB (NOF^1). All NOF and OI metrics are standardized per stock. From the estimated SVAR model, we obtain the structural IRF of the return to a one-percent-point shock to the NOF. Both returns and NOF are computed over regular time intervals of 1, 5, 60, and 300 seconds. Estimated impacts are reported both in basis points (IRF) and in relative terms to the standard deviation of the quote midpoint returns (IRF/σ(R)). We also report the corresponding interquartile range, and the number of stocks (out of 50) for which we get a positive and statistically significant impact. Bold format means statistically significant at least at the 5% level. ***, **, * means statistically different at the 1%, 5%, and 10% level than the corresponding statistic for the HFTs. Our sample consists on the constituents of the NIFTY-50 from April to July 2015.

Panel A: NOF

	Trader type											
	HFTs				AATs				NATs			
Bar length (seconds)	1	5	60	300	1	5	60	300	1	5	60	300
Price impact (IRF)	**0.65**	**0.87**	**1.13**	**3.27**	**0.44** ***	**0.76** ***	**3.91** ***	**9.67** ***	**0.41** ***	**0.70** ***	**2.45** ***	**5.23** ***
(IQR)	(0.23)	(0.17)	(0.88)	(2.22)	(0.16)	(0.14)	(1.31)	(4.03)	(0.16)	(0.17)	(0.52)	(1.51)
IRF/σ(R)	**0.35**	**0.30**	**0.13**	**0.18**	**0.24** ***	**0.27** ***	**0.43** ***	**0.50** ***	**0.21** ***	**0.24** ***	**0.28** ***	**0.29** ***
(IQR)	(0.04)	(0.05)	(0.08)	(0.09)	(0.04)	(0.04)	(0.14)	(0.17)	(0.03)	(0.04)	(0.04)	(0.09)
Sig. IRF>0 (# stocks)	50	50	45	30	50	50	50	50	50	50	50	45
Panel B: NOF^5(L) and OI												
(a) Shock to NOF^5 (L)												
Price impact (IRF)	**0.28**	**0.40**	**1.07**	2.48	**0.27**	**0.54** ***	**2.85** ***	**7.97** ***	**0.32** **	**0.63** ***	**1.94** ***	**4.73** ***
(IQR)	(0.12)	(0.11)	(0.75)	(1.95)	(0.08)	(0.12)	(1.11)	(3.90)	(0.17)	(0.16)	(0.51)	(1.58)
IRF/σ(R)	**0.16**	**0.15**	**0.12**	0.14	**0.16**	**0.19** ***	**0.31** ***	**0.41** ***	**0.18** ***	**0.22** ***	**0.22** ***	**0.26** ***
(IQR)	(0.02)	(0.03)	(0.08)	(0.10)	(0.04)	(0.04)	(0.10)	(0.21)	(0.02)	(0.03)	(0.04)	(0.07)
Sig. IRF>0 (# stocks)	50	50	39	13	50	50	50	47	50	50	50	44
(b) Shock to OI												
Price impact (IRF)	**0.55**	**0.80**	0.63	0.16	**0.41** ***	**0.65** ***	**2.15** ***	**5.57** ***	**0.32** **	**0.44** ***	**1.21** ***	2.83 ***
(IQR)	(0.16)	(0.15)	(0.30)	(1.06)	(0.19)	(0.20)	(0.47)	(1.84)	(0.13)	(0.14)	(0.32)	(1.03)
IRF/σ(R)	**0.31**	**0.29**	0.07	0.01	**0.23** ***	**0.23** ***	**0.25** ***	**0.30** ***	**0.17** ***	**0.15** ***	**0.14** ***	0.16 ***
(IQR)	(0.06)	(0.06)	(0.04)	(0.05)	(0.04)	(0.02)	(0.05)	(0.06)	(0.03)	(0.02)	(0.03)	(0.06)
Sig. IRF>0 (# stocks)	50	50	23	0	50	50	50	43	50	50	50	13

TABLE IX
Quote-midpoint impact and message type

We estimate the cross-sectional median quote midpoint impact of a shock to the order flow imbalance controlling for the type of message: (1) market orders and marketable limit orders (Panel A); (2) submissions of non-marketable limit orders (Panel B), and (3) revisions and cancellations (monitoring) (Panel C). OI stands for the share imbalance of marketable order submissions, NOF(S) stands for the share imbalance of non-marketable limit order submissions, and NOF(M) stands for the imbalance of revisions and cancellations of standing limit orders (monitoring activity). All the imbalance metrics are expressed in relative terms to the total volume of the orders involved, and standardized per stock. For each stock, we estimate a SVAR model for continuously-compound quote midpoint return and the imbalance metrics. In computing the NOF components, we consider messages placed either at or within the prevailing best quotes of the LOB (NOF^1). The variables are defined over regular time intervals of 1, 5, 60, and 300 seconds. The SVAR assumes one-way contemporaneous causality from imbalances to returns. We report the cross-sectional median permanent quote midpoint impact both in basis points (IRF) and in relative terms to the standard deviation of the quote midpoint returns (IRF/σ(R)), the corresponding cross-sectional interquartile range, and the number of stocks (out of 50) for which we get a positive and statistically significant impact. Bold format means statistically significant at least at the 5% level. ***, **, * means that a given statistic in the NOF(S) or NOF(M) panel is statistically different than the corresponding statistic in the OI panel at the 1%, 5%, and 10% level of statistical significance (according to Wincoxon's ranksum tests). Our sample consists on the constituents of the NIFTY-50 from April to July 2015.

Panel A: OI	Time bar length (seconds)			
	1	5	60	300
Price impact (IRF)	**0.60**	**0.94**	**2.76**	**6.72**
(IQR)	(0.18)	(0.15)	(0.65)	(1.93)
IRF/σ(R)	**0.32**	**0.33**	**0.32**	**0.35**
(IQR)	(0.05)	(0.07)	(0.05)	(0.09)
Sig. IRF>0 (# stocks)	50	50	50	50
Panel B: NOF(S)				
Price impact (IRF)	**0.54**	**0.94**	**3.35** ***	**8.03** ***
(IQR)	(0.12)	(0.15)	(1.05)	(3.14)
IRF/σ(R)	**0.31** *	**0.34**	**0.39** ***	**0.43** ***
(IQR)	(0.06)	(0.05)	(0.07)	(0.15)
Sig. IRF>0 (# stocks)	50	50	50	50
Panel C: NOF(M)				
Price impact (IRF)	**0.40** ***	**0.57** ***	**1.95** ***	**5.25** ***
(IQR)	(0.12)	(0.16)	(0.90)	(2.89)
IRF/σ(R)	**0.21** ***	**0.20** ***	**0.23** ***	**0.28** ***
(IQR)	(0.03)	(0.03)	(0.12)	(0.16)
Sig. IRF>0 (# stocks)	50	50	50	40

TABLE X
Order flow toxicity: NOF

This table provides aggregated statistics on the estimation of the dynamic regression model in equation [15] on the relationship between illiquidity and the absolute order flow imbalance over regular intraday time intervals (in this case, 1-second and 5-second time bars). As dependent variable, we use either the weighted by time relative quoted spread (RQS), the weighted by volume relative effective spread (RES), or the illiquidity ratio of Amihud (2002) (AIR). RQS and RES are in basis points. AIR is multiplied by one million. As for the order flow imbalance, we use the NOF^1 (see "first specification"); this order-flow-based metric takes into account order submissions, revisions, and cancellations affecting orders placed or standing at or within the prevailing best quotes (l=1). We also consider the decomposition of the NOF into components due to market (and marketable limit) orders, commonly known as "order imbalance" (*OI*), and non-marketable limit orders (NOF^1(L)) (see "second specification"). All NOF and OI metrics are standardized per stock. As controls, the model includes the first lag of the dependent variable, dummies for the initial and final trading hours of each trading session, the logarithm of the volume in shares, and the standard deviation of the quote midpoint changes (as a proxy for volatility). All the non-deterministic explanatory variables are lagged one period. Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015. The model is estimated per stock by Newey and West (1994). We report the cross-sectional average coefficient for the variables of interest, aggregated t-statistics based on Chordia et al. (2005), and the percentage of stocks for which the coefficient of interest is statistically significant at the 1% and 5% level. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

	NOF^1(b)		NOF^1(L)(b)		OI(b)	
Dependent variable	1s bars	5s bars	1s bars	5s bars	1s bars	5s bars
(a) RQS (b+1)						
coef*100	-7.29 ***	-4.32 ***	-13.77 ***	-7.60 ***	-0.47	0.35
t-test	-14.39	-6.59	-24.83	-12.55	-1.26	1.17
Sig.>0 at 1% (5%) (# stocks)	0 (0)	3 (5)	0 (0)	0 (0)	9 (11)	18 (22)
(b) RES (b+1)						
coef*100	-26.61 ***	-17.00 ***	-31.83 ***	-20.52 ***	-5.73 ***	-4.08 ***
t-test	-40.33	-21.74	-43.59	-26.21	-5.82	-3.49
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	0 (0)	0 (0)	4 (5)	4 (5)
(c) AIR (b+1)						
coef*100	-26.45 ***	-17.88 ***	-36.72 ***	-25.85 ***	14.65 ***	17.68 ***
t-test	-15.51	-12.67	-20.68	-19.42	7.47	11.93
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	0 (0)	0 (0)	39 (41)	49 (49)

TABLE XI
Order flow toxicity: NOF and trader types

This table provides aggregated statistics on the estimation of the dynamic regression model in equation [15] on the relationship between illiquidity and the absolute order flow imbalance. We control for trader type. In particular, we distinguish between proprietary ATs (HFTs), agency ATs (AATs), and non-ATs (NATs). As dependent variable, we use either the weighted by time relative quoted spread (RQS), the weighted by volume relative effective spread (RES), or the illiquidity ratio of Amihud (2002) (AIR). RQS and RES are in basis points. AIR is multiplied by one million. As for the order flow imbalance, we use the NOF^1 of each type of trader. The NOF^1 is an order-flow-based metric takes into account order submissions, revisions, and cancellations affecting orders placed or standing at or within the prevailing best quotes (l=1). All NOF metrics are standardized per stock. Control variables are the same used in Table X. All the non-deterministic explanatory variables are lagged one period. Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015. The model is estimated per stock by Newey and West (1994). We report the cross-sectional average coefficient for the variables of interest, aggregated t-statistics based on Chordia et al. (2005), and the percentage of stocks for which the coefficient of interest is statistically significant at the 1% and 5% level. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

	HFTs' NOF^1		AATs' NOF^1		NATs' NOF^1	
Dependent variable	1s bars	5s bars	1s bars	5s bars	1s bars	5s bars
(a) RQS (b+1)						
coef*100	7.44 ***	5.24 ***	-3.04 ***	-3.43 ***	-8.58 ***	-4.28 ***
t-test	25.79	13.75	-7.28	-5.68	-19.05	-5.84
Sig.>0 at 1% (5%) (# stocks)	49 (49)	45 (46)	1 (1)	1 (2)	0 (0)	2 (3)
(b) RES (b+1)						
coef*100	13.23 ***	8.35 ***	-7.13 ***	-10.26 ***	-13.33 ***	-9.98 ***
t-test	34.04	15.88	-12.84	-13.14	-21.47	-11.01
Sig.>0 at 1% (5%) (# stocks)	49 (49)	44 (44)	3 (3)	0 (0)	0 (0)	0 (0)
(c) AIR (b+1)						
coef*100	24.10 ***	14.72 ***	-8.21 ***	-10.93 ***	-12.14 ***	-2.09
t-test	20.90	12.93	-5.89	-7.75	-5.66	0.58
Sig.>0 at 1% (5%) (# stocks)	49 (49)	46 (47)	2 (2)	0 (0)	2 (2)	14 (15)

TABLE XII
Order flow toxicity: NOF(L), OI, and trader types

This table provides aggregated statistics on the estimation of the dynamic regression model in equation [15] on the relationship between illiquidity and the absolute both marketable and non-marketable order flow imbalance. We control for trader type. In particular, we distinguish between proprietary ATs (HFTs) (Panel A), agency ATs (AATs) (Panel B), and non-ATs (NATs) (Panel C). Dependent and control variables are the same as in Tables X, and XI. As for the order flow imbalance, we use the NOF^1 of each type of trader. The NOF^1 is an order-flow-based metric takes into account order submissions, revisions, and cancellations affecting orders placed or standing at or within the prevailing best quotes (l=1). We decompose each trader types' NOF^1 into components due to market (and marketable limit) orders, commonly known as "order imbalance" (*OI*), and non-marketable limit orders ($NOF^1(L)$). While the OI is computed from trades, the NOF(L) is computed from submissions, revisions, and cancellations of limit orders. All NOF metrics are standardized per stock. All the non-deterministic explanatory variables are lagged one period. Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015. The model is estimated per stock by Newey and West (1994). We report the cross-sectional average coefficient for the variables of interest, aggregated t-statistics based on Chordia et al. (2005), and the percentage of stocks for which the coefficient of interest is statistically significant at the 1% and 5% level. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

	NOF^1(L)(b)		OI(b)	
Panel A: HFTs	1s bars	5s bars	1s bars	5s bars
(a) RQS (b+1)				
coef*100	4.57 ***	3.83 ***	9.18 ***	10.18 ***
t-test	16.97	10.80	29.27	24.11
Sig.>0 at 1% (5%) (# stocks)	46 (46)	41 (41)	50 (50)	50 (50)
(b) RES (b+1)				
coef*100	8.51 ***	6.57 ***	19.64 ***	20.11 ***
t-test	22.63	13.20	46.03	34.74
Sig.>0 at 1% (5%) (# stocks)	44 (44)	41 (42)	50 (50)	50 (50)
(c) AIR (b+1)				
coef*100	19.48 ***	14.19 ***	19.13 ***	16.01 ***
t-test	18.43	13.83	12.26	15.08
Sig.>0 at 1% (5%) (# stocks)	44 (46)	46 (47)	49 (50)	50 (50)
Panel B: AATs				
(a) RQS (b+1)				
coef*100	-7.60 ***	-5.85 ***	4.18 ***	2.24 ***
t-test	-19.75	-10.71	8.96	3.17
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	49 (49)	26 (33)
(b) RES (b+1)				
coef*100	-11.55 ***	-11.50 ***	5.83 ***	4.50 ***
t-test	-22.19	-15.89	8.69	4.61
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	46 (47)	31 (34)
(c) AIR (b+1)				
coef*100	-12.82 ***	-10.63 ***	6.31	2.21 *
t-test	-11.80	-10.06	1.34	1.70
Sig.>0 at 1% (5%) (# stocks)	2 (2)	2 (2)	18 (21)	20 (23)
Panel C: NATs				
(a) RQS (b+1)				
coef*100	-13.89 ***	-12.20 ***	1.60 ***	0.17
t-test	-40.27	-23.02	3.56	0.16
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	32 (36)	8 (8)
(b) RES (b+1)				
coef*100	-9.24 ***	-8.46 ***	-8.93 ***	-8.99 ***
t-test	-18.93	-11.34	-14.04	-10.55
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	0 (0)	0 (0)
(c) AIR (b+1)				
coef*100	-25.30 ***	-18.60 ***	4.63	5.83 ***
t-test	-19.94	-13.62	0.10	4.34
Sig.>0 at 1% (5%) (# stocks)	1 (1)	0 (0)	11 (13)	26 (33)

TABLE XIII
Order flow toxicity: message type

This table provides aggregated statistics on the estimation of the dynamic regression model in equation [15] on the relationship between illiquidity and the absolute both marketable and non-marketable order flow imbalance. We split non-marketable order flow into submissions (S) and revisions and cancellations (monitoring or M). Additionally, we distinguish between proprietary ATs (HFTs) (Panel A), agency ATs (AATs) (Panel B), and non-ATs (NATs) (Panel C). Dependent and control variables are the same as in Tables X, XI, and XII. As for the order flow imbalance, we use the NOF^1 of each type of trader, which is an order-flow-based metric takes into account order submissions, revisions, and cancellations affecting orders placed or standing at or within the prevailing best quotes (l=1). We decompose each trader types' NOF^1 into components due to market (and marketable limit), commonly known as "order imbalance" (*OI*), and non-marketable limit orders (NOF^1(L)). Furthermore, we split the later into a component due to new submissions (NOF^1(S)) and a component due to monitoring of standing orders (NOF^1(M)). All the NOF and OI metrics are standardized per stock. All the non-deterministic explanatory variables are lagged one period. Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015. The model is estimated per stock by Newey and West (1994). We report the cross-sectional average coefficient for the variables of interest, aggregated t-statistics based on Chordia et al. (2005), and the percentage of stocks for which the coefficient of interest is statistically significant at the 1% and 5% level. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

	NOF(S)		NOF(M)		OI	
Dependent variable	1s bars	5s bars	1s bars	5s bars	1s bars	5s bars
Panel A: HFTs						
(a) RQS (b+1)						
coef*100	4.10 ***	3.83 ***	6.33 ***	4.72 ***	8.52 ***	9.72 ***
t-test	14.77	9.57	21.12	11.56	26.66	22.57
Sig.>0 at 1% (5%) (# stocks)	47(48)	44 (46)	50 (50)	47 (48)	50 (50)	50 (50)
(b) RES (b+1)						
coef*100	5.51 ***	6.97 ***	12.33 ***	7.88 ***	18.24 ***	19.02 ***
t-test	14.67	12.92	30.12	14.04	42.23	32.22
Sig.>0 at 1% (5%) (# stocks)	43 (43)	42 (44)	50 (50)	48 (48)	50 (50)	50 (50)
(c) AIR (b+1)						
coef*100	19.40 ***	18.68 ***	19.54 ***	11.81 ***	16.21 ***	13.28 ***
t-test	16.34	16.79	19.29	11.68	9.77	12.54
Sig.>0 at 1% (5%) (# stocks)	48 (49)	47 (48)	48 (48)	47 (48)	46 (46)	50 (50)
Panel B: AATs						
(a) RQS (b+1)						
coef*100	-5.23 ***	-4.52 ***	0.87	-1.69 ***	4.14 ***	2.39 ***
t-test	-16.42	-8.26	1.48	-4.46	8.96	3.40
Sig.>0 at 1% (5%) (# stocks)	3 (4)	2 (2)	24 (30)	3 (5)	49 (49)	29 (35)
(b) RES (b+1)						
coef*100	-6.20 ***	-5.68 ***	4.32 ***	-0.88 **	5.35 ***	4.52 ***
t-test	-14.51	-7.63	9.17	-2.36	8.03	4.62
Sig.>0 at 1% (5%) (# stocks)	6 (6)	2 (2)	42 (44)	6 (10)	44 (47)	31 (35)
(c) AIR (b+1)						
coef*100	-6.45 ***	-3.07 ***	3.91	0.25 *	5.76	1.85
t-test	-8.47	-4.34	0.93	-1.84	1.16	1.64
Sig.>0 at 1% (5%) (# stocks)	4 (4)	5 (5)	17 (18)	7 (7)	18 (20)	19 (23)
Panel C: NATs						
(a) RQS (b+1)						
coef*100	-13.01 ***	-12.02 ***	-2.08 ***	-2.93 ***	1.81 ***	0.24
t-test	-38.23	-23.18	-9.43	-8.87	3.90	0.23
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	1 (1)	0 (1)	33 (36)	7 (8)
(b) RES (b+1)						
coef*100	-8.12 ***	-8.26 ***	0.32	0.32	-8.61 ***	-8.84 ***
t-test	-16.91	-11.34	0.84	0.74	-13.70	-10.40
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	16 (19)	17 (19)	0 (0)	0 (0)
(c) AIR (b+1)						
coef*100	-22.26 ***	-16.28 ***	-3.70 ***	-5.38 ***	4.93	5.99 ***
t-test	-18.34	-13.87	-5.70	-8.19	0.20	4.40
Sig.>0 at 1% (5%) (# stocks)	1 (1)	0 (0)	1 (2)	0 (1)	12 (12)	29 (34)

Appendix

TABLE AI
Quote-midpoint impact by type of order (case *l*=5)

We report the cross-sectional median permanent quote midpoint impact of a shock to the net order flow (NOF). We decompose the NOF into two pieces: NOF(L) considers order submissions, revisions, and cancellations of non-marketable limit orders; OI is the trade-initiator-based order imbalance. In computing the NOF, we use all messages placing non-marketable limit orders or modifying limit orders standing at or within the prevailing five best ask and bid quotes of the LOB. All NOF and OI metrics are standardized per stock. For each stock, we estimate a SVAR model for the continuously compound midpoint returns (in basis points), NOF(L), and OI. We impose one-way contemporaneous causality running from order flow to returns. Then, we obtain the structural IRF of the return to a one-percent-point shock to each component of the NOF. All variables involved are computed over regular intervals of 1, 5, 60, and 300 seconds. We report the cross-sectional median permanent quote midpoint impact both in basis points (IRF) and in relative terms to the standard deviation of the quote midpoint returns (IRF/σ(R)), the corresponding cross-sectional interquartile range, and the number of stocks (out of 50) for which we get a positive and statistically significant impact. Bold format means statistically significant at least at the 5% level. ***. **, * means a given statistic in the NOF(L) panel is statistically different than the corresponding statistic in the OI panel at the 1%, 5%, and 10% level of statistical significance (according to Wincoxon's ranksum tests). Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015.

	Time bar length (seconds)			
	1	5	60	300
	Shock to NOF(L)			
Price impact (IRF)	**0.42**	**0.88**	**3.62** ***	**8.74** ***
(IQR)	(0.11)	(0.19)	(1.20)	(2.72)
IRF/σ(R)	**0.23**	**0.31**	**0.42** ***	**0.48** ***
(IQR)	(0.07)	(0.07)	(0.11)	(0.12)
Sig. IRF>0 (# stocks)	50	50	50	50
	Shock to OI			
Price impact (IRF)	**0.42**	**0.88**	**2.53**	**5.55**
(IQR)	(0.10)	(0.18)	(0.56)	(1.70)
IRF/σ(R)	**0.23**	**0.32**	**0.29**	**0.31**
(IQR)	(0.05)	(0.06)	(0.04)	(0.06)
Sig. IRF>0 (# stocks)	50	50	50	49

TABLE AII
Quote-midpoint impact by trader type (case *l*=5)

We report cross-sectional median quote midpoint impact estimates of a shock to the order flow imbalance for three types of traders: proprietary ATs (HFTs), agency ATs (AATs), and non-ATs (NATs). The estimated impact (in basis points) is obtained per stock using a SVAR model for continuously compound quote midpoint returns (r) and trader-type-specific measures of order flow imbalance. We impose one-way contemporaneous causality from the order flow to r. In Panel A, we use the net order flow (NOF) per trader type (the SVAR model has 4 equations). In Panel B, we decompose the NOF per trader type into NOF(L), based on the non-marketable limit orders, and OI, based on market or marketable limit orders (the SVAR model has 7 equations). In both cases, we compute the NOF using all messages placed at or within the prevailing five best ask and bid quotes of the LOB (NOF^5). From the estimated SVAR model, we obtain the structural IRF of the return to a one-percent-point shock to the NOF. Both returns and NOF are computed over regular time intervals of 1, 5, 60, and 300 seconds. Estimated impacts are reported both in basis points (IRF) and in relative terms to the standard deviation of the quote midpoint returns (IRF/σ(R)). We also report the corresponding interquartile range, and the number of stocks (out of 50) for which we get a positive and statistically significant impact. Bold format means statistically significant at least at the 5% level. ***, **, * means statistically different at the 1%, 5%, and 10% level than the corresponding statistic for the HFTs. Our sample consists on the constituents of the NIFTY-50 from April to July 2015.

Panel A: NOF

	Trader type											
	HFTs				AATs				NATs			
Bar length (seconds)	1	5	60	300	1	5	60	300	1	5	60	300
Price impact (IRF)	**0.52**	**0.75**	**1.58**	**3.27**	**0.29** ***	**0.50** ***	**3.32** ***	**9.67** ***	**0.40** ***	**0.77**	**2.35** ***	**5.23** ***
(IQR)	(0.16)	(0.16)	(0.71)	(2.22)	(0.10)	(0.12)	(1.40)	(4.03)	(0.21)	(0.19)	(0.43)	(1.51)
IRF/σ(R)	**0.29**	**0.25**	**0.19**	**0.18**	**0.16** ***	**0.18** ***	**0.37** ***	**0.50** ***	**0.22** ***	**0.28** ***	**0.28** ***	**0.29** ***
(IQR)	(0.06)	(0.05)	(0.08)	(0.09)	(0.03)	(0.06)	(0.13)	(0.17)	(0.03)	(0.04)	(0.04)	(0.09)
Sig. IRF>0 (# stocks)	50	50	50	30	50	50	50	50	50	50	50	45

Panel B: NOF^5(L) and OI

	1	5	60	300	1	5	60	300	1	5	60	300
(a) Shock to NOF^5 (L)												
Price impact (IRF)	**0.26**	**0.40**	**1.45**	**3.79**	**0.15** ***	**0.34** ***	**2.43** ***	**7.03** ***	**0.31** ***	**0.66** ***	**2.02** ***	**4.26** **
(IQR)	(0.07)	(0.10)	(0.78)	(1.91)	(0.05)	(0.12)	(0.88)	(3.40)	(0.16)	(0.16)	(0.32)	(1.53)
IRF/σ(R)	**0.14**	**0.14**	**0.16**	**0.20**	**0.09** ***	**0.12** ***	**0.27** ***	**0.35** ***	**0.18** ***	**0.23** ***	**0.24** ***	**0.24** **
(IQR)	(0.06)	(0.04)	(0.08)	(0.09)	(0.04)	(0.05)	(0.10)	(0.16)	(0.02)	(0.04)	(0.03)	(0.09)
Sig. IRF>0 (# stocks)	48	50	49	29	46	50	50	49	49	50	50	39
(b) Shock to OI												
Price impact (IRF)	**0.56**	**0.80**	0.48	-0.34	**0.40** ***	**0.66** ***	**2.24** ***	**5.35** ***	**0.32** ***	**0.43** ***	**1.08** ***	2.57 ***
(IQR)	(0.18)	(0.15)	(0.37)	(1.18)	(0.18)	(0.23)	(0.45)	(1.86)	(0.15)	(0.16)	(0.37)	(1.08)
IRF/σ(R)	**0.31**	**0.28**	0.05	-0.02	**0.23** ***	**0.23** ***	**0.26** ***	**0.30** ***	**0.17** ***	**0.16** ***	**0.13** ***	0.14 ***
(IQR)	(0.06)	(0.06)	(0.04)	(0.06)	(0.04)	(0.02)	(0.05)	(0.08)	(0.02)	(0.02)	(0.04)	(0.06)
Sig. IRF>0 (# stocks)	49	50	9	0	49	50	50	39	49	50	46	8

TABLE AIII
Quote-midpoint impact by message type (case *l*=5)

We estimate the cross-sectional median quote midpoint impact of a shock to the order flow imbalance controlling for the type of message: (1) market orders and marketable limit orders (Panel A); (2) submissions of non-marketable limit orders (Panel B), and (3) revisions and cancellations (monitoring) (Panel C). OI stands for the share imbalance of marketable order submissions, NOF(S) stands for the share imbalance of non-marketable limit order submissions, and NOF(M) stands for the imbalance of revisions and cancellations of standing limit orders (monitoring activity). All the imbalance metrics are expressed in relative terms to the total volume of the orders involved. For each stock, we estimate a SVAR model for continuously-compound quote midpoint return and the imbalance metrics. In computing the NOF components, we consider messages placed either at or within the prevailing five best quotes of the LOB (NOF^5). The variables are defined over regular time intervals of 1, 5, 60, and 300 seconds. The SVAR assumes one-way contemporaneous causality from imbalances to returns. We report the cross-sectional median permanent quote midpoint impact both in basis points (IRF) and in relative terms to the standard deviation of the quote midpoint returns (IRF/σ(R)), the corresponding cross-sectional interquartile range, and the number of stocks (out of 50) for which we get a positive and statistically significant impact. Bold format means statistically significant at least at the 5% level. ***, **, * means that a given statistic in the NOF(S) or NOF(M) panel is statistically different than the corresponding statistic in the OI panel at the 1%, 5%, and 10% level of statistical significance (according to Wincoxon's ranksum tests). Our sample consists on the constituents of the NIFTY-50 from April to July 2015.

Panel A: OI	Time bar length (seconds)			
	1	5	60	300
Price impact (IRF)	**0.51**	**0.90**	**2.62**	**5.86**
(IQR)	(0.17)	(0.21)	(0.58)	(1.79)
IRF/σ(R)	**0.28**	**0.32**	**0.30**	**0.32**
(IQR)	(0.03)	(0.06)	(0.05)	(0.06)
Sig. IRF>0 (# stocks)	50	50	50	49
Panel B: NOF(S)				
Price impact (IRF)	**0.40** ***	**0.82**	**3.63** ***	**9.50** ***
(IQR)	(0.09)	(0.20)	(1.18)	(3.56)
IRF/σ(R)	**0.23** ***	**0.29**	**0.42** ***	**0.54** ***
(IQR)	(0.07)	(0.09)	(0.11)	(0.18)
Sig. IRF>0 (# stocks)	50	50	50	50
Panel C: NOF(M)				
Price impact (IRF)	**0.24** ***	**0.49** ***	**2.83**	**8.46** ***
(IQR)	(0.08)	(0.16)	(1.25)	(4.64)
IRF/σ(R)	**0.13** ***	**0.17** ***	**0.32**	**0.46** ***
(IQR)	(0.03)	(0.03)	(0.11)	(0.18)
Sig. IRF>0 (# stocks)	50	50	50	50

TABLE AIV
Order flow toxicity: NOF (case *l*=5)

This table provides aggregated statistics on the estimation of the dynamic regression model in equation [15] on the relationship between illiquidity and the absolute order flow imbalance over regular intraday time intervals (in this case, 1-second and 5-second time bars). As dependent variable, we use either the weighted by time relative quoted spread (RQS), the weighted by volume relative effective spread (RES), or the illiquidity ratio of Amihud (2002) (AIR). RQS and RES are in basis points. AIR is multiplied by one million. As for the order flow imbalance, we use the NOF^5 (see "first specification"); this order-flow-based metric takes into account order submissions, revisions, and cancellations affecting orders placed or standing at or within the prevailing five best quotes (l=5). We also consider the decomposition of the NOF into components due to market (and marketable limit) orders, commonly known as "order imbalance" (*OI*), and non-marketable limit orders (NOF^5(L)) (see "second specification"). All the NOF and OI metrics are standardized per stock. As controls, the model includes the first lag of the dependent variable, dummies for the initial and final trading hours of each trading session, the logarithm of the volume in shares, and the standard deviation of the quote midpoint changes (as a proxy for volatility). All the non-deterministic explanatory variables are lagged one period. Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015. The model is estimated per stock by Newey and West (1994). We report the cross-sectional average coefficient for the variables of interest, aggregated t-statistics based on Chordia et al. (2005), and the percentage of stocks for which the coefficient of interest is statistically significant at the 1% and 5% level. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

	First specification		Second specification			
	NOF^5(b)		NOF^5(L)(b)		OI(b)	
Dependent variable	1s bars	5s bars	1s bars	5s bars	1s bars	5s bars
(a) RQS (b+1)						
coef*100	-6.17 ***	-3.89 ***	-8.61 ***	-5.40 ***	-0.47	0.05
t-test	-16.72	-7.66	-23.62	-11.77	-1.49	0.39
Sig.>0 at 1% (5%) (# stocks)	0 (0)	3 (4)	0 (0)	1 (2)	11 (12)	14 (18)
(b) RES (b+1)						
coef*100	-17.60 ***	-12.48 ***	-18.33 ***	-13.36 ***	-6.55 ***	-4.88 ***
t-test	-35.50	-19.30	-37.23	-21.78	-6.93	-4.20
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	0 (0)	0 (0)	3 (4)	4 (5)
(c) AIR (b+1)						
coef*100	-28.26 ***	-22.33 ***	-32.39 ***	-24.44 ***	12.31 ***	14.95 ***
t-test	-19.39	-14.79	-23.21	-18.44	6.66	11.25
Sig.>0 at 1% (5%) (# stocks)	0 (0)	1 (1)	0 (0)	0 (0)	37 (39)	49 (49)

TABLE AV
Order flow toxicity: NOF and trader types (case *l*=5)

This table provides aggregated statistics on the estimation of the dynamic regression model in equation [15] on the relationship between illiquidity and the absolute order flow imbalance. We control for trader type. In particular, we distinguish between proprietary ATs (HFTs), agency ATs (AATs), and non-ATs (NATs). As dependent variable, we use either the weighted by time relative quoted spread (RQS), the weighted by volume relative effective spread (RES), or the illiquidity ratio of Amihud (2002) (AIR). RQS and RES are in basis points. AIR is multiplied by one million. As for the order flow imbalance, we use the NOF^5 of each type of trader. The NOF^5 is an order-flow-based metric takes into account order submissions, revisions, and cancellations affecting orders placed or standing at or within the prevailing five best quotes (l=5). All the NOF metrics are standardized per stock. Control variables are the same used in Table X. All the non-deterministic explanatory variables are lagged one period. Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015. The model is estimated per stock by Newey and West (1994). We report the cross-sectional average coefficient for the variables of interest, aggregated t-statistics based on Chordia et al. (2005), and the percentage of stocks for which the coefficient of interest is statistically significant at the 1% and 5% level. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

	HFTs' NOF^5		AATs' NOF^5		NATs' NOF^5	
Dependent variable	1s bars	5s bars	1s bars	5s bars	1s bars	5s bars
(a) RQS (b+1)						
coef*100	2.28 ***	0.33	-5.18 ***	-5.42 ***	-6.64 ***	-2.14 ***
t-test	8.04	0.82	-17.21	-12.68	-16.21	-2.79
Sig.>0 at 1% (5%) (# stocks)	45 (45)	17 (18)	0 (0)	0 (0)	0 (0)	7 (8)
(b) RES (b+1)						
coef*100	2.66 ***	-1.16 **	-11.19 ***	-12.55 ***	-11.35 ***	-6.72 ***
t-test	7.68	-2.21	-26.97	-21.59	-20.43	-7.84
Sig.>0 at 1% (5%) (# stocks)	40 (42)	12 (13)	0 (0)	0 (0)	0 (0)	0 (0)
(c) AIR (b+1)						
coef*100	4.79 ***	-0.97	-21.50 ***	-21.15 ***	-10.24 ***	1.13 ***
t-test	4.77	-1.44	-17.49	-17.22	-5.12	2.58
Sig.>0 at 1% (5%) (# stocks)	34 (36)	12 (14)	0 (0)	0 (0)	1 (2)	21 (26)

TABLE AVI
Order flow toxicity: NOF(L), OI, and trader types (case *l*=5)

This table provides aggregated statistics on the estimation of the dynamic regression model in equation [15] on the relationship between illiquidity and the absolute both marketable and non-marketable order flow imbalance. We control for trader type. In particular, we distinguish between proprietary ATs (HFTs) (Panel A), agency ATs (AATs) (Panel B), and non-ATs (NATs) (Panel C). Dependent and control variables are the same as in Tables AI, and AII. As for the order flow imbalance, we use the NOF^1 of each type of trader. The NOF^1 is an order-flow-based metric takes into account order submissions, revisions, and cancellations affecting orders placed or standing at or within the prevailing five best quotes (l=5). We decompose each trader types' NOF^5 into components due to market (and marketable limit) orders, commonly known as "order imbalance" (*OI*), and non-marketable limit orders (NOF^5(L)). While the OI is computed from trades, the NOF(L) is computed from submissions, revisions, and cancellations of limit orders. All the NOF and OI metrics are standardized per stock. All the non-deterministic explanatory variables are lagged one period. Our sample consists on the constituents of the NIFTY-50, the official market index of the NSE of India, from April to July 2015. The model is estimated per stock by Newey and West (1994). We report the cross-sectional average coefficient for the variables of interest, aggregated t-statistics based on Chordia et al. (2005), and the percentage of stocks for which the coefficient of interest is statistically significant at the 1% and 5% level. ***, **, * means statistically significant at the 1%, 5%, and 10% level, respectively.

	NOF^5(L)(b)		OI(b)	
Panel A: HFTs	1s bars	5s bars	1s bars	5s bars
(a) RQS (b+1)				
coef*100	0.89 ***	0.12	8.84 ***	10.16 ***
t-test	3.39	0.48	32.14	24.99
Sig.>0 at 1% (5%) (# stocks)	27 (27)	12 (16)	50 (50)	50 (50)
(b) RES (b+1)				
coef*100	-0.25	-1.58 ***	19.53 ***	20.18 ***
t-test	0.46	-2.77	51.85	36.10
Sig.>0 at 1% (5%) (# stocks)	20 (22)	7 (7)	50 (50)	50 (50)
(c) AIR (b+1)				
coef*100	2.14 ***	-1.51	19.77 ***	15.86 ***
t-test	2.93	-1.14	16.20	16.18
Sig.>0 at 1% (5%) (# stocks)	26 (31)	14 (16)	50 (50)	50 (50)
Panel B: AATs				
(a) RQS (b+1)				
coef*100	-5.87 ***	-5.94 ***	3.42 ***	1.97 ***
t-test	-20.79	-14.63	8.32	2.92
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	48 (49)	25 (27)
(b) RES (b+1)				
coef*100	-10.55 ***	-11.80 ***	4.84 ***	4.12 ***
t-test	-27.18	-21.69	8.09	4.39
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	42 (44)	28 (34)
(c) AIR (b+1)				
coef*100	-18.97 ***	-18.72 ***	5.01	1.59 *
t-test	-17.72	-17.20	1.18	1.66
Sig.>0 at 1% (5%) (# stocks)	0 (0)	0 (0)	19 (19)	19 (23)
Panel C: NATs				
(a) RQS (b+1)				
coef*100	-8.68 ***	-6.15 ***	0.41	-0.42
t-test	-27.64	-11.23	1.35	-0.69
Sig.>0 at 1% (5%) (# stocks)	0 (0)	2 (2)	17 (22)	3 (6)
(b) RES (b+1)				
coef*100	-4.68 ***	-3.77 ***	-9.66 ***	-9.24 ***
t-test	-9.69	-4.72	-16.64	-11.19
Sig.>0 at 1% (5%) (# stocks)	1 (1)	3 (3)	0 (0)	0 (0)
(c) AIR (b+1)				
coef*100	-17.15 ***	-8.80 ***	1.01 *	3.93 ***
t-test	-13.81	-4.17	-1.72	3.74
Sig.>0 at 1% (5%) (# stocks)	1 (1)	6 (6)	6 (6)	25 (27)