Identifying High Frequency Trading activity without proprietary data

Bidisha Chakrabarty, Carole Comerton-Forde, and Roberto Pascual

July 2022

Abstract

Public databases do not identify high frequency traders (HFT), so researchers use proxies. We assess the reliability of commonly used proxies by benchmarking them against true HFT metrics. All proxies are highly correlated and can identify HFT activities in general, and their liquidity demanding and supplying functions, in regular times as well as episodes of high and/or low HFT activity. Their reliability does not depend on the time windows of data aggregation. *Strategic Runs* (Hasbrouck and Saar, 2013) works best to isolate HFT-specific activities.

JEL classification*:* G10

Keywords: *High-frequency trading; HFT proxies; Quote intensity; Strategic Runs; Monitoring Intensity; Message traffic; Cancellations; Liquidity demand; Liquidity supply*

[☆] The authors acknowledge a grant from the NSE-NYU Stern Initiative on the Study of Indian Financial Markets. We thank Tirthankar Patnaik (Chief Economist of the National Stock Exchange of India) for expert advice on the market structure and trading protocols at the NSE. We thank Allen Carrion, Jing-Chi Chen, Linda Chen (EFMA discussant), Thierry Foucault, Lewen Guo, Joel Hasbrouck (NSE-NYU discussant), Kose John, Andreas Park (NFA discussant), Andriy Shkilko (Finance Forum discussant), Konstantin Sokolov, Chen Yao (CICF discussant), and participants at the China International Conference in Finance, European Financial Management Association meetings, Indian Institute of Management (Udaipur) workshop, Northern Finance Association conference (2021), NSE-NYU Conference, NSE Workshop on Trading and Technology, University of Memphis, University of Technology (Sydney), and Women in Market Microstructure Conference for useful comments. We gratefully acknowledge guidance from Professors Joel Hasbrouck and Gideon Saar in adapting the logic of their low-latency (strategic runs) metric to our data.

High frequency traders (HFTs) contribute a significant proportion of global market volume and it is important to understand how they affect markets.[1] While early evidence (Brogaard, Hendershott, and Riordan, 2014; Conrad, Wahal, and Xiang, 2015) documents significant positive effects of these traders, recent research (Aquilina, Budish, and O'Neill, 2022; Koraczyk and Murphy, 2019) paints a more nuanced picture where HFTs also have (some) deleterious effects. To understand HFTs' effects, one needs to identify trades involving HFTs (versus non-HFTs). Public databases (e.g., NYSE's Trade and Quote database; Refinitiv's Tick History Data) do not provide trader identities. Researchers, therefore, often infer HFT activity using various proxies.

Given the growing importance and volume of HFT research and the use of proxies to infer HFTs' activities, in this study we conduct a comprehensive examination of the popular HFT proxies used in the literature.[2] Our purpose is to determine the efficacy of each proxy in identifying HFT activities, examine what type of HFT activities (liquidity demand or supply or both) they capture, and how well each proxy works to identify HFT activity when there are episodes of unusually high and/or low HFT liquidity demand or HFT liquidity supply. Finally, we evaluate which proxy, if any, may be more successful in isolating high-speed activity that is attributable to HFTs but not to other types of traders.

---

[1] O'Hara (2015) provides an excellent review of how algorithmic traders (ATs) and HFTs have changed markets. In a 2012 testimony to the US Senate Banking Committee, the Tabb Group estimates HFT in US equities at 55% of volume (https://www.banking.senate.gov/imo/media/doc/TabbTestimony92012.pdf), while across EU venues the European Securities and Market Authority (ESMA, 2014) estimates HFT at 24% to 43% for equities. In our literature review we group together AT and HFT although in our tests, we carefully distinguish between these two groups.

[2] Researchers also use other methods to identify HFTs such as direct identification in proprietary databases (e.g., Brogaard, Hendershott and Riordan, 2014; Kirilenko, Kyle, Samadi, and Tuzun, 2017; Comerton-Forde, Malinova, and Park, 2018; Boehmer, Li, and Saar, 2019; Brogaard, Hagströmer, Nordén, and Riordan, 2015), latency changing events (e.g., Conrad, Wahal, and Xiang, 2015; Hendershott, Jones and Menkveld, 2011; Ye, Yao, and Gai, 2013; Riordan and Storkenmaier, 2012; Chakrabarty, Moulton, and Pascual, 2017), or events that are expected to cause the most disruptions in HFTs' connectivity (Shkilko and Sokolov, 2020). These methods have their own disadvantages. Most researchers do not have access to proprietary datasets so these studies are not replicable, whereas replicability is a particularly important aspect of any scientific investigation.

To do this, we run a horserace of the various proxies and compare them with precisely identified HFT activity. For this exercise to be successful, we require correct HFT identification. This is made possible by data from the National Stock Exchange of India (NSE) which provides precise HFT identification at a message level, time-stamped at the nanosecond frequency.[3] Specifically, in our data we can identify whether orders are entered using the exchange-provided algorithmic order entry and management system, and further identify for each such algorithmic orders whether it is submitted for a proprietary account (high frequency traders or HFTs) or a client account (agency algorithm traders or AATs). The traders who do not use algorithmic order entry and management are the non-algorithmic traders or NATs, are also identified in our data. The strength of our data is that we know with certainty the source of the message, and can accurately attribute each message to an HFT, AAT or NAT. We do not need to rely on any proxies such as high cancellations or fast response times, such as those used to create other proprietary data sets like the ones provided by EUROFIDAI, Nasdaq, or the Canadian regulator IIROC. However, we note that in our data we are unable to identify individual accounts, so we cannot track any individual HFT firm over time. For each message we can identify whether it originated from algorithmic or non-algorithmic order entry mode, and additionally identify if it is for proprietary or agency trading. Therefore, we can cleanly identify HFT activity.

Using these HFT-identified data, we proceed with our examination as follows. First, we select a comprehensive set of measures that are popular in the literature to proxy for HFT activity.[4] We identify nine popular measures: (1) *Message Traffic* (*Mess*) is the number of messages,

[3] The NSE ($2.55 trillion) is the 4th exchange in the world, ranking just after the Toronto Stock exchange ($2.6 trillion) in capitalization. https://www.tomorrowmakers.com/stocks/top-10-stocks-exchanges-world-listicle
[4] Several of these measures are designed to capture AT rather than HFT activity specifically. We choose to include all measures in our investigation to cast a wide net to examine which type of trader activity each one captures. In Section 3, we evaluate which measure specifically captures HFT activity, versus other algorithmic trading activity.

including order submissions, revisions, and cancellations; (2) *Cancellations* (*Can*) is the number of limit order cancellations; (3) *Monitoring Intensity* (*MonInt*) is the number of cancellations plus the number of revisions; (4) *Fleeting Orders* (*FleetOrd*) is the number of orders cancelled or revised within 100 milliseconds after submission; (5) *Quote Intensity* (*QuoteInt*) is the number of changes in either the best bid quote, best ask quote, or depth at the best quotes; (6) Flickering quotes (*Flick*) is the standard deviation of the quote midpoint over 100 millisecond time intervals; (7) *Speed of Response* (*SResp*) is the number of responses to quote improvements within 100 milliseconds, (8) *Strategic Runs* (*SRuns*) is the number runs or sequences of linked revision or cancellation plus resubmission messages, within a given interval, and (9) Immediate-or-cancel orders (*IOC*).[5] We label these measures as HFT metrics and detail their construction in Section 2.

We construct each of these nine metrics in two ways – one using the HFT identifier in our data, the other without considering the HFT identifier and using all messages, as is commonly done in the literature when authors use proxies of HFT. To minimize the effect of possible outliers, we normalize each metric by subtracting its mean and scaling by its standard deviation. We label the metrics constructed from the HFT identifier in the data as *true* and the metric created from unflagged (total) order flow as *proxy*. For example, the *true Mess* metric is the number of messages from HFTs while the *proxy Mess* metric is the total number of messages. If message traffic is a good proxy for HFT activity, we should see a high correlation between *true Mess* and *proxy Mess*.

We begin our investigation by verifying that the nine metrics differ between the three trader categories, and in the direction expected. For example, while for HFTs the average cross-sectional message traffic or *Mess* in our sample is 654.08, the corresponding number for AATs is 204.12

[5] We adapt the *FleetOrd* measure of Hasbrouck and Saar (2009) and *SRuns* measure of Hasbrouck and Saar (2013) to the institutional features of the NSE. We discuss the institutional features in Section 1 and the measures in Section 2.

and for NATs it is 55.76.[6] Clearly HFTs generate significantly more message traffic than AATs or NATs. Likewise, HFTs are the fastest in responding to quote improvements (*SResp*), averaging 38.27 responses per minute vs. 11.9 for AATs and 0.55 for NATs. For *IOC*, the difference between HFTs and AATs is much lower than the difference between these algorithmic traders and the NATs. Overall, the results in this table indicate that the metrics we have constructed align with our expectations that HFTs are the highest speed traders. This validation exercise leads to the next steps: to compare the performance of the various *proxies* against the *true* metrics.

First, we show that the popular HFT metrics are highly correlated with each other. We compute both Pearson and Spearman correlations to address any possible non-linearities or outlier issues with the distribution of these metrics. Using both the *true* and the *proxy* measures, we find that the coefficients range between 40% and 90% (36% to 89%) for Pearson (Spearman) correlations and are statistically significant. More importantly, we examine the correlations between each HFT *proxy* and its corresponding *true* metric to assess how well each proxy captures the analogous true HFT activities. We find that for all nine metrics, correlations are high, ranging from 68% to 98% for Pearson and 74% and 99% for Spearman coefficients. All coefficients are significant at the 99% level of confidence. This provides strong evidence that the popular metrics used to represent HFT activity accurately represent true HFT activity and all of them are successful in capturing HFT activity.

We next investigate which types of HFT activities these metrics capture. To establish HFTs' effect on market quality, researchers often examine whether HFTs contribute more to liquidity supply or demand. Theory models (e.g., Baldauf and Mollner, 2020; Menkveld and Zoican, 2017) and empirical works (e.g., Brogaard and Garriott, 2019) distinguish between high

[6] We discuss data filters when we describe the construction of each metric. The numbers cited here are limited to messages placed within ten ticks from the best quotes and when data are aggregated over a 60 second time window.

frequency market makers who supply liquidity and have a positive effect versus high frequency opportunistic traders who demand liquidity and often have deleterious effect on market quality. Using both the *true* and *proxy* metrics in regression frameworks, we find that all the HFT metrics are good at explaining *trades* in which HFTs supply liquidity as well as *trades* in which HFTs are liquidity takers. Taking a step further, we next investigate whether these metrics can also explain other contributions of HFTs to the limit order book (LOB). We use three measures to capture HFTs' liquidity provision: the percentage of time HFTs are present at the best quotes, the time-weighted HFTs' contribution to the best quoted depth, and the time-weighted HFTs' contribution to the accumulated depth at the five best LOB levels. We find that the HFT metrics, *true* as well as *proxies*, explain HFT liquidity provision in the limit order book using all three measures.

Earlier we alluded to research which documents that HFTs' effect on market outcomes may vary depending on episodes of unusually high and/or low liquidity demand and/or supply. In fact, these unusual episodic liquidity spikes and HFTs' possible role in exacerbating these episodes is of current regulatory interest. After the Flash Crash (studied in Kirilenko et al., 2017), US regulators approved the limit-up limit-down (LULD) mechanism by which steep short-lived price spikes created in markets due to fast trading could be curbed.[7] Given the increasing and topical interest in these episodic spikes, we next examine how the HFT metrics perform during periods of unusually high or unusually low HFT liquidity demand or supply.

In a regression framework we find that all the *true* metrics as well as all the *proxy* metrics effectively identify HFT activity during periods of unusually high or low HFT liquidity demand

[7] The LULD mechanism restricts trades within an allowable band of 5% for large caps and active ETFs and 10% for other NMS stocks. If trading does not resume within the band within 15 seconds of a price band violation, trading switches to a 5-minute trading halt. The LULD was approved on May 31, 2012 and went into effect on a pilot basis on April 8, 2013. See SEC release 34-67091, May 31, 2012. Market-wide circuit breakers were triggered an historically unprecedent four times in March 2020, which is now being investigated according to a Congressional Research report, available at https://crsreports.congress.gov/product/pdf/IN/IN11339

as well as HFT liquidity supply. When the HFT liquidity demand and supply are both low (high), all metrics reach minimum (maximum) levels. For instance, when data are aggregated over 30-second time windows, the estimated average normalized *Mess* during ordinary times is -0.01. When both HFTs' liquidity demand and supply are unusually low, the *true* (*proxy*) *Mess* falls by –0.47 (–0.51). Keeping HFT demand at low, if HFT liquidity supply increases to unusually high levels, the *true* (*proxy*) *Mess* metric falls only by –0.15 (–0.21). This is in line with our expectation that when supply increases, keeping demand at the same (low) level, the HFT metric should increase if it is to be a valuable signal. Likewise, if HFT liquidity supply is held at an unusually low level and demand increases to unusually high, we again see the *true* (*proxy*) *Mess* metric coefficient increases to –0.18 (–0.21) from the comparison values of –0.47 (–0.51). Finally, if both demand and supply reach unusually high levels, the *true* (*proxy*) *Mess* metric increases to 0.43 (0.41). These patterns hold for all nine HFT metrics, indicating that all of them correctly signal the movements of the true metrics during these episodic spikes in HFT liquidity demand and/or supply.

So far, our results validate the reliability of the HFT proxy metrics in tracking true HFT activity. We round out our inquiry by asking a follow up question that includes the other two trader groups – AATs and NATs – within the ambit of our investigation. How closely are these proxies related to the activity of other trader groups? To answer this question, we first examine how the HFT *proxy* metrics correlate with the same metrics constructed for the AATs and NATs. For example, for the *Mess* metric, we construct AAT's *true Mess* from the message traffic of AATs only and examine its correlation with the HFT *proxy Mess*. If the AATs' order flow correlates with the HFTs' order flow, then both metrics should show significant correlation.

This exercise show that the correlations between HFTs' and AATs' order flow is much higher than the analogous correlations between HFTs' and NATs' order flow. However, all the correlations are statistically significant. This is not unexpected. Studies show that HFTs respond to incoming order flow from other traders, especially institutional traders (van Kervel and Menkveld, 2019) who are in both the AAT and the NAT category. Further, Comerton-Forde, Malinova and Park (2018) show that HFTs acting as informal market makers take turns to trade against the total incoming order flow. Brogaard, Hendershott and Riordan (2014) show that HFTs impose adverse selection costs on other investors by trading with them when they (the HFTs) have better information. Therefore, the positive correlations between HFT's order flow and AATs' and NATs' order flows are to be expected. What is heartening here is that the correlations of the HFT *proxy* metrics are higher with the AATs than with the NATs. Of note is the fact that of the nine HFT metrics examined, *SRuns* consistently show the lowest correlations with AATs and NATs. Recall that for an HFT metric to be a "good" proxy, it is reasonable to expect that HFTs should be the main driver of that proxy. So, a lower correlation with the other trader types, especially NATs who do not use the same technology or trading strategies as HFTs, is a desirable trait.

In the final step of our analysis, we examine which proxy (or proxies) is (are) driven primarily by HFT activity. If such a proxy can be identified, it can be used by researchers to focus on HFT activity when trader identifiers are not available in the data. To do this, we use a two-stage regression approach. In the first stage, we run a regression, stock by stock, of each HFT *true* metric on the corresponding AAT and NAT *true* metrics and save the residuals, which can be interpreted as the component of HFT activity that is uncorrelated with AAT and NAT activities. In the second stage, we regress these residuals on the corresponding HFT *proxy* (for example, the residual of HFT *true Mess* on *Mess*) using a pooled regression specification with intra-day controls. We do this in turn for AATs and NATs too. The results of this exercise show that while the residual HFT metrics remain significantly correlated with all nine HFT proxies, it shows the best fit for *SRuns*.

With a coefficient estimate of 77.59%, the largest *t*-statistic, and the highest R^2, *SRuns* stands out as the best among the nine proxies In addition, the R^2 of the second stage regression for *SRuns* is negligible for AATs (0.01) and NATs (0.00), indicating that the correlation, despite being statistically significant, is economically irrelevant. Therefore, *SRuns* appears to be the HFT *proxy* metric that is primarily driven by HFT (but not AAT or NAT) activities and can best identify HFT activities in data that do not identify trader types.

Finally, we note that when each of these metrics are created, researchers have the choice to aggregate data over particular time windows. For example, when creating *Mess*, one can aggregate data over a millisecond, a second, 30 minutes, or daily levels, to name a few. In research into high-speed trading, authors have often sought fine granularity in data to be able to observe the speed of response of HFTs. However, public data sources may not always record data at the finest granularity levels. Since our data are timestamped to the nanosecond, we also constructed the *true* and *proxy* metrics by using various time windows of data aggregation, starting at 30 seconds and going all the way up to the daily level. We do not find any qualitative difference in our conclusions based on the time window of data aggregation.

The rest of this paper is organized as follows. Section 1 describes the NSE market, our data, and provides some characteristics of our sample. Section 2 describes the construction of the HFT metrics. Section 3 presents results, and Section 4 concludes.

1. NSE market, data, and sample

1.1 NSE market

The NSE began operations as a fully electronic limit order book in 1995, and by 2018 it ranked as the 4^{th} (10^{th}) largest exchange in the world in terms of number of trades (dollar volume) according to statistics released by the World Federation of Exchanges. The Indian equity market is relatively unfragmented. There are only two lit exchanges, the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). There are no dark pools or other alternative trading venues. Since commencing operations in 1994 the NSE has come to dominate the BSE on many dimensions. The NSE lists larger companies – the BSE lists nearly three times as many companies despite the two exchanges having similar total market capitalizations.[8] The NSE also captures 78% to 80% of the trading activity during our sample period, including most of the activity of the largest stocks in India, those in our sample.[9] As a newer exchange, the NSE is more technologically advanced than the BSE. For example, colocation was introduced at the BSE in November 2012, almost three years after the NSE. HFTs are also more active on the NSE. Therefore, examining algorithmic trading behavior on the NSE allows us to capture a large swathe of the total sample of algorithmic trades in the Indian market. Orders are executed on the centralized limit order book of the NSE based on a price-time priority basis. There are no designated market makers.[10] Trading starts at 9:00 A.M. with a 15-minute pre-opening period followed by a call auction. Continuous trading occurs from 9:15 A.M. to 15:30 P.M.[11]

The main financial market regulator in India – the Securities and Exchange Board of India, (SEBI) – permitted the introduction of algorithmic trading in April 2008. But the high latency of computer networks between the exchange and members kept AT levels relatively low on the NSE in the initial period. This changed once colocation was introduced in January 2010. The specific

[8] https://www.world-exchanges.org/home/index.php/statistics/annual-statistics.

[9] Further details are available at: https://www.sebi.gov.in/sebi_data/attachdocs/1463726488005.pdf. By many measures, the NSE outpaces the BSE For some comparisons between these two exchanges, see https://economictimes.indiatimes.com/markets/stocks/news/what-listing-of-bse-and-nse-means-for-the-exchanges-and-investors/articleshow/56787104.cms BSE listings can trade on the NSE and vice-versa, using unlisted trading privilege (UTP), very similar to how NYSE and Nasdaq listings can trade in either exchange using the UTP. https://tradingqna.com/t/most-of-the-companies-are-listed-on-both-exchanges-so-which-exchange-to-choose/6562

[10] Since 2013, market makers are permitted in certain categories of index futures, but not in equity securities, which is the focus of our paper.

[11] The NSE displays on its website real-time information of the top five ask and bid quotes (price and depth).

form of colocation on the NSE allows traders to rent rack-space within the exchange premises next to exchange servers. The service is offered with a tiered fee structure, where a fixed sum is charged for a fixed amount of bandwidth (measured as number of messages per second) and increased to the next tier if the orders per messages exceed the lower tier allowance.[12] Because of this particular type of tiered pricing, where the fees are based on tranches of messages, traders on the NSE find it cost-effective to revise existing orders when they wish to change quoted prices or depth, rather than cancel orders and place new orders, which is more common in the US markets.[13] After the introduction of colocation the share of ATs in total traded value gradually increased from 10% in 2009, to about 55% in 2013 (Aggarwal and Thomas, 2014). Nawn and Banerjee (2019) report that 95% of the order messages and 43% of the trading volume in the 50 largest NSE-listed stocks comes from ATs, which include both proprietary (HFTs) and agency ATs (AATs).

1.2 Data

Our data is provided by the NSE. The dataset of orders and trades is comprehensive, with details of each transaction for each trading day stored with the date and exact timestamp at the nanosecond. There is a message file and a trade file. The message file contains information on every order, including submissions of market and limit orders, and cancelations and revisions of standing limit orders.[14] Each order has a unique code which allows us to track its history overtime. Limit orders are by far the most common type of order. For limit orders, we know the direction (buy or sell), limit price, the displayed size, and the hidden size. The trade files provide information on each individual trade, including the size, the price, and the time at which the trade took place.

[12] https://www.nseindia.com/trade/platform-services-co-location-facility

[13] Nikolsko-Rzhevska Nikolsko-Rzhevskyy and Black (2020) show that even in the US the number of order revisions is on the rise.

[14] There are identifiers for orders with special conditions, such as iceberg orders (which shows the displayed and the hidden portions of the total order), on-stop, or immediate-or-cancel (IOC).

Most importantly the trade files also provide the code of the orders involved in the trade, which allows us to correlate orders with order executions. An incoming aggressive order may execute against one or several limit orders on the opposite side of the book. Where an incoming order 'walks up or down the book', the trade is reported in several entries, one for each passive order executed.

Important to our study is the trader identification process these data allow. The NSE data includes flags that allow us to identify trader types (not the individual identity of each trading firm). Each order identifies whether it was entered using the exchange-provided algorithmic order entry and management system. Therefore, we can identify for every order whether it originates from an AT or an NAT. Further, the data also identify for every AT whether it is for a client account (agency algorithmic trading or AAT) or a proprietary account (high frequency trading or HFT).[15] This enables us to classify each order as belonging to one of three mutually exclusive and exhaustive groups: HFTs, AATs, and NATs. In addition to more precisely pinpointing HFT, our identification allows traders to switch their type and is thus temporally dynamic, which is different from the assumption in many other proprietary databases (the Paris HFT data from the EUROFIDAI, for example) where HFT classification, once assigned, does not change. Our identification process allows us to accurately capture AAT and HFT even within a single firm, for example, a bulge bracket broker who does both agency trading (AAT) and proprietary trading (HFT). Since we have message level identifiers, for a firm that has an HFT trading desk but is not a pure play HFT, we can separate the order flow coming from the HFT trading desk from the rest of its order flow. The Nasdaq HFT data, for example, cannot make such a distinction.

[15] Algorithmic trading by proprietary trading firms is the SEC's (2010) definition of HFT.

1.3 Sample

We use the 50 stocks in the Nifty-50 Index of NSE as our sample for the three month period from May to July 2015. The Nifty-50 Index tracks the movements of the top 50 blue-chip companies by market capitalization that are traded on the NSE. Although this index includes only 50 of the 1600 companies that trade on the NSE, it captures over 66% of the float-adjusted market capitalization and is therefore considered a true reflection of the Indian stock market.[16] Using our data, we re-construct the full order book for each stock at each timestamp in a day, and determine the accuracy of our re-constructed order book by comparing the trades generated from our reconstructed order book with the actual trades files, similar to the procedure followed in Aggarwal, Panchapagesan, and Thomas (2020).

Table I shows the characteristics of our full sample, as well as for above and below median firms. Average market capitalization of the firms in our full sample is about Rs. 1135 billion, which translates to 17.87 billion US dollars per the INR/USD exchange rate in May 2015.The average market capitalization for the above median firms is Rs. 1827 billion and Rs. 442 billion for the below median firms. The difference is large and significant. Likewise, depth at the best quotes (in thousands of rupees) for the above median firms is over twice as large (1,032,000) as that for the below median firms (512,500). Limit order submissions (101,250) far outnumber market order submissions (2,850), as in most developed markets (e.g., Brogaard, Hendershott, and Riordan, 2019 report that only 5% of all messages are market order submissions). As mentioned earlier, due to the tiered pricing of fees charged for algorithmic order entry, in the NSE limit order revisions are far more prevalent (1,130,780) compared to cancellations (65,570). Interestingly, the

[16] https://tradebrains.in/nifty-50-companies-list/

difference between the above median and below median firms is greater in the category of revisions than cancellations.

[Table I]

To obtain a better sense of the prevalence of HFT activity in the NSE market, as well as to benchmark with HFT activity studied in other markets around the world, we next present some descriptive statistics of HFT contribution to various metrics in the limit order book in our sample. Table II shows those metrics.

[Table II]

In Panel A of Table II, we show HFTs' contribution to the order flow. HFTs are engaged in 23.02% of all trades. There is very little difference in participation between the above-median (22.49%) and below-median (24.08%) firms. These participation rates are lower than those observed in the US but are comparable to levels reported in Canada and Australia. Brogaard, Hendershott and Riordan (2014) report HFT accounts for around 42% of trading in large stocks but only 18% in small stocks on the Nasdaq in 2008 to 2009, Brogaard, Hendershott and Riordan (2019) report that HFT accounts for around 20% of trading in Canada for a 2012-2013 sample, and ASIC (2015) report that HFT accounts for around 27% of trading in Australia in the March quarter of 2015.

Reflecting the trend in the overall market (Table I), HFTs also submit more limit orders than market orders. Notably, HFTs also cancel fewer orders (70.16%) compared to their order revisions (87.30%), reflecting the economic rationale provided by the NSE's message-fee pricing structure as discussed in Section 1.1. Consistent with results observed in other markets, HFTs account for a smaller fraction of submissions, cancellations, and revisions in smaller firms.

In Panel B of Table II, we provide cross-sectional average daily statistics on HFTs' contribution to liquidity supply in the NSE limit order book. We measure HFT contribution using three alternative measures: (a) the percentage of time HFTs quote either the best ask prices and/or the best bid prices; (b) the time-weighted percentage of the total depth at the best quotes that is posted by HFTs; and (c) the time-weighted percentage of the accumulated total depth at the five best levels of the LOB that is posted by HFTs. HFTs spend 47.07% of the time at best quotes in the full sample. For the above-median stocks, this reduces to 40.03% while for the below-median stocks it increases to 53.83%, which is a statistically significant difference, and likely reflects the fact that there is more competition for liquidity provision in the larger cap stocks. HFTs contribute 22.62% to the best quoted depth and 27.60% to the accumulated depth at the first five levels of the order book. These levels are similar to those reported for the Australian market, where HFTs account for 20% of the depth at the best prices and 25% of the depth at the first three levels of the order book (ASIC, 2013).

2. HFT metrics

Studies that use HFT proxies create them from characteristics in the data that are likely to be associated with HFTs but not with non-algorithmic traders. Several of these proxies are not exclusive to HFTs but may characterize all algorithmic traders. Since HFTs are a subset of algorithmic traders, we opt to use metrics with a broader scope. We evaluate nine popular measures used in the literature to proxy for HFT activity.[17] We compute all nine metrics for each stock (i) and time interval (t) in two alternative ways using our high-frequency message-level data:

[17] We note here that some of the proxies used in the literature are normalized by the number of trades. We opt not to normalize our proxies by the number of trades because Yao and Ye (2018) robustly show that normalizing by trades renders these proxy metrics, in their case message-to-trade ratio, a poor cross-sectional proxy for HFTs' liquidity provision. In fact, stocks with more liquidity provided by HFTs have *lower* message-to-trade ratios.

- *true* HFT metrics: Using all messages identified in our database as originating from HFTs.
- *proxy* HFT metrics: Using all messages, irrespective of the trader type.

Our first metric, *Message Traffic* or $Mess_{i,t}$, is the number of messages for stock *i* within interval *t*. Messages include order (limit, market, IOC, etc.) submissions, price or size revisions of limit orders, and cancellations. We include only messages at or less than 10 ticks from the prevailing best quotes.[18] This metric is similar in spirit to the algorithmic trading metric of Hendershott, Jones, and Menkveld (2011). Message traffic as a proxy for high-speed traders has also been used by Boehmer, Li, and Saar (2018), and Weller (2018), and is based on the notion that these traders are the main contributors to message traffic (e.g., SEC, 2014). Biais and Foucault (2014) provide an excellent review of the logic for why message traffic serves as a measure of AT activity in general and HFT activity as its subset.

Our second metric is the number of *Cancellations* ($Can_{i,t}$). This measure captures only cancellations at or less than 10 ticks from the prevailing best quotes. Although cancellations are part of the *Mess* metric, the HFT literature has often honed in on the fact that the faster connections and technology possessed by HFTs allow them to rapidly cancel orders, which is why in modern markets approximately nine out of ten orders end up being cancelled. Yao and Ye (2018), in examining why trading speed matters, argue that in the landscape of speed competition in liquidity provision, there is a justified focus on order cancellation to avoid the risk of being adversely selected (Hoffmann, 2014; Jovanovic and Menkveld, 2015). Liquidity providers post quotes at which they will trade. When new information arrives, their quotes become stale, but if they have a speed advantage, they can quickly cancel the stale quotes and avoid being adversely selected. HFTs possess the technology to quickly cancel such stale quotes. Hence order cancellations are a

[18] Our results are not sensitive to this choice; but excluding non-aggressive orders reduces noise in the measure.

hallmark of HFTs. The use of this proxy is supported by several empirical studies showing that HFTs' cancellation rates of standing limit orders are much higher than those of non-HFTs (e.g., Friederick and Payne, 2015; Jørgensen, Skjeltorp, and Ødegaard, 2016; Malinova, Park, and Riordan, 2018; Dahlström, Hagströmer, and Nordén, 2018, and Chakrabarty, Hendershott, Nawn, and Pascual, 2020).[19]

Our third metric, *Monitoring Intensity* ($MonInt_{i,t}$), is the number of limit order updates (revisions and cancellations) at or within 10 ticks of the prevailing best prices, for stock i within interval t. This metric captures active risk management by HFTs. HFTs take advantage of their low monitoring costs, enhanced information processing capabilities, and superior speed to actively revise their orders in response to market events, incoming news, or upon detecting toxic order flow (Hoffmann, 2014; Janovic and Menkveld, 2016). This intense monitoring activity results in high order update-to-trade ratios which is another hallmark of HFT activity (Liu, 2009).

Our fourth metric, *Fleeting Orders* ($FleetOrd_{i,t}$), was first proposed by Hasbrouck and Saar (2002). In examining the order book of the Island ECN (later incorporated into Nasdaq) they found that about 25% of all orders are cancelled within two seconds. They dubbed these 'fleeting orders.' Although they do not explicitly relate fleeting orders to HFT, they associate their increasing use to several factors including technological progress. In the two decades since their study, trading technology has substantially sped up. To reflect the realities of the NSE market, we compute fleeting orders as the number of orders updated (either cancelled or revised) within 100

[19] Recall that revisions are more prevalent on NSE. We include revisions in our subsequent metrics, but do not include revisions along with cancellations to facilitate comparison with studies that have used only cancellations as a proxy for HFT activity.

milliseconds after submission.[20] Cartea, Payne, Penalva, and Tapia (2019) also use this proxy in their empirical study.

Our fifth metric, *Quote Intensity* ($QuoteInt_{i,t}$), is the number of quote mid-point or best depth changes, i.e., number of updates in either the best offer or bid quotes, or the best bid depth or best ask depth changes. This metric was introduced by Conrad, Wahal, and Xiang (2015). *QuoteInt* measures quote flickering, which is the result of placing and quickly canceling or revising orders.[21] Therefore, it should be strongly correlated with other proxies, such as *FleetOrd*, *MonInt*, or *Can*. While the previous metrics require message level data, $QuoteInt_{i,t}$ only requires high-frequency data on the best quotes, and is less computationally intensive.

Our sixth metric is *Flick*, which is proposed in Hasbrouck (2018) and captures quote flickering. This measure may be computed as the standard deviation of the best ask, best bid, or quote midpoint (in levels, not changes) over extremely short time intervals. We report results using the quote midpoint and compute its standard deviation over 100 millisecond time intervals, and then aggregate over 30 seconds, 300 seconds, and so on. Results are robust to using the best bid or best ask quotes instead of the quote midpoint.

[20] Hasbrouck and Saar (2009) identify fleeting orders based exclusively on the time to cancellation. This measure can be broadened by using the time to first message change. After all, a revised order is equivalent, from the economic perspective, to a cancellation and resubmission. Since the goal of the metric is to capture low latency activity, it makes sense to focus on the first event altering an order, irrespective of whether it is a cancellation or a revision. Nonetheless, our findings are robust to the use of the original definition of fleeting order. They are also robust to the use of alternative time aggregation windows, from 25 milliseconds to 2 seconds (Hasbrouck and Saar's choice).

[21] In the theoretical model of Baruch and Glosten (2013), quote flickering results from limit order traders managing their undercutting exposure by rapidly cancelling their quotes and replacing them with randomly chosen ones. Empirically, Hasbrouck (2018) shows that quote flickering arises from the strategies and interactions (such us successive undercutting) of high-frequency market makers. Hasbrouck and Saar (2009) associate quote flickering with trading strategies whereby technologically sophisticated traders chase prices or search for hidden or latent liquidity. In the theoretical models of Hoffmann (2014) and Jovanovic and Menkveld (2016), quote flickering results from active risk management by fast liquidity providers. Finally, flickering quotes may result from manipulative practices such as quote stuffing and spoofing, often attributed to HFTs (e.g., Egginton, Van Ness, and Van Ness, 2016).

The seventh metric is the *Speed of Response* ($SResp_{i,t}$), which is the number responses in the limit order book that happen within 100 ms of an NBBO quote improvement (i.e., decrease in the best ask or increase in the best bid).[22] Hasbrouck and Saar (2013) assess HFTs' speed of response (see their Figure 2) using three types of responses: (a) a limit order submission on the same side as the improvement; (b) a cancellation of a limit order on the same side, and (c) an execution against the improved quote. We follow a similar approach but also include (d) a cancellation on the opposite side, and (e) a revision on the same side that results in an increase in aggressiveness, as valid responses as well.[23]

The success of HFTs' strategies relies on their response speed to different market signals (Foucault, Hombert, and Roçu, 2016; Baron, Brogaard, Hagströmer, and Kirilenko, 2018). Accordingly, they acquire low-latency technology and pay for low-latency services, such as colocation or direct access to data feeds, to be able to respond to events faster than other traders, including other HFTs (Ding, Hanna, and Hendershott, 2014). Therefore, it is reasonable to expect that an increase in HFT activity will improve the speed of response. Chakrabarty, Jain, Shkilko and Sokolov (2020) show that when HFTs are impacted by a ban on direct market access, their speed of response declined significantly, indicating that this metric correlates well with HFT activity.

Our eighth metric is based on a measure proposed by Hasbrouck and Saar (2013). This metric – *Strategic Runs* ($SRuns_{i,t}$) – captures dynamic limit order strategies at the millisecond frequency, mostly attributable to HFTs. A strategic run is a set of linked messages. In Hasbrouck

[22] Our inferences hold if we use the time to response instead of the number of responses within a given time.

[23] We treat revisions that increase the aggressiveness of an order standing on the same side of the book as a cancellation plus a resubmission, therefore, matching this update with cases (a)+(c) above. As an alternative definition, we also consider the average speed, in seconds, with which there is a response in the limit order book after a quote improvement. Our findings with this alternative metric are the same.

and Saar's original formulation for the Nasdaq ITCH database, a run starts with a non-marketable limit order submission that is later cancelled. This cancellation is then linked with a subsequent resubmission of either a passive or a marketable order if the latter occurs within 100*ms* after the cancellation, is in the same direction, and is for the same size (adjusted for earlier partial executions or revisions). The run builds forward linking subsequent cancellations and resubmission messages and terminates when either (a) the last linked order is fully executed, or (b) the last linked order is partially executed, revised, or cancelled and there is no new message for which we can impute a link. The first run starts with the first order submitted on a given trading session. All orders that comprise the strategic run are marked as 'used' and not considered for other strategic runs. The second run starts with the first 'unused' limit order. After building the second run, we mark its components as 'used.' The process continues until there are no 'unused' submission messages. This metric has been used by Bartlett and McCrary (2019), Ersan and Ekinci (2016), Boehmer, Li, and Saar (2018), and Chordia and Miao (2020).

In our context, we amend the algorithm of Hasbrouck and Saar (2013) to accommodate order revisions and, thus, broaden the definition of *SRun*, as we do with the definition of fleeting order.[24] Recall that at the NSE, the total number of messages per co-located server connection is constrained by their tranche fee structure. As a result, limit order revisions, that count as a single message, are much more common than cancellations plus resubmissions, which count as two messages.[25] Therefore, connections between orders with different IDs that in the Nasdaq market might result from cancellations and ultra-fast replacements, could show up in the NSE data as fast revisions of the same order ID. Additionally, if we understand Hasbrouck and Saar's algorithm as

[24] We thank Professors Joel Hasbrouck and Gideon Saar for providing detailed instructions on how they computed their strategic runs using ITCH data and the logic that would parallel to the NSE market.
[25] Although in US markets cancellations are still more common that revisions, revisions are growing and are now thrice as common as executions (Nikolsko-Rzhevska, Nikolsko-Rzhevskyy, and Black, 2020).

an attempt to identify messages that are part of a "smart order" algorithm, by accounting for revisions we reduce the error associated with inferring a strategic run relying exclusively on orders that have different IDs.

To align with the logic of Hasbrouck and Saar (2013), we define a strategic run as a sequence of linked messages that satisfy the following conditions: (a) the average time between subsequent messages with the same order ID (i.e., revisions) must be less than 500 milliseconds; (b) there must be no revision with a size increase; (c) two messages with different IDs are linked if and only if the orders involved have the same size, direction, and the submission of the second order occurs no more than 100*ms* after the cancellation of the first order, and (d) the "length" of the run (i.e., the number of linked messages) must be at least equal to 10.[26] Once we have the list of all strategic runs, we compute $SRuns_{i,t}$ as the number of strategic runs that stock *i* experiences in interval *t*.[27]

Our last (ninth) measure is a particular order type – the immediate-or-cancel orders (*IOC*) – that is used mainly by traders who are latency-conscious. IOCs are limit orders combined with the instruction to either fill the order immediately or to instead cancel it. Studies (Aquilina, Budish, and O'Neill, 2021; Comerton-Forde, Malinova, and Park, 2018) show that IOCs are aggressive orders used by traders with sophisticated trading technology who wish to trade on short-lived information. In modeling HFTs, Baldauf and Moller (2020) show that liquidity taking HFTs or 'snipers' use IOCs. Our data identify orders that are received at the exchange as IOCs. We use this

[26] Our findings persist if we reduce the average time between revisions to 100 or 250*ms* or if we increase it to 1 second. Similarly, our findings persist if we consider a minimum length of 5 or 20 messages.

[27] Hasbrouck and Saar (2013, p. 659) time-weight $SRuns_{i,t}$. Time-weighting works as follows: if a given run is active throughout interval t, it adds 1 to $SRuns_{i,t}$. If it is active only 10% of *t*, it adds 0.1. Hasbrouck and Saar (2013) argue that this minimizes measurement errors. We find that time-weighting reduces the variability of the metric, especially when we consider large intervals for aggregation, which results in a lower correlation with other proxies and a slightly lower performance in signaling unusually high or low HFT activity. We therefore opt for the unweighted version.

flag to construct the true metric using the number of IOCs entered by HFTs, and the proxy metric by using all IOC orders.[28]

3. Results

Before examining the efficacy of the HFT proxies and presenting our results, we perform one more benchmarking exercise to ensure that what we identify as HFT indeed aligns with our *a priori* expectation of HFT behavior. We construct all nine of our HFT metrics described above, using the actual order flow of each trader type – HFT, AAT, and NAT. If *Mess* is a good proxy for HFT behavior, for example, then we expect that *Mess* constructed from the message traffic of the HFTs identified in our data will be much higher than *Mess* constructed using the message traffic of the NATs in our data. Table III reports the result of this exercise.

[Table III]

Table III shows that while average cross-sectional *Mess*, aggregated over one-minute time intervals and limited to orders standing within 10 ticks from the prevailing best quotes, is 654.08 for HFTs, the comparable numbers are 204.12 for AATs and 55.76 for NATs. Clearly *Mess* is economically and statistically significantly higher for HFTs than the other two trader types. *Can* for HFTs (16.51) and AATs (17.27) are significantly higher than those of the NATs (2.10).[29] We expect that access to algorithmic order management available for both HFTs and AATs creates these similar levels of cancellation for these two groups. Additionally, Subrahmanyam and Zheng (2015) find that HFT firms do not cancel orders more frequently than non-HFT firms, which is

[28] We also disaggregated the IOC metric by executed versus unexecuted IOCs; results are similar.

[29] It is important to explain the differences between the statistics in Tables II and III. In Table III order flow is limited to everything that happens within 10 ticks from the prevailing best quotes. In Table II, we use all messages to give an overall representation of stock-level activity before imposing any filters. So, the fact that in Table III we have similar number of cancellations for HFTs and AATs indicates that HFTs cancel non-aggressive orders more often than AATs, and more often than their own aggressive orders (which they likely choose to revise, rather than cancel given the fee structure in the NSE market).

also a similar feature in our data. For all other metrics as well, the numbers for the HFTs are much larger than the comparable numbers for the AATs and NATs, although for *IOC*, both types of algorithmic traders use these significantly more than the NATs. This finding suggests that *IOC* could be a good proxy for overall AT activity, but it may fail to isolate HFT activity from AAT activity.

3.1 Correlations between true and proxy metrics

The HFT and non-HFT metrics in Tables II and III are computed using the flags provided in the data, which allow direct identification of the trader type behind each message. Most researchers, however, do not have data with HFT identification. So, we now move to the main purpose of this study - to examine how these metrics fare if constructed from overall order flow data without any HFT identification (i.e., *proxy* measures) and begin by asking three related questions: (a) how well do the *proxy* metrics correlate amongst themselves, (b) how well do the *true* metrics correlate amongst themselves, and (c) how well do the *proxy* metrics correlate with the *true* metrics. (a) and (b) address whether these metrics are good substitutes for each other and (c) addresses the more important question of whether the *proxy* metrics serve as good substitutes for the *true* metrics.

To construct these correlations, we first standardize each series by subtracting the mean and scaling by its standard deviation, and then filter each series for intraday deterministic patterns by regressing the time series of each metric per stock on fifteen 25-minute interval dummies (recall that the NSE continuous trading session starts at 9:15 A.M. and ends at 15:30 P.M.). We use the residuals of those regressions to compute the Pearson correlations reported in Table IV Panels A and B. To address any possible non-linearities or outliers in the time series of these metrics, we also compute Spearman correlations and report them in Table IV Panels C and D. In each panel of

Table IV, the upper triangular matrix reports results using 300-second time aggregation windows, while the lower triangular matrix reports correlations with 30-second time aggregation windows.

[Table IV]

In Table IV, Panel A shows the Pearson correlations between the *proxy* HFT metrics while Panel B reports analogous results for the *true* HFT metrics. The results show that all metrics are highly positively correlated with each other for both the *proxy* (Panel A) and the *true* (Panel B) HFT metrics. The pairwise correlations align with expectations. For example, for the 300s time intervals, *Mess* and *MonInt* show one of the highest correlations (86.52%), which is expected since cancellations and revisions of standing limit orders are the main components of message traffic. Likewise, *QuoteInt* and *Flick* are highly correlated (81.3%). While *QuoteInt* captures the volatility of the quote midpoint return, *Flick* is based on the variance of the quote midpoint in levels. Therefore, the high correlation reinforces that if we use the same input (quote midpoint) to compute them, both metrics provide similar information. The highest correlation is between *QuoteInt* and *SResp* (88.33%), suggesting that most of the best quote improvements (that increase *QuoteInt*) receive a fast response in less than 100ms (which increases *SResp*). Finally, fleeting orders are usually placed at or near the best quotes and increase message traffic, which explains the high correlation of *FleetOrd* with *QuoteInt* (75%), and *Mess* (79.25%). *FleetOrd* is highly correlated with *SRuns* (79.84%) as strategic runs often involve sequences of nested fleeting orders. Panel B of Table IV reveals that, among the *true* metrics the highest correlations involve the same metrics. All conclusions with Pearson correlations are supported by the Spearman correlations reported in Panels C and D, indicating that the distributional properties of the HFT metrics do not affect our conclusions in any material way.

Next, we proceed to ask how well the *proxies* correlate with the *true* measures. In other words, if the *proxy* metrics are to be useful, they should show positive, high, and significant correlations with their corresponding *true* metrics. Table V Panels A and B provide those results for Pearson and Spearman correlations, respectively. The HFT proxies indeed capture HFT activity. The Pearson correlation between the *true* and *proxy* measures of each metric is high, ranging between 67.94% (for the *Flick* metric at 30 seconds aggregation) and 98.45% (for *SRuns* at 1500 seconds aggregation). All are significant at the 99% level of confidence and for all windows of time aggregation (30, 60, 300, 900, and 1800 seconds). The results are generally stronger for the Spearman correlations in Panel B. So, our initial answer to the question of how well HFT proxy metrics measure true HFT activity appears to be an unqualified "very well."

[Table V]

3.2 A closer look: HFT liquidity demand and supply

While early research generally documents that HFT activity improves market outcomes (Brogaard, Hendershott, and Riordan, 2014; Carrion, 2013; Menkveld, 2013), as the evidence has developed over time, the picture that has emerged is more nuanced. Studies show that if the activity of these fast traders is separated out by their liquidity demand versus supply functions, conclusions about their beneficial effects on markets are sometimes not supported (Chakrabarty et al., 2020; Shkilko and Sokolov, 2020; Brogaard, Hendershott, and Riordan, 2019). So, we next examine how these metrics perform when specifically trying to capture the *trade-based* liquidity demanding versus liquidity supplying roles of HFTs.

To do so, we estimate pooled regression models of trade-based HFT liquidity supply (or "making") and liquidity demand (or "taking") on *true* HFT metrics (Table VI, Panels A and B) and *proxy* HFT metrics (Panels C and D). Liquidity supply is the percentage of trades in which

HFTs provide liquidity while liquidity demand is the percentage of trades in which HFTs demand liquidity. We regress each of these HFT liquidity demand/supply measures on each HFT activity metric (*true* metrics in Panels A and B and *proxy* metrics in Panels C and D) at a time, including as control variables the first lag of the dependent variable, a dummy for the first 30-minutes of trading, and a dummy for the last 30 minutes of trading.

[Table VI]

Specifically, we estimate a pooled regression model of HFT liquidity demand/supply in a specific time interval (30 and 300 seconds reported) on HFT metrics computed over the same time interval, i.e.,

$$HFT(D/S)_{it} = \alpha + \beta HFT\_Metric_{it} + \gamma HFT(D/S)_{it-1} + \delta D_t + \varepsilon_{it}, \tag{1}$$

where $HFT\ (D\ /\ S)_{it}$ is the measure of HFT liquidity demanding trades or HFT liquidity supplying trades, respectively, for firm *i* in time interval *t* and HFT_Metric_{it} is in turn the *true* HFT metric created from HFT-identified order flow (Panels A and B) and the *proxy* HFT metric created from all order flow (Panels C and D). All continuous variables are winsorized at the 1% on the right tail of the distribution and then normalized per stock by subtracting the mean and dividing by the standard deviation. D_t are time-of-day dummies computed at the same frequency as the time aggregation window for each equation. We include stock fixed effects and standard errors are clustered by stock and time intervals (Thompson, 2011). We report the estimates of the coefficient of interest β.

Panel A of Table VI shows how the *true* HFT metrics explain trades in which HFTs supply liquidity. For each time interval and for all nine HFT metrics, the estimated coefficients are positive and significant at the 99% level of confidence (*SResp* at 30s intervals being the only exception). This indicates that the true HFT metrics are good at capturing HFT liquidity supplying trades. Panel B examines how the *true* HFT metrics explain HFTs liquidity demanding trades. The results here are similar to those in Panel A and confirm that all nine HFT *true* metrics can signal HFTs' liquidity demanding trades. Results hold for both the 30 and 300 second time aggregation windows reported in these panels.

Panels C and D show how the *proxy* HFT metrics explain HFTs liquidity supplying and liquidity demanding trades, respectively. Again, for all metrics and for both the 30-second (except for *SResp*) and the 300-second levels of time aggregation, the estimated coefficients are significant at the 99% level of confidence and show the expected signs. Overall, the results in Table VI show that all the HFT proxies are very good at identifying times when HFTs are involved in liquidity supplying as well as liquidity demanding trades.

We repeat this exercise, but instead of trades, we examine alternative popular measures of HFTs' contribution to liquidity supply in the limit order book. We use three measures of liquidity supply: (a) the percentage of time HFTs are present at the best bid and ask quotes, (b) the time-weighted HFTs' contribution (as a percentage) to the best quoted depth, and (c) the time-weighted HFTs' contribution (as a percentage) to the quoted accumulated depth at the top five levels of the order book. We estimate pooled regressions similar to Eq (1) but instead of the $HFT(D/S)_{it}$ variable we now use the variables described in (a) to (c) above. Results are reported in Table VII.

[Table VII]

In Panels A through C, we report results using the *true* HFT metrics. The coefficients indicate that all nine *true* HFT metrics show the expected signs (positive) and are significant at the 99% level of confidence. In Panels D through F, we report results using the *proxy* HFT metrics. Consistent with the earlier results, all three panels show that the HFT *proxy* metrics have the right (expected) signs and are also significant at the 99% level of confidence. In sum, Table VII demonstrates that the HFT *proxies* robustly capture HFT's contribution to liquidity in the limit order book.

3.3 Periods of unusually high/low HFT liquidity demand/supply

One area of interest in high-speed traders' activities is their role during episodic spikes or troughs in liquidity demand and/or supply (Kirilenko et al., 2017). In the following tests, we address the performance of the HFT metrics during episodes of unusually high or low liquidity demand and/or supply. We define an unusually high (low) HFT contribution when the corresponding metric is above (below) the 75th (25th) percentile of its empirical distribution for each stock. Combining occurrences of unusually high and low supply and unusually high and low demand, we generate four dummy variables: (Dlow, Slow) is a dummy variable that equals 1 when HFTs' liquidity demand and supply are both unusually low, 0 otherwise; (Dlow, Shigh) is a dummy variable that equals 1 when HFTs' liquidity demand is unusually low but HFTs' liquidity supply is unusually high, 0 otherwise; (Dhigh, Slow) is a dummy variable that equals 1 when HFTs' liquidity demand is unusually high but HFTs' liquidity supply is unusually low, 0 otherwise; and (Dhigh, Shigh) is a dummy variable that equals 1 when HFTs' liquidity demand and supply are both unusually high, 0 otherwise. Our aim is to examine whether the HFT metrics, both the *true* and more importantly the *proxies*, can correctly signal HFT activity in situations where such high/low demand/supply conditions occur.

We cast these dummy variables in a regression framework similar to that described in Eq (1) and include additional dummy variables for the first 30 minutes and last 30 minutes of the trading session (Open and Close, respectively) as control variables. We use pooled regression models with stock fixed effects and double-cluster standard errors (by stock and time of day). We report our findings for the 30-second time aggregation windows. As in previous analyses, all continuous variables are winsorized at the 1% on the right tail of the distribution and then normalized. For brevity of reporting, we present results for two HFT activity measures – percentage of time HFTs are present at the best quotes, and HFTs' contribution to the limit order book quoted depth. Results are reported in Table VIII.

[Table VIII]

Panel A (C) presents the results with the *true* (*proxy*) HFT metrics for the percentage of time HFTs are present at the best quotes. For instance, when HFT liquidity demand and supply are both low, the coefficient of the *true* (*proxy*) *MonInt* metric is –0.38 (-0.41). Keeping demand at low, if HFT liquidity supply increases to unusually high, the *true* (*proxy*) *MonInt* metric coefficient increases to –0.24 (–0.27). This is in line with our expectation that when HFT liquidity supply increases, keeping HFT liquidity demand at the same (low) level, the HFT metric should increase if it is to be a valuable signal of HFT activity. Likewise, if HFT liquidity supply is held at an unusually low level and demand increases to unusually high, we again see the *true* (*proxy*) *MonInt* metric coefficient increases to –0.14 (–0.16). This pattern of movement holds for all nine HFT metrics, indicating that all nine proxies correctly signal the movements of the true metrics during these episodic spikes in HFT liquidity demand and/or supply. Consistently, all metrics show minimum (maximum) levels when both HFT liquidity demand and supply are unusually low (high). Analogous conclusions hold when comparing Panels B and D which report the results for HFT contribution to the depth in the limit order book. Overall, this table indicates that the HFT *proxies* are also capable of signaling the correct direction of HFT activity during times when HFT demand/supply is unusually high/low, although no proxy can single out liquidity demand from liquidity supply.

3.4 Can the proxies isolate HFTs from other traders?

While our investigation so far has validated the reliability of all nine HFT *proxy* metrics in tracking true HFT activity, our final aim is to identify whether any of these *proxy* metrics can pinpoint primarily HFT activity. After all, if the proxies also signal the activities of other types of traders, then their reliability to signal HFT activity will be limited. Therefore, in this final segment of our investigation, we widen the scope of our inquiry to include the other two trader groups: AATs and NATs. We ask if these proxies are also related to the other trader groups, and more importantly, if one proxy (or multiple proxies) can reliably signal primarily HFT activity (and not the activities of the other trader groups).

As a first step, we examine how the HFT *proxy* metrics correlate with the same metrics constructed from the order flow of the AATs and NATs. For example, for the *MonInt* metric, we construct the AAT *true MonInt* from the message traffic of AATs only, and examine its correlation with HFT *proxy MonInt*. If AATs' order flow is correlated with HFTs' order flow, then AATs' *true MonInt* should show significant correlation with HFTs' proxy *MonInt*. Table IX shows the results of this exercise.

[Table IX]

Panel A (B) shows the cross-sectional average correlations between the HFT *proxy* metrics and AAT (NAT) *true* metrics. The correlations are generally higher (although they continue to remain lower for *SRuns*) between the *proxies* and AATs' activity than between the *proxies* and NATs' activity. This is reassuring since AATs and HFTs share similar trading technologies while NATs do not. However, all the correlations are statistically significant. This is not unexpected given what the existing literature documents about HFTs' reactions to other traders' order flow. For example, Yang and Zhu (2020) show that their use of high-speed trading technology allows

HFTs to front- and back-run other traders, which induces correlated order flow, consistent with the main question in Hirschey (2021) regarding whether HFTs anticipate buying and selling pressure. Hirschey (2021) finds that the correlation between HFT and non-HFT trades is particularly strong at certain times (for example when non-HFTs are impatient and thus less focused on disguising order flow) and for certain types of stocks (for example, those in which information leakage is higher). What aligns with *a priori* expectations and is therefore encouraging in these results is that the correlations of the HFT *proxy* metrics with AATs' activity are greater in magnitude than the correlations of the HFT *proxy* metrics with NATs' activity. Of note is the fact that of all nine HFT metrics, *SRuns* and *IOC* show the lowest correlations.

As explained earlier, a "good" HFT proxy should rely on HFT activity as its primary driver. By showing a lower correlation with the other trader groups, *SRuns* provides a promising start. From here we proceed to examine if we can isolate, for each of the HFT *true* metrics, the component that is uncorrelated with AAT and NAT activity.

To do so we estimate a two-stage regression model. For the first pass, we run a pooled regression, stock by stock, of the HFT *true* metric on the AAT and NAT *true* metrics. The residuals of this stage-one regression, which can be interpreted as the component of HFT activity that is uncorrelated with AAT and NAT activities, is then used in the second stage, where we regress these residuals on the corresponding HFT *proxy*. For example, for the *MonInt* proxy, the residual of HFT *true_MonInt* is regressed on HFT *MonInt* using a pooled regression specification with dummies that control for regular intraday patterns, stock fixed effects, and double-clustered standard errors. We do this in turn for AATs and NATs. Results are reported in Table X.

[Table X]

Panel A shows the results for the 30-second time aggregation window, and Panel B shows the corresponding results for the 300-second time window. Both panels show that while the residual HFT metric remains significantly correlated with the HFT metric for all proxies, it shows the best fit for *SRuns* with a coefficient of 83.16%, the largest *t*-statistic, and the highest R^2. In addition, for this metric, the R^2 of the second stage regression is also negligible for AATs (0.01) and NATs (0.00), indicating that the correlation, despite being statistically significant, is economically irrelevant. Thus, *SRuns* appears to be the HFT *proxy* metric that is primarily driven by HFT and not AAT or NAT activities. The other candidate, *IOC*, does not appear to have much power to distinguish between HFT and AAT activity.

4. Conclusion

We test how popular metrics that aim to capture HFT activity perform in identifying actual HFT activities. This is an important investigation because many researchers do not have access to data that provide HFT identifiers and therefore use proxies to arrive at conclusions about how HFT affects markets. Since HFTs contribute a significant fraction of all trading volume, their impact on market quality is substantial, and regulators depend upon such empirical evidence to enact rules to facilitate or limit HFT activities. Using data from the NSE, which provides precise identification of HFTs (as well as other trader types – AATs and NATs), we compute the nine most popular HFT metrics in two ways – *true* metrics using the identifiers in the data and *proxy* metrics using all order flow ignoring the HFT identifiers.

We find that both the *true* and *proxy* HFT metrics can distinguish well between HFT versus non-HFT activities, and these metrics are highly correlated with each other. Thus, researchers using any one of the popular metrics without HFT identifiers can sufficiently capture HFT activity. When we separate HFT activity into liquidity demand versus supply, the proxies do well in identifying HFT liquidity demanding trades, HFT liquidity supplying trades, as well as three other popular measures that are generally used to denote HFT contribution to the limit order book. We also find that these metrics all perform well in signaling HFT liquidity demand and/or supply during periods of unusually high and/or low HFT liquidity demand and/or supply.

Another useful finding is that the results generally hold for different time windows over which data are aggregated. Data granularity does not appear to be a factor in the proxies' reliability. This finding suggests that researchers are less constrained in their choice of data aggregation.

Finally, we find that of all nine HFT proxy metrics, the *Strategic Runs* (*SRuns*) metric of Hasbrouck and Saar (2013) is the one that is primarily driven by HFT and not the other trader groups' order flow. This validates recent studies like Chordia and Miao (2020) and Boehmer, Li, and Saar (2018) that use this metric to proxy for low latency trader (i.e., HFT) activities when trader type is not identified in the data. By robustly cataloging the performance of popular HFT proxies used in the literature, the market conditions under which they work, and which proxy best captures the activity of HFTs, we hope to foster more research into the role of HFTs, even when researchers do not have HFT identifiers in the data.

References

Aggarwal, Nidhi, and Susan Thomas. 2014. The causal impact of algorithmic trading on market quality. Working paper, IGIDR WP 2014-023 (http://oii.igidr.ac.in:8080/xmlui/handle/2275/323)

Aggarwal, Nidhi, Venkatesh Panchapagesan, and Thomas, Susan. 2020. When do regulatory interventions work? Available at SSRN (https://ssrn.com/abstract=3761163)

Aitken, Michael J., Angelo Aspris, Sean Foley, and Frederick H. de B. Harris. 2018. Market fairness: The poor country cousin of market efficiency. *Journal of Business Ethics* 147: 5-23.

Aquilina, Matteo and Budish, Eric B. and O'Neill, Peter. 2021. Quantifying the High-Frequency Trading 'Arms Race': A Simple New Methodology and Estimates. *Quarterly Journal of Economics*, Forthcoming.

Australian Securities and Investments Commission. 2013. Dark liquidity and high-frequency trading. Report 331, March (https://asic.gov.au/regulatory-resources/find-a-document/reports/rep-331-dark-liquidity-and-high-frequency-trading/)

Australian Securities and Investments Commission. 2015. Review of high-frequency trading and dark liquidity. Report 452, October. (Australian Securities and Investments Commission. 2015. Review of high-frequency trading and dark liquidity. Report 452, October).

Baron, Matthew, Jonathan Brogaard, Björn Hagströmer, and Andrei Kirilenko. 2019. Risk and return in high-frequency trading. *Journal of Financial and Quantitative Analysis* 54.3: 993-1024.

Bartlett, Robert P., and Justin McCrary. 2019. How rigged are stock markets? Evidence from microsecond time stamps. *Journal of Financial Markets* 45: 37-60.

Baruch, Shmuel and Lawrence R. Glosten. 2013. Fleeting Orders. Available at SSRN (https://ssrn.com/abstract=2278457)

Biais, Bruno, and Thierry Foucault. 2014. HFT and market quality. *Bankers, Markets & Investors* 128: 5-19.

Biais, Bruno, Thierry Foucault, and Sophie Moinas. 2015. Equilibrium fast trading. *Journal of Financial Economics* 116.2: 292-313.

Boehmer, Ekkehart, Kingsley Fong, and Julie Wu. 2021. Algorithmic trading and market quality: international evidence. *Journal of Financial and Quantitative Analysis*, 56.8: 2659-2688.

Boehmer, Ekkehart, Dan Li, and Gideon Saar. 2018. The competitive landscape of high-frequency trading firms. *The Review of Financial Studies* 31.6: 2227-2276.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. 2014. High-frequency trading and price discovery. *The Review of Financial Studies* 27.8: 2267-2306.

Brogaard, Jonathan, Björn Hagströmer, Lars Nordén, and Ryan Riordan. 2015. Trading fast and slow: Colocation and liquidity. *The Review of Financial Studies* 28.12: 3407-3443.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. 2017. High frequency trading and the 2008 short-sale ban. *Journal of Financial Economics* 124.1: 22-42.

Brogaard, Jonathan, Allen Carrion, Thibaut Moyaert, Ryan Riordan, Andriy Shkilko, and Konstantin Sokolov. 2018. High frequency trading and extreme price movements. *Journal of Financial Economics* 128.2: 253-265.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. 2019. Price discovery without trading: Evidence from limit orders. *The Journal of Finance* 74.4: 1621-1658.

Carrion, Allen. 2013. Very fast money: High-frequency trading on the NASDAQ. *Journal of Financial Markets* 16.4: 680-711.

Cartea, Álvaro, Richard Payne, José Penalva, and Mikel Tapia. 2019. Ultra-fast activity and intraday market quality. *Journal of Banking and Finance* 99: 157-181.

Chakrabarty, Bidisha, Pamela C. Moulton, and Roberto Pascual. 2017. Trading system upgrades and short-sale bans: Uncoupling the effects of technology and regulation. *Journal of Empirical Finance* 43: 74-90.

Chakrabarty, Bidisha, Terrence Hendershott, Samarpan Nawn, and Roberto Pascual. 2021. Order exposure in high frequency markets. Available at SSRN (https://ssrn.com/abstract=3074049)

Chakrabarty, Bidisha, Pankaj Jain, Andriy Shkilko, and Konstantin Sokolov. 2021. Unfiltered market access and liquidity: Evidence from the SEC Rule 15c3-5. *Management Science*, 67.2: 1183-1198.

Chordia, Tarun, and Bin Miao. 2020. Market efficiency in real time: Evidence from low latency activity around earnings announcements. *Journal of Accounting and Economics* 70.2-3: 101335.

Comerton-Forde, Carole, Katya Malinova, and Andreas Park. 2018. Regulating dark trading: Order flow segmentation and market quality. *Journal of Financial Economics* 130.2: 347-366.

Conrad, Jennifer, Sunil Wahal, and Jin Xiang. 2015. High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics* 116.2: 271-291.

Dahlström, Petter, Björn Hagströmer, and Lars L. Nordén. 2018. Determinants of limit order cancellations. Available at SSRN (https://ssrn.com/abstract=3012831).

Ding, Shengwei, John Hanna, and Terrence Hendershott. 2014. How slow is the NBBO? A comparison with direct exchange feeds. *The Financial Review* 49: 313-332.

Egginton, Jared, Bonnie F. Van Ness and Robert A. Van Ness. 2016. Quote stuffing. *Financial Management* 45.3, 583-608.

Ersan, Oguz, and Cumhur Ekinci. 2016. Algorithmic and high-frequency trading in Borsa Istanbul. *Borsa Istanbul Review* 16.4: 233-248.

ESMA. "High-frequency trading activity in EU equity markets", Economic Report #1, 2014. (https://www.esma.europa.eu/sites/default/files/library/2015/11/esma20141_-_hft_activity_in_eu_equity_markets.pdf)

Foucault, Thierry, Johan Hombert, and Ioanid Roşu. 2016. News trading and speed. *The Journal of Finance* 71.1: 335-381.

Friederick, Sylvain, and Richard Payne. 2015. Order-to-trade ratios and market quality. *Journal of Banking & Finance* 50: 214-223.

Hagströmer, Björn, and Lars Nordén. 2013. The diversity of high-frequency traders. *Journal of Financial Markets* 16.4: 741-770.

Hasbrouck, Joel. 2018. High-frequency quoting: short-term volatility in bids and offers. *Journal of Financial and Quantitative Analysis* 53.2: 613-641.

Hasbrouck, Joel, and Gideon Saar. 2002. Limit orders and volatility in a hybrid market: The Island ECN. Available at SSRN (https://ssrn.com/abstract=310940).

Hasbrouck, Joel, and Gideon Saar. 2009. Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets* 12.2:143–172.

Hasbrouck, Joel, and Gideon Saar. 2013. Low-latency trading. *Journal of Financial Markets* 16.4: 646-679.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld. 2011. Does algorithmic trading improve liquidity? *The Journal of Finance* 66.1: 1-33.

Hirschey, Nicholas. 2021. Do High-Frequency Traders Anticipate Buying and Selling Pressure? *Management Science* 67.3: 3321-3345.

Hoffman, Peter. 2014. A dynamic limit order market with fast and slow traders. *Journal of Financial Economics* 113.1: 156-169.

Jovanovic, Boyan, and Albert J. Menkveld. 2015. Middlemen in limit order markets. Available at SSRN (https://ssrn.com/abstract=1624329).

Jørgensen, Kjell, Johannes Skjeltorp, and Bernt Arne Ødegaard. 2016. Throttling hyperactive robots – Order-to-trade ratios at the Oslo Stock Exchange. *Journal of Financial Markets* 37: 1-16.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun. 2017. The flash crash: High‐frequency trading in an electronic market. *The Journal of Finance* 72.3: 967-998.

Korajczyk, Robert A. and Murphy, Dermot. 2019. High Frequency Market Making to Large Institutional Trades. *Review of Financial Studies* 32.3: 1034-1067.

Malinova, Katya, Andreas Park, and Ryan Riordan. 2018. Do retail investors suffer from high frequency traders? Available at SSRN (https://ssrn.com/abstract=2183806).

Menkveld, Albert J. 2013. High frequency trading and the new market makers. *Journal of Financial Markets* 16.4: 712-740.

Menkveld, A.J. 2018. High-frequency trading as viewed through an electron microscope. *Financial Analysts Journal* 74.2: 24-31.

Nawn, S., and A. Banerjee. 2019. Do proprietary algorithmic traders withdraw liquidity during market stress? *Financial Management* 48.2: 641-676.

O'Hara, Maureen. 2015. High frequency market microstructure. *Journal of Financial Economics* 116.2: 257-270.

Riordan, Ryan, and Andreas Storkenmaier. 2012. Latency, liquidity and price discovery. *Journal of Financial Markets* 15.4: 416-437.

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

Shkilko, Andriy, and Konstantin Sokolov. 2020. Every cloud has a silver lining: Fast trading, microwave connectivity and trading costs. *Journal of Finance* 75.6: 2899-2927.

Subrahmanyam, Avanidhar, and Hui Zheng. 2016. Limit order placement by high-frequency traders. *Borsa Istanbul Review* 16.4: 185-209.

Thompson, Samuel B. 2011. Simple formulas for standard errors that cluster by both firm and time" *Journal of Financial Economics* 99.1: 1-10.

van Kervel, Vincent, and Albert J. Menkveld. 2019. High‐frequency trading around large institutional orders. *Journal of Finance* 74.3: 1091-1137.

Weller, Brian M. 2018. Does algorithmic trading reduce information acquisition? *The Review of Financial Studies* 31.6: 2184-2226.

Yang, Liyan, and Haoxiang Zhu. 2020. Back-running: Seeking and hiding fundamental information in order flows. *Review of Financial Studies* 33.4: 1484-1533.

Ye, Mao, Chen Yao, and Jiading Gai. 2013. The externalities of high frequency trading. Available at SSRN (https://ssrn.com/abstract=2066839).

TABLE I
Sample Statistics

This table presents the daily average statistics of stock characteristics, message traffic composition, and liquidity for the 50 stocks in our sample, the 25 largest, and 25 smallest, by average market capitalization. Our sample consists of the 50 constituents of the NIFTY-50, the official market index of the National Stock Exchange (NSE) of India, for the period May to July 2015. For each variable, we report the mean and standard deviation (in parentheses) of the daily means across stocks. ***, **, * indicate statistically significant difference between the 25 largest and the 25 smallest at the 1%, 5%, and 10% levels, respectively, according to the rank-sum test for equality of medians.

Statistic	All stocks	25 largest	25 smallest
Market capitalization (billions of rupees (rs))	1,134.96	1,827.08	442.83 ***
	(951.37)	(909.16)	(152.86)
Price (rs)	949.76	1,091.11	808.42 **
	(988.58)	(915.21)	(1056.43)
Quote midpoint realized volatility (x10000)	8.08	7.53	8.63 ***
	(1.19)	(1.02)	(1.09)
Relative spread (bps)	10.10	8.49	11.71 ***
	(3.42)	(2.82)	(3.24)
Quoted depth at the best quotes (rs/1000)	772.62	1,032.74	512.50 ***
	(452.26)	(480.63)	(213.61)
Number of trades (/1000)	48.48	64.40	32.55 ***
	(29.05)	(30.74)	(15.82)
Volume (rs/1000000)	1,519.19	2,183.17	855.21 ***
	(1095.13)	(1123.67)	(517.17)
Messages (/1000)	1,300.67	1,825.88	775.45 ***
	(944.71)	(890.56)	(674.04)
Limit order submissions (/1000)	101.25	126.90	75.60 ***
	(48.26)	(42.82)	(39.37)
Market order submissions (/1000)	2.85	3.57	2.13 ***
	(1.77)	(2.03)	(1.11)
Cancellations (/1000)	65.57	81.42	49.72 ***
	(34.18)	(29.75)	(31.26)
Revisions (/1000)	1,130.78	1,613.72	647.83 ***
	(876.90)	(841.50)	(613.11)

TABLE II
HFT contribution

Panel A presents daily average statistics of high-frequency traders' (HFTs') contribution to trading activity and message traffic for the 50 stocks in our sample, the 25 largest, and 25 smallest, by average market capitalization. Our sample consists of the 50 constituents of the NIFTY-50, the official market index of the National Stock Exchange (NSE) of India, for the period May to July 2015. In Panel B, we report cross-sectional average daily statistics of HFTs' contribution to liquidity supply in the limit order book, where contribution is measured by three alternative metrics: (a) the percentage of time HFTs are at either the best ask and/or the best bid quote; (b) the percentage of the total depth at the best quotes that is posted by HFTs; and (c) the percentage of the accumulated total depth at the 5 best levels of the limit order book that is posted by HFTs. In both panels, we provide for each metric the mean and standard deviations (in parentheses) of daily mean across stocks. In addition, in Panel A we also report the percentage for the whole market. ***, **, * indicate statistically significant difference between the 25 largest and the 25 smallest at the 1%, 5%, and 10% levels, respectively, according to the rank-sum test for equality of medians.

Panel A: Order flow (cross-sectional average daily statistics)

	All stocks		25 largest		25 smallest		
	Mean	%	Mean	%	Mean		%
Number of trades initiated by (/1000)	11.16	23.02	14.48	22.49	7.84	***	24.08
	(6.15)		(6.02)		(4.26)		
Volume (rs/1000000)	393.79	25.92	553.89	25.37	233.68	***	27.32
	(283.27)		(290.43)		(161.47)		
Messages (/1000)	1,084.73	83.40	1,556.39	85.24	613.08	***	79.06
	(845.90)		(809.42)		(585.03)		
Limit order submissions (/1000)	51.29	50.65	66.62	52.50	35.95	***	47.56
	(31.16)		(27.62)		(27.02)		
Market order submissions (/1000)	0.19	6.75	0.24	6.85	0.14	***	6.58
	(0.13)		(0.14)		(0.10)		
Cancellations (/1000)	46.01	70.16	59.74	73.38	32.27	***	64.90
	(29.05)		(26.16)		(25.41)		
Revisions (/1000)	987.22	87.30	1,429.74	88.60	544.70	***	84.08
	(792.13)		(764.29)		(537.58)		

Panel B: Liquidity supply (cross-sectional average daily statistics)

	All stocks	25 largest	25 smallest	
Time at the best quotes (%)	47.07	40.30	53.83	***
	(15.55)	(10.45)	(17.01)	
Contribution to best quotes' depth (%)	22.62	20.99	24.24	**
	(5.84)	(5.56)	(5.77)	
Contribution to the book's depth (%)	27.60	27.42	27.77	
	(5.01)	(4.98)	(5.14)	

TABLE III
HFT activity metrics per trader type

This table reports the cross-sectional average statistics for the nine popular HFT metrics used in (or inspired by) the extant literature to capture HFT activity. Metrics are computed over 1-minute intervals and for three types of market participants: high-frequency traders (HFTs), agency algorithmic traders (AATs), and non-algorithmic traders (NATs). The nine proxy metrics are: the number of messages (*Mess*), where messages include submissions, revisions and cancellations of orders within 10 ticks from the prevailing best quotes; the number of cancellations (*Can*) of limit orders standing within 10 ticks from the best quotes; monitoring intensity (*MonInt*), defined as the sum of all limit order updates (revisions plus cancellations) as long as the orders are standing within 10 ticks from the best quotes; the number of fleeting orders (*FleetOrd*), where a fleeting order is defined as a limit order that is either cancelled or revised in less than 100 milliseconds (ms) after submission; quote intensity (*QuoteInt*) is the sum of all changes in the best ask and bid quotes or depth; speed of response (*SResp*) is the number of limit order book responses within 100 ms following NBBO quote improvements (i.e., decrease in the best ask or increase in the best bid). As valid responses we consider a limit order submitted on the same side as the improvement, a cancellation on the same side, an execution against the improved quote, a cancellation on the opposite side, and a revision on the same side resulting in an increase in aggressiveness of the limit order), and the number of strategic runs (*SRuns*), where a strategic run is a sequence of linked messages (revisions and cancellations plus resubmissions) that satisfy the following conditions: the average distance between subsequent messages for the same order ID must be less than or equal to 500ms, the size of the order must not increase, and orders with different IDs are linked if and only if they have the same size, direction, and the time between the cancellation of the first order and the submission of the second order is less than 100ms; immediate-or-cancel orders (*IOC*). For each metric, we report the mean and standard deviation (in parentheses) of the daily means across stocks. ***, **, * indicate statistically significant difference between the HFTs statistics and the corresponding statistic for either the AATs or the NATs at the 1%. 5%, and 10% levels, respectively. Our sample consists of the 50 constituents of the NIFTY-50, the official market index of the National Stock Exchange (NSE) of India, for the period May to July 2015.

	Trader type		
Metric	HFT	AAT	NAT
Mess	654.08	204.12 ***	55.76 ***
	(405.44)	(86.16)	(38.29)
Can	16.51	17.27	2.10 ***
	(10.23)	(7.76)	(1.43)
MonInt	293.18	61.62 ***	6.87 ***
	(192.78)	(29.73)	(5.82)
FleetOrd	82.24	26.89 ***	6.46 ***
	(58.93)	(12.20)	(4.46)
QuoteInt	137.91	88.84 ***	36.72 ***
	(80.81)	(30.74)	(22.56)
Flick(x1000)	4.30	2.16 **	1.07 ***
	(4.55)	(1.69)	(0.75)
SResp	38.27	11.93 ***	0.55 ***
	(36.87)	(8.81)	(0.42)
SRuns	10.49	0.51 ***	0.02 ***
	(9.39)	(0.44)	(0.03)
IOC	10.85	8.05 **	0.17 ***
	(6.51)	(3.75)	(0.11)

TABLE IV
Correlations between HFT proxies

This table provides cross-sectional time series correlations between the *proxies* for HFT activity in Panel A, and between the *true* metrics in Panel B, computed at different levels of aggregation (30-second and 300-second time windows). True metrics are computed using the actual HFTs' order flow as flagged in our database. We filter the series for intraday deterministic patterns by regressing the time series of each metric per stock on fifteen 25-minute interval dummies (the NSE continuous trading session starts at 9:15 and ends at 15:30). We use the residuals of those regressions to compute the correlations reported in this Table. The metrics are: the number of messages (*Mess*), where messages include submissions, revisions and cancellations of orders within 10 ticks from the prevailing best quotes; the number of cancellations (*Can*) of limit orders standing within 10 ticks from the prevailing best quotes; monitoring intensity (*MonInt*), defined as the sum of all limit order updates (revisions plus cancellations) as long as the orders are standing within 10 ticks from the best quotes; the number of fleeting orders (*FleetOrd*), where a fleeting order is a limit order that is either cancelled or revised in less than 100 milliseconds (ms) after its submission; quote intensity (*QuoteInt*) is the sum of all changes in the best ask and bid quotes or depth; speed of response (*SResp*) is the number of limit order book responses within 100 ms following NBBO quote improvements, and the number of strategic runs (*SRuns*), where a strategic run is a sequence of linked messages (revisions and cancellations plus resubmissions) that satisfy the following conditions: the average distance between subsequent messages for the same order ID must be less than or equal to 500ms, the size of the order must not increase, and orders with different ID are linked if and only if they have the same size, direction, and the time between the cancellation of the first order and the submission of the second order is less than 100ms; immediate-or-cancel orders (*IOC*). Our sample consists of the 50 constituents of the NIFTY-50, the official market index of the National Stock Exchange (NSE) of India, for the period May to July 2015. ***, **, * indicate statistically different from zero at the 1%, 5%, and 10% levels, respectively.

Panel A: Cross-sectional average Pearson correlations across HFT proxies (filtered for intraday deterministic patterns)

300 sec. (upper triangular matrix) and 30 sec. (lower triangular matrix) intervals

HFT proxy	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
Mess		71.20 ***	86.52 ***	79.25 ***	85.56 ***	69.73 ***	78.91 ***	66.72 ***	44.00 ***
Can	66.10 ***		70.64 ***	70.43 ***	68.01 ***	58.15 ***	64.40 ***	50.98 ***	43.16 ***
MonInt	84.12 ***	66.13 ***		67.92 ***	68.30 ***	56.98 ***	66.39 ***	53.98 ***	34.84 ***
FleetOrd	75.69 ***	66.56 ***	62.83 ***		75.01 ***	65.10 ***	70.30 ***	79.84 ***	50.09 ***
QuoteInt	82.65 ***	60.47 ***	63.63 ***	68.79 ***		81.32 ***	88.33 ***	68.11 ***	51.63 ***
Flick	63.54 ***	50.24 ***	50.62 ***	58.60 ***	73.96 ***		83.94 ***	50.55 ***	53.43 ***
SResp	72.35 ***	55.68 ***	59.17 ***	62.98 ***	83.19 ***	76.38 ***		54.10 ***	45.92 ***
SRuns	62.16 ***	42.73 ***	48.87 ***	71.63 ***	60.62 ***	40.98 ***	42.99 ***		32.99 ***
IOC	36.63 ***	32.41 ***	28.18 ***	42.13 ***	42.21 ***	45.13 ***	36.49 ***	24.35 ***	

Panel B: Cross-sectional average Pearson correlations across "true" HFT metrics (filtered for intraday deterministic patterns)

HFT proxy	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
Mess		61.95 ***	84.92 ***	73.05 ***	79.26 ***	70.60 ***	74.56 ***	64.14 ***	33.95 ***
Can	57.79 ***		67.43 ***	63.45 ***	50.30 ***	49.21 ***	53.07 ***	47.22 ***	34.87 ***
MonInt	82.43 ***	62.24 ***		62.51 ***	61.10 ***	59.22 ***	62.72 ***	49.78 ***	27.87 ***
FleetOrd	69.22 ***	58.90 ***	57.40 ***		63.08 ***	62.44 ***	62.73 ***	80.50 ***	38.03 ***
QuoteInt	76.19 ***	44.35 ***	56.25 ***	56.06 ***		74.01 ***	82.80 ***	68.14 ***	36.92 ***
Flick	64.85 ***	42.85 ***	53.19 ***	55.85 ***	62.97 ***		75.16 ***	55.13 ***	38.77 ***
SResp	67.82 ***	45.46 ***	55.42 ***	54.68 ***	76.07 ***	62.10 ***		51.78 ***	39.90 ***
SRuns	59.97 ***	40.10 ***	45.27 ***	72.22 ***	58.58 ***	46.01 ***	40.42 ***		27.15 ***
IOC	27.67 ***	24.92 ***	21.61 ***	30.73 ***	29.81 ***	28.32 ***	30.04 ***	18.53 ***	

TABLE IV
Correlations between HFT proxies (Cont.)

Panel C: Cross-sectional average Spearman correlations across HFT proxies (filtered for intraday deterministic patterns)

300 sec. (upper triangular matrix) and 30 sec. (lower triangular matrix) intervals

HFT proxy	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
Mess		71.55 ***	87.46 ***	81.50 ***	86.12 ***	70.82 ***	79.06 ***	65.21 ***	47.93 ***
Can	68.27 ***		71.69 ***	70.56 ***	70.40 ***	62.04 ***	67.17 ***	49.35 ***	48.62 ***
MonInt	86.35 ***	70.31 ***		71.72 ***	72.03 ***	61.54 ***	69.90 ***	56.42 ***	40.32 ***
FleetOrd	77.21 ***	62.99 ***	67.42 ***		77.68 ***	65.99 ***	72.64 ***	77.22 ***	52.18 ***
QuoteInt	84.36 ***	66.54 ***	70.96 ***	73.17 ***		81.62 ***	87.95 ***	60.59 ***	56.13 ***
Flick	68.80 ***	59.00 ***	60.52 ***	62.86 ***	78.18 ***		86.12 ***	47.31 ***	55.53 ***
SResp	65.16 ***	53.57 ***	57.82 ***	61.34 ***	73.77 ***	70.53 ***		52.44 ***	50.67 ***
SRuns	52.35 ***	35.91 ***	45.42 ***	60.99 ***	44.88 ***	35.27 ***	36.63 ***		31.90 ***
IOC	45.26 ***	41.40 ***	38.61 ***	47.93 ***	53.16 ***	51.23 ***	40.46 ***	22.85 ***	

Panel D: Cross-sectional average Spearman correlations across "true" HFT metrics (filtered for intraday deterministic patterns)

HFT proxy	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
Mess		62.07 ***	85.92 ***	75.52 ***	81.90 ***	75.11 ***	75.30 ***	62.95 ***	37.11 ***
Can	59.27 ***		65.74 ***	64.05 ***	57.80 ***	54.90 ***	57.55 ***	45.81 ***	39.98 ***
MonInt	84.92 ***	62.36 ***		66.18 ***	68.34 ***	66.27 ***	66.28 ***	52.48 ***	32.08 ***
FleetOrd	71.70 ***	56.50 ***	62.05 ***		68.61 ***	65.50 ***	65.06 ***	78.76 ***	41.53 ***
QuoteInt	79.26 ***	56.22 ***	67.15 ***	64.99 ***		78.82 ***	85.53 ***	59.98 ***	43.28 ***
Flick	73.18 ***	51.09 ***	65.03 ***	59.93 ***	71.47 ***		75.57 ***	52.91 ***	40.93 ***
SResp	61.76 ***	46.26 ***	54.58 ***	54.70 ***	70.64 ***	56.97 ***		49.66 ***	43.17 ***
SRuns	51.03 ***	33.24 ***	42.84 ***	62.47 ***	43.81 ***	39.72 ***	34.84 ***		26.10 ***
IOC	36.31 ***	34.47 ***	31.30 ***	39.82 ***	44.39 ***	33.83 ***	34.06 ***	19.08 ***	

TABLE V
Correlations between HFT proxies and true metrics

This table provides cross-sectional time series correlations between the *proxies* for HFT activity and the corresponding *true* metrics, computed at different levels of aggregation, from 30-second to 1500-second time windows. True metrics are computed using the actual HFTs' order flow as flagged in our database. We filter the series for intraday deterministic patterns by regressing the time series of each metric per stock on fifteen 25-minute interval dummies (the NSE continuous trading session starts at 9:15 and ends at 15:30). We use the residuals of those regressions to compute the correlations reported in this Table. The metrics are: the number of messages (*Mess*), where messages include submissions, revisions and cancellations of orders within 10 ticks from the prevailing best quotes; the number of cancellations (*Can*) of limit orders standing within 10 ticks from the prevailing best quotes; monitoring intensity (*MonInt*), defined as the sum of all limit order updates (revisions plus cancellations) as long as the orders are standing within 10 ticks from the best quotes; the number of fleeting orders (*FleetOrd*), where a fleeting order is a limit order that is either cancelled or revised in less than 100 milliseconds (ms) after its submission; quote intensity (*QuoteInt*) is the sum of all changes in the best ask and bid quotes or depth; speed of response (*SResp*) is the number of limit order book responses within 100 ms following NBBO quote improvements, and the number of strategic runs (*SRuns*), where a strategic run is a sequence of linked messages (revisions and cancellations plus resubmissions) that satisfy the following conditions: the average distance between subsequent messages for the same order ID must be less than or equal to 500ms, the size of the order must not increase, and orders with different IDs are linked if and only if they have the same size, direction, and the time between the cancellation of the first order and the submission of the second order is less than 100ms; immediate-or-cancel orders (*IOC*). Our sample consists of the 50 constituents of the NIFTY-50, the official market index of the National Stock Exchange (NSE) of India, for the period May to July 2015. ***, **, * indicate statistically different from zero at the 1%, 5%, and 10% levels, respectively.

Panel A: Pearson correlation

	Metric								
Interval	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flickering*	*SResp*	*SRuns*	*IOC*
30s	97.40 ***	77.22 ***	97.97 ***	92.34 ***	92.54 ***	74.23 ***	93.91 ***	98.96 ***	82.09 ***
	(1.89)	(13.98)	(1.71)	(7.19)	(3.97)	(9.22)	(7.69)	(1.71)	(11.13)
60s	97.45 ***	78.52 ***	98.08 ***	92.98 ***	92.42 ***	77.25 ***	94.70 ***	99.02 ***	81.04 ***
	(1.88)	(13.38)	(1.69)	(6.71)	(4.02)	(9.10)	(6.90)	(1.65)	(11.92)
300s	97.41 ***	80.33 ***	98.25 ***	93.73 ***	91.96 ***	82.01 ***	95.70 ***	99.09 ***	79.76 ***
	(1.90)	(12.16)	(1.56)	(6.20)	(4.08)	(8.27)	(5.21)	(1.71)	(12.68)
900s	97.27 ***	80.40 ***	98.28 ***	94.03 ***	91.16 ***	81.45 ***	95.82 ***	99.02 ***	79.63 ***
	(2.08)	(12.52)	(1.59)	(5.85)	(4.66)	(8.48)	(4.50)	(2.32)	(11.94)
1500s	97.19 ***	80.24 ***	98.27 ***	94.04 ***	90.87 ***	81.70 ***	95.72 ***	98.99 ***	79.36 ***
	(2.19)	(12.80)	(1.62)	(5.93)	(4.82)	(8.49)	(4.48)	(2.51)	(12.00)

Panel B: Spearman correlation

Interval	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flickering*	*SResp*	*SRuns*	*IOC*
30s	96.77 ***	78.21 ***	97.13 ***	92.28 ***	88.93 ***	67.94 ***	93.81 ***	98.02 ***	81.83 ***
	(2.51)	(10.12)	(2.36)	(6.22)	(5.41)	(6.20)	(6.03)	(1.93)	(9.28)
60s	96.92 ***	79.56 ***	97.40 ***	92.97 ***	89.72 ***	71.53 ***	94.84 ***	97.83 ***	84.14 ***
	(2.32)	(9.45)	(2.14)	(5.54)	(5.22)	(6.47)	(4.98)	(2.39)	(7.88)
300s	96.93 ***	80.39 ***	97.67 ***	93.62 ***	90.33 ***	77.70 ***	95.72 ***	98.00 ***	85.92 ***
	(2.13)	(9.10)	(1.85)	(4.67)	(5.06)	(6.63)	(4.11)	(2.98)	(6.37)
900s	96.72 ***	80.02 ***	97.74 ***	93.49 ***	89.71 ***	78.69 ***	95.69 ***	98.30 ***	85.42 ***
	(2.31)	(9.50)	(1.82)	(4.84)	(5.58)	(7.21)	(4.16)	(3.13)	(6.63)
1500s	96.60 ***	79.56 ***	97.72 ***	93.43 ***	89.36 ***	78.75 ***	95.50 ***	98.45 ***	84.82 ***
	(2.39)	(10.21)	(1.85)	(4.98)	(5.86)	(7.50)	(4.43)	(3.20)	(7.03)

TABLE VI
HFTs' liquidity taking and making

This table provides coefficient estimates of pooled regression models of trade-based HFT liquidity supply (or "making") and liquidity demand (or "taking") on *true* HFT activity metrics (Panels A and B) and HFT activity *proxies* (Panels C and D). Liquidity making is the percentage of trades in which HFTs provide liquidity. Liquidity taking is the percentage of trades in which HFTs demand liquidity. We regress each of these metrics on each HFT activity true metric or proxy at a time, including as control variables the first lag of the dependent variable, a dummy for the first 30-minutes of trading, and a dummy for the last 30 minutes of trading. The HFT activity metrics are: the number of messages (*Mess*); the number of cancellations (*Can*); monitoring intensity (*MonInt*), defined as the sum of all limit order updates (revisions plus cancellations); the number of fleeting orders (*FleetOrd*), that is, orders that are either cancelled or revised in less than 100 milliseconds (ms) after its submission; quote intensity (*QuoteInt*) is the sum of all changes in the best ask and bid quotes or depth; speed of response (*SResp*) is the number of limit order book responses within 100 ms following NBBO quote improvements, the number of strategic runs (*SRuns*), where a strategic run is a sequence of linked messages (revisions and cancellations plus resubmissions), and immediate-or-cancel orders (*IOC*). We report findings with 30-second and 300-second time windows. All continuous variables are winsorized at the 1% on the RHS of the distribution. We only report the estimated coefficients for the variable of interest. The model includes stock fixed effects, and we estimate double-clustered (by stock and time of day) standard errors (reported in parenthesis) and t-statistics (Thompson, 2011). Our sample consists of the 50 constituents of the NIFTY-50, the official market index of the National Stock Exchange (NSE) of India, for the period May to July 2015. ***, **, * indicate statistically different from zero at the 1%, 5%, and 10% levels, respectively.

Bar length		*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
Panel A: True HFT metrics - HFTs making liquidity										
30s	Coef.x100	16.94 ***	14.56 ***	13.16 ***	13.39 ***	19.41 ***	16.91 ***	1.10	7.71 ***	64.47 ***
	sd.	(0.55)	(0.49)	(0.56)	(0.47)	(0.76)	(0.53)	(0.98)	(0.47)	(0.73)
	t-stat	*30.56*	*29.64*	*23.31*	*28.52*	*25.45*	*31.96*	*1.12*	*16.38*	*88.67*
300s		12.69 ***	10.96 ***	9.82 ***	10.30 ***	14.95 ***	14.30 ***	11.40 ***	7.37 ***	47.91 ***
		(0.83)	(0.74)	(0.77)	(0.75)	(0.83)	(0.82)	(0.78)	(0.63)	(1.29)
		15.28	*14.77*	*12.71*	*13.82*	*18.06*	*17.41*	*14.56*	*11.68*	*37.26*
Panel B: True HFT metrics - HFTs taking liquidity										
30s	Coef.x100	22.28 ***	17.12 ***	18.17 ***	13.94 ***	27.97 ***	21.78 ***	10.77 ***	9.41 ***	1.04
	sd.	(1.09)	(0.70)	(0.81)	(0.78)	(1.46)	(1.16)	(1.71)	(0.60)	(0.71)
	t-stat	*20.38*	*24.48*	*22.33*	*17.76*	*19.19*	*18.77*	*6.29*	*15.64*	*1.47*
300s		27.85 ***	21.15 ***	22.46 ***	17.71 ***	35.04 ***	28.44 ***	33.38 ***	13.75 ***	6.82 ***
		(1.16)	(0.75)	(0.90)	(0.90)	(1.63)	(1.42)	(1.68)	(0.84)	(0.98)
		23.95	*28.10*	*24.88*	*19.58*	*21.53*	*20.05*	*19.83*	*16.33*	*6.99*
Panel C: HFT proxies - HFTs making liquidity										
30s	Coef.x100	16.78 ***	15.19 ***	14.04 ***	13.35 ***	16.39 ***	19.98 ***	0.95	8.04 ***	49.93 ***
	sd.	(0.53)	(0.54)	(0.58)	(0.46)	(0.57)	(0.63)	(0.99)	(0.48)	(0.74)
	t-stat	*31.87*	*28.32*	*24.31*	*29.13*	*28.75*	*31.69*	*0.96*	*16.74*	*67.69*
300s		11.79 ***	11.39 ***	10.50 ***	9.70 ***	9.71 ***	11.82 ***	11.16 ***	7.69 ***	30.43 ***
		(0.83)	(0.85)	(0.80)	(0.77)	(0.87)	(1.01)	(0.80)	(0.65)	(0.98)
		14.17	*13.43*	*13.10*	*12.53*	*11.13*	*11.72*	*14.02*	*11.91*	*31.03*
Panel D: HFT proxies - HFTs taking liquidity										
30s	Coef.x100	22.49 ***	19.75 ***	19.00 ***	15.91 ***	25.09 ***	28.80 ***	11.22 ***	10.10 ***	3.65 ***
	sd.	(1.11)	(0.76)	(0.85)	(0.78)	(1.28)	(1.20)	(1.67)	(0.60)	(0.76)
	t-stat	*20.35*	*26.05*	*22.40*	*20.48*	*19.60*	*24.04*	*6.71*	*16.92*	*4.84*
300s		27.50 ***	23.48 ***	23.38 ***	19.59 ***	29.68 ***	34.35 ***	33.89 ***	14.71 ***	9.24 ***
		(1.18)	(0.89)	(0.93)	(0.87)	(1.47)	(1.23)	(1.67)	(0.82)	(0.91)
		23.25	*26.48*	*25.12*	*22.55*	*20.21*	*28.02*	*20.29*	*17.85*	*10.17*

TABLE VII
HFTs' contribution to the limit order book

This table provides coefficient estimates of pooled regression models of nine metrics of HFTs' contribution to the liquidity supply in the limit order book (LOB) of the National Stock Exchange of India (NSE) on *true* HFT activity metrics (Panels A to C) and HFT activity *proxies* (Panels D to F). In Panels A and D, HFTs' contribution is measured by the percentage of time HFTs post the best ask or bid quote (or both); in Panels B and E, we use the time-weighted percentage of the depth at the best quotes that is provided by HFTs; in Panels C and F, we use the time-weighted percentage of the accumulated depth at the five best levels of the LOB that is supplied by the HFTs, weighted by time. We regress the corresponding metric for HFT' contribution to liquidity on each HFT activity true metric or proxy at a time, including as control variables the first lag of the dependent variable, a dummy for the first 30-minutes of trading, and a dummy for the last 30 minutes of trading. The HFT activity metrics are: the number of messages (*Mess*); the number of cancellations (*Can*); monitoring intensity (*MonInt*), defined as the sum of all limit order updates (revisions plus cancellations); the number of fleeting orders (*FleetOrd*), that is, orders that are either cancelled or revised in less than 100 milliseconds (ms) after its submission; quote intensity (*QuoteInt*) is the sum of all changes in the best ask and bid quotes or depth; speed of response (*SResp*) is the number of limit order book responses within 100 ms following NBBO quote improvements, the number of strategic runs (*SRuns*), where a strategic run is a sequence of linked messages (revisions and cancellations plus resubmissions); and immediate-or-cancel orders (*IOC*). We report findings with 30-second and 300-second time windows. All continuous variables are winsorized at the 1% on the RHS of the distribution. We only report the estimated coefficients for the variable of interest. The model includes stock fixed effects, and we estimate double-clustered (by stock and time of day) standard errors (reported in parenthesis) and t-statistics (Thompson, 2011). Our sample consists of the 50 constituents of the NIFTY-50 for the period May to July 2015. ***, **, * indicate statistically different from zero at the 1%, 5%, and 10% levels, respectively.

Bar length		*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
Panel A: True HFT metrics - HFTs' presence at the best quotes (% of time)										
30s	Coef.x100	20.72 ***	16.61 ***	16.53 ***	12.21 ***	27.84 ***	17.45 ***	9.25 ***	8.32 ***	12.57 ***
	sd.	(1.28)	(0.80)	(0.97)	(0.76)	(1.79)	(1.23)	(1.61)	(0.64)	(0.66)
	t-stat	*16.15*	*20.79*	*17.03*	*16.10*	*15.52*	*14.15*	*5.73*	*13.10*	*18.96*
300s		22.85 ***	18.98 ***	18.23 ***	13.99 ***	30.60 ***	23.02 ***	25.97 ***	11.49 ***	12.86 ***
		(1.43)	(0.89)	(1.14)	(0.92)	(2.05)	(1.53)	(2.00)	(0.95)	(0.78)
		15.96	*21.39*	*16.05*	*15.28*	*14.94*	*15.09*	*13.00*	*12.06*	*16.41*
Panel B: True HFT metrics - HFTs' contribution to the best quotes depth (%)										
30s	Coef.x100	24.29 ***	20.96 ***	20.36 ***	14.50 ***	32.41 ***	20.54 ***	13.32 ***	8.94 ***	15.98 ***
	sd.	(1.40)	(0.90)	(1.05)	(0.83)	(2.06)	(1.24)	(1.99)	(0.69)	(0.80)
	t-stat	*17.32*	*23.39*	*19.36*	*17.52*	*15.70*	*16.62*	*6.70*	*13.04*	*19.88*
300s		26.03 ***	22.84 ***	21.73 ***	15.54 ***	35.68 ***	25.71 ***	34.33 ***	11.91 ***	16.43 ***
		(1.45)	(0.88)	(1.14)	(0.87)	(2.19)	(1.50)	(2.24)	(0.89)	(0.77)
		18.01	*26.05*	*19.05*	*17.95*	*16.28*	*17.18*	*15.32*	*13.39*	*21.21*
Panel C: True HFT metrics - HFTs' contribution to the LOB depth (%)										
30s	Coef.x100	12.90 ***	10.11 ***	10.26 ***	6.91 ***	13.39 ***	10.26 ***	6.61 ***	4.85 ***	8.38 ***
	sd.	(1.07)	(0.68)	(0.76)	(0.54)	(1.21)	(0.96)	(0.93)	(0.46)	(0.47)
	t-stat	*12.08*	*14.80*	*13.47*	*12.76*	*11.08*	*10.71*	*7.12*	*10.42*	*17.71*
300s		20.82 ***	15.87 ***	16.57 ***	10.70 ***	20.62 ***	16.58 ***	19.50 ***	9.54 ***	10.20 ***
		(1.34)	(0.80)	(0.94)	(0.69)	(1.61)	(1.39)	(1.54)	(0.81)	(0.47)
		15.52	*19.79*	*17.59*	*15.43*	*12.78*	*11.93*	*12.66*	*11.81*	*21.62*

TABLE VII (Cont.)
HFTs' contribution to the limit order book

Bar length		*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
Panel D: HFT proxies - HFTs' presence at the best quotes (% of time)										
30s	Coef.x100	20.93 ***	19.24 ***	17.40 ***	13.90 ***	24.51 ***	25.39 ***	9.59 ***	8.97 ***	13.30 ***
	sd.	(1.32)	(0.96)	(1.02)	(0.87)	(1.63)	(1.47)	(1.59)	(0.66)	(0.70)
	t-stat	*15.88*	*19.94*	*17.01*	*16.06*	*15.08*	(17.27)	*6.05*	*13.69*	*19.14*
300s		22.69 ***	21.41 ***	19.18 ***	15.43 ***	25.00 ***	25.75 ***	26.36 ***	12.33 ***	12.36 ***
		(1.45)	(1.12)	(1.17)	(1.06)	(1.84)	(1.65)	(2.00)	(0.98)	(0.74)
		15.62	*19.14*	*16.42*	*14.57*	*13.62*	(15.62)	*13.21*	*12.55*	*16.72*
Panel E: HFT proxies - HFTs' contribution to the best quotes depth (%)										
30s	Coef.x100	24.50 ***	23.57 ***	21.25 ***	16.74 ***	28.70 ***	33.21 ***	13.52 ***	9.80 ***	16.50 ***
	sd.	(1.46)	(1.12)	(1.12)	(0.95)	(1.90)	(1.70)	(1.97)	(0.72)	(0.84)
	t-stat	*16.80*	*21.05*	*19.02*	*17.68*	*15.14*	(19.52)	*6.87*	*13.56*	*19.74*
300s		25.66 ***	24.66 ***	22.53 ***	17.39 ***	29.53 ***	33.35 ***	34.23 ***	13.00 ***	15.37 ***
		(1.49)	(1.23)	(1.18)	(0.99)	(1.99)	(1.66)	(2.25)	(0.92)	(0.72)
		17.21	*19.99*	*19.04*	*17.53*	*14.84*	(20.10)	*15.20*	*14.09*	*21.42*
Panel F: HFT proxies - HFTs' contribution to the LOB depth (%)										
30s	Coef.x100	12.35 ***	10.27 ***	10.29 ***	7.77 ***	12.05 ***	14.02 ***	6.64 ***	5.13 ***	8.71 ***
	sd.	(1.05)	(0.72)	(0.79)	(0.63)	(1.07)	(1.08)	(0.92)	(0.49)	(0.51)
	t-stat	*11.76*	*14.25*	*13.04*	*12.35*	*11.21*	(12.93)	*7.23*	*10.44*	*17.18*
300s		19.10 ***	15.40 ***	16.33 ***	11.42 ***	17.11 ***	18.65 ***	19.06 ***	9.96 ***	9.48 ***
		(1.31)	(0.91)	(0.98)	(0.81)	(1.42)	(1.25)	(1.52)	(0.84)	(0.51)
		14.54	*16.89*	*16.71*	*14.11*	*12.07*	(14.90)	*12.54*	*11.81*	*18.42*

TABLE VIII
Unusually high and low HFT activity

We examine the behavior of nine popular HFT activity metrics (*true* in Panels A and B and *proxies* in Panels C and D) in periods of unusually high or low liquidity supply and demand by HFTs. The true metrics are based on the actual order flow flagged in our database; proxies are computed from the total order flow. The HFT activity metrics are: the number of messages (*Mess*); the number of cancellations (*Can*); monitoring intensity (*MonInt*), defined as the sum of all limit order updates (revisions plus cancellations); the number of fleeting orders (*FleetOrd*), that is, orders that are either cancelled or revised in less than 100 milliseconds (ms) after submission; quote intensity (*QuoteInt*) is the sum of all changes in the best ask and bid quotes or depth; speed of response (*SResp*) is the number of limit order book responses within 100 ms following NBBO quote improvements, the number of strategic runs (*SRuns*), where a strategic run is a sequence of linked messages (revisions and cancellations plus resubmissions), and immediate-or-cancel orders (*IOC*). Each metric is regressed on a set of indicator variables that identify episodes in which either HFTs' contribution to liquidity supply and/or demand is unusually high or low. HFT liquidity demand is measured by the percentage of trades initiated by the HFTs. HFT liquidity supply is measured as either the percentage of time HFTs post the best ask quote, the best bid quote or both (Panels A and C) or the time-weighted percentage of the accumulated depth at the five best levels of the book that is provided by the HFTs (Panels B and D). An unusually high (low) HFT contribution occurs when the corresponding metric is above (below) the 75th (25th) percentile of its empirical distribution, per stock. Combining unusually high and low supply and demand, we generate four dummy variables that are of main interest: (Dlow, Slow) is a dummy variable that equals 1 when HFTs' liquidity demand and supply are both unusually low, 0 o/w; (Dlow, Shigh) is a dummy variable that equals 1 when HFTs' liquidity demand is unusually low but HFTs' liquidity supply is unusually high, and so on and so forth. As control variables we use dummies for the initial 30 minutes and last 30 minutes of the trading session (Open and Close, respectively). We use pooled regressions with stock fixed effects, and we double-cluster (by stock and time of day) standard errors and t-statistics (Thompson, 2011). We report our findings with 30-second time aggregation. All continuous variables are winsorized at the 1% on the RHS of the distribution. Our sample consists of the 50 constituents of the NIFTY-50, the official market index of the National Stock Exchange of India, for May to July 2015. ***, **, * indicate statistically different from zero at the 1%, 5%, and 10% levels.

Panel A: HFT true metrics (30s bars) - presence at the best quotes (% time)

	Mess	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
(Dlow, Slow)	-0.47 ***	-0.35 ***	-0.38 ***	-0.32 ***	-0.48 ***	-50.26 ***	-0.41 ***	-0.24 ***	-0.43 ***
(Dhigh, Slow)	-0.18 ***	-0.17 ***	-0.14 ***	-0.11 ***	-0.22 ***	-10.78 *	-0.12 ***	-0.13 ***	-0.38 ***
(Dlow, Shigh)	-0.27 ***	-0.25 ***	-0.24 ***	-0.23 ***	-0.17 ***	-42.04 ***	-0.15 ***	-0.13 ***	-0.38 ***
(Dhigh, Shigh)	0.43 ***	0.31 ***	0.32 ***	0.28 ***	0.66 ***	32.50 ***	0.62 ***	0.21 ***	-0.09 ***
Open	0.48 ***	0.33 ***	0.30 ***	0.47 ***	0.47 ***	94.43 ***	0.44 ***	0.63 ***	0.34 ***
Close	0.13 ***	0.10 ***	0.16 ***	0.06 ***	0.16 ***	34.46 ***	0.24 ***	0.10 ***	-0.05 ***
Intercept	-0.01 *	0.00	0.00	-0.01 ***	-0.04 ***	-3.72 ***	-0.08 ***	-0.04 ***	0.09 ***
R2	0.09	0.05	0.05	0.05	0.13	0.15	0.11	0.05	0.04
Obs.	2,418,750	2,418,750	2,418,750	2,399,929	2,413,186	2,418,750	2,231,221	2,418,750	2,418,750

Panel B: HFT true metrics (30s bars) -contribution to the LOB depth (%)

	Mess	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
(Dlow, Slow)	-0.44 ***	-0.31 ***	-0.36 ***	-0.28 ***	-0.43 ***	-47.06 ***	-0.39 ***	-0.21 ***	-0.36 ***
(Dhigh, Slow)	-0.13 ***	-0.11 ***	-0.12 ***	-0.07 ***	-0.01	-5.70 **	-0.02	-0.08 ***	-0.32 ***
(Dlow, Shigh)	-0.15 ***	-0.14 ***	-0.12 ***	-0.16 ***	-0.20 ***	-23.84 ***	-0.21 ***	-0.07 ***	-0.27 ***
(Dhigh, Shigh)	0.68 ***	0.50 ***	0.53 ***	0.44 ***	0.78 ***	55.10 ***	0.76 ***	0.34 ***	0.03
Open	0.45 ***	0.30 ***	0.27 ***	0.45 ***	0.44 ***	92.15 ***	0.41 ***	0.61 ***	0.35 ***
Close	0.16 ***	0.12 ***	0.18 ***	0.08 ***	0.17 ***	37.38 ***	0.24 ***	0.11 ***	-0.02
Intercept	-0.05 ***	-0.04 ***	-0.03 ***	-0.04 ***	-0.07 ***	-8.48 ***	-0.09 ***	-0.06 ***	0.04 ***
R2	0.10	0.05	0.06	0.06	0.12	0.15	0.11	0.05	0.03
Obs.	2,418,750	2,418,750	2,418,750	2,399,929	2,413,186	2,418,750	2,231,221	2,418,750	2,418,750

TABLE VIII (Cont.)
Unusually high and low HFT activity

Panel C: HFT proxies (30s bars) - presence at the best quotes (% time)

	Mess	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
(Dlow, Slow)	-0.51 ***	-0.41 ***	-0.41 ***	-0.37 ***	-0.52 ***	-50.58 ***	-0.43 ***	-0.25 ***	-0.45 ***
(Dhigh, Slow)	-0.21 ***	-0.20 ***	-0.16 ***	-0.13 ***	-0.27 ***	-13.68 ***	-0.12 ***	-0.14 ***	-0.37 ***
(Dlow, Shigh)	-0.34 ***	-0.30 ***	-0.27 ***	-0.27 ***	-0.31 ***	-37.88 ***	-0.17 ***	-0.14 ***	-0.44 ***
(Dhigh, Shigh)	0.41 ***	0.35 ***	0.33 ***	0.31 ***	0.52 ***	47.73 ***	0.62 ***	0.22 ***	-0.05 **
Open	0.56 ***	0.44 ***	0.36 ***	0.56 ***	0.63 ***	93.25 ***	0.47 ***	0.63 ***	0.49 ***
Close	0.23 ***	0.16 ***	0.18 ***	0.15 ***	0.41 ***	45.01 ***	0.23 ***	0.10 ***	0.07 ***
Intercept	-0.01 *	-0.01	0.00	-0.02 ***	-0.05 ***	-6.68 ***	-0.08 ***	-0.04 ***	0.07 ***
R2	0.11	0.07	0.06	0.07	0.14	0.17	0.11	0.05	0.06
Obs.	2,418,750	2,418,750	2,418,750	2,399,929	2,413,186	2,418,750	2,231,221	2,418,750	2,418,750

Panel D: HFT proxies (30s bars) - contribution to the LOB depth (%)

	Mess	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
(Dlow, Slow)	-0.46 ***	-0.36 ***	-0.38 ***	-0.32 ***	-0.45 ***	-45.89 ***	-0.41 ***	-0.22 ***	-0.37 ***
(Dhigh, Slow)	-0.12 ***	-0.09 ***	-0.11 ***	-0.07 ***	-0.06 ***	-2.24	-0.02	-0.08 ***	-0.29 ***
(Dlow, Shigh)	-0.21 ***	-0.20 ***	-0.15 ***	-0.21 ***	-0.29 ***	-30.76 ***	-0.23 ***	-0.08 ***	-0.33 ***
(Dhigh, Shigh)	0.64 ***	0.51 ***	0.53 ***	0.48 ***	0.67 ***	68.14 ***	0.74 ***	0.36 ***	0.09 ***
Open	0.53 ***	0.43 ***	0.33 ***	0.54 ***	0.61 ***	90.50 ***	0.44 ***	0.61 ***	0.50 ***
Close	0.26 ***	0.18 ***	0.21 ***	0.16 ***	0.43 ***	47.37 ***	0.24 ***	0.11 ***	0.10 ***
Intercept	-0.06 ***	-0.04 ***	-0.04 ***	-0.05 ***	-0.08 ***	-10.50 ***	-0.08 ***	-0.06 ***	0.02 ***
R2	0.11	0.07	0.07	0.07	0.13	0.18	0.11	0.06	0.05
Obs.	2,418,750	2,418,750	2,418,750	2,399,929	2,413,186	2,418,750	2,231,221	2,418,750	2,418,750

TABLE IX
Correlation between HFT proxies and other traders' activity

We examine whether HFT activity proxies used or suggested by the extant literature correlate with the activities of other trader types. This table provides cross-sectional time series correlations between alternative HFT activity proxies, computed using all messages, and the corresponding metrics computed using the messages of agency algorithmic traders (AATs - Panel A) and non-algorithmic traders (NATs - Panel B). We report results with 30-second and 300-second time aggregation windows. We filter the series for intraday deterministic patterns by regressing the time series of each metric per stock on fifteen 25-minute interval dummies (the NSE continuous session starts at 9:15 and ends at 15:30). We use the residuals of those regressions to compute the correlations reported here. The HFT metrics are: the number of messages (*Mess*); the number of cancellations (*Can*); monitoring intensity (*MonInt*), defined as the sum of all limit order updates (revisions plus cancellations); the number of fleeting orders (*FleetOrd*), that is, orders that are either cancelled or revised in less than 100 milliseconds (ms) after its submission; quote intensity (*QuoteInt*) is the sum of all changes in the best ask and bid quotes or depth; speed of response (*SResp*) is the number of limit order book responses within 100 ms following NBBO quote improvements, the number of strategic runs (*SRuns*), where a strategic run is a sequence of linked messages (revisions and cancellations plus resubmissions), and immediate-or-cancel orders (*IOC*). Our sample consists of the 50 constituents of the NIFTY-50, the official market index of the National Stock Exchange (NSE) of India, for the period May to July 2015. ***, **, * indicate statistically different from zero at the 1%, 5%, and 10% levels, respectively.

Panel A: Cross-sectional averge correlation between AAT metrics and HFT metrics

Bar size	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
30s	68.24 ***	40.47 ***	60.02 ***	50.67 ***	56.55 ***	59.41 ***	61.50 ***	31.37 ***	44.96 ***
	(7.86)	(8.50)	(8.40)	(7.27)	(6.87)	(8.03)	(14.16)	(13.03)	(5.75)
300s	66.56 ***	41.41 ***	59.52 ***	51.34 ***	56.80 ***	69.32 ***	70.28 ***	35.38 ***	51.65 ***
	(9.42)	(9.54)	(10.45)	(9.50)	(8.46)	(10.17)	(10.89)	(9.55)	(8.95)

Panel B: Cross-sectional averge correlation between NAT metrics and HFT metrics

Bar size	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
30s	34.02 ***	15.44 ***	19.58 ***	24.03 ***	26.84 ***	43.07 ***	32.98 ***	10.03 ***	5.03 ***
	(14.08)	(6.53)	(11.35)	(8.69)	(8.82)	(14.42)	(7.44)	(12.94)	(8.39)
300s	35.16 ***	22.54 ***	21.52 ***	32.16 ***	30.64 ***	52.79 ***	44.90 ***	9.05 ***	4.52 ***
	(17.24)	(10.42)	(13.25)	(13.96)	(12.46)	(19.48)	(10.56)	(8.86)	(4.08)

TABLE X
HFT proxies and trader types' uncorrelated activity

We summarize a two-stage regression model to study how the component of the high-frequency traders (HFTs), agency algorithmic traders (AATs), and non-algorithmic traders (NATs) activity that is uncorrelated with other traders' activities relate to the HFT activity proxies suggested by the extant literature. In the first stage, we run stock by stock regressions of a given HFT activity metric (computed using the actual HFTs messages, flagged in our database) on the equivalent metrics for AATs and NATs. The residuals measure the component of the HFT activity for that stock that is uncorrelated with the activity of other market participants. In the second stage, we run a pooled regression with stock fixed effects and double-clustered (by stock and time of day) standard errors of the residual HFT activity from the first stage on the corresponding HFT activity proxy (computed using the total message traffic), and include as controls dummies for the initial 30 minutes and last 30 minutes of the trading session. The same process is repeated for AATs and NATs activity metrics, but now in the first stage AATs (NATs) activity is regressed on HFTs and NATs (HFTs and AATs) activity. The activity metrics for each trader type are: the number of messages (*Mess*); the number of cancellations (*Can*); monitoring intensity (*MonInt*), defined as the sum of all limit order updates (revisions plus cancellations); the number of fleeting orders (*FleetOrd*), that is, orders that are either cancelled or revised in less than 100 milliseconds (ms) after its submission; quote intensity (*QuoteInt*) is the sum of all changes in the best ask and bid quotes or depth; speed of response (*SResp*) is the number of limit order book responses within 100 ms following NBBO quote improvements, the number of strategic runs (*SRuns*), where a strategic run is a sequence of linked messages (revisions and cancellations plus resubmissions), and immediate-or-cancel orders (*IOC*). We report our findings with 30-second windows (Panel A) and 300-second windows (Panel B). We report the coefficient of the variable of interest in the 2nd stage pooled regression, its standard deviation (in parenthesis), and the adjusted R^2 for the first and second stages. All continuous variables are winsorized at the 1% on the RHS of the distribution. Our sample consists of the 50 constituents of the NIFTY-50, the official market index of the National Stock Exchange of India (NSE), for the period May to July 2015. ***, **, * indicate statistically different from zero at the 1%, 5%, and 10% levels, respectively.

Panel A: 30-second bars

Trader type	Statistic	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
HFT	Coef.x100	42.77 ***	50.77 ***	56.97 ***	50.12 ***	58.69 ***	14.08 ***	40.49 ***	83.16 ***	55.12 ***
	t-stat	(27.67)	(25.19)	(33.13)	(31.64)	(26.70)	(13.99)	(27.65)	(65.33)	(43.46)
	R^2 (2nd stage)	0.34	0.29	0.48	0.35	0.44	0.05	0.29	0.79	0.35
	R^2 (1st stage)	0.47	0.12	0.34	0.31	0.24	0.50	0.43	0.13	0.13
AAT	Coef.x100	12.54 ***	53.87 ***	13.47 ***	21.06 ***	28.51 ***	19.19 ***	20.89 ***	10.01 ***	50.12 ***
	t-stat	(14.00)	(19.53)	(14.67)	(21.50)	(12.76)	(12.23)	(13.53)	(9.91)	(24.35)
	R^2 (2nd stage)	0.04	0.33	0.04	0.09	0.11	0.07	0.08	0.01	0.30
	R^2 (1st stage)	0.95	0.86	0.96	0.84	0.82	0.89	0.96	0.84	0.91
NAT	Coef.x100	5.04 ***	10.15 ***	4.30 ***	12.55 ***	3.97 ***	24.56 ***	8.43 ***	1.28 ***	2.80 ***
	t-stat	(10.51)	(12.74)	(10.20)	(14.67)	(8.28)	(24.43)	(5.78)	(6.74)	(7.23)
	R^2 (2nd stage)	0.02	0.01	0.00	0.04	0.03	0.10	0.01	0.00	0.01
	R^2 (1st stage)	0.65	0.53	0.57	0.52	0.65	0.76	0.80	0.29	0.26
	Obs.	2,419,500	2,419,500	2,419,500	2,413,936	2,400,679	2,419,500	2,231,961	2,419,500	2,419,500

TABLE X (Cont.)
HFT proxies and trader types' uncorrelated activity

Panel B: 300-second bars

Trader type	Statistic	*Mess*	*Can*	*MonInt*	*FleetOrd*	*QuoteInt*	*Flick*	*SResp*	*SRuns*	*IOC*
HFT	Coef.x100	39.62 ***	46.36 ***	53.65 ***	42.52 ***	51.61 ***	7.74 ***	29.82 ***	78.94 ***	44.76 ***
	t-stat	(21.60)	(22.04)	(27.18)	(23.46)	(19.63)	(6.84)	(20.66)	(49.02)	(25.17)
	R^2 (2nd stage)	0.31	0.26	0.45	0.28	0.39	0.03	0.21	0.75	0.25
	R^2 (1st stage)	0.51	0.19	0.39	0.39	0.35	0.67	0.58	0.17	0.20
AAT	Coef.x100	9.85 ***	47.41 ***	10.42 ***	17.86 ***	22.01 ***	12.66 ***	15.50 ***	9.26 ***	50.13 ***
	t-stat	(10.19)	(17.47)	(10.69)	(14.71)	(9.48)	(12.30)	(10.20)	(7.48)	(19.05)
	R^2 (2nd stage)	0.05	0.29	0.05	0.09	0.09	0.05	0.06	0.01	0.32
	R^2 (1st stage)	0.93	0.87	0.95	0.84	0.77	0.92	0.97	0.77	0.89
NAT	Coef.x100	5.08 ***	10.41 ***	3.71 ***	12.38 ***	1.87 ***	19.91 ***	6.47 ***	1.20 *	1.27 **
	t-stat	(9.68)	(10.33)	(8.34)	(14.37)	(2.73)	(16.06)	(4.52)	(1.74)	(2.15)
	R^2 (2nd stage)	0.02	0.02	0.00	0.04	0.04	0.08	0.00	0.00	0.01
	R^2 (1st stage)	0.68	0.57	0.63	0.61	0.66	0.80	0.82	0.24	0.29
	Obs.	241,950	241,950	241,950	241,449	241,950	241,950	240,910	241,950	241,950