Order Exposure in High Frequency Markets*

Bidisha Chakrabarty, Terrence Hendershott, Samarpan Nawn, Roberto Pascual*

November, 2017

Abstract

Theory predicts that uninformed traders hide limit orders to avoid free-option risk while informed traders hide to delay information revelation. Evidence from non-high frequency markets supports the free-option narrative. We advance the study of order exposure to high frequency markets. Using detailed data that identify hidden order placement by high-frequency traders (HFTs) vis-à-vis other algorithmic and non-algorithmic traders, we find that HFTs use small share sizes to hide orders near the best quotes. HFTs' hidden orders have shorter time to completion, higher fill rates, lower implementation shortfall, and overall lower information content. Collectively our results show that extant models do not explain the order exposure choice of HFTs and calls for new theory. In that direction, we test and find that compared to other trader groups, HFTs' aggressive hidden limit orders more often undercut standing orders at or near the best quotes.

Keywords: Iceberg orders, hidden volume, high frequency trading, limit order book

JEL Classification: G11; G12; G14; G15, G24

*: Bidisha Chakrabarty (bidisha.chakrabarty@slu.edu) is at the John Cook School of Business, Saint Louis University, St. Louis, MO 63108, U.S.A. Terrence Hendershott (hender@haas.berkeley.edu) is at the Haas School of Business, University of California, Berkeley, CA 94720-1900, U.S.A. Samarpan Nawn (samarpann12@iimcal.ac.in) is at the Indian Institute of Management, Joka, Kolkata, West Bengal 700104, India. Roberto Pascual (rpascual@uib.es) is at the University of Balearic Islands, Valldemossa Road, 07122 Palma, Mallorca, Balearic Islands, Spain

* This working paper is part of the NSE-NYU Stern School of Business Initiative for the Study of the Indian Capital Markets. The authors acknowledge the support of the initiative. The views expressed in this Working Paper are those of the authors and do not necessarily represent those of the National Stock Exchange of India (NSE) or New York University (NYU). We thank seminar participants at the Auckland University of Technology and Saint Louis University for useful comments.

1. Introduction

Financial markets are becoming more opaque.[1] While market opacity has been extensively studied (e.g., Bloomfield and O'Hara, 1999; Boehmer, Saar, and Yu, 2005; Hendershott and Jones, 2005) traders' motives for hiding orders have received less attention. Existing research on order exposure focuses on non-high frequency settings. We study traders' order exposure decision in high frequency markets. We test extant theories and find that while they explain the order exposure decision of non-high frequency traders (non-HFTs), they generally do not fit the patterns we observe in the hidden order usage of HFTs.

In order driven markets, liquidity is supplied via limit orders that provide a free option to trade (Copeland and Galai, 1983). If liquidity providers are uninformed, placing large limit orders exposes them to (a) picking off risk (e.g., Foucault, 1999), (b) adverse selection (e.g., Glosten, 1996), and (c) front running by parasitic traders (e.g., Harris, 1997). In such settings, hiding the size of orders (fully or partially) limits their option value (e.g., Buti and Rindi, 2013). Empirical evidence supports this view (Bessembinder, Payanides, and Venkataraman, 2009 (BPV henceforth); Pardo and Pascual, 2012). Theory models (Kaniel and Liu, 2006; Goettler, Parlour, and Rajan, 2009) suggest that informed traders may also use limit orders. In this case, they may use hidden limit orders to delay information revelation (Moinas, 2010) or limit the expropriation of informational rents (Boulatov and George, 2013). This narrative finds support in experimental market settings (Bloomfield, O'Hara, and Saar, 2015; Gozluklu, 2016).

The extant evidence, however, does not address markets populated by HFTs who are a significant source of liquidity supply in modern financial markets (Hagströmer and Nordén, 2013; Brogaard, Hendershott, and Riordan, 2014).[2] While studies examine their market-making role, the literature has not examined whether HFTs are contributing to the increasing opacity of financial markets by using hidden limit orders. Which raises the question: Should we expect HFTs to use hidden limit orders?

[1] According to the SEC, in January 2017, the total hidden volume in US exchanges, defined as the total volume of trades against hidden orders divided by the total volume of all trades was around 30%, compared to 15% in 2012. https://www.sec.gov/marketstructure/datavis/ma_exchange_hiddenvolume.html#.WfH5IRNSxE4. In addition, 18% of all trades happen in dark pools. https://www.sec.gov/news/statement/shedding-light-on-dark-pools.html. Orders may be completely hidden, as in the US markets, or partially hidden (iceberg orders), as in Euronext Paris, National Stock Exchange of India, etc. Throughout this paper, we use the terms hidden orders and iceberg orders interchangeably. Note that all results presented are based on iceberg orders.

[2] In the US, HFTs contribute over 50% of all equity trades (Brogaard, Hendershott, and Riordan, 2014). The Aite Group estimates that in 2015 algorithmic trading was 32 % of equity volume in Asia, 36% in Russia and 21% in Brazil (http://aitegroup.com/report/equities-market-structure-evolution-commoditization-system-reboot)

If HFTs are uninformed traders, the traditional reasons why these traders hide orders should not apply to HFTs. HFTs face lower order exposure risk compared to other traders. They use smaller share sizes (O'Hara, 2015) and can monitor markets in near-continuous time (Biais and Foucault, 2014). Instead of losing time priority by placing hidden orders, they should place displayed orders and quickly cancel or update their quotes as market conditions necessitate.[3] Some recent studies, however, conclude that HFTs are informed, or at least their trades carry information. This may be because they are better at interpreting and/or reacting to public information (Chordia, Green, and Kottimukkalur, 2015, Bechwitz, Keim, and Massa, 2015), they anticipate and trade on other investors' order flow (Hirschey, 2016), or their low latency operations enable them to front- and/or back run the trades of institutional traders (Angel and McCabe, 2013). If HFTs trade on (short-lived) information, then they should be less likely to use hidden orders since non-exposure costs time priority and delays execution speed. Thus, irrespective of whether HFTs are informed or uninformed traders, the above arguments indicate that HFTs should make little use of hidden orders.

To test this hypothesis, we need order-level data that (a) identify HFT, and (b) flag hidden orders. Commonly used data in the HFT literature do not have such identifiers. We have access to data from the National Stock Exchange of India (NSE) that provide rich details on trader accounts and identify them as proprietary algorithmic trader (i.e., HFTs), agency algorithmic traders (AATs), or non-algorithmic traders (NATs). The NSE allows traders to place iceberg orders and our data identify both the displayed and the hidden portions of each iceberg order. During our sample period, HFTs contribute about 33% of the total daily volume on the NSE.[4] Using these data, we find that HFTs make extensive use of iceberg orders. In large-cap firms 10.38% (9.83%) of all limit orders (share volume) submitted by HFTs are iceberg orders. For mid-cap firms 36.0% (34.42%) of all orders (share volume) by HFTs are iceberg orders. Corresponding numbers for small-cap stocks are 15.84% (15.23%).

The free-option risk theory suggests that large limit orders are more likely to be hidden. Using Euronext Paris data from April 2003, a period with little high frequency trading, BPV find that the concealment option is indeed advantageous for placing large orders. Buti and Rindi (2013) indicate that front running is higher when order size is larger, which could motivate traders to hide such orders. So we examine whether HFTs place large

[3] Hidden orders lose time priority in the order book. If two orders are placed at the same price, the displayed order will be placed ahead of the hidden order, even if the displayed order came in later.
[4] https://www.nseindia.com/research/content/1314_BS6.pdf

orders using the iceberg provision. And we find that they do not. In large-cap stocks, while NATs place large order sizes (1139.59 shares average) as iceberg orders corroborating the BPV finding, HFTs use much smaller iceberg orders (459.58 shares average). For displayed orders, the patterns reverse. HFTs use large displayed orders (1150.50) while NATs use smaller sizes (309.27). In fact, 76.28% (5.11%) of HFTs' hidden (displayed) limit orders in large firms are placed in the under-50-shares category. For mid and small firms, the percentage of hidden orders in the under-50-shares increases to 98.72% and 83.96% respectively. The pattern reverses for NATs. It is clear that HFT iceberg orders are concentrated in small sized orders.

A look inside the different layers of the order book reveals further contrasts between hidden and displayed order placement by HFTs and NATs. We analyze order placement in four layers of the book – better than the prevailing best quotes, at the best quotes, up to five ticks away from the best quote, and the rest of the book. We find that while 46.03% of HFTs hidden orders in large stocks are placed at or better than the best quotes, only 1.50% of their displayed orders are placed at or better than the best quotes. In fact, over 97.72% of HFT's hidden orders in large stocks are within the five best ticks while NATs place about 39.12% of their iceberg orders away from the five best ticks. HFTs hidden orders in small stocks are rarely placed away from the five best ticks. NATs show the opposite pattern, placing the bulk of their hidden orders far away from the best quotes. Similar results obtain for accumulated share volume.

It therefore appears that HFTs' order non-exposure patterns are different from what the prior literature documents. What explains these findings? Previous work (BPV, De Winne and d'Hondt, 2007) find that stock characteristics, order attributes, and prevailing market conditions are related to both the decision to hide orders, and the size of the order that is hidden. We use a logit regression to model the order exposure decision and a Tobit model for the amount of shares to expose, and find that although all three trader categories respond similarly, there are interesting differences. First, HFTs' reaction is stronger than the AATs' and NATs,' indicating that order, stock, and market characteristics impact the order exposure of HFTs more than they do for the other traders. Second, HFTs are more likely to hide orders when spreads are wider, which may be prompted by their use of hidden orders to compete for liquidity supply by offering incremental price improvement. For example, when spreads are binding at one tick, price improvement is not feasible. Third, HFTs make greater use of hidden orders towards the end of the trading sessions. The higher volume at closing may reduce the likelihood of hidden volume detection, which sophisticated HFT algorithms can exploit.

In follow up analyses we use an ordered logit model to test if the hidden orders placed by HFTs differ in their execution probability compared to hidden orders by other trader categories. We find that HFTs' use hidden orders effectively in that their hidden orders have a higher execution probability despite the fact that hidden orders lose time priority. We also examine the efficiency with which HFTs use hidden orders. To do so we model the time to full execution of hidden orders using survival analysis (BPV; Lo, MacKinlay, and Zhang, 2002). This analysis is particularly relevant in our context, since the NSE allows iceberg orders where each successive tranche of the order is displayed after the previous tranche executes fully, thereby mechanically inducing a protracted time to completion. Results show that even after controlling for differences in order size and aggressiveness, hidden orders placed by HFTs take shorter time to fully execute compared to both AATs and NATs.

So, contrary to expectations based on existing theory, HFTs use hidden orders extensively, with strategic placement in the order book grid, which increases their efficiency of execution. But at what cost? We next present some estimates of the costs that HFTs face in their hidden order execution. To compute execution costs, we apply the implementation shortfall metric of Perold (1988) to non-marketable limit orders. This metric has two components – an effective cost piece that is akin to price impact, and an opportunity cost component of non-execution. Our results show that HFTs face higher effective cost, especially for hidden orders. This is expected since HFTs use aggressive hidden limit orders which should have larger price impact. However, their opportunity cost of non-execution is lower indicating less adverse price movement after their hidden order submissions. When we combine these two components, the latter effect dominates and, overall, HFT hidden orders have a lower implementation shortfall.

Theory models of order exposure are conditioned on the information set of the traders who use such orders, as discussed earlier. In our next set of analyses we address the information content of HFTs' hidden orders. Here we present three sets of measures. First we compute the permanent price impact of each trader group's message traffic variables and show the impulse response functions of HFTs' versus the two other traders groups' hidden orders (as well as displayed orders, cancellations, and trades). We follow Brogaard, Hendershott, and Riordan (2017) and estimate an extended version of Hasbrouck's (1991a,b) Vector Autoregressive (VAR) model in event time. We find that trades have the highest permanent impact, with HFT trades having the largest impact among the three trader categories. HFTs' hidden orders, however, have the lowest permanent impact, consistent with HFTs using aggressive market orders (not hidden orders) to trade on time sensitive information. Second, we probe this issue further by decomposing the order-flow related component of the efficient price variance into shares attributable to each trader (HFT, AAT, NAT) and order type (displayed, hidden) combination, extending the approach in Hasbrouck (1991b). Consistent with our earlier result, here too we find that HFTs' hidden orders explain the smallest portion of order-flow related price variation. Our third and final results in this section are the information shares (Hasbrouck, 1995) of each trader/order type combination. We find that the information share of HFTs' hidden orders is lower than both AATs' and NATs'.

Collectively, our results indicate that while NATs' hidden orders – large orders hidden away from the best quotes – comports with existing evidence of order exposure choice (BPV), the pattern of hidden order use by HFTs is different. Existing models do not explain why HFTs should use such orders, which calls for new theory to explicitly model the order exposure choice of HFTs. To that end, and as a first step we empirically investigate one possible reason HFTs may use hidden orders – to undercut standing quotes and compete to supply liquidity without revealing their presence at the top of the order book.[5] We use a logit model to study this possibility and define an undercutting order as a limit order that (a) is placed immediately after another submission on the same side of the market, (b) comes in under 10 milliseconds of the previous order, and (c) improves the price of the previous one. We find that compared to AATs and NATs, HFTs are more likely to use hidden orders to undercut existing orders at or near the best quotes, after controlling for other determinants of the use of hidden orders.

To our knowledge, this is the first study to document that, in contrast to conclusions drawn from existing theories of order exposure, HFTs make extensive use of hidden orders. The NSE database with its detailed trader classification and iceberg order flags makes this inquiry possible. This database also provides some other key advantages. The NSE is a consolidated market handling over 80% of all equity volume. Since there are no dark pools operating in that market, traders who wish to hide their trading interest have to use the non-

[5] Offering minimal price improvement to undercut standing quotes and move up in the order queue may enhance liquidity supply and narrow the bid-ask spread, or adversely impact other liquidity suppliers if such quotes are used to persistently jump ahead of standing orders. Since HFTs do not have any fiduciary obligation towards the traders whose quotes they undercut, our tests we do not address the illegal practice of "front running," where the undercutting party has a fiduciary obligation to the party whose orders are undercut.

display option in this lit market.[6] HFTs are fairly active on the NSE, contributing about one in three trades. Unlike the US markets with no specific date when HFT was introduced, the period of HFT introduction on the NSE is clear, December 2009, when the NSE began colocation. Finally, the NSE is one of the top ten equity exchanges by share volume. [7] All of these factors make it a good setting in which to test the order exposure decision of HFTs.

We structure the paper as follows. Section 2 describes the institutional details of the NSE market, the identification of trader account types and a description of the sample. Section 3 provides descriptive statistics and univariate tests of hidden order use by trader categories, including limit order book (LOB) evidence. Section 4 examines the determinants of the order non-exposure decision by trader categories, and Section 5 provides results on implementation shortfall analyses, price discovery and information shares. Section 6 presents results on HFTs' hidden orders used for undercutting, and Section 7 concludes.

2. Institutional features of the NSE and sample selection

2.1 Trading protocol and iceberg orders

As of April 2016, more than 1300 listed securities traded on the NSE representing greater than 80% of the total domestic traded volume (SEBI, 2016).[8] Trading is completely automated and order driven with no designated market makers, similar to the Nasdaq (U.S.), Euronext (Paris), and the Xetra (Germany). The electronic LOB market operates on a price-exposure-time priority basis. Information about quoted prices and sizes and executed trades (price and size) are disseminated by the exchange on a continuous time basis, with traders able to view the five best bid and ask quotes in real time. The market opens with a call auction that runs for 15 minutes, after which trading proceeds using a continuous order matching system.

Like many other stock exchanges, the NSE allows traders to hide a part of the order volume by choosing an iceberg option when entering the order.[9] The minimum exposure for any incoming order is 10% of the total volume. Once that portion is executed, another 10%

[6] Degryse, Tombeur, and Wuyts (2015) study the substitutibility between hidden orders in lit venues and dark pools in the Dutch market and find that dark trading negatively impacts hidden order trading, but not the other way around. Since hidden orders in lit venues are detectable, they may not be a good substitute for dark trading.
[7] https://www.world-exchanges.org/home/index.php/statistics/annual-statistics
[8] See SEBI Bulletin at http://www.sebi.gov.in/cms/sebi_data/attachdocs/1463726488005.pdf
[9] Aitken et al. (2001) show that 28% of trading volume on the Australian Stock Exchange is hidden. Hasbrouck and Saar (2009, 2013) find that, in US markets, 15% to 20% of the orders are executed against hidden volume. De Winne and d'Hondt (2007), and Bessembinder, Panayides, and Venkataraman (2009) show that around 45% of the order volume on Euronext is hidden.

(of the original order volume) is automatically displayed. Orders are prioritized based on price, exposure, and then time. Thus, at any price point, only the lit portion of the iceberg order will be filled and then other displayed orders in the queue at the same price point but entered later receive priority. The hidden portion of an earlier order is filled only after an incoming order has exhausted all displayed size at that price, including orders that arrive after the hidden order was submitted. Thus, the iceberg order provision of the NSE is identical to that used on the Euronext (BPV) and unlike the INET trading platform of Nasdaq in the U.S. which allows traders to fully hide an incoming order.

2.2 Trade and quote data

We obtain the trade and quote data from two daily files that the NSE provides for each day's message traffic. One of these files contains every message for each stock that traded that day including the ticker symbol, price, quantity and timestamp in jiffies (one jiffy is 33.3564 picoseconds or $(1/2\hat{}16)^{th}$ of a second). The message traffic includes order entry, modification, execution and cancellation events. Messages also include information about the *Client* and *Order Entry Mode* flags and information about other order modifier conditions, such as iceberg features (if any), stop loss price (if any), etc. The other (smaller) file contains analogous information for each trade.

2.3 Trader category identification in the NSE data

The message traffic data identify the account types of the traders that operate in the NSE. The data have three *Client* account classification types – *Custodian*, *Proprietary* and *Others*. The *Custodian* flag is used for traders who are members of the exchange but do not conduct their own clearing or settlement. Primarily this group comprises of foreign institutional investors, mutual funds, and financial institutions. The *Proprietary* flag applies to members of the exchange who trade for their own proprietary accounts. Interestingly, this group often functions as voluntary intermediaries (i.e., market makers) at the exchange. Finally the *Other* flag applies to all other customers of the exchange who employ their own clearing member. This group includes domestic corporations and retail traders, among others.

In addition to the trader category identification, the data also provide an additional flag for the *Order Entry Mode* used to interact with the NSE's limit order market. The flag for *Algorithmic Trader* applies if order entry and management is done using an algorithm; a *Non-Algorithmic Trader* flag applies if a trader uses manual order entry and management. The intersection of the three *Client* types with the two *Order Entry Modes* enables us to identify six distinct trader categories. Our particular focus in this study is on the *Proprietary* client using *Algorithmic* order entry mode to trade on their own account. That is the definition of HFTs, which we are able to cleanly identify in our data. We group other traders who use *Algorithmic* order entry into the agency algorithmic trader (AAT) category and all traders who do not use *Algorithmic* order entry mode as non-algorithmic traders (NATs). We present our results by these three groups – HFTs, AATs, and NATs.

2.4 Sample selection

Our study is about the order exposure decisions of HFTs and the literature shows that HFTs have a greater propensity to trade large stocks (Brogaard et al. 2014). To ensure even consideration of both HFTs and non-HFTs, we select a (market cap) stratified sample of 100 stocks as follows. We begin with the 1254 listed stocks in the NSE in September 2013, filter out 286 stocks that are not in continuous trading session in our sample period October to December 2013 (61 trading days). We also exclude firms that (i) have a closing price of Rs. 1 or lower, (ii) have fewer than 100 trades per day on average, (iii) trade less than 1000 shares a day, (iv) have a traded value per day of less than Rs. 100000 over the sample period, (v) have market-cap values in the Bloomberg and CMIE Prowess databases that diverge by over 10%, (vi) are involved in NSE or MSCI index changes. These filters reduce our universe of stocks to 695. We sort these stocks by their market capitalization and group them into deciles. From each decile we select 10 stocks to generate the sample of 100, with 30 large-cap stocks, 40 mid-cap stocks and 30 small-cap stocks. All company information come from the CMIE Prowess (analogous to Compustat), a database of Indian firms which covers approximately 80% of the NSE stocks (Kahraman and Tookes, 2017). Panel A of Table 1 shows the descriptive statistics of our sample.

[Insert Table 1 here]

The average firm in our sample has over 448 billion rupees market capitalization (about 7 billion USD per the exchange rate on 06/2017). The large-cap firms have a market capitalization of about 1465 billion rupees (22 billion USD), which is smaller than the large cap firms in the NASDAQ HFT dataset, where the average large cap firm is valued at 52.47 billion USD (Brogaard et al., 2014). Volume and number of trades are higher, and relative spread (ratio of the quoted spread to the quote midpoint) is much smaller for the large firms than mid-sized and the small firms, as expected. While both the accumulated displayed and hidden depths in the LOB are higher for large firms than mid- and small-sized firms, the differences are larger for displayed than for hidden depth.

To benchmark our direct identification of HFTs against much of the literature that uses proxies for HFT activity, in Panel B of Table 1 we report message traffic and cancellation statistics by trader categories and across the three market cap groups. Comparing across each row, we see that HFTs account for much greater message traffic (defined as the sum of submissions, cancellations, and revisions) either than the AATs or the NATs in the large cap stocks, but not in the mid-sized or the small stocks. However, when we scale message traffic by the number of trades executed, HFTs show a bigger presence even in the mid- and small-cap firms. This preponderance of HFTs to generate large message traffic volume echoes similar findings from the US equity markets (Hendershott, Jones, and Menkveld, 2011).

3. Hidden order use by trader category

3.1 Relative importance of Iceberg orders

In this section, we provide an in-depth look at iceberg order use by the different trader categories – HFTs, AATs and NATs. We begin with an examination of the relative importance of hidden versus displayed orders for these three trader groups. To do so, we examine the placement of displayed versus hidden orders in the LOB by constructing the LOB for each stock following the procedure outlined in Appendix A. We then compute the accumulated displayed and non-displayed depth, both in the number of orders and in share volume. Table 2 reports the results.

[Insert Table 2 here]

In Panel A we show the proportion of iceberg limit orders (ILOs) relative to all limit orders submitted, both for the number of orders and the volume of shares. Comparing across the first row, 10.38% (9.83%) of all orders (volume) submitted by HFTs in large cap stocks are ILOs. Although HFT message traffic is largest in the large cap stocks (Panel B of Table 1), we find that HFTs' use of ILOs is greater for mid-cap stocks. They place 36% (34.42%) of all orders (share volume) as ILOs in mid-cap stocks. In Panel B we show each trader category's share of both displayed limit orders (DLOs) and ILOs. HFTs account for 34.67% of DLOs but only 9.28% of ILOs in the large stocks. NATs, on the other hand, place 36.07% of their limit orders in large stocks as displayed and 66.29% as iceberg orders.

We next examine the order size distribution of displayed and iceberg orders by trader category. This investigation is motivated by the fact that prior literature shows hidden orders to generally be large sized (BPV). We define trade size categories in total shares for both displayed and hidden orders and use the two-sample Kolmogorov-Smirnov (Massey, 1951) test to compare the order size distributions of ILOs and DLOs submitted by the different trader categories. Table 3 shows the hidden and displayed order sizes placed by HFTs, AATs and NATs for large cap (Panel A), mid cap (Panel B) and small cap (Panel C) firms.

[Insert Table 3 here]

In Panel A (large cap firms), for example, the 76.28% under HFTs for ILOs indicate that 76.28% of HFT's iceberg orders are placed in the under-50-shares size category. By comparison, HFTs place only 5.11% of their displayed shares in this smallest share-size category and use larger share sizes when they fully expose their trading interest. Looking across the same row, we find that the pattern reverses for the NATs. These traders place more (65.99%) of their displayed shares and less of their iceberg shares (29.13%) in this smallest size-category. Looking down each column, we find that the largest proportion of HFT's ILOs are in the smallest size category and this declines steeply as we move up to larger share brackets, with the largest (over 2500 shares) category receiving only 0.05% of the total ILOs. Displayed shares, on the other hand, are more concentrated around the middle three categories (100-200, 200-500, and 500-1000 shares). NATs, by contrast, show a similar concentration around the middle order-size categories, but for their iceberg orders.

In mid-cap (Panel B) and small cap (Panel C) firms, HFTs place the majority of their iceberg orders – 98.72% and 83.96% respectively – in the smallest (under 50) share size category. The corresponding numbers for NATs – 31.53% and 22.77% - show that the NATs do not hide as much of their orders in small share sizes. So in the use of order sizes as well, we find a stark contrast between the HFTs and the NATs. While the NAT's order size choice for hiding their trading interest is consistent with previous literature, HFTs behave in quite the opposite way.

In each panel we also report the average size of iceberg and displayed orders. Previous work on hidden order exposure (e.g., BPV) lead us to expect that traders who wish to make large liquidity-motivated trades take advantage of the hidden order option. We find that while NATs behave in this expected fashion, placing large order sizes (1139.59) as iceberg orders compared to their smaller displayed orders (309.27) in large stocks, HFTs do the opposite. They use large displayed orders (1150.50) and comparatively smaller (459.58) iceberg orders.

These small-sized hidden orders placed by HFTs bear out O'Hara's (2015) prescient summing up of the relationship between HFTs, small trades, and the ability to conceal trading interest that "small trade sizes reflect the influence of HFTs because [these] "silicon traders" can spot (and exploit) human traders by their tendency to trade in round numbers, [and] all trading is converging to ever smaller sizes and is being hidden whenever possible."

3.2 Disaggregated look at the layers of the order book

Position in the limit order queue is valuable. While Hoffman (2014) refers to this as "time is money," Moallemi (2014) models the value of positions in the limit order queue. For HFTs, whose profits depend on being the fastest, the position in the limit order grid is of paramount importance. Hence, in this section, we examine where HFTs place their orders in the LOB. For these analyses, we build the order book at every order submission time and identify the position of order placement at four layers – price improving or better than the standing best bid and ask quotes ("Better"), the best bid and ask ("At"), up to the first five ticks from the best bid and ask ("Near") and the rest of the book ("Far"). Table 4 presents hidden and displayed order placement by the different trader categories across the three firm size groups.

[Insert Table 4 here]

Comparing along corresponding cells in Panels A and B of this Table, we find that while 25.11% of HFTs hidden orders in large stocks are placed at "Better" than the best quotes, less than 0.5% (0.47% in Panel B) of their displayed orders are price-improving. Within Panel A, we find that while 97.72% (25.11% + 20.92% + 51.68%) of HFT's hidden orders in large stocks are within the five best ticks, the comparable fraction for NATs is 65.70%. In fact, in all three firm size groups, HFTs place a greater proportion of hidden orders at or better than the best quotes. For the small stocks, HFTs rarely place any hidden orders away from the five best ticks. NATs show the exact opposite pattern, placing the bulk of their hidden orders far away from the best quotes.

For displayed order placement, shown in Panel B, we find the opposite pattern. Both HFTs and NATs place a bigger proportion of their displayed orders away from the best quotes. While HFTs use both the near and far regions of the LOB to place displayed orders, NATs concentrate their displayed orders mostly far from the best quotes. These results are mirrored by the share volume placement. Non-parametric tests show that the difference in ILO and DLO use is significant for all three trader categories.

[Insert Figure 1 here]

Figure 1 plots the estimated cross-sectional daily average probabilities of hidden order submission by HFTs, AATs, and NATs, conditional on the order size and aggressiveness, for the large cap stocks. It is clear that while HFTs (Fig. 1a.) have a higher probability of placing small sized hidden orders at all distances from the best quotes (at), they have the highest likelihood of placing such orders at the best quotes, followed by near the best quotes. Their use of hidden orders of larger size is significantly less. The pattern is the reverse for both AATs (Fig. 2b.) and NATs (Fig. 2c.), who use larger hidden order sizes and further away from the best quotes.

4. Determinants and efficiency of order exposure

De Winne and d'Hondt (2007) model the order (non)-exposure decision of traders on Euronext Paris as a function of the prevailing market conditions such as depth in the LOB, bid-ask spread, time of the day, as well as to order characteristics such as price aggressiveness and the total order size. BPV add to these factors and examine both the decision to hide, as well as the amount (of shares) to hide. We follow the more comprehensive approach of BPV and model each trader category's order exposure decision using logistic regressions, and the amount of shares to hide using Tobit regressions.

We run separate models for HFTs, AATs and NATs and run stock specific regressions on an order-by-order basis and report the cross-sectional average estimates of the variables. The *t*-statistic for testing the significance of each variable is computed using the Chordia, Roll, and Subrahmanyam (2005) method. This method accounts for possible cross-correlations in the individual stock regressions. Assuming that the pairwise residual correlations are constant across stocks, Chordia et al. (2005) show that the usual standard error of the aggregate estimate is inflated by a factor $[1+(N\text{-}1)\rho]^{0.5}$, where N is the number of stocks and ρ is the common cross correlations. Since order arrival times vary across stocks, the regression residuals are not synched in time. To address this, we measure the average residual for each stock over 15-minute periods, and estimate ρ as the average of 580 pairs of cross-correlation. The set of common explanatory variables for both the models include attributes of the incoming order (aggressiveness, order size), state of the order book (same and opposite side depth, LOB imbalance, relative spread etc.) and prevailing market conditions (volatility, trading frequency etc.). We estimate model [1] below:

$$
\begin{aligned}
Y_i = {} & \alpha_0 + \alpha_1 Aggr_i + \alpha_2 Ordersize_i + \alpha_3 Spread_i + \alpha_4 DepthSame_i + \\
& +\alpha_5 DepthOpp_i + \alpha_6 Volatility_i + \alpha_7 Waittime_i + \alpha_8 TradeFreq_i + \\
& +\alpha_9 HiddenSameSide_i + \alpha_{10} LOBImbalance_i + \alpha_{11} TradeSize_i + \\
& +\alpha_{12} MktVolatility_i + \alpha_{13} Lasthalfhourdummy_i + \acute{U}_i
\end{aligned}
\qquad [1]
$$

where for the logit model $Y_i = 1$ if i^{th} order is hidden, 0 otherwise; for the tobit model Y_i = hidden shares in the i^{th} order divided by avg. trading volume. Results are reported in Table 5.

[Insert Table 5 here]

In both models we focus only on limit orders and exclude marketable and market orders, since the exposure decision is relevant for traders submitting limit orders that wait in the order book instead of being executed immediately. This is similar to the methodology in BPV to facilitate a comparison of the results.

The empirical specifications for the independent variables (in both panels) capture the state of the LOB (inside spread and displayed depth, cumulative order book imbalance, and revelation of hidden orders at the inside quotes), trading conditions for each stock (volatility, trading frequency, and waiting time between recent order arrivals), order attributes (price aggressiveness and order size) and control variables (market volatility and time-of-the-day effects). For comparability across stocks, we normalize order size and trade size by dividing the actual observations by the stock's average daily trading volume. Appendix B lists the definitions of all variables used in each table. For this and all following tables, the estimation sample consists of data for December 2013, and only includes the 30 largest stocks in our full sample (in which HFTs are reasonably active) to ensure adequate number of observations for the models to converge.[10]

In Table 5, both Panels A and B, we first note that most of the order attribute and market condition variables are useful in explaining traders' exposure decisions, both in terms of whether to hide an order as well as how much to hide. More importantly, unlike in the previous analysis, most of these decision variables have the same direction and significance for all trader categories, and are consistent with the results in BPV (see their Table 5).

[10] For the subsample of the 30 largest stocks in our sample, iceberg limit orders represent 15% of the total volume (12.3% of all non-marketable limit orders) submitted across all stock-days.

The positive and significant coefficient on price aggressiveness (in both panels) indicate that all three categories of traders show an interest in assuming positions before their private information becomes public. Thus they place orders closer to the prevailing best quotes, but hide them so as not to expose their trading interest. Notably, the coefficient is much larger (2865.76) for HFTs than for AATs (511.34) and NATs (65.77). The positive sign on relative spread for HFTs indicates that they choose to hide their orders when the bid-ask spread is wide, consistent with protecting themselves from high adverse selection risk. This result aligns with the findings in BPV. AATs and NATs, in contrast, show a negative albeit weak coefficient on relative spread, which reflects the findings in De Winne and d'Hondt (2007). Hidden orders are less used by all three trader categories when depth at the best quote on the same side is greater, most likely reflecting the fact that a longer same side depth costs time priority, in which case a hidden order would be pushed to the back of the queue. HFTs show a negative relationship of the waiting time between order arrivals and the decision to hide an order, which is the opposite of the results for NATs. Like BPV, we interpret this as slower order arrival rate implies a lower likelihood that a subsequent order arrives at the same price, so that the loss of time priority due to order non-exposure is less costly. While HFTs successfully use this market characteristic to place hidden orders, NATs show the opposite behavior both for their decision to hide as well as for the amount of shares to hide. Overall, our results are consistent with BPV and De Winne and d'Hondt (2007) and reflect that while HFTs choose very different sizes for hidden orders and layers of the LOB in which to place them, they react similarly to stock characteristics, order attributes, and market conditions as identified in the previous literature.

To examine the likelihood of ILO execution, specifically contrasting HFTs with the other trader groups, we next estimate an ordered Logit model, as in Ranaldo (2004) and Pascual and Veredas (2009). The regression equation [2] is:

$$\begin{aligned} EXEC_i = {} & \alpha_0 + \alpha_1 Aggr_i + \alpha_2 Ordersize_i + \alpha_3 ILO_i + \alpha_4 HFT_i + \alpha_5 AAT_i + \alpha_6 ILOHFT_i + \\ & + \alpha_7 ILOAAT_i + \alpha_8 Spread_i + \alpha_9 DepthSameSide_i + \alpha_{10} DepthOpp_i + \\ & + \alpha_{11} LOBImbalance_i + \alpha_{12} Lasthalfhourdummy_i + \alpha_{13} OrderImbalance_i + \\ & + \alpha_{14} TradeFreq_i + \alpha_{15} Momentum_i + \alpha_{16} Volatility_i + \acute{U}_i \end{aligned} \qquad [2]$$

where the dependent variable (EXEC) is an ordinal variable that takes three possible values: EXEC = 1 indicates that the limit order is cancelled before execution; EXEC = 2 indicates that the limit order is partially executed and then cancelled; EXEC = 3 indicates that the limit order is fully executed. We exclude market and marketable limit orders and drop fleeting orders (Hasbrouck and Saar, 2009), because they are not intended to be executed. Revisions of non-executed orders are treated as the same order while revisions of partially-executed orders are treated as new submissions. Appendix B lists all other variable definitions. The model is estimated on a stock-by-stock basis with the coefficients and significance levels aggregated based on Chordia, et al. (2005). Table 6 reports the results.

[Insert Table 6 here]

The buy and sell limit orders in the two column show consistent results. The coefficient of interest is the dummy on ILO*HFT, which shows the execution probability of a hidden order placed by HFTs after controlling for all covariates found to affect hidden order placement (in Table 5) as well as trader categories (dummy for AAT and appropriate interactions are included to control for trader categories). Hidden orders placed by HFTs have a positive and significant coefficient for both buy (2.58) and sell (1.73) orders, indicating that HFTs use hidden orders effectively so that these orders, which lose time priority per the exchange trading rules, still have a higher execution probability. In Panel B we show the execution probability of ILOs (and DLOs) by HFTs versus AATs and NATs. We find that compared to AATs and NATs, HFTs have higher rate of execution of their orders, both displayed and non-displayed, submitted beyond the best quotes. Perhaps this suggests HFTs' ability to anticipate order flow and short-term order imbalances.

To complement the previous analysis on execution probability, we also examine the time to full execution of hidden orders placed by HFTs using survival analysis. Survival analysis can accommodate an important feature of limit order execution times: censored observations. If an order is cancelled 30 minutes after submission, then apparently it provides little information about the execution time, but the fact that it survived for 30 minutes is useful information. Such information contained in non-executed orders is used survival analysis. We model the determinants of execution of buy and sell limit orders separately and report the cross-sectional average estimates of the variables. The t-statistic for testing the significance of each variable is computed using the Chordia et al. (2005) method. We estimate the following model:

$$\begin{aligned} TIME_i = {} & \alpha_0 + \alpha_1 MQLP_i + \alpha_2 LastTradeBuyDum_i + \alpha_3 DepthSame_i + \alpha_4 DepthSame^2{}_i + \\ & + \alpha_5 DepthOpp_i + \alpha_6 OrderSize_i + \alpha_7 TradeFreq_i + \alpha_8 RelTradeFreq_i + \\ & + \alpha_9 ILO_i + \alpha_{10} * HFT_i + \alpha_{11} * AAT_i + \alpha_{12} * ILOHFT_i + \alpha_{13} * ILOAAT_i + \acute{U}_i \end{aligned} \quad [3]$$

where $TIME_i$= time to full execution of the i^{th} order, or the time survived in the book for a cancelled or expired order, with a positive censorship dummy.

The model covariates are the same as in the previous analysis, and control for stock, order book, and market conditions, as well as the order placement strategy of the other trader groups. As in the previous analysis, we exclude market and marketable limit orders and also filter out fleeting orders. Revisions of non-executed orders are treated as the same order while revisions of partially-executed orders are treated as new submissions. The econometric specifications follow BPV and Lo et al. (2002) and model an accelerated failure time specification of limit order execution times under the generalized gamma distribution. The models are estimated on a stock-by-stock basis, and we report aggregated coefficients and significance levels based on Chordia, et al. (2005). We report the results in Table 7.

[Insert Table 7 here]

As in Table 6, here too the buy and sell limit orders in the two columns show consistent results. The coefficient of interest is the dummy on ILO*HFT, which shows the time to full execution of an iceberg order placed by HFTs after controlling for all covariates found to affect ILO placement (in Tables 5 and 6) as well as trader categories (dummy for AAT and appropriate interactions are included to control for trader categories). Iceberg orders placed by HFTs have a negative and significant coefficient for both buy (-3.61) and sell (-2.76) orders, indicating that HFTs' hidden orders take shorter time to fully execute compared to AATs (the ILO_AAT dummy is also negative but about half the magnitude compared to HFTs). It is interesting that the intercept, which captures the effect for the residual trader group – NATs – is positive and significant. The combined results from Tables 7 and 8 show that HFTs manage their hidden orders such that they have a higher probability of, and lower time to, execution.

5. Implementation shortfall and information content of hidden orders

So far we have documented that HFTs efficiently place their hidden orders so that their time to execution is lower and execution probability is higher. But at what cost? We next examine the costs HFTs face in their hidden order execution. To compute execution costs, it is important to note that iceberg orders are single (or parent) orders that are broken up into a sequence of smaller (child) orders. As the parent orders are executed, they are recorded in the data as multiple smaller transactions in a correlated sequence of orders.

However, as Perold (1988) pointed out, the cost incurred by the trader is not a function of a single transaction but rather the entire sequence of child orders. To accommodate this order splitting in cost computation, Perold (1988) introduced the "implementation shortfall" metric to measure transaction cost for the parent order. Implementation shortfall compares the value of a paper portfolio with no transaction costs to the real portfolio obtained by actual trading and has been used in empirical work by Keim and Madhavan (1997), BPV, and Engle, Ferstenberg, and Russell (2012), among others.

5.1 Implementation shortfall

We evaluate the effective costs of execution and the opportunity costs of non-execution costs of ILOs and DLOs using the implementation shortfall (ISF) approach of Perold (1988). ISF is the sum of effective cost of execution or price impact (*PRI*) and opportunity costs of non-execution (*OPC*):

$$ISF = PRI + OPC = \kappa s\left(\bar{p} - q_0\right) + \left(1-\kappa\right)s\left(q_c - q_0\right) \quad [4]$$

The *PRI* component for a buy order is the difference between the average execution price ($\bar{p}$) and the mid-quote at the time of order submission (q_0), multiplied by the amount of shares executed (κs), where *s* is the order size (in shares) and κ is the fill rate of the order. The *OPC* for a buy order is the difference between the closing price on the day of order submission (q_c) and q_0, multiplied by the unexecuted part of the order $\left(1-\kappa\right)s$. Metrics for sell orders are analogously computed but conveniently signed.

Results are based on non-marketable limit orders. We exclude market, marketable limit, and fleeting orders (Hasbrouck and Saar, 2009). Revisions of standing limit orders are common in our data so we treat revisions of non-executed orders as the same order. In such cases, the *ISF* is computed using *s* as the order size after the last revision. Revisions of partially-executed orders are treated as new submissions. After computing the *ISF*, *PRI*, and *OPC* for each order, we regress them on order attributes, market conditions during the 30 minutes prior to order submission, and trader-category dummies. We estimate the model [5] using OLS with White-robust standard errors, on a stock-by-stock basis,

$$\begin{aligned} ISF_i = \alpha &+ \beta_A Aggr_i + \beta_S OrderSize_i + \beta_B BuyOrder_i + \beta_I ILO_i + \\ &+ \beta_{HFT} HFT_i + \beta_{AAT} AAT_i + \beta_{IH} ILOHFT_i + \beta_{IA} ILOAAT_i + \\ &+ \beta_T TrdFreq_i + \beta_T Volat_i + \varepsilon_i \end{aligned} \quad [5]$$

where the sub-index *i* represents the i^{th} order. See Appendix B for variable definitions.

We report median estimated coefficients across stocks, the percentage of statistically significant coefficients, and the percentage of significant and positive coefficients. For the execution cost component we provide results conditional on partial execution (fill rate > 0%); for the opportunity costs component, we provide results conditional on non-full execution (fill rate < 100%). Note that a fully executed order has zero opportunity cost, and a completely non-executed order has zero execution cost. The results of the implementation shortfall analyses are reported in Table 8.

[Insert Table 8 here]

In Panel A we show the total implementation shortfall. The coefficient on the ILOHFT dummy captures the shortfall measure for hidden orders placed by HFTs, controlling for order attributes (for example order aggressiveness, size, etc.) for order types (ILO only, since the other is displayed orders) and for trader categories (HFT and AAT, since there are three trader categories). The ILOHFT dummy is negative, is significant for 39.29% of the sample, and positive only for 10.71% of these stocks. Thus, the majority of the stocks in our sample show a negative implementation shortfall for hidden orders placed by HFTs. Note that the ILO dummy has a positive coefficient with the majority of the stocks showing positive and significant estimates. This is consistent with the finding in BPV that hidden orders in general have higher implementation shortfall, but our results show that this is not the case for hidden orders placed by HFTs.

To probe how HFTs manage to achieve reduced shortfall for their hidden orders, we disaggregate the metric into its two components – effective costs of execution (Panel B) and opportunity cost of non-execution (Panel C). In Panel B, for all fill rates (which includes all orders submitted), HFT effective cost is positive for hidden orders (coefficient of 0.0126 on the ILOHFT dummy) and this is significant for 71.43% of the sample stocks, and significantly positive for a majority (64.29%). If we consider only those orders with fill rates greater than zero (that is orders which had to be at least partly executed), both the magnitude and the percentage significantly increase. These results indicate that HFTs face a higher effective cost when their hidden orders are executed. In Panel C, we find that the opportunity cost of non-execution of (ILOHFT dummy) is negative (coefficient of -0.0714) for HFTs' hidden orders and this result is stronger for fill rates under 100%. This indicates that although HFTs face higher execution costs for hidden orders, their non-execution costs are lower, and larger in magnitude, and in sum the latter (Panel C) effect dominates the former (Panel B), leading to an overall lower implementation shortfall (Panel A) result.

5.2 Information content: Impulse response functions

The evidence from non-high-frequency markets indicates that hidden orders are generally uninformed (BPV, 2009). So we next turn to the information content of hidden orders placed by HFTs and compare that to the hidden orders of AATs and NATs. To measure the information content of each group's trades, we first calculate the permanent price impact of each group's trades. Unlike the multi-market settings in, for example, Huang (2002) and Barclay, Hendershott, and McCormick (2003), where price impact computations are plagued with difficulties in trade-quote alignment, the fact that the NSE handles over 80% of the equity volume in Indian markets provide us the advantage of a consolidated market.

We estimate an extended version of the Vector Autoregressive (VAR) model in Hasbrouck (1991a). The model is defined in event time (t), where an event may be a limit order submission, cancellation of a standing limit order, or a trade. Revisions that improve (degrade) prices or increase (decrease) quoted depth are treated as limit order submissions (cancellations). We distinguish between HFTs, AATs, and NATs, and for each trader group we consider two types of orders – iceberg and displayed. As a result of these partitions, the VAR model has 13 equations: one for the quote midpoint return and 12 for order-flow related variables. The optimal number of lags is determined using the Schwarz' Bayesian Information Criterion for each stock-day. We exclude stock-days that have less than 20 orders for each variable, where a variable is a trade, order display choice, or cancellation. The trade variable take the value +1 (-1) for buyer- (seller-) initiated trades. Displayed, iceberg, or cancellation variables that happen on the ask (bid) side of the LOB take the value (-1) +1. We reset the trading process at the end of each day, resetting all lagged values to zero. The model is estimated in event time, not transaction time, so contemporaneous correlation is negligible. Nonetheless, we compute the IRFs such that any correlation is taken into account (See methodological details in Appendix C). In Table 9 Panel A we report the impulse responses obtained from the estimation process described above.

[Insert Table 9 here]

The impulse response functions (IRF) indicate the permanent price impact of an innovation for each trader group (in the columns) and for each type of event (along the rows), computed as continuously compounded returns and presented in basis points. Estimates are cross-sectional averages with standard errors clustered by stock and day (Thompson, 2011).

As expected, trades have the largest coefficients for all trader categories. Among the three trader groups, HFT trades have the largest coefficient (1.2271) and this is significantly different from both NATs and AATs (boldfaced coefficients). Of greater interest to us, however, is the IRF of hidden orders placed by HFTs, compared to the other two trader groups. Here we find that HFTs' hidden orders have lower long-term price impact (coefficient of 0.1913 in Panel A) which is not significantly different from either AATs' (0.2401) or NATs' (0.2170).

Results in Table 9 Panel A suggest that ILOs are more informative than DLOs for all trader types. Notice, however, that these results do not control for the aggressiveness of limit order. Earlier tests show that HFTs use more aggressive ILOs and that the likelihood of their hiding orders increases with order aggressiveness. To test whether the Panel A results are affected by order aggressiveness, in Panel B, we show the impulse response functions for the same trader groups but after controlling for order aggressiveness. We classify as aggressive (non-aggressive) any limit order placed at or within (beyond) the prevailing best quotes. Our results survive: ILOs placed by HFTs have a significantly lower permanent price impact than similarly aggressive ILOs placed by other traders.

5.3 Information content: Efficient price variance decomposition

The IRFs reported in Panels A and B of Table 9 indicate that although HFTs have a significant impact on prices, this impact is smaller than that of either AATs or NATs. But how important is this impact in the overall price formation process? If different trader categories do not trade as frequently in each stock, then the IRF results will not be a good indicator of the information conveyed by the hidden orders of each trader category in the aggregate price formation process. To address this, we follow Hendershott and Riordan (2013) and decompose the variance of the efficient price into the portion of total price discovery that is correlated with HFTs, AATs, and NATs.

For each stock-day, we decompose the efficient price variance due to the order flow (OF) into its components. We estimate the efficient (or long-run) variance using Hasbrouck (1991b) approach. Using the Vector Moving Average representation of a VAR model for quote midpoint changes and order flow, this method provides an estimate of the efficient variance that can be split into an order-flow-related component and an order-flow-unrelated component. We estimate the share of each trader-message type on the order-flow-related component, which is of interest to us in examining the information content of hidden order of

HFTs. As with the IRF computation (Panels A and B of Table 9), the model is defined in event time. An event may be a limit order submission, a cancellation of a standing limit order, or a trade (market or marketable limit order). We report the results in Panel C of Table 9. We report the estimated cross-sectional average shares on the order-flow-related efficient variance with standard errors clustered by both stock and day (Thompson, 2011).

Overall, trades explain 67.05% of the OF-related price variance and DLOs explain another 25.95%. ILOs explain fewer than 8%. Looking across the columns, if we focus on all orders (last row of Panel A) it is clear that HFTs' orders (25.02%) contribute less than both AAT (34.54%) and NAT (40.44%) orders. Finally, the iceberg orders of HFT contribute the smallest (0.46%) to OF-related price variation. These results show that even after accounting for trader categories, in aggregate HFTs' hidden orders convey less information into price, when compared with either their displayed orders, or with the displayed and hidden orders of AATs and NATs. The boldfaced coefficients show that these differences between HFTs and NATs are statistically significant.

5.4 Information shares

Our third and final verification that HFTs' hidden orders do not appear to be informationally motivated comes from estimates of the information shares of these orders vis-à-vis their other orders, as well as the orders of the other trader groups. To do this, we use the approach proposed in Hasbrouck (1995). Although much of the literature, including Hasbrouck (1995) used this set-up to test the information shares in multi-market setting, this has also been used, for example in Hendershott and Riordan (2013), to assess the information shares across different trader categories. The econometric approach of the Hasbrouck (1995) model assumes that HFT, AAT and NAT quotes share a common efficient price process and the information share attributable to each of these trader categories is the relative contribution of their innovations in the common efficient price of the asset. Table 10 presents the results.

[Insert Table 10 here]

HFTs' average information share is 30.85% for their displayed orders and 6.13% for their hidden orders. Both AATs and NATs have greater information shares for their hidden orders, at 7.62% and 11.87% respectively. Tests of statistical significance show that these differences – both between HFTs and AATs and between HFTs and NATs – are significant. Overall, the results in Tables 9 and 10 point to the fact that the information conveyed by the hidden orders placed by HFTs is less than either their displayed orders, or the hidden (and displayed) orders of the other two trader categories.

6. Undercutting

If not informationally motivated, then why do HFTs hide orders? In this section we address this question. Our evidence so far shows that unlike the large liquidity-motivated hidden orders of NATs, HFTs usually place small hidden orders close to the best quotes. One possible use of such orders could be to undercut standing orders without being detected. HFTs with their super-fast computers are in a position to anticipate order flow (Angel and McCabe, 2013) and trade ahead of other investors' order flow (Hirschey, 2016). Do they use hidden orders to this end? We address this question as follows.

We identify all limit order submissions (including revisions) that offer price improvement for a standing limit order and/or increase the size of standing limit orders, for each trader/order type. We define an undercutting limit order as a limit order that (a) is placed immediately after another submission on the same side of the market, (b) comes in under 10 milliseconds of the previous order, and (c) improves the price of the previous one. We present results using undercutting orders restricted to the five best quotes; however our conclusions remain unchanged if we consider only the best quotes. Results are in Panel A of Table 11.

[Insert Table 11 here]

It is clear that of the three trader categories, HFTs use the highest proportion (5.0208% or 5.6019%) of hidden orders to undercut orders within five ticks of the standing best quotes, both at a lower level of stock activity (at least 20 orders of each type – hidden or displayed – by each trader category – HFT, AAT, and NAT per stock-day) or a higher level of stock activity (at least 50 orders, per order-type and trader category, as defined above). Not surprisingly, they also use displayed orders, 3.009% or 2.6037% depending on the level of activity in stocks, to trade ahead of standing quotes. Expectedly, NATs show the least amount of such undercutting activity, both for hidden and displayed orders.

This evidence, while illustrative, does not take account of market conditions. From BPV and our earlier regression results, we know that order exposure is affected by both stock and market attributes. Thus, we next estimate a logit regression to examine whether the observed higher rates of undercutting by HFTs' hidden orders remain after controlling for market conditions and the state of the order book. The dependent variable in this model is a dummy that takes the value of one if an order is an undercutting order as defined earlier, zero otherwise. The first two control variables describe the characteristics of the undercut order. First we consider the displayed size of the undercut order. We expect that when the undercut order has a larger displayed size, HFTs are more likely to jump ahead of it. Second we consider the aggressiveness of the undercut order. Aggressiveness is defined as the number of ticks away from the best quote on the same side. The further the undercut order is from the best quotes, in other words less aggressive, the less likely it is to be undercut. Thus we expect a negative relationship between the aggressiveness of the undercut order and its chance of being undercut. We include the trader types HFT and AAT (NAT is captured in the intercept) and the interaction of trader categories with the iceberg option, plus relative spread, depth on same and opposite sides, and volatility, all as defined earlier. Finally, we include a variable that gauges the possibility of hidden order detection. The variable *HidVolDetected* is a dummy that takes the value of one if the presence of hidden volume in same side has been revealed, zero otherwise. Hidden volume is revealed at the time an undercutting order is placed if the quantity that has been traded at the prevailing best quote is greater than the displayed depth, which is only possible if there was additional (iceberg) volume at the best quotes (e.g., Pardo and Pascual, 2012).

Panel B of Table 11 presents the results. Displayed size of the undercut order is positively related to the likelihood of an order used for undercutting, confirming that larger orders are more likely to be undercut. Likewise, when an order is closer to the top of the book (more aggressive), it is more likely to be undercut (shown by the negative and significant coefficient on *Aggr_of_FR-Order*). *HidVolDetected* is positive indicating that when traders can infer the presence of hidden volume at the best quotes, they are more likely to place orders to trade ahead of these hidden orders. In fact, the odds ratio shows that this likelihood is 1.5666 times (or 50.66% more) compared to the use of displayed orders by NATs (we use the DLOs of NATs as the reference group for all odds ratio calculation in this Panel).

The main variable of interest is *HFTILO*. The coefficient on this variable is 0.4149 and significant at the 1% level. Compare this to the negative coefficients on *AATILO* and *NATILO*. Clearly, HFTs use hidden orders for undercutting, while the two other trader groups are less likely to do the same. In fact, the odds ratio for HFTs is greater than 1 (1.5142) while for both other groups it is lower than 1, indicating that while HFTs use hidden orders to undercut the standing quotes at or near the top of the order book, the two other groups are less likely to use hidden orders for the same purpose.

7. Conclusion

Regulators, market operators (exchanges), and investors all agree that transparency is a desirable property in financial markets. At the same time, research shows that there is such a thing as too much transparency.[11] Thus, all major exchanges allow traders to hide their trading interest by placing hidden orders. To avoid a "corner solution" where everyone chooses to hide all trading intent, hidden orders face a penalty in the form of losing time priority (for similarly priced orders, a hidden order is always in the back of the queue relative to a displayed order, even if the latter entered the system later). Research on hidden orders generally conclude that patient liquidity providers use the option to hide when they want to transact large quantities while avoiding picking off risks (for example, BPV, De Winne and d'Hondt 2007, Buti and Rindi, 2013).

These findings come from non-high frequency markets, or models that do not account for the use of hidden orders by HFTs. Given that HFTs are the majority of traders in some markets (the US, Japan, and Europe, for example) and an increasing fraction in many others (India, China, for example), whether and how they use the option to hide orders should be of interest. In this paper we provide, to our knowledge, the first comprehensive account of hidden order use by HFTs. This study is made possible by our access to data from the NSE – the largest exchange in India that handles over 80% of the equity volume – which identifies in rich detail the types of traders as well as the order handling system they use. With these information, we can precisely identify HFTs – proprietary traders who use algorithmic order entry and management systems – and examine their hidden order use. The other advantage of the NSE data is that, unlike the Trade and Quote data of the NYSE, for example, it provides

[11] Bloomfield and O'Hara (1999) examine market transparency in a study tellingly titled "Market transparency: Who wins and who loses?" In this laboratory experiment they determine the effects of trade and quote disclosure on market efficiency, bid-ask spreads, and trader welfare. They find that although trade disclosure increases the informational efficiency of prices, it also increases opening bid-ask spreads by reducing market-makers' incentives to compete for order flow. As a result, trade disclosure benefits market makers at the expense of liquidity traders and informed traders. Additionally, they examine quote disclosure and find no discernible effects on market performance. Asquith, Au, Covert, and Pathak (2013) find that the introduction of the TRACE reporting system for bond markets helped some investors and dealers through a decline in price dispersion, while harming others through a reduction in trading activity.

real time message traffic which makes it possible to rebuild the LOB for each stock in continuous time.

We find that HFTs make extensive use of hidden orders. They do not appear to use hidden orders to avoid picking-off risk but instead use small order sizes, placed nearer the top of the book using the non-display option. This pattern is different from the NATs, who hide large orders and place them further away from the best quotes.

Although the market conditions, firm characteristics, and order book state that explain HFTs' order non-exposure are consistent with prior literature (BPV, for example), we find that HFTs are more skilled at minimizing the implementation shortfall of their hidden orders by reducing the opportunity costs of non-execution as well as improving the probability of execution. We address the information content of HFTs' hidden orders using three different measures to capture the information conveyed by such orders - the permanent price impact as represented by impulse response functions, decomposition of the efficient price variation into the order-flow related component, and the Hasbrouck (1995) information share measure. All three metrics indicate that hidden orders placed by HFTs have lower information content than their displayed orders, as well as the hidden orders of the other two trader groups. Collectively these evidence show that HFTs' pattern of hidden order use do not align with theoretical models of order exposure, and make a case for new theory. To that end, we show that HFTs' hidden order use is closer to the results in Hirschey (2016) – used to jump ahead of other investors' orders. By presenting new evidence on the use of hidden orders by HFTs, we believe this study makes a useful contribution to the literature.

References

Aitken, Michael J., Hank Berkman, and Derek Mak. 2001. The use of undisclosed limit orders on the Australian Stock Exchange, *Journal of Banking & Finance* 25 (8): 1589-1603

Anand, Amber, and Daniel. G. Weaver. 2004. Can order exposure be mandated? *Journal of Financial Markets* 7(4), 405-426.

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

Asquith, Paul, Andrea Au, Thomas Covert, and Parag A. Pathak. 2013. The market for borrowing corporate bonds. *Journal of Financial Economics* 107(1): 155-182.

Barclay, Michael J., Terrence Hendershott, and D. Timothy McCormick. 2003. Competition among trading venues: Information and trading on electronic communications networks." *The Journal of Finance* 58(6): 2637-2665.

Bessembinder, Hendrik, Marios Panayides, and Kumar Venkataraman. 2009. Hidden liquidity: an analysis of order exposure strategies in electronic stock markets. *Journal of Financial Economics* 94(3), 361-383.

Biais, Bruno, and Thierry Foucault. 2014. HFT and market quality. *Bankers, Markets & Investors* 128(1): 5-19.

Bloomfield, Robert, and Maureen O'Hara. 1999. Market transparency: Who wins and who loses? *The Review of Financial Studies* 12(1): 5-35.

Bloomfield, Robert, Maureen O'Hara, and Gideon Saar. 2015. Hidden liquidity: Some new light on dark trading. *Journal of Finance* 70(5), 2227-2274.

Boehmer, Ekkehart, Gideon Saar, Lei Yu. 2005. Lifting the veil: An analysis of pre-trade transparency at the NYSE. *Journal of Finance*, *60*(2), 783–839.

Boulatov, Alex. and Thomas. J. George. 2013. Hidden and displayed liquidity in securities markets with informed liquidity providers. *Review of Financial Studies* 26(8), 2096-2137.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. 2014. High frequency trading and price discovery. *Review of Financial Studies* 27, 2267-2306.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. 2017. Price discovery without trading: Evidence from limit orders. Working paper.

Buti, Sabrina, and Barbara Rindi. 2013. Undisclosed orders and optimal submission strategies in a limit order market. *Journal of Financial Economics* 109(3), 797-812.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam. 2005. Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics* 76(2): 271-292.

Chordia, Tarun, Clifton Green, and Badrinath Kottimukkalur. 2017. Rent Seeking by Low Latency Traders: Evidence from Trading on Macroeconomic Announcements. *Review of Financial Studies*. Forthcoming.

Degryse, Hans, Geoffrey Tombeur, and Gunther Wuyts. 2015. Two shades of opacity: Hidden orders versus dark trading. Working paper.

De Winne, Rudy, and Catherine D'hondt. 2007. Hide-and-seek in the market: placing and Detecting hidden orders. *Review of Finance* 11(4): 663-692.

Easley, David, Marcos M. López de Prado, and Maureen O'Hara. 2012. The volume clock: Insights into the high-frequency paradigm. *The Journal of Portfolio Management* 39(1): 19-29.

Engle, Robert, Ferstenberg Robert, and Russell Jeffrey. 2012. Measuring and modeling execution cost and risk. *The Journal of Portfolio Management* 38(2): 14-28.

Foucault, Thierry. 1999. Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets*, 2, 99-134.

Gai, Jiading, Yao Chen, and Mao Ye. 2013. The externalities of high frequency trading. Working paper.

Glosten, Larry. 1996. Is the electronic limit order book inevitable? *The Journal of Finance*, 49(4), 1127-1161.

Goettler, Ronald, Christine Parlour, and Uday Rajan. 2009. Informed traders and limit order markets. *Journal of Financial Economics*, 93(1): 67–87.

Goldstein, Michael A., Andriy V. Shkilko, Bonnie F. Van Ness, and Robert A. Van Ness. 2008. Competition in the market for NASDAQ securities. *Journal of Financial Markets* 11(2): 113-143.

Gozluklu, Arie. 2016. Pre-trade transparency and informed trading: Experimental evidence on undisclosed orders. *Journal of Financial Markets*, 28: 91–115.

Harris, Larry. 1997. Order exposure and parasitic traders. Unpublished working paper. University of Southern California, Los Angeles, CA.

Hasbrouck, Joel. 1991a. Measuring the information content of stock trades, *Journal of Finance* 46(1): 179–207.

Hasbrouck, Joel. 1991b. The summary informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 4(3), 571–595.

Hasbrouck, Joel. 1995. One security, many markets: Determining the contributions to price discovery, *Journal of Finance*, 50(4), 1175–1199.

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.

Hautsch, Nikolaus, and Ruihong Huang. 2012. On the dark side of the market: identifying and analyzing hidden order placements. Working paper.

Hendershott, Terrence, Charles M. Jones. 2005. Island goes dark: Transparency, fragmentation, and regulation. *Review of Financial Studies*, *18*(3), 743-793.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld. 2011. Does algorithmic trading improve liquidity? *The Journal of Finance* 66(1): 1-33.

Hendershott, Terrence, and Ryan Riordan. 2013. Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis* 48(4): 1001-1024.

Hirschey, Nicholas. 2016. Do high-frequency traders anticipate buying and selling pressure? Working Paper.

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

Huang, Roger D. 2002. The quality of ECN and Nasdaq market maker quotes. *The Journal of Finance* 57(3): 1285-1319.

Huang, Roger D. and Hans R. Stoll. 1992. The Design of Trading Systems: Lessons from Abroad, *Financial Analysts Journal* 48(5): 49-54.

Jain, Archana, Pankaj K. Jain, Thomas H. McInish, and Michael McKenzie. 2013. Worldwide reach of short selling regulations. *Journal of Financial Economics* 109(1): 177-197.

Jain, Pankaj K. 2005. Financial market design and the equity premium: Electronic versus floor trading. *The Journal of Finance* 60(6): 2955-2985.

Kahraman, Bige, and Heather Tookes. 2017. Trader leverage and liquidity. *The Journal of Finance*. Forthcoming.

Kaniel, Ron, and Hong Liu. 1996. So what orders do informed traders use? *Journal of Business* 79(4): 1867-1913.

Keim, Donald B., and Ananth Madhavan. 1997. Transactions costs and investment style: an inter-exchange analysis of institutional equity trades. *Journal of Financial Economics* 46(3): 265-292.

Lewis, Michael, and Dylan Baker. *Flash boys*. Simon & Schuster Audio, 2014.

Lo, Andrew W., A. Craig MacKinlay, and June Zhang. 2002. Econometric models of limit-order executions. *Journal of Financial Economics* 65(1): 31-71.

Massey Jr. Frank J. 1951. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association* 46(253): 68-78.

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

Moallemi, Ciamac. 2014. The Value of Queue Position in a Limit Order Book, Working paper.

Moinas, Sophie. 2010. Hidden limit orders and liquidity in order driven markets. Unpublished working paper. TSE Working Paper Series 10-147.

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

Pardo, Angel, and Roberto Pascual. 2012. On the hidden side of liquidity. *The European Journal of Finance*, 18(10): 949-967.

Pascual, Roberto, and David Veredas. 2010. Does the open limit order book matter in explaining informational volatility? *Journal of Financial Econometrics*, 8(1): 57–87.

Perold, Andre F. 1988. The implementation shortfall: Paper versus reality. *The Journal of Portfolio Management* 14(3): 4-9.

Rindi, Barbara. 2008. Informed traders as liquidity providers: Anonymity, liquidity and price formation. *Review of Finance*, 12(3): 497–532.

Sofianos, George. 1995. Specialist gross trading revenues at the New York Stock Exchange. Working paper.

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. 2016. High-frequency trading around large institutional orders. Working paper.

von Beschwitz, Bastian, Donald B. Keim, and Massimo Massa. 2015. First to 'read' the news: News analytics and high frequency trading. Working paper.

Table I
Sample descriptive statistics

This table provides daily cross-sectional average statistics for 100 stocks listed on National Stock Exchange (NSE) of India. Cross-sectional averages are computed from daily averages per stock. The sample period is October to December 2013 (61 trading days). The sample comprises market-capitalization-based subsamples of 30 (largest), 40 (medium), and 30 (smallest) stocks. Market capitalization is daily average in billions of Rupees. Volume is in 10,000-share units, number of trades is in 100-trade units, depth is in 1000-share units, and Price is in Rupees. Daily volatility is {(maximum price/minimum price) –1}x100. The relative bid-ask spread is the ratio of the quoted spread to the quote midpoint, in basis points. The relative effective spread is two times the difference between the average trade price and the quote midpoint divided by the quote midpoint. Displayed (hidden) depth is the accumulated displayed (non-displayed) depth in the limit order book (LOB). MT is message traffic or the number of order messages (sum of submissions, cancellations, and revisions) in 1000-message units. We provide two proxies for HFT: the ratio of MT to trades (MT/Trd) and cancellations to trades (CAN/Trd). Share in MT denotes each trader type's share in message traffic. Liquidity metrics are generated from 1-minute snapshots of the LOB and averaged across observations. Statistical significance is evaluated using the non-parametric Wilcoxon rank-sum test. In Panel A, "***", "**", "*" on "Mid" ("Small") indicate statistically different from the "Large" ("Mid") subsample at the 1%, 5%, and 10% levels, respectively. In Panel B, significance under the AATs column tests for the difference between HFTs and AATs, and in the NATs column tests the differences between algorithmic traders (HFTs and AATs) and NATs.

Panel A: Sample statistics

		Market-capitalization-based subsamples		
	Full sample	Large	Mid	Small
Market capitalization (billions)	448.19	1464.64	20.2 ***	2.39 ***
Volume ('0000)	86.54	227.73	40.2 ***	7.12 **
Number of trades ('00)	106.99	315.89	25.88 ***	6.23 ***
Volatility	42.36	32.96	44.37 ***	49.08
Relative bid-ask spread (bsp)	44.24	8.7	42.87 ***	81.61 ***
Displayed depth ('000)	103.57	203.27	81.65 ***	33.08 ***
Hidden depth ('000)	25.62	50.58	19.5 ***	8.81
Price (Rupees)	309.76	606.47	255.94 ***	84.8 ***

Panel B: Message traffic per trader type and subsample

		Trader types		
Subsample	Variable	HFTs	AATs	NATs
Large	MT	1191.19	139.26 *	43.28 ***
	MT/Trd	223.79	29.25 ***	2.51 ***
	CAN/Trd	10.32	1.99 ***	0.27 ***
	Share in MT	57.81	25.91 ***	16.28 ***
Mid	MT	6.37	11.19 *	5.10 ***
	MT/Trd	300.95	107.99	3.03 ***
	CAN/Trd	30.96	1.77 *	0.54 ***
	Share in MT	16.72	43.16 ***	40.11 ***
Small	MT	0.77	3.51 ***	1.31 ***
	MT/Trd	94.75	146.18 ***	3.62 ***
	CAN/Trd	9.77	1.53	0.73 ***
	Share in MT	5.97	57.23 ***	36.80 ***

Table II
Use of iceberg orders

For a market-capitalization representative sample of 100 of NSE-listed stocks, this table provides cross-sectional average daily statistics on the use of iceberg limit orders (ILOs) and displayed limit orders (DLOs) per trader category. We distinguish between high frequency traders (HFTs), agency algorithmic traders (AATs) and non-algorithmic traders (NATs) and provide statistics for subsamples of the largest (30), medium (40), and smallest (30) stocks in our sample. Our sample period is October to December 2013. In Panel A, we show the proportion of ILOs, both in the number of orders, and the accumulated volume, relative to all limit orders submitted. In Panel B, we provide each trader category's share of both ILOs and DLOs. Significant difference in medians between HFTs and AATs are shown beside AAT numbers and between all algorithmic traders (HFTs and AATs) and NATs are shown beside NAT numbers, using the non-parametric Wilcoxon rank-sum test. ***, **, * indicate statistically different at the 1%, 5%, and 10% levels respectively.

Panel A: Iceberg orders: Relative importance by type of trader and market capitalization subsample

		HFTs		AATs		NATs	
Variable	Subs.	Ord.	Vol.	Ord.	Vol.	Ord.	Vol.
% ILOs	Large	10.38	9.83	24.31 ***	32.47 ***	8.86 *	30.17 ***
	Mid	36.00	34.42	15.97	26.40	14.48 **	34.03
	Small	15.84	15.23	3.36 ***	7.77 ***	13.38 ***	32.42 ***

Panel B: Market shares of ILOs and DLOs for each trader type (%)

Variable	Subs.	Orders	Volume	Orders	Volume	Orders	Volume
DLOs	Large	34.67	55.84	21.59 *	8.09 ***	43.74 *	36.07 ***
	Mid	4.27	3.00	15.59 **	3.35 ***	80.14 ***	93.65 ***
	Small	1.55	0.75	19.90 ***	2.17 ***	78.54 ***	97.08 ***
ILOs	Large	9.28	3.69	49.73 ***	30.02 ***	40.99 **	66.29 ***
	Mid	18.90	8.14	13.45	6.34 **	67.65 ***	85.52 ***
	Small	5.80	2.49	4.36 ***	1.57 ***	89.84 ***	95.95 ***

Table III
Order size

We provide cross-sectional average daily statistics on the empirical distribution of the size of iceberg limit orders (ILOs) and displayed limit orders (DLOs) in the NSE. The sample consists of 100 stocks listed on the NSE between October and December 2013 that we split into three market capitalization groups: large caps (Panel A), mid-sized (Panel B), and small caps (Panel C), of sizes 30, 40, and 30 stocks, respectively. We distinguish between high frequency traders (HFTs), agency algorithmic traders (AATs) and non-algorithmic traders (NATs). The analysis is based on order-by-order data that we group according to the full (displayed plus non-displayed) order size. Trade size categories are defined in total (both displayed and hidden) shares. We use the two-sample Kolmogorov-Smirnov (Massey, 1951) test to compare the order size distributions of ILOs and DLOs submitted by the different trader categories. We provide the percentage of ILOs and DLOs in each order-size category per trader category.

	HFTs		AATs		NATs	
Order size distrib. (%)	DLOs	ILOs	DLOs	ILOs	DLOs	ILOs
Panel A. Large caps						
(0,50]	5.11	76.28	60.17	55.23	65.99	29.13
(50,75]	0.79	10.15	10.91	8.25	1.52	2.69
(75,100]	1.19	0.55	4.18	6.25	11.14	11.19
(100,200]	22.01	2.24	11.42	12.11	6.36	11.98
(200,500]	46.53	7.91	10.40	11.33	9.03	22.26
(500,1000]	19.02	2.39	1.54	4.03	3.26	10.79
(1000,2500]	2.82	0.44	0.73	2.03	1.46	6.07
>2500	2.53	0.05	0.65	0.77	1.24	5.89
HFTs vs. AATs/NATs (p-value)				0.00		0.00
DLOs vs. ILOs (p-value)		0.00		0.00		0.00
Average size (sh.)	1150.50	459.58	345.36 ***	880.65 *	309.27 **	1139.59 ***
Panel B. Mid-sized caps						
(0,50]	62.98	98.72	71.60	63.84	51.81	31.53
(50,75]	3.19	1.03	9.48	6.67	1.86	1.69
(75,100]	7.19	0.24	8.60	4.61	14.21	13.44
(100,200]	8.23	0.01	4.74	8.79	9.87	10.01
(200,500]	6.09	0.00	4.56	8.82	12.86	19.83
(500,1000]	1.55	0.00	0.59	4.00	5.00	10.06
(1000,2500]	0.81	0.00	0.27	2.11	2.38	6.61
>2500	9.96	0.00	0.15	1.17	2.01	6.84
HFTs vs. AATs/NATs (p-value)				0.00		0.00
DLOs vs. ILOs (p-value)		0.00		0.00		0.00
Average size (sh.)	207.85	94.35	96.68	1342.57 ***	396.23 ***	1247.97 ***
Panel C. Small caps						
(0,50]	46.51	83.96	87.99	75.64	47.95	20.77
(50,75]	4.19	15.44	1.43	3.16	1.58	1.75
(75,100]	29.71	0.58	7.82	3.20	14.85	16.32
(100,200]	12.24	0.00	1.62	6.11	11.10	12.46
(200,500]	5.85	0.01	0.84	5.62	15.76	22.51
(500,1000]	1.07	0.00	0.20	3.02	5.23	12.22
(1000,2500]	0.41	0.00	0.08	2.16	2.15	6.96
>2500	0.01	0.00	0.03	1.10	1.38	7.03
HFTs vs. AATs/NATs (p-value)				0.00		0.00
DLOs vs. ILOs (p-value)		0.00		0.00		0.00
Average size (sh.)	127.81	99.06	49.52	740.69 ***	319.05 ***	1196.78 **

Table IV
Iceberg order and hidden volume placement in the order book

We examine the placement of iceberg limit orders (ILOs) and displayed limit orders (DLOs), in Panels A and B respectively, both by the number of orders and the share volume. We build snapshots of the limit order book (LOB) at the time of each new order submission and group the LOB levels into four segments: (a) better than the standing quotes ("Better"), (b) at the best quotes ("At"), (c) from the best quotes up to 5 ticks away ("Near"), and (d) the rest ("Far"). The sample consists of 100 stocks listed on the NSE between October and December 2013 split into three market capitalization groups: largest (30), mid-sized (40), and smallest (30) stocks. We distinguish between high frequency traders (HFTs), agency algorithmic traders (AATs) and non-algorithmic traders (NATs). Each statistic reported is the time series mean of the daily proportion of orders at the four LOB level groups for all stocks taken together. We average ask and bid quotes. Statistical tests compare the medians of corresponding groups across Panels A and B, using the non-parametric Wilcoxon rank-sum test. (***, **, * indicate statistically different at the 1%, 5%, and 10% level respectively).

Panel A: ILOs placement

		Order placement			Volume placement		
Subsample	Aggressiveness	HFTs	AATs	NAT	HFTs	AATs	NAT
Large	Better	25.11 ***	10.81 ***	12.05 ***	26.64 ***	7.55 ***	6.80 ***
	At	20.92 ***	37.58 ***	17.72 ***	30.55 ***	36.57 ***	29.11 ***
	Near	51.68 ***	38.03 **	31.11 ***	39.88 ***	34.73 ***	29.80 ***
	Far	2.28 ***	13.57 ***	39.12 ***	2.93 ***	21.15 ***	34.30 ***
Mid	Better	70.14 ***	43.57 ***	20.27 ***	72.01 ***	13.29	10.57 ***
	At	15.66	26.76 **	16.92	13.31 ***	45.20 ***	22.68 ***
	Near	14.02 ***	22.57 ***	27.35 ***	14.43 ***	28.07 ***	24.65
	Far	0.19 ***	7.10	35.45 ***	0.25 ***	13.43	42.10 ***
Small	Better	82.33 ***	49.34 ***	25.98 ***	85.07 ***	26.96 *	16.43 ***
	At	5.60 ***	15.31	14.39 ***	4.97 ***	33.12 ***	17.45
	Near	11.82 ***	31.62 ***	27.31 ***	9.79 ***	32.62 ***	24.34 **
	Far	0.25 ***	3.74 ***	32.31 ***	0.18 ***	7.29 ***	41.78 ***

Panel B: DLOs placement

Subsample	Aggressiveness	HFTs	AATs	NAT	HFTs	AATs	NAT
Large	Better	0.47	2.43	4.32	0.08	0.64	2.40
	At	1.03	5.01	7.26	0.83	3.27	21.00
	Near	9.17	35.53	13.24	5.92	21.80	18.87
	Far	89.32	57.03	75.18	93.18	74.29	57.72
Mid	Better	36.17	23.06	11.56	3.28	13.71	7.26
	At	16.55	22.18	16.80	5.12	30.54	32.09
	Near	31.36	49.53	22.17	37.25	43.89	24.66
	Far	15.92	5.23	49.47	54.35	11.87	35.99
Small	Better	30.88	24.52	13.17	21.05	17.53	9.32
	At	14.77	13.24	12.23	15.24	17.47	18.59
	Near	37.23	55.37	20.82	40.17	46.78	22.77
	Far	17.12	6.87	53.78	23.54	18.22	49.32

***,**,* means statistically different than the corresponding statistic in Panel B at the 1%, 5

Table V
The order exposure decision

We study the determinants of the order (non-) exposure decision of high frequency traders (HFTs), agency algorithmic traders (AATs) and non-algorithmic traders (NATs). We use logistic models (Panel A) of order characteristics and market conditions to study the choice between submitting an iceberg limit order (ILO) and a fully displayed limit order (DLO). We exclude all market and marketable limit orders. The dependent variable equals one (zero) if the NAT submits an ILO (DLO). We use Tobit models (Panel B) of order characteristics and market conditions to study the decision of how much volume of a limit order is hidden. The dependent variable here is the amount of shares hidden, normalized by the stock's average daily trading volume. Appendix A lists the definitions of all variables. The models are estimated on a stock-by-stock basis, and we report aggregated coefficients and t-statistics using the approach in Chordia, Roll, and Subrahmanyam (2005). The estimation sample for this table consists of the 30 largest stocks (in which HFTs are reasonably active) from our main sample of 100 stocks listed on the NSE. The sample period is December 2013. ***, **, * indicate significance at the 1%, 5%, and 10% level respectively.

Panel A: Decision to hide - logistic model	ATs					
Variable	HFTs		AATs		NATs	
Intercept	-3.9108	***	-0.8195	**	-1.8061	***
Price aggressiveness	2865.7587	***	511.3416	***	65.7729	***
Total order size	31.7138	**	19.9858	***	18.3290	***
Relative spread	1558.2250	***	-69.7108		-4.0103	
Depth same side	-586.9779	***	-216.5916	***	-88.1710	***
Depth opposite side	39.8854		50.2558	***	-30.9239	**
Stock volatility	-0.0141		-0.0031		-0.0062	***
Waiting time	-50.3939	*	24.9165		15.5722	**
Trade frequency	-1.5337		-0.4582		-0.7669	**
Hidden same side	-3.0559		0.0679		-0.2246	
LOB order imbalance	15.7592		0.4677		-0.2394	
Last trade size	-3.4383	***	-2.0167	**	-0.4277	*
Market volatility	-0.0017	*	-0.0014		-0.0001	
Last half hour indicator	572.6601	***	72.4503		-169.1852	***
Panel B: Magnitude of hidden volume - Tobit model						
Intercept	-0.0041		-0.0007	**	-0.0031	***
Price aggressiveness	0.2880	***	0.0726	***	0.0607	**
Total order size	0.0043		0.0067	***	0.0055	***
Relative spread	0.1933	***	-0.0168		0.0709	
Depth same side	-0.0479	**	-0.0461		-0.0332	*
Depth opposite side	0.0051		0.0035		-0.0278	**
Stock volatility	0.5508		0.0208		-0.0501	***
Waiting time	-0.0060		0.0014		0.0049	***
Trade frequency	-0.0075		-0.0002		-0.0005	
Hidden same side	-0.0874		-0.0033		0.0004	
LOB order imbalance	0.0007		0.0010		0.0016	
Last trade size	-0.0003		-0.0003		-0.0001	
Market volatility	0.0000		0.0000		0.0000	
Last half hour indicator	0.0544		0.0843		-0.0466	

***,**,* means statistically significant at the 1%, 5% and 10% level, respectively

Table VI
Likelihood of order execution

We study the determinants of execution of iceberg (ILOs) and displayed limit orders (DLOs) in the NSE. We distinguish between high frequency traders (HFTs), agency algorithmic traders (AATs) and non-algorithmic traders (NATs). To model order execution likelihood, in Panel A we use an ordered Logit model, where the dependent variable (EXEC) is an ordinal variable that takes three possible values: EXEC = 1 indicates that the limit order is cancelled before execution; EXEC = 2 indicates that the limit order is partially executed and then cancelled; EXEC = 3 indicates that the limit order is fully executed. Appendix A lists all variable definitions. The models are estimated on a stock-by-stock basis, and we report aggregated coefficients and significance levels based on Chordia, Roll, and Subrahmanyam (2005). In Panel B we show ILO and DLO execution likelihood conditioned upon the level of the limit order book (LOB) where the order is placed. We consider three levels relative to the best quotes. The estimation sample consists of the 30 largest stocks (in which HFTs are reasonably active) from our main sample of 100 stocks listed on the NSE. The sample period is December 2013. In Panel A ***, **, * indicate significance at the 1%, 5% and 10% level, respectively. In Panel B ***, **, * indicate significantly different from HFTs at the 1%, 5% and 10% level, respectively.

Panel A: Likelihood of execution - Ordered probit model

Variable	Limit order to buy Coef.		Limit order to sell Coef.	
Aggressiveness	273.7968	***	159.7378	***
Order size	-2405.2770	**	-2055.9926	**
ILO (dummy)	-0.4301	**	-0.2509	
HFT (dummy)	-2.2576	***	-2.2935	***
Agency-AT (dummy)	-1.6361	***	-1.4190	***
ILO x HFT	2.5816	***	1.7313	***
ILO x Agency-AT	1.3648	***	0.9437	***
Relative spread	530.1330	***	490.8146	***
Depth same side	-85.8532	***	-59.6091	**
Depth opposite side	62.5362	***	72.1912	***
LOB imbalance	-0.1518	***	0.1550	***
Last half hour (dummy)	0.2398	***	0.2658	***
Order imbalance	-0.1464	***	0.1131	**
Trading frequency	1.1414	**	1.4850	**
Momentum	7.7690		3.9942	
Volatility	4897.55		6661.67	

Panel B: Likelihood of execution and order placement

Placement/trader type	ILOs All		ILOs w/HVol		DLOs All	
At or within the best quotes:						
HFT	79.10		59.61		79.42	
AAT	86.82		75.22	***	71.19	**
NAT	85.34		71.14	**	86.48	
Within the 2nd and 5th best quotes:						
HFT	83.42		64.83		48.40	
AAT	56.15	***	44.17	***	32.61	**
NAT	66.53	***	56.38	***	73.12	***
Beyond the 5th best quote						
HFT	81.84		70.11		6.98	
AAT	25.47	***	20.83	***	24.75	***
NAT	51.78	***	47.43	***	50.51	***

Table VII
Time to completion: Survival analysis

We study the determinants of the time to full execution of non-marketable limit orders at the NSE. We exclude market and marketable limit orders. We also drop fleeting orders (as defined by Hasbrouck and Saar, 2009). Revisions of non-executed orders are treated as the same order. Revisions of partially-executed orders are treated as new submissions. . Appendix A lists the definitions of all variables. The table reports the estimated parameters of an econometric model of time-to-completion using survival analysis. We follow Bessembinder et al. (2009) and Lo, et al. (2002). The model describes an accelerated failure time specification of limit order execution times under the generalized gamma distribution. The model is estimated on a stock-by-stock basis, and we report aggregated coefficients and significance levels based on Chordia, Roll, and Subrahmanyam (2005). The estimation sample for this table consists of the 30 largest stocks (in which HFTs are reasonably active) from our main sample of 100 stocks listed on the NSE of India and the sample period is December 2013. Explanatory variables are defined in the Appendix. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Limit order to buy		Limit order to sell	
Variable	Coef.	Sign.	Coef.	Sign.
Intercept	16.7115	***	17.0854	***
Midquote-limit price	2.8183	**	-1.6744	***
Last trade buy indicator	0.0762	*	-0.0918	
Same side depth	227.3927	***	221.0423	**
Same side depth squared	-169.5214	**	-151.7108	**
Opposite side depth	-196.9867	***	-227.5512	***
Order size (total)	47.1514	***	37.3681	**
Trading frequency	-14.2400	**	-10.3429	*
Relative trading frequency	-1.5036	***	-1.4494	***
ILO (dummy)	1.4503	***	1.1420	***
HFT (dummy)	2.7756	***	2.4768	***
AAT (dummy)	0.4430		0.0509	
ILO x HFT	-3.6125	***	-2.7638	***
ILO x AAT	-1.5064	***	-1.1791	**

Table VIII
Implementation shortfall of iceberg orders

We present the effective costs of execution and the opportunity costs of non-execution costs of iceberg orders (ILOs) and displayed limit orders (DLOs) in the NSE using the implementation shortfall (IMPS) approach of Perold (1988). Execution cost for a buy order is the difference between the average execution price and the mid-quote at the time of order submission, multiplied by the amount of shares executed. The opportunity cost for a buy order is the difference between the closing price on the day the order is cancelled or expires and the quote midpoint at the time the order is submitted, multiplied by the unexecuted part of the order (in shares). Metrics for sell orders are analogously computed but conveniently signed. We regress each cost component on order attributes (order aggressiveness, total size, buyer order indicator, and ILO indicator), market conditions during the 30 minutes prior to order submission (trading frequency and realized volatility), and trader-category dummies. We estimate regressions for the whole IMPS, but also for the execution cost component, and the opportunity costs component separately. Appendix A lists the definitions of all variables. Models are estimated on a stock-by-stock basis. We report median estimated coefficients across stocks, the percentage of statistically significant coefficients, and the percentage of significant and positive coefficients. Note that a fully executed order has zero opportunity cost, and a fully cancelled order has zero execution cost. For the execution cost component we provide results conditional on partial execution (fill rate > 0%); for the opportunity costs component, we provide results conditional on non-full execution (fill rate < 100%). The estimation sample for this table consists of the 30 largest stocks (in which HFTs are reasonably active) from our main sample of 100 stocks listed on the NSE of India and the sample period is December 2013. Explanatory variables are defined in the Appendix. We consider only non-marketable limit orders. Revisions of non-executed orders are treated as the same order. Revisions of partially-executed orders are treated as new submissions.

Panel A: Implementation shortfall

	All fill rates	
Variable	Coef.	%Signif.(pos.)
Intercept	0.0638	92.86 (60.71)
Aggressiveness	1.2400	82.14 (46.43)
Order size	-78.4506	75.00 (14.29)
BuyOrder	-0.1703	96.43 (28.57)
ILO	0.0121	53.57 (39.29)
HFT	0.0348	78.57 (50.00)
AAT	0.0141	75.00 (50.00)
ILOHFT	-0.0445	39.29 (10.71)
ILOAAT	-0.0016	64.29 (35.71)
Trading frequency	0.1042	64.29 (50.00)
Volatility	77.2356	60.71 (32.14)

Panel B: Effective costs

	All fill rates		Fill rate >0%	
Intercept	-0.0315	92.86 (0.00)	-0.0151	85.71 (0.00)
Aggressiveness	0.4152	89.29 (75.00)	12.9157	92.86 (92.86)
Order size	-5.8493	82.14 (0.00)	-70.6004	82.14 (0.00)
BuyOrder	0.0035	89.29 (78.57)	0.0021	78.57 (53.57)
ILO	-0.0139	89.29 (10.71)	-0.0128	71.43 (10.71)
HFT	0.0297	89.29 (89.29)	0.0081	67.86 (64.29)
AAT	0.0338	92.86 (92.86)	0.0267	92.86 (92.86)
ILOHFT	0.0126	71.43 (64.29)	0.0503	75.00 (71.43)
ILOAAT	0.0134	92.86 (78.57)	0.0184	85.71 (78.57)
Trading frequency	-0.0095	82.14 (7.14)	-0.0074	75.00 (14.29)
Volatility	-164.6489	85.71 (10.71)	-68.6062	39.29 (0.00)

Table VIII Cont.)
Implementation shortfall of iceberg orders

Panel C: Opportunity costs of non-execution

	All fill rates		Fill rate <100%	
Intercept	0.0799	92.86 (64.29)	0.1433	92.86 (64.29)
Aggressiveness	0.0653	75.00 (35.71)	0.0021	75.00 (42.86)
Order size	0.6778	57.14 (25.00)	4.9032	53.57 (25.00)
BuyOrder	-0.1714	89.29 (28.57)	-0.2998	89.29 (28.57)
ILO	0.0432	67.86 (53.57)	0.1359	53.57 (39.29)
HFT	-0.0033	78.57 (32.14)	-0.0358	71.43 (21.43)
AAT	-0.0182	67.86 (14.29)	-0.0744	78.57 (14.29)
ILOHFT	-0.0714	46.43 (10.71)	-0.1022	50.00 (10.71)
ILOAAT	-0.0292	82.14 (32.14)	-0.0633	67.86 (25.00)
Trading frequency	0.1159	71.43 (53.57)	0.4274	67.86 (53.57)
Volatility	193.9898	57.14 (32.14)	239.0775	67.86 (35.71)

Table IX
Impulse-response functions and order-flow related variance decomposition

In Panels A and B we provide stock-day average impulse response functions (IRF) from an extended VAR (Hasbrouck,1991a). In Panel C we estimate the efficient variance using the Hasbrouck (1991b) approach and decompose the efficient variance into an order-flow-related component and an order-flow-unrelated component. For all models we use order level data for December 2013 on the 30 largest stocks in our representative sample of 100 NSE-listed stocks. The models are defined in event time (t), where an event may be a limit order submission, cancellation, or trade. Revisions that improve (degrade) prices or increase (decrease) quoted depth are treated as limit order submissions (cancellations). We distinguish between high frequency traders (HFTs), agency algorithmic traders (AATs), and non-algorithmic traders (NATs). We differentiate between iceberg (ILOs) and displayed limit orders (DLOs). As a result of these partitions, the models have 13 equations: one for the quote midpoint return and 12 for order-flow related variables. The optimal number of lags is determined using the Schwarz' Bayesian Information Criterion. "Trade" variables are signed +1 (-1) for buyer- (seller-) initiated trades. "DLO", "ILO" or "Cancellation" variables that happen on the ask (bid) side of the LOB are signed (-1) +1. We assume the trading process restarts each day, resetting all lagged values to zero. Standard errors are clustered by both stock and day (Thompson, 2011). In Panel B, present the IRF tests but controlling for order aggressiveness. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively. In Panels B and C, we boldface those coefficients for AATs and NATs that are significantly different from corresponding coefficients for HFTs.

		Trader type		
Message	All traders	HFT	AAT	NAT
Panel A: Continously-compound return (in basis points) IRF				
Trades		1.2271 ***	**0.7259** ***	**0.8582** ***
		(0.1382)	(0.1017)	(0.1474)
DLO		0.0816 **	0.0568 ***	**0.1640** ***
		(0.0318)	(0.0099)	(0.0260)
ILO		0.1913 ***	0.2401 ***	0.2170 ***
		(0.0536)	(0.0328)	(0.0308)
Cancellations		0.0793 ***	0.0454 ***	**0.1233** ***
		(0.0291)	(0.0117)	(0.0254)
Panel B: IRF - controlling for aggressiveness				
Trades		1.1591 ***	**0.7273** ***	**0.8583** ***
		(0.1261)	(0.1039)	(0.1485)
DLOa		0.2512 ***	0.2410 ***	**0.6221** ***
		(0.0505)	(0.0288)	(0.0696)
DLOna		0.0111 *	-0.0014	-0.0002
		(0.0064)	(0.0052)	(0.0039)
ILOa		0.1778	**0.3523** ***	**0.4907** ***
		(0.1132)	(0.0445)	(0.0543)
ILOna			-0.0351	-0.0239 **
			(0.0225)	(0.0116)
Cancellations		0.0623 ***	0.0502 ***	**0.1163** ***
		(0.0196)	(0.0100)	(0.0238)
Panel C: OF-related efficient variance (OFEV) decomposition				
Trades	67.05	16.09 ***	**21.39** ***	**29.57** ***
		(1.69)	(3.13)	(2.24)
Limit orders	25.95	6.18 ***	**9.25** ***	**10.52** ***
		(1.03)	(1.21)	(0.93)
Iceberg orders	7.84	0.46 **	**5.68** ***	**1.69** ***
		(0.18)	(0.87)	(0.14)
Cancellations	-0.84	2.29 ***	**-1.78** **	**-1.34** ***
		(0.72)	(0.73)	(0.25)
All orders		25.02	34.54	40.44

Table X
Information shares

The table reports the average stock-day information shares (IS) for different types of traders and orders in the NSE. Information shares are estimated using Hasbrouck (1995) approach. We report lower bound (minimum), upper bound (maximum), and average information shares for three types of traders: proprietary ATs (hereafter, HFTs), agency ATs (hereafter, AATs), and non-ATs (hereafter, NATs). Moreover, we distinguish between iceberg orders (ILOs) and fully displayed limit orders (DLOs). On a one-second frequency, we obtain the best quotes for each trader category and order type. The price path of each trader category and order type pair is given by the quote midpoint prevailing at the end of each second. Using the IS approach, we decompose the variation in the unobserved common efficient price into individual components attributable to specific trader and order type. Our main purpose is to examine the fraction of price discovery attributable to ILOs and how much of it is attributable to HFTs' and ATs' orders. We use order level data for December 2013 on the 30 largest stocks in our representative sample of 100 NSE-listed stocks. ***, **, * next to a HFTs' or ATs' IS indicates that the IS statistic is significantly different from the corresponding NATs' IS statistic for the same order type.

		Information shares (%)		
Trader type	Order	Min.	Max.	Avg.
HFTs	DLO	15.87	45.83	30.85
	ILO	5.91	6.34 ***	6.13 **
AATs	DLO	8.81 ***	34.44 ***	21.62 ***
	ILO	5.00	10.25 ***	7.62 **
NATs	DLO	16.22	47.62	31.92
	ILO	6.36	17.39	11.87

***, ** means statistically differnt than the NAT's statistic at the 1% and 5% level, respectively

Table XI
Undercutting using iceberg orders

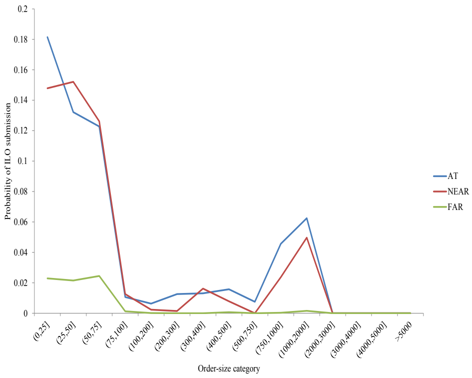
We present the proportions of hidden (and displayed) orders used for undercutting by the three trader categories (in Panel A) and use a logit regression model to study the likelihood of undercutting by these three trader groups (in Panel B). We define an undercutting limit order as a limit order that (a) is placed immediately after another submission on the same side of the market, (b) comes in under 10 milliseconds of the previous order, and (c) improves the price of the previous one. We present results using undercutting orders restricted to the five best quotes. We divide the total number of undercutting orders of each type – hidden and displayed – placed by each trader category – HFT, AAT, and NAT – by all orders submitted of a given type by each trader category. We present those fractions in Panel A. In Panel B we present the coefficients and odds ratios of the logit regression where the dependent variable is a dummy that takes the value of 1 if the order is an undercutting order, 0 otherwise. Appendix A lists the definitions of all control variables. The models are estimated on a stock-by-stock basis, and we report aggregated coefficients and t-statistics using the approach in Chordia, Roll, and Subrahmanyam (2005). The estimation sample for this table consists of the 30 largest stocks (in which HFTs are reasonably active) from our main sample of 100 stocks listed on the NSE of India and the sample period is December 2013. ***, **, * indicate significance at the 1%, 5%, and 10% level respectively.

Panel A: Descriptive statistics on undercutting (% of orders)

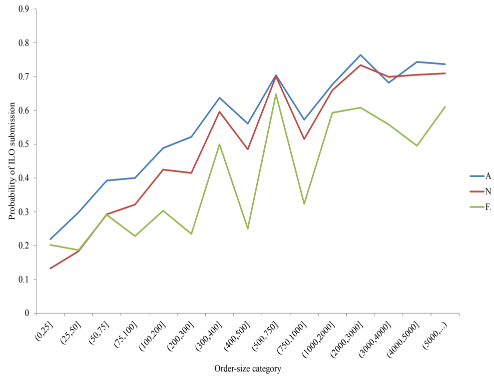
First case: At least 20 orders per category and stock-day					
Order	TraderType	Bid side		Ask side	
ILO	HFT	5.0208	***	5.4092	***
	AAT	3.2303	***	3.4066	***
	NAT	0.8173	***	0.8078	***
DLO	HFT	3.0091	***	3.2427	***
	AAT	4.6264	***	5.0179	***
	NAT	1.1373	***	1.1707	***
Second case: At least 50 orders per category and stock-day					
Order	TraderType	Bid side		Ask side	
ILO	HFT	5.6019	***	6.0651	***
	AAT	3.3964	***	3.4847	***
	NAT	0.8088	***	0.8025	***
DLO	HFT	2.6037	***	2.7307	***
	AAT	5.1687	***	5.5820	***
	NAT	1.0611	***	1.0792	***

Panel B: Logit model on undercutting

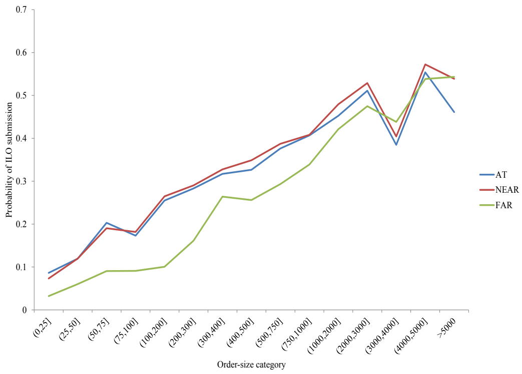
Variable	Coef.		Odds ratio	CRS t-stat
DispSize_of FR_Order	0.0004	***	1.0004	10.03
Aggr_of_FR_Order	-0.0744	***	0.9283	-119.14
HFT	0.7620	***	2.1425	39.49
AAT	0.9856	***	2.6794	40.69
HFTILO	0.4149	***	1.5142	7.67
AATILO	-0.1902		0.8268	-0.06
NATILO	-0.5556	***	0.5737	-3.96
HidVolDetected	0.4489	***	1.5666	66.72
Spread	0.0300	***	1.0304	39.78
DepthSame/100	0.3798	***	1.4620	10.71
DepthOpposite/100	-0.9478	***	0.3876	-9.27
Volatility*10000	0.0134	***	1.0135	22.57
Intercept	-4.0663	***		-183.04



1.a. HFTs' probability of ILO submission



1.b. AATs' probability of ILO submission



1.c. NATs' probability of ILO submission

Figure 1

Probability of submitting an ILO conditional on order size and aggressiveness

We plot estimated cross-sectional daily average probabilities of iceberg limit order (ILO) submission in the NSE conditional on order size and order aggressiveness. We distinguish between high frequency traders (HFTs), agency algorithmic traders (AATs) and non-algorithmic traders (NATs). The sample consists of the 30 largest stocks from our size-stratified sample of 100 stocks listed on the NSE between October and December 2013. We combine the limit order book levels into three groups: at the best quotes ("At"); from the best quotes up to 5 ticks away ("Near"), and the rest ("Far"). For each order size, level of aggressiveness, and type of trader, we provide the percentage of iceberg orders of all the non-marketable limit orders submitted. Figures 1.a, 1.b, and 1.c provide the findings for HFTs, AATs, and NATs respectively.

Appendix A: Building the Limit Order Book of the National Stock Exchange of India

The National Stock Exchange of India (NSE) provides two types of files: order files and trade files. Order files contain all the message traffic. Each order has an identification code that allows us to follow the history of the order from submission to execution/cancellation/expiry. Messages are time-stamped to the nearest microsecond. For each order, we know the type of message (new submission, revision, or cancellation); the type of order (limit order or market orders); the type of trader submitting the order (HFT, AAT or NAT, based on the identification of trader accounts described in the accompanying paper); the order direction (buy or sell), and the total size of the order. For limit orders we also know the limit price, and for iceberg limit orders (ILOs) we know both the total size and the displayed size (hidden volume is the difference between the total size and the displayed size). The file also identifies orders with special conditions: immediate-or-cancel, and on stop. The trade files provide, for each trade, the buy and sell orders matched; the type of trader submitting each order; the trade size, and the trade price.

We start each day assuming the limit order book (LOB) is empty. We use the registers of the opening auction from the order file to build the LOB pre-allocation. Orders in the NSE are sorted by price-time priority, with market orders having priority over limit orders, no matter the time of submission. The trade files provide the information about orders matched at the allocation price of the opening call auction. Non-allocated market orders at the end of the auction time are transformed into limit orders at the allocation price. If there are no trade registers associated with the opening auction it indicates that there was no allocation price. In such cases, market orders are stored at the closing price of the previous session. The result is the initial snapshot of the LOB for the corresponding day.

Then, we update the state of the LOB conditioned on each and every posterior message (new submission, revision, or cancellation) during the continuous session. We match the order and the trade files, checking that every market order and every marketable limit order submitted have their corresponding trade registers. By doing so, we can also discern the actual direction of each trade, i.e., whether the trade is buyer- or seller-initiated.

During the continuous session ILOs orders are allowed. As in other markets around the world, the hidden part of the iceberg orders loses time priority against displayed limit orders. Accordingly, every time the displayed volume unit of an iceberg order is exhausted, the emerging new displayed volume unit moves to the end of the queue of all standing limit orders at the same price. Our program allows us to obtain snapshots of both the displayed and the hidden components of the LOB at every instant during the continuous session.

Order revisions are the most common type of message in the NSE. These revisions can change the order size, the limit price, or both. Some of these updates can change the priority of the execution of the order. In particular, increases in volume will cause losing time priority. Decreases in volume, however, will not change priority. Obviously, increases (decreases) in the limit price of a standing limit order to buy (sell) will increase price priority. Changes in hidden volume with no change in displayed volume are possible. In that case, the displayed part of the iceberg order does not lose time priority. We update the state of the LOB after each revision to reflect these changes in price-time priority.

Changes in the type of order, from “on stop” to ordinary or the other way around are possible, but not very frequent. When an “on stop” order changes to ordinary order, it is treated as a new submission. When an ordinary order changes to “on stop”, it is removed from the LOB. Orders on stop can be revised while not activated. Once activated, a new register indicates the change in status and the final conditions under which the order reaches the LOB. At that point, the order is treated as an ordinary new submission. Immediate-or-cancel orders only change the book if executed and, therefore, generate a trade.

The best proof that our program works is that the resulting LOB file and the trade file perfectly match. When a marketable limit or a market order is submitted, the associated sequence of trade registers is consistent with what can be inferred by matching the incoming aggressive order with the price-time priority sorted orders standing in the LOB, and controlling for hidden volume. Additionally, there are no inconsistencies between the timing of order flow events and the timing of the associated trades.

Appendix B: List of variables and scaling factors

Variable	Definition	Multiplier
Depth same side	displayed depth at the best bid (ask) for a buy (sell) order divided by the avg. daily trading vol	
Depth opposite side	displayed depth at the best ask (bid) for a buy (sell) order divided by the avg. daily trading vol	
Trade frequency	number of Shares Traded in the last hour	1/1000000
Hidden same side	number of hidden shares at the best quote on the same side (i.e., at the bid side for a buy order) revealed by the most recent transaction	1/1000
Last half hour indicator	indicator variable that equals 1 for orders submitted in the last hour of the trading day else 0	
Last trade buy indicator	indicator variable that equals 1 If the last trade is buyer initiaed 0 otherwise	
Last trade size	size of the most recent transaction divided by the average daily trading volume	100
LOB order imbalance	% difference between displayed liquidity in the best five prices on the buy and sell side of the book, suitably signed (i.e., the variable is positive when same size liquidity exceeds opposite	1/100000
Market volatility	Sum of Squares of logarithm of NIFTY50 price returns over the last 60 minutes	10000000
Midquote-limit price	Difference between the mid quote and the limit price of the order	
Momentum	Quote mid point return in last 5 minutes	
Order Imbalance	Number of Shares traded in trades initiated by Buy orders minus Number of Shares traded in trades initiated by Sell orders divided by number of shares traded in last 5 minutes	
Price aggressiveness	distance of the order's limit price from the opposite quote price, suitably signed (a higher value indicates a more aggressively priced order) divided by the quote midpoint	
Relative trading frequen	Num of shares traded in last 30 minutes divided by Num of shares traded in last 60 minutes	
Same side depth squarec	Depth same side*Depth same side	100
Stock volatility	Sum of Squares of logarithm of stock quote's midpoint returns over the last 5 minutes	
Total order size	total (displayed plus hidden) size of the order divided by average daily trading volume	100
Trade frequency	number of Shares traded per second in the last 60 minutes	1/1000
Waiting time	avg. time between last 3 order message arrivals on the same side, reseting the clock daily	1/10000000

Appendix C: Methodological details of estimation

Table IX

Impulse-response function (VAR model)

We investigate the permanent price impact (informational content) of different types of orders by different trader category. As order types, we consider market and marketable limit orders (Trades), displayed non-marketable limit orders (DLO), non-marketable iceberg orders (ILO), and cancellations of standing limit orders. We consider three types of traders: HFTs, AATs, and NATs. To estimate the permanent price impact of each type of order, we use the VAR approach of Hasbrouck (1991a), as extended by Fleming, Mizrach, and Nguyen (2015) and Brogaard, Hendershott, and Riordan (2017). The model is defined in event time, where each order is an observation (t), and estimated per stock-day. We assume the trading process restarts each day, resetting all lagged values to zero. The model is,

$$\begin{aligned}
r_t &= \sum_{j=1}^{n} \alpha_j^0 r_{t-j} + \sum_{j=0}^{n} \beta_j^{0,1} X_{t-j}^1 + \sum_{j=0}^{n} \beta_j^{0,2} X_{t-j}^2 + \cdots + \sum_{j=0}^{n} \beta_j^{0,12} X_{t-j}^{12} + \varepsilon_t \\
X_t^1 &= \sum_{j=1}^{n} \alpha_j^1 r_{t-j} + \sum_{j=1}^{n} \beta_j^{1,1} X_{t-j}^1 + \sum_{j=1}^{n} \beta_j^{1,2} X_{t-j}^2 + \cdots + \sum_{j=1}^{n} \beta_j^{1,12} X_{t-j}^{12} + \mu_t^1 \\
X_t^2 &= \sum_{j=1}^{n} \alpha_j^2 r_{t-j} + \sum_{j=1}^{n} \beta_j^{2,1} X_{t-j}^1 + \sum_{j=1}^{n} \beta_j^{2,2} X_{t-j}^2 + \cdots + \sum_{j=1}^{n} \beta_j^{2,12} X_{t-j}^{12} + \mu_t^2 \qquad [1] \\
\vdots &= \qquad\qquad\qquad \vdots \\
X_t^{12} &= \sum_{j=1}^{n} \alpha_j^{12} r_{t-j} + \sum_{j=1}^{n} \beta_j^{12,1} X_{t-j}^1 + \sum_{j=1}^{n} \beta_j^{12,2} X_{t-j}^2 + \cdots + \sum_{j=1}^{n} \beta_j^{12,12} X_{t-j}^{12} + \mu_t^{12}
\end{aligned}$$

or in compact form

$$A_0 y_t = \sum_{j=1}^{n} A_j y_{t-j} + \xi_t$$

where $y_t' = \left(r_t, X_t\right)$ is the 1x13 vector of contemporaneous dependent variables and $\xi_t' = \left(\varepsilon_t, \mu_t\right)$ is the 1x13 vector of innovations to the dependent variables; r_t is the continuously compound quote midpoint return expressed in basis points; X_t is a vector of order-flow related variables. By combining the 3 types of traders and the 4 types of events/orders, we have 12 possible order-flow categories (X^1 *to* X^{12}): NAT/AT/HFT – Trade, NAT/AT/HFT – DLO, NAT/AT/HFT – ILO, NAT/AT/HFT – Cancellation.

Each X^k can take one of three possible values: 0, 1 or -1: X^k = 1(-1) if order t is a buy (sell) order of type k and zero otherwise. Because the model is defined in event time, whenever $X^k = 1$, $X^z = 0$ $\forall z{\neq}k$. The number of lags (n) is stock-day specific and determined using the Schwarz' Bayesian Information Criterion (SBIC), which Lütkepohl (2005, p. 148-152) shows provides consistent estimates of the true lag order.

As in Hasbrouck (1991a) original VAR model, we assume contemporaneous causality running from the order flow to the changes in prices. Accordingly, the 13x13 matrix A_0 equals

$$A_0 = \begin{pmatrix} 1 & -\beta_0^{0,1} & -\beta_0^{0,2} & -\beta_0^{0,3} & \cdots & -\beta_0^{0,12} \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

and the variance-covariance matrix of the residuals becomes,

$$Var(\xi_t) = \begin{pmatrix} \sigma_\varepsilon^2 & \bar{0} \\ \bar{0}' & \Omega \end{pmatrix}$$

where σ_ε^2 is the variance of the innovation to r; $\bar{0}$ is a 1x12 vector of zeros, and Ω is the variance-covariance matrix of X.

Because the model is defined in event time, Ω is near-diagonal. For a representative stock-day, the average contemporaneous correlation across μ^k innovations is about 0.1%. We follow Brogaard et al. (2017) in computing orthogonalized and order independent IRFs. The IRF for trades and orders (cancellations) is computed for a unitary positive (negative) shock at period t = 0, assuming the model is in a steady state (i.e., all lagged variables equal to zero), and the subsequent price impact (in basis points) is accumulated over the next 20 periods. The model is estimated for each stock-day, and we report the average IRF across stock-days. Statistical significance is clustered by stock and day (e.g., Thompson, 2011).

Table IX
Order Flow-related efficient variance decomposition

The analysis summarized in Table X provides estimates of the average information content of particular types of orders submitted by particular types of traders. That analysis, however, should not be interpreted in terms of overall contributions to price discovery, since we ignore the frequency with which each event takes place. In Table XI, we follow Hasbrouck (1991b) to estimate the relative contribution of the different (trader category, order type) binomials to the component of the long-run variance of the stock attributable to the order flow.

From the VMA representation of a VAR model similar to [1] but with only two variables (quote midpoint changes and trades), Hasbrouck (1991b) obtains an estimate of the long-run variance of the corresponding asset, say Σ, which he further decomposes into a trade-related component, due to the innovations to the trading process (i.e., μ in [2]) and a trade-unrelated component, due to the innovations to the quote midpoint changes (i.e., ε in [1]). Because of the one-directional causality assumption (from trades to quotes), ε are contemporaneously uncorrelated with μ.

Specifically, the VAR model in [1] can be re-written as

$$A_0 y_t = \sum_{j=1}^{n} A_j y_{t-j} + \varepsilon_t \rightarrow A(L) y_t = \varepsilon_t .$$

where L is the lag operator, that is, $L^j y_t = y_{t-j}$, and *A(L)* is a lag polynomial, that is, $A(L) = A_0 - \sum_{j=1}^{n} A_j L^j$. Its VMA representation would be

$$y_t = \Psi(L)\xi_t \qquad [2]$$

In expanded form, the VMA in [2] is as follows,

$$
\begin{aligned}
r_t &= \sum_{j=1}^{\infty} \theta_j^0 \varepsilon_{t-j} + \sum_{j=0}^{\infty} \phi_j^{0,1} \mu_{t-j}^1 + \sum_{j=0}^{n} \phi_j^{0,2} \mu_{t-j}^2 + \cdots + \sum_{j=0}^{\infty} \phi_j^{0,12} \mu_{t-j}^{12} + \varepsilon_t \\
X_t^1 &= \sum_{j=1}^{\infty} \theta_j^1 \varepsilon_{t-j} + \sum_{j=1}^{\infty} \phi_j^{1,1} \mu_{t-j}^1 + \sum_{j=1}^{\infty} \phi_j^{1,2} \mu_{t-j}^2 + \cdots + \sum_{j=1}^{\infty} \phi_j^{1,12} \mu_{t-j}^{12} + \mu_t^1 \\
X_t^2 &= \sum_{j=1}^{\infty} \theta_j^2 \varepsilon_{t-j} + \sum_{j=1}^{\infty} \phi_j^{2,1} \mu_{t-j}^1 + \sum_{j=1}^{\infty} \phi_j^{2,2} \mu_{t-j}^2 + \cdots + \sum_{j=1}^{\infty} \phi_j^{2,12} \mu_{t-j}^{12} + \mu_t^2 \\
\vdots &= \qquad\qquad\qquad\qquad \vdots \\
X_t^{12} &= \sum_{j=1}^{\infty} \theta_j^{12} \varepsilon_{t-j} + \sum_{j=1}^{\infty} \phi_j^{12,1} \mu_{t-j}^1 + \sum_{j=1}^{\infty} \phi_j^{12,2} \mu_{t-j}^2 + \cdots + \sum_{j=1}^{\infty} \phi_j^{12,12} \mu_{t-j}^{12} + \mu_t^{12}
\end{aligned}
\qquad [3]
$$

Notice that [3] keeps the contemporaneous causality flow from orders to trades.

We define the order-flow related component of the long-run variance of the stock (I_x) as

$$
I_x = \frac{Var\left(E\left[\Delta m_t \mid \mu_t\right]\right)}{Var(\Delta m_t)} = \frac{\sigma_{\Delta m,\mu}^2}{\sigma_{\Delta m}^2} \qquad [4]
$$

Now assume that any non-diagonal element in $Var(\mu_t) = \Omega$ is negligible; they actually are, as we have explained before. Let $\phi_j^0 = \left(\phi_j^{0,1}, \phi_j^{0,2}, \ldots, \phi_j^{0,12}\right)$ be the row vector of order flow related coefficients at lag j in the r_t equation of the VMA model [3]. Therefore, the row vector of cumulated impacts of order-flow-related unitary shocks is $\phi = \sum_{j=0}^{\infty} \phi_j^0 = \left(\sum_{j=0}^{\infty} \phi_j^{0,1}, \sum_{j=0}^{\infty} \phi_j^{0,2}, \ldots, \sum_{j=0}^{\infty} \phi_j^{0,12}\right)$. Similarly, the cumulated impact of a unitary order flow unrelated shock is $\theta = 1 + \sum_{j=1}^{\infty} \theta_j^0$. Hasbrouck (1991b) shows that the long run variance (the variance of the efficient price) can be computed from the VMA coefficients as $\sigma_{\Delta m}^2 = \phi \Omega \phi' + \theta^2 \sigma_\varepsilon^2$, and the order flow related efficient variance can be decomposed as

$$
\sigma_{\Delta m,\mu}^2 = \phi \Omega \phi' = \left(\sum_{j=0}^{\infty} \phi_j^{0,1}\right)^2 \sigma_{\mu_1}^2 + \left(\sum_{j=0}^{\infty} \phi_j^{0,2}\right)^2 \sigma_{\mu_2}^2 + \ldots + \left(\sum_{j=0}^{\infty} \phi_j^{0,12}\right)^2 \sigma_{\mu_{12}}^2 \ , [5]
$$

that is, the sum of the variance of the IRFs of order flow related shocks.

In Table XI we provide the contribution of each (trader category, order type) binomial's related shocks to the long-run variance component in eq. [5].

Table X
Information shares

Following Hasbrouck (1995), we estimate stock-day information shares (IS) for different trader-category/order-type combinations. In contrast to analyses run in event time, Hasbrouck's IS approach evaluates price discovery using trader-category/order-type specific quotes collected at regular time intervals. As noted in Brogaard et al. (2017), event time analyses do not account for tiny differences in the response of different traders to new public information releases, which result in the subsequent price discovery being attributed to the fastest traders. Thus, the IS uses a more conservative timing approach to price discovery.

We compute trader-category/order-type specific quote midpoints prevailing at the end of each second. We consider three types of traders (HFTs, AATs, and NATs) and two types of orders (DLOs and ILOs). For each trader category, we collect the best ask and bid quotes supported by standing DLOs and compute the quote midpoint by averaging the best ask and bid quotes. In case there are no DLOs standing on the LOB for that trader category, the observation is replaced by the closest preceding non-missing observation. For ILOs, we proceed in the same way.

Hasbrouck's (1995) approach decomposes the variance of the underlying efficient price into components attributable to the different trader-category/order type pairs, the so-called "information shares". The first step of this methodology estimates a Vector Error-Correction (VEC) Model for each stock-day, under the assumption that the quote midpoints are co-integrated. The VEC model is reported in eq. [6]

$$
\begin{aligned}
\Delta q_t^{HI} &= \alpha^{HI\prime}\beta q_{t-1} + \sum_{j=1}^{n}\phi_j^{HI,HI}\Delta q_t^{HI} + \sum_{j=1}^{n}\phi_j^{HI,HD}\Delta q_t^{HD} + \sum_{j=1}^{n}\phi_j^{HI,AI}\Delta q_t^{AI} + \sum_{j=1}^{n}\phi_j^{HI,AD}\Delta q_t^{AD} + \sum_{j=1}^{n}\phi_j^{HI,NI}\Delta q_t^{NI} + \sum_{j=1}^{n}\phi_j^{HI,ND}\Delta q_t^{ND} + \varepsilon_t^{HI} \\
\Delta q_t^{HD} &= \alpha^{HD\prime}\beta q_{t-1} + \sum_{j=1}^{n}\phi_j^{HD,HI}\Delta q_t^{HI} + \sum_{j=1}^{n}\phi_j^{HD,HD}\Delta q_t^{HD} + \sum_{j=1}^{n}\phi_j^{HD,AI} r_t^{AI} + \sum_{j=1}^{n}\phi_j^{HD,AD}\Delta q_t^{AD} + \sum_{j=1}^{n}\phi_j^{HD,NI}\Delta q_t^{NI} + \sum_{j=1}^{n}\phi_j^{HD,ND}\Delta q_t^{ND} + \varepsilon_t^{HD} \\
\Delta q_t^{AI} &= \alpha^{AI\prime}\beta q_{t-1} + \sum_{j=1}^{n}\phi_j^{AI,HI}\Delta q_t^{HI} + \sum_{j=1}^{n}\phi_j^{AI,HD}\Delta q_t^{HD} + \sum_{j=1}^{n}\phi_j^{AI,AI}\Delta q_t^{AI} + \sum_{j=1}^{n}\phi_j^{AI,AD}\Delta q_t^{AD} + \sum_{j=1}^{n}\phi_j^{AI,NI}\Delta q_t^{NI} + \sum_{j=1}^{n}\phi_j^{AI,ND}\Delta q_t^{ND} + \varepsilon_t^{AI} \\
\Delta q_t^{AD} &= \alpha^{AD\prime}\beta q_{t-1} + \sum_{j=1}^{n}\phi_j^{AD,HI}\Delta q_t^{HI} + \sum_{j=1}^{n}\phi_j^{AD,HD}\Delta q_t^{HD} + \sum_{j=1}^{n}\phi_j^{AD,AI}\Delta q_t^{AI} + \sum_{j=1}^{n}\phi_j^{AD,AD}\Delta q_t^{AD} + \sum_{j=1}^{n}\phi_j^{AD,NI}\Delta q_t^{NI} + \sum_{j=1}^{n}\phi_j^{AD,ND}\Delta q_t^{ND} + \varepsilon_t^{AD} \\
\Delta q_t^{NI} &= \alpha^{NI\prime}\beta q_{t-1} + \sum_{j=1}^{n}\phi_j^{NI,HI}\Delta q_t^{HI} + \sum_{j=1}^{n}\phi_j^{NI,HD}\Delta q_t^{HD} + \sum_{j=1}^{n}\phi_j^{NI,AI}\Delta q_t^{AI} + \sum_{j=1}^{n}\phi_j^{NI,AD}\Delta q_t^{AD} + \sum_{j=1}^{n}\phi_j^{NI,NI}\Delta q_t^{NI} + \sum_{j=1}^{n}\phi_j^{NI,ND}\Delta q_t^{ND} + \varepsilon_t^{NI} \\
\Delta q_t^{ND} &= \alpha^{ND\prime}\beta q_{t-1} + \sum_{j=1}^{n}\phi_j^{ND,HI}\Delta q_t^{HI} + \sum_{j=1}^{n}\phi_j^{ND,HD}\Delta q_t^{HD} + \sum_{j=1}^{n}\phi_j^{ND,AI}\Delta q_t^{AI} + \sum_{j=1}^{n}\phi_j^{ND,AD}\Delta q_t^{AD} + \sum_{j=1}^{n}\phi_j^{ND,NI}\Delta q_t^{NI} + \sum_{j=1}^{n}\phi_j^{ND,ND}\Delta q_t^{ND} + \varepsilon_t^{ND}
\end{aligned}
\tag{6}
$$

where $q_t' = (q_t^{HI}, q_t^{HD}, q_t^{AI}, q_t^{AD}, q_t^{NI}, q_t^{ND})$ is the transposed quote-midpoint vector. For each stock-day we obtain the optimal lag length (n in eq. [6]) using the SBIC. We determine the number of linearly independent co-integration relationships (the co-integration rank) using the trade statistic proposed by Johansen (1995). Under the assumption that the difference of any two quote midpoint series in q is co-integrated of order (1,1), the co-integration rank should be equal to five. This is actually the case for all stock-days except for 10 cases. We exclude those abnormal stock-day observations. For the same reason, the co-integrating matrix β should look like

$$\beta = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

We do not restrict the β coefficients, but our estimates corroborate the above assumption about this matrix with the co-integrating vectors being the difference between two of the quote midpoints in q. Finally, the error-correction vector $\alpha^{j\prime} = \left(\alpha^{j,HI}, \alpha^{j,HD}, \alpha^{j,AI}, \alpha^{j,AD}, \alpha^{j,ND}, \alpha^{j,HI}\right)$, for j = {HI, HD, AI, AD, ND, HI} captures the sensitivity of the j-th quote to deviations from other trader-category/order-type quotes.

The VEC model [6] can be written in a more compact form as

$$\Delta q_t = \alpha\beta' q_{t-1} + B(L)\Delta q_{t-1} + \varepsilon_t \qquad [6']$$

The VMA representation of [6'] is

$$\Delta q_t = \Psi(L)\varepsilon_t \qquad [7]$$

Co-integration entails $\beta'\Psi(1) = 0$, with $\Psi(1) = \sum_{j=1}^{\infty} \Psi_j$ (e.g., Engle and Granger, 1987). Under the assumption in [6], Hasbrouck (1995) shows that all the rows of the impact matrix $\Psi(1)$ are identical

$$\Psi(1) = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_6 \end{pmatrix} = \begin{pmatrix} \psi \\ \vdots \\ \psi \end{pmatrix}$$

the long-run impact becomes $\psi\varepsilon_t$, and the long-run variance is

$$Var(\Delta m_t) = Var(\psi\varepsilon_t) = \psi\Omega\psi' \qquad [8]$$

To solve the identification problems that arise when the contemporaneous correlation between innovations is non-negligible, Hasbrouck (1995) suggests using the Cholesky factorization of $\Omega = FF'$, so that the IS for a given innovation is

$$IS_j = \frac{\left([\psi F]_j\right)^2}{\psi\Omega\psi'} \qquad [9]$$

where $[\psi F]_j$ is the j-th element of the row vector ψF.[12] The resulting factorization, however, depends on the order of the variables in the q_t vector. Equation [9] will allocate a greater IS to the first quote in vector q_t.

Hasbrouck (1995) proposes to obtain upper and lower bounds on the IS of each quote by rotating the ordering of the variables in the q vector. Unfortunately, that implies that the IS approach can only determine the contribution of each market or quote within a range. The width of this range depends on the contemporaneous correlation across quotes (e.g., Huang, 2002).

Baillie, Booth, Tse, and Zabotina (2002) and de Jong (2002) both show that the price impact vector ψ and $\alpha_\perp$, the orthogonal vector of the error-correction term $\alpha_\perp\alpha' = 0$, are equal up to a scale factor π, $\psi = \pi\alpha_\perp$, that drops out in the IS measure in [9]. This result largely simplifies the computation of the IS since it is not necessary to obtain the VMA representation of the VEC model. Using this result, we compute the upper and lower bounds of the IS of each trader-category/order-type pair as

$$IS_j = \frac{\left([\alpha_\perp F]_j\right)^2}{\alpha_\perp\Omega\alpha_\perp'} \qquad [10]$$

[12] With correlated innovations the ISs are not identified since the covariance terms could be arbitrarily allocated between quotes.

As in former analyses, the ISs are estimated for each stock-day, and we report the average IS across stock-days. Statistical significance is clustered by stock and day (e.g., Thompson, 2011).