# Price discovery - contribution of proprietary and agency algorithmic traders

Samarpan Nawn[a]

Ashok Banerjee[b,*]

Current Version: November 2017

* Corresponding Author Email: ashok@iimcal.ac.in

[a] PhD student, Finance & Control, Indian Institute of Management Calcutta, India

[b] Finance & Control, Indian Institute of Management Calcutta, India

We thank Avanidhar Subrahmanyam, Suresh Sundaresan, David Yermack and Gautam Mitra for their feedback on the paper as part of the IIM Calcutta Finance Research Workshop 2016 and the seminar participants of ICFMCF 2017 at Indian Institute of Technology Kharagpur for some useful suggestions and its organising committee for the Best paper award.

# Price discovery - contribution of proprietary and agency algorithmic traders


## Abstract

We investigate the relative roles of limit orders from proprietary algorithmic traders (PAT) and agency algorithmic traders (AAT) in the discovery of security prices in National Stock Exchange (NSE) of India. Our results suggest that AAT contributes more to price discovery compared to PAT. This challenges the modern notion that “speed is information,” since PAT includes the high-frequency traders (HFT). As AAT are primarily hired by institutional investors to execute their orders, our result shows that stock specific information is still in the hands of those traders who were traditionally viewed as informed. We also find no evidence to support the popular notion that fast traders use limit orders to “mislead” market participants about future price movements.



*JEL Classification code: G14*
*Keywords: HFT, limit orders, quote, market manipulation*

## 1. Introduction

One of the major functions of a capital market is price discovery of securities (O'Hara 2003). Algorithmic traders' contribution to this task in an order driven market is understudied and prompts further investigation. To understand the price discovery impact of algorithmic traders, it is essential to consider their various types. Algorithmic traders can be broadly classified into two categories, agency algorithmic traders (AAT) and proprietary algorithmic traders (PAT). AAT primarily provide execution services to clients for reducing their trading costs of large trades. PAT, on the other hand, look to profit from the trading process itself and are mainly known for its subgroup High frequency traders (HFT).

The motivation of the two types of traders is very different. Although the clients of AAT can practically be anyone, more often they are institutional investors such as mutual funds, foreign institutional investors, insurance companies, etc., who need and can afford the services. It is perceived by many that institutional investors are typically more informed (Badrinath et al., 1995; Chakravarty, 2001) and thus orders from AAT, which involve executing the instructions of institutional investors, can contribute heavily to price discovery. The view that institutional investors are informed is, however, traditional. Modern view suggests that order execution speed is information in today's market (O'Hara, 2015), implying orders of PAT are indeed a bigger source of price discovery. We attempt to examine the orders of which group of traders – AAT or PAT – is more information driven.

Academic research on AT/HFT using developed market data faces two problems (Biais and Foucault, 2014). The first problem is the difficulty with the identification of AT/HFT. Most of the exchanges do not provide order level information on algorithmic trading, and researchers have largely been forced to use some proxies to identify machine trading (Hasbrouck and Saar, 2013; Hendershott et al., 2011). While some studies employ flagged data from Canada (Brogaard et al., 2016) and Germany (Hendershott and Riordan 2013), the segregation is only dichotomous (HFT vs. non-HFT or AT vs. non-AT). Having only one class of machine traders can potentially convolute many findings as proprietary and agency algorithmic traders are quite dissimilar. The second problem is that developed markets are by and large fragmented, and most studies employ data from a single exchange. Machine traders are likely to take positions in multiple exchanges at the same time, and thus results based on data from a single exchange are often inadequate to conclude about their behaviour. The data we choose to employ address both concerns. Firstly, NSE[1] of India provides flagged order level data on whether the order is from an algorithmic terminal or not and whether the order is for a client account or a proprietary account. Secondly, the stock market in India has only two exchanges, and NSE itself has 75% volume of the country's cash equities segment[2]. Thus the Indian market can be considered largely unfragmented.

Along with the two algorithmic trading groups, we consider the non algorithmic traders (NAT) as our third and final group. We use the Hasbrouck (1995) framework for several co-integrated price series to find out the information shares of each group of traders. Brogaard et al. (2016) also use Hasbrouck (1995) method in their dichotomous data (HFT/non-HFT) to conclude that HFT's limit orders contribute significantly to the price discovery process than non-HFTs limit orders. But they conclude so by looking at the simple averages of the best bid and best ask provided by two trading groups. We argue the price measure should consider the quantities at the limit order book (LOB) along with the respective prices, to make the metric a closer representation of the prices offered by each trading group. The price measure should be a consensus average price measure of that trading group. It is also of interest to make the

[1] NSE is one of the top 15 stock exchanges of the world and the approximate market capitalization of $1.7 Trillion. In terms of size, NSE is very close to Deutsche Boerse, the primary German exchange. http://economictimes.indiatimes.com/markets/stocks/news/the-17-most-valuable-stock-exchanges-in-the-world/articleshow/54013184.cms accessed on 20th Nov 2017

[2] http://www.moneycontrol.com/stocks/marketstats/turnover/ accessed on 5th July 2017

selection wider by considering top five quotes instead of just the top, as these form the freely visible and hence decision inducing part of the LOB[3].

Why the quantities should matter is explained with the help of an illustration. We show a hypothetical (but realistic) limit order book (Figure 1) for a stock at one particular point in time. The bid price is 99, and the ask is 101. The average best price (using Brogaard et al., 2016) is 100 for both HFTs and non-HFTs. A closer look at the LOB suggests that HFT is ask side heavy and non-HFT is bid side heavy. The weighted price for HFT is 100.2 (using only the top quote) and for non-HFT is 99.8. If we use the top five quotes then the weighted consensus price for each trading group diverges even further and becomes 101.4 and 99.6 respectively.

**Figure 1**

| | Bid side | | | Ask side | |
|---|---|---|---|---|---|
| | Quantity | Price | | Quantity | Price |
| Non-HFT | 30 | 99 | Non-HFT | 20 | 101 |
| HFT | 10 | 99 | HFT | 15 | 101 |
| Non-HFT | 40 | 99 | HFT | 50 | 101 |
| Non-HFT | 30 | 99 | HFT | 70 | 101 |
| Non-HFT | 30 | 98 | HFT | 50 | 102 |
| Non-HFT | 20 | 97 | Non-HFT | 30 | 102 |
| Non-HFT | 15 | 96 | HFT | 25 | 103 |
| Non-HFT | 30 | 95 | Non-HFT | 25 | 104 |
| HFT | 20 | 95 | Non-HFT | 30 | 104 |
| Non-HFT | 10 | 95 | HFT | 20 | 104 |
| | | | Non-HFT | 10 | 105 |
| | | | HFT | 20 | 105 |

The contribution of a trading group to price discovery should also be considered on a depth-adjusted basis. Let us assume that trading group A contributes only 10% of the total depth available at the top 5 quotes, while trading group B contributes 70%. Also suppose that

[3] Cao et al. (2009) find the information content of both the mid-quote and quotes behind the top is significant

on the Hasbrouck metric, group A's information share is 35% and group B's is 40%. It would be inappropriate to conclude only on the basis of information share that group B contributes more to price discovery. On a depth adjusted basis, group A's limit orders convey more information.

In a depth-adjusted way, we find that limit orders of AAT contribute the most to price discovery. This result challenges the notion of "speed is information" and suggests that information in stock prices are still incorporated through the investors who were traditionally viewed as informed since AAT primarily executes the orders of institutional investors. Interestingly, using Brogaard et al. (2016) definition of best prices, we find the information share of NAT is maximum. We discuss in the results section why our method provides different findings.

Analyzing the limit orders of algorithmic traders is also important from a regulatory point of view. Regulators[4] seem to suggest that machine traders manipulate stock prices with the excessive use of order cancellations and order revisions. This criticism is labelled strongly against the HFT. Since most of the market manipulation strategies (e.g., layering and spoofing) involve quotes just behind the top, our study examines how limit orders supplied by various trading groups in top five quotes affect the stock price. Cao et al. (2009) and Harris and Panchapagesan (2005) find the imbalance between buy and sell side of the limit order book is associated positively with short-term future returns. We compute buy minus sell order imbalance of the limit order book for the three trading classes and regress one-second future returns on them. If all trading groups post only genuine or non-spoofing orders then the relationship between their limit order imbalance and future returns is expected to be positive as per Cao et al. (2009). However, if any trading group regularly manipulates the market with

[4] Securities and Exchange Commission (SEC) and Financial Industry Regulatory Authority, Inc. (FINRA) in the U.S., Securities and Exchange Board of India (SEBI) in India and also regulators of the other countries

limit orders and they do it successfully then the corresponding relationship between their limit order imbalance and future returns will not be positive.

Our results find the coefficient of order imbalance of PAT is positive and significant. Our evidence suggests PAT do not use limit orders to manipulate or mislead the market and their limit order imbalance imparts more information on the future stock price movement than the limit order imbalance of other traders.

Our sample covers 134 stocks divided into 50 large capitalization (LCAP) and 84 small and mid cap (SMCAP) stocks. This helps us examine the impact of three trading groups on two different market capitalization classes. Although the literature has mostly concentrated on large stocks in analyzing AT/HFT, we find that AAT does contribute significantly to the price discovery process in small and mid cap stocks also.

We contribute to the literature in more than one way. *First*, we use exact identification marks to separate types of algorithmic traders, and our results are based on an unfragmented market. *Second*, we use consensus average price measures of each trading group instead of simple averages to recognize the effect of order size on price discovery process. *Third*, it highlights the differential impact of the proprietary and agency algorithmic traders on the discovery of security prices. Not much is known about AAT from existing literature. Although some studies distinguish between proprietary and agency algorithmic trading (Hagströmer and Nordén 2013; Hasbrouck and Saar 2013), the focus is more on proprietary algorithms and HFT. We find that AAT group serves a very important function of the market – helping the price discovery process more than other trading groups, including the PAT/HFT. *Fourth*, we also find that limit order imbalance of PAT in the top few quotes positively predicts future stock return. This should allay the fears that these traders indulge in market manipulating tactics such as spoofing etc. The coefficient is not only positive and significant but is also on average ten times more than the coefficient for NAT.

Rest of the paper is organized the following way. Section 2 looks at the related literature and hypotheses. Section 3 describes data and section 4 the summary statistics. Section 5 discusses the empirical results. Section 6 concludes.

## 2. Motivation and hypothesis

The presence of algorithmic trading in stock markets has intrigued academic researchers. The research on AT has concentrated on the issues of liquidity (Hendershott et al. 2011; Hendershott and Riordan 2013) and price discovery. The literature so far has documented mainly positive effects of AT/HFT on price discovery. Using the 2003 introduction of auto quote at the NYSE as an instrument for establishing causality from AT, Hendershott et al. (2011) show that AT is related to faster price discovery[5]. Brogaard et al. (2014) document that HFT plays an important role in price discovery using the 2008-2009 Nasdaq sample of 26 HFTs. O'Hara, Yao, and Ye (2014) show that trades of 100 or fewer shares, sizes commonly used by HFTs and broker execution algorithms, account for a substantial portion of price discovery. Boehmer, Fong, and Wu (2015) find that introduction of co-location facilities is associated with faster price discovery. However, Biais et al. (2014) suggest that HFT often trade on market data (i.e., order flow, volume, etc.) rather than producing information themselves. They add that without HFT, new information would be anyways incorporated into prices, albeit a bit slowly (in a few seconds rather than in a few milliseconds). Jovanovic and Menkveld (2016) suggest that HFT trading on information which is soon to be public can have mixed welfare effects. While on the one hand, it can increase gains from trade but on the other, it can impose adverse selection costs on others.

[5] See also Hendershott and Riordan (2009)

Brogaard et al. (2014) also note that the information set used by HFT is often short-lived and if the information would anyway become public in a few seconds then the adverse selection costs can make the welfare gains small or negative. Hirschey (2013) finds evidence that HFT extracts information from the order flow of other traders and trades ahead of them.

Hagstromer and Norden (2013) and Hasbrouck and Saar (2013) discuss the difference between proprietary and agency algorithmic trading without highlighting the differential impact these two groups of traders have on the market. Hendershott and Riordan (2009) suggest that examining the different type of ATs can provide insight into their respective application to particular trading strategies. Nawn and Banerjee (2017) find PAT have a positive impact on market liquidity and their behaviour mimics that of the traditional market makers. They also find that AAT withdraw liquidity during short-term market stress. Till date, literature has not considered which type of the algorithmic traders contributes more to price discovery.

Brogaard et al. (2016) suggest that price discovery is now primarily through quotes than through trades[6]. O'Hara (2015) argues that with changing market structure and availability of real-time market data feed, HFTs can turn public information to private information, albeit for a very short time. Over this short time, the information could well become more order related than asset related. Further, Haldane (2011) suggests that to be uninformed is to be slow. He asserts that in a high-speed world, being informed means observing and reacting to market movements faster than the others. This line of argument brings us to believe that PAT/HFT would contribute the most to price discovery.

The view – speed is information - challenges the traditional description of informed traders. Academic research has consistently found, over the last couple of decades, that institutional

[6] Also see Bloomfield et al. (2005) and Kaniel and Liu (2006)

traders are informed, primarily because they have greater ability and sophistication to unearth information through fundamental analysis. Chakravarty (2001) finds that trades generated by institutions account for a disproportionately large cumulative stock price change relative to their share of trades. Badrinath et al. (1995) find returns on the portfolio of stocks with the highest level of institutional ownership lead returns on portfolios of stocks with lower levels of institutional ownership[7].

Chan and Lakonishok (1995) find institutions like to split their orders, possibly to hide information from other market participants and to avoid price concessions. The introduction of AT provided the institutions with algorithmic execution services to perform the same. Domowitz and Yegerman (2006) suggest that orders executed through algorithms have lower implementation shortfall costs than other orders, providing evidence on how agency algorithmic traders provide value to their clients. Engle et al. (2012) use execution data from Morgan Stanley algorithms to study the trade-offs between algorithm aggressiveness and the mean and dispersion of execution cost. The fact AAT primarily execute the orders of institutional traders' leads us to the possibility that AAT contribute the most to price discovery. The two contradictory views bring us to an important question to explore - which category of traders contribute more to price discovery and bring us to the following testable hypothesis:

**Hypothesis 1 (H1)**: PAT contribute the most to price discovery

Relatedly, HFT are accused of "playing" with the limit orders to mislead other market participants[8]. Regulators seem to suggest that HFT manipulate the stock prices with the

[7] Also see Arbel and Strebel (1983), Lee (1992) and Sias and Starks (1997)

[8] The common market manipulative strategies include layering, spoofing, smoking, etc. Layering is a strategy where a trader makes and then cancels orders that they never intend to execute in hopes of influencing the stock price. Spoofing is placing large orders behind the best quote in the direction opposite to his actual interest. These orders are not meant to execute but to spoof other market participants that there is news in one direction. Smoking is posting inviting limit orders to attract slow traders and then revising the orders to less generous conditions by the time the market orders from slow traders arrive.

excessive use of order cancellations. They are believed to post artificial orders just behind the best bid and ask to manipulate stock prices. For example, if an HFT wants to buy, he may temporarily place large sell orders above the best ask, which is not meant for execution. The large limit sell order may scare others investors who might presume that there is a negative sentiment in the stock and update their price quotes down; the HFT can then buy at a lower price to accomplish his original interest. If a trading group uses such strategies, a direct implication is that their limit order imbalance would be a poorer predictor of future short-term return in comparison to the imbalances of other trading groups. Thus, another interesting issue would be to test whether HFT tries to manipulate stock markets by posting deceptive limit orders.

**Hypothesis 2 (H2):** PAT limit order imbalance a poorer predictor of future short-term returns than the imbalances of other trading groups.

## 3. Data

NSE is a pure order-driven market without any designated market maker. Trading takes place by an order matching mechanism – matching the marketable sell orders with best available limit buy orders and matching marketable buy orders with best available limit sell orders. After a 15 minute pre-opening session, the exchange is ready for continuous trading at 9:15 AM. Trading is conducted on weekdays excluding public holidays, in a single continuous session from 9:15 AM to 3:30 PM. For all traded securities, the exchange freely displays live, five best bid and ask quotes and the number of shares demanded or offered at those quotes.

Direct market access and algorithmic trading have been allowed in India since April 2008. However, high-frequency and automated trading, in reality, took off in India with the launch of NSE's co-location services[9] in January 2010.

[9] Co-location service allows renting rack space to traders with low latency connectivity to the exchange with the mandatory power supply, cooling and security requirements of the industry

Each order message in the data, obtained from NSE, is earmarked by an algorithmic flag and a client flag. The two flags help identify whether an order was from an algorithmic terminal and whether the order was for a client account or a proprietary account. Combining the two flags, the traders can be partitioned into three mutually exclusive and exhaustive classes – Proprietary algorithmic traders (PAT), Agency algorithmic traders (AAT) and Non algorithmic traders (NAT).

We consider two disjoint months of data for our work – October 2012[10] and March 2013. The choice of sample months was due to an event on 1st January 2013 - the exchange reduced co-location charges by 50% to make algorithmic trading feasible for small and medium traders, who till then were unable to afford the facility. The event could potentially change the nature of algorithmic trading in the market. We did not want our results to be materially affected by this and thus choose sample months on either side of the event avoiding two immediate months. We find our results are independent of the event and similar across October and March. NSE of India has 1541 stocks listed during our sample period. We exclude the stocks whose median daily traded volume was less than Rs 100000, the median traded quantity was less than 1000, the median number of trades was less than 100 and the median price was less than Rs 1 over the two month sample period. We also exclude stocks whose market capitalization was not available in Bloomberg during our sample period. We were left with 670 stocks in the population after applying these filters. We select one-fifth of the stocks randomly from the population to form our sample. The chosen 134 stocks are then sorted by market capitalization. The largest 50 out the 134 is marked as large capitalization stocks, whereas the smallest 50 is marked as small capitalization stocks. The rest 34 stocks are grouped as mid capitalization stocks. The mid cap and small cap group are similar in terms of number of trades, the number of shares traded, price etc. and hence for the rest of the paper, these

[10] We exclude 5th Oct 2012 from our sample as the day was in spotlight due to a mini flash crash of several large cap stocks

are bunched together as mid and small (84) capitalization group (SMCAP). In sync with the literature on price discovery, we use time interval of 1 second. We build the limit order book at 1 second frequency for our sample of 134 stocks for two months.

## 4. Summary Statistics

Table 1 summarizes broad features of our sample. The large capitalization (LCAP) stocks naturally have higher average prices, more rupee volume and number of trades than the small and mid capitalization (SMCAP) stocks. Our dichotomous classification of stocks would help us contrast the algorithmic trader's participation in big and small stocks.

**Table 1**

| | Market Capitalization Group | |
|---|---|---|
| | LCAP | SMCAP |
| Market Capitalization (In Rs. Millions) | 141842.93 | 59325.92 |
| Price | 456.17 | 208.30 |
| Number of Shares traded | 312459.38 | 397228.14 |
| Money value of trades (In Rs. Millions) | 128.49 | 44.74 |
| Number of transactions | 6800.89 | 3633.76 |

The table provides daily cross-sectional average statistics on our sample of 134 stocks. Cross-sectional averages are computed over daily averages per stock. Our sample period includes 2 months, October 2012 and March 2013. The 134 stocks are partitioned into 50 large capitalization (LCAP) and 84 small and mid capitalization (SMCAP) stocks.

Table 2 charts the differences in order and trade attributes among three trader groups – AAT, PAT and NAT. We find that for the LCAP stocks, the PAT group, by far, sends the most number of messages to the exchange, followed by AAT and then NAT. The pattern is similar for SMCAP stocks. The order and trade sizes are lowest for AAT for both the market capitalization groups. This is particularly significant as AAT provide execution services to clients which primarily involve executing large orders split through algorithms. Message to trade ratio, an important metric to understand the functioning of a trading class, is very small for NAT and is about 30 to 50 times higher for the algorithmic groups. The HFT characterise the PAT group and this is evident in their more frequent messaging pattern than the AAT.

**Table 2**

| | LCAP | | | SMCAP | | |
|---|---|---|---|---|---|---|
| | AAT | PAT | NAT | AAT | PAT | NAT |
| Number of orders | 6171.6 | 8965.8 | 7818.9 | 883.9 | 887.2 | 2537.1 |
| Number of messages | 52316.2 | 237196.3 | 17484.5 | 9821.2 | 11504.2 | 3795.3 |
| Size of orders | 170.6 | 597.7 | 292.7 | 108.5 | 188.6 | 404.8 |
| Number of trades | 4038.5 | 1367.0 | 5912.9 | 411.8 | 233.8 | 1857.3 |
| Size of trades | 52.9 | 71.3 | 79.4 | 57.5 | 64.9 | 142.6 |
| Message to trade | 42.5 | 146.0 | 3.3 | 120.9 | 200.3 | 3.5 |

The table provides daily cross-sectional average statistics on our sample of 134 stocks. Cross-sectional averages are computed over daily averages per stock. Our sample period includes 2 months, October 2012 and March 2013. PAT stands for proprietary algorithmic traders, AAT stands for agency algorithmic traders and NAT stands for non algorithmic traders

**Table 3**

| | LCAP | SMCAP |
|---|---|---|
| Quotes 1 to 5 | | |
| NAT | 67.8% | 86.7% |
| PAT | 18.3% | 5.5% |
| AAT | 13.9% | 7.8% |
| Top Quote | | |
| NAT | 73.2% | 87.0% |
| PAT | 7.6% | 3.9% |
| AAT | 19.2% | 9.1% |
| Quotes 2 to 5 | | |
| NAT | 67.2% | 86.6% |
| PAT | 20.0% | 5.8% |
| AAT | 12.8% | 7.6% |

The table provides daily cross-sectional average statistics on our sample of 134 stocks. Cross-sectional averages are computed over daily averages per stock. Our sample period includes 2 months, October 2012 and March 2013. PAT stands for proprietary algorithmic traders, AAT stands for agency algorithmic traders and NAT stands for non algorithmic traders. The numbers represent the share of order book depth across the three trading groups.

We examine the percentage of order book depth, at the one-second interval, supplied by the different group of traders at the top 5 quotes. For the 50 LCAP stocks, NAT accounts for

68% of the depth while algorithmic traders account for the rest 32%[11]. The share of PAT (AAT) within AT is 18% (14%). The distribution over the top 5 quotes is not similar though for the two algorithmic players. While PAT accounts for only 8% of the top quote depth, the number increases to 20% when considering depth at quotes in positions 2 to 5. For AAT the pattern is completely reversed as the percentage decreases from 19% to 13%[12]. PAT seem to position their quotes just behind the best prices. These positions are significant for two reasons -a) any quote stationed there are less likely to be traded instantaneously compared to the top quote, and b) NSE displays top 5 positions live to all traders free of charge. So, substantial orders placed at positions 2 to 5 possibly can be used for market manipulative purposes. We discuss, in our results section, the analysis of our second research question (H2) related to this issue. Overall we observe that non algorithmic traders are undisputed leaders in supplying depth over the LOB (Table 3).

## 5. Empirical results

### *5.1 H1 – Price discovery*

We use the method developed by Hasbrouck (1995) to identify relative information content of different co-integrated price series. We briefly discuss the method[13] here. We consider three representative price series corresponding to price quotes of three trading groups and we discuss the choice of price series later.

Let $X_t$ be the vector of these three price series. Each price series is individually non-stationary, but they share a common underlying trend as they are the prices of the same underlying security. Thus the difference between any pair of them is stationary and they are co-integrated. The co-integration implies that they may be represented in a vector error correction model (VECM) of order k

[11] The pattern of order book depth for the SMCAP stocks is similar
[12] This reversal is not so pronounced in case of SMCAP stocks
[13] See Hasbrouck (1995) and Huang (2002) for a detailed discussion

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_1^k л_i X_{t-i} + \varepsilon_t \,, \qquad (1)$$

Where $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \Omega$, $\alpha$ is the error correction vector, $\beta$ is the co-integrating vector and $л_i$ are matrices of autoregressive coefficients. This model can also be represented in a vector moving average (VMA) process

$$X_t = X_0 + \psi(1) \sum_{i=1}^t \varepsilon_i + \psi * (L)\varepsilon_t, \qquad (2)$$

The matrix $\psi(1)$ contains the sums of all the moving average coefficients, and $\psi^*(L)$ contains elements that are scalar polynomials in argument L (lag operator). In the expression of $X_t$ in (2), the first and the last terms are constant and stationary respectively. The middle term captures the stochastic trend common to all the price series and is the driving force that results in co-integration. This term captures the permanent impact of new information on prices and excludes all transient effects.

Under this set up, we propose to use the Hasbrouck information share, which for the jth price series is defined as

$$S_j = \frac{\Psi_j^2 F_{jj}}{\Psi\Omega\Psi'} \qquad (3)$$

Where $\Psi_j$ is the jth element of $\Psi$, one of the identical rows of $\Psi(1)$ and F is the lower triangular Cholesky factorization of $\Omega$. By using matrix F, the components of the innovations are orthogonalized and by suitably placing one price at a time to first and last position (in vector X) one gets the maximum and minimum information share, respectively. The information share reported for each price series is the average of the maximum and minimum information share for the same. This procedure is widely adopted in the literature (Booth et al., 2002, Cao et al., 2009).

Our choice of X vector comprises of three price series corresponding to three trading groups. The price series we consider is the order size weighted average quotes sitting in the

LOB within top five quotes, coming from a particular trading group. For the PAT group, the price series at any particular time is defined in the following way,

$$P_{PAT} = \sum_{j=1}^{5}(QD_j^{PAT} * P_j^{Buy} + QS_j^{PAT} * P_j^{Sell}) / \sum_{j=1}^{5}(QD_j^{PAT} + QS_j^{PAT}) \quad (4)$$

Where $QD_j$ is quantity demanded at price level j, $QS_j$ is quantity supplied at price level j. The index j runs from 1 to 5 representing the top 5 prices. For example, when j = 1 it would represent the top quotes and $P_1^{Buy}$ would be the best bid price and $P_1^{Sell}$ would be the best ask price. The price series for the other two trading groups are defined analogously.

We argue that this price measure provides a consensus price for the trading group. Brogaard et al. (2016) also apply the Hasbrouck (1995) method to get price discovery contribution from HFT and non-HFT, but they choose only the simple average best price of HFT (and non-HFT) as the representative price corresponding to the particular trading group. We have already explained how considering only the best prices ignore the information present in quantities and other prices spread across the LOB (figure 1).

The weighted average consensus price would better estimate the average informativeness of a trading class. Suppose, there are 15 ‘A’ traders out of which one is a super trader and is informed. The rest 14 are uninformed. Also suppose there are 10 ‘B’ traders, each of them being less informed than the super trader but more informed than the other 14 ‘A’ traders. Now Brogaard et al. (2016) method in the Hasbrouck price discovery metric would almost at all times pick the super trader’s quote while computing trader group ‘A’s price discovery. This would imply that ‘A’ traders are more informed, whereas the average informativeness of ‘B’ traders is more.

For the 50 LCAP stocks (Panel A of Table 4), the average information share of AAT is 42% and is nearly the same as that of NAT (43%). The information share of PAT is much lower at 16%. We find (in Table 3) that the depth share of AAT in top 5 quotes is 14%

whereas that of NAT is 68%. Despite the clear dominance of NAT over AAT in occupying LOB positions, we find that the information share is nearly similar for these two trader groups. Thus, on a depth adjusted basis, the price discovery contribution (ratio of information share over depth share) of AAT is 3 whereas that of NAT is 0.6. Similarly, for PAT, the number is 0.9. The depth adjusted information share measure shows how much a participant contributes to price discovery with respect to their mere presence. The pattern of price discovery is similar for the SMCAP stocks with the highest contribution from AAT on a depth adjusted basis.

The sheer volume of limit orders would naturally have an effect on the information shares. The NAT contributes 68% of the depth in the top 5 quotes for the LCAP stocks. With such a large depth share, it is imperative that their limit orders would contribute a significant part of the price discovery. To counter that effect we consider the depth adjusted information ratios and results dramatically change. The ratios indicate that limit orders of AAT introduce information in stock prices the most, for both the LCAP and SMCAP groups. The orders of PAT score much lower information shares - while PAT's contribution numbers are greater than NAT, the fact that they lag much behind AAT is interesting and entails further explanation.

We find that the distribution of PAT's orders over the top 5 quotes is not uniform (Table 3). The majority of their orders are stored at positions 2 to 5 avoiding the top quote. AAT, on the other hand, provide more depth orders at top quote than at positions 2 to 5. AAT, in general, seem to be the more informed and therefore position themselves at top quote more often as the probability of execution is highest at the top. PAT, while being present at top 5 quotes, look to avoid the top quote in particular. As their orders do not contribute much to the price discovery, it seems that they do not trade on the basis of stock specific information. They rather wait for profitable trading opportunities such as the change in market conditions, change in order flow, etc. Nawn and Banerjee (2017) find that PAT increase the supply of limit orders at top few quotes when the short-term stock volatility is high - exploiting an opportunity to earn extra bid-ask spread during these times. AAT execute the orders of institutional investors who spend considerable time and efforts in collecting stock specific information. This is reflected in the results - AAT contribute the most to the price discovery of securities.

**Table 4**

Panel A

| | | NAT | | | PAT | | | AAT | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | MIN | MAX | AVE | MIN | MAX | AVE | MIN | MAX | AVE |
| Large market capitalization | Mean | 35% | 51% | 43% | 11% | 21% | 16% | 35% | 49% | 42% |
| | S.D. | 11% | 8% | | 5% | 6% | | 9% | 13% | |
| | IS Depth adjusted | | | 0.6 | | | 0.9 | | | 3.0 |
| Mid & Small market capitalization | Mean | 39% | 52% | 46% | 11% | 19% | 15% | 34% | 46% | 40% |
| | S.D. | 13% | 10% | | 5% | 6% | | 11% | 14% | |
| | IS Depth adjusted | | | 0.5 | | | 2.7 | | | 5.1 |

Panel B

| | | NAT | | | PAT | | | AAT | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | MIN | MAX | AVE | MIN | MAX | AVE | MIN | MAX | AVE |
| Large market capitalization | Mean | 48% | 72% | 60% | 14% | 30% | 22% | 11% | 30% | 20% |
| | S.D. | 18% | 10% | | 8% | 19% | | 4% | 9% | |
| Mid & Small market capitalization | Mean | 57% | 69% | 63% | 9% | 17% | 13% | 20% | 29% | 25% |
| | S.D. | 12% | 13% | | 5% | 11% | | 12% | 11% | |

The table provides Hasbrouck information share (IS) estimates on our sample of 134 stocks. Our sample period includes 2 months, October 2012 and March 2013. PAT stands for proprietary algorithmic traders, AAT stands for agency algorithmic traders and NAT stands for non algorithmic traders. For panel A, the price series corresponding to a trading group for computing IS is the depth weighted average price across the top 5 quotes of the LOB for that group. For panel B, the price series corresponding to a trading group for computing IS is the simple average of the best price on bid and ask side respectively for that group.

We also discuss the results using the Brogaard et al. (2016) definition of price series (Panel B of Table 4). By taking only the simple average best prices, we ignore the information contained in the quantities demanded and supplied in the LOB and also the information present in the quotes just behind the top. We find the average information share of the NAT is more than 60%. The information shares of PAT and AAT are close to 20% each under this method. Judging by the reported standard deviations, we note that the information share numbers for PAT are statistically not significant, while those for AAT and NAT are. Notably, these information share numbers are very similar to the depth share at top quotes (Table 3). The trading group which has the maximum presence at the top of the book will automatically dominate others in terms of information share, under this method.

In order to make our results comparable to the price discovery analysis of Brogaard et al. (2016), we re-arrange our data to construct two trading classes – PAT (analogous to HFT) and OTHERS (analogous to non-HFT). We present the information share estimates using the weighted average consensus price (Table 5 Panel A) and the best quotes only (Table 5 Panel B) for the two trading groups. Expectedly, the informativeness of the individual classes AAT and NAT gets withered when considered as a single group OTHERS. As a consequence, the information share of PAT increases under both methods compared to the three trading group set up. The depth adjusted information share for PAT is higher than that of OTHERS for both LCAP and SMCAP. If we had only looked at this dichotomous classification (PAT and OTHERS) then we would have also concluded that PAT contributes the most to price discovery. However, our main results (Table 4) clearly show that this is not the case.

**Table 5**

Panel A

| | | PAT | | | OTHERS | | |
|---|---|---|---|---|---|---|---|
| | | MIN | MAX | AVE | MIN | MAX | AVE |
| Large market capitalization | Mean | 23% | 34% | 29% | 66% | 77% | 72% |
| | S.D. | 10% | 8% | | 8% | 10% | |
| | IS Depth adjusted | | | 1.6 | | | 0.9 |
| Mid & Small market capitalization | Mean | 34% | 44% | 39% | 56% | 66% | 61% |
| | S.D. | 15% | 13% | | 13% | 15% | |
| | IS Depth adjusted | | | 7.0 | | | 0.7 |

Panel B

| | | PAT | | | OTHERS | | |
|---|---|---|---|---|---|---|---|
| | | MIN | MAX | AVE | MIN | MAX | AVE |
| Large market capitalization | Mean | 18% | 36% | 27% | 64% | 82% | 73% |
| | S.D. | 7% | 15% | | 8% | 10% | |
| Mid & Small market capitalization | Mean | 30% | 38% | 34% | 62% | 70% | 66% |
| | S.D. | 15% | 15% | | 15% | 15% | |

The table provides Hasbrouck information share (IS) estimates on our sample of 134 stocks.. Our sample period includes 2 months, October 2012 and March 2013. PAT stands for proprietary algorithmic traders, OTHERS stand for agency algorithmic traders and non algorithmic traders. In panel A, the price series corresponding to a trading group for computing IS is the depth weighted average price across the top 5 quotes of the LOB for that group. In panel B, the price series corresponding to a trading group for computing IS is the simple average of the best price on bid and ask side respectively for that group.

*5.2 H2-Market manipulation*

We compare the quantities available in the limit order book in both buy and sell side and construct an imbalance measure between the two for each trading group. Our objective here is to identify whether PAT uses their quotes to mislead the market about the future price movement. We want to define the time unit which suitably characterizes the “future” returns for testing our hypothesis. Market manipulative quotes are expected not to last long and are likely to be cancelled very quickly. SEC data show that 23% of all cancellations happen within 50 milliseconds ($\frac{1}{20}$ th of a second) of order placement and by 500 milliseconds ($\frac{1}{2}$ of a second) the number increase to 38%[14]. Thus by one second (1000 milliseconds), we expect the market manipulative quotes to achieve what they are meant for. For example, consider the case of a spoofing trader whose real intention is to buy and therefore has posted a large limit sell order just above the ask price. The speed at which modern markets operate, one second is possibly enough time for the price to be temporarily driven down by the "manipulative" large limit sell order only to be pushed up again by a "real" buy order from the spoofing trader. Of course, when he posts the final buy order, in parallel he cancels the large spoofing sell order. The example explains the assumed non-positive relation between lagged limit order imbalance of spoofing traders and one-second returns.

We regress mid quote returns on the lagged limit order imbalance measures of three trading groups. As controls, we include lagged values of mid quote returns due to negative autocorrelation in returns (Roll, 1984) and return persistence (Poterba and Summers, 1988). Formally, the empirical model considered here is:

$$r_t = \propto + \beta_1 PATLOBIMB_{t-1} + \beta_2 AATLOBIMB_{t-1} + \beta_3 NATLOBIMB_{t-1} + \gamma_1 r_{t-1} + \gamma_2 r_{t-2} + \gamma_3 r_{t-3} + \varepsilon_t \qquad (5)$$

[14] https://www.sec.gov/marketstructure/research/highlight-2013-05.html#.Wg_eL1uCzIU accessed on 20th November 2017

**Table 6**

| | | α | β1 | β2 | β3 | γ1 | γ2 | γ3 |
|---|---|---|---|---|---|---|---|---|
| Top quote | LCAP | 0.00 | 48.89 | 17.4 | 3.33 | -5.84 | -1.34 | -0.17 |
| | | (0.89) | (9.95) | (3.48) | (2.89) | (-3.47) | (-1.09) | (-0.52) |
| | n=50 | [18,6] | [47,1] | [30,0] | [30,1] | [0,43] | [2,20] | [5,13] |
| | SMCAP | 0.00 | 117.72 | 85.48 | 5.19 | -5.12 | -2.75 | -2.15 |
| | | (0.13) | (2.90) | (1.93) | (1.01) | (-2.29) | (-1.52) | (-1.27) |
| | n=84 | [19,8] | [51,4] | [44,4] | [30,4] | [0,51] | [2,38] | [1,28] |
| Quote 2 to 5 | LCAP | -0.00 | 5.49 | 3.8 | 0.71 | -5.43 | -1.57 | -0.05 |
| | | (-0.99) | (5.19) | (2.74) | (2.65) | (-5.13) | (-1.89) | (-0.98) |
| | n=50 | [1,15] | [42,0] | [34,1] | [32,0] | [0,43] | [2,27] | [4,12] |
| | SMCAP | -0.00 | 9.33 | 5.83 | 1.16 | -7.71 | -3.36 | -2.6 |
| | | (-0.63) | (1.86) | (0.92) | (2.18) | (-5.16) | (-2.84) | (-2.83) |
| | n=84 | [6,27] | [44,11] | [29,6] | [46,1] | [0,69] | [1,54] | [0,57] |
| Top 5 quotes | LCAP | 0.00 | 7.44 | 6.8 | 1.59 | -5.44 | -1.65 | -0.05 |
| | | (0.07) | (7.38) | (5.54) | (4.93) | (-5.13) | (-1.90) | (-1.01) |
| | n=50 | [10,10] | [48,0] | [44,0] | [40,0] | [0,43] | [2,27] | [4,12] |
| | SMCAP | -0.00 | 22.55 | 11.43 | 1.7 | -6.86 | -3.36 | -2.6 |
| | | (-0.60) | (3.64) | (1.99) | (3.16) | (-5.19) | (-2.83) | (-2.84) |
| | n=84 | [8,22] | [62,3] | [46,2] | [55,3] | [1,69] | [1,54] | [0,57] |

The table presents the GMM estimates from the regressions estimated for each of the 134 stocks based on one-second time intervals. The regression model is

$$r_t = \propto + \beta_1 PATLOBIMB_{t-1} + \beta_2 AATLOBIMB_{t-1} + \beta_3 NATLOBIMB_{t-1} + \gamma_1 r_{t-1} + \gamma_2 r_{t-2} + \gamma_3 r_{t-3} + \varepsilon_t$$

$r_t$ = return at time interval t, $PATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from PAT in the LOB at time t, $AATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from AAT in the LOB at time t, $NATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from NAT in the LOB at time t. The first number in each cell is the cross-sectional median of the estimated regression coefficients. The second number (in parentheses) is the median of the t-statistics from the 50 individual regressions. The third group involving two numbers (in brackets) is the number of positive and significant t-statistics and the number of negative and significant t-statistics out of n stocks.

**Table 7**

| | | α | β1 | β2 | β3 | γ1 | γ2 | γ3 |
|---|---|---|---|---|---|---|---|---|
| Top quote | LCAP | 0.00 | 47.29 | 35.96 | 5.09 | -5.43 | -1.77 | -0.19 |
| | | (0.5) | (6.03) | (3.1) | (2.56) | (-2.62) | (-0.85) | (-0.52) |
| | n=50 | [17,6] | [41,0] | [39,0] | [30,1] | [0,34] | [3,13] | [2,6] |
| | SMCAP | 0.00 | 93.59 | 171.46 | 5.82 | -5.38 | -3.15 | -1.31 |
| | | (0.31) | (1.42) | (1.65) | (0.72) | (-1.5) | (-1.25) | (-0.68) |
| | n=84 | [10,9] | [35,7] | [41,4] | [24,5] | [0,38] | [2,24] | [3,22] |
| Quote 2 to 5 | LCAP | -0.00 | 6.17 | 4.39 | 2.38 | -6.75 | -2.17 | -0.39 |
| | | (-1.12) | (5.2) | (1.74) | (2.96) | (-4.56) | (-1.85) | (-0.92) |
| | n=50 | [2,17] | [33,0] | [26,0] | [30,0] | [0,44] | [1,26] | [6,11] |
| | SMCAP | -0.00 | 12.77 | 4.43 | 4.64 | -7.49 | -4.05 | -2.43 |
| | | (-0.38) | (1.23) | (0.39) | (1.6) | (-2.81) | (-1.87) | (-1.56) |
| | n=84 | [4,13] | [29,4] | [17,6] | [39,1] | [0,59] | [2,47] | [2,39] |
| Top 5 quotes | LCAP | -0.00 | 10.35 | 11.39 | 2.99 | -6.76 | -2.16 | -0.40 |
| | | (-0.56) | (5.82) | (4.71) | (3.61) | (-4.57) | (-1.88) | (-0.92) |
| | n=50 | [3,10] | [41,0] | [39,0] | [37,0] | [0,44] | [1,26] | [6,12] |
| | SMCAP | -0.00 | 24.84 | 29.83 | 4.89 | -7.45 | -4.00 | -2.42 |
| | | (-0.33) | (1.6) | (1.43) | (2.31) | (-2.79) | (-1.86) | (-1.56) |
| | n=84 | [5,14] | [39,2] | [34,3] | [48,1] | [0,59] | [2,47] | [2,39] |

The table presents the GMM estimates from the regressions for the subsample where order cancellation rate of PAT is greater than its third quartile. The regressions are estimated for each of the 134 stocks based on one-second time intervals. The regression model is

$$r_t = \propto + \beta_1 PATLOBIMB_{t-1} + \beta_2 AATLOBIMB_{t-1} + \beta_3 NATLOBIMB_{t-1} + \gamma_1 r_{t-1} + \gamma_2 r_{t-2} + \gamma_3 r_{t-3} + \varepsilon_t$$

$r_t$ = return at time interval t, $PATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from PAT in the LOB at time t, $AATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from AAT in the LOB at time t, $NATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from NAT in the LOB at time t. The first number in each cell is the cross-sectional median of the estimated regression coefficients. The second number (in parentheses) is the median of the t-statistics from the 50 individual regressions. The third group involving two numbers (in brackets) is the number of positive and significant t-statistics and the number of negative and significant t-statistics out of n stocks.

We estimate equation (5) for all 134 stocks using generalized method of moments (GMM) and obtain t-statistics that are robust to heteroskedasticity and autocorrelation (Newey and West method (1987)). We consider three variants of limit order imbalance – a) imbalance in top quote, b) imbalance in quotes 2 to 5 and c) imbalance in top 5 quotes. Our results show that PAT limit order imbalance is a positive predictor of immediate short-term future return. The coefficient of interest ($\beta_1$) is positive and significant consistently across the three variants of order imbalance and two market capitalization groups (Table 6). $\beta_1$ is, on the average, twice the value of the $\beta_2$ (the coefficient of AAT) and ten times the value of $\beta_3$ (the coefficient of NAT). For example, considering the LCAP stocks and order imbalance at the top quote, the cross-sectional average value of $\beta_{1,}$ $\beta_2$ and $\beta_3$ are respectively 48.89, 17.4 and 3.33, with average T-statistics 9.95, 3.48 and 2.89 respectively. Out of the 50 stocks, 47 stocks have a positive and significant value of $\beta_1$ and only 30 stocks have a positive and significant value of $\beta_2$ or $\beta_3$. Thus we find PAT limit order imbalance not only positively predicts future short-term return, the prediction is better than those of AAT or NAT. Our results contradict the apprehension of regulators that PAT (or HFT) manipulate the stock prices with their limit orders. The results, in fact, establish that order imbalance of PAT is at least as authentic and informative predictor of short-term future price movement as the order imbalance of AAT or NAT.

The $\beta_1$ estimate provides an average over the entire sample identifying whether PAT in general use limit orders to manipulate market prices. One may argue, PAT may not manipulate at all times but at particular times only. SEC and FINRA allege that HFT firms engage in manipulative strategies involving the use of order cancellations[15]. It is more likely that PAT manipulates market when their order cancellation rates are high. To identify such a suitable subsample, we consider the data in groups of one-minute time intervals and compute

[15] High-Frequency Trading: Background, Concerns, and Regulatory Developments: www.fas.org/sgp/crs/misc/R43608.pdf accessed on 20th November 2017

the order cancellation rate of PAT as the number of cancellation messages by PAT in the time interval over the total number of messages by PAT in the interval. The time intervals for which the order cancellation rate is greater than its third quartile is used to construct the chosen subsample.

We re-estimate the model (5) for this subsample and examine whether the estimated $\beta_1$ indicates any market manipulation. Table 7 provides the details. On the average $\beta_1$ still positively predicts short-term returns in this subsample. For example, for the LCAP stocks, and order imbalance taken from top 5 quotes, the coefficient for 41 stocks are positive and significant here compared to 48 in the whole sample results. The average coefficients of PAT imbalance in Table 7 are close to the coefficients of AAT imbalance and are substantially large compared to the coefficients of NAT. Overall, we find the results in Table 6 are robust and PAT do not appear to use market manipulative strategies even when they are most likely to do so.

The results of Table 6 suggests that order imbalance of PAT is a better predictor of short-term returns than the imbalances of other traders. It is interesting to understand whether PAT posses any information or they gather information from imbalances of other investors. Goldstein et al. (2017) suggest HFT is more likely to supply liquidity on that side of the order book which is already thick. Dahlstrom et al. (2017) find more than other traders, HFT use existing depth imbalance to track fundamental values. To test the source of PAT's information content, we partition the sample into two subsamples – S1, where PAT's order imbalance is unlikely to be affected by the order imbalance of other traders and S2, where it may be affected and run the model (5) on S1 and S2 separately.

**Table 8**

| | | α | β1 | β2 | β3 | γ1 | γ2 | γ3 |
|---|---|---|---|---|---|---|---|---|
| Top quote | LCAP | 0.00 | 47.55 | 21.72 | 3.1 | -5.2 | -1.07 | -0.36 |
| | | (0.99) | (7.72) | (2.72) | (2.83) | (-2.7) | (-0.62) | (-0.62) |
| | n=50 | [18,5] | [45,0] | [33,0] | [30,0] | [0,41] | [3,9] | [2,10] |
| | SMCAP | 0.00 | 108.89 | 49.19 | 9.28 | -5.35 | -2.79 | -1.55 |
| | | (0.29) | (2.17) | (1.36) | (0.85) | (-1.81) | (-1.33) | (-0.92) |
| | n=84 | [15,6] | [47,3] | [37,5] | [33,2] | [0,46] | [2,35] | [3,22] |
| Quote 2 to 5 | LCAP | -0.00 | 0.74 | 1.33 | 0.37 | -6.26 | -2.19 | -0.24 |
| | | (-0.62) | (4.17) | (1.01) | (0.94) | (-5.84) | (-2.12) | (-1.27) |
| | n=50 | [7,9] | [26,12] | [16,3] | [15,6] | [0,46] | [2,30] | [4,19] |
| | SMCAP | -0.00 | -2.65 | 1.57 | 0.11 | -6.94 | -3.77 | -3.04 |
| | | (-0.57) | (-0.59) | (0.56) | (0.22) | (-4.53) | (-2.78) | (-2.64) |
| | n=84 | [7,20] | [9,25] | [18,7] | [19,11] | [0,67] | [1,58] | [2,53] |
| Top 5 quotes | LCAP | -0.00 | 2.1 | 7.24 | 1.09 | -6.3 | -2.27 | -0.25 |
| | | (-0.54) | (6.18) | (4.69) | (2.99) | (-5.84) | (-2.12) | (-1.25) |
| | n=50 | [10,10] | [31,2] | [37,0] | [34,0] | [0,46] | [2,30] | [4,19] |
| | SMCAP | -0.00 | 4.8 | 8.49 | 1.36 | -6.93 | -3.77 | -3.04 |
| | | (-0.53) | (0.9) | (1.5) | (2.41) | (-4.55) | (-2.79) | (-2.65) |
| | n=84 | [4,19] | [31,4] | [37,3] | [44,2] | [1,67] | [1,58] | [2,53] |

The table presents the GMM estimates from the regressions for the subsample S1 where i) the sign of PAT limit order imbalance and the sign of limit order imbalance of other traders are different or ii) the sign of the two are same but absolute value of the imbalance of others is less than 25% of the absolute value of the imbalance of PAT. The regressions are estimated for each of the 134 stocks based on one-second time intervals. The regression model is

$$r_t = \propto + \beta_1 PATLOBIMB_{t-1} + \beta_2 AATLOBIMB_{t-1} + \beta_3 NATLOBIMB_{t-1} + \gamma_1 r_{t-1} + \gamma_2 r_{t-2} + \gamma_3 r_{t-3} + \varepsilon_t$$

$r_t$ = return at time interval t, $PATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from PAT in the LOB at time t, $AATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from AAT in the LOB at time t, $NATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from NAT in the LOB at time t. The first number in each cell is the cross-sectional median of the estimated regression coefficients. The second number (in parentheses) is the median of the t-statistics from the 50 individual regressions. The third group involving two numbers (in brackets) is the number of positive and significant t-statistics and the number of negative and significant t-statistics out of n stocks.

**Table 9**

| | | α | β1 | β2 | β3 | γ1 | γ2 | γ3 |
|---|---|---|---|---|---|---|---|---|
| Top quote | LCAP | 0.00 | 60.43 | 20.5 | 2.58 | -5.98 | -1.19 | -0.31 |
| | | (0.93) | (6.6) | (2.76) | (1.99) | (-2.63) | (-0.73) | (-0.24) |
| | n=50 | [17,5] | [44,0] | [33,0] | [26,1] | [0,36] | [0,11] | [3,8] |
| | SMCAP | 0.00 | 105.01 | 127.5 | 10.77 | -5.65 | -2.52 | -2.17 |
| | | (0.11) | (2.1) | (1.53) | (1.07) | (-1.8) | (-1.11) | (-1.06) |
| | n=84 | [16,10] | [44,6] | [39,2] | [22,5] | [1,43] | [2,22] | [2,22] |
| Quote 2 to 5 | LCAP | -0.00 | 17.32 | 3.22 | 0.56 | -5.56 | -1.53 | -0.06 |
| | | (-0.84) | (9.49) | (1.96) | (1.33) | (-3.99) | (-1.46) | (-0.9) |
| | n=50 | [5,15] | [46,0] | [29,3] | [22,4] | [0,40] | [1,23] | [4,18] |
| | SMCAP | -0.00 | 45.48 | 7.73 | 0.81 | -6.82 | -3.07 | -2.25 |
| | | (-0.59) | (3.2) | (0.85) | (0.97) | (-3.48) | (-1.98) | (-1.92) |
| | n=84 | [8,22] | [58,2] | [23,5] | [30,10] | [0,67] | [2,47] | [0,49] |
| Top 5 quotes | LCAP | 0.00 | 17.05 | 6.39 | 0.89 | -5.56 | -1.54 | -0.07 |
| | | (0.24) | (10.35) | (4.27) | (2.57) | (-4.00) | (-1.46) | (-0.9) |
| | n=50 | [6,9] | [46,0] | [37,2] | [29,0] | [0,40] | [1,23] | [4,17] |
| | SMCAP | -0.00 | 51.55 | 15.72 | 1.22 | -6.88 | -3.02 | -2.25 |
| | | (-0.49) | (4.22) | (1.65) | (1.41) | (-3.49) | (-1.98) | (-1.94) |
| | n=84 | [7,20] | [64,1] | [41,2] | [39,7] | [1,66] | [2,47] | [0,50] |

The table presents the GMM estimates from the regressions for the subsample S2 where the sign of PAT limit order imbalance and the sign of limit order imbalance of other traders are the same and the absolute value of the imbalance of other traders is more than or equal to 25% of the absolute value of the imbalance of PAT.. The regressions are estimated for each of the 134 stocks based on one-second time intervals. The regression model is

$$r_t = \propto + \beta_1 PATLOBIMB_{t-1} + \beta_2 AATLOBIMB_{t-1} + \beta_3 NATLOBIMB_{t-1} + \gamma_1 r_{t-1} + \gamma_2 r_{t-2} + \gamma_3 r_{t-3} + \varepsilon_t$$

$r_t$ = return at time interval t, $PATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from PAT in the LOB at time t, $AATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from AAT in the LOB at time t, $NATLOBIMB_t$ = defined as the difference between the number of shares at available at bid side minus the number of shares at available at ask side among the orders coming from NAT in the LOB at time t. The first number in each cell is the cross-sectional median of the estimated regression coefficients. The second number (in parentheses) is the median of the t-statistics from the 50 individual regressions. The third group involving two numbers (in brackets) is the number of positive and significant t-statistics and the number of negative and significant t-statistics out of n stocks.

S1 includes the cases where i) the sign of the PAT limit order imbalance and the sign of limit order imbalance of other traders are different or ii) the sign of the two are same but absolute value of the imbalance of other traders is less than 25% of the absolute value of the imbalance of PAT. S2 includes all other cases which do not fall in S1. Table 8 and Table 9 respectively present the results for estimating model (5) on S1 and S2. For both subsamples, $\beta_1$ is on the average positive, but the significance is much greater for the sample S2. For example, for the SMCAP stocks, and order imbalance taken from top 5 quotes, $\beta_1$ from S1 is 4.8 (with average T-statistic 0.9), and that from S2 is 51.55 (with average T-statistic 4.22). The number of stocks out 84 for which the individual $\beta_1$ is significant is 31 for S1 and 64 for S2. When PAT order imbalance is not influenced by the imbalances of other traders, its predictable power diminishes. Whereas when PAT limit order imbalance is in sync with the imbalances of other traders, with PAT possibly drawing information from the order imbalances of other traders, its predictable power improves sharply.

## 6. Conclusion

Price discovery contribution of algorithmic traders is largely understudied, the primary reason being the dearth of specific data which explicitly identifies algorithmic trading. In the situation where the data is available, it generally identifies one class of machine traders. There are at least two broad categories of algorithmic traders – proprietary and non-proprietary, and the motivation for each is very different. We use the data provided by NSE of India that identifies orders by three mutually exclusive and exhaustive classes of traders – proprietary algorithmic traders, agency algorithmic traders and non algorithmic traders, and estimate their contribution to price discovery.

We find that order and trade sizes are lowest for the AAT among the three trading groups. This is because AAT provide execution services to institutional investors which primarily involve executing large orders split through algorithms. We also find that in a depth-adjusted way, the AAT provide maximum contribution to price discovery. This is a significant result as it shows that stock specific information probably still lies with “traditional” informed investors, the institutional investors. This result challenges the popular notion that “speed is information” as PAT (the superset of fast traders HFT) load low on price discovery.

We also examine whether there is any merit in the accusation by some, including the regulators, that HFT manipulate the market by excessive usage of order cancellations. The allegation is HFT provides certain short duration orders for misleading the market participants regarding the market sentiment. We test this by regressing the one-second return on lagged limit order imbalance for the three trading groups. If any trading group use the limit order book to temporarily obscure prices, then their limit order imbalance will be a poor predictor of short-term future return. But we find PAT’s order imbalance has a positive and highly significant coefficient in predicting the short-term return. The coefficient is even greater than that of AAT or NAT order imbalance. The result suggests that HFT are not generally harmful to the market.

## References:


Arbel, A., & Strebel, P. (1983). Pay attention to neglected firms!. *The Journal of Portfolio Management*, *9*(2), 37-42.

Badrinath, S. G., Kale, J. R., & Noe, T. H. (1995). Of shepherds, sheep, and the cross-autocorrelations in equity returns. *Review of Financial Studies*, *8*(2), 401-430.

Biais, B., & Foucault, T. (2014). HFT and market quality. *Bankers, Markets & Investors*, *128*, 5-19.

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

Boehmer, E., Fong, K. Y., & Wu, J. J. (2015). International evidence on algorithmic trading.

Booth, G. G., Lin, J. C., Martikainen, T., & Tse, Y. (2002). Trading and pricing in upstairs and downstairs stock markets. *Review of Financial Studies*, *15*(4), 1111-1135.

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

Brogaard, J., Hendershott, T., & Riordan, R. (2016). Price discovery without trading: Evidence from limit orders.

Cao, C., Hansch, O., & Wang, X. (2009). The information content of an open limit-order book. *Journal of futures markets*, *29*(1), 16-41.

Chakravarty, S. (2001). Stealth-trading: Which traders' trades move stock prices?. *Journal of Financial Economics*, *61*(2), 289-307.

Chan, L. K., & Lakonishok, J. (1995). The behavior of stock prices around institutional trades. *The Journal of Finance*, *50*(4), 1147-1174.

Dahlström, P., Hagströmer, B., & Norden, L. L. (2017). Determinants of Limit Order Cancellations, Working paper

Domowitz, I., & Yegerman, H. (2006). The cost of algorithmic trading: A first look at comparative performance. *The Journal of Trading*, *1*(1), 33-42.

Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251-276.

Goldstein, M. A., Kwan, A., & Philip, R. (2017). High-frequency trading strategies, Working Paper

Hagströmer, B., & Norden, L. (2013). The diversity of high-frequency traders. *Journal of Financial Markets*, *16*(4), 741-770.

Haldane, A, 2011. The race to zero. Bank of England speeches, given to the International Economic Association 16th World Congress, July 8.

Harris, L. E., & Panchapagesan, V. (2005). The information content of the limit order book: evidence from NYSE specialist trading decisions. *Journal of Financial Markets*, *8*(1), 25-67.

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The journal of Finance*, *50*(4), 1175-1199.

Hasbrouck, J., & Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, *16*(4), 646-679.

Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity?. *The Journal of Finance*, *66*(1), 1-33.

Hendershott, T., & Riordan, R. (2009). Algorithmic trading and information. *Manuscript, University of California, Berkeley*.

Hendershott, T., & Riordan, R. (2013). Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis*, *48*(04), 1001-1024.

Hirschey, N. 2013. Do high-frequency traders anticipate buying and selling pressure? Working Paper.

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

Jovanovic, B., & Menkveld, A. J. (2016). Middlemen in limit order markets, Working Paper.

Kaniel, R., & Liu, H. (2006). So what orders do informed traders use?. *The Journal of Business*, *79*(4), 1867-1913.

Lee, C. M. (1992). Earnings news and small traders: An intraday analysis. *Journal of Accounting and Economics*, *15*(2-3), 265-302.

Nawn, S., & Banerjee, A. (2017). Do Proprietary Algorithmic Traders Withdraw Liquidity During Market Stress? Working Paper.

Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.Econometrica, 55(3), 703-708.

O'Hara, M. (2003). Presidential address: Liquidity and price discovery. *The Journal of Finance*, *58*(4), 1335-1354.

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

O'hara, M., Yao, C., & Ye, M. (2014). What's not there: Odd lots and market data. *The Journal of Finance*, *69*(5), 2199-2236.

Poterba, J. M., & Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. Journal of financial economics, 22(1), 27-59.

Robert, E., Robert, F., & Jeffrey, R. (2012). Measuring and modeling execution cost and risk. *The Journal of Portfolio Management*, *38*(2), 14-28.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. The Journal of Finance, 39(4), 1127-1139.

Sias, R. W., & Starks, L. T. (1997). Return autocorrelation and institutional investors. *Journal of Financial economics*, *46*(1), 103-131.