Do High Frequency Traders Provide or Drain Liquidity?
A Study of the Market Pre-Opening Period on the
Tokyo Stock Exchange

Preliminary and incomplete

Mario Bellia, SAFE - Goethe University *
Loriana Pelizzon, Goethe University and Ca’ Foscari University of Venice †
Marti G. Subrahmanyan, Leonard N. Stern School of Business - New York University ‡
Jun Uno, Waseda University and Ca’ Foscari University of Venice §
Darya Yuferova, Rotterdam School of Management - Erasmus University ¶

March 2015

Abstract
Liquidity provision and price discovery are two important functions of financial markets. The fundamental changes witnessed by equity markets over the past decade, in particular due to the growing presence of High Frequency Traders (HFTs), have prompted a re-examination of how these two functions have been affected. In this vein, the key questions that we examine in this paper are (a) whether High Frequency Traders (HFTs) in equity markets provide or drain liquidity, and (b) whether HFTs merely amplify noise or lead to an improvement in the price discovery process, during the market pre-opening period. To address these questions, we exploit the natural experiment of the introduction of a low latency trading platform by the Tokyo Stock Exchange (TSE) in January 2010 to study the HFT order submission strategies. We utilize a novel data-set on server IDs provided by the TSE, which allows us to distinguish HFT from non-HFT orders, for our empirical analysis. HFTs take advantage of TSE’s low-latency trading facility and place a number of new orders and revisions up to two seconds, and cancel existing orders as little as 130 milliseconds, prior to the opening time. We document that aggressive small orders entered earlier in the pre-opening period have a higher likelihood of cancellation, and less aggressive orders entered later have a higher likelihood of revision. These results suggest that quote updates by HFTs in the last instants prior to the opening have characteristics similar to those of high frequency liquidity providers. At the same time, strategic order submission by HFTs causes a delay of 460 milliseconds in price discovery in the pre-opening period since 3 years after "Arrowhead" inception. However, cancellations arriving in the last milliseconds prior to the opening do not increase the volatility of the pre-opening quotes. Overall, we conclude that HFTs do provide liquidity while at the same time slowing down the price discovery process marginally, without magnifying the price fluctuations, during the pre-opening period.

Key-words: High Frequency Traders (HFTs), Order Submission, Order Cancellation, Pre-Opening, Price Discovery, Liquidity Provision

*bellia@safe.uni-frankfurt.de, Theodor W. Adorno Platz 3, 60323 Frankfurt am Main, Germany
†pelizzon@safe.uni-frankfurt.de, Theodor W. Adorno Platz 3, 60323 Frankfurt am Main, Germany
‡msubrahm@stern.nyu.edu, NY 10012-1126 New York, USA.
§juno@waseda.jp 1-4-1 Nihombashi, Chuo-ku, Tokyo 103-0027, Japan
¶dyuferova@rsm.nl, PO Box 1738, 3000 DR Rotterdam, The Netherlands
1. Introduction

During the past decade, global equity markets have been fundamentally altered due to the vast improvements in the speed of trading and the consequent fragmentation of market activity. Among other changes, traditional market makers have been replaced by high frequency traders (HFTs), in most markets.\(^1\) This replacement has had a dramatic impact on the behavior of liquidity providers in financial markets. There has been intense debate and scrutiny by investors, market makers, exchanges, and regulators regarding the advantageous, even unfairly advantageous position of HFTs in global markets.\(^2\) However, we still know very little about the determinants of high frequency liquidity provision. Key questions that we ask in this research are whether HFTs provide or drain liquidity, and whether HFTs amplify noise or lead to an improvement in the price formation process, during the market pre-opening period. We contribute to the literature on high frequency trading with a clear focus on liquidity provision and price discovery in the pre-opening period.

Our research follows earlier work in two distinct areas. The first relates to findings regarding the microstructure of trading activity in the market pre-opening period, while the second relates to the impact of HFTs. The pattern of market pre-opening trading has been studied in the earlier literature (e.g., by Amihud and Mendelson (1991), Biais, Hillion, and Spatt (1999), Ciccotello and Hatheway (2000), Madhavan and Panchapagesan (2000)). However, much of this literature is dated, and is based on research conducted well before the rapid growth of HFTs, over the course of the past decade or so. It is, therefore, necessary to re-examine trading activity in the pre-opening period once again, given the vast changes that have occurred, since the advent of HFT activity.

The literature on HFTs is relatively sparse, given that these institutional changes are recent and the data are only recently becoming available. However, this literature is growing rapidly.\(^3\) It should be noted, however, that the focus of most of the literature is the continu-

---

\(^1\)See Hendershott and Riordan (2009), Brogaard (2010), Jovanovic and Menkveld (2011), and Raman and Yadav (2014) for details.

\(^2\)See Lewis (2014) for a humorous, popular, albeit one-sided discussion of HFT behavior.

\(^3\)For a review of the burgeoning literature, see Jones (2013) and Biais and Foucault (2014).
ous trading session, rather than the pre-opening session of the trading day. SEC (2010) separates HFT strategies into the following four groups: market making (as in Menkveld (2013), Brogaard, Hagströmer, Norden, and Riordan (2013)), arbitrage (as in Foucault, Kozhan, and Tham (2014)), directional strategies (as in Hirschey (2013), Brogaard, Hendershott, and Riordan (2014), and Scholtus, van Dijk, and Frijns (2014)), and structural strategies (as in McInish and Upson (2012)). The first strategy provides liquidity to the market, while the latter three consume liquidity from the market. Baron, Brogaard, and Kirilenko (2012) and Hagströmer and Norden (2013) empirically confirm the separation of HFTs into those who mainly use limit orders and those who mainly use market orders. For the purpose of this paper, we focus on market making by HFT in the pre-opening period. In an early study, Menkveld (2013) analyzes transactions of a large HFT firm that is active on the NYSE-Euronext and Chi-X markets, right after Chi-X started as an alternative trading venue for European financial markets. He shows that, in 80% of the cases, HFTs provided liquidity on both markets, during the continuous trading session. In an event study framework, Brogaard, Hagströmer, Norden, and Riordan (2013) show that liquidity providers are willing to pay for higher trading speed (using a premium co-location service which allows traders to co-locate their servers near the exchange’s matching machine with upgraded transmission speed), and that this is beneficial for overall market liquidity.

It has also been documented in this literature that the opening price is characterized by higher volatility relative to other times during the trading day, including the closing.\textsuperscript{4} This conclusion has been attributed to many factors, including the accumulation of information, the concentration of orders overnight, and market-maker intervention.\textsuperscript{5} We are able to shed new light on this phenomenon by employing a rich, new database to study how high frequency liquidity providers build their positions before the market opening, and how they increase


\textsuperscript{5}Gomber, Arndt, Lutat, and Uhle (2011), Menkveld (2013), and Kirilenko, Kyle, Samadi, and Tuzun (2014) document typical behavior of HFT during the continuous trading session starting with a zero inventory position at the beginning of the trading day. However, they do not describe how they prepare their positions during the pre-opening session, in anticipation of the continuous trading session.
the efficiency of price formation at the market opening.\textsuperscript{6} To the best of our knowledge, these are issues that have not been studied so far in the literature in any depth, a lacuna that we aim to fill.

In January 2010, the TSE implemented a major improvement in its trading architecture by introducing a low latency platform, known as "Arrowhead", along with a new design of the intra-day auction, a fundamental change that could affect the behavior of traders.\textsuperscript{7} We view this change as a natural experiment, in which the introduction of the Arrowhead system is an exogenous event that triggered an abrupt change in the behavior of the traders. To test our hypotheses, we use a novel data-set on server IDs provided by the Tokyo Stock Exchange (TSE) to distinguish the originators of orders. We classify the servers into two groups based on their trade-to-quote ratio and the cancellation rate. Since HFT liquidity providers are sensitive to latency (as noted by Hasbrouck and Saar (2013) and Brogaard, Hendershott, and Riordan (2014)), they are likely to use servers with the lowest trade-to-quote ratio and the highest cancellation rate. Thus, we can classify the types of order submitters based on the server used, and determine whether HFTs are liquidity makers or liquidity takers, and whether their activity leads to better price discovery in the pre-opening period. Our data-set is unique in that it provides complete coverage of HFT in a large international equity market, compared to prior studies that obtained limited information about a market, either from one HFT firm (as in Menkveld (2013)) or from a smaller market (as in Brogaard, Hagström, Norden, and Riordan (2013)).

We distinguish between three types of market participant who benefit from the increased speed of trading. First are the HFTs, who engage in liquidity provision during the course of the continuous trading session, but also build their positions during the pre-opening period by submitting a set of limit orders to the book, as early as the inception of trading, because the

\textsuperscript{6}For studies on HFT and market quality see Hendershott, Jones, and Menkveld (2011), Easley, de Prado, and O’Hara (2012), Hendershott and Riordan (2013), Malinova, Park, and Riordan (2013), Boehmer, Fong, and Wu (2014), and Brogaard, Hendershott, and Riordan (2014).

\textsuperscript{7}Other papers that investigate the effect of increased exchange latency are Riordan and Storkenmaier (2012), Menkveld and Zoican (2013) and Ye, Yao, and Gai (2013).
time priority of orders in the continuous session is important for them. They modify their orders accordingly thereafter, in anticipation of the opening price. This characterization allows us to distinguish orders submitted by HFTs for market-making purposes from orders submitted for liquidity-taking purposes.

Second are the institutional investors, who are willing to execute large orders at the market opening but will not enter them into the order book until the very last moment (perhaps the last millisecond prior to the opening), as these orders may have a significant impact on the opening price.\(^8\) The early entry of large orders has clear disadvantages: large orders attract other participants and induce other investors to react sooner, causing a deterioration in the execution price of large orders. Additionally, in most markets, there is no time-priority applied to orders submitted during the pre-opening period and executed at the opening single price auction.

Third, are aggressive investors may enter "noisy" orders and cancel them right before the execution takes place.\(^9\) The term "noisy" connotes a type of order that uses an aggressive limit price to send a signal to investors on the opposite side, to induce them to provide liquidity. Indeed, some investors may have an incentive to enter false orders with aggressive limit prices to elicit a favorable response from true orders from the opposite side. While this strategy does not always work to the advantage of the aggressive investor, it may serve to add noise to the pre-opening quotes.

All three types of market participant benefit from a low latency of trading. The faster execution of trades enables them to delay their final action until very close to the market opening. Therefore, the noise effects may prevail up to the final seconds in the pre-opening period. If so, it is useful to investigate the type of order submission that causes a deterioration of the pre-opening quotes.

Our empirical results for the TSE show a dramatic shift in order submission up to the last seconds before the opening time. This phenomenon also affects the efficiency of pre-opening

---

\(^8\)Kraus and Stoll (1972), Chan and Lakonishok (1993), and Chiyachantana, Jain, Jiang, and Wood (2004) study the price impact of institutional trades.

\(^9\)For a study about price manipulation please see Ye, Yao, and Gai (2013)
quotes as predictors of opening prices. The submission of new orders rises to its highest frequency a bit earlier than does the cancellation of orders. During the last minute before opening, 16% of orders submitted have an impact on mid-quotes. However, reverses in the sign of the price change from the previous day’s close make up only 0.6% of all events. This indicates that most orders that have an impact on quotes generate only minor changes. The unbiasedness of "pre-quotes" reaches its highest level 680 milliseconds before the opening time right after the inception of new system in January-March 2010, and 220 milliseconds in April-May 2013, when the HFT participation rate exceeded 50% (Hosaka (2014)). The path by which the unbiasedness approaches its highest level was much smoother in April and May 2013 than in January to March 2010, providing evidence that high-frequency quote updates contribute to price discovery.

We find that aggressive small orders entered in the earlier part of the pre-opening period have a higher likelihood of cancellation, and less aggressive orders entered in the later period have a higher likelihood of revision. The concentration of order submissions towards the end of the pre-opening time causes a delay in price discovery. However, cancellations occurring in the last 500 milliseconds do not increase the volatility of the pre-opening quotes.

The purpose of disseminating pre-opening quotes is to provide a good indication of the current opening price. However, the results for the TSE indicate that it might mislead market participants. Increasing competition among low-latency trading creates a complex price formation process that all investors need to know about. Institutions utilizing algorithmic trading tools need to make a careful assessment of the pre-opening order submission activities.

Our empirical design and hypotheses are presented in Section 2. The empirical results, including the data description, the characterization of pre-opening quote behavior, and the results on order revisions and cancellations, are described in Section 3. Section 4 concludes.

2. Empirical design and Hypotheses

2.1. A new trading platform

On January 4, 2010, the TSE launched a new trading system named Arrowhead. The main features of this system are (i) accelerated computer-processing speeds, (ii) a co-location...
service that reduces the physical distance between market participants (investors as well as brokerage firms), and (iii) removing the three second interval from intra-day matching. Thus, January 2010 can be viewed as the time of arrival of a new trading paradigm in Japan.

Each trading day, the TSE starts receiving orders from brokers at 8 am, and the single price auction for the market opening begins at 9 am. As soon as it receives orders, the TSE disseminates the pre-opening quotes to the market. The pre-opening quotes consist of ask and bid prices and their associated quantities. In the case of the TSE, the best bid and ask prices during the pre-opening period are determined in a fashion that is different from that of the best bid and ask prices during the continuous session. In particular, the pre-opening best bid and ask prices are the respective prices at which the demand and supply schedules (two step-functions with cumulative volume on $x$-axis and price on $y$-axis) intersect. The lowest (highest) price between the two is the reported best bid (ask) price. Thus, the best bid and ask during the pre-opening period are the most likely (possible) opening prices. They are not the lowest sell limit price and highest buy limit price as in the continuous session.

Table 1 shows the relative frequencies of order types in the whole period and relevant sub-periods. In the entire pre-opening period, new orders make up about 60%, and cancellations and price revisions, roughly 20% each. The number of shares for new orders is, on average, 4,244, for cancellations, 3,955, and for price revisions, 1,913, respectively. In the last ten minutes and the last one minute of the pre-opening period, the share of new orders drops to less than 50%, and those of cancellations and price revisions increase accordingly.

INSERT TABLE 1 HERE.

2.2. Design of the empirical study

We select our universe of stocks from the constituents of the TOPIX100 index, which comprises of the stocks on the TSE's first section with the highest liquidity. The literature on HFTs, such as Uno and Shibata (2012), Menkveld (2013), and Brogaard, Hendershott, and Riordan (2014) suggest that large institutional investors and traders generally prefer high-liquidity stocks in Europe, the US, and Japan. Among the top 100 stocks, we exclude three stocks that have larger trading volumes in exchanges other than the TSE, since the
The focus of our study is the trading system on this exchange.\textsuperscript{10}

The sample period we select for our analysis is between April 1 and May 31, 2013. In this period, the volatility of the stock market rose after the new Governor of the Bank of Japan, Haruhiko Kuroda, announced a new aggressive Quantitative Easing policy. A number of unexpected events occurred in this period, making the role of the pre-opening quotes even more crucial than at any other time. For purposes of benchmarking, we will refer to the period from November 2009 though March 2010 as the comparative (control) period, since the TSE introduced the Arrowhead system on January 4, 2010. This initial month of January 2010 gives us the opportunity to examine the turning point of the TSE’s platform change and its effect on order submission behavior, with the other months being used for robustness checks to capture the announcement effect.

We exclude stock-days when special quotes are disseminated before or during the single price auction, because orders submitted during the pre-opening period do not meet the opening price rules. We keep track of the mid-quote between the best ask and bid second-by-second during the pre-opening period. We use three data sources for our analysis, in addition to the identities of the server IDs: Nikkei Tick Data, Thomson-Reuters Tick History and TSE Order Submission Data. The third dataset allows us to keep track of the order submission timing as well as updates of each order during the market pre-opening. It also includes unique identifiers for the server connections between brokers and the exchange. The order submission characteristics of each server reflect whether a user is a HFT or not. Typically, HFTs request their broker to provide access to an exclusive server; if so, we can identify the server IDs that are most likely used by individual HFTs, as described in detail later.

2.3. Hypotheses

Among market participants, many, particularly institutional investors, seek execution of their early orders for the day in the opening single price auction. There are some important exceptions, however. First, HFT liquidity providers do not aim to execute their orders at

\textsuperscript{10}The three stocks are Murata, Nintendo, and Nihon Densan.
the opening price. Instead, they try to build a set of limit orders to prepare for the market-making activity in the continuous session that follows.\textsuperscript{11} Second, some aggressive investors may attempt to manipulate the price and send a false signal to the other market participants. They will eventually cancel these orders prior to the initiation of the auction. Therefore, we distinguish between the following three types of market participant.

\textit{Institutional investors} are those with large orders who may cause a significant market impact. They will not enter their orders until the very last moment, because there would be a clear disadvantage to doing it so earlier: larger orders attract other participants and induce other investors to react sooner and cause a deterioration in the execution price of large orders for the investor. Additionally, in most markets, there is no time priority applied to orders submitted during the pre-opening period. (However, the original time priority in the pre-opening session is activated in the following continuous session, in the event the order is not executed.) Thus, there is a limited benefit from early submission of orders by such large institutions, and there is a potential cost. The only significant disadvantage of waiting and submitting the order at the very last second is that institutional investors, such as pension funds and mutual funds, generally do not use a co-location service to place their orders (as documented by Gomber, Arndt, Lutat, and Uhle (2011)), and thus, cannot take too much risk by delaying their order submission beyond the opening auction.

\textit{Aggressive investors} are those who may enter an order with an aggressive limit price to send a signal to investors on the opposite side, and induce them to provide liquidity. This type of order may include false orders with aggressive limit prices aimed at eliciting a favorable response from true orders from the opposite side. While this does not always work to the advantage of the aggressive investor, it may serve to add noise to the pre-opening quotes and become a source of cancellations.

\textit{Adaptive investors}, essentially HFTs, are those who utilize algorithmic trading to submit many orders/revisions in response to the observed interaction between aggressive and other

\textsuperscript{11}Brogaard (2010), Menkveld (2013) and Brogaard, Hendershott, and Riordan (2014) provide detailed analyses of the trading strategies employed by high frequency traders in the continuous trading session.
adaptive investors during the pre-opening period. The current cutting-edge technology employed by the TSE permits investors to monitor their order submission activities and adjust their trading strategies to the varying state of the order book, during the pre-opening period, since the trading platform of the TSE permits a latency of order submission of less than two milliseconds. Also, high frequency liquidity providers may build their positions during the pre-opening period, in anticipation of their view of the opening price. The liquidity providers maintain a set of limit orders in the book so that their revision or cancellation could affect multiple orders. This type of HFT submits its limit orders as early as the inception of the pre-opening period, because the time priority of orders in the post-opening period is important for them. Although the time priority of orders does not have any effect on execution in the opening auction, the original time priority in the pre-opening session is activated in the following continuous session. It is crucial to ensure, for market-making purposes, therefore, that a given trader’s orders have first (higher) priority at each limit price, as this will maximize the turnover of his inventory and increase his trading profit.\footnote{This nature of HFT market makers is also pointed out by Easley, Lopez de Prado, and O’Hara (2011). Raman and Yadav (2014) also discuss traders’ inventory and order revision, in the context of HFTs.} These characteristics are unique to HFTs who are adaptive investors and distinguish them from those of other types of HFTs. Indeed, the manner in which HFTs modify or cancel their limit orders moment-by-moment is the key behavior on which we focus in this study.\footnote{Hasbrouck and Saar (2009), Menkveld (2013) and Kirilenko, Kyle, Samadi, and Tuzun (2014).}

All three types of participant benefit from the speed of trading. A lower latency of order turnaround permits all of them to delay their final actions until the very last moment of the pre-opening period. Therefore, the noise effects of trading may prevail up to the final seconds of the pre-opening period. Among the three types of market participant, large institutions are the least likely to be the very last player in the game, because, as mentioned earlier, most of them do not have a co-location service, and high-frequency liquidity providers may act up to the last milliseconds to optimize their limit order positioning. Thus, we investigate the timing of order submission as well as the type of order submission, such as the size of the order and single or multiple orders, so as to distinguish between orders from various players.
Low latency leads to a delay of the order submissions, revisions, and cancellations decisions towards the end of the pre-opening period for all three types of market participants discussed above. In this paper, we focus on the role of HFTs in price discovery and liquidity provision. Therefore, we formulate two sets of hypothesis.

**Hypothesis 1: HFT and Price Discovery**

HFT activity a) leads to a delay in the price discovery process, b) decreases the noise in the opening price, and c) has a smaller impact on the best bid and offer quotes than non-HFT activity.

**Hypothesis 2: HFT and Liquidity Provision**

HFT activity in the pre-opening period leads to the building of positions in anticipation of the continuous trading session, which, in turn, contributes to the provision of liquidity.

3.1. Empirical Analysis: Data Description and Pre-Opening Quote Behavior

3.1.1. Server IDs

As mentioned earlier, the novel data provided by TSE are the unique IDs of the virtual servers. A virtual server is a logical device that needs to be set up between the computer systems of the market participant and the exchange, in order to send/receive data to/from each other. There is a limitation in terms of the number of message submission per second for each server, so that heavy users such as HFTs use multiple servers to execute their orders. However, the ID assigned to a particular virtual server is fixed during the period of our analysis. In our analysis, we use data from the pre-opening as well as continuous session to compute important measures such as the trade-to-quote ratio (number of trades to quotes) and the cancellation rate (number of cancellations divided by the total number of messages), both of which are widely recognized as proxies for HFT trading patterns, as argued by Hendershott, Jones, and Menkveld (2011); Brogaard (2010); Menkveld (2013); Brogaard, Hendershott, and Riordan (2014). We choose a threshold of (less than) 25% for the trade-to-quote ratio, and (higher than) 20% for the cancellation rate, as in Ferber (2012) and Hosaka (2014), to construct our sample of proxy servers.\(^\text{14}\)

\(^{14}\)Hosaka (2014) uses the same information to examine characteristics of order flow from HFTs. He finds
We identify 3,663 servers that were used for trading our universe of stocks, of which 875 servers pass our threshold criteria. In our sample, the median trade-to-quote ratios of HFT and non-HFT servers are 15.3% and 28.9%, respectively, and the median cancellation rates are 28.3% vs. 5.8%. The median numbers of messages (new orders, revisions and cancellations) for HFT and non-HFT servers are 530 and 133 per day per stock, respectively.

3.1.2. Pre-opening order flow

Figure 1, Panel A, shows new orders entered every second as a percentage of the total new orders during the pre-opening period. Grey bars are orders from HFT and black bars are those from non-HFT servers. In the first five minutes of the pre-market opening period, which starts at 8 am and end at 9 am, 70% of the orders submitted during the entire period are entered. The order submission slows down after the first five minutes, and is again reactivated ten minutes before the official opening time. The high level of order submissions in the first five minutes indicates the accumulation of orders overnight as well as the advantages of being among the first priority orders at each limit price for liquidity providers. Figure 1, Panel B, shows the order submission activities by HFTs during the pre-opening period. It clearly shows a peak in the very beginning of the period and picks up again very close to the opening time, vastly exceeding the number of orders submitted by non-HFTs.

Figures 2, Panels A and B, show new order submissions and cancellations as a percentage of total orders by non-HFTs and HFTs in the last ten minutes of the pre-market opening period. In Figure 2A, new orders from non-HFTs start to increase to an average of 0.2% from 0.05% per second during the last ten minutes before 9 am. Those from HFTs (Figure 2B) rise just before the opening time. A rise in order cancellations (black line) happens suddenly, one second before 9 am for both HFTs and non-HFTs. The percentage of cancellation messages increases from less than 0.1% to 0.48% and 0.45% respectively. It is interesting to note that cancellations from both HFTs and non-HFTs reach their peak at the very last second. That orders from HFTs participate at “best quotes” with a higher probability than those from non-HFTs.
This indicates that our classification of non-HFTs might include some traders who also use the co-location service, albeit in a less intensive manner. We investigate this in detail at the millisecond scale and present the results in Figure 3. We confirm that the cancellations indeed occur less than one second before 9 am. The cancellation phenomenon starts at 500 milliseconds before 9 am and peaks out at 130 milliseconds before 9 am. This action would not be possible in the absence of a low latency trading environment, and hence, leads to our conclusion that some of the non-HFTs in our classification could also be using the co-location service.

3.1.3. Deviation of mid-quotes from the opening price

We find in the previous section that the number of order submissions rises right before the opening time. We look into the movements of pre-opening quotes between 8 am and 9 am to see how quickly a pre-opening quote approaches the opening price for the day. For this purpose, we compute the relative deviation of mid-quotes from the opening price for each stock on each day using equation (1):

\[
Deviation = \left( \frac{M_{d,s}}{O_d} - 1 \right) \times 100
\]  

(1)

where \( M_{d,s} \) is mid-quote at time \( s \) on day \( d \), \( O_d \) is opening price on day \( d \). First, we compute equation (1) second-by-second per stock per day. Then we calculate the second-by-second medians.

Figure 4 shows the second-by-second movements of the pre-opening quotes. During the first five minutes, the deviation of the pre-opening quote declines rapidly from above 2% to between 0.6% and 0.7%. This means that significant amounts of order submissions during this period contribute to price discovery. However, after 8:05 am, the deviation becomes almost flat with some spikes and then it resumes its adjustment toward the opening price after 8:50 am. It gets down to 0.22% before the opening time, which is still a little bit wider than a half-spread, on average, for the sample stocks during the trading session. According
to Figure 4, the observed pattern of the price deviation is consistent with our hypotheses regarding order submission strategies employed by the three main players. It shows that lower latency does not attenuate the reduction of the deviation between the pre-opening quotes and the opening price. Hence, the orders submitted after 8:50 am play an important role in price discovery.

INSERT FIGURE 4 HERE.

3.1.4. Aggressiveness of limit orders

The aggressiveness of limit orders can be characterized by a comparison between the limit price and the prevailing mid-quotes (Biais, Hillion, and Spatt (1995)). Since TSE’s mid-quotes are an expected opening price, a negative (positive) deviation of sell (buy) limit price indicates a high aggressiveness of the order submission strategy. Table 3 shows the aggressiveness of limit sell (buy) price of orders submitted or revised/cancelled from HFTs and non-HFTs during the last 60 seconds of the pre-opening period. The limit prices of new sell(buy) orders submitted by HFTs show an average 5.53% (6.78%) absolute deviation from the mid-quote, but those by non-HFTs shows a much higher absolute deviation of 32.07% (35.06%). Cancelled sell(buy) orders by HFTs exhibit a 3.51% (4.53%) absolute deviation from the mid-quote, while cancelled sell(buy) orders by non-HFTs show 26.50% (29.85%). These results are consistent with the notion that HFTs liquidity providers adjust their orders to the expected opening price. Orders from non-HFTs do not have similar characteristics.

INSERT TABLE 3 HERE.

Figures 5, Panels A and B, show the relative limit prices of the cancelled orders in the last second of our sample period. We separate those observations in the period between 8:59:59 and 9:00:00 into two equal segments: those occurring more than or less than 500 milliseconds before 9 am. More than 80% of cancelled buy and sell limit orders have limit prices within a plus or minus 10% deviation from the mid-quote. It should be noted that limit sell (buy) orders that have limit prices lower (higher) than the mid-quote exhibit more cancellations less than 500 milliseconds before 9 am. Also, limit orders that must be included
in the opening transaction are cancelled more often. Figures 5, Panels C and D show the sub-sample from Figures 5 in which the relative limit price is within plus/minus 5% of the mid-quote. Less than 500 milliseconds before 9 am, sell limit orders with limit prices 1% lower than the mid-quote comprise 40% of the sub-sample of cancellations. During this period, buy limit orders with limit prices 1% higher than the mid-quote comprise 60% of the sub-sample of cancellations. This supports our conjecture that the surge in cancellations comes from high-frequency liquidity traders who wish to avoid executing their orders at the opening price.

INSERT FIGURE 5 AND 6 HERE.

Figure 6 shows how often submitted orders have an impact on the prevailing quotes. Out of 91,139 orders entered at 8:59 am during April and May 2013, about 16% have an impact on the mid-quotes. However, reverses in the sign of the price change from the previous day’s close make up only 0.6% of all events. This indicates that most orders that have an impact on quotes generate only minor changes. Figures 7, Panels A and B, depict the numbers of orders that have an impact on the ask or bid quotes double in the last two seconds of the pre-opening period. Figures 8, Panels A and B, shows that the average impact on the ask and bid quotes do not change in the last two seconds.

In sum, although we observe sudden increases in cancellations less than 500 milliseconds before 9 am, they do not have a significant price impact on the prevailing quotes. This suggests that the surge in cancellations occurs for position building and position risk management purposes. We confirm this interpretation in the following sub-section.

INSERT FIGURE 7 AND 8 HERE.

3.1.4. Tests of unbiasedness of the pre-opening quotes

We repeat the test of price efficiency with the pre-opening quotes using an unbiasedness regression that has been used widely in the literature, as a test of average predictability. Specifically, this test is used by Biais, Hillion, and Spatt (1999) to characterize the extent to which there is learning and price discovery in the pre-opening period. They use the closing
price of the day as a proxy for the equilibrium price $v$. We modify their framework for our purpose and estimate equation (2) as follows:

$$
\nu - E(\nu|I_0) = \alpha_t + \beta_t [P_t - E(\nu|I_0)] + Z_t
$$

where $\nu$ is the opening price (instead of the closing price used in Biais, Hillion, and Spatt (1999)), $P_t$ is the pre-opening mid-quote, and $E(\nu|I_0)$ is the previous day’s closing price. The distribution of the change in price, from the previous day’s close to the mid-quote, varies with time as the opening time approaches. The variance of the noise in the mid-quote is also likely to vary with time. In this spirit, we estimate the unbiasedness regression using the specification in (2), for each second as well as for each 10 milliseconds for each stock over our sample period. Thus, we analyze for each point in the time, the distribution across days of the mid-quote. If the pre-opening mid-quote is an unbiased estimator of the opening price, the coefficient $\beta_t$ in the specification should be insignificantly different from 1. We hypothesize that the earlier in the pre-opening period the coefficient $\beta_t$ equals 1, the greater is the price efficiency of the pre-opening quote. We also analyze the pattern of the root mean square error (RMSE) over the pre-opening period. This analysis allows us to quantify the information content of the pre-opening prices: the lower the RMSE, the greater the information content.

Figure 9, Panel A, shows the average of the coefficient, $\beta_t$, and the bands of plus or minus two sigma of cross-sectional standard errors, over time. The mean coefficient is significantly different from one until two seconds before 9 am, and becomes insignificantly different from one only one second before 9 am in the sample from April and May 2013. In order to investigate, price discovery at the millisecond level, we run the same regression for three different periods for each 10 milliseconds. In particular, we analyze data from November-December 2009, January-March 2010, and April-May 2013. Inclusion of the two additional periods allows us to test changes in the price discovery process due to the introduction of low latency trading platform "Arrowhead" and implementation of several other institutional changes such as co-location service (see Uno and Shibata (2012)). Thus, January 2010 can be viewed as the time of arrival of a new trading paradigm in Japan. The implementation of
the new trading platform with a change in the latency, and the new design of the pre-opening
auction caused a shift in the behavior of all traders. This structural change created room for
the HFTs to exploit the breakthrough in the latency. Thus, this natural experiment is ideal
for assessing the effect of the latency regime on price informativeness: Reducing the latency
potentially increases the speed of quote flow, which in turn may lead to an improvement on
the accuracy of the price, better liquidity and greater speed of price discovery.

To test these hypotheses, we investigate whether the observation that the coefficient
becomes one, one second before 9 am was equally valid at that point in time in the earlier
regime to check if there has been a structural change after the introduction of the “Arrowhead”
system. Figure 9, Panels B, shows that the coefficient becomes one 220 milliseconds before
9 am in the period of April and May 2013, 680 milliseconds before 9 am in the period of
January-March 2010, but never reaches one in the period of November and December 2009.
The comparison between 2013 and 2010 suggests that introduction of the "Arrowhead" and
its increasing usage by HFTs delayed price discovery by 460 milliseconds. From 2010 to 2013,
the proportion of orders coming through co-location servers more than tripled, from 10-15%
to above 50%. Although, the moment the beta coefficient in 2013 becomes one is delayed,
the beta coefficient reaches 0.9 level in 2013 much earlier than that in 2010. The convergence
path of 2010 shows a stepwise trend, which indicates a symptom of caution in the quote
submission from the HFTs. The fact that beta does not reach one in 2009 is indicative of a
slow price discovery and a weak accuracy of opening price. 32 stocks out of 97 in our sample
are subject to a tick size change which was effective at January 2010, larger tick size may also
contribute to the difference between opening price and mid-quotes. The analysis for RMSE
indicates that noise in the pre-opening quotes is bigger in 2013 than in 2010. It needs further
investigation. Overall results indicate that price efficiency has improved in the low latency
regime. The new latency regime generate a new environment for all players, but the learning
process to efficiently exploit the improved speed requires time and for a careful calibration
of the algorithm, while human intervention cannot benefit from the improvement in speed,
since it is too rapid, in any case. In the first month since the inception of the new trading
system on January 2010, the orders from co-location servers were about 15%, whereas they exceeded 50 as of May 2013 (as documented by Hosaka (2014)).

HFTs were already present in the TSE before 2010, but with limited ability to have an impact on prices (see Uno and Shibata (2012)). The natural experiment that we analyze shows that the introduction of Arrowhead system was an exogenous event that triggered several consequences: the accuracy of price, the need for adaption by HFTs, a reduction of price dispersion and an improvement of liquidity. However, we caution that given the design of the experiment and the absence of a control group, we cannot say anything about causality. These findings are consistent with the hypothesis that high-frequency quote updates contribute to price discovery.

INSERT FIGURE 9 (PANEL A AND B) HERE.

3.2. Revisions and Cancellations

3.2.1. Determinants of revisions/cancellations

One of the characteristics of trading in the pre-opening period documented in Section 3.1 is the flurry of new orders, cancellations, and revisions as time approaches the 9 am opening time. The TSE order file data allow us to investigate the determinants of revisions and cancellations from a history of order status during the pre-opening period. We estimate a probit model in order to investigate the motivations behind the revisions and cancellations.

One possible motivation behind revisions and cancellations is for the investor to adjust the limit price as a reaction to changes in the expected opening price. The intent of the revision in this case is to increase the probability of execution in the opening auction. This is a valid strategy for investors who wish to execute their orders in the opening auction. This motivation would suggest that these investors would increase the number of order revisions closer to 9 am.

In contrast to the above behavior, HFT liquidity providers do not aim to execute their orders at the opening price. They adjust the limit prices of their orders to be surrounding the expected opening price. This means that they want their stance regarding liquidity provision to be neutral with respect to the opening price. They then submit their orders as soon as the
TSE starts receiving orders. However, it is crucial for them to have a higher time priority for their orders at each limit price in the book in order to enable them to have a quicker turnover of their position. Therefore, HFT orders are made up of a set of both buy and sell limit orders. Adjustments are triggered by changes in the prevailing quotes, which happen on a continuous basis during the pre-opening period. Since revisions of orders are free of charge, HFTs can keep their high time priority until the very last second.

There are at least two reasons for cancellations by HFTs. One arises when limit prices are deep in the money, i.e., the buy orders are well above the current mid-quote, and sell orders are well below the current mid-quote. Cancelling deep-in-the-money orders may have an impact on the prevailing best quotes. This type of order may be entered earlier to influence other order submitters. Another possibility is the initiation of a set of cancellations and placement of new orders by HFTs. This is a faster procedure for changing their limit prices, since revising orders submitted earlier takes more time than entering a set of cancellations and new orders. However, we cannot separate these two motivations due to data limitations regarding customer IDs.

We now test the following hypothesis by estimating a probit model of cancellations and revisions. For limit sell (buy) orders from HFT liquidity providers with a limit sell (buy) price which is lower (higher) than an expected opening price, the lower (higher) the relative limit price is, the more likely it will be canceled. To test this hypothesis, we estimate a probit model for cancellations and revisions. Based upon the order flow analysis in the previous sections, the estimation period is the 2 second window between 8:59:58 and 8:59:59. We estimate the specification:

$$\rho_j = \alpha + \beta_1 \frac{|\text{Limit price}_j - \text{Midquote}_j|}{\text{Midquote}_j} \times 1\{\frac{\text{Limit price}_j - \text{Midquote}_j}{\text{Midquote}_j} > 0\} +$$
$$+ \beta_2 \frac{|\text{Limit price}_j - \text{Midquote}_j|}{\text{Midquote}_j} \times 1\{\frac{\text{Limit price}_j - \text{Midquote}_j}{\text{Midquote}_j} < 0\} +$$
$$+ \beta_3 \frac{1}{\text{RevFreq}_j} + \beta_4 \ln(\text{Elapsed Time}_j + 1) + \beta_5 \text{Size}_j + \beta_6 \text{Depth}_j + \epsilon_j$$

---

The TSE does not provide customer IDs for individual orders.
The dependent variable of equation (3), $\rho_j$, takes the value one when an order is cancelled, and zero otherwise. There are six explanatory variables: the coefficients $\beta_1^- (\beta_1^+)$ capture the sensitivity of the cancellation action to the aggressiveness of the limit price, measured by the relative deviation of the limit price from the mid-quote. We analyze two separate cases when the limit price is above and below the mid-quote, respectively. The third variable, with the coefficient $\beta_2$, is the inverse of the revision frequency, the fourth one, with the coefficient $\beta_3$, is the logarithm of the elapsed time from the original submission time, and the fifth one, with the coefficient $\beta_4$, is the size of the order and the last one, with the coefficient $\beta_5$, is the depth, which is the average ask (bid) size just before the sell (buy) order submission. Market orders are excluded from the sample.

The orders submitted through HFT and non-HFT servers are separated when we estimate the probit model of equation (3). We expect that the coefficient of $\beta_1^- (\beta_1^+)$ for orders from HFT liquidity providers should be larger (in absolute terms) than for other orders because these traders are more sensitive to the expected opening price. In the case of sell orders, we expect $\beta_1^- > \beta_1^+$ because HFT liquidity providers do not aim to obtain execution at the opening price. In the case of buy orders, we expect $\beta_1^- < \beta_1^+$ for the same reason.

Table 4, Panel A (sell orders), shows the results of the estimation for cancelled sell orders. About one third of sell orders submitted by HFTs in this period are cancelled. The coefficient $\beta_1^-$ of limit sell orders from HFT liquidity providers is positive, and statistically significant at the 1% significance level. This means that the lower is the sell limit price (which means that it is most likely to be executed at the opening), the more likely the order is to be cancelled. This is consistent with the behavior of HFT liquidity providers. The coefficient in the sample of orders from HFT liquidity providers is almost 67% larger than those from non-HFT traders. HFT liquidity providers are more likely to cancel their orders when the limit sell price is below the prevailing mid-quotes. Other controlling variables such as the coefficient of revision frequency ($\frac{1}{RevFreq}$) and elapsed time are also positive and significant at the 1% level of significance. The cancelled orders are less frequently revised and tend to sit for a longer time in the order book. This indicates that the orders are submitted early.
The sizes of the orders are smaller. These characteristics are common to the order strategy of both the HFT and non-HFT players.

Table 4, Panel A (buy orders), shows the results for cancelled buy orders. Nearly one third of buy orders submitted by HFT liquidity providers in this period are cancelled. The coefficient $\beta_1^+$ of limit buy orders from HFT liquidity providers is statistically significant at the 1% level and positive. This means that the higher is the buy limit price (which means that it is most likely to be executed at the opening), the more likely the order is to be cancelled. This is also consistent with the behavior of HFT liquidity providers. The coefficient $\beta_1^+$ in the sample of orders from HFT liquidity providers is about 50% larger than those from of non-HFT traders. HFT liquidity providers are more likely to cancel orders when the limit buy price is above the prevailing mid-quotes. The coefficient of revision frequency ($\frac{1}{RevFreq}$) is positive and significant at the 1% level, and that of elapsed time is also positive and significant at the 1% level of significance. These results are the same as those for the cancelled sell orders. Like cancelled sell orders, cancelled buy orders have the property of being less frequently updated, with a longer elapsed time and a smaller size of orders for both HFT and non-HFT orders.

Table 4, Panel B (sell orders), shows the estimation results of equation (3) for revised sell orders. Only 20% of sell orders submitted by HFT liquidity providers in this period consist of revisions. In the case of HFT orders, the coefficients $\beta_1^-$ is insignificant, and $\beta_1^+$ is positive and significant at the 1% level. In case of non-HFT orders, however, both coefficients $\beta_1^-$ and $\beta_1^+$ are positive and significant at the 1% level. Non-HFT traders are more likely to revise the limit price of their orders when the limit sell price is above the mid-quote, but HFTs is less likely to revise them. Similar results hold for revised buy orders in Table 4, Panel B (for buy orders). The buy orders revised by non-HFT players have limit prices that are lower than the prevailing mid-quote. But this is not the case for HFTs: they leave their buy limit orders that have limit price lower than the current mid-quote, without making any changes. This suggests that HFTs do not aim to execute their orders at the opening price.
3.2.2. The relation between mid-quote change and HFT order flow

Lastly, we examine the relationship between changes in the mid-quote and submitted order types by HFTs and non-HFT players. Table 5 shows that, in most cases, we do not find significant price changes except new sell and new buy orders from non-HFT traders, and constant revision of buy orders from HFTs. The average price change caused by new sell orders from non-HFT traders is -1.79%, a large, statistically significant negative price impact, but that by HFT is only 0.10%, which is also not significantly different from zero. The average price change caused by new buy orders from non-HFT traders is 1.87%, a statistically significant positive price impact, but that by HFTs is -0.48% which is, again, not significantly different from zero. The revisions for buy orders made by HFTs causes a 2.0% price change in the case of quantity revision, and a 2.6% price change, in the case of a limit price revision. These cases require more careful examination in future research. Overall, however, our results indicate that the last second cancelation and revision do not cause a significant impact on the prevailing quotes.

INSERT TABLE 5 HERE.

4. Conclusion

The institution of a market pre-opening period is an important feature of many stock markets today. A key question we ask in this research is whether high-frequency quote revisions that occur during the pre-opening period amplify noise or lead to an improvement in the price formation. A flurry of order flows come in just a blink before the market opening at 9 am. New orders come in the last second, but order cancellations start to increase less than 500 milliseconds before 9 am and continue up until 130 milliseconds before 9 am.

The number of orders that have an impact on the prevailing quotes increases in the last two milliseconds prior to the opening time. However, this does not increase the volatility of the pre-opening quotes. This means that the size of the impact on the mid-quotes is relatively small. Return reversal from the previous day’s close happens with a probability of 0.6%. We interpret this as a harmless adjustment made by high-frequency liquidity providers.
This interpretation is confirmed by the fact that the coefficients estimated using the unbiasedness regression reach the point at which they are statistically indifferent from one at the very last second before 9 am. After the advent of the low-latency trading facility, the improvement in the pre-opening price efficiency (unbiasedness) is delayed until half a second before the opening.

A probit analysis of the order submission shows that limit price aggressiveness and the history of limit price revisions are related to the likelihood of cancellation at the last minute. Smaller sizes, earlier submission times, and more aggressive limit prices are all related to a higher probability of cancellation. Overall, the results do not support the notion that the order submission strategies employed by HFT cause a deterioration in price formation in the market.

In order to investigate the reasons for the flurry of cancellations that occurs at the very last moment, we perform a similar analysis for the pre-opening period of the afternoon session in the TSE which is not included in this version of the paper. We find that revisions of orders occur more often than cancellations in the pre-opening of the afternoon session. The closing price from the morning session provides a less noisy estimation of the opening price of the afternoon session. This suggests that the flurry of cancellations occurs due to uncertainty about the opening price at the beginning of the trading day.

Our findings in the paper can be confirmed in cases where the same stock is traded in different venues within the same time-zone. In Japan, off-exchange venues such as proprietary trading systems (PTS) do not attract order flows in the time prior to the official trading time of the TSE. The US and Europe, where multiple venues compete against each other, may provide researchers with additional opportunities to seek further insights on the factors driving our results. We plan to work on this in our next project, involving data from Europe (Eurofidai) and the U.S. (NASDAQ).
References


**Table 1: Distribution of order flow**

This table shows the distribution of the order flow for 97 stocks from TOPIX100 during the sample period from April to May 2013. We report the average number of orders, relative frequency of orders, and the average size of the orders in shares submitted during the whole pre-opening period (8:00:00.000 - 8:59:59.999), during the last ten minutes of the pre-opening period (8:50:00.000 - 8:59:59.999), and during the last one minute of the pre-opening period (8:59:00.000 - 8:59:59.999). All orders are grouped according to their types: new orders, quantity revisions (reduction in the order size), limit price revisions, and cancellations (withdrawals of orders). The data on the order flow are provided by Tokyo Stock Exchange.

<table>
<thead>
<tr>
<th></th>
<th>New orders</th>
<th>Quantity Revisions</th>
<th>Price Revisions</th>
<th>Cancellations</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average # of orders</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:00:00.000 - 8:59:59.999</td>
<td>117.20</td>
<td>1.70</td>
<td>38.30</td>
<td>39.00</td>
<td>196.20</td>
</tr>
<tr>
<td>Relative frequency of orders</td>
<td>59.70%</td>
<td>0.90%</td>
<td>19.50%</td>
<td>19.90%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Average size of orders in shares</td>
<td>4,244.70</td>
<td>1,913.20</td>
<td>3,955.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:50:00.000 - 8:59:59.999</td>
<td>97.60</td>
<td>2.90</td>
<td>67.20</td>
<td>71.20</td>
<td>238.90</td>
</tr>
<tr>
<td>Relative frequency of orders</td>
<td>40.90%</td>
<td>1.70%</td>
<td>28.10%</td>
<td>29.80%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Average size of orders in shares</td>
<td>5,744.80</td>
<td>2,548.90</td>
<td>4,328.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:59:00.000 - 8:59:59.999</td>
<td>173.00</td>
<td>5.90</td>
<td>110.20</td>
<td>110.60</td>
<td>399.70</td>
</tr>
<tr>
<td>Relative frequency of orders</td>
<td>43.30%</td>
<td>1.50%</td>
<td>27.60%</td>
<td>27.70%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Average size of orders in shares</td>
<td>4,782.80</td>
<td>1,946.60</td>
<td>3,992.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Server characteristics HFT / non-HFT groups (TOPIX 100, April and May 2013)

This table shows server characteristics for HFT and non-HFT server groups based on the order flow for 97 stocks from TOPIX 100 during the sample period (April and May 2013). We divide servers into two groups: non-HFT and HFT. We use the following conditions to classify HFT server. Condition (1) is that cancellation ratio should be larger than 20%. Condition (2) is that the trade-to-quote ratio should be smaller than 25%. We report total number of servers, median number of daily messages per stock, median trade-to-quote and cancellation ratios. The data on the order flow are provided by Tokyo Stock Exchange.

<table>
<thead>
<tr>
<th></th>
<th>Non-HFT</th>
<th>HFT</th>
</tr>
</thead>
<tbody>
<tr>
<td># of servers</td>
<td>2,788</td>
<td>875</td>
</tr>
<tr>
<td>Median # of daily messages per stock</td>
<td>133</td>
<td>530</td>
</tr>
<tr>
<td>Median trade-to-quote ratio</td>
<td>28.90%</td>
<td>15.30%</td>
</tr>
<tr>
<td>Median cancellation ratio</td>
<td>5.80%</td>
<td>28.90%</td>
</tr>
</tbody>
</table>
Table 3: Comparison of the relative limit price of the orders submitted by HFT and non-HFT

This table shows the comparison of the limit order strategies employed by HFT and non-HFT for 97 stocks from TOPIX100 during the sample period (April and May 2013) for the last minute of the pre-opening period (8:59:00.000 - 8:59:59.999). We show the mean and standard deviation for the relative limit price of the four types of orders: new orders, quantity revisions (reduction in the order size), limit price revisions, and cancellations (withdrawals of orders), where the relative limit price is determined as follows:

\[
\text{Relative limit price} = \frac{|\text{Limit price} - \text{Midquote}|}{\text{Midquote}}.
\]

We analyze the relative limit order prices separately for the sell limit orders (Panel A) and buy limit orders (Panel B). The data on the order flow and server IDs used to classify traders into HFT and non-HFT are provided by Tokyo Stock Exchange.

### Panel A. Absolute difference between limit price and the mid-quote for sell orders

<table>
<thead>
<tr>
<th></th>
<th>New orders</th>
<th>Quantity Revisions</th>
<th>Price Revisions</th>
<th>Cancellations</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.53%</td>
<td>3.17%</td>
<td>2.18%</td>
<td>3.51%</td>
<td>4.12%</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>18.18%</td>
<td>10.96%</td>
<td>5.87%</td>
<td>11.03%</td>
<td>14.24%</td>
</tr>
<tr>
<td># of observations</td>
<td>60,250</td>
<td>28,936</td>
<td>23,008</td>
<td>20,751</td>
<td>132,945</td>
</tr>
<tr>
<td>Non-HFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>32.07%</td>
<td>7.31%</td>
<td>7.88%</td>
<td>26.50%</td>
<td>22.28%</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>44.47%</td>
<td>19.08%</td>
<td>19.91%</td>
<td>39.02%</td>
<td>38.26%</td>
</tr>
<tr>
<td># of observations</td>
<td>106,551</td>
<td>39,700</td>
<td>36,821</td>
<td>19,340</td>
<td>202,412</td>
</tr>
</tbody>
</table>

### Panel B. Absolute difference between limit price and the mid-quote for buy orders

<table>
<thead>
<tr>
<th></th>
<th>New orders</th>
<th>Quantity Revisions</th>
<th>Price Revisions</th>
<th>Cancellations</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>6.78%</td>
<td>3.73%</td>
<td>2.44%</td>
<td>4.53%</td>
<td>5.00%</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>21.39%</td>
<td>13.17%</td>
<td>7.40%</td>
<td>15.64%</td>
<td>17.15%</td>
</tr>
<tr>
<td># of observations</td>
<td>71,790</td>
<td>36,087</td>
<td>27,801</td>
<td>24,037</td>
<td>150,715</td>
</tr>
<tr>
<td>Non-HFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>35.06%</td>
<td>6.89%</td>
<td>6.73%</td>
<td>29.85%</td>
<td>24.10%</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>45.94%</td>
<td>19.82%</td>
<td>19.43%</td>
<td>40.99%</td>
<td>40.07%</td>
</tr>
<tr>
<td># of observations</td>
<td>109,983</td>
<td>40,284</td>
<td>37,098</td>
<td>22,787</td>
<td>210,152</td>
</tr>
</tbody>
</table>
Table 4: Probit model for canceled and revised sell and buy orders

This table shows the estimation results (coefficients and z-statistics) of the probit model estimated separately for canceled (Panel A) and revised (Panel B) buy and sell orders from HFT and non-HFT for 97 stocks from TOPIX100 during the sample period (April and May 2013) for the last one minute of the pre-opening period (8:59:00.000 - 8:59:59.999). The probit model specification is given by equation (4):

\[
\rho_j = \alpha + \beta_1 \left| \frac{\text{Limit price}_j - \text{Midquote}_j}{\text{Midquote}_j} \right| \times 1\{ \text{Limit price}_j - \text{Midquote}_j > 0 \} + \\
+ \beta_1 \left| \frac{\text{Limit price}_j - \text{Midquote}_j}{\text{Midquote}_j} \right| \times 1\{ \text{Limit price}_j - \text{Midquote}_j < 0 \} + \\
+ \frac{1}{\text{RevFreq}_j} + \beta_2 \ln(\text{ElapsedTime}_j + 1) + \beta_3 \text{Size}_j + \beta_4 \text{Depth}_j + \epsilon_j \quad (3)
\]

The dependent variable \( \rho_j \) takes the value 1 when an order \( j \) is canceled (Panel A) or revised (Panel B), and zero otherwise. There are six independent variables: the aggressiveness of the limit price which is measured as the relative deviation of the limit order price from the mid-quote at the time of order submission (we analyze separately two cases when the limit price is above and below the mid-quote), the inverse of revision frequency (\( \frac{1}{\text{RevFreq}_j} \)), the elapsed time from the original submission time (\( \text{ElapsedTime}_j \)), the size of the order in Japanese yen (\( \text{Size}_j \)), and the depth (\( \text{Depth}_j \)), which is the best ask (bid) size just before the sell (buy) order submission (averaged over a second in case of multiple orders submitted during that second). Market orders are excluded from the sample. The data on the order flow and server IDs used to classify traders into HFT and non-HFT are provided by Tokyo Stock Exchange.

<table>
<thead>
<tr>
<th>Panel A. Probit model: Canceled sell and buy orders</th>
<th>Canceled sell orders</th>
<th>Canceled buy orders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HFT</td>
<td>Non-HFT</td>
</tr>
<tr>
<td></td>
<td>Coef</td>
<td>z-stat</td>
</tr>
<tr>
<td>Limit price(_j) - Midquote(_j) (&gt;0) (\frac{\text{Limit price}_j - \text{Midquote}_j}{\text{Midquote}_j})</td>
<td>-0.015</td>
<td>-26.36</td>
</tr>
<tr>
<td>Limit price(_j) - Midquote(_j) (&lt;0) (\frac{\text{Limit price}_j - \text{Midquote}_j}{\text{Midquote}_j})</td>
<td>0.303</td>
<td>7.008</td>
</tr>
<tr>
<td>(1/\text{RevFreq}_j)</td>
<td>4.231</td>
<td>63.084</td>
</tr>
<tr>
<td>(\text{ElapsedTime}_j)</td>
<td>0.185</td>
<td>39.149</td>
</tr>
<tr>
<td>(\text{Size}_j)</td>
<td>-0.167</td>
<td>-49.658</td>
</tr>
<tr>
<td>(\text{Depth}_j)</td>
<td>-0.013</td>
<td>-5.372</td>
</tr>
<tr>
<td>(\text{Intercept})</td>
<td>-2.902</td>
<td>-77.589</td>
</tr>
<tr>
<td>McFadden (R^2)</td>
<td>0.406</td>
<td>0.612</td>
</tr>
<tr>
<td># Obs with (\rho_j = 0)</td>
<td>15,033</td>
<td>9,213</td>
</tr>
<tr>
<td># Obs with (\rho_j = 1)</td>
<td>5,861</td>
<td>4,437</td>
</tr>
</tbody>
</table>
Table 3: Probit model for cancelled and revised sell and buy orders (continued)

<table>
<thead>
<tr>
<th></th>
<th>HFT</th>
<th>Non-HFT</th>
<th>HFT</th>
<th>Non-HFT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef</td>
<td>z-stat</td>
<td>Coef</td>
<td>z-stat</td>
</tr>
<tr>
<td></td>
<td>Revised sell</td>
<td></td>
<td>Revised buy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>orders</td>
<td></td>
<td>orders</td>
<td></td>
</tr>
<tr>
<td>Limit price - Midquote (&gt;0)</td>
<td>-30.382</td>
<td>-12.641</td>
<td>-2.171</td>
<td>-0.899</td>
</tr>
<tr>
<td>Limit price - Midquote (&lt;0)</td>
<td>-5.055</td>
<td>-1.468</td>
<td>23.255</td>
<td>9.673</td>
</tr>
<tr>
<td>1/RevFreq</td>
<td>-6.186</td>
<td>-52.188</td>
<td>-5.324</td>
<td>-55.174</td>
</tr>
<tr>
<td>ElapsedTime</td>
<td>-0.034</td>
<td>-5.000</td>
<td>-0.040</td>
<td>-7.033</td>
</tr>
<tr>
<td>Size</td>
<td>0.070</td>
<td>42.638</td>
<td>0.067</td>
<td>39.971</td>
</tr>
<tr>
<td>Depth</td>
<td>-0.031</td>
<td>-11.132</td>
<td>-0.046</td>
<td>-14.476</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.935</td>
<td>24.664</td>
<td>0.017</td>
<td>24.373</td>
</tr>
<tr>
<td>McFadden $R^2$</td>
<td>0.347</td>
<td>0.376</td>
<td>0.320</td>
<td>0.306</td>
</tr>
<tr>
<td># Obs with $\rho_j = 0$</td>
<td>16,269</td>
<td>11,875</td>
<td>18,973</td>
<td>13,211</td>
</tr>
<tr>
<td># Obs with $\rho_j = 1$</td>
<td>4,295</td>
<td>1,775</td>
<td>4,925</td>
<td>2,131</td>
</tr>
</tbody>
</table>
Table 5: Average mid-quote changes from the orders submitted by HFT and non-HFT

This table shows the comparison of the limit order strategies employed by HFT and non-HFT for 97 stocks from TOPIX100 during the sample period (April and May 2013) for the last one minute of the pre-opening period (8:59:00.000 - 8:59:59.999). We show the mean, standard deviation, and t-statistics for the average change in the mid-quote as a result of the order submission for the four type of orders: new orders, quantity revisions (reduction in the order size), limit price revisions, and cancellations (withdrawals of orders). We analyze the average mid-quote changes separately for the sell limit orders (Panel A) and buy limit orders (Panel B). We include only those seconds in the sample when we observe the mid-quote change. The data on the order flow and server IDs used to classify traders into HFT and non-HFT are provided by Tokyo Stock Exchange.

<table>
<thead>
<tr>
<th></th>
<th>New orders</th>
<th>Quantity Revisions</th>
<th>Price Revisions</th>
<th>Cancellations</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. Average mid-quote changes as a result of sell orders</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.105</td>
<td>-0.496</td>
<td>-0.612</td>
<td>-0.803</td>
<td>-0.284</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>93.295</td>
<td>29.134</td>
<td>64.717</td>
<td>144.593</td>
<td>93.034</td>
</tr>
<tr>
<td>t-statistics</td>
<td>0.234</td>
<td>-1.189</td>
<td>-1.187</td>
<td>-0.658</td>
<td>-0.929</td>
</tr>
<tr>
<td># of observations</td>
<td>42,998</td>
<td>20,057</td>
<td>15,752</td>
<td>14,019</td>
<td>92,826</td>
</tr>
<tr>
<td>Non-HFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-1.795</td>
<td>0.900</td>
<td>0.699</td>
<td>-1.260</td>
<td>-0.858</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>229.073</td>
<td>265.797</td>
<td>278.203</td>
<td>90.480</td>
<td>234.950</td>
</tr>
<tr>
<td>t-statistics</td>
<td>-2.041</td>
<td>0.498</td>
<td>0.352</td>
<td>-1.553</td>
<td>-1.273</td>
</tr>
<tr>
<td># of observations</td>
<td>67,876</td>
<td>21,674</td>
<td>19,598</td>
<td>12,436</td>
<td>121,584</td>
</tr>
<tr>
<td><strong>Panel B. Average mid-quote changes as a result of buy orders</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-0.476</td>
<td>2.016</td>
<td>2.539</td>
<td>-0.908</td>
<td>0.526</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>105.879</td>
<td>37.069</td>
<td>41.012</td>
<td>177.106</td>
<td>102.049</td>
</tr>
<tr>
<td>t-statistics</td>
<td>-1.029</td>
<td>8.653</td>
<td>8.631</td>
<td>-0.666</td>
<td>1.737</td>
</tr>
<tr>
<td># of observations</td>
<td>52,321</td>
<td>25,315</td>
<td>19,135</td>
<td>16,882</td>
<td>113,653</td>
</tr>
<tr>
<td>Non-HFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>1.869</td>
<td>0.266</td>
<td>0.067</td>
<td>-2.364</td>
<td>0.796</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>174.286</td>
<td>90.172</td>
<td>90.384</td>
<td>197.941</td>
<td>133.650</td>
</tr>
<tr>
<td>t-statistics</td>
<td>2.818</td>
<td>0.433</td>
<td>0.109</td>
<td>-1.442</td>
<td>1.857</td>
</tr>
<tr>
<td># of observations</td>
<td>69,061</td>
<td>23,566</td>
<td>21,361</td>
<td>14,575</td>
<td>128,563</td>
</tr>
</tbody>
</table>
Figure 1: Flow of order submission in the pre-opening period

Figure 1 depicts second-by-second order flow for 97 stocks from TOPIX 100 during the sample period (April and May 2013). The Tokyo Stock Exchange starts receiving orders at 8 am and starts the call auction at 9 am. Average percentage of total number of orders is the number of total orders at each second divided by the total number of orders submitted in the whole pre-opening period (8:00:00.000 - 8:59:59.999). Y-axis represents the percentage of pre-opening total new orders, and X-axis represents time in seconds between 8 am and 9 am. 80000 means 8:00:00. In Panel A the Grey bar depicts the average percentage of the total number of orders submitted through non HFT servers and black bar through HFT servers during the sample period per second. In Panel B reports only the average percentage of the total number of orders submitted through HFT servers. The data on the order flow are provided by Tokyo Stock Exchange.

(A) Order flow from HFT and non-HFT during the preopening period

(B) Order flow from HFT during the preopening period
Figure 2: Flow of order submission (new orders and cancellation) in the last ten minutes of the pre-opening session

Figure 2, Panel A and B, show the second-by-second average number of new orders and cancellations for non-HFT and HFT respectively for 97 stocks from TOPIX 100 during the sample period (April and May 2013). Average percentage is the number of total orders at each second divided by the total number of orders submitted for the last 10 minutes of the pre-opening period (8:50:00.000 - 8:59:59.999) Y-axis represents the percentage of pre-opening total new orders, and X-axis represents time in seconds between 8:50am and 9 am. 85000 means 8:50:00. The data on the order flow are provided by Tokyo Stock Exchange.

(A) Non HFT order submission (new orders and cancellation)

(B) HFT order submission (new orders and cancellation)
Figure 3: Order flow during the last one second

Figure 3 shows four types of order submission activities: new orders (black line), price revision (dotted line), cancellation (dark grey line) and quantity revision (light grey line) in the last second of the pre-opening period (8:59:59.000 - 8:59:59.999) at the millisecond scale for 97 stocks from TOPIX 100 during the sample period (April and May 2013). The data on the order flow are provided by Tokyo Stock Exchange.
Figure 4: Deviation from the opening price

Figure 4 shows the deviation of the pre-opening quote from the opening price computed at each second in the whole pre-opening period (8:00:00.000 - 8:59:59.999) for 97 stocks from TOPIX 100 during the sample period (April and May 2013). Deviation is defined as follow, as reported in equation (1):

\[ \text{Deviation} = \left( \frac{M_{t,s}}{O_t} - 1 \right) \times 100 \]

Deviation is computed per second per day per stock and then averaged at each second. The data on the order flow are provided by Tokyo Stock Exchange.
Figure 5, Panel A and B, shows the relative limit price of the cancelled sell (buy) orders in the last second of the pre-opening period (8:59:59.000 - 8:59:59.999) with the X-axis in the unit of 10% for 97 stocks from TOPIX 100 during the sample period (April and May 2013). Relative limit price is defined as follow:

\[
\text{Relative limit price} = \frac{| \text{Limit price} - \text{Midquote} |}{\text{Midquote}}
\]

Negative number value of the relative limit price means that the limit sell price is lower than the prevailing mid-quote, thus it is immediately executable. The dotted line shows observations in the first half of 8:59:59 (8:59:59.000 - 8:59:59.499) and the black line shows the last half of 8:59:59 (8:59:59.500 - 8:59:59.999). The data on the order flow are provided by Tokyo Stock Exchange.
**Figure 5: Relative limit price of cancelled sell and buy orders**

Figure 5, Panel C and D, is an enlarged version of Figure 5 (A and B) with finer X-axis with a 1% unit. This shows the distribution of the sample in Figure 5 in which the relative limit price is between plus and minus 5% of the mid-quote in the last second of the pre-opening period (8:59:59.000 - 8:59:59.999) for 97 stocks from TOPIX 100 during the sample period (April and May 2013). Relative limit price is defined as follow:

\[
\text{Relative limit price} = \frac{|\text{Limit price} - \text{Midquote}|}{\text{Midquote}}
\]

Negative number value of the relative limit price means that the limit sell price is lower than the prevailing mid-quote, thus it is immediately executable. The dotted line shows observations in the first half of 8:59:59 (8:59:59.000 - 8:59:59.499) and the black line shows the last half of 8:59:59 (8:59:59.500 - 8:59:59.999). The data on the order flow are provided by Tokyo Stock Exchange.
Figure 6: Mid-quote changes in the pre-opening period

Figure 6 shows the percentage of price changes caused by submission of orders for the last minute of the pre-opening period (8:59:59.000 - 8:59:59.999) for 97 stocks from TOPIX 100 during the sample period (April and May 2013). The events: ‘up’ means that the submitted order generates a positive impact on the mid-quote; ’down’ means that the submitted order generates a negative impact on the mid-quote. The data on the order flow are provided by Tokyo Stock Exchange.
Figure 7: Number of events (stock-second) that have an impact on the ask and bid price

Figure 7, Panel A and B, shows that the number of stock-second that experienced a negative or positive average impact on the ask (bid) price for the last minute of the pre-opening period (8:59:59.000 - 8:59:59.999). We count stock-days per each second in which market impact is observed for 97 stocks from TOPIX 100 during the sample period (April and May 2013). The data on the order flow are provided by Tokyo Stock Exchange.

(A) Impact on ask price

(B) Impact on bid price
Figure 8: Average impact on ask and bid price during the last minute of the pre-opening period

Figure 8 (A and B) shows that the median market impact on the ask (bid) price in each second where we observe changes in mid-quotes for the last minute of the pre-opening period (8:59:59.000 - 8:59:59.999) for 97 stocks from TOPIX 100 during the sample period (April and May 2013). We separate events which have positive or negative changes on the ask price from the previous second. The data on the order flow are provided by Tokyo Stock Exchange.
Figure 9: Results of unbiasedness regression

Using mid-quotes at each second, we estimate equation (2):

\[ \nu - E(\nu|I_0) = \alpha_t + \beta_t [P_t - E(\nu|I_0)] + Z_t \]

where \( \nu \) is the opening price (instead of the closing price used in Biais, Hillion, and Spatt (1999), \( P_t \) is the pre-opening mid-quote, and \( E(\nu|I_0) \) is the previous day’s closing price. We estimate equation (2) for every second in the last ten minutes of the pre-opening period (8:50:00.000 - 8:59:59.999) and for each of the 97 stocks from TOPIX 100 in April and May 2013 for panel A. The averages of the \( \beta \) coefficient and RMSE across stocks are shown in figure 9 panel A. If the pre-opening mid-quote is an unbiased estimator of an opening price, the \( \beta \) coefficient equals 1. RMSE (root mean square error) quantifies the informational content of the preopening prices. Panel B depicts the unbiasedness regression, for every ten millisecond, estimate in the last ten seconds (8:59:50.000 - 8:59:59.999) during the pre-opening period. Three periods are displayed: Nov - Dec 2009, Jan - Mar 2010 and Apr - May 2013. The tick-by-tick data time-stamped to millisecond are provided by Thompson Reuters.
We estimate equation (2) also every ten millisecond in the last ten seconds of the pre-opening period (8:59:50.000 - 8:59:59.999) and for each of the 97 stocks from TOPIX 100 in 3 periods: November and Dec 2009, from January to March 2010 and April and May 2013 for panel B. The averages of the $\beta$ coefficient and RMSE across stocks are shown in figure 9 panel B. If the pre-opening mid-quote is an unbiased estimator of an opening price, the $\beta$ coefficient equals 1. RMSE (root mean square error) quantifies the informational content of the preopening prices. The tick-by-tick data time-stamped to millisecond are provided by Thompson Reuters.

**Figure 9: Results of unbiasedness regression (continued)**

![Diagram of Beta Coefficient](image)

![Diagram of RMSE](image)