Low-Latency Trading and Price Discovery: Evidence from the Tokyo Stock Exchange in the Pre-Opening and Opening Periods

Mario Bellia, *SAFE - Goethe University* [*]
Loriana Pelizzon, *Goethe University and Ca' Foscari University of Venice* [†]
Marti G. Subrahmanyam, *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* [¶‖]

July 2016

Abstract

We study whether the presence of low-latency traders (including high-frequency traders (HFTs)) in the pre-opening period contributes to price discovery and liquidity provision in the subsequent opening call auction. We empirically investigate these questions using a unique dataset based on server IDs provided by the Tokyo Stock Exchange (TSE), one of the largest stock markets in the world. Our data allow us to develop a more comprehensive classification of traders than in the prior literature, and to investigate the behavior of the different categories of traders, based on their speed of trading and inventory holdings. We find that HFTs dynamically alter their presence in different stocks and on different days; therefore, we focus on HFT activity only when traders utilize their low-latency capacity. We find that, in spite of the lack of immediate execution, about one quarter of HFTs participate in the pre-opening period, and contribute significantly to price discovery. They also contribute to liquidity provision in the opening call auction. In line with the previous literature, we also document that HFTs contribute to price discovery and are liquidity consumers during the continuous period. However, this result is driven by the three quarters of HFTs that were inactive in the pre-opening period. In contrast, those that were active in the pre-opening period contribute to liquidity provision in the subsequent continuous session. This indicates that, while HFTs contribute to both price discovery and liquidity provision, there is considerable heterogeneity in their contributions to both.

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

[*] 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
[‖] We are grateful to Jonathan Brogaard, Austin Gerig, Björn Hagströmer, Joel Hasbrouck, Frank Hatheway, Terry Hendershott, Andrei Kirilenko, Mark van Achter, anonymous high-frequency traders, and participants at the FMA European Conference 2015, the 4th International Conference on the Industrial Organization of Securities and Derivatives Markets: High Frequency Trading, the SAFE Microstructure Workshop, Goethe University, the Conference on Securities Markets Trends, Risks and Policies, CONSOB-BAFFI CAREFIN, Bocconi University, the Swiss Society for Financial Market Research SGF Conference, the Nippon Finance Association Meeting, and the CFS Conference on High Frequency Trading, for helpful suggestions. We also thank the Tokyo Stock Exchange for providing us with anonymous detailed account-level data, which formed the basis of the research reported on in this paper. This work was carried out with the generous financial support of EUROFIDAI, which we appreciate. We also thank the Research Center SAFE, funded by the State of Hessen Initiative for Research (LOEWE), for financial support. Darya Yuferova also gratefully acknowledges the Vereniging Trustfonds Erasmus Universiteit Rotterdam for supporting her research visit to NYU Stern.

1. Introduction

Global equity markets have been fundamentally altered in the past decade due to vast improvements in the speed of trading and the consequent fragmentation of market activity. Among other changes, in most markets, traditional market makers have been replaced by high-frequency traders (HFTs), operating at the level of a few milliseconds or even microseconds.[1] This increase in trading speed allows markets to operate far beyond human capabilities, given that the average time it takes for a human to blink varies from 300 to 400 milliseconds. These technological developments have had a dramatic impact on the behavior of liquidity providers and consumers in financial markets, and have implications for price discovery. The resulting changes have led to intense debate and scrutiny from investors, market makers, exchanges, and regulators regarding the advantageous, even unfairly advantageous, status of HFTs in global markets.[2]

Regulators in many countries have been debating, and in some cases have implemented, new regulations on HFTs in recent years. A financial transaction tax has been adopted by France, Italy, and Canada. While there are many aspects of HFT activity and its impact on global equity markets, two issues come to the fore in any policy discussion on the role of HFTs. The first is whether HFTs contribute to price discovery in the sense that they improve the incorporation of new information into asset prices in terms of speed and accuracy. The second is whether they contribute to an improvement in market liquidity, which would permit market participants to enter and exit a position in an asset rapidly, and at a minimal cost.

These issues have to be posed in the context of the trading schedules of equity markets. Many equity markets around the world have several distinct periods during the trading

[1]See Brogaard (2010), Jovanovic and Menkveld (2015), Hendershott and Riordan (2013), and Raman and Yadav (2014), for evidence of this.

[2]See Lewis (2014) for a popular account of this perspective.

day: the pre-opening period, when quotes are placed and trades may or may not occur, an opening call auction, when buy and sell orders are crossed to determine an opening price, and a continuous trading period, when normal trading activity occurs, with posted quotes, orders, and trades. These three periods have distinct characteristics and the extent of HFT activity may vary across them. In this paper, we aim to investigate the HFTs' ability to contribute to price discovery and liquidity provision in these three different periods.

It is important to examine the first two periods before analyzing the continuous period, because the pre-opening period has very different characteristics. The pre-opening period is the first time in the day that information accumulated overnight gets incorporated into market prices, even if only quotes are posted and no trades occur. In contrast, the opening call auction is the first time in the day (after the previous day's closing) that market prices can incorporate new information accumulated overnight, based on actual trades. The continuous trading period, on the other hand, provides a much longer period of trading, but in the context of the arrival of new information, which may affect market fundamentals. HFTs' contributions to price discovery and liquidity may well differ among the three periods. We focus, in this paper, on HFT activity in the pre-opening period and the opening call auction, using the continuous trading period for comparative purposes. Given the growing presence of very fast traders in the market, the manner in which price discovery occurs during the pre-opening period is a crucial issue to be investigated, as is the liquidity provision during the opening auction.

The main questions we address in this paper are related to the role of fast traders, i.e., HFTs, in the pre-opening period, and the difference in trader behavior between this period, on the one hand, and the opening call auction and continuous trading period that follow. First, we investigate whether, in the absence of trading, HFTs still participate in the pre-opening period and, if they do participate, (i) whether they are more or less active in the pre-opening period than during the continuous period that follows, and (ii) how and precisely when they participate during the opening auction. Second, we investigate how the presence of HFTs in the pre-opening period contributes to price discovery in the continuous trading period that ensues. Third, we study whether HFTs provide liquidity in the opening auction, and compare the liquidity provision that occurs with that in the continuous trading period.

In order to empirically investigate these questions, we use a unique dataset provided by the Tokyo Stock Exchange (TSE), one of the largest stock markets in the world and the market with the largest presence of HFT activity: 55.3% compared to 49% in the U.S. market and 35% in the European market, as of 2012 (as documented by Hosaka (2014)). The TSE is also unique in relation to other major stock exchanges in the world, since it has a market share of over 90% of all Japanese stocks, while the NASDAQ and the New York Stock Exchange, for example, have a market share of well under 30% of their markets. Also, in terms of information flows, especially when stocks are traded in multiple locations, the TSE trades in a different time zone than the major stock exchanges in Europe and the U.S., which also trade some Japanese stocks directly or in the form of depositary receipts, permitting cleaner identification.

In the TSE, the execution of orders is not permitted during the pre-opening period, and hence buy/sell schedules cannot be crossed. In fact, traders cannot seek immediacy in this period; hence, HFTs that have the advantage of moving more quickly than other traders in reacting to new information or order flow cannot employ their superior ability to achieve speedy execution. This may result in almost no presence of HFTs in the opening auction period, although this warrants empirical scrutiny. Therefore, it is interesting to investigate the role played by HFTs during these periods. Ideally, it would be useful to obtain a trader-level classification of trading configuration, speed, and inventory, for each stock on a daily basis, rather than a classification based on overall trading characteristics.

In this study, we get closer to the level of granularity required to completely characterize HFT activity. Our methodology, which is entirely different from those used by prior researchers to identify HFT activity, is based on a novel dataset of virtual server (VS) IDs that cover *all* orders entered by traders in the TSE. A VS is a logical device that needs to be set up between the computer systems of the market participant and the exchange, such that they may send/receive data to/from one another. The unique dataset used in this paper is one of the most comprehensive on HFT activity, used in the literature thus far. Our methodology permits us to disentangle the types of traders, on a stock-by-stock, day-by-day, and period-by-period basis. It is comparable to studies having the trader identification data and, hence, offers several advantages for researchers that are worth highlighting. [3]

First, our data relate to trading information at the disaggregated level of the individual servers used for trading. Based on server usage, we are, therefore, able to infer account-level trading. We are, however, different from prior studies in the microstructure literature, that lack access to such disaggregated data, and are either forced to rely on an HFT/non-HFT flag or, when they do use account-level information, are able to cover only a small sample of the market, and even then typically focus only on the continuous period.

Second, given the granularity of our data, one can check whether there are differences between trader activity in the pre-opening period, the opening auction, and the continuous trading period that ensues, based on trader type, as determined by VS usage. In turn, our data allow us to measure the impacts of different types of traders on price discovery and liquidity provision. Thus, only with our data (or with trader identification data) can one

[3]The study closest in spirit to ours is one by Brogaard, Hagströmer, Norden, and Riordan (2015), who use subscription data for different speeds of co-location services as a screening device for HFTs. They distinguish between traders based on their usage of the low-latency facility, but do not have the relevant information on the server configurations of individual trading desks that we have. Our data are far more granular than theirs and permit a more detailed day-by-day and stock-by-stock analysis.

shed any light on the price discovery and liquidity consequences of HFT activity.

Third, our data permit a comprehensive classification scheme, which applies to the trading data on the stock-day-period basis. As we show in our paper, traders tend to switch their type from one day to another, and from one stock to the next; thus, the comprehensive nature of our data allows us to move away from the ad hoc assumption of immutable HFT classification: *"once an HFT, forever an HFT."*

Finally, our data cover all 1,702 stocks traded on the First Section of the TSE, comprising over 94.8% of the overall trading volume and 98.6% of the market capitalization of Japanese stocks, with reliable stamp data. Our analysis, therefore, refers to virtually the whole Japanese market, and not a subset of it as has been used in almost every prior study.

Using the granular data available to us, we classify traders into twelve subgroups based on speed and inventory behavior during the continuous period. In terms of speed, we identify three subgroups, namely FAST, MODERATE, and SLOW; in terms of inventory, we identify four subgroups, namely LARGE, MEDIUM, SMALL, and NOTRADE, based on end-of-day inventory. Although these two characteristics, speed and inventory, are generally used to identify HFTs, it is presumed that they are related to each other.

We investigate the issues defined above using the perspective of a policymaker who wishes to assess the benefits of HFT activity in the equity market in the pre-opening period and the opening call auction. We find that one quarter of HFTs trading within the continuous period also participate in the pre-opening and opening auction. However, even if they comprise only a minority of HFTs, their participation is highly relevant in terms of the number of quotes they make. We examine the contribution HFTs make to price discovery and find that they do contribute to it during the pre-opening period. Indeed, compared to other types of traders, they are the group that contribute the most. We also compare HFTs to other categories that differ in terms of either speed or inventory, to investigate which of those two characteristics matters more for price discovery, and conclude that, in the pre-opening period, speed matters more. We also examine the price discovery contribution made by HFTs in the continuous period and show that, as a group, they contribute to price discovery during that period as well (in line with previous literature, as in Brogaard, Hendershott, and Riordan (2014). However, the subgroup of HFTs that actively participate in the pre-opening section does not contribute to price discovery during the continuous period.

In terms of the HFTs' liquidity contribution during the opening call auction, our analysis shows that there is one. In line with the previous literature, however, we find that HFTs consume liquidity during the continuous period. The subgroup that actively participates during the pre-opening period, though, contributes to liquidity provision during both the opening call auction and the continuous period. For all three metrics of HFT activity - participation, price discovery, and liquidity provision - we perform several detailed analyses, aiming to disentangle the contribution of HFTs conditional on systematic and idiosyncratic volatility. Overall, we find that HFTs do not harm market quality when systematic or idiosyncratic volatility is high.

The outline of the paper is as follows. In Section 2, we survey the literature on price discovery, liquidity provision, and HFTs. In Section 3, we provide a detailed description of the research issues, the hypotheses we investigate, and the empirical methodology we employ. Section 4 describes the institutional details of the TSE, particularly for HFTs, and the special features of our database. In Section 5, we describe the data-filtering procedures we use to identify the 12 trader groups based on activity during the continuous period. Our empirical analysis and results are presented in Section 6. Section 7 presents our robustness check and Section 8 concludes.

2. Literature review

The recent HFT-specific theoretical literature deals with the speed advantage of HFTs in terms of both information processing and trading. Most of it focuses only on the continuous trading period. Their greater speed allows HFTs to react more quickly to public news than other traders (as in Jovanovic and Menkveld (2015), Biais, Foucault, and Moinas (2015), and Foucault, Hombert, and Roşu (2016)). Cespa and Foucault (2011) describe a new mechanism whereby dealers use the prices of other securities as information that generates spillover effects in terms of both price and liquidity, while Gerig and Michayluk (2014) differentiate HFTs from other traders in terms of their ability to monitor a large number of securities contemporaneously and, therefore, better predict future order flow. Pagnotta and Philippon (2011) analyze speed and fragmentation in a model in which exchanges invest in trading speed, finding that competition among trading venues increases investor participation, but leads to an excessive level of speed. Aït-Sahalia and Saglam (2014) explain that the low-latency environment increases the rates of quotation and cancellation on both sides of the market, and find that an increase in volatility reduces HFT activity. Biais, Foucault, and Moinas (2015) suggest that fast traders increase negative externalities, and thus adverse selection, crowding out slower traders. Jovanovic and Menkveld (2015) develop a model in which the ability of HFTs to process and react to new information more quickly than other market participants can generate both beneficial and deleterious effects.

The recent theoretical work of Budish, Cramton, and Shim (2015) advocates frequent batch auctions instead of the continuous auction that is currently predominant in global financial markets, a fairly radical departure from the prevailing regime. Frequent batch auctions coming at an interval of, say, every second, eliminate the arms race, because they both reduce the value of tiny speed advantages for HFTs and transform competition on speed into competition on price. The authors' model predicts narrower spreads, deeper markets, and increased social welfare. Another theoretical work, by Fricke and Gerig (2014), studies the optimal interval of the auction cycle based on earlier work by Garbade and Silber (1979). Their model predicts that an asset will be liquid if it has (1) low price volatility, (2) a large number of public investors, and (3) a high correlation between its and other assets' returns. These papers evoke shades of the debate on the switch from the current continuous auction to a periodic auction, which may reduce the speed advantage of HFTs.

Our paper provides empirical insights on HFT behavior in the batch auction setting. Also, to our knowledge, there are no papers that investigate the impact of HFT activity on the price discovery process in the *pre-opening period* that transitions into the opening batch auction. Our paper aims to fill this void. We are able to shed new light on this phenomenon by employing a rich, new database to study how HFTs place their orders before the market opening, and whether they increase the efficiency of price formation at the market opening. Our research follows earlier work in two distinct areas of the academic literature. The first relates to the microstructure of trading activity in the market pre-opening period, while the second relates to the impact of HFTs on price discovery. Although the pattern of the market pre-opening trading has been studied in the earlier literature (e.g., by Amihud and Mendelson (1991), Biais, Hillion, and Spatt (1999), Cao, Ghysels, and Hatheway (2000), Ciccotello and Hatheway (2000), Madhavan and Panchapagesan (2000), and Barclay and Hendershott (2003)), much of this literature is dated, and is based on research conducted well before the rapid growth in the number of HFTs over the course of the past decade or so. It is, therefore, necessary to examine trading activity in the pre-opening period once again, given the dramatic changes that have occurred since the advent of HFT activity.

To cite one example, the seminal work of Biais, Hillion, and Spatt (1999) emphasizes the difference between the price discovery processes in the pre-opening and continuous periods. Specifically, they test whether pre-opening quotes reflect noise (as orders can be revised or cancelled at any time before the opening auction) or true information. They find that, in the earlier period of the pre-opening period, quotes are likely to be pure noise. However, closer to the opening auction, the evidence is consistent with quotes reflecting information. They argue that there are two possible reasons for the large component of noise in the early part of the pre-opening period. First, noise could reflect the complexity of the price discovery process, in the absence of trade execution. Second, the manipulative behavior of traders could be contaminating the price discovery process. However, these reasons may no longer apply, due to the advent of rapid changes in information technology and the creation of a very fast trading environment, well known in the literature for encouraging HFT activity. Moreover, those authors do not distinguish between the different types of traders, based on their trading speed and inventory behavior.

Barclay and Hendershott (2003) analyze price discovery during the after-hours and pre-opening periods using U.S. stock data. They find that a larger degree of price discovery occurs during the pre-opening period than during the after-hours period. However, in the U.S. market, the execution of orders is possible during the pre-opening period, which is not the case in the TSE. Also, these authors do not distinguish between the different types of traders, and specifically between HFT and non-HFT order flow. To our knowledge, the only paper that investigates the specific behavior of different types of traders during the pre-opening period is that of Cao, Ghysels, and Hatheway (2000), which concentrates on market maker behavior. They find that non-binding pre-opening quotations of NASDAQ market makers convey information for price discovery in the absence of trading, although there was no fast trading in the period they consider.[4]

[4]According to Cao, Ghysels, and Hatheway (2000), dealers can trade during the pre-opening period via the electronic communication networks (ECNs). However, in practice, this trading activity is very low.

The body of empirical studies on HFT trading activities is growing rapidly.[5] It should be noted, however, that the focus of most of the literature is the *continuous trading* period, rather than the *pre-opening* period of the trading day. Baron, Brogaard, and Kirilenko (2012) estimate the profitability of high-frequency trading, while Hagströmer and Norden (2013) empirically confirm the categorization of HFTs into those that are engaged in market-making activities and those that are merely opportunistic traders. Menkveld (2013) analyzes the 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 stocks. He shows that, in 80% of the cases, HFTs provided liquidity on both markets, during the continuous trading period. In an event study framework, Brogaard, Hagströmer, Norden, and Riordan (2015) show that liquidity providers are willing to pay for higher trading speed (using a premium co-location service that allows traders to co-locate their servers near to the exchange's matching engine with upgraded transmission speed), and that this is beneficial for overall market liquidity. Finally, Gomber, Arndt, Lutat, and Uhle (2011), Menkveld (2013), and Kirilenko, Kyle, Samadi, and Tuzun (2015) document the typical behavior of HFTs during the continuous trading period, starting with a zero-inventory position at the beginning of the trading day. Some strategies employed by HFTs can consume liquidity from the market. McInish and Upson (2013) document an example of the structural strategy employed by HFTs and attempt to estimate the profits from this strategy, while Hirschey (2013) and Scholtus, van Dijk, and Frijns (2014) document the strategies of HFTs around news and macro announcements. Foucault, Kozhan, and Tham (2015) show that fast arbitrageurs can undermine liquidity by exploiting arbitrage opportunities in the FX market.

Studies on HFTs and market quality include Hendershott and Moulton (2011), Hender-

[5]For reviews of the burgeoning literature, see Jones (2013) and Biais and Foucault (2014).

shott, 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). However, none of these studies describe how HFTs prepare their positions during the pre-opening period, in anticipation of the continuous trading period, nor do they investigate the behavior of HFTs that carry inventories overnight. In contrast to the prior literature, the particular emphasis of this paper is on HFT behavior in the pre-opening period: If HFTs indeed have superior information-processing ability then it will be advantageous for them to place orders in the pre-opening period as well.

In summary, our paper is related to the previous and current literature on HFTs, but differs from them on several dimensions. First, it relies on a unique characterization of HFTs that is derived from the specifics of the trading technology (as described in detail in Section 5 below), rather than relying merely on trading metrics. Second, we use the *whole* market sample to identify different trader groups on the TSE. Other papers have relied on reasonably complete information but for a much smaller subset of the market. Our reliance on the identification of server IDs permits us to get around the problem of limited access to client-specific trading data, and yet obtain complete data for the whole market. Third, we focus on the pre-opening period to test hypotheses regarding the effectiveness of price discovery as a consequence of HFT activity.

3. Research issues, hypotheses and methodology

We take the perspective of a policymaker who wishes to assess the benefits of HFT activity in equity markets. Hence, our perspective is positive: What is the impact of HFTs on the functioning of such markets?[6] There are two principal potential benefits of such activity for

[6]We defer normative issues such as the optimal strategy for an HFT, including order placement, size, time etc., to future research.

markets: price discovery and liquidity. Price discovery refers to the speed and unbiasedness of the process by which new information, revealed through trading, is incorporated into the market price of an asset. Liquidity refers to the speed and cost of entering and exiting an existing position in the asset.

3.1. Research issues

The quoted or traded price of an asset changes in response to the arrival of new information. In addition, a trade or even a quote by one market participant, either in response to the information or for other reasons such as liquidity, may influence other market participants to change their quotes or orders. Hence, in a market where multiple players are posting quotes or placing orders, prices are constantly changing for both reasons: information and the trading itself. A natural question that arises with regard to assessing the effectiveness of any market innovation or structural change is whether it influences the speed with which the market price incorporates the influences of information and trading, and also whether and how rapidly that price converges to its equilibrium value. While equilibrium values may be difficult to assess in fast-moving markets, with the constant arrival of new information, it may be easier to assess how closely prices adjust to the price that is going to prevail at a particular point in the future, when several trades are likely to occur. For example, if quotes are being placed in the pre-opening period, a natural question would be whether a quote for a particular asset moves the price towards or away from the price that is going to prevail in the opening call auction. This is the concept of price discovery that we will explore in detail in this paper:

Definition. *Price discovery is defined as the effect of a given quote, trade, or type of trader on the deviation of the mid-price from the price that is going to prevail at a particular point in the future.*

A parallel question arises in relation to whether a given quote or order improves or deteriorates the liquidity in the market, as defined by its impact on immediacy and/or the price impact of the execution of orders. In general, a trade or order that reduces the effective bid-ask spread can be said to improve liquidity. To illustrate, if more quotes or orders are being placed on one side of the market than the other, there is a greater imbalance, and hence a decline in liquidity. For instance, if a particular type of trader decreases this imbalance between demand and supply, she is said to improve liquidity. In that sense, liquidity provision (or consumption) measures the extent to which a particular trader mimics the behavior of a designated market maker (who provides liquidity) or an informed trader (who consumes liquidity). Hence, the definition of liquidity in our context is:

Definition. *Liquidity is defined as the effect of a given quote, trade, or type of trader on the reduction in the imbalance between the demand and supply of an asset in the market.*

Given these broad conceptual definitions of price discovery and liquidity, we frame the research issues that we seek to investigate:

1. Do HFTs participate in the pre-opening period, despite the fact that no trading occurs during that period of time? If so, what is their level of activity?
2. Do the same HFTs participate in all three time segments, pre-opening, opening call auction, and the continuous periods?
3. Do HFTs participate in the opening call auction and, if so, with what intensity?
4. What is HFTs' contribution to price discovery in the pre-opening and continuous periods?
5. What is HFTs' contribution to liquidity in the opening call auction and continuous periods?

3.2. Hypotheses

We address the above questions in terms of specific empirical hypotheses, which are listed below:

Hypothesis 1. ***(HFTs are the most active participants in the pre-opening, opening call auction, and continuous periods.)***

This hypothesis establishes that HFTs are significant players in the market.

Hypothesis 2. ***(Of all types of traders, HFTs contribute the most to price discovery in the pre-opening and continuous periods.)***

Although there is no trading in the pre-opening period, the placement of quotes by the various traders has an impact on price movements during that period. The opening call auction that follows is a result of the price discovery process during the pre-opening period.

Hypothesis 3. ***(Of all types of traders, HFTs contribute the most to liquidity in the opening call auction and the continuous period.)***

Liquidity can be measured only when there is actual placement of orders and trading occurs, and not with the mere posting of quotations. Hence, it is relevant to speak of liquidity only in the case of the opening call auction and the continuous period.

3.3. Methodology

3.3.1. HFT participation

The key issue in addressing Hypothesis 1 is the presence of HFTs during the pre-opening section, characterized by the absence of trading. We also aim to test whether they have a presence that is larger than that of the other types of traders, and whether a significant proportion of HFTs participate in the pre-opening section. We investigate this hypothesis by looking first at their presence in the pre-opening period, the opening auction, and the continuous period. We define the trader presence ratio as the percentage of traders in trader group l that submit at least one quote for stock j, on day k, in any of the sections of the day, as depicted in Figure 1.

More formally,

$$PR_{p,j,k,l} = \frac{\text{Number of trading desks}_{p,j,k,l}}{\sum_{l} \text{number of trading desks}_{p,j,k}} \tag{1}$$

where p is the period considered (pre-opening, opening auction, continuous) for stock j, on day k, for the trader group l. Therefore, *Total number of trading desks*$_{j,k,l}$ is the number of traders in trader group l that submit at least one quote for stock j on day k.

We then investigate their presence in the pre-opening (and the continuous) period using the activity ratio based on the volume of quotes:

$$QAR_{p,j,k,l} = \frac{\text{Number of quotes}_{p,j,k,l}}{\sum_{l} \text{number of quotes}_{p,j,k}} \tag{2}$$

where p is one of the two periods considered, pre-opening or continuous (there are no quotes in the opening auctions, only trades), and where the ratio relates to stock j, day k, and trader group l.

For the opening auction, we calculate the activity ratio by considering the number of shares traded, instead of the number of quotes:

$$TAR_{p,j,k,l} = \frac{\text{Number of shares traded}_{p,j,k,l}}{\sum_{l} \text{number of shares traded}_{p,j,k}} \tag{3}$$

where p is one of the periods considered (opening auction or continuous), for stock j, day k, and trader group l. We calculate this other ratio for the continuous period as well.

3.3.2. Price discovery

The key issue in addressing Hypothesis 2 regarding price discovery is the informativeness of trading by HFTs. We investigate which trader groups contribute to the price discovery process, and compare the extent of their contributions using order-by-order data and associated quoted (traded) price changes. In this manner, we take advantage of our detailed data, as we can pinpoint an order that moves the quoted (traded) price, and thus we can identify which trader group submitted that order and the type of the order.

We measure the amount of new information incorporated into stock prices using the weighted price contribution, WPC (e.g., Barclay and Warner (1993), Cao, Ghysels, and Hatheway (2000), and Barclay and Hendershott (2003)). The WPC measures how much the prices deviate from the price that will prevail at a later date, for instance between the pre-opening period and the opening call auction. We apply the same methodology for both the pre-opening and the continuous period. Specifically, as we mentioned earlier, for the continuous period we concentrate our analysis only on the first thirty minutes because we wish to compare the behavior of traders between the pre-opening and the continuous period with a minimal effect from new information.

First, we define the price discovery contribution as the amount by which an incoming order moves the quoted (traded) price closer to the reference price. Thus, we compute the price discovery contribution (PDC) on an order-by-order (trade-by-trade) basis, as follows:

$$PDC_{i,j,k} = Deviation_{i,j,k} - Deviation_{i-1,j,k} \tag{4}$$

where $Deviation_{i,j,k}$ is calculated in two different ways for the pre-opening and continuous periods, but with a reduction in the deviation being viewed as price discovery.

For the continuous period, the deviation of the trading prices from the price at 9:30 am is calculated as

$$Deviation_{i,j,k} = \left| \frac{P_{i,j,k}}{P930_{j,k}} - 1 \right| \times 100 \tag{5}$$

where $P_{i,j,k}$ is the trading price at the time of the i-th transaction, for stock j, on day k, and $P930_{j,k}$ is the price at 9:30 am for stock j on day k. In order to determine the return during the first 30 minutes of continuous trading, we use the average trading price between 9:30 and 9:35 to avoid the bid-ask bounce problem. $Deviation_{i,j,k}$ in this case is the absolute deviation from the price at 9:30 am, of the traded price at the time of the i-th trade, for stock j on day k (see equation (5)). Similarly, $Deviation_{i-1,j,k}$ is the absolute deviation from the price at 9:30 am, of the traded price at the time of the $(i-1)$-th trade, for stock j on day k.

The amount by which $Deviation_{i,j,k}$ is lower than $Deviation_{i-1,j,k}$ is the contribution to price discovery made by the order that initiates the i-th trade. Orders that initiate the transaction are new market orders and new or revised limit orders that either lock in or "cross" the prevailing bid-ask spread.[7] We discard those transactions from the continuous trading period for which we cannot identify the initiating order.[8] We refer to orders that initiate a transaction as "aggressive orders" (as in Biais, Hillion, and Spatt (1995), Ranaldo (2004), Duong, Kalev, and Krishnamurti (2009), and Yamamoto (2011)). For the pre-opening, $Deviation_{i,j,k}$ is defined as

$$Deviation_{i,j,k} = \left| \frac{M_{i,j,k}}{O_{j,k}} - 1 \right| \times 100 \tag{6}$$

[7]Locked limit orders are orders with the limit buy (sell) price equal to the best bid (ask) price, while crossed limit orders are orders with limit buy (sell) price greater (smaller) than the best ask (bid) price (see Cao, Ghysels, and Hatheway (2000)).

[8]If an order imbalance causes a larger price change than the pre-specified amount (e.g., the maximum price change between two trades is 70 Japanese Yen in the price range 3000-5000 Japanese Yen), the TSE stops continuous trading and conducts a call auction. The TSE disseminates special quotes to notify the market about the trading halt. In our sample, less than 1% of the trades fall into this category.

where $M_{i,j,k}$ is the quoted price at the time of arrival of order i for stock j on day k, and $O_{j,k}$ is the opening price for stock j on day k.

$Deviation_{i,j,k}$ in this case is the absolute deviation of the quoted price from the opening price, immediately after order i is entered for stock j on day k (see equation (6)). $Deviation_{i-1,j,k}$ is the absolute deviation of the quoted price from the opening price, immediately before order i is entered for stock j on day k. The amount by which $Deviation_{i,j,k}$ is lower than $Deviation_{i-1,j,k}$ is the contribution to price discovery made by order i.

Among the orders submitted during the pre-opening period, we can identify those orders with the potential to impact the prevailing quotes. We call them "aggressive orders" (as for the continuous period described above). The TSE uses unique rules for determining the best pre-opening bid and ask quotes. These rules are different from those applied in the continuous period and are briefly explained in Section 4 and in Appendix A. There are four cases in which we categorize an order as aggressive: (1) all market orders; (2) a limit buy order with a limit price greater than or equal to the prevailing best bid; (3) a limit sell order with a limit price less than or equal to the prevailing ask; and (4) any orders submitted at a time when the best bid equals the best ask.[9] When $PDC_{i,j,k}$ as defined in equation (4) is negative, the deviation is reduced and the quoted (traded) price moves closer to the reference price.

We define the WPC for stock j, day k, and order i as

$$WPC_{i,j,k} = \frac{PDC_{j,k}}{\sum\limits_{j} |PDC_{j,k}|} \times \frac{PDC_{i,j,k}}{PDC_{j,k}} \tag{7}$$

[9]Such a situation occurs when the cumulative amount of buy orders equals that of sell orders. Thus, the next order must cause an imbalance between buy and sell orders, and make the best ask higher than the best bid price. We refer to such orders as "locked orders." Cao, Ghysels, and Hatheway (2000) analyze locked/crossed market quotes during the NASDAQ pre-opening period. In the TSE's pre-opening period, market best quotes may be locked, which means that the best ask equals the best bid, but crossed quotes (which means that the best bid is greater than the best ask) never happen, by rule.

where $PDC_{i,j,k}$ is the price discovery contribution of order i, for stock j, on day k; $PDC_{j,k}$ is the accumulated price discovery contribution for stock j, on day k. The first term of WPC is the weighting factor for the stock on day k. The second term is the percentage contribution of price discovery made by order i to the total price discovery, during either the pre-opening or the continuous period, for stock j on day k. Since the size of PDC varies for each stock and each day, the relative contribution adjusts for the scale difference across stocks as well as across trading days, and the first factor adjusts for the relative importance of price discovery across stocks on day k. When $PDC_{j,k}$ equals zero, we do not compute WPC for stock j on day k. We winsorize $PDC_{i,j,k}$ at the 0.1% and 99.9% levels.

There are several approaches in the existing literature, for measuring such informativeness. For example, Hasbrouck (1995) develops a methodology for estimating information shares. In addition to that, Van Bommel (2011) summarizes three prominent alternatives that have been used widely in the literature: (1) the variance ratio; (2) the R^2 of the regression of unbiasedness of prices; and (3) the weighted price contribution (WPC), which is the one used in this paper. In simple terms, the variance ratio compares the variances in, say, the pre-opening period and the continuous period, and the test looks at whether that ratio is equal to one. The R^2 of the unbiasedness regression tests, for example, whether the evolution of prices in the pre-opening period is an unbiased predictor of the prices in the call auction. Van Bommel (2011) argues that, if the price process is a driftless martingale, only the WPC is an unbiased estimator for the return variance explained during a time interval. On top of this, we prefer using the WPC methodology to the other metrics for a variety of reasons. First, we do not take a position on price efficiency or the long-term fundamental value of the asset. Rather, we measure the deviation with respect to the price observed later, at a given point in time. Our concept of price discovery is, therefore, different from the information share measure developed by Hasbrouck (1995) and Hasbrouck (2002), which requires the notion of an unobservable efficient price. Second, in view of the granularity of our data, we are able to provide a detailed analysis of how each action by a given trader can move the price, all other influences being held fixed. Given our ability to identify the actions of individual traders, we are not constrained to merely document a statistical relationship between trading intensity and subsequent price movements.

3.3.3. Liquidity

Since our data provide information at the millisecond level, on trade size, trade direction, and the initiator of the trade, we can study, in detail, for each trade, whether liquidity is provided or consumed, and by which trader. There are several alternative proxies that could be used to analyze market liquidity, but for our purpose here we define liquidity provision (consumption) depending on whether a trader acts like a market maker (taker). In other words, during the continuous period, a liquidity provider is one that posts orders that do not initiate trades: orders that are not market orders or marketable limit orders. We look, in detail, at all such orders and calculate, for each trader group, for each trade, for each stock, and on each day, the following liquidity provision (LP) and liquidity consumption (LC) ratios for the continuous period:

$$LP_{i,j,k,l} = \frac{\text{Number of shares traded}_{i,j,k,l} \mid \text{Trader}_l \text{ does not initiate trade}}{\text{Total traded volume of first 30 minutes of continuous period}_{j,k}} \tag{8}$$

$$LC_{i,j,k,l} = \frac{\text{Number of shares traded}_{i,j,k,l} \mid \text{Trader}_l \text{ initiates trade}}{\text{Total traded volume of first 30 minutes of continuous period}_{j,k}} \tag{9}$$

where LP represents the *liquidity provision* and LC the *liquidity consumption* of trade i for stock j on day k.

We then define the *net liquidity provision* for the *continuous period* by calculating the difference between the two ratios, that is

$$NLP_{i,j,k,l} = LC_{i,j,k,l} - LP_{i,j,k,l} \tag{10}$$

We calculate a similar measure for the opening call auction. In this case, we assume that the trader is a liquidity provider when she is trading in the opposite direction to the market. Otherwise, she is a liquidity consumer. We define the direction of the trade by analyzing the difference between the opening call auction price P_a and the closing price of the day before P_c. When $P_a > P_c$ ($P_a < P_c$), a selling (buying) trade is trading against the market and is therefore providing liquidity, and a buying (selling) trade is demanding liquidity. Based on this definition, we calculate, for the *opening call auction*,

$$LP_{i,j,k,l} = \frac{\text{Number of shares}_{i,j,k,l}\text{traded against the market}}{\text{Total traded volume of the auction}_{j,k}} \tag{11}$$

$$LC_{i,j,k,l} = \frac{\text{Number of shares}_{i,j,k,l}\text{ traded in the same direction as the market}}{\text{Total traded volume of the auction}_{j,k}} \tag{12}$$

We then calculate the net liquidity provision for the opening call auction by calculating the difference between the two ratios, as in equation (10) above.

3.3.4. Testing methodology

In the previous subsections, we have defined the different measures we use to investigate the three hypotheses. For these measures, we report, in the results section below, their means, and test whether these ratios are statistically different to those of the other trader groups by estimating the following panel regression:

$$Y_{p,j,k,l} = a_0 + \sum a_l * I_l + e_{p,j,k,l} \tag{13}$$

where $Y_{p,j,k,l}$ is the investigated measure, that is, participation measured by QAR or TAR, price discovery by WPC, and liquidity provision by NLP. p is the period considered (pre-opening, opening auction, continuous) for stock j, on day k, for trader group l. I_l is a dummy variable that equals 1 for trader group l. We then test whether the coefficients a_l are statistically different from zero, and statistically different from one trader group to another.

We also perform a second test on whether the HFT presence, the price discovery contribution, and the liquidity provision are statistically larger than for all the other trading groups considered as non-HFTs, in line with the analysis presented in the previous literature. More formally, we estimate the following regression:

$$Y_{p,j,k,l} = a_0 + a_{l(p,j,k,l)} * I_{HFT} + e_{p,j,k,l} \tag{14}$$

where p is the period considered (pre-opening, opening auction, continuous), for stock j, on day k, for trader group l, and I_{HFT} is a dummy variable that equals 1 for the trader group FAST/SMALL (HFT).

4. Institutional structure and data description

4.1. Institutional structure of HFT on the TSE

The trading schedule for the TSE is characterized by two separate trading periods, one in the morning and one in the afternoon, along with two pre-opening periods (one early in the morning and the other in the middle of the day).[10]

Figure 1 provides a representation of the different trading periods. In each trading period, morning and afternoon, the opening price is determined by a single price auction ("Itayose" in Japanese) that kicks off at 9 am in the case of the morning session, and at 12:30 pm in the case of the afternoon session, based on buy and sell orders accumulated during the pre-opening periods that precede them.[11]

INSERT FIGURE 1 HERE

[10]In Japan, there are two private venues, SBI Japannext and ChiX-Japan, whose trading periods start before 9 am; however, very little trading occurs before 9 am. It should be mentioned that, for the Nikkei Stock Index Futures, traded on the Singapore Stock Exchange (SGX) in Singapore, trading starts at 8:45 am, Tokyo time, and such trading may also contribute to price discovery.

[11]Single price auctions are used to determine closing prices as well. The rules for the establishment of the closing call auction price are slightly different from those for the opening call auction price. See TSE (2015).

The opening time of the TSE is not randomized, as happens on some European exchanges such as Deutsche Bourse. As soon as the opening price of a stock is determined, the continuous trading period for the stock commences. These three periods, the pre-opening, the opening call auction, and the continuous trading period are connected seamlessly. In the pre-opening period, investors are free to submit orders as they do in the continuous period. There are two general types of orders allowed on the TSE: limit orders and market orders.[12] Each trading day, the TSE starts receiving orders from its member-brokers at 8 am (12:05 pm for the afternoon session), and does so without executing any orders until the single price auction for the market opening begins, at 9 am (12:30 pm).

As soon as the TSE receives orders, it disseminates the pre-opening quotes, not just the best ask and best bid, but 10 quotes above and below the best quotes, to the market.[13] Every time it receives an order, the pre-opening quotes are refreshed. The best ask (bid) is identified as the smallest (largest) ask (bid) price at which the cumulative depth of the ask (bid) schedule is greater than the cumulative depth of the bid (ask) schedule (the determination of the best bid and ask, and the rules, are explained in detail in Appendix A). The principle underlying the order matching is based on price and time priority in the continuous period. In the opening call auction, however, time priority is ignored. That is, all orders placed in the pre-opening period, before the opening price has been determined, are regarded as simultaneous orders.

The opening auction determines the price at which the largest number of executions (in terms of number of shares) is possible. There are three conditions to be met at the auction:

[12] Traders can specify that an order only be eligible for execution at the opening auction. Should it not be executed at the opening auction, such an order will be cancelled automatically, rather than being moved to the continuous trading period.

[13] A subscriber to the full quotes service will be able to see information (price and quantity) on the entire book. However, the quantities for the best ask and the best bid will be the same as for the standard service.

(1) All market orders must be executed at the opening price. (2) Orders with a sell limit price lower than the opening price, and buy limit price higher than the opening price, must be executed. (3) Buy and sell orders with limit prices equal to the opening price must be executed for the entire amount of either the buy or the sell side. The third condition suggests that, often, orders on either side, whose limit prices are equal to the opening price, are not fully executable. If this happens, the TSE allocates the available shares to participating member firms on a pro-rated basis.[14] As explained above, the speed (or timing) of order submission does not matter during the pre-opening period because of the lack of time priority in the opening call auction. Therefore, investors care about the timing of their submission only due to the third condition above and the treatment of unfilled orders in the continuous period, when time priority is activated. It should be noted that the feature of the opening call auction, whereby there is no time priority for limit orders submitted during the pre-opening period, can cause delayed order submissions, price revisions, and cancellations, until just before market opening.

4.2. Server IDs and data

We use order data covering the complete history of an order (new order entry, execution, revision of quantity or price, and cancellation in the pre-opening and continuous trading periods) obtained from the TSE. Each historic record is time-stamped at the millisecond level, and includes information on order type, side (buy or sell), number of shares, limit price, unique order number, and server ID (VS). We also use data on the filled orders, and unfilled orders (cancelled immediately) for the opening call auctions, during the period of our study, between April 1 and May 31 2013. We focus on April and May 2013 since the

[14]For further details of pro-rated allocation refer to TSE (2015, pp. 28–20).

volatility of the stock market rose around that time, after the new governor of the Bank of Japan, Haruhiko Kuroda, announced a new aggressive quantitative easing (QE) policy for the bank. Given the structural change in the regime, a number of unexpected events occurred during this period, rendering the role of the pre-opening quotes more crucial than at any other time.

The unique feature of this study is that we use novel data provided by the TSE, which include the unique IDs of the VSs (Appendix D describes the typical setup of VSs). We find that 5,580 such servers were used in the sample period considered. However, we observe that orders pertaining to the same stock are sometimes submitted by different servers; for example, an order submitted by VS "A" has been cancelled by VS "B." To determine the relationship between servers, we therefore investigate the *entire* universe of stocks traded on the TSE's First Section in the period considered, a total of 1,702 stocks.[15] By combining all the order and trade information in our sample period, we identify 3,021 groups, which we call traders.[16] Figure 2 depicts the sizes of the traders based on the number of VSs they employ. Among the 3,021 traders, 329 utilize between 2 and 41 VSs, while the rest (2,692) use only a single VS.[17]

INSERT FIGURE 2 HERE

Thanks to our matching procedure, our database is similar to the one with the anonymous trader identifiers, because we are able to track the traders' behavior through time. To our knowledge there are only few studies in the HFT literature that use a database with

[15]Stocks listed on the TSE are split into different sections based on their market capitalization, the number of shareholders, and other parameters. The First Section of the TSE includes relatively large companies, and comprises about 94.8% of the overall trading volume and 98.6% of the market capitalization.

[16]In Appendix D, we describe how we identify "traders".

[17]In contrast to Brogaard, Hagströmer, Norden, and Riordan (2015), who use the *grade* of the co-location service as a categorizing device for measuring the speed requirements of traders, we focus instead on how traders *configure* their respective trading environments.

user/trader identifier: Hagströmer and Norden (2013), Malinova et al. (2013), Menkveld (2013), Brogaard et al. (2015), Baron et al. (2016), and Korajczyk and Murphy (2015). The data typically come from Nordic (NASDAQ OMX) or Canadian (Toronto Stock Exchange) markets, with the exception of Menkveld (2013) who focuses on the Dutch market. However, market fragmentation poses an important limitation for the abovementioned studies and, in particular, for the coverage of platforms on which a particular stock is traded. Even if the data cover not only regulated exchanges, but also alternative trading venues, such additional data are limited to the country under investigation, thus excluding, for instance, the data regarding traders' behavior in respect to Canadian stocks being traded in the U.S. markets. Moreover, even if information from different platforms is available, it is not a trivial task to synchronize the time stamps of trades and quotes across different venues, which makes it virtually impossible to analyze which particular order has moved a price. In contrast to the Nordic and Canadian equity markets, the Japanese market exhibits a very low degree of fragmentation. In particular, only 5.5% of the total trading volume (2013) was hosted by private venues other than the TSE. Besides that, when the Japanese market is open, all major markets, including those in North America and Europe, are closed. The SGX opens at 8.45am and starts trading the future contract on NIKKEY.

Last but not least, the Japanese market is a bigger market than the Nordic and Canadian and thus likely to attract more HFT activity. In fact, as mentioned above, according to Hosaka (2014) the TSE market has an HFT presence of 55% of trades, compared to 49% in the U.S. and 34% in Europe. All in all, our data have an important advantage over the data previously used in the literature as they allow us to avoid issues that arise from market fragmentation and therefore allow us to provide a more complete picture of HFT activity. To our knowledge, we are also the first to utilize trade/quote data with trader identifiers to analyze the pre-opening period. It should be stressed that our data are composed of all the stocks in the TOPIX100, the index of major stocks, in Japan. Additionally, we do not have the issue of other stock exchanges trading the same stock at the same time.

Table 1 presents the characteristics of the traders, based on the trading environment of the 1,702 stocks in our universe. As the table shows, the median number of stock traded, per day, per group of VSs, ranges from 183 when only one server is used, to 963 when 40 servers are used. The median number of stocks traded, per server, decreases monotonically from the group that uses only 1 server (for 183 stocks) to the group that uses 20-29 servers (for 16 stocks). Increasing the number of servers beyond 30 leads to an increase in the number of stocks allocated to each server. We also investigate the maximum speed capabilities of the different traders as well as the end-of-day inventories. We measure the speed as the minimum time that elapsed between two consecutive order submissions for the same stock. Traders with just a single server place orders with a median speed of 12 seconds, and a median inventory of 100%. These characteristics match those of retail and wholesale brokers, which typically have several buy-side customers. For traders that use multiple servers, as the number of servers used by a trader increases from 2 to 41 the median speed moves to 1 millisecond; therefore, as soon as the number of servers used increases, speed increases as well. The median inventory varies considerably across traders, but is negatively related to the number of servers used, reflecting the variety of investment horizons among them, and indicating that traders with higher speeds usually hold less inventory at the end of the day.

INSERT TABLE 1 HERE

In the TSE, some traders, such as HFTs, use multiple VSs exclusively, because of a limitation on the number of quotes submitted per second for each server.[18] Using multiple

[18] The TSE provides three levels of service, with a maximum of 60, 40, and 20 quotes per second, respectively. We refer to all messages, including cancellations, in our definition of quotes. According to a prominent

servers, each trader optimizes the performance of the trading operations for their subset of stocks. Some traders operate within a specific group of stocks every day, in which case they may fix the allocation of stocks to each server. Other traders may change part of their allocation on a day-by-day basis. As the table shows, by using multiple servers, the traders are able to increase their trading speed significantly.

4.3. The universe of stocks and the sample period

For our analysis on market quality, we limit the sample of stocks to the constituents of the TOPIX100 index, during April and May 2013. This index is comprised of stocks from the TSE's First Section, with high liquidity and relatively large market capitalization, and comprises over 60% of its market capitalization. Of the TOPIX100 stocks, we exclude three that have larger trading volumes in exchanges other than the TSE, since the focus of our study is the trading system on this exchange.[19]

In our analysis, we exclude stock-days for which special quotes are disseminated before or during the single price auction, because orders submitted during the pre-opening period do not meet the normal opening price rules, in such cases. For the purpose of our analysis, we concentrate on three periods: the pre-opening section excluding the first 10 minutes, the opening auction, and the first 30 minutes of the continuous period. We exclude the first 10 minutes of the pre-opening period for the following reason: During the first 10 minutes, the limit order book accumulates many orders that were waiting overnight for the beginning of the pre-opening period of the TSE at 8 am. The arrival times of these orders are not directly related to the traders' actual submission decisions. Therefore, in our analysis of the

HFT, for a trader that wishes to be truly anonymous at least 20 VSs are necessary in order to implement a strategy of trading 1,500 stocks all at once. If the HFT also needs to cancel several orders immediately after submitting new ones, an additional 20 VSs may be required, making a total of 40 VSs necessary to support intensive HFT activity across multiple stocks.

[19]The three excluded stocks are Murata, Nintendo, and Nihon Densan.

pre-opening period, we focus on the remaining 50 minutes, during which traders monitor pre-opening quotes and make order submission decisions accordingly. We focus on the first 30 minutes of the continuous trading period, as we wish to analyze the difference in trader behavior, based on the same information, but in different trading periods: the pre-opening period, the opening call auction, and the continuous trading period. If we extended the sample to the full continuous trading period, we would contaminate our analysis with new information arriving in the market later in the trading day, and harm the comparability with the other periods.

5. HFT identification

The data we have allow us to track trader behavior, cluster traders into several different groups and, in particular, identify HFTs.

5.1. HFT identification strategy

A useful guideline defining the features of HFTs has been presented by theU.S.'s Securities and Exchange Commission (SEC). The document, (SEC, 2010, p. 45), lists five characteristics of HFTs:

1. *"Use of extraordinarily high speed and sophisticated programs for generating, routing, and executing orders."*
2. *"Use of co-location services and individual data feeds offered by exchanges and others to minimize network and other latencies."*
3. *"Very short time-frames for establishing and liquidating positions."*
4. *"Submission of numerous orders that are cancelled shortly after submission."*
5. *"Ending the trading day in as close to a flat position as possible (that is, not carrying significant, unhedged positions overnight)."*

Motivated by this list of characteristics, we use both speed and inventory to classify traders. These two metrics are closely related to all five characteristics listed above: speed matches characteristics 1, 2, and 4 above, while inventory matches characteristics 3 and 5. The speed is largely determined by the trading infrastructure in which each trading desk invests (the number of servers, the software programs used, the quality of servers installed, etc.), and which is not easily replaceable in the short run, whereas inventory is closely related to trading styles, such as those exhibited by buy-side investors, market makers, and arbitrageurs.

With these two measures in mind, we are able to investigate how the different traders' behavior affects the pre-opening period. One issue we have to address in our classification is whether the members of different groups of traders remain the same across time and stocks. To our knowledge, HFTs engage in a variety of strategies that do not necessarily remain the same from one day to the next or across stocks. In fact, HFTs implement multiple algorithms depending on whether they believe that the liquidity-taking or the liquidity-making strategy offers more profitable opportunities for a given stock on a given day. Therefore, we can assume that traders can engage in different types of trading strategies on a stock-by-stock and day-by-day basis.

Keeping this type of strategic behavior in mind, we compute our metrics on a per-stock, per-day basis, for all trading desks. Our aim is to investigate how the behavior of high-speed traders affects the pre-opening price. As far as we know, all the empirical studies in the literature, except ASIC (2013) and Benos and Sagade (2016), *assume* that HFTs behave in an identical manner on every day, and for every stock.

5.1.1. Speed

We empirically measure the minimum elapsed time between two consecutive order submissions for the same stock, without any restrictions, for a combination of two order types (i.e., any two out of new orders, cancellations, and revisions during our sample period), as a measure of speed.[20] We refer to our measure of time elapsed, using the time stamp of the order, as speed. In other words, a high speed means a short elapsed time between order submissions. Some authors refer to this concept as latency, but we prefer to use speed since we take latency to have connotations of the technological limits of speed, rather than the actual speed realized. Henceforth, we simply call our measure "speed".

A realization of a very high speed has to be supported by the appropriate trader's infrastructure. Hence, the number of servers a trader uses is a crucial determinant of her speed. As noted earlier, we observe varying numbers of servers, ranging from 1 to 41, used by the same trader, in our sample period. We also find that the number of stocks allocated to an individual server is negatively associated with the speed of the trader and vice versa. Appendix B provides a detailed analysis of the relationship between speed and quotes per server.

5.1.2. Inventory

The other major classification variable we employ is the inventory of the trader. Trader inventory is estimated as the (absolute) ratio of the buy volume minus the sell volume at the end of day k divided by the total trading volume of the trader on that day. Many empirical studies report that the key characteristic of HFT liquidity providers is a flat inventory position at the end of each trading day (Menkveld (2013), Kirilenko, Kyle, Samadi, and Tuzun (2015), and SEC (2014)). To investigate this issue further, we compute the end-of-day inventory for each trader and for each stock, as well as intraday inventory, volatility of intraday inventory, and the number of times during the day that the inventory crosses zero,

[20] Hasbrouck and Saar (2013) measure low-latency activity by identifying "strategic runs," which are linked submissions, cancellations, and executions that are likely to be part of a dynamic strategy. However, their data do not enable them to identify individual traders as we can.

per trader-stock-day.

5.1.3. *Classification*

We classify all traders according to observed speed and inventory during the continuous trading period for each stock-day. We apply the following classification scheme: We divide all traders, based on their speed, into three groups: FAST, MODERATE, and SLOW. For each stock-day, the SLOW group includes traders with a speed greater than 60 seconds. We then split the remainder of the speed distribution relative to the median. Therefore, the FAST group includes traders whose speed is less than the median, and the MODERATE group includes traders whose speed is greater than the median, but smaller than or equal to 60 seconds. Where we are unable to compute the speed due to the absence of multiple orders for the same stock on the same day, we treat the trader as a SLOW trader. Therefore, our definition of speed is not nominal speed but relative speed. This is in line with what Baron, Brogaard, Hagströmer, and Kirilenko (2016) find, in that it is relative speed and not nominal speed that drives differences in performance across HFTs: it is not being fast that allows an HFT to capture trading opportunities but being fastest. This is also in line with the theoretical literature on HFTs: Biais, Foucault, and Moinas (2015), Budish, Cramton, and Shim (2015), and Menkveld and Zoican (2015).

We also divide all traders into four groups based on their inventory for each stock-day: LARGE, MEDIUM, SMALL, and NOTRADE. Inventory at the end of the day is calculated by the number of shares bought minus the number of shares sold during the day, as a fraction of the total volume of shares bought and sold in the day for that particular stock. In particular, if a trader's inventory is equal to 100%, we consider the trader to be a LARGE inventory trader. If a trader's inventory is not computable, we consider the trader to be a NOTRADE agent. The rest of the distribution is split on a stock-day basis relative to the median to form the MEDIUM and SMALL inventory groups. It is important to note that we differentiate between a trader who ends a particular day with a flat inventory as a result of buy and sell activity throughout the day, and a NOTRADE agent. It should also be noted that NOTRADE agents include traders that submit orders, but whose orders are not filled. Table 2 briefly summarizes our scheme.

INSERT TABLE 2 HERE

5.2. HFT identification: summary statistics

Based on the classification of the groups we report in Table 2, we calculate several summary statistics on the behavior of the different groups of traders as reported in Table 3.

INSERT TABLE 3 HERE

Table 3 presents the summary statistics for speed and inventory for each group under our classification procedure. The median minimum speed in the FAST group varies, across different inventory subgroups, from 9 to 4 milliseconds. The MODERATE group exhibits a much higher median minimum speed, ranging from 3.3 to 6.8 seconds. The SLOW group has a median minimum speed of above 300 seconds. Note that similar patterns hold if we focus on the bottom 1%, 5%, or 10% of the speed distribution, instead of the specified minimum speed.

By construction, the LARGE inventory subgroup always has a 100% end-of-day inventory, meaning that, during the day, traders either only buy or only sell the stock. Traders from the MEDIUM inventory subgroup tend to end their trading day with an inventory of around 65% of the total volume traded, while traders from the SMALL inventory subgroup can end up with an inventory as low as 15%. Patterns for the median intraday inventory and its volatility are consistent with the end-of-day inventory. Besides that, we also estimate the number of times during the day that the signed inventory crosses zero. FAST/SMALL (HFTs) and MODERATE/SMALL exhibit the largest median number of zero crossings. Based on the speed and inventory classifications, one can consider FAST/SMALL traders as HFTs, as indicated in brackets in all the tables.

We emphasize that we use information from the continuous period on the same stock-day to describe trader behavior in the pre-opening period. This is motivated by changes in the traders' strategies from one day to another (see Table 4 for the transition frequency matrix of trader strategies). In particular, on average, only in 27.76% of cases do traders remain in the same group from one active stock-day to the next. The most persistent group is the SLOW/LARGE group (52.00%). Among FAST traders, the greatest persistence is observed for the FAST/SMALL (HFTs) group (40.85%). Within the same speed group, ignoring the differences in inventory, we observe more persistence: on average, traders tend to remain in the same speed group in 62.75% of the cases. Traders tend to remain in the same inventory group in 46.60% of the cases, on average, ignoring the speed dimension, with the largest contribution to this persistence coming from the LARGE inventory group.

INSERT TABLE 4 HERE

In the robustness section (Section 7.1), for comparison purposes, we present the results we obtain when we apply a classification scheme following Brogaard, Hagströmer, Norden, and Riordan (2015), a modification of the Kirilenko, Kyle, Samadi, and Tuzun (2015) approach, which splits traders into two groups, namely HFTs and non-HFTs, based on three criteria: end-of-day inventory, inventory at the end of each minute, and volume traded. As shown in the robustness section, this classification does not always identify very fast traders and their activity during the pre-opening period.

6. Results

We now present the results of our empirical investigation, in line with the hypotheses presented in Section 3 above. The presentation consists of three subsections, dealing with (1) the participation of HFTs at three different times, the pre-opening period, the opening auction, and the continuous trading period; (2) the contribution of HFTs to price discovery during the pre-opening and continuous trading periods; and (3) the impact of HFT trading on liquidity in the opening call auction and the continuous trading period. In the robustness analysis, we also check whether HFTs change their behavior in terms of (1), (2), and (3) in different volatility regimes.

6.1. HFT participation

Hypothesis 1. ***(HFTs are the most active participants in the pre-opening, opening call auction and continuous trading periods.)***

We investigate the participation of the different trading groups, and specifically those that we identify as HFTs (the FAST/SMALL group), in the pre-opening period, the opening auction, and the first 30 minutes of the continuous period, using the whole continuous period as a benchmark. We first compute the presence ratio (PR) defined in equation (1). Table 5 shows the PRs of all twelve groups in the three periods.

INSERT TABLE 5 HERE

The PRs of HFTs during the pre-opening period and the opening call auction are 26.96% and 17.47%, respectively. These results address the question of whether HFTs participate in the pre-opening period, despite the fact that no trading occurs at that point in time. Our analysis shows that, even though there is no trading, one quarter of HFTs do participate in the pre-opening period. Table 5 also shows that in the continuous period (the first 30 minutes) 83.65% are present, confirming that the large majority of the HFT traders do participate during the early part of the trading day. Table 5 also shows that 16.16% of HFTs actively participate in all three periods: pre-opening, opening auction, and continuous. The table also reports a distinction between HFTs that actively trade in the opening auction (Active-w-Trade) along with those that are present in the pre-opening period but do not trade during the opening auction (Active-w/o-Trade), and on the other hand those that are Non-Active in both the pre-opening period and the opening auction, but do participate in the continuous period: in the continuous period, almost all of the Active-w-Trade HFTs also participate in the continuous period, and the same applies to Active-w/o-Trade HFTs. However, the majority of HFTs present in the continuous period are Non-Active (58.91%). Figure 3 shows the presence of HFT traders during different periods of the trading day. In particular, we show that 26.96% of HFTs are present in the pre-opening period. Out of these traders, 65% (=17.47/26.96) execute their orders in the opening call auction. Among those traders that execute their orders in the opening call auction, 93% (=16.16/17.47) are also present in the first 30 minutes of the continuous trading period.

INSERT FIGURE 3 HERE

Going beyond mere presence, we next compute their level of activity, based on quotes submitted by traders in the different periods, that is, the QAR defined in equation (2). In Table 6, Panel A, we report the average QAR.

INSERT TABLE 6 HERE

The panel shows that HFTs represent the trader group with the largest proportion of orders submitted, with about 20.58% of the total orders submitted during the pre-opening period, defined as averages, which splits into about 4.98% by the Active-w/o-Trade and 15.60% by the Active-w-Trade group, respectively. For comparison, the table shows that HFTs are also the trading group submitting the majority of quotes during the continuous period (42.05%), almost double the percentage submitted during the pre-opening period. This indicates that the majority of the quotes in the continuous period are submitted by HFTs that are not present during the pre-opening period (Non-Active).[21] The variation of the QAR within the speed and inventory categories also suggests that both the speed and inventory dimensions are important when analyzing the impacts of the different trader groups on market quality.

In Table 6, Panel B, we perform a formal test of whether this activity is also statistically the greatest (see equation (13)). We estimate the coefficient using the Active-w-Trade SLOW/SMALL group as a base case. We observe that Active-w-Trade HFTs are confirmed as being statistically significantly present, with the largest coefficient among all groups during the pre-opening period and the first 30 minutes of the continuous period. The test of significance is based on clustered standard errors (by stock). We also considered robust standard errors and the results were similar. We also performed a test of whether Active-w-Trade HFTs have a statistically larger presence during the pre-opening period than Active-w-Trade FAST/MEDIUM, MODERATE/SMALL, and MODERATE/MEDIUM traders. We reject the null hypothesis that the coefficients are the same at the 1% level, for Active-w-Trade

[21] We also analyze the behavior of traders based on aggressive orders alone. Please see Appendix C for details.

MODERATE/MEDIUM traders. However, we cannot reject the hypothesis that quoting activity is the same for FAST/SMALL (HFTs), FAST/MEDIUM, and MODERATE/SMALL traders. Put differently, Active-w-Trade HFTs are among the top three trader groups in terms of quoting activity during the pre-opening period, while Non-Active HFTs are the most active group during the continuous period.

Finally, we investigate whether HFTs as a trader group have a QAR that is larger than non-HFTs. The results are reported in Table 6, Panel C, and show that HFTs have a ratio of quote activity that is 6.11% larger than that of non-HFTs in the pre-opening period (12.09% larger in the continuous period). HFTs that are Active-w-Trade have a quote activity that is 11.39% larger than all other traders, as a group. However, Non-Active HFTs have a quote activity that is 28.10% larger than all the other traders in the continuous period. This test indicates that, even if there is a huge dispersion in their participation across stocks and across days, there is always a significant fraction of quotes posted by HFTs in both the pre-opening and continuous periods.

So far, we have presented a comparative analysis of the pre-opening and continuous periods. We repeat the above analysis for participation in the opening auction, using trades rather than quotes, to measure the activity ratio TAR, defined in equation (3). The results are reported in Table 7.

INSERT TABLE 7 HERE

Table 7, Panel A, shows that HFTs do participate significantly in the opening auction: 36.65% of the volume traded during the auction relates to HFTs, either as sellers or buyers. Compared with the other trader groups, they trade the second largest fraction. In line with previous empirical evidence, HFTs also contribute greatly to the volume traded during the continuous period. The same panel shows a large difference between the volume traded by

HFTs present in the auction, and those present only during the continuous period. Those that are not active during the auction (Non-Active) are responsible for 44.40% of the total volume of trades during the continuous period, while those that are active during the opening auction (Active-w-Trade) account, on average, for 19.79%.

We next perform a test of whether HFT participation is statistically different from zero, and not driven by outliers (see equation (13)). These results are reported in Table 7, Panel B. As shown, the presence of HFTs in all periods is significantly different from the TAR of the base-case group (Active-w-Trade SLOW/SMALL), and also different from zero, at the 1% level. HFTs are the second among the trader groups participating in the opening auction, in terms of TAR, while the most active group are the FAST/MEDIUM traders (the F-test we perform shows the difference between the two groups' trading activity to be statistically different from zero).

We also perform a general comparison of HFTs versus non-HFTs, by grouping all the other trader groups into one, and the results are reported in Table 7, Panel C, showing that, during the opening call auction, HFTs exhibit a level of trading activity that is 16.23% higher than the level of the other traders, and showing this difference in presence to be statistically different from zero. During the continuous period, HFTs have a trading activity level that is 17.50% larger than that of the other traders. However, Non-Active (Active-w-Trade) HFTs have a trading activity level 38.42% (12.86%) larger than that of the other traders.

6.2. Price discovery

Hypothesis 2. ***(Of all types of traders, HFTs contribute the most to price discovery in the pre-opening and continuous trading periods.)***

Our data allow us to measure the PDC by individual order and, in particular, by individual aggressive order, so that we can aggregate the WPC according to the trader group that submitted the order, to show the proportion of the price contribution made by a particular trading group. We calculate the WPC as described in equation (7). The results of the WPC analysis are reported in Table 8.

INSERT TABLE 8 HERE

Table 8, Panel A, shows that the main contribution to the WPC comes from FAST/MEDIUM (-26.66%) traders, followed by the HFTs (-16.23%) and MODERATE/SMALL (-13.61%) traders, during the pre-opening period. Moreover, we find that HFTs that are Active-w-Trade during the auction contribute 13.67% to the WPC, while those that are Active-w/o-Trade account for only 2.56%. The panel also shows the results of the WPC analysis for the first 30 minutes of the continuous trading period. HFTs, in this case, contribute 35.47% to price discovery, the largest contribution among all the trading groups. However, there is a lot of heterogeneity among HFTs, with big differences between those that actively participate in the pre-opening period and the opening auction, and the Non-Active HFTs. Non-Active HFTs are responsible for -37.99% of the WPC during the continuous period. On the contrary, among the traders that are active during pre-opening period, Active-w-Trade HFTs no longer contribute to price discovery, and if anything detract from it (+3.85%).

Table 8, Panel B, presents the results of a formal test of whether the price discovery caused by HFTs is indeed the largest among the trader groups. We estimate the regression using the Active-w-Trade SLOW/SMALL group as the base case. The largest price discovery contribution during the pre-opening period is made by the FAST/MEDIUM traders, followed by the Active w-Trade HFTs. In the continuous period, the Non-Active HFTs show the largest WPC, while the Active-w-Trade HFTs show WPCs that are not statistically different from zero. Therefore, the statistical analysis indicates that Active-w-Trade HFTs do not contribute to price discovery, but neither do they harm it. Table 8, Panel C, investigates whether HFTs as a trader group have a WPC larger than that of non-HFTs. It shows that HFTs as a group contribute more to price discovery than the other trader groups (-3.70%, which is statistically significant at the 1% level), in the pre-opening period. However, only Active-w-Trade HFTs show a significant amount of price discovery that is statistically larger than that made by non-HFTs during the pre-opening period. Active-w/o-Trade HFTs actually reduce price discovery during the pre-opening period. In the continuous period, HFTs contribute significantly to price discovery, with this contribution coming largely from Non-Active HFTs. Active-w-Trade HFTs have an impact that is not statistically significant, while Active-w/o-Trade HFTs deteriorate price discovery. These results confirm the heterogeneity of HFTs and the difference in their contributions to market quality, conditional on the period considered: pre-opening versus continuous.[22]

One of the characteristics of the opening price is that its volatility is larger than that of the closing price (Stoll and Whaley, 1990; Amihud and Mendelson, 1991; Berkman, Koch, Tuttle, and Zhang, 2012). This stylized fact has been attributed to many factors, including the accumulation of information and concentration of orders overnight, with attendant consequences for asymmetry of information, and market-maker intervention. Amihud and Mendelson (1991) conclude that the higher volatility of the opening price of the morning session is due to the staleness of the reference price, which is the closing price of the previous day (18 hours prior to opening). However, in this section, the WPC analysis for the pre-opening period is based on the auction price, which may be less stale. To check whether this assumption affects the overall results for the major contributors to price discovery during the pre-opening period, we consider the average price between 9:30 and 9:35, rather than

[22]The results of the test of whether the contributions from trader groups are significantly different across stock-date are very similar to Table 8, Panel B (the results are reported in Table IA 2.1 of Internet Appendix IA 2).

the auction price, which is used for investigating price discovery for the continuous period. Our analysis confirms that the selection of the reference price does not affect our conclusions (with the results reported in Table IA 2.2 of Internet Appendix IA 2).

In the robustness section below, we perform several other complementary analyses related to price discovery. First, we investigate the impact of a particular order type on price discovery (similarly to Barclay and Warner (1993) and Chakravarty (2001)). Second, we investigate price discovery by looking at the deviation of the pre-opening quoted price from the opening price, computed for each second of the entire pre-opening period, in different stock groups. Finally, we test the price efficiency of the pre-opening quotes, using an unbiasedness regression that has been widely used in the literature (e.g., Biais, Hillion, and Spatt (1999), Barclay and Hendershott (2003), and Barclay and Hendershott (2008)).

6.3. Liquidity provision

Hypothesis 3. ***(Of all types of traders, HFTs contribute the most to liquidity in the opening call auction and the continuous trading period.)***

In this subsection, we investigate the contribution of HFTs to liquidity provision in the opening auction and the continuous period. We use the Net Liquidity Provision measure (NLP) defined in equation (10) based on the liquidity-demanding and liquidity-supplying trading volumes, relative to the total trading volume during the opening auction, and the first 30 minutes of the continuous period. Table 9 presents the NLPs during the opening auction and the continuous period.

INSERT TABLE 9 HERE

Table 9, Panel A, shows that HFTs provide liquidity to the market during the opening auction, but consume liquidity during the continuous period. Again, we find considerable heterogeneity among HFTs: Active-w-Trade HFTs provide liquidity during both the opening auction and the continuous period, while Active-w/o-Trade HFTs contribute to liquidity provision during the continuous period. However, Non-Active HFTs largely consume liquidity. This result is similar to what we find for the continuous period in relation to price discovery. The result can be theoretically justified if we keep in mind the canonical microstructure models of Kyle (1985), Glosten and Milgrom (1985) and others, in which informed investors consume liquidity and uninformed investors and market makers supply liquidity. Indeed, in the Kyle model, uninformed liquidity traders supply liquidity by selling/buying for other reasons, such as liquidity needs, while informed investors take advantage of this liquidity provision by executing their informed trades against them.

In Table 9, Panel B, we perform the regression analysis (see equation (13)) and the results confirm those reported in Panel A. Non-Active HFTs significantly consume liquidity and the other HFTs contribute to liquidity in both the opening auction and the continuous period. We perform an F-test to compare liquidity provision in the opening call auction by FAST/SMALL (HFTs), MODERATE/SMALL, and MODERATE/MEDIUM traders. We show that the main liquidity provider in the opening call auction is the MODERATE/MEDIUM trader group, while HFTs share second place with the MODERATE/SMALL traders. In Table 9, Panel C, we compare HFTs to non-HFTs and find that in the pre-opening period they contribute significantly to liquidity, while in the continuous trading period they largely consume liquidity. This result, however, is driven completely by Non-Active HFTs.

To conclude, Active-w-Trade HFTs are among the main liquidity providers in the opening call auction and the continuous period. Non-Active HFTs are major liquidity demanders in the continuous period.

6.4. Summary of results

Table 10 summarizes our results with respect to how HFTs behave. HFTs include all FAST/SMALL traders, Active-w-Trade HFTs are traders who participate in the pre-opening, call auction, and continuous periods, and Non-Active HFTs are those who do not participate in either the pre-opening or the call auction, but do participate in the continuous period. Panel A shows that only one quarter of HFTs participate, and their market share of quotes is 20.58% but their share of trades in the call auction is larger than their share of quotes. This implies that HFT orders play a central role in price discovery. This can be confirmed by their WPC (-16.23%), which is the second largest WPC contribution of all the trader groups (see Table 8, Panel C). It is worth noting that HFTs as a group provide liquidity in the call auction but consume liquidity in the continuous period.

Panels B and C reveal that the decomposition of HFTs based on activity during the pre-opening period is crucial in determining liquidity provision. Panel B shows that Active-w-Trade HFTs provide liquidity in the call auction and the continuous trading period. Panel C shows that, contrary to Active-w-Trade HFTs, Non-Active HFTs consume liquidity, while playing a greater role in price discovery during the continuous trading period. We confirm the heterogeneity among HFTs and their roles in price formation and liquidity provision.

INSERT TABLE 10 HERE

7. Robustness checks

7.1. An alternative classification scheme

For purposes of comparison, we present the results obtained when we apply a classification scheme following Brogaard, Hagströmer, Norden, and Riordan (2015), a modification of the Kirilenko, Kyle, Samadi, and Tuzun (2015) approach, which splits traders into two groups: HFTs and non-HFTs. In particular, under this classification, a trader is defined as an HFT in a particular stock if and only if, on at least 50% of the active days, she satisfies the following criteria: First, her end-of-day inventory is no greater than 10% of her trading volume, for that stock, on that day. Second, her inventory at the end of each minute is no greater than 15% of her trading volume for that stock on that day. Third, her trading volume in that stock, on that day, is in the top quartile of the total trading volume for all traders in that stock on that day. This classification scheme is applied to only one month, April 2013, as there was a change in the server ID definition at the beginning of May 2013.

INSERT TABLE 11 HERE

Table 11 presents a summary of trader characteristics based on this classification scheme. In particular, the median speed for HFTs is 2 milliseconds, compared to 4 milliseconds for FAST/SMALL traders. HFTs are characterized by a 0.95% end-of-day inventory, in contrast to the 15.20% end-of-day inventory of our FAST/SMALL traders. Note that HFTs also have a lower intraday inventory and lower volatility of intraday inventory than FAST/SMALL traders. The median number of times that the inventory crosses zero for HFTs equals 14 times per day, while the number is twice per day for FAST/SMALL traders. Based on this classification scheme, we identify five traders as HFTs. Under this classification, there is little activity coming from HFTs in the pre-opening period. Note that most of the observations are marked as non-HFTs, suggesting that the Kirilenko, Kyle, Samadi, and Tuzun (2015) scheme is a stricter (narrower) classification of HFTs than the classification we propose. We believe that the diversity of market participants in the TSE better suits our more comprehensive approach than the narrower alternative scheme proposed earlier.

INSERT TABLE 12 HERE

Table 12 shows how the two classification schemes compare to one another. In particular, we show that traders classified as HFTs under the Kirilenko, Kyle, Samadi, and Tuzun (2015) scheme are most likely to fall into either the FAST/SMALL or the MODERATE/SMALL group. Clearly, the Kirilenko, Kyle, Samadi, and Tuzun (2015) scheme has a narrowly specified definition of HFTs, and fails to capture the subtle differences in the activities of other groups. We, therefore, prefer our HFT identification strategy, given that the Kirilenko, Kyle, Samadi, and Tuzun (2015) scheme seems less appropriate for the TSE market, at least with the current thresholds in place.

7.2. HFT behavior in different volatility regimes

In this robustness check, we analyze whether the above results are affected by market-wide (systematic) and stock-specific (idiosyncratic) movements in stock prices from one day to the next. In fact, the SEC (2010) and others express concern about the impact of HFTs on market quality, especially during times of stress. In our analysis, we distinguish between systematic and idiosyncratic volatility. In this spirit, we investigate whether we can observe differences in the participation, price discovery, and liquidity provision of HFTs, when the daily price dispersion of the market is large, and when the daily price dispersion of a single stock is large, even when market volatility is low. We measure market volatility by the difference between the daily highest value of the index TOPIX and the lowest value, and we measure stock volatility by the difference between the daily highest price of the stock and the lowest price. We then calculate the median during the sample period of the aforementioned difference for the market and stock volatilities. We set a systematic volatility dummy to one when the value of the market volatility is above the median, and zero otherwise. We set the stock volatility dummy to one when the value of the stock volatility is above the median, and zero otherwise. We estimate the idiosyncratic volatility dummy as the difference between the stock volatility dummy and the product of the systematic volatility dummy and the stock volatility dummy.[23]

By interacting the systematic and idiosyncratic volatility dummies with the trader group dummies, we test the hypothesis of whether there is any difference in HFT behavior, during the days when there is high systematic or idiosyncratic volatility:

$$Y_{p,j,k,l} = a_0 + \sum \left(a_l * I_l + b_l * I_l * SysVol_k + c_l * I_l * IdioVol_{j,k}\right) + SysVol_k + IdioVol_{j,k} + e_{p,j,k,l} \tag{15}$$

where $Y_{p,j,k,l}$ is the investigated measure, that is, participation measured by QAR and TAR, price discovery by WPC, and liquidity provision by NLP. p is the period considered (pre-opening, opening auction, continuous) for stock j, day k, and trader group l. I_l is a dummy variable that equals 1 for the trader group l. $SysVol_k$ is a dummy variable that equals 1 if the high minus low range for the TOPIX index on day k is higher than the median. $IdioVol_k$ is a dummy variable that equals 1 if the high minus low range for stock j on day k is higher than the median and $SysVol_k$ is not equal to 1 on day k. We then test whether the coefficients b_l and c_l are statistically different from zero. In Tables 13, 14, 15, and 16, we

[23] We acknowledge that using the median may be not enough to identify stressful periods. However, since we only have 41 days in our sample, tests based on stricter criteria may suffer from a lack of power, as happens with the test in Brogaard, Hendershott, and Riordan (2014). Nevertheless, in unreported results, we repeat the analysis using the 90th percentile instead of the median and, if anything, the results become stronger.

report b_l (Panel A) and c_l (Panel B) from the regressions for the quoting activity, trading activity, price discovery, and liquidity provision of traders in different volatility regimes.

INSERT TABLE 13 HERE

The results in Table 13 confirm that all HFTs, except Non-Active HFTs, increase their quoting activity relative to other trader groups, in times of high systematic and idiosyncratic volatility, with this effect being more pronounced for idiosyncratic volatility. Non-Active HFTs do not change their behavior.

The results of the trading activity analysis are reported in Table 14. The analysis also shows that the trading activity during the opening auction and the continuous trading period of Active-w-Trade HFTs increases with systematic volatility, while other HFTs do not change their behavior. Idiosyncratic volatility also leads to an increase in the trading activity of HFTs relative to other trader groups, with the exception of Non-Active HFTs, who do not change their behavior.

INSERT TABLE 14 HERE

This is consistent with the analysis of the QAR: Non-Active HFTs tend to behave in a similar manner when volatility is higher. This finding is consistent with the idea that idiosyncratic volatility is indicative of the larger amount of stock-specific information, rather than stress alone. These results confirm the results of Brogaard, Hendershott, and Riordan (2014) for the continuous period. However, by considering both the systematic volatility and the idiosyncratic volatility, our analysis shows that, in both the pre-opening and the continuous section, HFTs do not reduce their participation on days when systematic or idiosyncratic volatility is high.

Next we investigate whether the above results are affected by different systematic and idiosyncratic volatility regimes. We find that most of the HFTs do not change their contribution to price discovery across different volatility regimes, with only two exceptions: First, Non-Active HFTs contribute more to price discovery during the first 30 minutes of the continuous trading period when systematic volatility is high. Second, Active-w-Trade HFTs contribute more to price discovery during the pre-opening period in times of high idiosyncratic volatility. Remarkably, *none* of the HFTs decrease their price discovery contribution in regimes with high systematic/idiosyncratic volatility.

INSERT TABLE 15 HERE

We next investigate whether there are differences in HFTs' liquidity provision across different volatility regimes. We observe that HFTs' liquidity provision in the opening call auction is not affected by volatility. We show that, in times of high systematic/idiosyncratic volatility, Non-Active HFTs consume less liquidity than in times of low systematic/idiosyncratic volatility. Active-w/o-Trade HFTs also increase their liquidity provision in times of high idiosyncratic volatility. The results are reported in Table 16. In brief, HFTs' liquidity provision does not decrease in highly volatile periods.

INSERT TABLE 16 HERE

To sum up, we do not find any evidence that HFTs decrease their presence in the market, or stop facilitating price discovery or providing liquidity, in times of high systematic or idiosyncratic volatility.

7.3. Alternative price discovery measures

In Section 3.3.2 we investigated price discovery by looking at the WPC measure, without analyzing the role different orders play in price discovery. Moreover, the analysis sums up the WPC for the period without distinguishing between the different impacts of price discovery second by second during the pre-opening period, paying no attention to the speed of price discovery. Finally, we look at the unbiasedness of the pre-opening quotes, another measure the literature proposes for investigating price discovery.

7.3.1. Price discovery by order type

Our data allow us to measure PDC by individual order, so that we can aggregate WPC according to the trader group that submitted the order, and show the proportion of the price contribution made by a particular trading group and order type (similarly to Barclay and Warner (1993) and Chakravarty (2001)).

INSERT TABLE 17 HERE

Which types of orders contribute most to price discovery? According to Table 17, the types of orders contributing most to the WPC are new limit and market orders submitted by Active-w-Trade traders. Cancellations of market orders and price revisions of limit orders also contribute to price discovery. On the other hand, quantity revisions and cancellations of limit orders increase the deviation of the quoted price from the opening price. Orders from Active-w/o-Trade traders have a marginal effect. Our overall results indicate that quote setting during the pre-opening period is conducted by the Active-w-Trade FAST/SMALL (HFTs), FAST/MEDIUM, MODERATE/SMALL, and MODERATE/MEDIUM groups. Therefore, traders with high speeds and small inventories are indeed the ones that contribute the most to price discovery during the pre-opening period, even though there is no trading in this period, and only a fraction of fast traders participate.

7.3.2. Speed of price discovery

One of the questions we aim to answer with this paper concerns price discovery. We have considered the pre-opening period without distinguishing the earlier part of the period from the minute just before the auction. However, the order submission activity during the pre-opening section varies as the auction time approaches. In particular, a detailed analysis of the order flow allows us to confirm that the number of order submissions rises right before the opening auction. For a detailed description of the the pre-opening order flow, including the quoting activity for our different categories, see Internet Appendix IA 1.

Therefore, we look more closely at the movements in the pre-opening quotes between 8:10 and 9:00 am to determine how quickly the pre-opening quotes approach the opening price for the day, closer to the auction time. For this purpose, we report in Figure 4 the absolute value of the relative deviation of the quoted prices from the opening price for each stock, on each day, calculated using equation (6) from Section 3.3.2.

INSERT FIGURE 4 HERE

Figure 4 shows the median of the second-by-second movements in the pre-opening quotes, across the 97 stocks. 10 minutes after the start of the pre-opening period, the deviation is around 0.66%. The deviation decreases slowly during the period from 8:10 to 9:00 am. 10 minutes before the opening auction, the deviation is around 0.39%. This suggests that around 60% (=0.39/0.66) of the total price discovery during the pre-opening period excluding the first 10 minutes, occurs during the last 10 minutes. The deviation diminishes to 0.06% just before the opening time. Hence, the orders submitted after 8:50 am play an important role in price discovery.

Next, we aim to answer the question of whether stocks with a greater presence of one trader group relative to another trader group show different patterns of price convergence of the quoted price to the opening price. We conduct this analysis in three steps: First, we investigate whether we observe a significant variation in the relative activity of different types of traders across stocks in terms of the proportion of aggressive order submissions. In particular, for each stock, we estimate the relative activity of each trader group as the number of aggressive quotes (quotes that could potentially have an impact on the quoted price) from each trader group relative to the number of aggressive quotes from all trader groups during the pre-opening period excluding the first 10 minutes, aggregated across stocks and days (see Table 18). FAST/SMALL (HFTs) exhibit wide variation in their activity from stock to stock for the pre-opening period excluding the first 10 minutes: from 5.80% to 58.65%.

INSERT TABLE 18 HERE

Second, based on the distribution of the relative activity of the traders, we separate the 97 stocks from the TOPIX100 into two groups: stocks for which the activity of FAST/SMALL traders (HFTs) during the pre-opening period excluding the first 10 minutes crosses a threshold of 30% (18 stocks), and all other stocks (79 stocks). Figure 5 presents the median absolute deviation of the quoted price from the opening price, per second of the pre-opening period, and for the last 10 minutes of the pre-opening period. Note that, for stocks that pass the 30% threshold, the median absolute deviation is always smaller than it is for stocks that do not pass the threshold. However, immediately before the opening auction, the absolute deviation is approximately the same for both stock groups. The maximum gap between the two series is 0.20%. During the last 10 minutes of the pre-opening period, the median gap size is around 0.10%, except in the last couple of seconds, during which the gap closes rapidly due to the convergence of the absolute deviation to the opening price of the second group of stocks. To sum up, the presence of the FAST/SMALL traders (HFTs) improves the efficiency of the price discovery process.

INSERT FIGURE 5 HERE

7.3.3. Unbiasedness of the quoted price

We test for the price efficiency of the pre-opening quotes across the two groups using an unbiasedness regression that has been widely used in the literature.[24] This technique is widely applied to characterize the extent to which there is learning and price discovery in the pre-opening period. In the cited contribution, the proxy of the equilibrium price v is usually the closing price of the day. We modify their framework for our purposes and estimate the following equation:

$$\nu - E\left(\nu|I_0\right) = \alpha_t + \beta_t\left[P_t - E\left(\nu|I_0\right)\right] + Z_t \tag{16}$$

where ν is the opening price (instead of the closing price used in Biais, Hillion, and Spatt (1999)), P_t is the pre-opening quoted price, and $E\left(\nu|I_0\right)$ is the previous day's closing price. The distribution of the change in price, from the previous day's close to the quoted price, varies over time as the opening time approaches. The amount of noise in the quoted price is also likely to vary with time. In this spirit, we estimate the unbiasedness regression using the specification shown in equation (16), for each 1-second interval and for each stock. If the pre-opening quoted price is an unbiased estimator of the opening price, the coefficient β_t in the specification should be insignificantly different from 1. We hypothesize that the earlier in the pre-opening period the coefficient β_t equals 1, the greater is the price efficiency of the pre-opening quote. We analyze the pattern of the value of the t-statistic, under the null hypothesis that β_t is equal to 1, over the pre-opening period. Figure 6 shows the β_t estimates

[24]Among other papers that use an unbiasedness regression to investigate price discovery are Biais, Hillion, and Spatt (1999), Barclay and Hendershott (2003, 2008), Comerton-Forde and Rydge (2006), and Chakrabarty, Corwin, and Panayides (2011).

for each second during the pre-opening period. From Figure 6, we can see that β_t steadily increases during the pre-opening period. However, β_t becomes insignificantly different from 1 only seconds before the opening call auction.

INSERT FIGURE 6 HERE

Furthermore, there are remarkable cross-section differences. To exploit them graphically, we follow the same approach as in Section 7.3.2. We divide the stocks into two groups based on the activity of FAST/SMALL traders (HFTs). Figure 7 shows the β_t estimates and t-statistic under the null hypothesis that β_t is equal to 1 for every second during the pre-opening period, for these two groups of stocks, for April and May 2013. Remarkably, the β_t for stocks subject to high activity from the FAST/SMALL traders (HFTs) differs insignificantly from 1 during the whole pre-opening period. On the contrary, the β_t for stocks subject to low activity from the FAST/SMALL traders (HFTs) increases slowly from 0.3 to 1, and only becomes insignificantly different from 1 seconds before the pre-opening. Overall, these results are consistent with FAST/SMALL traders (HFT) improving price discovery during the pre-opening period

INSERT FIGURE 7 HERE

8. Conclusion

The market pre-opening period and the batch auction are important features of many stock markets today. They are an ideal laboratory for investigating the potential role of HFTs in periodic batch auctions, when immediate execution is not possible. Our study examines activity in this trading period in the context of HFT activity, which has come to dominate global equity markets. Key questions we ask in this research are whether, in the absence of trading, very fast traders (including HFTs) still participate in the market, and how the presence of fast traders contributes to price discovery in the pre-opening period, and later on in the opening batch auction. In order to empirically investigate these questions, we use a unique dataset provided by the TSE, which allows us to develop a more comprehensive classification of traders than in the prior literature and to investigate the behavior of different categories of traders, based on their capability for high-speed trading.

We classify traders into three speed and four inventory groups (making a total of 12 groups) on a stock-day basis. We observe that, on average, in only 28% of cases do traders remain in the same speed/inventory group from one day to the next. We also show that FAST traders can act as both market makers (SMALL inventory) and position takers (LARGE inventory). It is, therefore, not appropriate to assume that HFTs always trade all stocks in the same manner, every day. Hence, our classification of traders based on both speed of trading and inventory, and varying across stocks and across days, is likely to throw additional light on the effect of HFT activity.

Our empirical results for the TSE show that FAST traders participate in the pre-opening period and in the opening auction to a lesser extent than in the continuous trading period. With respect to the total number of orders, however, FAST traders play a dominant role in the pre-opening period. They submit 51% of the total number of orders, while MOD-

ERATE and SLOW traders submit 42% and 7%, respectively. We find that FAST/SMALL traders, which we identify as high-frequency market makers, and FAST/MEDIUM traders, contribute the most to price discovery. These results indicate that HFTs contribute to price discovery, and lead the price formation process throughout the pre-opening period, through their intense activity in relation to new limit orders and price revisions. Cancellations of limit orders deteriorate price discovery but cancellations of market orders improve it. It is important to note that, due to the lack of immediacy of execution, the presence of FAST traders in the pre-opening period is smaller than in the continuous trading period. However, we find that a larger presence of FAST traders in the trading of a stock improves the price discovery process. Moreover, we show that FAST traders tend to select the stocks in which they are more active, strategically, based on the stocks' characteristics.

Our results suggest that three quarters of FAST/SMALL traders (HFTs) do not participate in the opening call auction. These traders are the most active players in the first 30 minutes of the continuous trading period (they are responsible for initiating around 30% of the trades). This suggests that the majority of fast traders prefer an environment in which immediate execution is possible. Our findings also suggest that FAST/SMALL traders (HFTs) that are active only during the continuous trading period (Non-Active) are responsible for the majority of the price discovery process and the majority of liquidity consumption. On the contrary, FAST/SMALL traders (HFTs) that are active during the pre-opening period, and execute their orders at the opening call auction (Active-w-Trade), are among the main liquidity suppliers.

To sum up, our results suggest that HFTs that participate in the pre-opening period are different from those that only participate in the continuous trading period. We emphasize the need for further research on how a switch to a periodic auction from the current continuous auction may impact the behavior of fast traders. Our findings offer some preliminary evidence in the context of the debate on the relative merits of periodic batch versus continuous auctions.

References

Aït-Sahalia, Y. and M. Saglam (2014). High frequency traders: Taking advantage of speed. *Working paper*.

Amihud, Y. and H. Mendelson (1991). Volatility, efficiency, and trading: Evidence from the Japanese stock market. *Journal of Finance 46*(5), 1765–1789.

ASIC (2013). Dark liquidity and high-frequency trading. *Australian Securities and Investments Commission: Report 331*.

Barclay, M. J. and T. Hendershott (2003). Price discovery and trading after hours. *Review of Financial Studies 16*(4), 1041–1073.

Barclay, M. J. and T. Hendershott (2008). A comparison of trading and non-trading mechanisms for price discovery. *Journal of Empirical Finance 15*(5), 839–849.

Barclay, M. J. and J. B. Warner (1993). Stealth trading and volatility: Which trades move prices? *Journal of Financial Economics 34*(3), 281–305.

Baron, M., J. Brogaard, B. Hagströmer, and A. Kirilenko (2016). Risk and return in high-frequency trading. *Available at SSRN*.

Baron, M., J. Brogaard, and A. Kirilenko (2012). The trading profits of high frequency traders. *Working paper*.

Benos, E. and S. Sagade (2016). Price discovery and the cross-section of high-frequency trading. *Journal of Financial Market*.

Berkman, H., P. D. Koch, L. Tuttle, and Y. J. Zhang (2012). Paying attention: overnight returns and the hidden cost of buying at the open. *Journal of Financial and Quantitative Analysis 47*(04), 715–741.

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

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

Biais, B., P. Hillion, and C. Spatt (1995). An empirical analysis of the limit order book and the order flow in the Paris Bourse. *Journal of Finance 50*(5), 1655–1689.

Biais, B., P. Hillion, and C. Spatt (1999). Price discovery and learning during the preopening period in the Paris Bourse. *Journal of Political Economy 107*(6), 1218–1248.

Boehmer, E., K. Y. Fong, and J. J. Wu (2014). International evidence on algorithmic trading. *Working paper*.

Brogaard, J. (2010). High frequency trading and its impact on market quality. *Working paper*.

Brogaard, J., B. Hagströmer, L. L. Norden, and R. Riordan (2015). Trading fast and slow: Colocation and market quality. *Review of Financial Studies 28*(12), 3407–3443.

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

Budish, E. B., P. Cramton, and J. J. Shim (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *Quarterly Journal of Economics 130*(4), 1547–1621.

Cao, C., E. Ghysels, and F. Hatheway (2000). Price discovery without trading: Evidence from the NASDAQ preopening. *The Journal of Finance 55*(3), 1339–1365.

Cespa, G. and T. Foucault (2011). Learning from prices, liquidity spillovers, and market segmentation. *CEPR Discussion Paper No. DP8350*.

Chakrabarty, B., S. A. Corwin, and M. A. Panayides (2011). When a halt is not a halt: An analysis of off-NYSE trading during NYSE market closures. *Journal of Financial Intermediation 20*(3), 361–386.

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

Ciccotello, C. S. and F. M. Hatheway (2000). Indicating ahead: Best execution and the NASDAQ preopening. *Journal of Financial Intermediation 9*(2), 184–212.

Comerton-Forde, C. and J. Rydge (2006). The influence of call auction algorithm rules on market efficiency. *Journal of Financial Markets 9*(2), 199–222.

Duong, H. N., P. S. Kalev, and C. Krishnamurti (2009). Order aggressiveness of institutional and individual investors. *Pacific-Basin Finance Journal 17*(5), 533–546.

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

Foucault, T., J. Hombert, and I. Roşu (2016). News trading and speed. *Journal of Finance 71*(1), 335–382.

Foucault, T., R. Kozhan, and W. W. Tham (2015). Toxic arbitrage. *HEC Paris Research Paper No. FIN-2014-1040*.

Fricke, D. and A. Gerig (2014). Too fast or too slow? Determining the optimal speed of financial markets. *DERA Working Paper Series*.

Garbade, K. D. and W. L. Silber (1979). Structural organization of secondary markets: Clearing frequency, dealer activity and liquidity risk. *The Journal of Finance 34*(3), 577–593.

Gerig, A. and D. Michayluk (2014). Automated liquidity provision. *Working paper*.

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

Gomber, P., B. Arndt, M. Lutat, and T. Uhle (2011). High-frequency trading. *Working paper*.

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

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

Hasbrouck, J. (2002). Stalking the "efficient price" in market microstructure specifications: an overview. *Journal of Financial Markets 5*(3), 1–19.

Hasbrouck, J. and G. Saar (2013). Low-latency trading. *Journal of Financial Markets 16*(4), 646–649.

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

Hendershott, T. and P. C. Moulton (2011). Automation, speed, and stock market quality: The NYSE's hybrid. *Journal of Financial Markets 14*(4), 568–604.

Hendershott, T. and R. Riordan (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*.

Hosaka, G. (2014). Analysis of high-frequency trading at Tokyo Stock Exchange. *Working paper*.

Jones, C. M. (2013). What do we know about high-frequency trading? *Working paper*.

Jovanovic, B. and A. J. Menkveld (2015). Middlemen in limit-order markets. *Working paper*.

Kirilenko, A. A., A. S. Kyle, M. Samadi, and T. Tuzun (2015). The Flash crash: The impact of high frequency trading on an electronic market. *Working paper*.

Korajczyk, R. A. and D. Murphy (2015). High frequency market making to large institutional trades. *Available at SSRN*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335.

Lewis, M. (2014). *Flash boys: A Wall Street revolt.* WW Norton & Company.

Madhavan, A. and V. Panchapagesan (2000). Price discovery in auction markets: A look inside the black box. *Review of Financial Studies 13*(3), 627–658.

Malinova, K., A. Park, and R. Riordan (2013). Do retail traders suffer from high frequency traders? *Working paper*.

McInish, T. H. and J. Upson (2013). The quote exception rule: Giving high frequency traders an unintended advantage. *Financial Management 42*(3), 481–501.

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

Menkveld, A. J. and M. A. Zoican (2015). Need for speed? Low latency trading and adverse selection. *Working paper*.

Pagnotta, E. and T. Philippon (2011). Competing on speed. *Working paper*.

Raman, V. and P. Yadav (2014). Liquidity provision, information, and inventory management in limit order markets: An analysis of order revisions. *Working paper*.

Ranaldo, A. (2004). Order aggressiveness in limit order book markets. *Journal of Financial Markets 7*(1), 53–74.

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

SEC (2010). Concept release on equity market structure. *Release No. 34-61358, File No. S7-02-10*.

SEC (2014). Equity market structure literature review. Part II: High frequency trading.

Stoll, H. R. and R. E. Whaley (1990). Stock market structure and volatility. *Review of Financial Studies 3*(1), 37–71.

TSE (2015). Tokyo Stock Exchange: Guide to TSE trading methodology.

Van Bommel, J. (2011). Measuring price discovery: the variance ratio, the r 2, and the weighted price contribution. *Finance Research Letters 3*(8), 112–119.

Yamamoto, R. (2011). Order aggressiveness, pre-trade transparency, and long memory in an order-driven market. *Journal of Economic Dynamics and Control 35*(11), 1938–1963.

Table 1: Traders' characteristics during the continuous trading period

This table shows characteristics of the trading infrastructure and behavior of traders on the Tokyo Stock Exchange, where 5,580 unique virtual server IDs are used by traders. We trace the usage of individual virtual servers and, during the continuous trading period, identify 3'048 trading desks (traders) using single (or multiple) server(s) for their trading. All traders are sorted into one of the six groups based on the number of servers they utilize. For each group, we describe the number of traders, median number of servers used per trade, number of stocks traded (in total and per server), median speed (minimum time elapsed between two consecutive orders for the same stock), median inventory (the median of the end-of-the-day inventory), median number of quotes (in total and per stock), and median volume share per day (the proportion of the buy volume plus the sell volume per trading desk). These characteristics are based on the continuous trading period activity for the period of April-May 2013, for 1,702 stocks on the Tokyo Stock Exchange. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

	Grouped by number of servers used					
	1	2-9	10-19	20-29	30-39	40-41
# of traders	2'718	215	80	19	11	5
# of servers	1	3	12.5	21	32	40
# of stocks traded in total	183	247	363	461	869	963
# of stocks traded per server	183	81	31	16	28	23
Speed	12.16	1.107	0.767	0.002	1.251	0.001
Inventory	100.0%	100.0%	99.2%	13.6%	75.0%	75.0%
# of quotes per stock-day	6	8	17	74	25	126
Volume share per stock-day	0.07%	0.17%	0.32%	0.73%	0.56%	0.92%

Table 2: Classification of traders

This table shows the traders' classification proposed in this paper. Specifically, we split all traders into 12 groups on a stock-day basis. To split traders, we use information from the continuous trading period on the same day. First, we divide all traders into 3 groups based on their speed (minimum time elapsed between two consecutive orders for the same stock): FAST, MODERATE, and SLOW. Second, we divide each speed group into 4 subgroups based on the traders' inventory (the absolute ratio of cumulative buy minus cumulative sell volume to cumulative buy plus sell volume at the end of the day): LARGE, MEDIUM, SMALL, and NOTRADE. The characteristics are given per group on a stock-day basis for the period of April and May 2013 for the 97 stocks from TOPIX100. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

SPEED	FAST	Traders with speed below the median (excluding all trader-stock-days for which the minimum speed is higher than 60 seconds)
	MODERATE	Traders with speed above the median (excluding all trader-stock-days for which the minimum speed is higher than 60 seconds)
	SLOW	Traders with speed greater than 60 seconds
INVENTORY	LARGE	Trader's inventory equals 100%
	MEDIUM	Trader's inventory above the median and less than 100% (excluding all trader-stock-days for which the inventory equals 100%)
	SMALL	Trader's inventory below the median and less than 100% (excluding all trader-stock-days for which the inventory equals 100%)
	NOTRADE	Trader submits orders that are not filled (zero trades - only quotes)

Table 3: Description of traders' characteristics

This table shows summary statistics for the classification of the traders during the continuous trading period according to the scheme proposed in Table 2 using information about speed and inventory from the same day's continuous trading period. Panel A shows the medians of the 1%, 5%, and 10% percentiles of the trader-stock-day speed distribution, and the median minimum speed (used for classification of traders). Panel B shows the median number of traders, median end-of-day inventory (used for classification of traders), median average intraday inventory and its volatility, and the median number of times during the day that the inventory crosses zero, per trader-stock-day. These characteristics are presented per group for the period of April and May 2013, for the 97 stocks from TOPIX100. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: Speed of traders

			Median of			
Speed	Inventory	# of traders	Min speed	P1 speed	P5 speed	P10 speed
FAST	LARGE	79	0.008	0.008	0.019	0.065
	MEDIUM	84	0.005	0.007	0.028	0.252
	SMALL (HFTs)	84	0.004	0.007	0.025	0.120
	NOTRADE	38	0.009	0.009	0.020	0.046
MODERATE	LARGE	95	5.004	5.005	5.993	8.979
	MEDIUM	76	4.003	4.010	7.893	15.815
	SMALL	71	3.298	3.383	7.422	15.057
	NOTRADE	48	6.801	6.801	7.586	9.544
SLOW	LARGE	212	1288.993	1288.993	1288.993	1288.993
	MEDIUM	39	540.659	540.659	540.659	543.852
	SMALL	33	531.157	531.157	531.157	535.923
	NOTRADE	37	333.359	333.359	333.359	333.651

Panel B: Inventory of traders

			Median of			
Speed	Inventory	# of traders	End-of-day inventory	Intraday inventory	Volatility of intraday inventory	# of zero crossings
FAST	LARGE	79	100.00%	68.75%	32.06%	0
	MEDIUM	84	66.67%	36.97%	20.90%	0
	SMALL (HFTs)	84	15.15%	15.33%	9.50%	2
	NOTRADE	38				
MODERATE	LARGE	95	100.00%	75.00%	33.33%	0
	MEDIUM	76	65.52%	37.29%	21.16%	1
	SMALL	71	16.67%	18.19%	11.08%	2
	NOTRADE	48				
SLOW	LARGE	212	100.00%	100.00%	35.36%	0
	MEDIUM	39	65.12%	42.11%	21.88%	1
	SMALL	33	15.56%	25.00%	17.35%	1
	NOTRADE	37				

Table 4: Transition matrix for trader classification

This table shows the transition matrix for the trader classification based on 97 stocks from TOPIX100 for April-May 2013. We split all traders into 12 groups on a stock-day basis, as described in Table 2, using information about speed and inventory from the same day's continuous trading period. Afterwards, we report the percentage of traders that either remain in the same group or move from one group to another between date $t-1$ (the last day when the trader was active in a particular stock) and date t for a particular stock. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

	Date t	FAST				MODERATE				SLOW			
Date $t-1$		LARGE	MEDIUM	SMALL (HFTs)	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE
FAST	LARGE	24.92%	14.74%	10.06%	8.24%	12.90%	5.20%	3.61%	4.92%	10.92%	1.50%	1.19%	1.79%
	MEDIUM	13.27%	30.97%	23.33%	1.84%	6.85%	8.81%	6.71%	1.33%	4.21%	1.20%	0.95%	0.54%
	SMALL (HFTs)	8.69%	23.02%	40.85%	1.66%	4.29%	6.44%	8.59%	1.02%	2.97%	0.98%	1.01%	0.48%
	NOTRADE	14.79%	3.71%	3.49%	35.62%	7.27%	1.50%	1.23%	16.97%	9.88%	0.86%	0.81%	3.86%
MODERATE	LARGE	10.52%	6.45%	4.09%	3.31%	24.06%	9.36%	6.28%	8.41%	18.81%	2.94%	2.35%	3.40%
	MEDIUM	5.00%	10.04%	7.51%	0.76%	11.36%	23.80%	20.33%	2.37%	8.26%	5.41%	3.99%	1.18%
	SMALL	3.85%	8.01%	10.43%	0.71%	8.10%	21.59%	28.38%	1.95%	6.99%	4.41%	4.28%	1.29%
	NOTRADE	8.01%	2.35%	1.86%	15.09%	16.31%	3.77%	2.86%	23.16%	16.67%	1.40%	1.35%	7.16%
SLOW	LARGE	4.18%	1.82%	1.37%	1.90%	9.17%	3.41%	2.66%	3.97%	52.00%	7.38%	5.78%	6.35%
	MEDIUM	2.76%	2.74%	2.20%	0.79%	7.06%	10.30%	7.89%	1.61%	35.70%	14.58%	10.62%	3.76%
	SMALL	2.82%	2.60%	2.86%	0.95%	6.83%	9.21%	9.67%	1.95%	34.33%	12.87%	11.64%	4.29%
	NOTRADE	3.90%	1.33%	1.09%	4.86%	9.09%	2.60%	2.55%	9.47%	33.86%	4.07%	3.98%	23.20%

Table 5: Presence ratios

This table shows presence ratios for trades classified during the continuous trading period according to the scheme proposed in Table 2 using information about speed and inventory from the same day's continuous trading period. We report the average presence ratio during the pre-opening period, opening call, first 30 minutes of the continuous trading period, and presence ratio for all three subperiods. Presence ratio is defined as the ratio of the number of traders that are active during each of the subperiods to the number of traders that are active during the whole continuous trading period averaged across stock-days. These ratios are presented per group for the period of April and May 2013, for the 97 stocks from TOPIX100. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

		Pre-opening period			Opening auction	First 30 minutes of the continuous trading period				All three subperiods
Speed	Inventory	Active-w/o-Trade	Active-w-Trade	Total	Active-w-Trade	Active-w/o-Trade	Active-w-Trade	Non-Active	Total	Active-w-Trade
FAST	LARGE	10.29%	6.16%	16.45%	6.16%	8.36%	4.94%	48.27%	61.57%	4.94%
	MEDIUM	10.94%	22.04%	32.98%	22.04%	9.86%	19.75%	53.79%	83.40%	19.75%
	SMALL (HFTs)	9.49%	17.47%	26.96%	17.47%	8.58%	16.16%	58.91%	83.65%	16.16%
	NOTRADE	6.80%		6.80%		5.10%		37.96%	43.06%	
MODERATE	LARGE	11.60%	6.66%	18.26%	6.66%	7.60%	4.38%	34.13%	46.11%	4.38%
	MEDIUM	19.36%	30.69%	50.05%	30.69%	16.38%	26.48%	33.42%	76.28%	26.48%
	SMALL	18.26%	31.82%	50.08%	31.82%	15.45%	28.79%	33.82%	78.06%	28.79%
	NOTRADE	5.74%		5.74%		2.95%		25.09%	28.04%	
SLOW	LARGE	9.21%	7.40%	16.61%	7.40%	3.49%	1.15%	16.52%	21.16%	1.15%
	MEDIUM	20.03%	20.09%	40.12%	20.09%	13.36%	11.28%	23.88%	48.52%	11.28%
	SMALL	18.54%	17.12%	35.66%	17.12%	11.84%	9.44%	24.46%	45.74%	9.44%
	NOTRADE	42.25%		42.25%		0.99%		3.85%	4.84%	

Table 6: Quoting activity

This table shows proportion of quoting activity (new orders, revisions, and cancellations) by 12 trader groups for the pre-opening period (8:10 - 8:59) and during the first 30 minutes of the continuous trading period. Traders are classified according to the scheme proposed in Table 2 using information about speed and inventory from the same day's continuous trading period. We also split traders into 3 categories: traders that do not participate in the pre-opening period (Non-Active), traders that participate during the pre-opening period, but do not trade in the pre-opening period auction (Active-w/o-Trade), and traders that participate during the pre-opening period and trade in the call auction (Active-w-Trade). Average quoting activity is presented per group for the period of April and May 2013, for the 97 stocks from TOPIX100. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: Average quoting activity

		Pre-opening period			First 30 minutes of the continuous trading period			
Speed	Inventory	Active-w/o-Trade	Active-w-Trade	Total	Active-w/o-Trade	Active-w-Trade	Non-Active	Total
FAST	LARGE	3.52%	2.23%	5.75%	1.25%	1.52%	7.16%	9.93%
	MEDIUM	3.59%	14.31%	17.90%	2.00%	11.09%	10.06%	23.15%
	SMALL (HFTs)	4.98%	15.60%	20.58%	2.54%	9.23%	30.28%	42.05%
	NOTRADE	1.10%		1.10%	0.36%		4.71%	5.07%
MODERATE	LARGE	2.83%	2.34%	5.17%	0.58%	0.46%	2.46%	3.50%
	MEDIUM	4.03%	12.23%	16.26%	0.98%	2.45%	2.06%	5.49%
	SMALL	3.64%	15.04%	18.68%	0.92%	3.15%	2.45%	6.52%
	NOTRADE	0.84%		0.84%	0.15%		1.39%	1.54%
SLOW	LARGE	2.48%	2.84%	5.32%	0.27%	0.08%	0.93%	1.28%
	MEDIUM	1.39%	2.24%	3.63%	0.23%	0.18%	0.28%	0.69%
	SMALL	1.11%	1.63%	2.74%	0.17%	0.13%	0.26%	0.56%
	NOTRADE	2.01%		2.01%	0.07%		0.15%	0.22%

Panel B: Quoting activity regression by date

Speed	Inventory	Pre-opening period			First 30-minutes of the continuous session			
		Active-w/o-Trade	Active-w-Trade	Total	Active-w/o-Trade	Active-w-Trade	Non-Active	Total
FAST	LARGE	1.89***	0.61***	0.86***	1.12***	1.39***	7.02***	3.20***
	MEDIUM	1.97***	12.69***	6.94***	1.86***	10.95***	9.93***	7.61***
	SMALL (HFTs)	3.36***	13.98***	8.28***	2.41***	9.09***	30.15***	13.91***
	NOTRADE	-0.53***		-0.91***	0.23**		4.58***	2.43***
MODERATE	LARGE	1.21***	0.72***	0.58***	0.44***	0.32***	2.32***	1.06***
	MEDIUM	2.41***	10.61***	6.12***	0.85***	2.32***	1.93***	1.72***
	SMALL	2.02***	13.42***	7.33***	0.79***	3.01***	2.32***	2.06***
	NOTRADE	-0.79***		-1.17***	0.01		1.25***	0.66***
SLOW	LARGE	0.86***	1.21***	0.65***	0.13***	-0.05***	0.80***	0.32***
	MEDIUM	-0.23***	0.61***	-0.20**	0.09***	0.05***	0.15***	0.12***
	SMALL	-0.52***		-0.64***	0.04***		0.12***	0.08***
	NOTRADE	0.38***			-0.07***		0.02	
	Constant	1.63***		2.01***	0.13***			0.11***
	# obs	66,885		66,885	105,105			105,105
	Adj R^2	0.510		0.281	0.757			0.391
	Clustered St.Err.	By stock		By stock	By stock			By stock

Panel C: Quoting activity regression for HFT by date

		Pre-opening period			First 30-minutes of the continuous session			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
F/S (HFTs)	Non-Active				28.10***			
	Active-w/o-Trade	0.23				-0.50**		
	Active-w-Trade		11.39***				6.39***	
	Total			6.11***				12.09***
	Constant	4.75***	4.22***	4.18***	2.18***	3.05***	2.84***	1.93***
	# obs	66,885	66,885	66,885	105,105	105,105	105,105	105,105
	Adj R^2	0.000	0.132	0.072	0.553	0.000	0.029	0.288
	Clustered St.Err.	By stock	By stock	By stock	By stock	By stock	By stock	By stock

Table 7: Trading activity

This table shows the proportion of trading activity (buy+sell volume) carried out by 9 trader groups (the NOTRADE category is omitted as these traders do not trade during the stock-day), for the opening call auction and during the first 30 minutes of the continuous trading session. The grand total across trader groups is 200%. Traders are classified according to the scheme proposed in Table 2 using information about speed and inventory from the same day's continuous session. We also split traders into 3 categories: traders that do not participate in the pre-opening period (Non-Active), traders that participate during the pre-opening period, but do not trade at the opening call auction (Active-w/o-Trade), and traders that participate during the pre-opening period and trade at the call auction (Active-w-Trade). Average trading activity is presented per group for the period of April and May 2013, for the 97 stocks from TOPIX100. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: Average trading activity

		Opening auction	First 30 minutes of the continuous trading period			
Speed	Inventory	Active-w-Trade	Active-w/o-Trade	Active-w-Trade	Non-Active	Total
FAST	LARGE	10.37%	1.44%	3.47%	6.75%	11.66%
	MEDIUM	53.64%	3.65%	23.11%	15.80%	42.56%
	SMALL (HFTs)	36.65%	4.69%	19.79%	44.40%	68.88%
MODERATE	LARGE	9.85%	1.74%	1.56%	5.24%	8.54%
	MEDIUM	30.36%	4.03%	9.78%	7.59%	21.40%
	SMALL	25.67%	3.67%	11.27%	11.03%	25.97%
SLOW	LARGE	18.62%	1.83%	0.64%	8.51%	10.98%
	MEDIUM	9.57%	1.61%	1.24%	2.80%	5.65%
	SMALL	5.28%	1.02%	0.75%	2.59%	4.36%

Panel B: Trading activity regression by stock-date							
		Opening auction	First 30-minutes of the continuous session				
Speed	Inventory	Active-w-Trade	Active-w/o-Trade	Active-w-Trade	Non-Active	Total
FAST	LARGE	5.09***	0.69***	2.73***	6.00***	2.44***
	MEDIUM	48.36***	2.90***	22.36***	15.05***	12.73***
	SMALL (HFTs)	31.37***	3.95***	19.05***	43.65***	21.51***
MODERATE	LARGE	4.57***	1.00***	0.81***	4.49***	1.40***
	MEDIUM	25.08***	3.29***	9.03***	6.84***	5.68***
	SMALL	20.39***	2.93***	10.52***	10.28***	7.20***
SLOW	LARGE	13.34***	1.09***	-0.11***	7.76***	2.21***
	MEDIUM	4.29***	0.86***	0.49***	2.05***	0.43***
	SMALL		0.28***		1.84***	
	Constant	5.28***	0.75***			1.45***
	# obs	28,665	85,995			85,995
	Adj R^2	0.493	0.684			0.356
	Clustered St.Err.	By stock	By stock			By stock

Panel C: Trading activity regression for HFT by stock-date						
		Opening auction	First 30-minutes of the continuous session			
		(1)	(2)	(3)	(4)	(5)
F/S (HFTs)	Non-Active		38.42***			
	Active-w/o-Trade			-2.82***		
	Active-w-Trade	16.23***			12.86***	
	Total					17.50***
	Constant	20.42***	5.98***	7.51***	6.93***	5.46***
	# obs	28,665	85,995	85,995	85,995	85,995
	Adj R^2	0.057	0.419	0.002	0.047	0.241
	Clustered St.Err.	By stock	By stock	By stock	By stock	By stock

Table 8: Weighted price discovery

This table presents the summary statistics for the weighted price discovery contribution (WPC) during the pre-opening period and first 30 minutes of the continuous trading period (see equation 6 and 7). WPC is attributable to aggressive orders. Aggressive orders in the pre-opening period are defined as follows: (1) all market orders; (2) limit buy orders with a limit price greater than or equal to the prevailing best bid; (3) limit sell orders with a limit price less than or equal to the prevailing ask; (4) any orders submitted when best bid equals best ask (zero imbalance).Aggressive orders made during the continuous trading session are defined as orders that initiate the transaction (the time stamp of the transaction should be equal to the time stamp of the new order entry or the time stamp of the price revision): (1) new market orders; (2) limit-to-market orders; (3) new or revised buy (sell) limit orders with a limit price greater (smaller) than the best ask (bid) price ("Cross"); (4) new buy (sell) limit orders with a limit price equal to the best ask (bid) price ("Lock"). We divide all traders into 12 groups on a stock-day basis, as described in Table 2, using information about speed and inventory from the same day's continuous session. We also split traders into 3 categories: traders that are not active during the pre-opening period (Non-Active), traders that participate during the pre-opening period, but do not trade at the opening call auction (Active-w/o-Trade), and traders that participate during the pre-opening period and trade at the call auction (Active-w-Trade). Panel A shows the WPC during the pre-opening period, while Panel B shows the WPC excluding the first 10 minutes of the pre-opening period, for 97 stocks from the TOPIX100 during the sample period of April and May 2013. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: Average WPC

		Pre-opening period			First 30 minutes of the continuous trading period			
Speed	Inventory	Active-w/o-Trade	Active-w-Trade	Total	Active-w/o-Trade	Active-w-Trade	Non-Active	Total
FAST	LARGE	-3.01%	-4.54%	-7.55%	-0.64%	-1.87%	-5.37%	-7.88%
	MEDIUM	-3.55%	-23.11%	-26.66%	-4.35%	-10.53%	-18.25%	-33.13%
	SMALL (HFTs)	-2.56%	-13.67%	-16.23%	-1.33%	3.85%	-37.99%	-35.47%
	NOTRADE	-0.50%		-0.50%				
MODERATE	LARGE	-1.68%	-3.99%	-5.67%	1.58%	0.82%	-2.48%	-0.08%
	MEDIUM	-2.95%	-9.90%	-12.85%	1.74%	7.77%	-8.38%	1.13%
	SMALL	-2.40%	-11.21%	-13.61%	-1.63%	-3.08%	-8.05%	-12.76%
	NOTRADE	-0.42%		-0.42%				
SLOW	LARGE	-2.55%	-5.12%	-7.67%	1.04%	0.11%	-7.33%	-6.18%
	MEDIUM	-1.32%	-3.31%	-4.63%	-0.88%	0.89%	-1.76%	-1.75%
	SMALL	-0.89%	-1.12%	-2.01%	-1.23%	-0.02%	-2.61%	-3.86%
	NOTRADE	-2.20%		-2.20%				

Panel B: WPC regression by date

Speed	Inventory	Pre-opening period: Active-w/o-Trade	Pre-opening period: Active-w-Trade	Pre-opening period: Total	First 30-minutes of the continuous session: Active-w/o-Trade	First 30-minutes of the continuous session: Active-w-Trade	First 30-minutes of the continuous session: Non-Active	First 30-minutes of the continuous session: Total
FAST	LARGE	-1.89***	-3.41***	-1.57***	-0.62	-1.85	-5.35	-1.34
	MEDIUM	-2.42***	-21.98***	-11.12***	-4.33***	-10.51*	-18.23***	-9.76***
	SMALL (HFTs)	-1.44***	-12.54***	-5.91***	-1.31	3.87	-37.97***	-10.53***
	NOTRADE	0.62		1.70***				
MODERATE	LARGE	-0.56	-2.87***	-0.63*	1.60**	0.84	-2.46	1.26
	MEDIUM	-1.82***	-8.78***	-4.22***	1.76	7.79**	-8.36***	1.66
	SMALL	-1.28***	-10.08***	-4.60***	-1.61*	-3.05	-8.03***	-2.96**
	NOTRADE	0.70*		1.78***				
SLOW	LARGE	-1.42***	-4.00***	-1.63***	1.06	0.13	-7.30***	-0.77
	MEDIUM	-0.20	-2.19**	-0.11	-0.86	0.92	-1.74	0.71
	SMALL	0.24		1.20***	-1.21*		-2.59***	
	NOTRADE	-1.08**						
	Constant	-1.12***		-2.20***	-0.02			-1.29***
	# obs	861		861	1,107			1,107
	Adj R^2	0.664		0.301	0.138			0.036
	Robust St.Err.	YES		YES	YES			YES

Panel C: WPC regression for HFT by date

		Pre-opening period (1)	Pre-opening period (2)	Pre-opening period (3)	First 30-minutes of the continuous session (4)	First 30-minutes of the continuous session (5)	First 30-minutes of the continuous session (6)	First 30-minutes of the continuous session (7)
F/S (HFTs)	Non-Active				-35.60***			
	Active-w/o-Trade	2.31***				2.46**		
	Active-w-Trade		-9.35***				7.85	
	Total			-3.70***				-9.13**
	Constant	-4.87***	-4.32***	-4.41***	-2.39***	-3.79***	-3.99***	-2.69***
	# obs	861	861	861	1,107	1,107	1,107	1,107
	Adj R^2	0.005	0.092	0.027	0.102	-0.000	0.004	0.018
	Robust St.Err.	YES	YES	YES	YES	YES	YES	YES

Table 9: Imbalance between liquidity demand and supply

This table shows the imbalance between liquidity consumption and supply among 9 trader groups (the NOTRADE category is omitted as these traders do not trade during the stock-day), for the opening call auction and during the first 30 minutes of the continuous trading session. Traders are classified according to the scheme proposed in Table 2 using information about speed and inventory from the same day's continuous session. We also split traders into 3 categories: traders that do not participate in the pre-opening period (Non-Active), traders that participate during the pre-opening period, but do not trade at the opening call auction (Active-w/o-Trade), and traders that participate during the pre-opening period and trade at the call auction (Active-w-Trade). In the case of the opening call auction, trading activities are considered to provide liquidity if traders trade in the opposite direction to the price movement and trading activities are considered to consume liquidity if traders trade in the direction of the price movement. In the case of the continuous session, trading activities are considered to consume liquidity if the order initiates the transaction, and supply liquidity otherwise. Orders that initiate the transaction (the time stamp of the transaction should be equal to the time stamp of the new order entry or the time stamp of the price revision) satisfy one of four conditions: (1) new market orders; (2) limit-to-market orders; (3) new or revised buy (sell) limit orders with a limit price greater (smaller) than the best ask (bid) price ("Cross"); (4) new buy (sell) limit orders with a limit price equal to the best ask (bid) price ("Lock"). We report the liquidity-demanding and liquidity-supplying trading volume relative to the total trading volume during the pre-opening call averaged across stock-days. For the imbalance between liquidity consumption and liquidity supply, we also report the significance levels of a *t*-test for whether the imbalance is significantly different from 0. ***, **, and * indicate 1%, 5%, and 10% significance levels, respectively. Liquidity consumption and liquidity provision are presented per group for the period of April and May 2013, for the 97 stocks from TOPIX100. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: Average liquidity imbalance

		Opening auction	First 30 minutes of the continuous trading period			
Speed	Inventory	Active-w-Trade	Active-w/o-Trade	Active-w-Trade	Non-Active	Total
FAST	LARGE	1.53%	-0.65%	-1.37%	-1.58%	-3.60%
	MEDIUM	3.13%	-1.23%	-7.71%	0.41%	-8.53%
	SMALL (HFTs)	-0.99%	-2.43%	-4.90%	14.73%	7.40%
MODERATE	LARGE	0.02%	-0.67%	-0.36%	-0.67%	-1.70%
	MEDIUM	-2.55%	-1.34%	-2.22%	2.35%	-1.21%
	SMALL	-1.22%	-1.12%	-2.05%	5.42%	2.25%
SLOW	LARGE	0.42%	-0.81%	-0.14%	4.84%	3.89%
	MEDIUM	-0.17%	-0.55%	-0.32%	1.65%	0.78%
	SMALL	-0.18%	-0.36%	-0.13%	1.18%	0.69%

Panel B: Liquidity imbalance regression by stock-date

		Opening auction	First 30-minutes of the continuous session			
Speed	Inventory	Active-w-Trade	Active-w/o-Trade	Active-w-Trade	Non-Active	Total
FAST	LARGE	1.70***	-0.52***	-1.24***	-1.45***	-1.43***
	MEDIUM	3.31***	-1.10***	-7.58***	0.53***	-3.07***
	SMALL (HFTs)	-0.81***	-2.31***	-4.78***	14.86***	2.23***
MODERATE	LARGE	0.19	-0.54***	-0.23***	-0.54***	-0.79***
	MEDIUM	-2.37***	-1.21***	-2.09***	2.48***	-0.63***
	SMALL	-1.04***	-0.99***	-1.93***	5.55***	0.52***
SLOW	LARGE	0.60**	-0.68***	-0.02	4.97***	1.07***
	MEDIUM	0.01	-0.42***	-0.19***	1.77***	0.03
	SMALL		-0.24***		1.31***	
	Constant	-0.18**		-0.13***		0.23***
	# obs	26,685		85,995		85,995
	Adj R^2	0.016		0.503		0.071
	Clustered St.Err.	By stock		By stock		By stock

Panel C: Liquidity imbalance regression for HFT by stock-date

		Opening auction	First 30-minutes of the continuous session			
		(1)	(2)	(3)	(4)	(5)
F/S (HFTs)	Non-Active		15.30***			
	Active-w/o-Trade			-2.53***		
	Active-w-Trade	-1.11***			-5.09***	
	Total					2.77***
	Constant	0.12***	-0.57***	0.09***	0.19***	-0.31***
	# obs	26,685	85,995	85,995	85,995	85,995
	Adj R^2	0.001	0.291	0.008	0.032	0.026
	Clustered St.Err.	By stock	By stock	By stock	By stock	By stock

Table 10: Summary of Results

This table summarizes the results on HFTs' behavior. HFTs include all FAST/SMALL traders, Active-w-Trade HFTs are those who participate in the pre-opening, call auction and continuous periods, and Non-Active HFTs are those who do not participate in either the pre-opening or call auction, but do participate in the continuous period.

	Pre-opening period	Call auction	Continuous trading period
Panel A: HFTs			
Presence ratio	26.96%	17.47%	83.65%
Quoting activity ratio	20.58%	-	42.05%
Trading activity ratio	-	36.65%	68.88%
WPC	-16.23%	-	-35.47%
Liquidity	-	-0.99%	7.40%
Panel B: Active-w-Trade HFTs			
Presence ratio	17.47%	17.47%	16.16%
Quoting activity ratio	15.60%	-	9.23%
Trading activity ratio	-	36.65%	19.79%
WPC	-13.67%	-	3.85%
Liquidity	-	-0.99%	-4.90%
Panel C: Non-Active HFTs			
Presence ratio	-	-	58.91%
Quoting activity ratio	-	-	30.28%
Trading activity ratio	-	-	44.40%
WPC	-	-	-37.99%
Liquidity	-	-	14.73%

Table 11: Classification scheme proposed by Kirilenko et al. (2015)

This table shows summary statistics for the classification of traders based on Kirilenko, Kyle, Samadi, and Tuzun (2015). In this case, we divide traders into two groups (HFTs and non-HFTs) using information from the continuous trading session of the same day. A trader is defined as an HFT in a particular stock if and only if, on at least 50% of the active days, she satisfies the following three criteria: (1) Her end-of-day inventory is no greater than 10% of her trading volume for that stock on that day. (2) Her inventory at the end of each minute is no greater than 15% of her trading volume for that stock on that day. (3) Her trading volume in that stock on that day is in the top quartile of total trading volume for all traders in that stock on that day. In addition, we require HFTs to be active in that stock for at least 10 of the days in our sample period. Panel A shows the median number of traders, the median of the 1%, 10%, and 50% percentiles of the trader-stock-day speed distribution, and the median minimum speed (used for the classification of traders). Panel B shows the median number of traders, median end-of-day inventory (used for the classification of traders), median average intraday inventory and its volatility, and the median number of times during the day that the inventory crosses zero, per trader-stock-day. These characteristics are presented per group for the period of April 2013, for the 97 stocks from TOPIX100. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: Speed of traders

		Median of			
	# of traders	Min speed	P1 speed	P5 speed	P10 speed
HFT	5	0.002	0.002	0.008	0.020
Non-HFT	905	1.917	1.973	4.102	9.165

Panel B: Inventory of traders

		Median of			
	# of traders	End-of-day inventory	Intraday inventory	Volatility of intraday inventory	# of zero crossings
HFT	5	0.95%	2.91%	2.27%	14
Non-HFT	905	100.00%	53.89%	21.60%	0

Table 12: Comparison of classifications

This table shows the summary comparison of the classification of traders proposed in this paper versus that based on Kirilenko, Kyle, Samadi, and Tuzun (2015) for 97 stocks from TOPIX100 during April 2013. The classification proposed in this paper splits traders into 12 groups on a stock-day basis, as reported in Table 2. The classification of traders based on Kirilenko, Kyle, Samadi, and Tuzun (2015) splits traders into two groups (HFTs and non-HFTs). A trader is defined as an HFT in a particular stock if and only if, on at least 50% of the active days, she satisfies the following three criteria: (1) Her end-of-day inventory is no greater than 10% of her trading volume for that stock on that day. (2) Her inventory at the end of each minute is no greater than 15% of her trading volume for that stock on that day. (3) Her trading volume in that stock on that day is in the top quartile of total trading volume for all traders in that stock on that day. In addition, we require HFTs to be active in that stock for at least 10 of the days in our sample period. We report the number of trader-stock-days in each group. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Speed	Inventory	HFT	Non-HFT
FAST	LARGE	10	124,237
	MEDIUM	66	134,622
	SMALL (HFTs)	7,701	130,443
	NOTRADE	0	68,333
MODERATE	LARGE	1	153,047
	MEDIUM	26	118,539
	SMALL	1,380	110,629
	NOTRADE	5	79,764
SLOW	LARGE	12	336,510
	MEDIUM	0	64,533
	SMALL	15	53,209
	NOTRADE	4	66,724

Table 13: Quoting activity regression in different regimes

		Pre-opening period		First 30-minutes of the continuous session		
Speed	Inventory	Active-w/o-Trade	Active-w-Trade	Active-w/o-Trade	Active-w-Trade	Non-Active
Panel A: Quoting activity regression by date: interaction effects with systematic volatility						
FAST	LARGE	-0.36***	0.53***	-0.13	-0.09	0.14
	MEDIUM	0.22	2.05***	0.14	0.61**	0.24
	SMALL (HFTs)	0.96***	1.87***	0.33***	1.11***	-0.32
	NOTRADE	-0.03		-0.01		-0.54***
MODERATE	LARGE	0.20*	0.55***	-0.03	0.03	-0.08
	MEDIUM	0.53***	0.21	0.04	-0.05	0.08*
	SMALL	0.45***	0.32	-0.01	-0.10	0.26***
	NOTRADE	0.33***		0.01		-0.12**
SLOW	LARGE	-0.09	0.70***	-0.02*	0.03***	-0.04
	MEDIUM	0.18**	-0.02	0.00	-0.01	0.02*
	SMALL	0.19**		-0.01		0.02
	NOTRADE	-0.02		0.02**		0.01
Panel B: Quoting activity regression by stock: interaction effects with idiosyncratic volatility						
FAST	LARGE	0.10	0.41***	0.02	-0.16	-0.33
	MEDIUM	0.71***	3.06***	0.30*	0.65	-0.63***
	SMALL (HFTs)	1.60***	3.85***	0.62***	1.88***	0.81
	NOTRADE	0.17*		0.06		-0.50*
MODERATE	LARGE	0.26*	0.43***	-0.01	0.00	-0.44***
	MEDIUM	0.33**	-0.04	0.00	-0.05	0.01
	SMALL	0.40***	-0.17	0.00	-0.05	0.32***
	NOTRADE	0.38***		0.03**		-0.35***
SLOW	LARGE	-0.36***	0.17	-0.06***	0.03***	-0.17***
	MEDIUM	0.10	-0.20*	-0.02	-0.01	0.01
	SMALL	0.03		-0.01		-0.00
	NOTRADE	-0.29**		0.02**		0.01

Table 14: Trading activity regression in different regimes

Speed	Inventory	Opening auction	First 30-minutes of the continuous session		
		Active-w-Trade	Active-w/o-Trade	Active-w-Trade	Non-Active
Panel A: Trading activity regression by stock-date: interaction effects with systematic volatility					
FAST	LARGE	0.21	-0.18	-0.20	-0.22
	MEDIUM	2.09**	0.05	1.51***	0.24
	SMALL (HFTs)	3.87***	0.34	2.24***	-0.46
MODERATE	LARGE	-2.94***	0.06	0.08	0.20
	MEDIUM	1.43*	0.13	0.21	0.48**
	SMALL	1.70***	0.09	0.04	0.06
SLOW	LARGE	1.78***	-0.19*	0.09	-0.49**
	MEDIUM	-0.33	-0.12	-0.09	0.09
	SMALL		-0.03		-0.29**
Panel B: Trading activity regression by stock-date: interaction effects with idiosyncratic volatility					
FAST	LARGE	-0.53	-0.10	-0.32	-0.58**
	MEDIUM	5.63***	0.66***	3.07***	-0.74*
	SMALL (HFTs)	8.02***	1.00***	4.88***	-0.58
MODERATE	LARGE	-0.48	0.00	0.05	-0.56***
	MEDIUM	-0.59	0.27	0.01	0.00
	SMALL	1.29*	0.34**	0.07	1.48***
SLOW	LARGE	-3.37***	-0.27**	0.03	-1.94***
	MEDIUM	-1.46***	-0.23**	-0.07	-0.26**
	SMALL		-0.10		-0.63***

Table 15: WPC regression in different regimes

		Pre-opening period		First 30-minutes of the continuous session		
Speed	Inventory	Active-w/o-Trade	Active-w-Trade	Active-w/o-Trade	Active-w-Trade	Non-Active
Panel A: WPC regression by date: interaction effects with systematic volatility						
FAST	LARGE	0.01	0.00	-0.02	0.03	-0.05
	MEDIUM	0.01	0.02	-0.05	0.24*	-0.10
	SMALL (HFTs)	0.02	0.00	0.02	0.13	-0.29*
	NOTRADE	0.02				
MODERATE	LARGE	0.01	-0.01	-0.00	0.01	-0.07*
	MEDIUM	-0.01	0.01	0.01	0.10*	-0.06
	SMALL	0.01	-0.02	0.00	-0.02	-0.05
	NOTRADE	0.01				
SLOW	LARGE	0.02	-0.04	-0.01	-0.01	-0.03
	MEDIUM	0.02	-0.00	0.01	-0.00	-0.00
	SMALL	0.01		0.01		-0.01
	NOTRADE	0.01				
Panel B: WPC regression by stock: interaction effects with idiosincratic volatility						
FAST	LARGE	-0.01	-0.01	0.02	0.04	-0.03
	MEDIUM	-0.00	-0.10**	-0.04	-0.06	-0.24**
	SMALL (HFTs)	0.00	-0.11***	0.04	0.19	-0.12
	NOTRADE	0.00				
MODERATE	LARGE	-0.01	-0.03	0.02	0.03	-0.05
	MEDIUM	-0.02	0.04	0.02	0.14*	-0.02
	SMALL	-0.00	-0.03	0.02	-0.19*	0.00
	NOTRADE	-0.00				
SLOW	LARGE	-0.01	-0.01	0.05***	0.05***	0.02
	MEDIUM	-0.00	-0.03	0.04*	0.02	0.05*
	SMALL	-0.01		0.04*		0.04*
	NOTRADE	-0.00				

Table 16: Liquidity imbalance regression in different regimes

		Opening auction	First 30-minutes of the continuous session		
Speed	Inventory	Active-w-Trade	Active-w/o-Trade	Active-w-Trade	Non-Active
Panel A: Liquidity imbalance regression by stock-date: interaction effects with systematic volaility					
FAST	LARGE	-0.58	0.24***	0.23	0.21
	MEDIUM	0.45	0.01	0.47	0.08
	SMALL (HFTs)	0.28	-0.02	-0.07	-1.18***
MODERATE	LARGE	0.33	0.08	0.04	0.30**
	MEDIUM	0.86	0.01	-0.15	0.19
	SMALL	-0.13	0.03		-0.26
				0.18	
SLOW	LARGE	0.58	0.08	0.06	-0.68***
	MEDIUM	-0.31	0.06	0.08	-0.12
	SMALL		0.04		-0.31**
Panel B: Liquidity imbalance regression by stock-date: interaction effects with idiosyncratic volatility					
FAST	LARGE	-1.47***	0.22**	0.39**	0.11
	MEDIUM	-1.88	-0.28**	0.69*	-0.45*
	SMALL (HFTs)	-0.46	-0.54***	-0.11	-2.21***
MODERATE	LARGE	-1.08	-0.01	0.02	0.23*
	MEDIUM	0.79	-0.08	-0.10	0.05
	SMALL	-0.37	-0.01		1.05***
				0.51***	
SLOW	LARGE	0.04	0.09	-0.06	-1.04***
	MEDIUM	-0.18	0.12	0.09	-0.24*
	SMALL		0.10		-0.22

Table 17: Contribution to weighted price discovery by order type during the pre-opening period

This table presents the summary statistics for the weighted price discovery contribution (WPC), the percentage amount by which an incoming aggressive order moves the prevailing mid-quote closer to the opening price divided by the accumulated price discovery contribution during the pre-opening period, as defined in equation (3). Aggressive orders are defined as follows: (1) all market orders; (2) limit buy orders with a limit price greater than or equal to the prevailing best bid; (3) limit sell orders with a limit price less than or equal to the prevailing ask; (4) any orders submitted when best bid equals best ask (zero imbalance). We distinguish between WPC for each of the 9 different types of orders. We divide all traders into 12 groups on a stock-day basis, as described in Table 2, using information about speed and inventory from the same day's continuous session. We also split traders into 2 categories: traders that participate during the pre-opening period, but do not trade at the opening call auction (Active-w/o-Trade), and traders that participate during the pre-opening period and trade at the call auction (Active-w-Trade). Panel A presents the WPC during the pre-opening period, while Panel B presents the WPC excluding the first 10 minutes of the pre-opening period, for 97 stocks from the TOPIX100 during the sample period of April and May 2013. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

		Total	Limit orders				Market orders				Zero imbalance
Speed	Inventory		New	Qty revision	Cancellation	Price revision	New	Qty revision	Cancellation	Price revision	
Panel A: Active-w/o-Trade											
FAST	LARGE	-3.01%	-4.39%	0.53%	2.06%	-1.12%	-0.03%	0.03%	-0.11%	0.00%	0.00%
	MEDIUM	-3.55%	-2.43%	0.34%	-0.25%	-0.88%	-0.12%	0.00%	-0.21%	0.00%	0.00%
	SMALL (HFTs)	-2.56%	-2.99%	0.21%	1.01%	-0.62%	-0.16%	0.00%	-0.01%	0.01%	0.00%
	NOTRADE	-0.50%	-0.85%	0.14%	0.32%	-0.05%	-0.04%	-0.01%	-0.01%	0.00%	0.00%
MODERATE	LARGE	-1.68%	-1.85%	0.08%	0.33%	-0.11%	0.02%	0.02%	-0.19%	0.01%	0.00%
	MEDIUM	-2.95%	-2.59%	0.04%	0.19%	-0.05%	-0.16%	0.00%	-0.32%	-0.04%	0.00%
	SMALL	-2.40%	-2.11%	0.04%	0.29%	-0.22%	-0.25%	-0.01%	-0.09%	-0.06%	0.00%
	NOTRADE	-0.42%	-0.61%	0.06%	0.15%	-0.01%	-0.02%	0.00%	0.01%	0.00%	0.00%
SLOW	LARGE	-2.55%	-2.69%	0.09%	0.15%	0.13%	-0.12%	0.00%	-0.10%	0.00%	0.00%
	MEDIUM	-1.32%	-1.25%	0.00%	0.02%	-0.03%	0.04%	0.00%	-0.11%	-0.01%	0.00%
	SMALL	-0.89%	-0.94%	0.00%	0.15%	0.00%	0.03%	-0.01%	-0.13%	0.02%	0.00%
	NOTRADE	-2.20%	-2.46%	0.18%	0.20%	-0.10%	0.01%	0.00%	-0.04%	0.00%	0.00%
Panel B: Active-w-Trade											
FAST	LARGE	-4.54%	-2.15%	-0.05%	0.16%	-0.02%	-3.10%	0.17%	0.43%	0.02%	0.00%
	MEDIUM	-23.11%	-9.92%	0.01%	-0.56%	-0.08%	-11.30%	-0.29%	-1.11%	0.14%	0.01%
	SMALL (HFTs)	-13.67%	-7.15%	0.14%	0.28%	-0.40%	-5.85%	-0.08%	-0.52%	-0.09%	0.00%
MODERATE	LARGE	-3.99%	-1.49%	-0.03%	-0.24%	-0.11%	-1.96%	-0.05%	-0.09%	-0.03%	0.00%
	MEDIUM	-9.90%	-4.51%	-0.05%	-0.02%	0.03%	-5.14%	-0.14%	-0.11%	0.03%	0.00%
	SMALL	-11.21%	-6.36%	-0.01%	-0.36%	-0.04%	-3.87%	-0.03%	-0.41%	-0.13%	0.00%
SLOW	LARGE	-5.12%	-1.94%	-0.01%	-0.08%	-0.03%	-3.07%	-0.05%	0.04%	0.02%	-0.01%
	MEDIUM	-3.31%	-0.82%	0.00%	-0.22%	-0.13%	-2.17%	0.00%	0.03%	0.00%	0.00%
	SMALL	-1.12%	-0.61%	0.02%	0.07%	0.06%	-0.36%	0.01%	-0.32%	0.01%	0.00%

Table 18: Aggressive orders across stocks

This table provides summary statistics for the aggressive orders across stocks. We divide all traders into 12 groups on a stock-day basis, as described in Table 2, using information about speed and inventory from the same day's continuous trading period. For each stock, we compute the proportion of aggressive orders (orders with the potential to impact the prevailing quotes) submitted by each group of traders relative to the total number of aggressive orders for a particular stock during the pre-opening period, excluding the first 10 minutes for April and May 2013 across 97 stocks from TOPIX100. Aggressive orders are defined as follows: (1) all market orders; (2) limit buy orders with a limit price greater than or equal to the prevailing best bid; (3) limit sell orders with a limit price less than or equal to the prevailing ask; (4) any orders submitted when best bid equals best ask. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Speed	Inventory	MIN	P5	P25	P50	P75	P95	MAX
FAST	LARGE	1.57%	2.54%	4.03%	6.40%	8.49%	12.32%	14.85%
	MEDIUM	12.27%	14.95%	19.20%	20.83%	23.62%	27.25%	35.76%
	SMALL (HFTs)	5.80%	9.94%	14.89%	19.09%	28.74%	43.73%	58.65%
	NOTRADE	0.02%	0.04%	0.23%	0.56%	0.99%	2.30%	3.29%
MODERATE	LARGE	1.33%	2.15%	3.17%	4.42%	5.75%	7.53%	8.49%
	MEDIUM	4.53%	7.82%	11.80%	14.02%	16.67%	19.40%	22.45%
	SMALL	5.82%	9.38%	12.75%	16.03%	20.08%	24.64%	27.62%
	NOTRADE	0.01%	0.04%	0.15%	0.29%	0.53%	1.26%	2.67%
SLOW	LARGE	0.69%	1.60%	3.26%	5.39%	8.07%	11.63%	17.27%
	MEDIUM	0.34%	0.89%	1.80%	3.53%	4.55%	6.63%	8.30%
	SMALL	0.30%	0.47%	1.26%	2.64%	3.61%	4.63%	6.10%
	NOTRADE	0.07%	0.17%	0.48%	1.07%	1.62%	3.30%	5.76%

Figure 1: Trading schedule of the TSE

This figure shows the trading schedule of the Tokyo Stock Exchange on a typical trading day. The two separate sessions (morning and afternoon) start, respectively, at 9 am and 12.30 pm. A call auction precedes the subsequent continuous period.

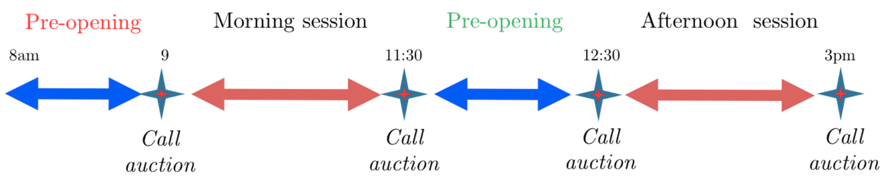


Figure 2: Graphical representation of usage of virtual servers by traders

This graph displays the relation between the number of virtual servers and the number of trading desks, during the period of April and May 2013, on the Tokyo Stock Exchange, for 1,702 stocks. The total number of virtual servers is 5,580 (all the dots in the figure), while the number of trading desks using one or more virtual servers is 3,021 (the colored groups in the figure). Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

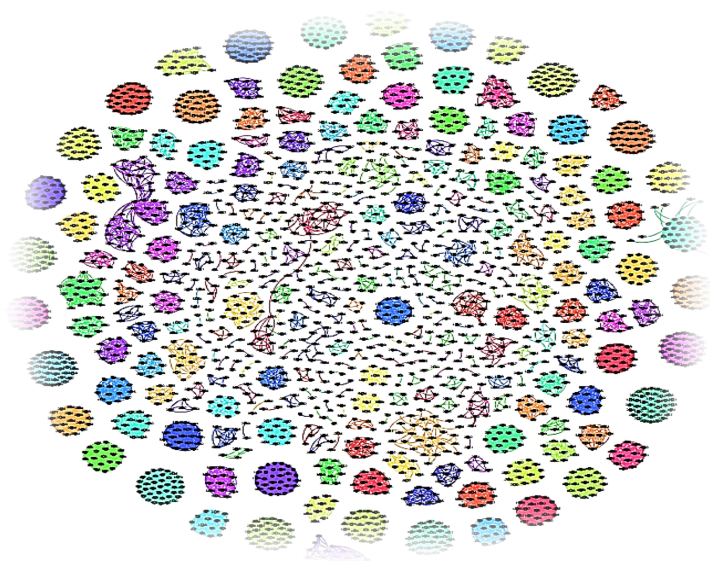


Figure 3: Activity of FAST/SMALL traders (HFTs) during the trading day

This diagram shows activity of FAST/SMALL traders (HFTs) classified during the continuous session according to the scheme proposed in Table 2 using information about speed and inventory from the same day's continuous session. We report the median proportions of FAST/SMALL traders (HFTs) based on their decision to participate or not during the pre-opening period, opening call, and first 30 minutes of the continuous trading session. These proportions are presented for the period of April and May 2013, for the 97 stocks from TOPIX100. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

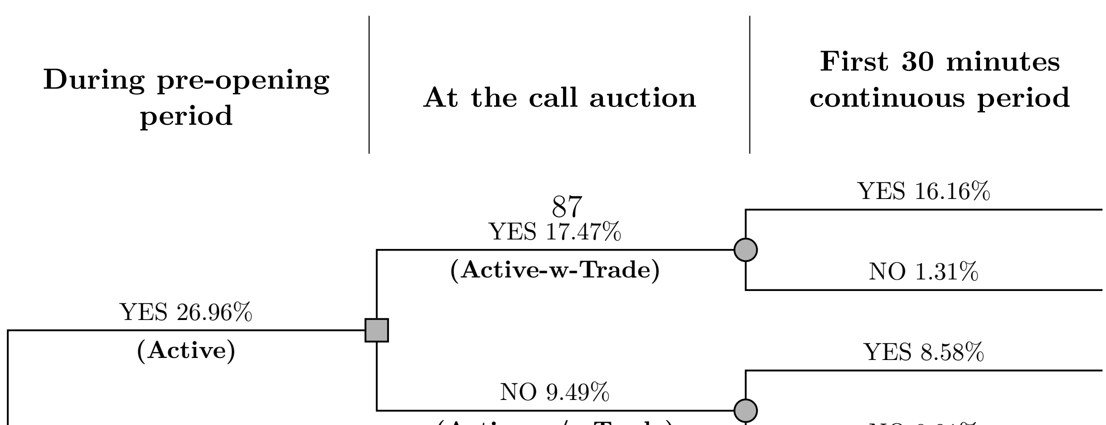


Figure 4: Deviation from the opening price

This figure shows the deviation of the pre-opening mid-quote from the opening price, computed for each second of the pre-opening period (8:10:00.000 - 8:59:59.999) for 97 stocks from the TOPIX100 during the sample period of April and May 2013. The deviation is defined as the percentage difference between the mid-quote, $M_{t,k}$, at time t on day k, and the opening price, O_k, on day k, as defined in equation (6). The deviation is computed per second per day per stock and then medians are calculated for each second. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

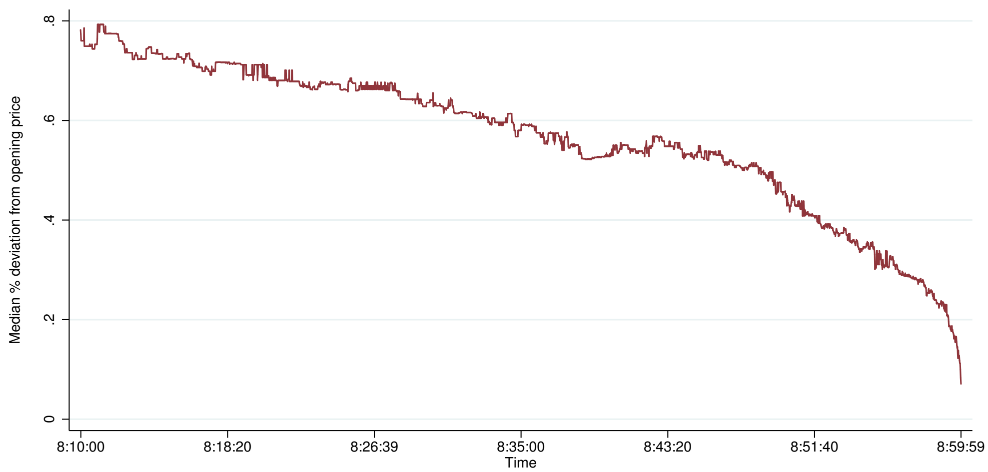
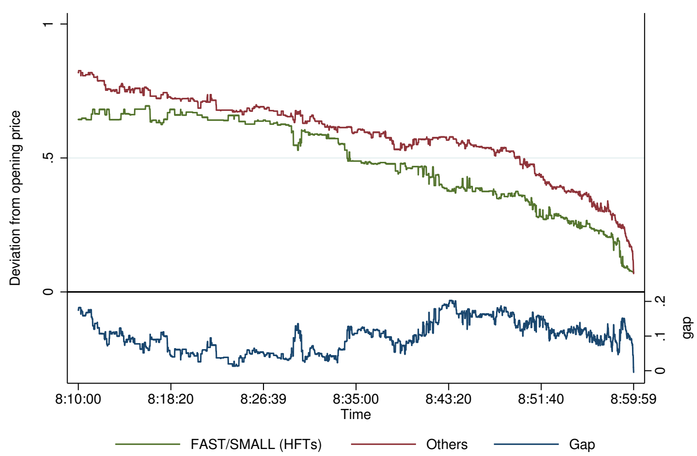


Figure 5: Comparison of the deviation from the opening price between stocks for which high-speed traders have different levels of participation

This figure shows, for two groups of stocks, the percentage deviation of the pre-opening mid-quote from the opening price, computed for each second of the pre-opening period (8:10:00.000 - 8:59:59.999), for 97 stocks from the TOPIX100, during the sample period of April and May 2013. We split stocks into two groups: the first group includes stocks for which aggressive activity by FAST&MODERATE/SMALL&MEDIUM traders passes a threshold of 30% (18 stocks). The second group includes all other stocks (79 stocks). Panel A displays the deviation for the entire pre-opening period for the two groups of stocks, while Panel B displays deviations for the last 10 minutes of the pre-opening period. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange, while quotes and trade data are obtained from the Thomson-Reuters Tick History Database.

Panel A - Deviation for the entire pre-opening period for the two groups of traders



Panel B - Deviation for the last 10 minutes for the two groups of traders

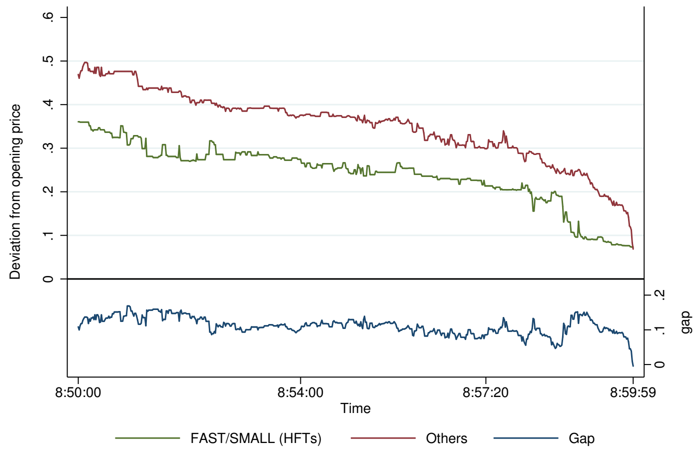


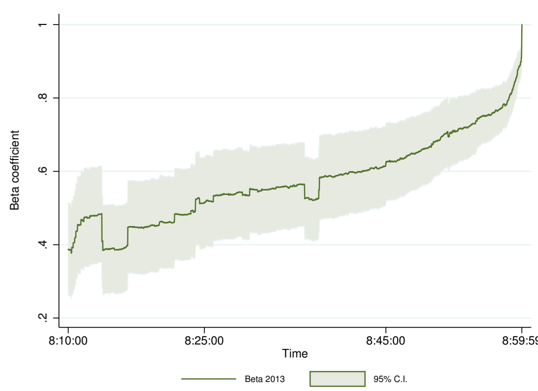
Figure 6: Tests of unbiasedness regressions of the pre-opening mid-quotes

Using mid-quotes, we estimate equation (16):

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

where v is the opening price, P_t is the pre-opening mid-quote, and $E(\nu|I_0)$ is the previous day's closing price, estimated every second during the pre-opening period (8:10:00.000 - 8:59:59.999) for each of the 97 stocks from the TOPIX100. The figures show the averages of the β coefficients (Panel A) and the t-statistics under the null hypothesis that β is equal to 1 (Panel B). The tick-by-tick data, time stamped to the millisecond, are obtained from the Thomson-Reuters Tick History Database.

Panel A - Beta coefficient



Panel B - T-statistic

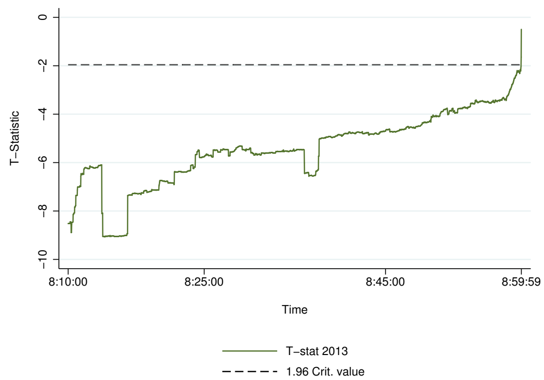


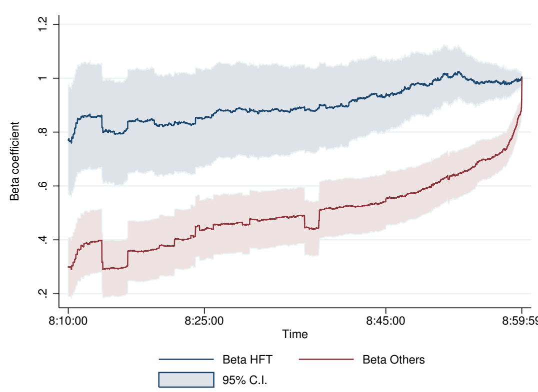
Figure 7: Comparison of the test of unbiasedness regressions between stocks with different levels of high-speed traders participation in the last 20 seconds

Using mid-quotes, at each 100-millisecond interval, we estimate equation 16:

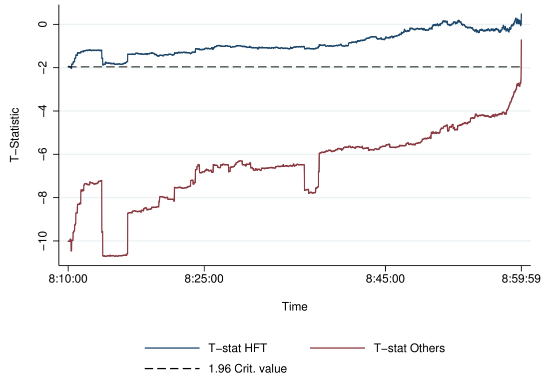
$$\nu - E\left(\nu|I_0\right) = \alpha_t + \beta_t\left[P_t - E\left(\nu|I_0\right)\right] + Z_t$$

where v is the opening price, P_t is the pre-opening mid-quote, and $E(\nu|I_0)$ is the previous day's closing price, computed at each second of the pre-opening period (8:10:00.000 - 8:59:59.999), for each of the 97 stocks from the TOPIX100 during the sample period of April-May 2013. We split stocks into two groups: the first group includes stocks for which aggressive activity of FAST&MODERATE/SMALL&MEDIUM traders passes a threshold of 30% (18 stocks). The second group includes all other stocks (79 stocks). The averages of the β coefficients are shown in Panel A. Panel B shows the t-statistics under the null hypothesis that β is equal to 1. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange, while quotes and trade data are obtained from the Thomson-Reuters Tick History Database.

Panel A - Beta coefficient for the two groups



Panel B - T-statistic for the two groups


T−Statistic
0
−2
−4
−6
−8
−10
8:10:00
8:25:00
8:45:00
8:59:59
Time
T−stat HFT
T−stat Others
1.96 Crit. value

APPENDIX

A. Best quotes during the pre-opening period

A.1. Mid-quote calculation

The pre-opening quotes consist of bid and ask prices and their associated quantities. In the case of the TSE, the best bid and ask prices are determined differently during the pre-opening period than during the continuous trading period. During the continuous trading period, the best bid is the highest available bid price, and the best ask is the lowest available ask price. This means that the bid and ask schedules do not intersect as the submission of a buy order with a limit price greater than the best available ask price will cause the immediate execution of that order and it will not join the queue in the limit order book.

On the contrary, during the pre-opening period no execution is allowed before the opening auction, when all orders are executed at a single price. Therefore, the best bid and ask prices reported during the pre-opening period are the respective prices at which the bid (demand) and ask (supply) schedules intersect. The best ask is identified as the smallest ask price at which the cumulative depth of the ask schedule is greater than the cumulative depth of the bid schedule. The best bid is identified as the largest bid price at which the cumulative depth of the bid schedule is greater than the cumulative depth of the ask schedule.[25]

This appendix illustrates how the best bid price and the best ask price are determined during the pre-opening period. First of all, the TSE computes the cumulative amount of eligible buy and sell orders at each price (depth). Usually, more buy orders are accumulated

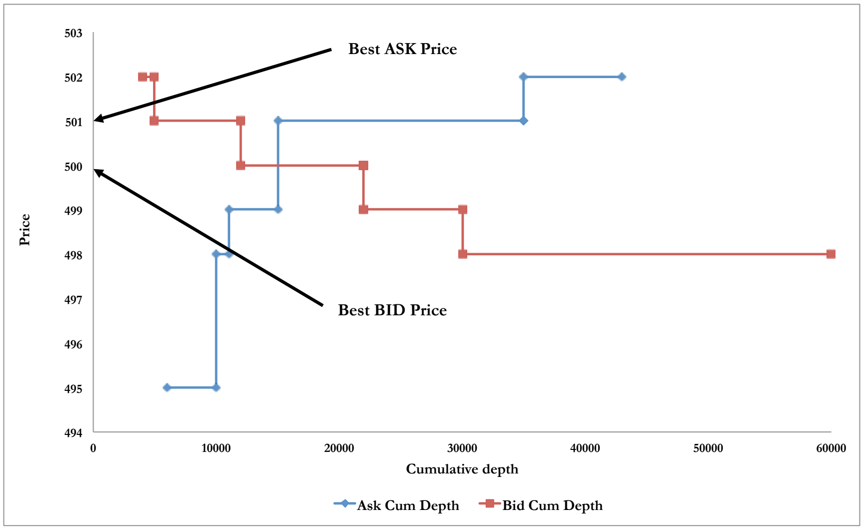
[25] We use two different sources for the best bid and ask prices in the pre-opening period. First, we use the TRTH data with a millisecond time stamp. However, there is a time stamp mismatch between the order flow data provided by the TSE and the TRTH best quotes time stamp. Therefore, for the analysis that requires exact matching between these two databases, we construct the best bid-offer ourselves on a tick-by-tick basis. This is a non-trivial task due to the multiple rules employed by the TSE. We verify the sequence of our best bid and ask estimates using the TRTH database, and ensure that our estimates are consistent with the TRTH best bid and ask prices time stamped without a time delay.

around lower prices and more sell orders are accumulated around higher prices so that there is a point at which the situation of "cumulative buy orders" being greater than "cumulative sells" turns into "cumulative buys" being less than or equal to "cumulative sells". The best bid is the highest bid price at which the cumulative bid depth is greater than the cumulative ask depth and the best ask is the lowest ask price at which the cumulative ask depth is greater than the cumulative bid depth.

Therefore, the best bid and ask prices reported during the pre-opening period are the respective prices at which the bid (demand) and ask (supply) schedules (two step functions with cumulative volume on the X-axis and price on the Y-axis) intersect. Either the best ask or the best bid price is the opening price, as a result of the single price auction explained in Section 4. In the pre-opening period, however, the cumulative amounts of buy and sell orders can be the same, particularly at the beginning of the pre-opening period when just a few orders have been entered. In these special situations, the TSE has another rule for determining the best bid and ask in the pre-opening period, which is based on yesterday's closing price and the upper or lower limit on the price of a stock. Refer to TSE (2015) for details.

Figure A.1: Determination of best bid and ask prices during the pre-opening period

This figure shows a hypothetical example of how the best bid price and the best ask price are determined during the pre-opening period. We plot bid (demand) and ask (supply) schedules with cumulative volume on the X-axis and price on the Y-axis. The blue line represents the ask schedule, while the red line represents the bid schedule. The best bid is the highest bid price at which the cumulative bid depth is greater than the cumulative ask depth. The best ask is the lowest ask price at which the cumulative ask depth is greater than the cumulative bid depth.



Best ASK Price
Best BID Price
Price
503
502
501
500
499
498
497
496
495
494
0
10000
20000
30000
40000
50000
60000
Cumulative depth
Ask Cum Depth
Bid Cum Depth

B. Speed model estimation

Due to the limitation on the number of quotes per second per server, the coverage of stocks and intensity of quotes of a trader determine the size of their operation. Our novel data on server IDs allow us to estimate the relation between speed, server configuration, and quote intensity, using the following equation:

$$\begin{aligned} Speed_{j,k,l} = a + b\ln(Quote_{j,k,l}) + \\ c\ln(Nstock_{k,l}/Nserver_l) + d\ln(MaxQuote_{k,l}) + \epsilon_{j,k,l} \end{aligned} \tag{C.1}$$

$Speed_{j,k,l}$ is the speed measure for stock j, day k, and trader l. $Quote_{j,k,l}$ is the number of quotes for stock j, day k, and trader l. $Nstock_{k,l}$ is the number of stocks traded on day k by trader l. $Nserver_l$ is the number of servers used by trader l (a fixed number during our sample period). $MaxQuote_{k,l}$ is the maximum number of quotes per second sent by trader l on day k.

The daily number of stocks per server indicates the trader's speed requirement. The number of quotes is used by other HFT studies to identify HFTs that engage in market making. The maximum number of quotes per second is another aspect of trading style; for example, an index arbitrager might execute a basket of 225 Nikkei Index constituents simultaneously. Our empirical measure of speed is limited by the time stamp unit of one millisecond, meaning that the distribution of observed elapsed time is clustered at one millisecond. Taking into account the censored nature of the dependent variable, we use a Tobit model to estimate equation (C.1).

Table B.1 shows a strong relation between the number of stocks per server, the total number of quotes, and the maximum number of quotes per second. The smaller the number of stocks per server, and the larger the number of quotes (maximum number of quotes per second), the lower is the speed. This result suggests that speed-based classification is equivalent to classification based on the total number of quotes.

Table B.1: Speed model estimation

Estimation, using Tobit regression, of the model in equation (C.1). $Speed_{j,k,l}$ is the speed measure for stock j, day k, and trader l. $Quote_{j,k,l}$ is the number of quotes for stock j, day k, and trader l. $Nstock_{k,l}$ is the number of stocks traded on day k by trader l. $Nserver_l$ is the number of servers used by trader l (a fixed number during our sample period). $MaxQuote_{k,l}$ is the maximum number of quotes per second sent by trader l on day k. Our sample consists of 97 stocks from TOPIX100 during April and May 2013. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Dependent variable: $Speed_{j,t,k}$		
	Coef	t-stat
$Constant$	5.44	571.65
$\ln(Quote_{j,t,k})$	-2.08	-1555.60
$\ln(Nstock_{t,k}/Nserver_k)$	0.41	263.60
$\ln(MaxQuote_{t,k})$	-1.35	-489.31
Left-censored obs	73,011	
Right-censored obs	0	
Uncensored obs	3,120,836	
Total obs	3,193,847	

C. Aggressive orders

Among the orders submitted during the pre-opening period, we can identify those orders with the potential to impact the prevailing quotes. We call them "aggressive orders" (as in Biais, Hillion, and Spatt (1995), Ranaldo (2004), Duong, Kalev, and Krishnamurti (2009), and Yamamoto (2011)). The TSE uses unique rules for determining the best pre-opening bid and ask quotes. These rules are different from those applied in the continuous trading period and are briefly explained in Section 4. There are four cases of orders that we categorize as aggressive: first, all market orders; second, a limit buy order with a limit price greater than or equal to the prevailing best bid; third, a limit sell order with a limit price less than or equal to the prevailing ask; fourth, any orders submitted at a time when the best bid equals the best ask.[26]

When an order that satisfies one of the abovementioned conditions is newly entered, modified, or cancelled, it has the potential to impact the prevailing quotes. Table C.1 shows the total number of orders from the 12 trader groups defined earlier. The largest proportion of aggressive orders comes from Active-w-Trade FAST/SMALL traders (HFTs). On average, they submit 128.24 aggressive orders (46.30 market orders and 80.21 limit orders). The next largest group of aggressive traders are the FAST/MEDIUM traders, who submit 85.79 aggressive orders (39.57 market orders and 44.76 limit orders). Note that our classification does not take into account trading share, such as top quartile of volume, and that only one quarter of FAST/SMALL traders (HFTs) participate in the pre-opening period but their submission of aggressive orders is significantly greater than that of the other groups. The

[26]Such a situation occurs when the cumulative amount of buy orders equals that of sell orders. Thus, the next order must cause an imbalance between buy and sell orders and make the best ask higher than the best bid price. We refer to such orders as "locked orders." Cao, Ghysels, and Hatheway (2000) analyze locked/crossed market quotes during the NASDAQ pre-opening period. In the TSE's pre-opening period, market best quotes may be locked, which means that the best ask equals the best bid, but crossed quotes (which means that the best bid is greater than the best ask) never happen, by rule.

ratios of aggressive limit orders to the total number of limit orders from these two most aggressive groups of traders are 27.82% and 31.80%, respectively. FAST/SMALL traders (HFTs) place the highest number of aggressive limit orders in relation to the total number of orders, which indicates their interest in affecting the price.

Table C.1: Aggressive orders during pre-opening period

Summary statistics for order aggressiveness for the 12 trader groups, during the pre-opening period excluding the first 10 minutes, for 97 stocks from the TOPIX100, during the sample period of April and May 2013. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Speed	Inventory	Total # of aggressive orders	# of market orders	# of aggressive limit orders	# of zero imbalance orders	Ratio of total order aggres-siveness	Ratio of limit order aggressiveness
Panel A: Active-w/o-Trade							
FAST	LARGE	13.20	0.41	12.03	0.76	46.46%	44.16%
	MEDIUM	17.75	0.90	16.24	0.61	46.46%	44.26%
	SMALL (HFTs)	28.01	0.51	26.47	1.02	47.92%	46.52%
	NOTRADE	2.83	0.07	2.50	0.26	35.14%	32.39%
MODERATE	LARGE	5.53	0.65	4.11	0.77	27.76%	22.20%
	MEDIUM	8.53	0.87	6.58	1.08	25.85%	21.19%
	SMALL	8.89	0.73	7.22	0.95	28.37%	24.32%
	NOTRADE	1.41	0.17	1.01	0.23	26.43%	20.43%
SLOW	LARGE	4.61	0.49	3.45	0.67	27.86%	22.42%
	MEDIUM	2.39	0.22	1.78	0.39	25.89%	20.67%
	SMALL	2.10	0.16	1.68	0.26	28.35%	24.02%
	NOTRADE	4.03	0.50	2.99	0.53	27.74%	22.21%
Panel B: Active-w-Trade							
FAST	LARGE	13.51	6.52	6.82	0.17	60.23%	43.33%
	MEDIUM	85.79	39.57	44.76	1.46	47.20%	31.80%
	SMALL (HFTs)	128.24	46.30	80.21	1.73	38.13%	27.82%
MODERATE	LARGE	12.28	6.00	6.04	0.24	51.24%	34.08%
	MEDIUM	54.43	22.02	30.75	1.66	39.51%	26.96%
	SMALL	77.02	29.07	45.83	2.12	37.44%	26.26%
SLOW	LARGE	14.37	8.24	5.90	0.22	74.99%	55.18%
	MEDIUM	8.79	4.41	4.08	0.31	57.75%	38.82%
	SMALL	6.40	3.09	3.10	0.21	58.13%	40.23%

D. Configuration of multiple virtual servers (VSs) used by one trader

On January 4, 2010, the TSE launched a new trading system named "Arrowhead", which reduced the order submission response time to 2 milliseconds. The main features of this system are (1) accelerated computer-processing speeds, (2) a co-location service that reduces the physical distance between market participants (investors as well as brokerage firms), eliminating the former transmission time of around 3 to 9 milliseconds between the TSE's "Arrowhead" and the customer's computer, and (3) the removal of the three-second delay in intra-day matching. Thus, January 2010 can be viewed as the month of introduction of a new trading paradigm in Japan.

VSs are used to send/receive data to/from the TSE. There are 5,580 servers in existence during our sample period. Most of them (2,692) are used as single servers and the rest as part of multiple-server configurations. When using multiple servers, each trader optimizes the configuration of servers so that she can maximize the performance of her trading activity. Some traders trade a specific group of stocks every day, in which case they may fix the allocation of stocks to each server. Other traders may change part of their allocation on a day-by-day basis. As Table 1 shows, by optimizing the number of stocks per server, a trader can increase her speed significantly. Figure D.1 illustrates one example of a server configuration.

Figure D.1: Illustration of a possible VS configuration for mimicking the TSE's matching engine

This figure shows an example of a potential server configuration. One trading desk (trader) uses four VSs to handle its order flow. The optimizing technique illustrated involves allocating stocks to individual servers with the aim of mimicking the allocation of stocks in the TSE's matching engine. This enables the trader to avoid conjecturing about the order submission task for a large number of stocks at a particular VS.