DARYA YUFEROVA

Price Discovery, Liquidity Provision, and Low-Latency Trading

Price Discovery, Liquidity Provision, and Low-Latency Trading

Price Discovery, Liquidity Provision, and Low-Latency Trading

Prijsontdekking, liquiditeitsprovisie, en snelle handel

THESIS

to obtain the degree of doctor from
Erasmus University Rotterdam
by the command of
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctoral Board

The public defense shall be held on
Thursday 30 June 2016 at 09:30 hours

by

Darya Yuferova
born in Novosibirsk, Russia

Erasmus University Rotterdam

Doctoral Committee

Promotor: Prof.dr. M.A. van Dijk

Other members: Prof.dr. A.J. Menkveld
Prof.dr. D.J.C. van Dijk
Prof.dr. M.J.C.M. Verbeek

Copromotor: Dr. D.G.J. Bongaerts

Erasmus Research Institute of Management – ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ERIM Electronic Series Portal: http://repub.eur.nl/pub

ERIM PhD Series in Research in Management, 379
ERIM reference number: EPS-2016-379-F&A
ISBN 978-90-5892-445-2
©2016, Darya Yuferova

Design: B&T Ontwerp en advies www.b-en-t.nl

This publication (cover and interior) is printed by haveka.nl on recycled paper, Revive®.
The ink used is produced from renewable resources and alcohol free fountain solution.
Certifications for the paper and the printing production process: Recycle, EU Flower, FSC, ISO14001.
More info: http://www.haveka.nl/greening

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the author.

To My Family

Acknowledgements

"The journey of a thousand miles begins with a single step," – Lao-Tzu.

Five exciting years have passed since I started as a PhD Candidate at the Rotterdam School of Management, Erasmus University. Now, I am staying one step away from reaching the goal that I was obsessed with for quite a long period of time — accomplishing a PhD degree. I have come a long way from the decision made by an undergraduate student in Novosibirsk State University to continue my education abroad to finilazing my PhD studies. I strongly believe that my "journey of a thousand miles" has just begun and many wonderful moments are still waiting for me in the future. Hereby, I would like to thank all those people who have helped me to make my "single step."

First and foremost, I would like to express my gratitude for my supervisors Mathijs van Dijk and Dion Bongaerts for all their support and guidance that has helped me to develop into a person I am now. Mathijs and Dion have always been around when I needed them most. The knowledge I have acquired from them is truly priceless. Needless to say, I have enjoyed working with them and hope that we will have other joint projects in the future.

I would also like to thank other members of my dissertation committee Joop Huij, Casper de Vries, Andre Lucas, Albert Menkveld, Loriana Pelizzon, Elvira Sojli, Marti Subrahmanyam, Marta Szymanowska, Sjoerd van Bekkum, Dick van Dijk, and Marno Verbeek for reading my work and given insightful comments and suggestions. I have been lucky enough to work on joint projects with excellent academic researchers and enjoy the endless opportunity to learn from them. Therefore, my special thanks go to my co-authors Mario Bellia, Loriana Pelizzon, Richard Roll, Dominik Rösch, Marti Subrahmanyam, and Jun Uno.

I cannot underestimate the importance of the working environment in the successful completion of my PhD studies. These warm words go to the faculty and staff members of the Finance Department. I am indebted to Mathijs Cosemans, Sarah Draus, Francisco Infante, Thomas Lambert, Anjana Rajamani, Fabrizio Spargoli, Elvira Sojli, Mark Van Achter, and Wolf Wagner for offering me their help in the refinement of the job market paper and the preparation for the job market process itself. The great outcome of my job market would not be possible without their help. I would also like to thank our secretaries Myra van Lissenberg and Flora Pos for their great organization skills that made my life much easier. I also dedicate my appreciation to the PhD students in Finance Department: Aleks, Eden, Jose, Lingtian, Marina, Rogier, Roy, Ruben, Teodor, and Vlado with whom we had a great time together.

I would also like to express my gratitude for faculty of the Economics Department at Novosibirsk State University who inspired me to pursue the academic career: Sergey Kokovin, Lubov Talysheva, and Alexander Tsyplakov.

Finally, special thanks go to my family. Words cannot express how grateful I am to my mother, father, step-father, and my grandmother who have always believed in me even when I have doubted myself and supported me to strive towards my goal.

И последняя, но от этого не менее важная и искренняя, благодарность моей семье. Словами невозможно передать, как я благодарна маме, папе, Володе, и бабушке за то, что всегда в меня верили, даже и особенно тогда, когда я сомневалась в себе, и поддерживали на пути к моей цели.

Darya Yuferova

Rotterdam

April 2016

Contents

List of Tables

List of Figures

Chapter 1

Introduction

During the past decades, international equity markets have been fundamentally altered due to the intensification of the globalization process (market interconnectedness) and vast increase in trading speed (emergence of low-latency trading). Researchers and industry pundits have paid increasing attention to these changes and the threats that they may pose to the international financial markets. Specifically, researchers in the field of market microstructure analyse how the trading process was affected by the above mentioned changes. In this dissertation, I investigate the consequences of these two changes with respect to price discovery and liquidity provision.

First, I investigate linkages across international equity markets with an emphasis on the role of liquidity and trading activity in the transmission of shocks to prices (see Chapter 2)[1]. Second, I examine how price discovery and liquidity provision were altered due to the emergence of algorithmic (low-latency) trading (see Chapters 3 and 4)[2].

[1] Chapter 2 is based on the paper entitled "The Propagation of Shocks across International Equity Markets: A Microstructure Perspective," which is a joint work with Dion Bongaerts, Richard Roll, Dominik Rösch, and Mathijs van Dijk (available at http://ssrn.com/abstract=2475518). I was actively involved in developing the conceptual framework, hypotheses, and methodology used in this paper. I was also responsible for conducting the majority of the data analyses and writing substantial parts of the paper.

[2] Chapter 3 is based on the paper entitled "Intraday Return Predictability, Informed Limit Orders, and Algorithmic Trading," which is my single-authored job market paper (available at http://ssrn.com/abstract=2686082).

Chapter 4 is based on the paper entitled "Low-Latency Trading and Price Discovery without Trading: Evidence from the Tokyo Stock Exchange in the Pre-Opening Period and the Opening Batch Auction," which is joint work with Mario Bellia, Loriana Pelizzon, Marti Subrahmanyam, and Jun Uno (available at http://ssrn.com/abstract=2705962). I was actively involved in developing the conceptual framework, hypotheses, and methodology used in this paper. I was also responsible for conducting the majority of the data analyses.

Interconnectedness of financial markets

Despite the general awareness of market interconnectedness (see Karolyi (2003) for review of the literature on shock propagation across financial markets), before the recent financial crisis occurred, few people recognized how strong the links between the different markets actually were. This global crisis shed light on the interconnectedness and interdependence between financial markets that differ both in terms of instruments traded and geographical location. The crisis is believed to have started in the U.S. housing market, in particular in the mortgage-backed securities market, and quickly spread to the other markets in the U.S. and then across borders. Soon the whole world suffered from a severe global financial crisis (e.g., Longstaff (2010); Rose and Spiegel (2010); Eichengreen, Mody, Nedeljkovic, and Sarno (2012); Bekaert, Ehrmann, Fratzscher, and Mehl (2014)).

Furthermore, the recent financial crisis has highlighted the importance of liquidity, since part of the financial distress during the sub-prime crisis was liquidity-induced (e.g., Brunnermeier (2008); Gorton (2009a,b)). Previous research shows that liquidity is priced both as a characteristic (Amihud and Mendelson (1986); Amihud (2002)) and a risk factor (Pastor and Stambaugh (2003); Acharya and Pedersen (2005)). Recent studies also propose that liquidity dry-ups may be related to the extreme price movements (e.g., Bernardo and Welch (2004); Morris and Shin (2004); Gârleanu and Pedersen (2007); Huang and Wang (2009); Cespa and Foucault (2014)).

For example, Brunnermeier and Pedersen (2009) propose a well-known mechanism that may lead to market crashes during the crisis: liquidity spirals. According to their model, there can be two types of mutually reinforcing spirals: margin spirals and loss spirals. Suppose a speculator receives an external funding shock and hits his funding constraint; in other words, he does not have enough capital to cover margin requirements. This forces him to sell the assets from his inventory position as soon as possible, even at fire sale prices, which lowers the market price of the asset. As prices are driven away from fundamentals, speculators will incur losses on their inventory positions (loss spiral) and be subject to higher margins (margin spiral). As a result, liquidity in the market evaporates and prices drop rapidly.

If shocks in low-liquidity environments can induce crises, or amplify crises, the damage to the real economy can be substantial. For example, according to Getter, Jickling, Labonte, and Murphy (2007), GDP growth can be damaged by a financial crisis through two channels: consumption and investment. Because of the crisis, the wealth of individual households declines, leading to a decrease in their consumption. Increased borrowing costs as well as limited access to capital markets will impact companies' investments in physical capital. In sum, it is important to understand what role market microstructure variables, such as liquidity and trading activity, play in the origination of financial shocks. Detailed understanding of how shocks originate and propagate is crucial in developing regulatory policies that decrease the fragility of financial system.

Motivated by the recent financial crisis, Chapter 2 of this dissertation investigates the intraday propagation of shocks to prices, liquidity, and trading activity across 12 equity markets around the world for the 1996 – 2010 period. We use the non-parametric approach proposed in Barndorff-Nielsen and Shephard (2006) to identify jumps in prices, liquidity (as measured by proportional quoted and effective spreads), and trading activity (as measured by turnover and market order imbalance) employing a 5-minute frequency. The findings show that shocks to prices and trading activity regularly spillover across equity markets within the same 5-minute interval. The results suggest that these shocks are driven by information rather than liquidity as they do not revert and are related to macroeconomic news announcements as well. On the contrary, shocks to liquidity are rare and tend to be isolated events. In other words, shocks to liquidity seem to be unrelated to shocks to prices and/or trading activity in the same market, as well as to shocks to liquidity in other markets. In sum, the results do not confirm the role of liquidity in the origination and propagation of financial shocks.

After the recent financial crisis, several regulatory measures were undertaken to decrease the vulnerability of financial markets (e.g., the Dodd-Frank Wall Street Reform and Consumer Protection Act in the U.S.). This act, among other changes, forbids proprietary trading by commercial banks to avoid unnecessary risk taking

(Volcker Rule) and disruptive trading practices such as “spoofing” (submitting orders without any intention to execute them in order to manipulate the price) — a technique actively used by algorithmic (low-latency) traders. The consequences of algorithmic (low-latency) trading are the focus of Chapters 3 and 4 of this dissertation.

Low-latency trading

One can argue that speed differentials were always present in the financial markets and, hence, no changes to price discovery and/or liquidity provision are expected. Indeed, financial historians can provide such examples dating back to 18th century. For example, Dempster, Wells, and Wills (2000) analyze trading and information dissemination on London and Amsterdam exchanges around that period: “During the 1700’s the distance between London and Amsterdam was three days travel.... Wilson [1941, 104] relates the story of small fishing smacks that were supposed to meet the English ships and speed back to Amsterdam with the latest news related to the Bubble.”

Although speed differentials clearly are not a new phenomenon, there are several reasons the emergence of low-latency trading may have altered financial markets significantly. First, the speed at which algorithmic traders operate is far beyond the human ability. For instance, in 2010 NASDAQ upgraded its system to allow operations at the nanosecond level. Second, algorithmic traders and specifically their subset high-frequency traders (HFTs) are nowadays responsible for the majority of equity turnover. Agarwal (2012) shows that in 2005 HFTs were responsible for only 20% of the U.S. equity turnover, but in 2010 their share increased to 50%-60%.

Biais and Foucault (2014) and O’Hara (2015) review the recent literature on the advantages and disadvantages of low-latency trading. First, algorithmic traders have crowded out traditional market makers and, thus, represent the majority of intermediaries in the modern equity markets (e.g., Menkveld (2013); Jovanovic and Menkveld (2015)). Contrary to traditional market makers, algorithmic traders are not required to make the market (provide liquidity) when necessary. Thus, fast liquidity providers may disappear from the market at a moment when they are needed most (e.g., during turmoil times), which in turn may lead to market fragility (Bongaerts and Van Achter (2016)). Second, algorithmic traders are likely to be informed traders given that they have superior technology to process information from the order flow and to trade on it (Cespa and Foucault (2011); Scholtus, van Dijk, and Frijns (2014); Foucault, Kozhan, and Tham (2015); Foucault, Hombert, and Roşu (2016); Menkveld and Zoican (2015)). Therefore, algorithmic traders may make prices more efficient and impose adverse selection costs on other market participants. The aggregate impact of low-latency traders on the welfare still remains unclear as the arms race between traders may lead to overinvestment in speed compared to social optimum (e.g., Ye, Yao, and Gai (2013); Biais, Foucault, and Moinas (2015); Pagnotta and Philippon (2015); Bongaerts, Kong, and Van Achter (2016)).

To sum up, algorithmic (low-latency) traders are a very controversial group of market participants who have provoked a lot of discussion by regulatory bodies (see SEC (2010); SEC (2014)), as well as academics, and industry pundits.[3] For instance, low-latency traders are blamed for the origination and amplification of the Flash Crash on May 6, 2010, when the U.S. market dropped by approximately 9% and rebounded back within 30-minutes. Kirilenko, Kyle, Samadi, and Tuzun (2015) provide empirical evidence that although low-latency traders did not directly instigate the Flash Crash, they certainly amplified it.

Despite the fact that low-latency traders have been present in the market for more than a decade, they still attract a lot of regulatory attention. In particular, on November 24, 2015, the U.S. Commodity Futures Trading Commission unanimously approves Regulation on Automated Trading (Reg AT). As part of the proposed rules, algorithmic traders should open their source code to regulators in order to ensure that their algorithms do not make markets more vulnerable. In Chapters 3 and 4, I analyze the role of algorithmic (low-latency) trading in price discovery and liquidity provision.

In Chapter 3, I examine how information is incorporated into prices in the limit order book markets and what role algorithmic traders play in this process. This re-

[3] Lewis (2014) presents a popular although one-sided view on low-latency traders.

search question is motivated by the following facts. First, the limit order book is the dominant market design for equity exchanges around the world (see Swan and Westerholm (2006)). Second, the traditional view in the market microstructure literature used to be that informed traders use only market orders (e.g., Glosten and Milgrom (1985); Kyle (1985); Glosten (1994)). Third, the emergence of algorithmic traders could alter the price discovery process substantially. I examine the strategic choice of informed traders for market versus limit orders for all NYSE-listed common stocks for the 2002 – 2010 period by analyzing intraday (one-minute) return predictability from market and limit order flows. My first finding indicates that informed limit (not market) orders are the dominant source of intraday return predictability. In order to establish causal relations between algorithmic trading and informational content of limit and market orders, I focus on the period surrounding permanent exogenous shock to algorithmic trading — NYSE Hybrid Market introduction, which is a technological change in market design that resulted in an increase in speed and automation. My second finding shows that the increase in algorithmic trading is associated with more informed trading through both market and limit orders. Overall, my results suggest that widely-used measures of informed trading/adverse selection neglect the lion's share of informed trading — informed trading through limit orders.

Chapter 4 of this dissertation focuses on the role of low-latency traders in the price discovery process during the pre-opening period, the first time of the day when prices could incorporate overnight information. For investors who want to execute their orders at the opening call, it is important to know whether under the presence of algorithmic traders pre-opening quotes are good predictors of the opening prices.[4] To the best of our knowledge, this is the first study that sheds light on the role of low-latency traders in the pre-opening period. In this chapter, we analyze whether low-latency traders participate in the market and whether they contribute to price discovery in the absence of trading. We use account level data for TOPIX100 constituents during April and May 2013 provided by the Tokyo Stock Exchange to answer these questions. First, we develop a new classification of traders, which is

[4]For studies on price discovery during pre-opening period, see Biais, Hillion, and Spatt (1999); Cao, Ghysels, and Hatheway (2000); and Barclay and Hendershott (2003).

more comprehensive than what has been used in the prior literature. Second, we show that although low-latency traders participate less in the pre-opening period than in the continuous trading session, they actively contribute to price discovery: largely via new limit orders and price revisions. Third, stocks in which low-latency traders are dominant are those for which pre-opening mid-quotes converge faster to the opening price and are less biased. In sum, low-latency traders are the dominant contributors to price discovery during the pre-opening period.

Overall, this dissertation contributes to the area of market microstructure, broadly defined. Prior studies have mainly focused on the U.S. equity market. An important advantage of studying international equity markets is the prospect of drawing policy lessons from variations in the ways these markets are organized and regulated. Several open questions in terms of the regulation of low-latency trading and optimal market design in the presence of low-latency traders remain. For example, Budish, Cramton, and Shim (2015) question whether continuous trading or periodic batch auctions are a better market design in the presence of low-latency traders. Another issue which remains unanswered was raised by SEC (2014) with respect to the definition of a high-frequency trader. SEC (2014) argues that current metrics that use account level data and are based on the high-frequency traders' characteristics outlined in SEC (2010) may be too narrow to capture the true range of high-frequency activity. At the same time, proxies for high-frequency trading calculated from publicly available data sources may be too broad and include activity that should not be classified as high-frequency trading. My future research agenda is related to the two questions highlighted above. In sum, I aim to derive policy implications for the regulation of low-latency trading activity in order to improve such functions of financial markets as price discovery and liquidity provision.

Chapter 2

The Propagation of Shocks Across International Equity Markets: A Microstructure Perspective*

2.1 Introduction

Since at least the stock market crash of October 1987, investors, policy makers, and researchers have been interested in whether and how shocks to one financial market spread to other markets. The Mexican, Asian, and LTCM crises in the 1990s were accompanied by the emergence of a large literature on international financial market linkages and financial contagion. The recent global financial crisis has further highlighted how shocks to certain financial markets can rapidly spread to markets for other asset classes and to markets in other countries. Yet, the channels through which financial market shocks originate and propagate across markets are not well understood.[1]

A growing body of theoretical research points at an important role for market

*This chapter is based on Bongaerts, Roll, Rösch, van Dijk, and Yuferova (2016) "The Propagation of Shocks Across International Equity Markets: A Microstructure Perspective" (available at http://ssrn.com/abstract=2475518). We are grateful to Yakov Amihud, Torben Andersen, Joachim Grammig, Charles-Albert Lehalle, Francis Longstaff, Albert Menkveld, Asani Sarkar, Ramabhadran Thirumalai, Michel van der Wel, Christian Voight, Avi Wohl, seminar participants at Erasmus University, and conference participants at the 5th Emerging Markets Finance Conference in Bombay, the 2014 Extreme Events in Finance conference in Royaumont, the 8th Financial Risks International Forum in Paris, the 2014 German Finance Association meeting in Karlsruhe, and the 2014 INFER workshop in Bordeaux for helpful comments. We thank Michel van der Wel for sharing the U.S. macro news announcements data. Van Dijk gratefully acknowledges financial support from the Vereniging Trustfonds Erasmus Universiteit Rotterdam and from the Netherlands Organisation for Scientific Research through a "Vidi" grant. This work was carried out on the National e-infrastructure with the support of SURF Foundation. We thank OneMarket Data for the use of their OneTick software.

[1]See, among others, Eun and Shim (1989), Roll (1989), Hamao, Masulis, and Ng (1990), and Lin, Engle, and Ito (1994) for early research on the propagation of financial market shocks; Reinhart and Calvo (1996), Forbes and Rigobon (2002), Bae, Karolyi, and Stulz (2003), and Hartmann, Straetmans, and De Vries (2004) for studies on contagion; Karolyi (2003) for a literature review; and Longstaff (2010) and Bekaert, Ehrmann, Fratzscher, and Mehl (2014) for analyses of the propagation of shocks across, respectively, markets for different asset classes and international equity markets during the recent crisis.

liquidity. In particular, recent theories feature "sudden liquidity dry-ups," "liquidity crashes," or "liquidity black holes" that arise through channels related to the supply of and/or demand for liquidity; in turn, these liquidity shocks induce shocks to security prices and spillovers to other markets.[2] Prominent accounts of the recent crisis (e.g., Brunnermeier (2008); Brunnermeier, Crockett, Goodhart, Persaud, and Shin (2009); Gorton (2009a,b)) emphasize the importance of these liquidity channels, but direct empirical evidence is limited.

In this paper, we aim to test the relevance of the liquidity channel for the origination and propagation of financial market shocks by taking a microstructure perspective. Specifically, we analyze why shocks to equity prices occur and whether and how they spread across markets by investigating their relation with shocks to market liquidity and trading activity, using microstructure data for 12 developed and emerging equity markets around the world over the period 1996-2011. To the best of our knowledge, we are the first to study cross-market linkages of stock prices jointly with liquidity and trading activity.[3] Our main alternative hypothesis to the liquidity explanation is that shocks are driven by information; i.e., shocks to prices may reflect economic news that could also be relevant for securities traded on other markets (e.g., King and Wadhwani (1990)).

Our microstructure perspective also involves analyzing the origination and propagation of shocks at a much higher frequency than prior work: 5-minute intervals within the trading day. Most studies to date study the interconnectedness of financial markets at the daily or even lower frequency (e.g., Bae, Karolyi, and Stulz (2003); Hartmann, Straetmans, and De Vries (2004); Longstaff (2010); Bekaert, Ehrmann, Fratzscher, and Mehl (2014); Pukthuanthong and Roll (2015)). However, a relatively low-frequency approach could miss spillovers at higher frequencies and fail to

[2]Recent theoretical studies on such liquidity channels include Kyle and Xiong (2001), Gromb and Vayanos (2002), Kodres and Pritsker (2002), Bernardo and Welch (2004), Morris and Shin (2004), Yuan (2005); Gârleanu and Pedersen (2007), Pasquariello (2007), Andrade, Chang, and Seasholes (2008), Brunnermeier and Pedersen (2009), Huang and Wang (2009), and Cespa and Foucault (2014).

[3] Several papers examine co-movement in liquidity within and across equity markets (e.g., Chordia, Roll, and Subrahmanyam (2000); Brockman, Chung, and Pérignon (2009); Zhang, Cai, and Cheung (2009); Karolyi, Lee, and Van Dijk (2012)) and co-movement in the turnover of individual U.S. stocks (e.g., Lo and Wang (2000) and Cremers and Mei (2007)), but none of these papers also studies stock price linkages.

uncover patterns in liquidity and/or trading activity that could help to explain the occurrence and propagation of shocks to prices within and across markets.[4] We note that for developed markets in recent years, the 5-minute frequency might no longer be perceived as high-frequency. But for emerging markets and for our full sample period 1996-2011, this seems a reasonable frequency to ensure sufficient trading in each interval as well as sufficient time for shocks to propagate to other markets.

Using global tick-by-tick trade and quote data from the Thomson Reuters Tick History (TRTH) database, we construct time-series at the 5-minute frequency of market-wide stock returns (based on midquotes), liquidity (quoted and effective spreads), and trading activity (turnover and order imbalance) for 12 equity markets over the period 1996-2011. We include both developed and emerging equity markets within three regions: America (Brazil, Canada, Mexico, and the U.S.), Asia (Hong Kong, India, Japan, and Malaysia), and Europe/Africa (France, Germany, South Africa, and the U.K.).

We identify shocks to prices, liquidity, and trading activity in each country using the jump measure of Barndorff-Nielsen and Shephard (2006), which is a statistical non-parametric method to test for jumps in a time-series. We propose a refinement of their method so that we are not only able to infer whether a jump occurred on a certain day, but also in which exact 5-minute interval. This approach allows us to create time-series of jumps in prices, liquidity, and trading activity at the 5-minute frequency for each equity market over the period 1996-2011 (based on data on over 5 billion transactions in total).

We first study the origination of shocks on the 12 equity markets in our sample. We find that 5-minute jumps in prices, quoted spreads, and order imbalance are frequent, while jumps in effective spreads and turnover are rare for most markets. The magnitudes of typical jumps in prices, quoted spreads, and order imbalance are large, at around 4 to 6 jump-free standard deviations.

[4] Some prior work does study intraday spillover effects of returns and/or volatility across markets (e.g., Hamao, Masulis, and Ng (1990); King and Wadhwani (1990); Lin, Engle, and Ito (1994); Susmel and Engle (1994); Ramchand and Susmel (1998); Connolly and Wang (2003)), but these studies generally measure returns and/or volatility over intervals of 15 minutes or one hour, look at a more limited sample of markets, and do not consider these variables jointly with liquidity and/or trading activity.

We find little evidence that jumps in prices are accompanied by jumps in liquidity, as measured by quoted spreads. This constitutes initial evidence that liquidity may not play a central role in the origination of price jumps. We do find a relation between jumps in prices and jumps in trading activity, as measured by order imbalance. Around 20% of the jumps in prices in our sample are accompanied by jumps in order imbalance on the same day, which is far more than expected if jumps in prices and order imbalance were independent. Close to 8% of price jumps happen simultaneously with order imbalance jumps in the same 5-minute interval, and almost all of these involve jumps in prices and order imbalance of the same sign. This finding could be an indication that at least some of the price jumps are driven by temporary price pressure effects (i.e., a liquidity demand channel), but could also be consistent with speculative trading around or portfolio rebalancing in response to the arrival of news (i.e., an information channel).

We carry out two specific tests to distinguish the liquidity and information hypotheses. First, we investigate whether there are reversals after jumps in prices (and after simultaneous jumps in prices and order imbalance). We find that, whether accompanied by jumps in order imbalance or not, price jumps represent sudden and permanent shocks to prices; there is no evidence of subsequent price reversals. Second, we examine whether jumps in prices (and simultaneous jumps in prices and order imbalance) occur around macroeconomic news announcements stemming from one of the countries in our sample. We find that a substantial fraction of the jumps in prices (and of the simultaneous jumps in prices and order imbalance) occur around such announcements. For example, in developed Europe, almost 40% of the jumps in prices and around 50% of the simultaneous jumps in prices and order imbalance happen within one hour after a macroeconomic news announcement.[5] The evidence that price jumps do not revert and often occur around macroeconomic news announcements is most consistent with the information channel.

We then investigate within-region and across-region spillover effects of jumps in

[5]These fractions are lower for other countries, primarily because U.S. macroeconomic news announcements yield the strongest results, and the most important U.S. announcements (e.g., GDP, nonfarm payroll employment) fall outside of the opening hours of the American and Asian markets.

prices, quoted spreads, and order imbalance. We document significant spillover effects at the 5-minute frequency for jumps in prices as well as for jumps in trading activity, based on correlations of the time-series of jumps in prices and order imbalance, taking into account the magnitude of the jump. These correlations are especially strong within Europe and between Europe and the U.S. However, jumps in quoted spreads are not correlated across different markets, which suggests that liquidity shocks do not propagate across markets and "sudden liquidity dry-ups" are mainly local phenomena.

We further estimate logit regressions with the jumps in prices on a particular market as the dependent variable to distinguish between same-country, within-region, and across-region spillover effects of jumps in prices and order imbalance. This analysis confirms our findings based on the correlations and furthermore provides evidence of the existence of spillover effects between jumps in prices and order imbalance not only within the same country but also within and across regions.

Overall, this paper finds little empirical support for theories in which liquidity plays a key role in the origination and propagation of financial market shocks. Jumps in equity prices are prevalent and large, and regularly coincide with jumps in order imbalance and with price jumps in other markets. However, price jumps do not revert and often happen around macroeconomic news announcements. Jumps in quoted spreads tend to be isolated events that are neither associated with jumps in prices nor with jumps in quoted spreads on other markets.

Of course, there are limitations to our analysis. Our focus is on the high-frequency origination and propagation of financial market shocks, so we may miss lower-frequency shocks to prices, liquidity, and trading activity. Nevertheless, our results also hold at the 15-minute and 1-hour frequencies (instead of the 5-minute frequency). Our evidence based on intraday data seems to at least challenge the widely held view that financial market liquidity can suddenly evaporate and thereby cause precipitous price drops and spillover effects to other markets. In fact, by analyzing shocks at relatively high frequencies, we stack the cards in favor of finding supportive evidence of a liquidity channel, since our approach allows us to identify price jumps that revert within the day, which lower frequency analyses might miss. Notwithstanding, our results indicate that sudden price shocks are predominantly driven by information.

Also, our liquidity measures are limited to quoted and effective spreads, which may not cover all relevant aspects of market liquidity. However, price impact measures estimated at the 5-minute frequency are extremely noisy and may be mechanically related to price changes. We do obtain similar results using a liquidity measure based on the number of stocks trading in an interval. In separate tests, we also find little evidence that shocks to a variety of proxies for funding liquidity (a potential liquidity supply channel) are associated with a relatively greater prevalence of jumps in prices, liquidity, or trading activity. Furthermore, it is hard to imagine that a true liquidity crash would not show up in quoted spreads.

Our primary contribution is to the literature on international financial market linkages and financial contagion. We add to this line of research by analyzing such linkages across international equity markets at the 5-minute frequency, and by offering a detailed analysis of the dynamics of liquidity and trading activity around shocks to equity prices. We thereby investigate the prediction of a number of recent theoretical studies that channels related to the supply of and/or demand for market liquidity play an important role in the propagation of financial market shocks. Moreover, we contribute to the literature on commonality in liquidity and trading activity by studying the degree of cross-market co-movement in large, sudden changes in liquidity and trading activity.

We believe that our paper sheds new light on a number of important issues. In today's complex, dynamic, and interconnected global financial system, it is important for investors, exchanges, and regulators to understand whether and how shocks are propagated from one financial market to another at high speed, what the role of liquidity and trading activity is in the occurrence and propagation of shocks to prices, and how strong cross-market linkages are within and across different regions. Our results may help investors to make better decisions regarding optimal portfolio diversification, financial institutions to develop better risk management policies, and exchange officials and regulators to develop better policies to reduce international financial fragility.

2.2 Data and methods

This section describes the data, variable definitions, and methods used in the paper. We obtain intraday data on trades and quotes (and their respective sizes) from the Thomson Reuters Tick History (TRTH) database. TRTH is provided by Securities Industry Research Centre of Asia-Pacific (SIRCA) and includes tick-by-tick data for trades and best bid-offer quotes stamped to the millisecond. The database is organized by Reuters Instrumental Codes (RICs), spans different asset classes, and covers more than 400 exchanges since 1996.[6]

To obtain a sample that is representative of global equity markets but still manageable in light of the vast size of the global tick-by-tick data, we pick four countries (with different levels of development) from each of three regions classified based on their time zone: America, Asia, and Europe/Africa.[7] In particular, we select Brazil, Canada, Mexico, and the U.S. from the American region; Hong Kong, India, Japan, and Malaysia from the Asian region; and France, Germany, South Africa, and the U.K. from the European/African region. We obtain the RICs for all common stocks that are traded on the major stock exchange (defined as the exchange that handles the majority of trading volume) in each of these countries from Datastream and then collect the RICs for all of these stocks that were part of the main local market index at some point during the sample period from 1996 till 2011 from the TRTH Speedguide (see Appendix A.1.1). Following Rösch, Subrahmanyam, and van Dijk (2015), we apply extensive data filters to deal with outliers and trades and quotes outside of the daily trading hours (details are in Appendix A.1.2).

[6] Recent papers that use the TRTH database include Boehmer, Fong, and Wu (2012), Lau, Ng, and Zhang (2012), Marshall, Nguyen, and Visaltanachoti (2012), Marshall, Nguyen, and Visaltanachoti (2013a,b), Boehmer, Fong, and Wu (2015), Fong, Holden, and Trzcinka (2014), Frino, Mollica, and Zhou (2014), Lai, Ng, and Zhang (2014), and Rösch, Subrahmanyam, and van Dijk (2015).

[7] We note that even within these regions there are small differences in time zones and trading hours.

2.2.1 Variable definitions

Our primary goal is to provide a microstructure perspective on the propagation of shocks across international equity markets and to test the liquidity vs. information explanations for why such shocks occur and spillover to other markets. Therefore, we focus on intraday data for returns, liquidity, and trading activity at the market-level. Specifically, we choose 5-minute intervals as our unit of observation, which seems to be a reasonable compromise between intervals that are sufficiently fine-grained to study the high-frequency propagation of price shocks and their relation to liquidity and trading activity on the one hand, and intervals that have enough trades to adequately measure trading activity and effective spreads (especially in the beginning of our sample period and for the emerging markets in our sample) and that are long enough to capture spillovers to other markets on the other hand. Our choice of 5-minute intervals is also motivated by Tauchen and Zhou (2011), who use the same frequency to analyze jumps in the S&P500 index (1986-2005), 10-year Treasury bonds (1991-2005) and the dollar/yen exchange rate (1997-2004). We discard overnight changes in prices, liquidity, and trading activity. In supplementary tests, we rerun all of our analyses at the 15-minute and 1-hour frequencies.

We first measure variables at the individual stock-level and then aggregate to the market-level. Following Chordia, Roll, and Subrahmanyam (2008), log returns are computed over 5-minute intervals based on midpoints between the quoted bid and ask prices (rather than based on the trade prices or on midquotes matched with the last trade in the interval) of individual stocks. Using midquote returns has two advantages. First, it avoids the bid-ask bounce problem that is inherent in returns based on trade prices. Second, it ensures that returns for every stock are computed over the same 5-minute interval despite differences in trading frequency across stocks.

We use proportional quoted spreads and proportional effective spreads ($PQSPR$ and $PESPR$) as measures of liquidity. While the former measures transaction costs only if the trade does not exceed the depth at the best bid-offer (BBO), the latter measures the actual transaction costs when a trade takes place. We compute $PQSPR$ based on quote data only, for the last BBO available for a given stock in a particular

5-minute interval. For $PESPR$, we first match trade and quote data and then compute the effective spread based on the last trade within a particular 5-minute interval as the difference between the trade price and the prevailing midquote. $PESPR$ is thus only available for 5-minute intervals with at least one trade. This restriction is not very onerous as in total there are more than 5 billion trades in our sample. We stay away from estimating price impact measures at the 5-minute frequency, since they tend to be very noisy and may be mechanically related to price changes. As a further test, we redo all of our analyses based on the number of stocks trading in a specific interval as an alternative market-wide liquidity measure. Motivated by the emerging literature on the link between market liquidity and funding liquidity (e.g., Brunnermeier and Pedersen (2009)), we also examine whether shocks to various measures of funding liquidity are associated with shocks to prices, liquidity, and trading activity.

We use turnover and order imbalance (OIB) to measure trading activity. We compute turnover as the total trading volume (in local currency) of a stock during the 5-minute interval, and scale this number by the aggregate market capitalization at the end of the previous year. To compute OIB, we need to determine whether a trade is buyer- or seller-initiated. We use the Lee and Ready (1991) algorithm to sign trades. We then compute the OIB of a given stock as the difference between buyer- and seller-initiated trading volume (in local currency) during the 5-minute interval, scaled by the aggregate market capitalization at the end of the previous year. We obtain data on aggregate market capitalization (in USD) and exchange rates from the World Bank website.

We aggregate our five main variables (returns, quoted and effective spreads, turnover, and order imbalance) to the market-level by taking an equally-weighted average of the stock-level variables for returns and spreads, and by summing up the scaled stock-level variables for turnover and order imbalance. To reduce the impact of stock-level noise and to secure a certain level of representativeness, we discard 5-minute intervals for a given market when there are fewer than ten stocks with a trade.

2.2.2 Jump measure (BNS)

There is a vast literature that studies spillover effects from one market to another as well as a plethora of different methods. For example, Bae, Karolyi, and Stulz (2003) define "coexceedances" as the simultaneous incidence of extreme returns (identified as those in the top or bottom 5% of the return distribution by country over the whole sample period) and model the determinants of such coexceedances using multinomial logit models. Hartmann, Straetmans, and De Vries (2004) use extreme value theory to show that the actual probability of a simultaneous crash on two markets is much higher than the expected probability under the assumption that extreme events are independent across markets. Chiang, Jeon, and Li (2007) use a dynamic conditional correlation (DCC) model, while Rodriguez (2007) employs a switching copula approach to document spillover effects.

In this paper, we follow Pukthuanthong and Roll (2015) and use a statistical jump measure to identify a shock.[8] Advantages of this method are that it adheres closely to the intuitive view of a shock to financial markets as a discontinuous event in an otherwise continuous time-series, that it does not require arbitrary definitions of extreme events, and that it is easy to compute and does not require the estimation of a large number of parameters. Furthermore, it can pinpoint the particular interval when the shock occurs and it can detect both country-specific shocks and shocks that are transmitted to other markets, without a need to make assumptions regarding the joint distribution of variables across multiple markets. Potential disadvantages are that on days with many observations in the tail of the full-sample distribution, it may not classify observations as jumps that could be regarded as extreme under different methods and, similarly, it may not identify "clumps" (series of changes in the variables of interest that may accumulate to a large change but do not constitute discontinuous jumps). To mitigate the latter concern, we also measure jumps at the 15-minute and 1-hour frequencies.

In this paper, we use the jump measure proposed by Barndorff-Nielsen and Shep-

[8]Various jump measures include those devised by Barndorff-Nielsen and Shephard (2006), Jiang and Oomen (2008), Lee and Mykland (2008), and Jacod and Todorov (2009).

hard (2006) [BNS] which is based on the ratio of scaled bipower (continuous) variation to squared variation and which is “by far the most developed and widely applied of the different [jump] methods” (Bollerslev, Law, and Tauchen (2008), p. 239) and the best jump measure in the simulations of Pukthuanthong and Roll (2015). The squared variation is obtained by summing up the squared 5-minute observations during a day, while the bipower variation is based on the scaled summation of the products of the absolute values of the current and lagged 5-minute observations. The bipower and squared variations on a particular day are similar in the absence of jumps, while the bipower variation is significantly smaller than the squared variation if the time-series has a jump on that day.

Under the null hypothesis of no jumps, the BNS measure follows a standard normal distribution, so statistical significance can be determined based on standard normal critical values. Since the time-series of jumps in prices, liquidity, and trading activity form the inputs of our subsequent analyses, the usual tradeoff between type I and type II errors is especially relevant in our setting. In particular, we are concerned about incorrectly classifying “normal” observations as jumps. To limit the type I error, we use a 0.1% significance level (instead of the common 10%, 5%, or 1% thresholds). Our time-series based on 5-minute intraday intervals over 1996-2011 contain sufficient observations (up to around 370,000) to still have the potential to detect a substantial number of jumps based on this strict statistical criterion.

For each day, we can thus identify whether there was a jump in any of these variables on any market. A drawback of the standard application of the BNS method is that it cannot pinpoint the exact 5-minute interval when the jump occurs. We thus propose a refinement of the BNS approach in the form of an algorithm that allows us to infer the exact interval in which the jump occurs. In short, for each day with a significant jump statistic for a certain variable, we identify the 5-minute return interval with the observation that has the greatest effect on the jump statistic and is greater in absolute terms than 1.96 jump-free standard deviations (i.e., the square root of the scaled bipower variation for that variable on that day). We classify such observations as jumps. It turns out that on all days in our sample for which the

BNS statistic is significant, there is at least one such observation. Subsequently, we remove it from the time-series of that variable on that day and again test for the occurrence of a jump on that day, repeating the procedure until no further jumps are detected. Appendix A.2 presents a more detailed description of this algorithm.[9]

2.3 Empirical results

This section first presents summary statistics for the returns, liquidity, and trading activity at the market-level (Section 2.3.1), followed by summary statistics of the BNS jump measures for each of these variables (Section 2.3.2). Subsequently, we investigate the link between jumps in prices, liquidity, and trading activity within each market (Section 2.3.3) and whether any such link is driven by liquidity or information (Section 2.3.4). Then, we study the propagation of shocks to prices, liquidity, and trading activity across equity markets within the same region and also across regions, for the same variable and across different variables (Section 2.3.5). We conclude this section with a discussion of a number of supplementary tests (Section 2.3.6).

2.3.1 Summary statistics

Table 2.1 shows the mean and the standard deviation of the 5-minute equally-weighted market returns, equally-weighted proportional quoted spreads ($PQSPR$) and effective spreads ($PESPR$), aggregate market turnover, and aggregate market order imbalance scaled by aggregate market capitalization (OIB) for each of the 12 markets.

Averaged across the 12 markets in our sample, the mean 5-minute return equals -0.1 basis points per 5-minute interval, with an average standard deviation of around 10 basis points. Average returns are slightly negative for 9 out of 12 countries, primarily because we include the recent crisis in our sample period and exclude overnight returns (Berkman, Koch, Tuttle, and Zhang (2012) show that intraday returns tend to be lower than overnight returns). The average mean $PQSPR$ ($PESPR$) across markets is equal to 0.49% (0.36%), with an average standard deviation of 0.34% (0.24%). As a comparison, Chordia, Roll, and Subrahmanyam (2011) report

[9]We thank Torben Andersen for his advice on this approach.

Table 2.1. Summary statistics of market returns, liquidity, and trading activity

This table shows the whole sample time-series mean and standard deviation of the 5-minute equally-weighted market returns in basis points, the 5-minute equally-weighted proportional quoted spreads (***PQSPR***) and effective spreads (***PESPR***) in percentage, the 5-minute market aggregate turnover in basis points, and the 5-minute market aggregate order imbalance scaled by aggregate market capitalization (***OIB***) in basis points for 12 equity markets over 1996-2011. We refer to Section 2.2 and Appendix A.1 for a detailed description of sample selection, data filters, and variable definitions. The final row presents the total number of valid 5-minute intervals in the sample (at least ten stocks should be traded in each particular interval to be included in the sample). Markets are grouped by region and are listed in alphabetical order within each region. Data to calculate these variables are from TRTH (trade and quote data) and the World Bank website (aggregate market capitalization and exchange rates). Only common stocks that were ever part of the major local equity index are included in the computation of market returns, quoted and effective spreads, order imbalance, and turnover (data on index constituents are from the TRTH Speedguide, while common stocks are identified through Datastream).

		America				Asia				Europe/Africa			
		Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
PRICE	Mean	-0.160	-0.047	0.085	0.015	-0.153	-0.385	-0.160	-0.229	-0.062	-0.113	0.055	-0.031
	St.Dev.	14.000	5.346	9.279	10.710	13.752	11.860	11.867	8.992	10.804	10.216	6.062	7.671
PQSPR	Mean	0.487	0.751	0.652	0.149	0.476	0.265	0.416	0.941	0.195	0.218	0.819	0.499
	St.Dev.	0.331	0.307	0.378	0.120	0.185	0.144	0.149	0.306	0.113	0.158	0.300	0.348
PESPR	Mean	0.358	0.471	0.414	0.088	0.428	0.232	0.350	0.749	0.179	0.159	0.505	0.336
	St.Dev.	0.243	0.215	0.253	0.066	0.186	0.118	0.112	0.177	0.110	0.108	0.248	0.294
Turnover	Mean	0.117	0.131	0.102	0.083	0.219	0.306	0.295	0.070	0.516	0.228	0.088	0.156
	St.Dev.	0.187	0.087	0.060	0.038	0.159	0.205	0.198	0.061	0.724	0.172	0.065	0.125
OIB	Mean	0.002	0.004	0.004	0.005	0.005	0.000	0.013	-0.001	0.003	0.000	0.000	-0.001
	St.Dev.	0.119	0.035	0.037	0.016	0.071	0.068	0.090	0.030	0.386	0.064	0.030	0.031
# Obs.		111,302	284,549	148,337	282,881	135,083	154,554	162,314	230,355	366,396	323,607	269,618	372,300

an average $PESPR$ of 0.0223% for NYSE stocks over 2001-2008, which is of roughly the same order of magnitude as the number of 0.088% reported for the U.S. in Table 2.1, especially when taking into account that spreads were considerably higher over the period 1996-2000. Averaged across markets, scaled turnover (OIB) is equal to 0.19 (0.003) basis points, with a standard deviation of 0.17 (0.08) basis points.

The final row of Table 2.1 shows the number of 5-minute intervals for which the various variables can be computed for each market; this number varies across markets according to the sample period available in TRTH, the opening hours, and the intensity of trading activity (since we discard 5-minute intervals during which fewer than ten stocks are traded). The average number of 5-minute intervals across all markets is 236,775. We transform the stock variables $PQSPR$ and $PESPR$ to a flow variable by taking 5-minute log-changes (in line with Pukthuanthong and Roll (2015), who compute shocks to prices based on the return series). We also take log-changes of turnover to construct a variable with a mean close to zero. We then compute the daily BNS jump measure for the five key variables of interest and use the algorithm described in Appendix A.2 to identify the exact 5-minute interval when a jump occurs in case the daily BNS statistic is statistically significant.

2.3.2 Frequency and magnitude of jumps in prices, liquidity, and trading activity

Panel A of Table 2.2 shows the total number of 5-minute intervals with jumps across variables and markets. Positive ("POS") and negative ("NEG") jumps are reported separately. We observe a substantial number of jumps in prices, $PQSPR$, and OIB. Averaged across all 12 markets, there are 196 (210) positive (negative) jumps in prices; 117 (65) positive (negative) jumps in $PQSPR$; and 256 (242) positive (negative) jumps in OIB. Jumps in these variables occur much more often than under the no jumps assumption. We reject the null hypothesis of no jumps if the BNS statistic for a particular day is below the 0.1% percentile of the standard normal distribution (one-sided test). Thus, the type I error (erroneously rejecting the null hypothesis of no jumps) is 0.1% of the total number of days in our sample. Put differently, over the entire 1996-2011 sample period we would expect to see four days being classified

Table 2.2. The frequency and magnitude of jumps in prices, liquidity, and trading activity

Panel A of this table shows the number of 5-minute intervals with a jump in the 5-minute equally-weighted market returns (***PRICE***), 5-minute log-changes in equally-weighted proportional quoted spreads (***PQSPR***) and effective spreads (***PESPR***), 5-minute log-changes in the market aggregate turnover (***TURNOVER***), and 5-minute market aggregate order imbalance scaled by the aggregate market capitalization (***OIB***) for 12 equity markets over 1996-2011. Panel B of this table shows the corresponding mean and standard deviation of the magnitude of the jump measured in terms of jump-free standard deviations (that is, the square root of the scaled bipower variation). The total number of 5-minute observations for each variable is shown in Table 2.1 Jumps are identified using the BNS jump statistic that is based on the ratio of the bipower (continuous) variation to the squared variation of the intraday observations for each variable (see Appendix A.2 for details). We classify a day as a day with a jump in a particular variable if the BNS measure is below the 0.1% percentile of the standard normal distribution. Subsequently, we follow an algorithm that allows us to pinpoint the exact 5-minute interval in which the jump occurs. The jumps are classified according to their sign: positive (POS) and negative (NEG). Markets are grouped by region and listed in alphabetical order within each region. Data to calculate these variables are from TRTH (trade and quote data) and the World Bank website (aggregate market capitalization and exchange rates). Only common stocks that were ever part of the major local equity index are included in the computation of market returns, quoted and effective spreads, order imbalance, and turnover (data on index constituents are from the TRTH Speedguide, while common stocks are identified through Datastream).

Panel A: Number of 5-minute intervals with a jump

		America				Asia				Europe/Africa			
		Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
PRICE	POS	33	132	109	140	433	19	500	244	201	162	148	227
	NEG	39	127	88	160	439	68	637	187	225	205	134	212
PQSPR	POS	6	189	110	38	35	38	191	167	27	107	222	278
	NEG	7	63	107	21	42	16	82	154	13	47	131	92
PESPR	POS	1	4	9	4	7	2	70	5	11	2	3	14
	NEG	1	3	2	6	7	1	11	15	25	2	0	12
Turnover	POS	5	10	4	11	30	36	29	4	17	11	0	15
	NEG	0	6	9	9	5	153	9	0	14	12	1	11
OIB	POS	304	383	54	129	324	77	205	232	590	246	410	115
	NEG	254	296	25	79	266	182	143	242	560	224	493	139

Table 2.2. The frequency and magnitude of jumps in prices, liquidity, and trading activity (continued)

Panel B: Magnitude of the jump (in jump-free standard deviations)

			America				Asia				Europe/Africa			
			Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
PRICE	POS	Mean	5.27	4.65	4.80	4.50	5.82	4.39	5.44	6.35	5.90	4.98	5.06	5.08
		St.Dev.	2.32	1.50	1.81	1.77	2.10	1.41	2.15	2.26	2.31	1.70	1.95	2.01
	NEG	Mean	-4.99	-4.54	-4.65	-4.36	-6.20	-4.50	-5.65	-6.98	-5.40	-5.27	-4.75	-5.42
		St.Dev.	1.57	1.92	1.61	1.39	2.37	1.39	2.17	2.95	2.26	1.90	1.80	2.13
PQSPR	POS	Mean	4.80	5.23	4.47	5.20	4.29	5.20	5.33	5.12	6.05	5.07	4.73	6.63
		St.Dev.	2.98	1.78	1.17	1.74	1.36	1.94	1.78	1.63	4.40	1.75	1.48	2.70
	NEG	Mean	-4.60	-3.90	-4.11	-5.32	-3.85	-4.19	-4.25	-4.17	-7.61	-4.17	-4.01	-5.05
		St.Dev.	2.20	1.16	1.32	2.08	1.25	1.24	1.11	1.32	6.87	1.25	1.10	2.24
OIB	POS	Mean	4.55	5.89	5.08	4.83	4.50	4.17	4.16	4.71	6.71	5.51	6.34	4.91
		St.Dev.	2.43	3.20	1.46	2.46	1.62	1.36	1.34	1.57	4.09	2.14	6.09	1.85
	NEG	Mean	-4.75	-5.88	-4.95	-5.04	-4.68	-4.53	-4.30	-4.67	-7.38	-5.36	-6.46	-4.97
		St.Dev.	3.09	2.75	1.55	2.00	1.63	1.67	1.31	1.60	4.48	2.17	4.06	1.78

as days with jumps under the null hypothesis of no jumps. However, the numbers of jumps in prices, $PQSPR$, and OIB are much higher. For example, in Germany there are 205 5-minute intervals with a negative jump in prices, which occur on 178 different days (compared to four days under the null hypothesis) or approximately 5.1% (compared to 0.1% under the null hypothesis) of all 3,523 trading days from 1999 to 2011 for which jumps could be estimated for Germany. The finding that jumps in prices, $PQSPR$, and OIB occur much more frequently than under the no jumps assumption is obtained for all markets in the sample. While positive and negative jumps in prices and order imbalance are equally likely, we identify almost twice as many positive as negative jumps in $PQSPR$. Intuitively, sudden evaporations of liquidity are more common than sudden liquidity improvements.

Jumps in $PESPR$ and turnover are considerably less prevalent than jumps in prices, $PQSPR$, and OIB. In fact, $PESPR$ (11 positive and 7 negative jumps on average across markets) and turnover (14 positive and 19 negative jumps on average across markets) almost never jump. With the notable exceptions of $PESPR$ for Japan and turnover for India, the number of days on which we identify jumps in $PESPR$ and turnover is only slightly greater than the type I error of our test. A potential explanation for the low number of jumps in $PESPR$ (as compared to jumps in $PQSPR$) is that $PESPR$ can only be measured when a trade occurs; rational investors observing a jump in quoted spreads could abandon the market and return when liquidity improves. Based on the results in Panel A of Table 2.2, we exclude the time-series of jumps in $PESPR$ and turnover from the remainder of our analyses.

Although these empirical patterns of jumps in the different variables are overall quite similar across markets, there is also considerable cross-market variation in the number of jumps for individual variables. For example, the number of positive (negative) 5-minute jumps in prices varies from 19 to 500 (from 39 to 637) across different markets; the number of positive (negative) jumps in $PQSPR$ varies from 6 to 278 (from 7 to 154); and the number of positive (negative) jumps in OIB varies from 54 to 590 (from 25 to 560). There is no clear pattern across developed and emerging markets. In unreported analyses (available from the authors), we also study the time-series development of the number of jumps by country and by variable and find little evidence of consistent patterns (e.g., trends or clustering).[10]

The jumps documented in Panel A of Table 2.2 are all statistically significant at a very high confidence level. However, market participants not only care about the frequency and statistical significance of shocks to financial markets, but also about their economic magnitude. Therefore, in Panel B of Table 2.2, we present summary statistics (means and standard deviations) of the magnitudes of the 5-minute market-wide jumps in prices, $PQSPR$, and OIB. To obtain a consistent measure of the magnitude of jumps across the different variables and markets, we assess the magnitude in terms of the number of "jump-free standard deviations" or the square root of the scaled bipower variation (since the bipower variation measures the variation of the continuous, i.e., non-jump, part of the process only).

It is clear from Panel B of Table 2.2 that the magnitudes of the jumps in prices, $PQSPR$, and OIB we detect using the BNS approach are large for all markets in the sample. The average jump magnitude for both negative and positive jumps in prices, $PQSPR$, and OIB is around five jump-free standard deviations, with a range in absolute terms from 3.85 (negative $PQSPR$ jumps in Hong Kong) to 7.61 (negative $PQSPR$ jumps in France) jump-free standard deviations.[11]

For jumps in prices, five jump-free standard deviations correspond to a 5-minute market-wide shock to equity prices of around 40 basis points, which signifies an economically large market-wide price shock over such a short interval (40 basis points is 400 times greater than the absolute value of the average 5-minute market return across markets). Jumps in $PQSPR$ of five jump-free standard deviations amount to a market-wide shock to quoted spreads of 42%, which is 83 times greater than the absolute value of the average 5-minute change in market-wide quoted spreads.

[10] We also find only limited evidence that jumps in prices, liquidity, and trading activity cluster during a trading day on a specific market. For example, averaged across the 12 markets, 89% of the days with a significant BNS statistic for the time-series of aggregate equity prices have only one price jump, 9% have two price jumps, and 2% have three or more price jumps.

[11] The theoretical probability of observing a five standard deviation shock to a normally distributed variable is 0.006 basis points. This probability corresponds to one 5-minute interval out of 1,744,277, or one 5-minute interval every 96 years (assuming six-hour trading days and 252 trading days per year). In other words, the observed frequency of such substantial shocks is much higher than the expected frequency under the assumption of normally distributed variables.

The results in Table 2.2 thus indicate that jumps in prices, $PQSPR$, and OIB are prevalent and large. In the next subsection, we examine the relation between jumps in prices, liquidity, and trading activity within each market.

2.3.3 Coinciding jumps in prices, liquidity, and trading activity within a market

Recent theoretical studies (referenced in footnote 2) suggest an important role for channels related to the supply of and/or demand for liquidity in the origination and propagation of price shocks. A common thread in these theories is that shocks to prices are accompanied by shocks to liquidity and/or trading activity. For example, price shocks can arise because financial intermediaries reduce the supply of liquidity in the face of funding constraints (e.g., Gromb and Vayanos (2002); Brunnermeier and Pedersen (2009)) or because of a surge in the demand for liquidity when wealth effects, loss limits, or hedging desires induce traders to sell (e.g., Kyle and Xiong (2001); Morris and Shin (2004); Andrade, Chang, and Seasholes (2008)). In several of these models, feedback loops (e.g., "liquidity black holes" or "liquidity spirals") can arise in which deteriorating market liquidity, tightening funding constraints, and selling reinforce each other, causing the decline in liquidity and prices to worsen over time.

As a first assessment of the importance of the liquidity channel for the origination and propagation of price shocks, we are therefore interested in whether price shocks tend to be accompanied by shocks to liquidity and/or trading activity.

We start by documenting the links among jumps in the different variables within each market. To that end, we treat a jump in prices (or in one of the other variables) as an event and examine whether there are jumps in liquidity and/or trading activity at the same time as the event (i.e., in the same 5-minute interval), before the event (from the beginning of the same trading day – or from the previous price jump on the same day – until the event), or after the event (from the event until the end of the same trading day – or until the next price jump on the same day). We refer to co-jumps on the same day as "coinciding" and to co-jumps in the same 5-minute interval as "simultaneous."

The results are in Table 2.3. Panels A and B assess whether price jumps (the event) are accompanied by jumps in, respectively, $PQSPR$ and OIB on the same market on the same day. Panel C assesses whether OIB jumps (the event) are accompanied by jumps in $PQSPR$ on the same market on the same day. The first two columns of each panel show the signs of the jumps in the variables under consideration. For example, in Panel A, the first column shows the sign of the price jump events ("POS" or "NEG"). The first two rows of Panel A show the number of positive or negative price jumps that are *not* associated with a jump in $PQSPR$ on the same market on the same day. The next four rows show the number of positive or negative price jumps that are accompanied by a "simultaneous" positive or negative jump in $PQSPR$ on the same market. The following four rows show the number of positive or negative price jumps that were preceded by a positive or negative jump in $PQSPR$ on the same market on the same day. The final four rows show the number of positive or negative price jumps that were followed by a positive or negative jump in $PQSPR$ on the same market on the same day. The structure of Panels B and C is the same.[12]

Panel A of Table 2.3 shows no consistent pattern in the coincidence of jumps in prices and jumps in $PQSPR$. Very few price jumps are accompanied by jumps in $PQSPR$, either in the same 5-minute interval or before or after the price jump on the same trading day. And even for markets for which prices and proportional quoted spreads regularly jump on the same day (such as Japan), there is no consistent pattern in the direction of the jumps. As an example, although all of the 19 $PQSPR$ jumps in Japan that accompany a negative price jump in the same 5-minute interval are of positive sign (in line with the prediction of the liquidity hypothesis that a price decline is associated with a sudden deterioration in liquidity), we also observe that 13 of the 16 $PQSPR$ jumps in Japan that accompany a positive price jump in the same 5-minute interval are positive, which is hard to reconcile with a liquidity story. Only

[12]We note that the sum of the numbers of price jumps in the columns of Panel A of Table 2.3 sometimes slightly exceeds the total number of price jumps for the respective market reported in Table 2.2 in case some price jumps are accompanied by more than one jump in $PQSPR$ on the same day. The fractions of coinciding jumps reported in this subsection are corrected for any such double counting.

Table 2.3. Coinciding jumps in prices, liquidity, and trading activity within a market

This table shows the number of jumps in variable 1 and variable 2 (where variable 1 and variable 2 refer to either ***PRICE***, ***PQSPR***, or ***OIB***) that occur on the same trading day (within/before/after the same 5-minute interval) for each of the 12 equity markets in our sample over 1996-2011. We treat either a positive or a negative jump in variable 1 as an event and we count the number of 5-minute intervals with jumps in variable 2 in the same interval as the event (i.e., simultaneously), before the event (that is, from the beginning of the same trading day – or from the previous jump in variable 1 on the same day – till the event) and after the event (that is, from the event till the end of the same trading day – or till the next jump in variable 1 on the same day). In each panel, the first two columns show the signs of the jumps in the variables under consideration. The first column shows the sign of the jumps in variable 1 (POS for positive price jumps and NEG for negative jumps in variable 1). In each panel, the first two rows show the number of positive or negative jumps in variable 1 that are not associated with a jump in variable 2 on the same market on the same day. The next four rows show the number of positive or negative jumps in variable 1 that are accompanied by a positive or negative jump in variable 2 on the same market in the same 5-minute interval. The following four rows show the number of positive or negative jumps in variable 1 that were preceded by a positive or negative jump in variable 2 on the same market on the same day. The final four rows show the number of positive or negative jumps in variable 1 that were followed by a positive or negative jump in variable 2 on the same market on the same day. In Panel A, jumps in variable 1 and variable 2 correspond to jumps in 5-minute equally-weighted market returns (***PRICE***) and 5-minute log-changes in equally-weighted proportional quoted spreads (***PQSPR***), respectively; in Panel B, jumps in variable 1 and variable 2 correspond to jumps in 5-minute equally-weighted market returns (***PRICE***) and 5-minute market aggregate order imbalance scaled by aggregate market capitalization (***OIB***), respectively; in Panel C, jumps in variable 1 and variable 2 correspond to jumps in 5-minute market aggregate order imbalance scaled by aggregate market capitalization (***OIB***) and 5-minute log-changes in equally-weighted proportional quoted spreads (***PQSPR***), respectively. We refer to the caption of Table 2.2 and to Appendix A.2 for a detailed description of the jump statistics. Markets are grouped by region and listed in alphabetical order within each region. Data are from TRTH, the World Bank website, and Datastream.

Panel A: Coinciding jumps in prices and *PQSPR*

	Sign of the jump in		America				Asia				Europe/Africa			
	PRICE	***PQSPR***	Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
Jumps in ***PRICE*** with no jumps in ***PQSPR*** on the same day	POS	NA	33	120	87	137	428	18	469	227	190	156	128	190
	NEG	NA	37	114	65	150	430	67	596	157	221	199	123	193
Simultaneous jumps in ***PRICE*** and ***PQSPR***	POS	POS	0	1	2	0	0	0	13	2	1	0	7	4
	POS	NEG	0	1	8	0	0	0	3	1	0	0	3	6
	NEG	POS	0	1	9	2	2	0	19	7	0	0	1	4
	NEG	NEG	0	0	4	0	0	0	0	3	0	1	0	1
Jumps in ***PRICE*** preceded by jump in ***PQSPR*** on same day	POS	POS	0	1	6	1	3	0	4	4	5	3	7	21
	POS	NEG	0	2	3	1	2	0	4	4	1	0	4	4
	NEG	POS	1	6	2	3	3	1	2	4	0	4	3	6
	NEG	NEG	0	4	2	2	6	0	9	13	0	1	0	1
Jumps in ***PRICE*** followed by jump in ***PQSPR*** on same day	POS	POS	0	2	1	1	0	0	4	6	2	1	0	13
	POS	NEG	0	4	5	1	0	1	2	0	3	2	4	1
	NEG	POS	1	1	5	0	1	0	7	11	1	2	7	4
	NEG	NEG	2	1	7	2	0	0	1	6	2	0	3	1

Table 2.3. Coinciding jumps in prices, liquidity, and trading activity within a market (continued)

Panel B: Coinciding jumps in prices and *OIB*

	Sign of the jump in		America				Asia				Europe/Africa			
	PRICE	*OIB*	Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
Jumps in *PRICE* with no jumps in *OIB* on the same day	POS	NA	25	109	106	120	369	12	419	226	118	104	104	190
	NEG	NA	33	106	85	149	359	42	510	173	137	147	109	179
Simultaneous jumps in *PRICE* and *OIB*	POS	POS	1	8	2	15	22	2	42	4	43	31	1	13
	POS	NEG	0	0	0	0	1	0	0	0	0	0	0	0
	NEG	POS	0	0	0	0	1	0	1	0	0	0	2	0
	NEG	NEG	2	3	1	3	22	17	58	3	34	28	0	14
Jumps in *PRICE* preceded by jump in *OIB* on same day	POS	POS	2	4	0	3	20	0	8	5	15	4	6	0
	POS	NEG	4	3	0	1	10	2	2	1	9	5	9	4
	NEG	POS	2	4	0	3	13	1	8	0	16	10	2	1
	NEG	NEG	1	7	0	2	16	8	6	5	15	4	5	3
Jumps in *PRICE* followed by jump in *OIB* on same day	POS	POS	3	5	1	0	19	2	21	5	26	13	13	13
	POS	NEG	5	7	0	0	8	0	5	4	17	4	18	9
	NEG	POS	3	5	1	0	16	1	27	1	9	5	19	7
	NEG	NEG	2	5	1	3	36	3	24	6	21	6	10	5

Panel C: Coinciding jumps in *OIB* and *PQSPR*

	Sign of the jump in		America				Asia				Europe/Africa			
	OIB	*PQSPR*	Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
Jumps in *OIB* with no jumps in *PQSPR* on the same day	POS	NA	301	352	50	128	319	75	195	222	576	236	366	108
	NEG	NA	253	259	25	74	263	174	136	221	552	216	451	116
Simultaneous jumps in *OIB* and *PQSPR*	POS	POS	0	1	0	0	0	0	3	1	1	0	0	0
	POS	NEG	0	0	0	0	0	0	0	0	1	0	1	0
	NEG	POS	0	1	0	0	0	0	2	4	0	1	0	1
	NEG	NEG	0	0	0	0	0	0	0	0	0	0	0	0
Jumps in *OIB* preceded by jump in *PQSPR* on same day	POS	POS	1	17	1	0	2	1	2	3	2	9	11	3
	POS	NEG	1	5	2	0	2	2	2	4	1	5	15	3
	NEG	POS	0	20	0	3	0	4	2	3	2	3	15	13
	NEG	NEG	0	3	0	1	2	0	1	5	0	2	7	2
Jumps in *OIB* followed by jump in *PQSPR* on same day	POS	POS	2	7	0	1	0	0	4	5	4	3	14	1
	POS	NEG	1	5	1	0	0	0	0	0	1	2	4	0
	NEG	POS	0	4	0	0	1	2	0	3	2	3	17	4
	NEG	NEG	0	3	0	1	0	1	0	2	1	1	6	0

6.9% of all price jumps in the sample are accompanied by a jump in $PQSPR$ on the same day, and this fraction drops to 2.2% for the same 5-minute interval. Moreover, only about half of the coinciding jumps in prices and $PQSPR$ are of opposite sign, as predicted by the liquidity hypothesis.[13]

Panel B of Table 2.3 shows a considerably stronger relation between jumps in prices and jumps in OIB. Not only do we observe a greater incidence of coinciding jumps in prices and OIB, these coinciding jumps also more often have the sign predicted by price pressure effects (a liquidity demand channel). In particular, Panel B shows that positive (negative) jumps in prices are regularly associated with positive (negative) jumps in OIB, especially when prices and OIB jump in the same 5-minute interval (as indicated by the higher numbers in the first and the last rows of the "Simultaneous jumps" section in Panel B). Across the whole sample, 19.3% of the jumps in prices are accompanied by a jump in OIB on the same day. Approximately 8% of all price jumps in the sample are accompanied by an OIB jump in the same 5-minute interval, and almost all of these involve same-sign jumps. The finding of regular co-jumps in prices and OIB of the same sign is consistent with the view that prices jump in part because of sudden shifts in the demand for liquidity, but it could also arise as a result of speculative trading around or portfolio rebalancing in response to the arrival of new information.

Panel C of Table 2.3 shows that the pattern of coincidences of jumps in $PQSPR$ and jumps in OIB is about as weak as in Panel A. In short, there is little evidence that jumps in OIB are related to jumps in $PQSPR$. Only 5.1% (0.28%) of the OIB jumps are accompanied by a $PQSPR$ jump on the same day (in the same 5-minute interval).

Overall, the results in Table 2.3 indicate that a non-trivial fraction of the 5-minute jumps in prices are accompanied by same-sign jumps in order imbalance, even within the same 5-minute interval. We find little evidence of such links between jumps in prices and jumps in $PQSPR$ and between jumps in $PQSPR$ and jumps in OIB.

To fully understand the strength of the relation between jumps in prices and

[13] This finding contrasts the results of Jiang, Lo, and Verdelhan (2011), who show that market liquidity shocks have significant predictive power for jumps in U.S. Treasury-bond prices.

jumps in OIB, we need to examine how likely simultaneous jumps in these variables are given the total number of jumps in prices and OIB. As an example, in Germany 28 out of the 205 negative price jumps are accompanied by jumps in OIB of the same sign in the same 5-minute interval. Put differently, approximately 14% of the negative jumps in prices on the German equity market are accompanied by a simultaneous negative jump in OIB. We need a metric to judge whether 14% is abnormally high relative to the benchmark where jumps in prices and jumps in OIB are completely independent. To construct such a metric, we conduct a statistical test to compare the empirically observed frequency of simultaneous jumps in prices and OIB to the theoretical frequency that we would observe if jumps in prices and OIB were independent. The test is based on the comparison of two binomial distributions. The first distribution has a probability of success equal to the empirically observed frequency of simultaneous jumps in prices and OIB. The second distribution has a probability of success equal to the theoretical frequency of such simultaneous jumps under the assumption of independence. We test whether these two probabilities are the same, against the alternative hypothesis that the empirical probability is greater than the theoretical probability.

Table 2.4 shows the number of simultaneous jumps in prices and OIB in the same 5-minute interval by market, as well as the associated empirical probability of simultaneous jumps, the theoretical probability of simultaneous jumps under the independence assumption, and a one-sided p-value of the binomial test described above. For example, for Germany the empirical probability of a jump in prices equals 11.36 basis points and of a jump in OIB equals 14.55 basis points (based on Table 2.2). Thus, under the assumption that jumps in prices and OIB are independent, the probability of observing a simultaneous jump in prices and OIB in the same 5-minute interval is 0.02 basis points (11.36 basis points × 14.55 basis points). However, Table 2.3 shows that simultaneous jumps in prices and OIB are observed in 59 5-minute intervals, which corresponds to an empirical probability of simultaneous jumps of 1.83 basis points. The final row of Table 2.4 shows that the p-value of the test that the empirical probability of simultaneous jumps (1.83 basis points) is equal to the

Table 2.4. The likelihood of simultaneous jumps in prices and order imbalance within a market

This table assesses whether the empirical frequency of simultaneous (that is, same 5-minute interval) jumps in 5-minute equally-weighted market returns (***PRICE***) and 5-minute market aggregate order imbalance scaled by aggregate market capitalization (***OIB***) exceeds the theoretical frequency under the assumption that jumps in prices are independent from jumps in ***OIB***, for each of the 12 equity markets in our sample over 1996-2011. The table shows the number of simultaneous jumps in prices and ***OIB***, the empirically observed frequency of such simultaneous jumps (that is, relative to the total number of 5-minute intervals for each market), and the theoretical probability of such simultaneous jumps under the assumption that jumps in prices and ***OIB*** occur independently (these probabilities are given in basis points). The next row of the table presents the ***p***-value of a statistical test on the equality of the empirically observed frequency and the theoretical probability. The null hypothesis is that the empirical and theoretical probabilities are equal (in other words, jumps between variables are independent), while the alternative is that the empirical probability is greater than the theoretical probability. Numbers in bold font indicate statistical significance at the 1% level or better (one-sided test). The final row indicates number of 5-minute intervals over which jumps could be computed (at least ten stocks should be traded in each particular interval to be included in the sample and at least 25 intervals should be valid during a particular day to start jumps computation). Markets are grouped by region and listed in alphabetical order within each region. We refer to the caption of Table 2.2 and to Appendix A.2 for a detailed description of the jump statistics. Data are from TRTH, the World Bank website, and Datastream.

	America				Asia				Europe/Africa			
	Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
# simultaneous jumps	3	11	3	18	46	19	101	7	77	59	3	27
Prob(simultaneous jumps) empirical (in bp)	0.27	0.39	0.22	0.64	3.42	1.23	6.25	0.30	2.10	1.83	0.11	0.73
Prob(simultaneous jumps) theoretical, under independence assumption (in bp)	0.03	0.02	0.01	0.01	0.29	0.01	0.15	0.04	0.04	0.02	0.04	0.01
p-value (one-sided test)	**0.006**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.072	**<0.001**
# Obs.	110,651	284,365	138,545	282,718	134,326	154,511	161,707	230,193	366,342	322,923	267,093	372,216

theoretical probability (0.02 basis points) is <0.001, which implies a clear rejection of the null hypothesis that jumps in prices and OIB on the German equity market are independent.

For all countries except South Africa, we reject the null hypotheses that jumps in prices occur independently from jumps in OIB at the 1% level or better. On some markets (Brazil and Mexico), the number of simultaneous jumps in prices and OIB is quite small, but on many other markets we document frequent simultaneous jumps in prices and OIB in the same 5-minute interval (most notably Japan, with 100 such cases). In other words, a significant fraction of price jumps is associated with simultaneous jumps in OIB, which suggests that studying such co-jumps can help us to understand why price jumps occur.

The evidence in this subsection suggests that price jumps occur independently of $PQSPR$ jumps, but not of OIB jumps. Although we thus find little support for the main prediction of the liquidity hypothesis that shocks to prices are accompanied by shocks to liquidity, the finding that a subset (around 8%) of price jumps occur simultaneously with OIB jumps could be consistent with a liquidity demand channel at least for this subset of price jumps. In the next subsection, we present two specific tests of the predictions of the liquidity and information hypotheses.

2.3.4 Jumps in prices and OIB Liquidity vs. information

The liquidity and information hypotheses offer competing explanations for why price jumps occur, and why they occur simultaneously with jumps in order imbalance. On the one hand, jumps in prices can occur as the result of the price pressure associated with large one-directional uninformed order flow when markets are less than perfectly resilient. On the other hand, a sudden and permanent price adjustment can occur as a result of new information arriving on the market that may also give rise to market-wide order imbalances – for example due to speculative trading or large-scale portfolio rebalancing. (We note that given the fact that many co-jumps in prices and OIB occur within the same 5-minute interval, it is hard to pin down causality or the exact sequence of these jump events.)

We conduct two empirical tests to distinguish between these hypotheses. First, we investigate whether prices exhibit a reversal after a price jump (and after a simultaneous jump in prices and OIB) in Section 2.3.4.1. The liquidity hypothesis predicts that price pressure is temporary and prices should revert, while the information hypothesis predicts that price adjustments are permanent and no reversal should be observed. Then, we examine whether jumps in prices (and OIB) are associated with macroeconomic news announcements, which represent the arrival of important information on the market (Section 2.3.4.2).

2.3.4.1 Price reversals after jumps in prices (and OIB)

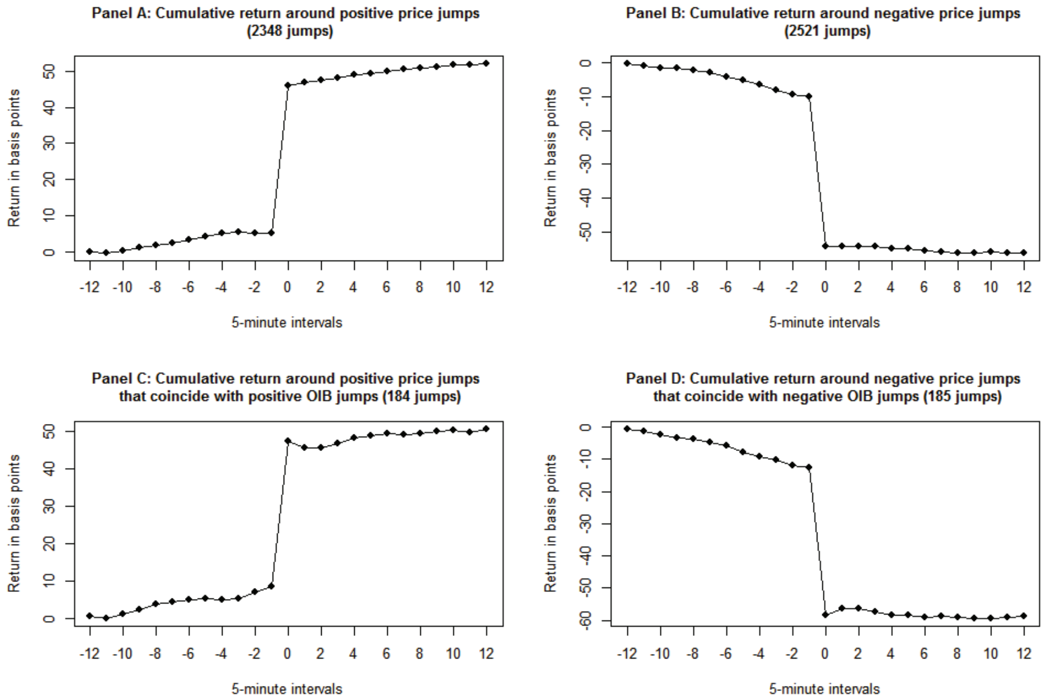
Figure 2.1 presents graphs of the cumulative market return in 5-minute intervals from one hour before ($t = -12$) until one hour after jumps ($t = +12$) in prices (positive jumps in Panel A and negative jumps in Panel B) and jumps in prices that are accompanied by jumps in OIB of the same sign in the same 5-minute interval (positive co-jumps in Panel C and negative co-jumps in Panel D), aggregated across all jumps on the 12 markets in our sample and measured in basis points.[14] The total number of jumps underlying Panels A and B is 2,348 and 2,521, respectively (obtained by aggregating the number of positive and negative jumps in prices across all markets from Table 2.2). The total number of jumps underlying Panels C and D is 184 and 185, respectively (obtained by aggregating the number of positive and negative simultaneous jumps in prices and OIB across all markets from Table 2.3). As also shown in Table 2.2, Figure 2.1 indicates that the average price jump is around 40-50 basis points, which is a substantial market-wide return over a 5-minute interval. Negative price jumps tend to be slightly larger than positive price jumps, but there is little indication that price jumps that are accompanied by same-sign jumps in OIB are of a different magnitude than price jumps in isolation.

The graphs in the four panels of Figure 2.1 also show that price jumps are truly sudden: there is a clear discontinuity relative to cumulative returns before the 5-minute interval of the jump – although there is some indication of a slight run-up in the same direction in the hour before the jump (the run-up is statistically significant

[14] We substitute missing data with zeroes in case of jumps for which we do not have data for the complete period from one hour before to one hour after the jump.

Figure 2.1. Behavior of prices around price jumps and simultaneous price and *OIB* jumps

This figure shows the cumulative 5-minute market-wide equally-weighted returns in basis points (averaged across all the price jumps in the 12 equity markets) from one hour before till one hour after either positive or negative jumps in price over 1996-2011. Panel A and Panel B present cumulative average returns around all price jumps in our sample, while Panel C and Panel D present cumulative average returns around jumps in price that coincide with jumps in ***OIB*** of the same sign in the same 5-minute interval. Cumulative returns are plotted for each 5-minute interval in the event window, with the price jump taking place at $t = 0$. We refer to the caption of Table 2.2 and to Appendix A.2 for a detailed description of the jump statistics. Data are from TRTH, the World bank website, and Datastream.



at the 5% level or better starting at $t = -8$, possibly suggesting a slight amount of information leakage). These patterns indicate that our identification of price jumps is quite clean; unreported results show that jumps in $PQSPR$ and in OIB represent similarly sudden and discontinuous changes in the variable of interest.

More importantly from the perspective of distinguishing the liquidity and information channels, there is little evidence of any reversal following either price jumps or simultaneous jumps in prices and OIB. If anything, there is some slight return continuation, especially after positive price jumps. In other words, price jumps tend to constitute permanent price changes, consistent with the hypothesis that price jumps (as well as simultaneous jumps in prices and OIB) occur due to the arrival of new information on the market rather than due to price pressure effects or other liquidity channels.

2.3.4.2 Macroeconomic news announcements and jumps in prices (and OIB)

The second test of the liquidity vs. information hypotheses aims to examine more directly whether price jumps (and simultaneous jumps in prices and OIB) are related to information events. In particular, we investigate whether jumps in prices (and OIB) are associated with macroeconomic news announcements from a number of different countries in our sample over the period 2001-2011, obtained from the Econoday database (the data on macroeconomic news announcements includes scheduled announcements regarding GDP, nonfarm payroll employment, producer and consumer price indices, etc.).[15] We manually select similar categories of macroeconomic news announcements as used in Andersen, Bollerslev, Diebold, and Vega (2003) and Opschoor, Taylor, Van der Wel, and van Dijk (2014) based on the description of the announcement. We only include announcements that fall within the opening hours

[15] We are grateful to Michel van der Wel for providing the data on U.S. macroeconomic news announcements over 2004-2009, as used in Opschoor, Taylor, Van der Wel, and van Dijk (2014), and for his advice on obtaining and filtering the data for the other years and for several of the other countries in our sample. We note that the Econoday database does not cover our full sample period 1996-2011, but starts in 2001. For some countries, coverage starts even later (for example, coverage of macroeconomic news announcements in China – which we include because of their relevance for Hong Kong – starts in 2007) and some of the other countries in our sample are not covered at all during our sample period.

of at least one of the markets in our sample. In total, we analyze 6,037 different macroeconomic news announcements from Canada, China, the European Monetary Union (EMU), France, Germany, Japan, the U.K., and the U.S., out of which 1,921 occur within the opening hours of the American markets, 2,304 occur within the opening hours of the Asian markets, and 4,751 occur within the opening hours of the European/African markets in our sample.[16]

We examine how many of the jumps in prices (and OIB) in our sample occur within a short window (from five minutes before till one hour after the event) around the release time of any of the macroeconomic news announcements we collected. We use a one-hour window after the announcements to allow for some time for the news to be incorporated in prices. One hour may seem like a long period of time to capture the response of U.S. markets to U.S. macroeconomic news announcements in recent years. However, for other markets, for the earlier years in our sample, and for news from other countries/regions, it may take more than a few minutes for the news to be fully incorporated into local prices. As a comparison, Lee (2012) uses a 30-minute post-announcement window in her analysis of jumps in market-wide and firm-specific U.S. equity prices around U.S. macroeconomic news announcements in the period 1993-2008.

Table 2.5 presents the results. The first line in the table shows the total number of macroeconomic news announcements we collected from around the globe that occurred within the opening hours of each of the 12 markets in our sample. The other four lines in the table show the total number of price jumps on each market over the period 2001-2011, the number of price jumps that occur within the event window around the macroeconomic news announcements, the total number of simultaneous jumps in prices and OIB on each market over the period 2001-2011, and the number of simultaneous jumps in prices and OIB that occur within the event window around the news announcements.

For all of the markets in our sample except Japan, our sample includes at least 500

[16] We aggregate multiple macroeconomic announcements with the same release time to one event, so the numbers of announcements reported in the text and in Table 2.5 refer to the number of unique release times.

Table 2.5. Simultaneous jumps in prices and order imbalance within a market and macroeconomic news announcements

This table presents the number of jumps in 5-minute equally-weighted market returns (***PRICE***) and the number of simultaneous jumps in ***PRICE*** and 5-minute market aggregate order imbalance scaled by aggregate market capitalization (***OIB***) that occur within a short event window around macroeconomic news announcements over 2001-2011. In total, we use data on 6,037 different macroeconomic news announcements from the American region (Canada and the U.S.), from the Asian region (China and Japan), and from the European region (EMU, France, Germany, and the U.K.). The event window around the macroeconomic news announcements is [-1,+12], measured in 5-minute intervals. The first row indicates the number of unique release times of macroeconomic news announcements that occur within the opening hours of the market. Markets are grouped by region and listed in alphabetical order within each region. Data are from TRTH, the World Bank website, and Datastream. Data on the macroeconomic news announcements are from the Econoday database.

	America				Asia				Europe/Africa			
	Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
# of macroeconomic news announcements within trading hours	1,871	1,212	1,164	1,212	618	2,183	118	1,280	4,524	4,524	4,337	4,524
# of jumps in ***PRICE***	72	154	195	193	714	85	745	341	257	303	197	250
# of jumps in ***PRICE*** in the window [-1,+12] around macro announcements	23	15	23	44	60	14	7	25	87	119	29	95
# of simultaneous jumps in ***PRICE*** and ***OIB***	3	11	3	12	42	18	62	5	55	54	0	26
# of simultaneous jumps in ***PRICE*** and ***OIB*** in the window [-1,+12] around macro announcements	2	1	1	5	2	4	3	0	30	29	0	11

news announcements from different countries that occur within the market's opening hours over the period 2001-2011. For most markets, a considerable fraction of the price jumps (and simultaneous jumps in prices and OIB) occur within one hour of a macroeconomic news announcement. Around 17% of the price jumps (and 31% of the simultaneous jumps in prices and OIB) on the American markets are associated with a macroeconomic news announcement. These news announcements are mainly European and U.S. announcements, though we note that the most important U.S. announcements (e.g., nonfarm payroll, employment, producer and consumer price indices) fall outside the opening hours of the American markets. For Asia, we find that 6% of the price jumps (and 7% of the simultaneous jumps in prices and OIB) occur within the event window. However, none of the U.S. macroeconomic news announcements and very few of the news announcements from China and Japan take place within the opening hours of the Asian markets. In other words, the vast majority of the macroeconomic news announcements reported in Table 2.5 for markets in Asia are announcements from Europe, which may be of comparatively little relevance for Asian markets.

For European markets, we find strong evidence that jumps in prices (and OIB) are related to macroeconomic news announcements. For example, for Germany, we document 303 5-minute intervals with price jumps over 2001-2011, of which 119 (or 40%) occur around one of the news announcements in our sample. Over the same period, we observe 54 5-minute intervals with simultaneous jumps in prices and OIB in Germany, of which 29 (or 54%) are in the event window surrounding one of the announcements. Across the three European markets in our sample, 37% of the price jumps and 52% of the simultaneous jumps in prices and OIB occur around an announcement. The relative strength of the results for European markets is likely driven by the fact that many of the U.S. macroeconomic news announcements – arguably the most influential in the world – fall within the opening hours of the European markets.[17]

[17] In an unreported analysis, we examine whether jumps in prices (and simultaneous jumps in prices and OIB) in Europe tend to occur around particular categories of U.S. macroeconomic news announcements. We find that especially nonfarm payroll employment, producer and consumer price indices, and initial unemployment claims announcements are often accompanied by jumps in prices

Across all 12 markets in the sample, 15% of the price jumps (and 30% of the simultaneous jumps in prices and *OIB*) are associated with a macroeconomic news announcement. We interpret this as evidence that a considerable fraction of the jumps in prices (and *OIB*) in our sample are associated with the arrival of important economic news, consistent with the information hypothesis. Of course, our results do not imply that we can trace each price jump to one of the many macroeconomic news announcements in our sample. However, we would like to point out that these announcements often involve relatively minor news events or news that was anticipated, and that many of the most important (notably U.S.) announcements do not occur within the trading hours of most markets in our sample. For European markets, which do tend to be open during U.S. macroeconomic news announcements, we find a much stronger association between price jumps and economic news. Furthermore, there is a host of other news events (e.g., unscheduled news announcements, policy speeches, industry news, local or global political news, acts of terrorism, natural or nuclear disasters) that could cause sudden shocks to equity prices but that are hard to measure in a consistent way. Our estimates are therefore likely to heavily underestimate the fraction of price jumps associated with news events.

Nonetheless, to examine whether there is stronger evidence in favor of the liquidity hypothesis for the jumps in prices (and *OIB*) that we are unable to relate to macroeconomic news, we repeat the price reversal analysis from Section 2.3.4.1 for the subsets of jumps in prices (and *OIB*) that do and that do not occur within the event window around one of the macroeconomic news announcements over 2001-2011. The results, which are unreported but available from the authors, show that the graphs of the cumulative market return from one hour before until one hour after price jumps are very similar for jumps in prices (and *OIB*) that are and that are not associated with macroeconomic news; there is no evidence of price reversals in either case. This finding suggests that even price jumps outside of the event window around the macroeconomic news announcements in our sample are mainly driven by information rather than liquidity.

(and *OIB*) in Europe.

Taken together, the evidence in this subsection based on return reversals surrounding price jumps (and simultaneous jumps in prices and OIB) and based on the occurrence of jumps in prices (and OIB) around macroeconomic news announcements is most consistent with the information hypothesis. In the next subsection, we assess whether and why jumps in prices, liquidity, and trading activity spill over across markets.

2.3.5 Spillovers in jumps in prices, liquidity, and trading activity across markets

So far, we have provided evidence on the prevalence of jumps in prices, liquidity, and trading activity, on coinciding jumps in different variables within one market, and on the main channel through which jumps in prices (and OIB) arise. We now turn to one of the main further goals of the paper: to analyze the role of liquidity and trading activity in the within-region and across-region propagation of shocks to financial markets. To the best of our knowledge, our paper is the first to study high-frequency spillover effects of shocks to liquidity and trading activity across equity markets, and to link these to spillovers of price shocks.

We start with presenting summary statistics for coinciding jumps in price, $PQSPR$, and OIB across markets within each of the three regions, followed by an examination of spillover effects within and across regions for each of the variables separately (Section 2.3.5.1). In Section 2.3.5.2, we aim to explain price jumps on one market based on variables from the same market, the same region, and other regions.

2.3.5.1 Coinciding jumps in prices, liquidity, and trading activity across markets

Table 2.6 reports the number of days on which one, two, or three or more markets within the same region exhibit a positive/negative jump in prices, $PQSPR$, or OIB. Here, we only analyze co-jumps by region since, for example, there is no overlap in trading hours between markets in America and in Asia and we exclude overnight changes in our variables.

In most instances, there is at most one market that has a jump in prices, $PQSPR$,

Table 2.6. Co-jumps in prices, liquidity, and trading activity across markets on the same day

This table presents the number of days on which one, two, or three markets within each region (America, Asia, and Europe/Africa) exhibit a jump in 5-minute equally-weighted market returns (***PRICE***), 5-minute log-changes in equally-weighted proportional quoted spreads (***PQSPR***), or 5-minute market aggregate order imbalance scaled by aggregate market capitalization (***OIB***), for 12 equity markets over 1996-2011. First, we use the BNS jump measure to identify days with jumps in each variable for each market. Then, we count the number of markets that have a jump of the same sign in the same variable on the same day, and distinguish between three cases: only one market has a jump in that variable on a certain day, two markets have a jump in that variable of the same-sign on the same day, and three or more markets have a jump in that variable of the same-sign on the same day. Jumps are classified according to their sign: positive (POS) and negative (NEG). We refer to the caption of Table 2.2 and to Appendix A.2 for a detailed description of the jump statistics. Data are from TRTH, the World Bank website, and Datastream.

	America						Asia						Europe/Africa					
	PRICE		*PQSPR*		*OIB*		*PRICE*		*PQSPR*		*OIB*		*PRICE*		*PQSPR*		*OIB*	
	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG	POS	NEG
=1	339	330	307	166	592	481	856	908	390	259	593	569	452	489	491	241	849	921
=2	9	15	4	3	16	11	102	135	8	1	39	57	71	56	20	2	117	116
>=3	0	1	0	0	1	0	5	6	0	0	2	3	15	21	0	0	9	11

or OIB during a particular day in a particular region, but there are also a considerable number of cases of two or more countries having a jump in the same variable of the same sign on the same day. For example, in the European/African region, we observe 566 days over our sample period on which at least one of the four markets in that region experiences a negative price jump. Out of those 566 days, 489 (86.4%) are days on which only one of the four markets faces a negative price jump, on 56 days (9.9%) two markets face a negative price jump, and on 21 days (3.7%) at least 3 markets face a negative price jump.

Similar results are obtained for positive price jumps and for negative and positive OIB jumps in Europe/Africa and for negative and positive jumps in prices and in OIB in Asia. Co-jumps in the same variable of the same sign on different markets within a region are much less likely in America. Across all 12 markets in the sample, 11.3% (8.7%) of all days with price (OIB) jumps exhibit same-sign price (OIB) jumps in at least two different markets within the same region. In contrast, we find very few occasions of co-jumps in $PQSPR$ on different markets within the same region. Across all markets, only 2.0% of the days with $PQSPR$ jumps exhibit same-sign $PQSPR$ jumps in more than one market. This finding suggests that shocks to liquidity do not tend to occur on multiple markets in the same time frame.

Overall, the results in Table 2.6 indicate that although the majority of jumps in prices, $PQSPR$, or OIB are market-specific, we regularly observe co-jumps in prices and OIB of the same sign on the same day across multiple markets in the Asian and European/African regions. However, jumps in $PQSPR$ on a given day are almost always contained to a single market.

In Table 2.7, we extend the analysis in Table 2.6 by presenting correlations of jumps in prices, $PQSPR$, and OIB at the 5-minute (instead of daily) frequency and not only across individual markets within each region, but also across markets in different regions. Table 2.7 shows contemporaneous spearman rank correlations for the 5-minute time-series of jumps in prices (Panel A), $PQSPR$ (Panel B), and OIB (Panel C) across different markets (during overlapping trading hours only). We take into account the sign, magnitude, and significance of the jumps by setting our

Table 2.7. Correlations of 5-minute jumps in prices, *PQSPR*, and *OIB* within and across regions

This table presents contemporaneous Spearman rank correlations between jumps in 5-minute equally-weighted market returns (***PRICE***), 5-minute log-changes in equally-weighted proportional quoted spreads (***PQSPR***), and 5-minute market aggregate order imbalance scaled by aggregate market capitalization (***OIB***) across 12 equity markets over 1996-2011. Panel A shows Spearman correlations of 5-minute jumps in prices across markets, Panel B shows Spearman correlations of 5-minute jumps in ***PQSPR*** across markets, and Panel C shows Spearman correlations of 5-minute jumps in ***OIB*** across markets. Correlations are computed based on overlapping trading hours only. We take into account the sign and the magnitude of the jumps by setting our jump variables equal to zero in 5-minute intervals without a jump in the respective variable, and to the signed magnitude of the jump (measured in jump-free standard deviations) in 5-minute intervals with a jump. Numbers in bold font indicate statistical significance at the 1% level or better. Markets are grouped by region and listed in alphabetical order within each region. We refer to the caption of Table 2.2 and to Appendix A.2 for a detailed description of the jump statistics. Data are from TRTH, the World Bank website, and Datastream.

Panel A: Spearman correlations of 5-minute jumps in *PRICE*

		America				Asia				Europe/Africa			
		Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
America	Brazil		**3.97%**	0.00%	0.00%					**5.11%**	**5.08%**	0.00%	**7.28%**
	Canada			0.00%	**1.87%**					**1.66%**	-0.87%	0.00%	0.91%
	Mexico				0.65%					**1.40%**	**2.46%**	0.00%	**1.57%**
	U.S.									**7.61%**	**7.65%**	0.01%	**7.64%**
Asia	Hong Kong						0.00%	0.02%	**10.47%**	0.01%	0.00%	0.00%	0.03%
	India							-0.01%	**1.58%**	**3.19%**	-0.01%	0.00%	-0.01%
	Japan								0.01%				
	Malaysia									0.01%	0.00%	0.00%	0.00%
Europe/Africa	France										**18.78%**	**1.86%**	**13.44%**
	Germany											**2.13%**	**15.71%**
	South Africa												**2.36%**
	U.K.												

Table 2.7. Correlations of 5-minute jumps in prices, *PQSPR*, and *OIB* within and across regions (continued)

Panel B: Spearman correlations of 5-minute jumps in *PQSPR*

		America				Asia				Europe/Africa			
		Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
America	Brazil		0.00%	0.00%	0.00%						0.00%	0.00%	0.00%
	Canada			0.00%	**0.85%**					-0.02%	-0.02%	-0.05%	-0.03%
	Mexico				0.00%					0.00%	0.00%	0.04%	0.01%
	U.S.									0.00%	0.00%	0.01%	0.00%
Asia	Hong Kong						0.00%	0.01%	0.00%	-0.02%	-0.03%	-0.05%	-0.01%
	India							-0.02%	0.00%	0.00%	0.00%	0.00%	0.00%
	Japan								0.05%				
	Malaysia									-0.04%	-0.07%	-0.03%	-0.10%
Europe/Africa	France										0.00%	0.00%	**1.00%**
	Germany											-0.01%	-0.01%
	South Africa												-0.01%
	U.K.												

Panel C: Spearman correlations of 5-minute jumps in *OIB*

		America				Asia				Europe/Africa			
		Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
America	Brazil		0.00%	-0.01%	**1.34%**					**1.91%**	0.77%	-0.99%	**2.65%**
	Canada			**2.02%**	**0.84%**					0.84%	0.71%	**-1.84%**	0.00%
	Mexico				0.00%					0.00%	0.00%	-0.01%	0.00%
	U.S.									**5.48%**	**3.62%**	-0.01%	**1.30%**
Asia	Hong Kong						0.01%	**1.92%**	-0.30%	**-2.45%**	-0.01%	0.02%	0.00%
	India							0.00%	0.00%	0.63%	0.01%	0.00%	0.00%
	Japan								**3.23%**				
	Malaysia									0.00%	0.00%	0.00%	0.00%
Europe/Africa	France										**8.88%**	0.43%	**4.08%**
	Germany											**0.76%**	**6.55%**
	South Africa												0.30%
	U.K.												

jump variables equal to zero in 5-minute intervals without a significant jump in the respective variable, and to the signed magnitude of the jump (measured in jump-free standard deviations) in 5-minute intervals with a jump. Bold correlations are significant at the 1% level or better. We do not report 5-minute correlations across markets in America and Asia since trading hours do not overlap.

The table shows that the time-series of signed price jumps are significantly correlated at the 5-minute frequency within the European/African region, and in particular within developed Europe. For example, the correlation between price jumps in Germany and the U.K. is equal to 15.71%. The correlations between price jumps on developed markets in Europe and South Africa are considerably smaller (around 2%) but still statistically significant. We note that since the vast majority of the observations of the 5-minute time-series of jumps are zero, high correlations are not to be expected and even very small correlations can be viewed as economically meaningful.

Price jumps on European markets are also significantly correlated with price jumps on American markets, especially with the U.S. (correlations around 7.5%), but also with Brazil, Canada, and Mexico (correlations in the range of 1-7%). Within the American region, we also observe several significant correlations in price jumps across different markets, though the economic magnitude of the correlations is more modest (up to 4%). Co-jumps in prices across markets in Asia are not a prominent phenomenon, with the notable exception of Hong Kong and Malaysia, which exhibit a significant correlation in price jumps of almost 10%. There is little evidence of co-jumps in prices across markets in Europe and Asia.

All in all, we find that 21 out of the 46 market-pairs in our sample exhibit significantly (at the 1% level) positive correlations in price jumps at the 5-minute frequency. We view this as evidence that, even at a very high-frequency, shocks to prices show economically meaningful spillover effects across equity markets around the world.

In contrast, Panel B of Table 2.7 shows almost no significant correlations in 5-minute jumps in $PQSPR$ across individual markets within and across regions. The exceptions are the correlations between $PQSPR$ jumps in Canada and the U.S. and between $PQSPR$ jumps in France and the U.K.. Both of these correlations are statistically significant, but at around 1% they are considerably smaller than the price jumps correlations in Panel A. These results confirm the conclusion from Table 2.6 that "sudden liquidity dry-ups" or "liquidity black holes" are mainly local phenomena that do not tend to spill over to other markets within or across regions.

The correlations between jumps in OIB across different markets presented in Panel C of Table 2.7 show a similar pattern as the price jump correlations in Panel A, although both economic and statistical significance are somewhat weaker. 14 out of the 46 market-pairs in our sample show significantly positive correlations. Jumps in OIB are significantly correlated within the European/African region and between developed Europe and the U.S., while – like price jumps – OIB jumps are only weakly correlated within the Asian region and across Europe/Africa and Asia. Although prior studies have identified links between shocks to prices on different equity markets, we believe we are the first to document that shocks to order imbalance can also be propagated across international equity markets at a high-frequency.

2.3.5.2 Coinciding jumps in prices, liquidity, and trading activity across markets and variables

We now build upon the analyses in Tables 2.6 and 2.7 by not only studying coinciding jumps in the same variable within and across regions, but also examining whether the likelihood of a price jump on a particular market can be explained by jumps in other variables on the same market and on different markets in the same region as well as in other regions. In other words, we attempt to answer the question of how price shocks are propagated from one market to another, with a specific focus on microstructure variables.

We adopt the method proposed by Bae, Karolyi, and Stulz (2003) and estimate logit models to explain the occurrence of price jumps on each individual market at the 5- minute frequency. The results are in Table 2.8. As dependent variable, we use an indicator variable of whether there was a price jump on a particular market i in a particular 5-minute interval. All of our logits are estimated separately for negative and positive price jumps, to allow for asymmetric effects depending on the sign of the jumps. As independent variables, we use an indicator variable of same-sign OIB

jumps on market i in the same 5-minute interval, indicator variables of whether at least one other market in the same region (labeled "not i" in Table 2.8) has a same-sign jump in prices or in OIB in the same 5-minute interval, and indicator variables of whether at least one market in a different region has a same-sign jump in prices or in OIB in the same 5-minute interval. Since the independent variables based on different markets than market i are only defined during overlapping trading hours, we only include indicator variables of jumps in prices and OIB in Europe/Africa in the logits explaining price jumps on American markets and on Asian markets, while jumps in prices and OIB in both America and Asia serve as independent variables in the logits for price jumps on European markets. Since our results so far indicate little role for liquidity in the occurrence and spillovers of price jumps, we exclude $PQSPR$ jumps from the logit models.[18]

Table 2.8 presents the marginal effects (in %) of the logit models, organized by region (Panel A: America; Panel B: Asia; Panel C: Europe/Africa) and by the sign of the price jumps within each panel (Part I: positive; Part II: negative). Bold numbers are significant at the 10% level or better. For each market in each region, we estimate one, two, or three logit models, depending on the number of regions with overlapping trading hours with that market. The first model includes only independent variables from the same region. The second and third models also include independent variables from one or two other regions – if there is any overlap in the trading hours. We note that the number of observations available for the estimation of the second and third models is substantially reduced relative to the first model.[19]

We hypothesize that the probability of negative (positive) price jumps on market

[18]In reported tests, we do include $PQSPR$ jumps in the logit models, but find that they can often not be estimated because of "separation problems" in the estimation. Put differently, if one of the independent variables could almost perfectly explain jumps in prices on market i, then numerically we observe fitted probabilities equal to either 0 or 1, which results in unreliable model estimation. For instance, if positive jumps in prices on market i never coincide during the same 5-minute interval with positive jumps in $PQSPR$ from another region, then having an indicator variable for positive jumps in $PQSPR$ from another region equal to 1 guarantees no positive jumps in prices on market i during that interval.

[19]We generally estimate two models for markets in America, one model for markets in Asia, and three models for markets in Europe/Africa, but have to discard some individual models for individual markets in case there is a separation problem in the estimation.

Table 2.8. Logit models to explain 5-minute jumps in prices

This table shows marginal effects (in %) of logit models to explain the occurrence of jumps in 5-minute equally-weighted market returns (***PRICE***) for each of the 12 equity markets in our sample over 1996-2011. As dependent variable, we use an indicator variable of whether there was a price jump on a particular market ***i*** in a particular 5-minute interval. As independent variables, we use an indicator variable of same-sign jumps in 5-minute market aggregate order imbalance scaled by aggregate market capitalization (***OIB***) on market ***i*** in the same 5-minute interval, indicator variables of whether at least one other market in the same region (labeled "not ***i***") has a same-sign jump in ***PRICE*** or in ***OIB*** in the same 5-minute interval, and indicator variables of whether at least one market in a different region has a same-sign jump in price or in ***OIB*** in the same 5-minute interval. Independent variables are defined only when at least one of the markets in region has overlapping opening hours with market ***i***. We cannot include American and Asian markets in the same model since there is no overlap in trading hours. Some of the independent variables are omitted from the model specification due to a separation problem in the estimation. Numbers in bold font indicate statistical significance at 10% level and less. The markets are grouped by region: Panel A presents the results for the American markets, Panel B for the Asian markets, and Panel C for the European markets. All logits are estimated separately for negative price jumps (Part I of each panel) and positive price jumps (Part II of each panel). Markets are listed in alphabetical order. We refer to the caption of Table 2.2 and to Appendix A.2 for a detailed description of the jump statistics. Data are from TRTH, the World Bank website, and Datastream.

Panel A: Logit models to explain 5-minute price jumps on American markets

Part I: Positive jumps in *PRICE*

Market *i*:	Brazil		Canada		Mexico		U.S.	
Model:	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
OIB POS i	0.34	**-0.03**	**1.86**	0.41	3.38	8.85	**10.89**	**5.36**
PRICE POS not i	0.26	0.04	0.02	0.02	0.60	0.37	0.09	**-0.05**
OIB POS not i	0.28	0.05	0.41	1.49	**-0.08**	**-0.16**	0.31	0.30
PRICE POS Europe/Africa		2.05		0.10		1.28		0.75
OIB POS Europe/Africa		0.09		**1.11**		0.43		**0.82**
# Obs.	97,109	26,446	279,646	69,864	137,766	31,579	279,291	69,778

Part II: Negative jumps in *PRICE*

Market *i*:	Brazil		Canada		Mexico		U.S.	
Model:	(1)		(1)	(2)	(1)		(1)	(2)
OIB NEG i	0.23		0.83	2.35	4.07		1.97	0.63
PRICE NEG not i	0.28		**1.79**	1.57	**-0.06**		**1.48**	0.97
OIB NEG not i	0.32		0.36	0.44			0.13	0.14
PRICE NEG Europe/Africa				0.31				**0.63**
OIB NEG Europe/Africa				0.03				**1.84**
# Obs.	97,109		279,646	69,864	137,766		279,291	69,778

Table 2.8. Logit models to explain 5-minute jumps in prices (continued)

Panel B: Logit models to explain 5-minute price jumps on Asian markets

Part I: Positive jumps in *PRICE*

Market *i*:	Hong Kong	India	Japan	Malaysia
Model:	(1)	(1)	(1)	(1)
OIB POS i	**6.10**	2.71	**18.85**	1.48
PRICE POS not i	**4.51**	0.25	**-0.43**	2.81
OIB POS not i	**-0.30**	**-0.01**	**-0.43**	**-0.11**
# Obs.	128,250	90,743	71,776	179,525

Part II: Negative jumps in *PRICE*

Market *i*:	Hong Kong	India	Japan	Malaysia
Model:	(1)	(1)	(1)	(1)
OIB NEG i	**7.69**	3.20	**41.64**	0.84
PRICE NEG not i	**7.67**		1.15	**4.73**
OIB NEG not i	-0.17	**-0.03**	0.63	-0.04
# Obs.	128,250	90,743	71,776	179,525

Table 2.8. Logit models to explain 5-minute jumps in prices (continued)

Panel C: Logit models to explain 5-minute price jumps on European/African markets

Part I: Positive jumps in *PRICE*

Market *i*:	France			Germany			South Africa	U.K.		
Model:	(1)	(2)		(1)	(2)	(3)	(1)	(1)	(2)	(3)
OIB POS i	**2.16**	**0.80**		**1.74**	2.21	7.74	0.16	**3.10**	3.12	**-0.04**
PRICE POS not i	**2.42**	**2.03**		**2.45**	2.61	0.34	0.53	**2.58**	**2.37**	0.30
OIB POS not i	**0.55**	**0.71**		**0.29**	0.35	0.24	0.03	**0.48**	0.18	1.06
PRICE POS America		0.17			0.19				**0.81**	
OIB POS America		**0.82**			**-0.06**				0.07	
PRICE POS Asia										**-0.04**
OIB POS Asia						**-0.05**				
# Obs.	364,975	78,574		320,337	72,869	72,582	25,1708	361,716	78,455	71,791

Part II: Negative jumps in *PRICE*

Market *i*:	France			Germany			South Africa	U.K.		
Model:	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(1)	(2)	(3)
OIB NEG i	**1.19**	**1.92**	1.61	**1.60**	0.60	2.22	**-0.04**	**0.53**	0.60	2.98
PRICE NEG not i	**2.32**	**2.74**	0.69	**3.58**	**5.44**	1.96	0.74	**2.25**	**3.27**	0.66
OIB NEG not i	**0.44**	0.41	0.96	**0.60**	**0.67**	0.40	0.24	**0.50**	0.22	0.91
PRICE NEG America		0.04			**0.95**				0.17	
OIB NEG America		**-0.05**			0.08				**-0.04**	
PRICE NEG Asia			2.62							**-0.07**
OIB NEG Asia			**-0.07**			**-0.05**				0.77
# Obs.	364,975	78,574	73,007	32,0337	72,869	72,582	251,708	361,716	78,455	71,791

i increases with negative (positive) jumps in OIB on the same market and with negative (positive) jumps in prices and OIB on other markets in the same and in other regions. In other words, the marginal effects in Table 2.8 are all expected to be positive.

The results of the logit models in Table 2.8 are consistent with our findings in Table 2.3 that price jumps on a particular market are linked to OIB jumps of the same sign on the same market in the same 5-minute interval. For 12 out of the 24 cases (negative and positive price jumps on 12 markets), we find a positive and significant marginal effect of OIB jumps on market i (based on the first logit model for each market). These effects are often economically substantial, especially for markets in Asia and Europe: they vary from 6.10% (positive price jumps in Hong Kong) to 41.64% (negative price jumps in Japan) in Asia and from 0.53% (negative price jumps in the U.K.) to 3.10% (positive price jumps in the U.K.) in Europe. In only two cases (Brazil and South Africa) do we observe significantly negative marginal effects of OIB jumps on the same market, but, at -0.03% and -0.04%, their economic magnitude is small.

Table 2.8 also confirms the results of the correlation analysis in Table 2.7. In particular, price jumps on other markets in the same region significantly increase the probability of a price jump on market i in 11 out of the 24 cases (based on the first logit model for each country). These effects are observed in all regions. For instance, price jumps on the other markets within the Europe/Africa region have positive and significant marginal effects on price jumps on market i varying from 2.25% to 3.58%. Only in two cases (Mexico and Japan) do we observe significantly negative marginal effects of jumps in prices on the other markets within the same region, but their economic magnitude is relatively small.

The results on the effect of OIB jumps on other markets in the same region on price jumps on market i are mixed for America and Asia. If anything, the significant marginal effects for this variable in Panels A and B suggest that price jumps on a particular market are associated with OIB jumps of the opposite sign on other markets in the same region. In contrast, for the three European markets in our sample, there is consistent evidence that the probability of price jumps on one market is positively related to same-sign OIB jumps on other markets in the same 5-minute interval. These marginal effects are all positive and significant within developed Europe, but are relatively modest, ranging from 0.29% to 0.60%.

The second and third logit models for each market in Table 2.8 assess cross-region spillovers of jumps in prices and OIB. Perhaps not surprisingly, the evidence for cross-region spillovers of price jumps is weaker and less consistent than for within-region spillovers. The marginal effect of price jumps in other regions is positive and significant only in few cases: for negative price jumps in the U.S. vis-à-vis the European/African region (marginal effect of 0.63%, see Panel A), for positive price jumps in the U.K. vis-à-vis the American region (0.81%, Panel C), and for negative price jumps in Germany vis-à-vis the American region (0.95%, Panel C). Similarly, in most cases, the effect of OIB jumps in other regions on price jumps on market i is not significant either in statistical or in economic terms, except for price jumps in Canada and the U.S. vis-à-vis OIB jumps in the European/African region (marginal effect between 0.82% and 1.84%, see Panel A) and for positive price jumps in France vis-à-vis OIB jumps in America (effect of 0.82%, Panel C). Some of the marginal effects in Table 2.8 are not in line with expectations. For example, the marginal effect of OIB jumps in Asia on the likelihood of positive price jumps in Germany is -0.05% (Panel C). Although some of these exceptions are statistically significant, their economic magnitude is small.

In sum, the results in Table 2.8 highlight that shocks to prices can be propagated from one market to another within a 5-minute horizon. Such propagation is especially strong across markets within the same region, although some cross-region effects are also observed. Furthermore, price jumps are regularly linked to same-sign OIB jumps on the same market, and, for Europe, also to same-sign OIB jumps on other markets in the same region.

To address the question whether the high-frequency propagation of shocks to equity prices across markets is driven by liquidity or by information, we repeat the price reversal analysis from Section 2.3.4.1 for the subsets of jumps in prices that only occur on one market and that occur simultaneously on at least two of the markets in our sample. Of the 2348 (2521) positive (negative) price jumps in our sample, 200 (253) occur simultaneously with a price jump on at least one other market. Unreported results show that the price reversal graphs are very similar for both subsets of price jumps. In other words, there is no evidence that price jumps that occur simultaneously in multiple markets exhibit reversals, consistent with the hypothesis that these jumps are primarily driven by information rather than liquidity.

2.3.6 Supplementary tests

Our analyses so far suggest that shocks to equity prices are prevalent and large, are linked to shocks to order imbalance, exhibit regular high-frequency spillovers across international markets, and are mainly driven by information rather than liquidity. In this section, we discuss the results of a number of supplementary tests that we carried out to evaluate the robustness of these conclusions.

2.3.6.1 Alternative frequencies

One potential limitation of our study is that we measure jumps in prices, liquidity, and trading activity at a relatively high frequency: 5-minute intervals within the trading day. Our choice for this frequency was motivated by our aim of a detailed, intraday analysis of the dynamics of liquidity and trading activity around financial market shocks and by issues concerning non-overlapping trading hours, overnight returns, and special features of the opening session that arise in analyses of cross-market spillovers at the daily frequency. Nonetheless, we repeat all of our analyses at the 15-minute and 1-hour frequencies to assess whether we may have missed lower-frequency shocks, or lower-frequency relations between shocks to prices, liquidity, and trading activity (results available from the authors). At these lower frequencies, the number of jumps is naturally smaller, but we still find quite frequent jumps in prices, quoted spreads, and order imbalance, while jumps in effective spreads are rare; we do now also observe regular jumps in turnover. The economic magnitudes of the jumps are still around five jump-free standard deviations. Similar to Table 2.3, we find virtually no evidence that price jumps are associated with $PQSPR$ jumps (nor with turnover jumps). In contrast to Table 2.3, we no longer observe a significant relation between price jumps and same-sign OIB jumps in the same interval at the 15-minute and 1-hour frequencies, which underlines the value of using a relatively high frequency to study the role of liquidity and trading activity around financial market shocks. Similar to Figure 2.1 and Table 2.5, we find that price jumps are not followed by reversals, and that a substantial fraction of price jumps occur around macroeconomic news announcements. Both pieces of evidence suggest that 15-minute and 1-hour price jumps are also primarily driven by information rather than liquidity. These price jumps also regularly spill over across markets, consistent with Tables 2.6 and 2.7. There is no evidence of spillovers in $PQSPR$ jumps to other markets, but both jumps in OIB and in turnover exhibit significant correlations across markets, especially for Europe. In short, we conclude that our inferences are not materially affected by redoing our analyses at a lower frequency within the trading day.

2.3.6.2 The dynamics of liquidity and trading activity around price jumps

Table 2.3 shows that price jumps are regularly associated with same-sign jumps in OIB in the same 5-minute interval, but not with $PQSPR$ jumps. However, it is possible that this result is affected by our approach to identify shocks to liquidity and order imbalance as discontinuous jumps. To obtain a broader picture of the behavior of liquidity and trading activity around price jumps, Figure 2.2 shows the cumulative change in $PQSPR$ (Panels A and B; in %) as well as the dynamics of OIB (Panels C and D; in basis points) from one hour before until one hour after positive and negative jumps in prices, aggregated across all jumps on the 12 markets in our sample. Panels A and B show that liquidity does fluctuate around price jumps; quoted spreads tend to fall slightly in the hour before a price jump, followed by a small upward blip in the 5-10 minutes before the price jump, and a more pronounced decline after the price jump. Nonetheless, the observed patterns seem hard to square with theories that propose a key role for liquidity in the origination of price shocks. First, the quoted spread effects are small. The blip in $PQSPR$ just before the price jump has a magnitude of 4-5 percentage points, which is much smaller than the average $PQSPR$ jump of around 42%. Second, liquidity tends to improve following a price jump, but

Figure 2.2. Behavior of *PQSPR* and *OIB* around price jumps

This figure shows the cumulative changes in 5-minute equally-weighted proportional quoted spreads ($PQSPR$), and 5-minute market aggregate order imbalance scaled by aggregate market capitalization (OIB) (averaged across all the price jumps in the 12 equity markets) from one hour before till one hour after either positive or negative jumps in price over 1996-2011, respectively. Panel A and Panel B present average cumulative changes in $PQSPR$ around positive and negative price jumps in our sample, while Panel C and Panel D present average OIB around positive and negative price jumps in our sample. Cumulative changes in $PQSPR$, OIB are plotted for each 5-minute interval in the event window. We refer to the caption of Table 2.2 and to Appendix A.2 for a detailed description of the jump statistics. Data are from TRTH, the World bank website, and Datastream.

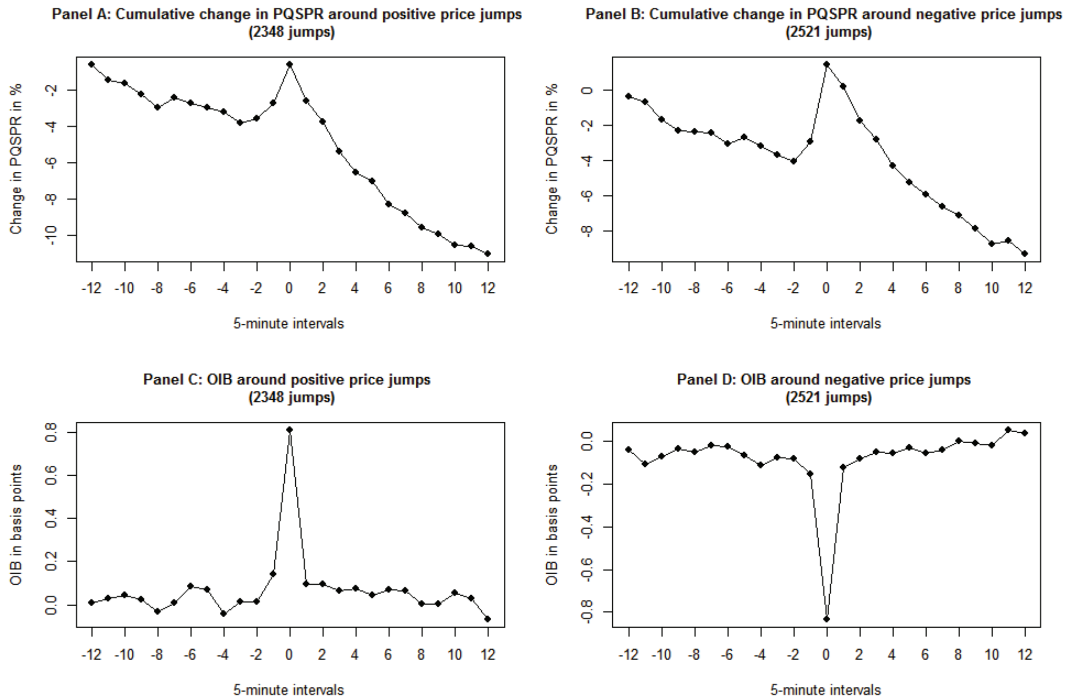


Figure 2.1 shows no evidence of any accompanying price reversal. Third, the liquidity patterns around price jumps are very similar for positive and negative price jumps. Fourth, there is little indication of "liquidity black holes" or "liquidity spirals" in the sense that feedback effects cause liquidity crashes to worsen over time. Rather, the observed patterns in quoted spreads in Figure 2.2 seem to accord well with an increase in adverse selection costs just before the arrival of economic news, and the resolution of asymmetric information following the news arrival.[20] Panels C and D of Figure 2.2 show a clear, once-off spike in OIB in the same direction as the price jump, in the same interval. Given the absence of reversals after price jumps, this pattern seems more consistent with speculative trading or portfolio rebalancing around the arrival of news than with temporary price pressure effects or with feedback loops in which initial price drops induce further selling.

2.3.6.3 Alternative liquidity measures

Liquidity is a multi-faceted concept and the liquidity measures used in this paper ($PQSPR$ and $PESPR$) may not cover all relevant aspects of liquidity. However, we would expect significant shocks to liquidity to also be reflected in quoted or effective spreads. Moreover, price impact measures suffer from large estimation errors at high frequencies and could be mechanically linked to price changes. In unreported analyses, we redo all of our analyses with an alternative liquidity measure based on the number of different stocks that traded on a specific market in a specific interval. This trade-based liquidity measure builds on the premise that stocks may not trade in a certain interval in part because of high trading costs. The trade-based measure should thus be positively associated with the level of market liquidity. Indeed, this measure is highly, but not perfectly, correlated with the illiquidity measures $PQSPR$ and $PESPR$; for most markets, these correlations roughly range from -0.4 to -0.7. On

[20] To the extent that the $PQSPR$ patterns in Panels A and B of Figure 2.2 are driven by macroeconomic news announcements, one might wonder about the scope for asymmetric information around such announcements. However, we note that Boudt and Petitjean (2014) document significant increases in trading costs and the demand for immediacy around macroeconomic news releases and price jumps in Dow Jones stocks. And Jiang, Lo, and Valente (2014) find that high-frequency trading adversely affects liquidity around macroeconomic news announcements. Alternatively, the blip in quoted spreads in Figure 2.2 could be due to concerns about inventory risk instead of adverse selection.

average, jumps in the trade-based measures are about as frequent as $PQSPR$ jumps, and are also of similar magnitude. Similar to Table 2.3, we find little evidence that price jumps coincide with jumps in the trade-based liquidity measure. Similar to Tables 2.6 and 2.7, jumps in the trade-based measure are almost always isolated events that do not spillover to other markets. Our conclusion that liquidity does not play more than a minor role the origination and propagation of prices shocks is thus not sensitive to the use of this alternative liquidity measure.

Several studies that model liquidity supply channels for the origination and propagation of financial market shocks feature an important role of funding liquidity. For example, in Brunnermeier and Pedersen (2009), "liquidity spirals" arise when financial intermediaries reduce the supply of liquidity in response to worsening funding liquidity (e.g., increasing margins) and when funding liquidity, in turn, is decreasing in market illiquidity. To more specifically test the implications of these liquidity supply channels, in Table 2.9, we examine whether shocks to prices, liquidity, and trading activity are associated with shocks to funding liquidity. This table assesses whether jumps in prices, $PQSPR$, and OIB are more likely to occur on days with a jump in the TED spread (the difference between the 3-month LIBOR and the 3-month T-bill rate, obtained from the Federal Reserve Bank of St. Louis), which is a common proxy for funding liquidity.[21] TED spread jumps are measured at the daily frequency, since funding liquidity may not be likely to exhibit sudden intraday changes, since one component of the TED spread (the LIBOR) is determined only once per day, and since we lack intraday data on the other component (the T-bill).

Panels A and B of Table 2.9 document the number and empirical frequency of days with negative price jumps, positive $PQSPR$ jumps, negative OIB jumps, and positive TED spread jumps for each of the 12 markets. We focus on jumps with these signs since theories on funding liquidity primarily associate a drop in funding liquidity with a drop in prices, a worsening of liquidity, and securities sales. The TED spread exhibits a positive jump on only 9 days over the entire sample period 1996-2011.

[21] We obtain similar results when, instead of the TED spread, we use the U.S. default spread (Baa-Aaa), the LIBOR, country-specific short-term interest rates, or country-specific banking industry index returns as proxies for funding liquidity.

Table 2.9. Jumps in the *TED* spread and jumps in prices, *PQSPR*, and *OIB*

This table shows the number of days with the jumps (Panel A), the probability of observing a day with a jump (Panel B), the number of days with coinciding jumps (Panel C), and the *p*-values of the test of whether the empirical probability is greater than the theoretical probability (Panel D) for negative jumps in ***PRICE***, positive jumps in ***PQSPR***, negative jumps in ***OIB***, and positive jumps in the ***TED*** spread. Jumps in ***PRICE***, ***PQSPR***, and ***OIB*** are computed at the 5-minute frequency on a daily basis, while jumps in the ***TED*** spread are computed at the daily frequency on a yearly basis. We refer to the caption of Table 2.2 and to Appendix A.2 for a detailed description of the jump statistics. Data are from TRTH, the World Bank website, Datastream, and Federal Reserve Bank of St. Louis.

	America				Asia				Europe/Africa			
	Brazil	Canada	Mexico	U.S.	Hong Kong	India	Japan	Malaysia	France	Germany	South Africa	U.K.
Panel A: Number of days with jumps												
PRICE < 0	37	107	69	139	411	56	552	177	186	172	115	181
PQSPR > 0	6	175	97	37	34	36	181	155	24	76	196	235
OIB < 0	153	256	18	71	222	154	121	195	473	192	384	126
TED > 0	6	9	6	9	9	3	9	9	9	9	9	9
# Obs.	1,441	3,967	1,960	3,952	3,665	2,560	3,856	3,892	4,011	3,401	3,176	3,990
Panel B: Probability of jumps in percentage												
PRICE < 0	2.57	2.70	3.52	3.52	11.21	2.19	14.32	4.55	4.64	5.06	3.62	4.54
PQSPR > 0	0.42	4.41	4.95	0.94	0.93	1.41	4.69	3.98	0.60	2.23	6.17	5.89
OIB < 0	10.62	6.45	0.92	1.80	6.06	6.02	3.14	5.01	11.79	5.65	12.09	3.16
TED > 0	0.42	0.23	0.31	0.23	0.25	0.12	0.23	0.23	0.22	0.26	0.28	0.23
Panel C: Coinciding jumps in the TED spread and either PRICE, PQSPR or OIB												
PRICE < 0; *TED* > 0	0	0	0	0	1	0	1	0	1	0	0	0
PQSPR > 0; *TED* > 0	0	0	1	0	0	0	0	2	0	1	0	1
OIB < 0; *TED* > 0	1	0	0	1	0	0	1	0	0	0	0	0
Panel D: *P*-value (empirical probability of coinciding jumps > theoretical probability under independence assumption)												
PRICE < 0; *TED* > 0	1.00	1.00	1.00	1.00	0.64	1.00	0.72	1.00	0.34	1.00	1.00	1.00
PQSPR > 0; *TED* > 0	1.00	1.00	0.26	1.00	1.00	1.00	1.00	**0.05**	1.00	0.18	1.00	0.41
OIB < 0; *TED* > 0	0.47	1.00	1.00	0.15	1.00	1.00	0.25	1.00	1.00	1.00	1.00	1.00

We note that the number of TED spread jumps reported in the table differs across countries because not all of these jumps occur within the available sample period for all countries, as the TRTH data coverage for some countries starts later than 1996 (see Appendix A.1.1). Panel C shows that negative price jumps, positive $PQSPR$ jumps, and negative OIB jumps almost never coincide with positive TED spread jumps on the same day. Panel D presents the results of a test of whether the empirical probability of such coinciding jumps is greater than the theoretical probability under the assumption that the jumps in the individual variables are independent (similar to Table 2.4). The results in this panel indicate very little evidence that shocks to prices, liquidity, and trading activity are associated with shocks to funding liquidity.

2.4 Conclusion

The recent financial crisis has highlighted the importance of global systemic risk in the current environment of globally integrated financial markets and fast trading technology. We conduct a study of the intraday propagation of shocks across 12 equity markets around the world at the 5-minute frequency over 1996-2011 – with a particular focus not only on shocks to prices, but also on shocks to liquidity (quoted and effective spreads) and trading activity (turnover and order imbalance). Our main purpose is to test the liquidity vs. information channels for the origination and propagation of financial market shocks.

Our findings are based on jump statistics in these five variables at the 5-minute frequency and can be summarized as follows. First, jumps in prices, quoted spreads, and order imbalance are large and occur much more often than jumps in effective spreads and turnover. Second, we document a significant association between jumps in prices and in order imbalance, while jumps in quoted spreads are independent from jumps in the other variables. Third, we show that jumps in prices and simultaneous jumps in prices and order imbalance are primarily driven by information rather than liquidity. Fourth, jumps in prices and order imbalance exhibit significant spillover effects across markets (even in the same 5-minute interval and especially for markets in Europe and the U.S.), but spillovers of jumps in quoted spreads to other markets are rare.

To sum up, our study provides evidence that the propagation speed of shocks across international equity markets is very high. In designing optimal financial regulation and risk management, policy makers and investors should not neglect microstructure effects related to the occurrence of price shocks. In particular, price shocks should not be viewed independently from shocks to trading activity. Shocks to liquidity, however, seem to play a less central role in the origination and propagation of price shocks than previously thought.

We leave further analyses of the speed and mechanism of the propagation of price shocks across markets for future research. In particular, recent advances in trading technology suggest that, in the later years of our sample period, the propagation of shocks across markets may take place at an even higher frequency than the one studied in this paper. Moving to a higher frequency of analysis would also allow for the estimation of daily vector autoregressions to get a better handle on causality, but will likely limit the sample to developed markets in recent years in order to construct meaningful measures of trading activity over such ultra-short horizons. Another potential extension would be to broaden the scope of the analysis beyond the 12 markets in our sample, which would enable an analysis of the determinants of the speed and the strength of the propagation of stocks across different (pairs of) markets.

Chapter 3

Intraday Return Predictability, Informed Limit Orders, and Algorithmic Trading*

3.1 Introduction

The limit order book is the dominant market design in equity exchanges around the world.[1] The prevalence of limit order book markets calls for a detailed understanding of how such markets function. In particular, understanding the price discovery process on these markets required a detailed study of the trader's choice between submissions of market and limit orders. The conventional wisdom in the microstructure literature used to be that informed traders use only market orders, while uninformed traders use both market and limit orders (for theoretical work see Glosten and Milgrom (1985); Kyle (1985); Glosten (1994); Seppi (1997)). Only recent studies explicitly consider the choice of informed traders for market or limit orders.[2]

*This chapter is based on Yuferova (2016) "Intraday Return Predictability, Informed Limit Orders, and Algorithmic Trading" (available at http://ssrn.com/abstract=2686082). I am grateful to Dion Bongaerts, Mathijs Cosemans, Sarah Draus, Thierry Foucault, Wenqian Huang, Lingtian Kong, Albert Menkveld, Marco Pagano, Christine Parlour, Loriana Pelizzon, Dominik Rösch, Stephen Rush, Asani Sarkar, Elvira Sojli, Mark Van Achter, Mathijs van Dijk, Wolf Wagner, Jun Uno, Marius Zoican, participants of the FMA 2015 Doctoral Consortium, participants of the PhD course on "Market Liquidity" in Brussels, and seminar participants at Erasmus University, NYU Stern, and Tinbergen Insitute for helpful comments. I gratefully acknowledge financial support from the Vereniging Trustfonds Erasmus Universiteit Rotterdam. I am also grateful for the hospitality of NYU Stern, where some work on this paper was carried out. This work was carried out on the National e-infrastructure with the support of SURF Foundation. I thank OneMarket Data for the use of their OneTick software.

[1] According to Swan and Westerholm (2006), 48% of the largest equity markets are organized as pure limit order book markets (e.g., Australian Stock Exchange, Toronto Stock Exchange, Tokyo Stock Exchange), 39% are organized as limit order books with designated market makers (e.g., New York Stock Exchange, Borsa Italiana), and the remaining 12% are organized as hybrid dealer markets (e.g., NASDAQ, Sao Paulo Stock Exchange) as of the beginning of 2000.

[2] For theoretical studies on the choice of uninformed traders between market and limit orders, see Cohen, Maier, Schwartz, and Whitcomb (1981), Chakravarty and Holden (1995), Handa and Schwartz (1996), Parlour (1998), Foucault (1999), Foucault, Kadan, and Kandel (2005), Goettler, Parlour, and Rajan (2005), and Roşu (2009); for theoretical studies on the choice of informed traders between market and limit orders see, Liu and Kaniel (2006), Goettler, Parlour, and Rajan (2009), and Roşu (2016); for empirical studies on the choice between market and limit orders on equity

Informed traders can submit a market order and experience immediate execution at the expense of the bid-ask spread (consume liquidity). Alternatively, informed traders can submit a limit order and thus bear the risk of non-execution, as well as the risk of being picked-off, but earn the bid-ask spread (provide liquidity).

The importance of the informed trader's choice between market and limit orders is emphasized by a heated public debate about whether one group of market participants poses negative externalities to another group of market participants due to informational asymmetries. This informational advantage is especially pronounced for traders with superior technologies for the collection and processing of information. Another feature that enhances informational inequality in the market is the ability to continuously monitor and respond to market conditions. Both characteristics are distinct characteristics of high-frequency traders (a subset of algorithmic traders). Consistently, several papers identify algorithmic traders' strategies that are disadvantageous for retail investors.[3] Previous research has focused on informed algorithmic trading via market orders with only one exception.[4] In sum, understanding *how* informed trading takes place and *what role* algorithmic traders play in this process are important questions to explore in modern market microstructure.

In this paper, I address these questions by studying intraday return predictability. Naturally, orders submitted by informed traders contain information about future price movements. If an informed trader actively uses market orders, an imbalance between buyer- and seller-initiated volume may be informative about future price movements. If an informed trader actively uses limit orders, the limit order book may contain information that is not yet incorporated into the price. Therefore, strategies employed by informed traders may induce intraday return predictability

markets see, Bae, Karolyi, and Stulz (2003), Anand, Chakravarty, and Martell (2005), Bloomfield, O'Hara, and Saar (2005), and Baruch, Panayides, and Venkataraman (2016); for empirical studies on the choice between market and limit orders on foreign exchange markets see, Menkhoff, Osler, and Schmeling (2010), Kozhan and Salmon (2012), and Kozhan, Moore, and Payne (2014).

[3]See for theoretical work, e.g., Biais, Foucault, and Moinas (2015); Foucault, Kozhan, and Tham (2015); Jovanovic and Menkveld (2015); Foucault, Hombert, and Roşu (2016); see for empirical work, e.g., Hirschey (2013); McInish and Upson (2013); Brogaard, Hendershott, and Riordan (2014); Foucault, Kozhan, and Tham (2015).

[4]Brogaard, Hendershott, and Riordan (2015) examine informed trading via both market and limit orders by high-frequency traders for the sample of 15 Canadian stocks from October 2012 to June 2013.

from market and limit order flows alike.

My main contribution to the literature is twofold. First, I contribute to the literature on intraday return predictability. I distinguish between two sources of intraday return predictability (inventory management and private information). My findings indicate that the main source of the intraday return predictability is private information embedded in limit orders. Furthermore, I show that this result holds for a wide cross-section of stocks and through a prolonged time period.[5] Second, my paper contributes to the ongoing debate on the role of algorithmic traders (especially it subset, high-frequency traders) in informed trading activity (see Biais and Foucault (2014) for review on high-frequency trading activity and market quality). My evidence suggests that an increased degree of algorithmic trading activity leads to an increased usage of both informed limit and informed market orders (with the main effect concentrated in market orders). Informed limit orders still remain the main source of the intraday return predictability even after increased degree of algorithmic trading activity.

The analysis is organized in two stages. First, I analyze intraday return predictability from market and limit order flows and separate the effect of informed trading from the effect of inventory management. Second, I analyze the impact of algorithmic trading on the choice between market and limit orders made by an informed trader. In particular, I exploit a quasi-natural experiment to establish a causal inference between algorithmic trading and intraday return predictability from market and limit order flows. I also test recent theories of the choice between informed trading through market versus limit orders by exploiting their predictions regarding differences between low and high volatility stocks.

Using tick-by-tick trade data and data on the first 10 best levels of the consolidated limit order book for the NYSE from the Thomson Reuters Tick History (TRTH) database, I construct a time series of mid-quote returns, market order im-

[5] For papers studying intraday return predictability from the limit order book in equity markets see Irvine, Benston, and Kandel (2000), Kavajecz and Odders-White (2004), Harris and Panchapagesan (2005), Cao, Hansch, and Wang (2009), Cont, Kukanov, and Stoikov (2013), and Cenesizoglu, Dionne, and Zhou (2014). However, none of these papers uses such comprehensive data as used in this paper.

balance, and snapshots of the first 10 best levels of the U.S. consolidated limit order book at the one-minute frequency at the individual stock level. The sample covers all NYSE-listed common stocks for the years 2002-2010. TRTH data used in this paper are very comprehensive. In particular, for the stocks under consideration, I have information for 1.36 billion trades and 8.54 billion limit order book updates.

Intraday return predictability from limit order book data can arise from two sources. First, inventory management (Hypothesis 1) may induce intraday return predictability by generating price pressure as a result of limited risk-bearing capacity of risk-averse liquidity providers (e.g., Stoll (1978); Menkveld (2013); Hendershott and Menkveld (2014)). Second, private information (Hypothesis 2) may also induce intraday return predictability (see Liu and Kaniel (2006); Goettler, Parlour, and Rajan (2009); Roşu (2016)). The latter source of return predictability is the main focus of this paper. I approach the problem of isolating private information source of intraday return predictability from two angles. First, inventory management should result in temporary price effects, while private information should result in permanent price effects. Therefore, controlling for lagged returns in predictive regressions allows me to separate inventory management effects from the effects of private information.

Second, I run a VAR model and decompose market and limit order flows into two components: inventory-related (fitted values) and information-related (surprises) components. The use of surprises as a proxy for informed market and limit order flows is motivated by the fact that both limit and market order flows are persistent (e.g., Hasbrouck (1991); Biais, Hillion, and Spatt (1995); Ellul, Holden, Jain, and Jennings (2003); Chordia, Roll, and Subrahmanyam (2005)) and that this persistence is attributable to reasons other than information (e.g., Degryse, de Jong, and van Kervel (2014)). Huang and Stoll (1997), Madhavan, Richardson, and Roomans (1997), and Sadka (2006) also use surprises in market order imbalance to isolate the adverse selection component of the bid-ask spread.

Combining these two approaches, I run the predictive regressions with lagged surprises in returns, lagged surprises in market order imbalance, and lagged surprises in depth concentration at the inner and outer levels of the ask and bid sides of the limit order book. In this specification any remaining inventory management effects should be captured by the coefficient of lagged surprises in returns. I use both market order flow and limit order book variables in the predictive regressions to capture the trader's choice between market and limit orders. Inclusion of market order imbalance is also motivated by Chordia, Roll, and Subrahmanyam (2005, 2008), who show that market order imbalance is predictive of future price movements.

The findings of the first part of the analysis indicate that the main source of intraday return predictability is private information (inventory management (lagged returns) accounts only for 30% of total predictive power as measured by the average incremental adjusted R^2 from the predictive regressions). In addition, the results indicate that informed trading through the limit order book accounts for 50% of return predictability that is 30% greater than a fraction of return predictability induced by informed trading through market orders.

The findings contradict the traditional view that only market orders are used for informed trading. Furthermore, the findings suggest that informed trading via market orders is of less importance than informed trading via limit orders.

In the second part of the analysis, I investigate how the presence of algorithmic traders affects the order choices made by informed traders. This is a non-trivial task as algorithmic traders endogenously determine the extent of their participation in each stock at each point in time. I follow the approach of Hendershott, Jones, and Menkveld (2011) and use the NYSE Hybrid Market introduction — a permanent technological change in market design[6] — as an instrumental variable to help determine the causal effects of algorithmic trading activity on intraday return predictability from informed market and limit order flows. The rollout to the Hybrid Market was implemented in a staggered way, which helps clean identification. I follow Hendershott, Jones, and Menkveld (2011) and Boehmer, Fong, and Wu (2015) and use the daily number of best bid-offer quote updates relative to the daily trading volume (in \$10,000) as a proxy for algorithmic trading activity on each stock-day.

I develop two competing hypotheses of the effects of algorithmic trading on in-

[6]NYSE Hybrid Market introduction allowed market orders to "walk" through the limit order book automatically and thus, increased automation and speed (Hendershott and Moulton (2011)).

formed traders' choices: the efficient technology hypothesis (Hypothesis 3) and the competition hypothesis (Hypothesis 4). On the one hand, the technological advantage of algorithmic traders makes limit orders more attractive to them as they are able to reduce pick-off risks better than the other market participants (the efficient technology hypothesis). On the other hand, competition between algorithmic traders for (trading on) the same information makes market orders more attractive to them as they guarantee immediate execution (the competition hypothesis).

The results show that algorithmic trading activity leads to increased informational content in both market and limit orders. However, an increase in the predictive power associated with limit order book variables (the efficient technology hypothesis) is smaller than the increase in predictive power associated with market order imbalance (the competition hypothesis). Although the evidence is consistent with both hypotheses, the effects of the competition hypothesis seem to dominate the effects of the efficient technology hypothesis. In other words, increased algorithmic trading activity is associated with a relative shift from liquidity provision (limit orders) to liquidity consumption (market orders) by informed traders.

Overall, my paper provides evidence that informed traders tend to act more often as liquidity providers (use limit orders), than liquidity demanders (use market orders). However, with an increased presence of algorithmic traders, the amount of informed liquidity provision increases less than the amount of informed liquidity consumption. One important implication of my analysis concerns measures of asymmetric information and/or informed trading (e.g., PIN measure by Easley, Kiefer, O'Hara, and Paperman (1996); adverse selection component of bid-ask spread by Glosten and Harris (1988) and Huang and Stoll (1997)), which have been used widely in studies on market microstructure, asset pricing, and corporate finance.[7] These measures are exclusively based on market orders, and thus neglect the lion's share of informed trading on the equity markets — informed trading via limit orders.

[7]E.g., Easley, Hvidkjaer, and O'Hara (2002), Vega (2006), Chen, Goldstein, and Jiang (2007), Korajczyk and Sadka (2008), Bharath, Pasquariello, and Wu (2009), and Easley, de Prado, and O'Hara (2012).

3.2 Hypotheses

In this section, I develop the hypotheses for the tests of the choice between limit and market orders by informed traders based on the evidence from intraday return predictability. In Section 3.2.1, I develop two hypotheses regarding the sources of intraday return predictability: the inventory management hypothesis and the private information hypothesis. In Section 3.2.2, I describe the hypotheses regarding the effect of algorithmic trading activity on the strategies employed by informed traders (the efficient technology hypothesis and the competition hypothesis). The effect of the realized volatility is described in Section 3.2.3.

3.2.1 Sources of intraday return predictability

Intraday return predictability from the limit order book can arise from two (not mutually exclusive) sources: inventory management and private information. Under the inventory management hypothesis, depth concentration at the inner levels of the limit order book indicates that a liquidity provider wants to unload inventory. This situation creates a temporary price impact that is reverted as soon as the inventory position of the liquidity provider is liquidated (e.g., Stoll (1978); Ho and Stoll (1981); Menkveld (2013); Hendershott and Menkveld (2014)). Indeed, a liquidity provider will be hesitant to immediately replenish the ask side of the limit order book as a large market buy order walks through the limit order book, because she would prefer to liquidate excessive inventory first. It is optimal for her to post aggressive limit orders on the bid side of the book, while on the ask side she will post a limit order deep in the limit order book. In this way, she encourages other market participants to sell her their stocks while discouraging them from buying from her. Therefore, I formulate the *inventory management hypothesis* as follows:

H1 (the inventory management hypothesis): Depth concentration at the inner levels of the ask (bid) sides of the limit order book is associated with decrease (increase) in future stock returns, with depth concentration at the outer levels having virtually no effect on future stock returns.

Under the traditional approach to the adverse selection problem in equity markets only inventory management should drive intraday return predictability from the limit order book. This approach is built under the assumption of informed traders only using market orders (e.g., Glosten and Milgrom (1985); Kyle (1985); Glosten (1994); Seppi (1997)), which may be an inadequate approximation of reality. Later studies build upon this initial work and allow both informed and uninformed traders to choose between the order types (Liu and Kaniel (2006); Goettler, Parlour, and Rajan (2009); Roşu (2016)).

Based on theoretical predictions from Goettler, Parlour, and Rajan (2009), an informed trader, who receives good news about a stock, has three different options to exploit this information. First, the trader can submit a buy market order. Second, the trader can submit a limit buy order at the inner level of the bid side of the limit order book; this limits execution probability, but saves transaction costs. Third, the trader can also submit a limit sell order at the outer levels of the ask side of the limit order book in combination with one of the two above mentioned orders to lock-in the benefit from the price difference. The opposite is true for the bad news scenario.

In reality, an informed trader's choice between market and limit orders depends on the strength of the signal received, the lifespan of the information, the ratio of informed to uninformed traders, etc. In the case of a weak and very short-lived signal, the trader is likely to use market orders. In the case of very strong signal that has a relatively long lifespan, the trader is likely to use limit orders at the inner and outer levels of the limit order book. In the case of the average signal with a short lifespan (which I believe is the dominant type of signal), the trader is likely to use a mixture of market and limit orders (see Table 3.1).

Therefore, I formulate the *private information hypothesis* as follows:

H2 (the private information hypothesis): Depth concentration at the inner levels of the ask side of the limit order book is associated with decrease in future stock returns, while depth concentration at the outer levels of the ask side of the limit order book is associated with increase in future stock returns. The opposite is true for the bid side of the limit order book.

The main purpose of this paper is to test the private information hypothesis and

Table 3.1. Expected signs of the coefficients for two sources of intraday return predictability

This table shows the expected behavior of informed trader conditional on the type of news received (Panel A) as well as variables and corresponding expected signs of the coefficients under private information and inventory management hypotheses in the following predictive regression of one-minute mid-quote return with lagged return, lagged market order imbalance (***MOIB***), and lagged depth concentration at the inner and outer levels of the ask and bid sides of the limit order book (***LOB***: ***Bid Inner***, ***Bid Outer***, ***Ask Inner***, ***Ask Outer***) as explanatory variables, see equation (3.1) (Panel B):

Panel A: Expected behavior of informed trader

Order type		Good news	Bad news
Market	Buy	×	
	Sell		×
Limit Bid	Inner levels	×	
	Outer levels		×
Limit Ask	Inner levels		×
	Outer levels	×	

Panel B: Expected signs under inventory management and private information hypothesis

Variable	Inventory Management	Private Information
Ret_{t-1}	NEG	NA
$MOIB_{t-1}$	POS	POS
$Bid\ Inner_{t-1}$	POS	POS
$Bid\ Outer_{t-1}$	NA	NEG
$Ask\ Inner_{t-1}$	NEG	NEG
$Ask\ Outer_{t-1}$	NA	POS

investigate the effect of algorithmic traders on the informed trader's choice between market and limit orders discussed in the next subsection.

3.2.2 Effect of algorithmic trading activity

During the past decade, a new group of market participants — algorithmic traders — has emerged and evolved into a dominant player responsible for the majority of trading volume. Algorithmic trading "is thought to be responsible for as much as 73 percent of trading volume in the United States in 2009" (Hendershott, Jones, and Menkveld (2011), p. 1). Therefore, it is a natural question to ask what role algorithmic traders are playing in informed trading process and to what extent their presence affects the informed trader's choice between market and limit orders.

Possessing private information is equivalent to having capacity to absorb and analyze publicly available information (including information from the past order flow) faster than other market participants (Foucault, Kozhan, and Tham (2015); Menkveld and Zoican (2015); Foucault, Hombert, and Roşu (2016)). Efficient information processing technology is a distinct feature of algorithmic traders, hence they are more likely to be informed than other market participants. However, ex ante it is not clear whether algorithmic traders would prefer to use market or limit orders to profit from their informational advantage.

On the one hand, limit orders are attractive for traders who can accurately predict execution probabilities, continuously monitor the market, and quickly adapt to market conditions. Algorithmic traders possess all of these characteristics. Thus, they may be inclined to use limit orders for informed trading.

On the other hand, competition among informed traders will lead to a faster price discovery and a shorter lifespan for the information obtained by the informed trader. Algorithmic traders compete for the same information by processing the same news releases or by analyzing past order flow patterns as fast as possible. In a competitive market, a trader must be the first in line to trade on information in order to profit from it. Given that only market orders can guarantee immediate execution, algorithmic traders may be inclined to use market orders for informed trading. Therefore, I formulate two competing hypotheses for the strategies employed by informed algorithmic traders:

H3 (the efficient technology hypothesis): The predictive power of informed market orders is lower for stocks subject to high algorithmic trading activity than for stocks subject to low algorithmic trading activity. On the other hand, the predictive power of informed limit orders is higher for stocks subject to high algorithmic trading activity than for stocks subject to low algorithmic trading activity.

H4 (the competition hypothesis): The predictive power of informed market orders is higher for stocks subject to high algorithmic trading activity than for stocks subject to low algorithmic trading activity. On the other hand, the predictive power of informed limit orders is lower for stocks subject to high algorithmic trading activity than for stocks subject to low algorithmic trading activity.

3.2.3 Effect of realized volatility

According to Goettler, Parlour, and Rajan (2009), informed traders may prefer market orders to limit orders at the inner levels of the limit order book for high volatility stocks and limit orders at the inner levels of the limit order book to market orders for low volatility stocks. The intuition is as follows. Posting a limit order is like writing an option (e.g., Copeland and Galai (1983); Jarnecic and McInish (1997); Harris and Panchapagesan (2005)). It is known that the sensitivity of the option price to the changes in the volatility of the underlying asset, i.e., vega (ν), is positive. In other words, the option price increases when the volatility of the underlying asset increases. In this way, the option writer gets compensated for the increased risk of option execution. Thus, the increased volatility of the stock will make limit orders riskier and hence, less profitable. In addition, market orders become more profitable due to picking off the stale limit orders posted by slow (and most likely uninformed) traders. And last but not least, in a highly volatile environment it is harder to distinguish between informed and uninformed market orders and hence, hiding informed trading is easier.

Given that on an intraday horizon, realized volatility based on the mid-quote returns is a good proxy for fundamental volatility, I formulate the *realized volatility hypothesis* as follows:

H5 (the realized volatility hypothesis): The predictive power of informed market orders is greater for high volatility stocks than for low volatility stocks. On the other hand,the predictive power of informed limit orders concentrated at the inner levels of the limit order book is greater for low volatility stocks than for high volatility stocks.

3.3 Data, Variables, and Summary Statistics

In this section, I describe the data, variables, and summary statistics. I obtain intraday consolidated data on trades and the 10 best levels of the limit order book for the U.S. market from the Thomson Reuters Tick History (TRTH) database. The TRTH database is provided by the Securities Industry Research Centre of Asia-Pacific (SIRCA). Data on trades and best bid-offer quotes are available since 1996. Data on the limit order book levels are available only from 2002 as the NYSE opened its limit order book to the public on January 24, 2002. The limit order book data provided by TRTH does not include order level information (e.g., no order submission, revision, or cancellation details), only the 10 best price levels and the depth on bid and ask sides of the book that is visible to the public. The data comes from the consolidated tape. In other words, the best bid-offer reported in the data is the best bid-offer for any exchange in the U.S. The same applies to the other levels of the limit order book.

TRTH data are organized by Reuters Instrumental Codes (RICs), which are identical to TICKERs provided by the Center for Research in Security Prices (CRSP). Merging data from CRSP and TRTH allows me to identify common shares that indicate the NYSE as their primary exchange and to use company specific-information (e.g., market capitalization, turnover, etc.). This study is limited to NYSE-listed stocks only as intraday return predictability from limit order book information as well as the behavior of the informed traders could be very sensitive to market design. Hence, it seems inappropriate to put, for example, the NASDAQ (hybrid dealer market) and NYSE (limit order book with designated market makers) data together.

The available data for the limit order book cover the period from 2002 to 2010. The joint size of the trade and limit order book data reaches 2.5 terabytes. In order to make the analysis feasible, I compute one-minute mid-quote returns and market order imbalances, and take snapshots of the limit order book at the end of each one-minute interval. I filter the data to discard faulty data entries and data entries outside continuous trading session (see the Appendix B for details).

3.3.1 Variable descriptions

In this section, I describe the variables used to study the choice of informed traders between market and limit orders by means of intraday return predictability from the limit order book. In particular, I look at the return predictability one-minute ahead. Therefore, I need intraday data on returns, market order imbalances ($\boldsymbol{MOIB}$), and limit order book data ($\boldsymbol{LOB}$) at one-minute frequency. For all the variables, I discard overnight observations.

I follow Chordia, Roll, and Subrahmanyam (2008) and compute one-minute log-returns ($\boldsymbol{Ret}$) based on the prevailing mid-quotes (average of the bid and ask prices) at the end of the one-minute interval, rather than the transaction prices or mid-quotes matched with the last transaction price. In this way I avoid the bid-ask bounce and ensure that the returns for every stock are indeed computed over a one-minute interval. I implicitly assume that there are no stale best bid-offer quotes in the sample, thus I consider a quote to be valid until a new quote arrives or until a new trading day starts.

To calculate a one-minute $\boldsymbol{MOIB}$, I match trades with quotes and sign trades using the Lee and Ready (1991) algorithm. TRTH data are stamped to the millisecond, therefore the Lee and Ready (1991) algorithm is quite accurate. In particular, a trade is considered to be buyer-initiated (seller-initiated) if it is closer to the ask price (bid price) of the prevailing quote. For each one-minute interval, I aggregate the trading volume in USD for buyer- and seller-initiated trades separately at the stock level. Thereafter, I subtract seller-initiated dollar volume from buyer-initiated dollar volume to obtain $\boldsymbol{MOIB}$.

There are multiple ways to describe the limit order book. Most of the papers that study intraday return predictability either focus on different levels of the limit order book or on the corresponding ratios of these levels between the ask and bid sides of the limit order book. For instance, Wuyts (2011), Cao, Hansch, and Wang (2009), and Cenesizoglu, Dionne, and Zhou (2014) use slopes and depth at different levels of the limit order book to summarize its shape. However, due to variation in the shape of the limit order book as well as in the number of available levels of the limit order book (in my sample the daily average number of levels can be as low as just six levels), I believe that definition of inner and outer levels by means of a relative threshold is more suitable than definition by means of the number of levels in the limit order book (e.g., levels from 2 to 5 are inner levels and levels from 6 to 10 are outer levels).

Examples of a relative approach to limit order book description are Cao, Hansch, and Wang (2009), who also use volume-weighted average price for different order sizes to describe the limit order book, and Kavajecz and Odders-White (2004), who use a so-called "near-depth" measure, which is a proportion of the depth close to the best bid-offer level relative to the cumulative depth within a certain price range.

For the purpose of testing the private information hypothesis, I focus on the ratios within the ask and bid sides separately, rather than across the ask and bid sides of the limit order book. I use a modification of the "near-depth" measure introduced by Kavajecz and Odders-White (2004). First, I compute a snapshot of the ask and bid sides of the limit order book at the end of each one-minute interval. Then, I define the inner depth concentration as cumulative depth lying between the mid-quote and one-third of the total distance between the 10th available limit price and the mid-quote relative to the total cumulative depth of the ask and bid side of the limit order book separately ($Ask\ Inner$ and $Bid\ Inner$). I define the outer depth concentration as cumulative depth lying between one-third and two-thirds of the total distance between the 10th available limit price and the mid-quote relative to the total cumulative depth of the ask and bid side of the limit order book separately ($Ask\ Outer$ and $Bid\ Outer$). Please refer to Table 3.2 for the summary of variables' descriptions.

My relative approach allows me to define inner and outer levels of the limit order book even if not all 10 levels are present for a particular stock at a particular time. Hence, I can define in unified fashion the levels that are close to the best bid-offer level, as well as the levels that are far away from the best bid-offer level across stocks

Table 3.2. Variables descriptions

This table shows the description of the variables used in the paper. Panel A shows variables for the first part of the analysis regarding intraday return predictability from market and limit order flows. Panel B shows variables for the second part of the analysis regarding the effect of algorithmic trading on the choice made by informed trader.

Panel A: Intraday return predictability

Variable	Description
Ret	One-minute log-returns based on the prevailing mid-quotes (average of the bid and ask prices) at the end of the one-minute interval at individual stock level.
MOIB	One-minute market order imbalance (buy volume minus sell volume) at individual stock level.
Bid Inner (*Ask Inner*)	Based on a snapshot of the bid (ask) side of the limit order book at the end of each one-minute interval, I define the inner depth concentration as cumulative depth lying between mid-quote and one-third of the total distance between 10th available limit price and mid-quote relative to the total cumulative depth of the bid (ask) side of the limit order book.
Bid Outer (*Ask Outer*)	Based on a snapshot of the bid (ask) side of the limit order book at the end of each one-minute interval, I define the outer depth concentration as cumulative depth lying between one-third and two-thirds of the total distance between 10th available limit price and mid-quote relative to the total cumulative depth of the bid (ask) side of the limit order book.

Panel B: Effect of algorithmic trading activity

Variable	Description
AT	The daily number of best bid-offer quote updates relative to the daily trading volume (in $10,000) on stock-day basis.
MCAP	Monthly average of the daily log of market capitalization in billions at individual stock level.
$1/PRC$	Inverse of monthly average of the daily closing price at individual stock level.
Turnover	Monthly average of the daily annualized turnover at individual stock level.
Volatility	Square root of monthly average of the daily high minus low range at individual stock level.

and through time.

3.3.2 Summary statistics

Table 3.3 presents summary statistics for the one-minute mid-quote returns (Ret), dollar market order imbalance ($MOIB$), and depth concentration at the inner levels ($Bid\ Inner$ and $Ask\ Inner$) and outer levels ($Bid\ Outer$ and $Ask\ Outer$) of the ask and bid sides of the limit order book (LOB), and cutoff points between the inner and outer levels of the limit order book measured relative to the mid-quote ($Bid\ Cutoff$ and $Ask\ Cutoff$) at the end of each one-minute interval for the whole period (from January 2002 to December 2010) and two sub-periods (from January 2002 to June 2006 and from July 2006 to December 2010). I start with winsorizing all variables at the 1% and 99% levels on a stock-day basis. Then, I compute averages of the one-minute observations for mid-quote returns (Ret), dollar market order imbalance ($MOIB$), and depth concentration at the inner and outer levels of the ask and bid sides of the limit order book per stock-day. Afterwards, I winsorize stock-day averages of the variables at the 1% and 99% levels based on the whole sample period or sub-periods and compute summary statistics.

The mean of the daily average one-minute mid-quote returns is -0.003 basis points for the whole sample period (see Panel A of Table 3.3). The average negative return is due to the inclusion of the recent financial crisis period in the sample. Indeed, in the first half of the sample period the average returns are 0.014 basis points, while in the second half of the period the average returns are -0.02 basis points. The mean of the daily average one-minute dollar market order imbalance is $4,133.34. This indicates that on average there is more buying than selling pressure in the market. However, this buying pressure is much more moderate at $840.15 – when I focus on the second half of the sample period due to the inclusion of the recent financial crisis.

Panel A of Table 3.3 also shows the depth concentration at the inner and outer levels separately of the ask and bid side of the limit order book for the whole sample period. The average proportion of the cumulative depth at the inner levels of the limit order book is 31.49% and 32.19% of the ask and bid side of the limit order book, respectively. The average proportion of the cumulative depth at the outer levels of

Table 3.3. Descriptive statistics

This table shows summary statistics of one-minute mid-quote returns, market order imbalance (***MOIB***), and depth concentration at the inner and outer levels of the ask and bid sides of the limit order book for the NYSE-listed common stocks during 2002-2010. Returns are reported in basis points, market order imbalance is reported in USD, depth concentration at the inner and outer levels is reported in percentage, cutoff points between inner and outer levels are reported in percentage relative to the mid-quote. For detailed description of the variables please refer to Table 3.2. To compute summary statistics, I follow the following procedure. First, I winsorize one-minute observations per stock-day at the 1% and 99% levels. Second, the average of the one-minute observations per stock-day is calculated for each variable. Third, I winsorize daily observations at the 1% and 99% levels for the whole sample period. Then, the summary statistics across all stock-days are computed for each variable. All the results are reported for the whole sample period (Panel A: Jan-2002 untill Dec-2010) and for the two sub-sample periods (Panel B: Jan-2002 untill Jun-2006 and Panel C: Jul-2006 untill Dec-2010). To be included in the sample, a stock should have NYSE as its primary exchange. Data on common stocks and primary exchange code are obtained from CRSP database (PRIMEXCH=N, and SHRCD=10 or 11, EXCHCD =1 or 31). Data on consolidated trades, quotes, and 10 best levels of the limit order book are provided by TRTH.

Panel A: Jan-2002 untill Dec-2010

	Ret	*MOIB*	*Bid Inner*	*Ask Inner*	*Bid Outer*	*Ask Outer*	*Bid Cutoff*	*Ask Cutoff*
Mean	-0.003	4,133.34	32.19%	31.49%	31.20%	31.36%	-1.43%	1.47%
St. Dev. Within	0.583	15,484.35	18.74%	17.47%	11.14%	10.92%	2.34%	2.63%
St. Dev. Between	0.160	5,930.18	11.43%	11.05%	6.40%	5.99%	2.27%	2.86%

Panel B: Jan-2002 untill Jun-2006

	Ret	*MOIB*	*Bid Inner*	*Ask Inner*	*Bid Outer*	*Ask Outer*	*Bid Cutoff*	*Ask Cutoff*
Mean	0.014	7,725.14	45.31%	42.66%	24.69%	25.49%	-2.45%	2.34%
St. Dev. Within	0.493	15,263.00	16.15%	15.35%	9.65%	9.45%	2.83%	3.20%
St. Dev. Between	0.181	9,590.36	11.70%	10.60%	6.09%	5.11%	3.01%	3.78%

Panel C: Jul-2006 untill Dec-2010

	Ret	*MOIB*	*Bid Inner*	*Ask Inner*	*Bid Outer*	*Ask Outer*	*Bid Cutoff*	*Ask Cutoff*
Mean	-0.020	840.15	20.47%	21.53%	37.03%	36.59%	-0.51%	0.68%
St. Dev. Within	0.664	14,398.41	10.64%	11.45%	8.69%	9.11%	0.72%	1.28%
St. Dev. Between	0.081	3,200.42	10.04%	10.71%	5.83%	6.28%	0.87%	1.34%

the limit order book is 31.36% and 31.20% of the ask and bid sides of the limit order book, respectively. Although the average depth concentration is very similar for the inner and outer levels for both ask and bid sides of the limit order book, depth concentration at the inner levels exhibits higher variation than depth concentration at the outer levels both in terms of within and between standard deviations. Notably, the ask and bid sides of the limit order book exhibit similar characteristics in terms of the depth concentration at the inner and outer levels.

Panel A of Table 3.3 also reports the cutoff points between inner and outer levels of the ask and bid sides of the limit order book measured as a percentage deviation from the mid-quote. For the whole sample period, the cutoff point (one-third of the total distance between the 10th available limit price and the mid-quote) is 1.47% and -1.43% of the ask and bid sides of the limit order book, respectively.

Sub-period analysis (see Panels B and C of Table 3.3) reveals that although on average through the whole sample period depth concentration at the inner and outer levels for both sides of the limit order book is similar, depth concentration at the inner levels tends to decrease over time, while depth concentration at the outer levels tends to increase over time.

In particular, in the first half of the sample period, depth concentration at the inner levels of the ask (bid) side of the limit order book is 42.66% (45.31%). In the second half of the sample period, depth concentration at the inner levels of the ask (bid) side of the limit order book is 21.53% (20.47%). In the first half of the sample period, depth concentration at the outer levels of the limit order book of the ask (bid) side of the limit order book is 25.49% (24.69%), while in the second half of the sample period it reaches 36.59% (37.03%).

This trend in the limit order book composition is also reflected in the cutoff points between the inner and outer levels of the limit order book. In particular, in the first half of the sample period, price levels of the limit order book are more dispersed than in the second half of the sample period. Hence, for the first half of the sample period I define inner depth as depth concentrated at price levels that do not differ from the mid-quote more than 2.34% (2.45%) of the ask (bid) side of the limit order book, respectively. The cutoff points for the second half of the period are 0.68% (0.51%) for the ask (bid) side of the limit order book, respectively.

This decreasing (increasing) trend in depth concentration at the inner (outer) levels of the limit order book can be also observed in Panel A of Figure 3.1. Panel B of Figure 3.1 shows the trend in cutoff points between the inner and outer levels of the limit order book.

The composition changes in the limit order book may be attributable to the different structural changes of the NYSE during the sample period such as autoquote introduction in 2003 (Hendershott, Jones, and Menkveld (2011)), NYSE Hybrid introduction in 2006-2007 (Hendershott and Moulton (2011)), Reg NMS implementation in 2007, and replacement of the specialist by designated market makers at the end of 2008.

3.4 Methodology

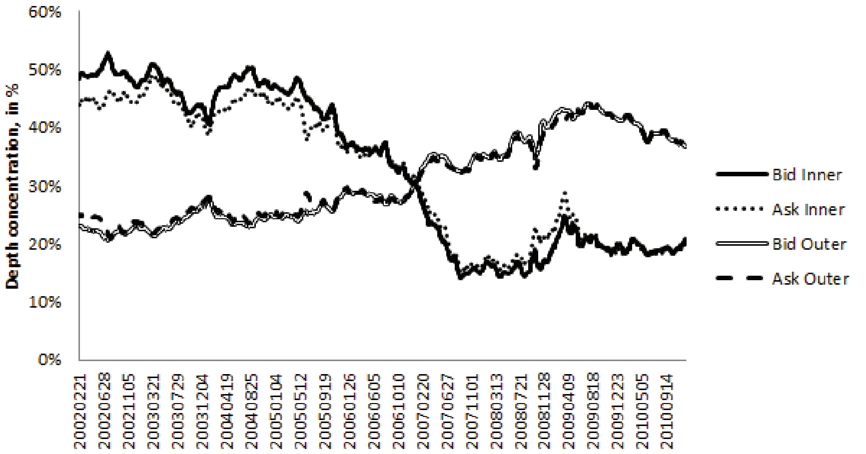
In this section, I describe the methodology used in the paper in order to investigate whether market and/or limit orders are used for informed trading. In particular, I empirically distinguish between two sources of intraday return predictability: inventory management (Hypothesis 1) and private information (Hypothesis 2). Given that the main goal of this paper is to investigate the informed trader's choice between market and limit orders, the latter source of the intraday return predictability is the one I focus on.

I run stock-day predictive regressions at one-minute frequency using one-minute mid-quote returns as the dependent variable. As explanatory variables I use lagged returns, lagged market order imbalance ($MOIB$), and lagged depth concentration at the inner and outer levels of the ask and bid sides of the limit order book. I include $MOIB$ in the model as I want to show that the LOB variables contain useful information for intraday return predictability beyond $MOIB$. Controlling for lagged returns allows me to differentiate between temporary effect (inventory management) and permanent effect (private information). The regression equation is given by:

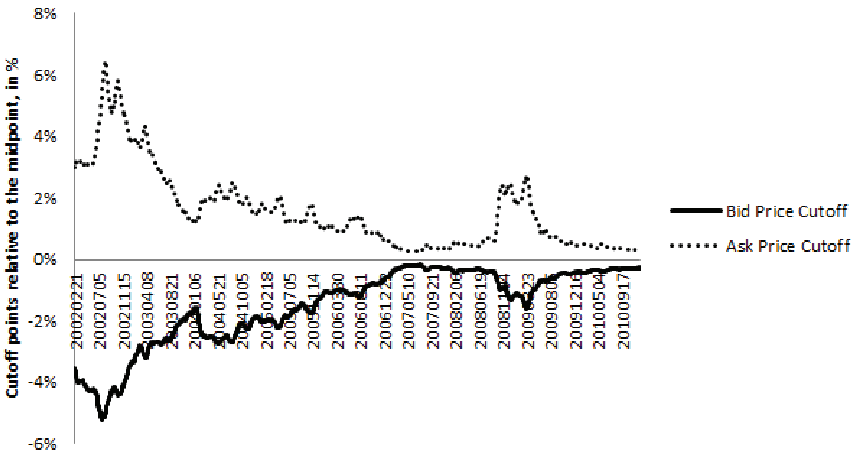
Figure 3.1. Limit order book composition

This figure shows the equally-weighted average of the depth concentration at the inner and outer levels for the ask and bid sides of the limit order book for the NYSE-listed common stocks during 2002-2010. Please refer to Table 3.2 for detailed variables description. I use the following procedure to construct time-series of the equally-weighted averages of the variables. First, I winsorize one-minute observations per stock-day at the 1% and 99% levels. Second, the average of the one-minute observations per stock-day is calculated for each variable. Third, I winsorize daily observations at the 1% and 99% levels for the whole sample period. Then, the summary statistics across all stock-days are computed for each variable. Then, I plot one-month moving average of each variable. To be included in the sample, a stock should have NYSE as its primary exchange. Data on common stocks and primary exchange code are obtained from CRSP database (PRIMEXCH=N, and SHRCD=10 or 11, EXCHCD =1 or 31). Data on consolidated trades, quotes, and 10 best levels of the limit order book are provided by TRTH.

Panel A: Depth concentration at the inner and outer levels of the limit order book



Panel B: Cutoff point between inner and outer levels of the limit order book



$$Ret_t = \alpha + \beta_1 Ret_{t-1} + \beta_2 MOIB_{t-1} + \beta_3 Bid\ Inner_{t-1} + \beta_4 Ask\ Inner_{t-1} + \tag{3.1}$$
$$+\beta_5 Bid\ Outer_{t-1} + \beta_6 Ask\ Outer_{t-1} + \epsilon_t$$

where Ret_t is the mid-quote return during the t-th one-minute interval, $MOIB_{t-1}$ is the dollar market order imbalance during the $(t-1)$-th one-minute interval, LOB_{t-1}: $Bid\ Inner_{t-1}$, $Ask\ Inner_{t-1}$, $Bid\ Outer_{t-1}$, $Ask\ Outer_{t-1}$ are the depth concentrations at the inner and outer levels of the ask and bid sides of the limit order book at the end of the $(t-1)$-th one-minute interval.

As a next step, I identify the private information component of the market and limit order flows and enhance the above mentioned methodology. Hasbrouck (1991) and Chordia, Roll, and Subrahmanyam (2005) show that $MOIB$ is positively autocorrelated. Moreover, Biais, Hillion, and Spatt (1995) and Ellul, Holden, Jain, and Jennings (2003) show that order flow is also persistent for limit orders. Biais, Hillion, and Spatt (1995) argue that there are three possible reasons for the order flow persistence: order splitting, imitation of other traders' behavior, and reaction to the public information in a sequential manner (e.g., due to the differences in trading speed). Degryse, de Jong, and van Kervel (2014) show that order flow persistence is caused by reasons other than private information. Previous empirical studies (e.g., Huang and Stoll (1997), Madhavan, Richardson, and Roomans (1997), Sadka (2006)) use unexpected changes in the market order flow in order to isolate information-related component. I extend this approach one step further and apply it to market and limit order flows. I argue that it is an appropriate extension as both market and limit order flows are persistent. Therefore, I use unexpected changes in the order flow for both market and limit orders as a proxy for the private information component of the order flow.

I obtain the surprises in returns, $MOIB$, and LOB variables by estimating stock-day $VAR(k)$ regression (number of lags, k, can take values from 1 to 5 and is selected by AIC criteria) and keeping the residual values:

$$X_t = \alpha + \sum_{l=1}^{l=k} \beta X_{t-l} + \epsilon_t \tag{3.2}$$

where X_t is a vector that includes Ret_t, $MOIB_t$, $Bid\ Inner_t$, $Ask\ Inner_t$, $Bid\ Outer_t$, and $Ask\ Outer_t$ measured at the t-th one-minute interval; ϵ_t is vector of residuals that includes the Ret_t^U, $MOIB_t^U$, $Bid\ Inner_t^U$, $Ask\ Inner_t^U$, $Bid\ Outer_t^U$, and $Ask\ Outer_t^U$.

In the remainder of the paper, the superscript U indicates a residual value from $VAR(k)$ rather than the variable itself. Misspecification of the $VAR(k)$ model may lead to some inventory effects ending up in the surprises. In order to address this issue, I include lagged surprises in returns as explanatory variable in the predictive regressions to capture return reversal, which is a distinct feature of the inventory management hypothesis. I run predictive regressions per stock-day with lagged surprises in returns, lagged surprises in $MOIB$, and lagged surprises in depth concentration at the inner and outer levels of the ask and bid sides of the limit order book as explanatory variables:

$$\begin{aligned} Ret_t = \alpha + \beta_1 Ret_{t-1}^U + \beta_2 MOIB_{t-1}^U + \beta_3 Bid\ Inner_{t-1}^U + \beta_4 Ask\ Inner_{t-1}^U + \\ + \beta_5 Bid\ Outer_{t-1}^U + \beta_6 Ask\ Outer_{t-1}^U + \epsilon_t \end{aligned} \tag{3.3}$$

3.5 Empirical Results

In this section, I provide empirical evidence for the informed trader's choice between market and limit orders by analyzing intraday return predictability from market and limit order flows (Section 3.5.1). Then, I discuss the role of algorithmic trading activity in the choice made by informed trader (Section 3.5.2). In Section 3.5.3, I provide supplementary analysis of the effects of realized volatility on the informed trader's choice.

3.5.1 Intraday return predictability

I start with examining whether limit order book variables are useful in predicting intraday returns without explicitly decomposing order flow into inventory- and information-related components. Table 3.4 presents estimation results of equation (3.1): predictive stock-day regressions of one-minute mid-quote returns on one-minute lagged mid-quote returns, one-minute lagged market order imbalance, and one-minute lagged depth concentration at the inner and outer levels of the ask and bid sides of the limit order book.

Panel A of Table 3.4 reports average coefficients with significance levels based on the average Newey-West t-statistics, as well as the proportion of the regressions that have significant individual t-statistics.[8] Ret_{t-1} is negatively related to the future returns. Such return reversals are in line with the inventory management hypothesis (Hypothesis 1). $MOIB_{t-1}$ is positively related to future stock returns (in line with, e.g., Chordia, Roll, and Subrahmanyam (2005, 2008)). In particular, the $MOIB_{t-1}$ coefficient is 4.65 and is positive and significant in 26.43% of the stock-day regressions. These results hold for the whole sample period as well as for the sub-periods.[9] The increase of one within standard deviation in $MOIB_{t-1}$ is associated with a 0.72 basis points increase in the future returns, which is equivalent to an increase of 1.24 within standard deviation for returns.

In line with the inventory management (Hypothesis 1) and informed limit orders (Hypothesis 2) hypotheses, depth concentration at the inner levels of the bid (ask) sides of the limit order book, $Bid\ Inner_{t-1}$ ($Ask\ Inner_{t-1}$) is positively (negatively) related to the future price movements. For the whole sample period, one within standard deviation increase in $Bid\ Inner_{t-1}$ ($Ask\ Inner_{t-1}$) corresponds to an increase of future returns by 0.35 basis points (decrease of future returns by -0.35 basis points), which is equivalent to an increase of 0.61 within standard deviation for

[8]To compute average Newey-West t-statistics, I do the following steps (following Rösch, Subrahmanyam, and van Dijk (2015)). First, I use a time series of the estimated coefficients for each stock to compute Newey-West t-statistics (Newey and West (1987)). Second, I average the cross-section of the Newey-West t-statistics to determine the average Newey-West t-statistics estimate.

[9]As a comparison, Rösch, Subrahmanyam, and van Dijk (2015) document that coefficient of $MOIB_{t-1}$ is 3.79 and is positive and significant in 30.07% of the predictive regressions using only lagged dollar market order imbalance over 1996-2010 for NYSE common stocks.

Table 3.4. Estimation results of the intraday return predictability from *MOIB* and *LOB*

This table shows the average estimation results of predictive regressions of one-minute mid-quote returns on lagged returns, lagged market order imbalance (***MOIB***), and lagged depth concentration at the inner and outer levels of the ask and bid sides of the limit order book for NYSE-listed common stocks during the sample period (2002-2010), see equation (3.1). I run this regression on the stock-day basis. The table reports average coefficients together with significance levels based on the average Newey-West t-statistics (Panel A), and adjusted R^2 decomposition (Panel B). Coefficient for order imbalance is scaled by 10^9. All other coefficients are scaled by 10^4. To compute average Newey-West t-statistic, I use a time-series of estimated coefficients for each stock to compute Newey-West t-statistics and average it across stocks. Individual regression t-statistics are used to determine the proportion of regressions that report significant coefficients (either positive or negative). The ordering of the variables used to decompose the adjusted R^2 is identical to the order in which they appear in the table. The last two rows show the total number of stock-day observations and the average number of stocks per day. To be included in the sample, a stock should have NYSE as its primary exchange. Data on common stocks and primary exchange code are obtained from CRSP database (PRIMEXCH=N, and SHRCD=10 or 11, EXCHCD =1 or 31). Data on consolidated trades, quotes, and 10 best levels of the limit order book are provided by TRTH. ***,**,* indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Coefficient estimates (dependent variable: Ret_t)

	Jan-2002 untill Dec-2010			Jan-2002 untill Jun-2006			Jul-2006 untill Dec-2010		
		% of significant and			% of significant and			% of significant and	
	Coef	Positive	Negative	Coef	Positive	Negative	Coef	Positive	Negative
Constant	0.048	8.29%	7.74%	0.111	8.50%	7.50%	-0.006	8.10%	7.96%
Ret_{t-1}	-0.011***	14.57%	20.23%	-0.013***	13.31%	20.24%	-0.009***	15.69%	20.21%
$MOIB_{t-1}$	4.650***	26.43%	2.48%	3.145***	24.87%	2.46%	6.278***	27.81%	2.50%
$Bid\ Inner_{t-1}$	1.894***	16.02%	4.80%	1.714***	16.31%	4.41%	2.055***	15.76%	5.14%
$Ask\ Inner_{t-1}$	-2.021***	4.41%	17.65%	-2.113***	3.41%	19.56%	-1.915***	5.30%	15.95%
$Bid\ Outer_{t-1}$	-0.155	7.21%	8.09%	0.154	7.92%	7.14%	-0.434**	6.57%	8.93%
$Ask\ Outer_{t-1}$	0.113	8.08%	7.27%	-0.168	7.20%	7.88%	0.356*	8.87%	6.74%
Adjusted R^2	1.64%			1.71%			1.58%		
# of stock-days	2,740,593			1,291,413			1,448,989		
Average # of stocks	1,228			1,167			1,289		

Table 3.4. Estimation results of the intraday return predictability from *MOIB* and *LOB* (continued)

Panel B: Adjusted R^2 decomposition (dependent variable: Ret_t)

	Jan-2002 untill Dec-2010		Jan-2002 untill Jun-2006		Jul-2006 untill Dec-2010	
	Adjusted R^2		Adjusted R^2		Adjusted R^2	
	Absolute	Relative	Absolute	Relative	Absolute	Relative
Constant						
Ret_{t-1}	0.53%	32.38%	0.54%	31.32%	0.53%	33.38%
$MOIB_{t-1}$	0.34%	20.66%	0.33%	19.28%	0.35%	21.98%
$Bid\ Inner_{t-1}$	0.21%	13.05%	0.23%	13.66%	0.20%	12.45%
$Ask\ Inner_{t-1}$	0.24%	14.74%	0.28%	16.08%	0.21%	13.42%
$Bid\ Outer_{t-1}$	0.16%	9.57%	0.17%	9.86%	0.15%	9.34%
$Ask\ Outer_{t-1}$	0.16%	9.60%	0.17%	9.79%	0.15%	9.44%
Total Inner	0.45%	27.79%	0.51%	29.74%	0.41%	25.87%
Total Outer	0.32%	19.17%	0.34%	19.65%	0.30%	18.78%
Total LOB	0.77%	46.96%	0.85%	49.39%	0.71%	44.65%
Total	1.64%	100.00%	1.71%	100.00%	1.58%	100.00%
# of stock-days	2,740,593		1,291,413		1,448,989	
Average # of stocks	1,228		1,167		1,289	

returns (decrease of 0.61 within standard deviation for returns).

However, the fact that $\textit{Bid Outer}_{t-1}$ ($\textit{Ask Outer}_{t-1}$) is negatively (positively) related to future price movements in the second half of the period cannot be explained under the inventory management hypothesis (Hypothesis 1), while it is true under the private information hypothesis (Hypothesis 2). Notably, the sign of $\textit{Bid Outer}_{t-1}$ ($\textit{Ask Outer}_{t-1}$) changes from insignificantly positive (negative) in the first half of the sample period to significantly negative (positive) in the second half of the sample period. In other words, informational content at the outer levels of the limit order book is lower in the first half of the sample period compared to the second half of the sample period. These results are also in line with increasing depth concentration at the outer levels of the limit order book and decreasing depth concentration at the inner levels of the limit order book over the sample period. For the whole sample period, one within standard deviation increase in $\textit{Bid Outer}_{t-1}$ ($\textit{Ask Outer}_{t-1}$) corresponds to decrease of future returns by -0.017 basis points (increase of future returns by 0.012 basis points), which is equivalent to decrease of 0.03 within standard deviation for returns (increase of 0.02 within standard deviation for returns).

Remarkably, the effects of the ask and bid sides of the limit order book are similar in terms of the absolute size of the coefficients. However, the median of daily correlation coefficients between $\textit{Bid Inner}_{t-1}$ and $\textit{Ask Inner}_{t-1}$ ($\textit{Bid Outer}_{t-1}$ and $\textit{Ask Outer}_{t-1}$) is quite low – at only 6.24% (2.21%). Put differently, the depth concentration of the ask and bid sides of the limit order book tend to vary largely independently from each other, thus their effects on future returns should not offset each other.

At the same time, Panel A of Table 3.4 shows a clear discrepancy in the absolute size of the coefficients between depth concentration at the inner and outer levels: 1.89 (-2.02) to -0.16 (0.11) of the bid (ask) side during the whole sample period, respectively.[10] This discrepancy could be due to the fact that outer levels are not

[10] A natural concern is that the inner and outer levels of the limit order book are negatively correlated by construction. If there is an extremely high correlation between depth concentration at the inner and outer levels of the limit order book, I can run into a multicollinearity problem. However, across all stock-days, these correlation coefficients never fall below -70%, and the median value is around -46% for both ask and bid sides of the limit order book.

likely to be used for inventory management. In addition, outer levels are used for informed trading if and only if an informed trader receives a relatively strong signal, which is unlikely to happen regularly on the market.

In order to measure the relative importance of market and limit order variables, I look at the R^2 decomposition of the predictive regressions. Panel B of Table 3.4 shows that the average adjusted R^2 of the predictive regressions is equal to 1.64% for the whole sample period. Adjusted R^2 attributable to $MOIB_{t-1}$ is 0.34% in absolute terms, which accounts for 20.66% of the total explanatory power. As a comparison, Chordia, Roll, and Subrahmanyam (2008) document an adjusted R^2 of 0.51% for predictive regressions using only lagged dollar market order imbalance for the 1993-2002 period, which is of the same order of magnitude as my estimate. Lagged return accounts for 32.38% of the total predictive power, while 46.96% of the total predictive power comes from the limit order book variables (with 27.79% attributable to the depth concentration at the inner levels of the limit order book and 19.17% attributable to the depth concentration at the outer levels of the limit order book).

My results are also consistent with Cao, Hansch, and Wang (2009), who document an increase in adjusted R^2 after inclusion of additional levels of the limit order book with a monotonic decrease of the added value for each additional level. My results are however at odds with Cont, Kukanov, and Stoikov (2013), who argue that only imbalances at the BBO level drive intraday return predictability. Despite the fact that Cao, Hansch, and Wang (2009) and Cont, Kukanov, and Stoikov (2013) also investigate intraday return predictability from the limit order book, the data used in their studies is quite limited. Specifically, Cao, Hansch, and Wang (2009) use one month of data on 100 stocks traded on the Australian Stock Exchange, while Cont, Kukanov, and Stoikov (2013) use one month of data on 50 stocks from S&P 500 constituents. Overall, my results allow me to draw more generalizable conclusions regarding intraday return predictability and observed time series and cross-sectional patterns.

The sub-period analysis yields the following results. Total predictive power of the regressions decreases slightly from 1.71% in the first half of the sample period to 1.58% in the second half. This decrease is attributable to the limit order book (adjusted R^2 decreases from 0.85% to 0.71%). The predictive power of the $MOIB$ increases slightly from 0.33% to 0.35%. This evidence is consistent with the fact that intraday return predictability from the limit order book is a persistent phenomenon during 2002-2010 for all NYSE-listed common stocks.

Next, I enrich the analysis discussed above in order to emphasize the importance of private information source of intraday return predictability. To determine the pure effect of private information on intraday return predictability from market and limit order flows, I follow the previous literature (e.g., Huang and Stoll (1997), Madhavan, Richardson, and Roomans (1997), Sadka (2006)) and use surprises in market and limit order flows to define the informational component of the order flows. I calculate surprises as residual values of the $VAR(k)$ regression on a stock-day basis with the number of lags determined by AIC criteria (see equation (3.2)). I then repeat the above-mentioned analysis with these surprises used as explanatory variables (see equation (3.3)). I use superscript U to refer to surprises in the variables.

Table 3.5 presents the average estimation results of this analysis. The results in Table 3.5 are similar to the results in Table 3.4, with the only exception of the depth concentration at the outer levels of the ask side of the limit order book, which is no longer significant during the second half of the period. Nevertheless, all the signs during the whole sample period and the second half of the sample period are consistent with the private information hypothesis (Hypothesis 2).

Based on the whole sample period, adjusted R^2 attributable to the $MOIB^U$ is 0.31% in absolute terms (20.92% in relative terms), while the adjusted R^2 attributable to surprises in LOB variables is 0.71% in absolute terms (47.21% in relative terms). The inner levels of the limit order book contribute 27.65% and outer levels contribute 19.56% of this predictive power.

All in all, this suggests that private information is the main source of the intraday return predictability: roughly 20% of this predictability is attributable to the informed market orders, roughly 50% is attributable to the informed limit or-

Table 3.5. Estimation results of the intraday return predictability from surprises in *MOIB* and *LOB*

This table shows the average estimation results of predictive regressions of one-minute mid-quote returns on lagged surprises in returns, lagged surprises in market order imbalance (***MOIB***), and lagged surprises in depth concentration at the inner and outer levels of the ask and bid sides of the limit order book for NYSE-listed common stocks during the sample period (2002-2010), see equation (3.3). Surprises are computed as residual values from $VAR(k)$ regression per stock-day, number of lags, k, can take values from 1 to 5 and is selected by AIC criteria (see equation (3.2)). Superscript U indicates that this a residual value from $VAR(k)$. I run this regression on the stock-day basis. The table reports average coefficients together with significance levels based on the average Newey-West t-statistics (Panel A), and adjusted R^2 decomposition (Panel B). Coefficient for order imbalance is scaled by 10^9. All other coefficients are scaled by 10^4. To compute average Newey-West t-statistic, I use a time-series of estimated coefficients for each stock to compute Newey-West t-statistics and average it across stocks. Individual regression t-statistics are used to determine the proportion of regressions that report significant coefficients (either positive or negative). The ordering of the variables used to decompose the adjusted R^2 is identical to the order in which they appear in the table. The last two rows show the total number of stock-day observations and the average number of stocks per day. To be included in the sample, a stock should have NYSE as its primary exchange. Data on common stocks and primary exchange code are obtained from CRSP database (PRIMEXCH=N, and SHRCD=10 or 11, EXCHCD =1 or 31). Data on consolidated trades, quotes, and 10 best levels of the limit order book are provided by TRTH. ***,**,* indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Coefficient estimates (dependent variable: Ret_t)

	Jan-2002 untill Dec-2010			Jan-2002 untill Jun-2006			Jul-2006 untill Dec-2010		
		% of significant and			% of significant and			% of significant and	
	Coef	Positive	Negative	Coef	Positive	Negative	Coef	Positive	Negative
$Constant$	-0.002	6.72%	6.19%	0.013	7.14%	6.26%	-0.017	6.35%	6.13%
Ret^U_{t-1}	-0.015***	13.38%	20.56%	-0.020***	11.86%	21.18%	-0.011***	14.74%	20.01%
$MOIB^U_{t-1}$	4.774***	25.25%	2.27%	3.213***	23.75%	2.34%	6.459***	26.58%	2.20%
$Bid\ Inner^U_{t-1}$	2.520***	15.00%	4.62%	2.617***	15.31%	4.19%	2.497***	14.73%	5.01%
$Ask\ Inner^U_{t-1}$	-2.571***	4.36%	16.24%	-2.775***	3.51%	17.74%	-2.394***	5.12%	14.90%
$Bid\ Outer^U_{t-1}$	-0.004	7.04%	7.76%	0.601	7.58%	6.76%	-0.434*	6.56%	8.66%
$Ask\ Outer^U_{t-1}$	0.049	7.76%	7.05%	-0.273	6.89%	7.44%	0.315	8.54%	6.70%
Adjusted R^2	1.50%			1.57%			1.44%		
# of stock-days	2,739,445			1,290,389			1,448,865		
Average # of stocks	1,228			1,166			1,289		

Table 3.5. Estimation results of the intraday return predictability from surprises in *MOIB* and *LOB* (continued)

Panel B: Adjusted R^2 decomposition (dependent variable: Ret_t)

	Jan-2002 untill Dec-2010		Jan-2002 untill Jun-2006		Jul-2006 untill Dec-2010	
	Adjusted R^2		Adjusted R^2		Adjusted R^2	
	Absolute	Relative	Absolute	Relative	Absolute	Relative
Constant						
Ret^U_{t-1}	0.48%	31.87%	0.49%	31.42%	0.47%	32.29%
$MOIB^U_{t-1}$	0.31%	20.92%	0.31%	19.83%	0.32%	21.97%
$Bid\ Inner^U{}_{t-1}$	0.20%	13.40%	0.22%	13.81%	0.19%	13.00%
$Ask\ Inner^U{}_{t-1}$	0.21%	14.25%	0.24%	15.34%	0.19%	13.20%
$Bid\ Outer^U{}_{t-1}$	0.15%	9.79%	0.15%	9.83%	0.14%	9.77%
$Ask\ Outer^U{}_{t-1}$	0.15%	9.77%	0.15%	9.78%	0.14%	9.78%
$Total\ Inner^U$	0.41%	27.65%	0.46%	29.15%	0.38%	26.20%
$Total\ Outer^U$	0.30%	19.56%	0.30%	19.61%	0.28%	19.55%
$Total\ LOB^U$	0.71%	47.21%	0.76%	48.76%	0.66%	45.75%
$Total^U$	1.50%	100.00%	1.57%	100.00%	1.44%	100.00%
# of stock-days	2,739,445		1,290,389		1,448,865	
Average # of stocks	1,228		1,166		1,289	

ders. Remaining 30% are stemming from inventory management concerns (lagged returns).

Furthermore, the evidence is consistent with the majority of informed trading taking place via limit orders contrary to the traditional view of informed trading taking place via market orders only.

3.5.2 Algorithmic trading and informed trader's choice

To this end, I provide evidence consistent with limit orders being actively used for informed trading. Furthermore, my findings suggest that informed limit orders are a prevalent source of intraday return predictability. I now examine the role of algorithmic trading activity in the choice made by the informed trader.

In particular, I identify the effects of algorithmic trading activity on intraday return predictability from the limit order book. The results of this section add to the ongoing debate on whether algorithmic traders improve or decrease market quality. Identifying the causal effects of the algorithmic trading activity is not a trivial task as the degree of algorithmic trading activity in each stock on each day is an endogenous choice made by the algorithmic trader. Therefore, I adopt an instrumental variable approach following Hendershott, Jones, and Menkveld (2011) to identify the causal effects of the algorithmic trading on limit order book informational content.

Since January 2002 when the NYSE opened its limit order book to public, there were two major technological advances in NYSE equity market design that impacted algorithmic trading activity: Autoquote in 2003 (Hendershott, Jones, and Menkveld (2011)) and NYSE Hybrid Market in 2006-2007 (Hendershott and Moulton (2011)). After the NYSE Hybrid Market introduction, orders were allowed to "walk" through the limit order book automatically, before this technological change market orders were executed automatically at the best bid-offer level only. I use the NYSE Hybrid Market introduction as an instrument for algorithmic trading activity that allows me to investigate the role of algorithmic traders in informed trading activity.

I obtain data on the NYSE Hybrid Phase 3 rollout, which was when the actual increase in the degree of automated execution and speed took place (Hendershott and Moulton (2011)) from Terrence Hendershott's website. This rollout was implemented in a staggered way from October 2006 until January 2007 (see Figure 3.2), which allows for a clean identification. My analysis is focused on the period around Hybrid introduction from June 2006 to May 2007. All stocks in the sample have CRSP data available during the whole period under consideration. I discard stocks with average monthly price bigger than \$1,000 and smaller than \$5. I winsorize all the variables at the 1% and 99% levels.

I consider the following proxy for algorithmic trading activity in the spirit of Hendershott, Jones, and Menkveld (2011) and Boehmer, Fong, and Wu (2015): AT, a daily number of best bid-offer quote updates relative to daily trading volume (in \$10,000).[11]

I follow Hendershott, Jones, and Menkveld (2011) and estimate the following IV panel regression with stock and day fixed effects (implicit difference-in-difference approach) and double-clustering of the standard errors (Petersen (2009)):

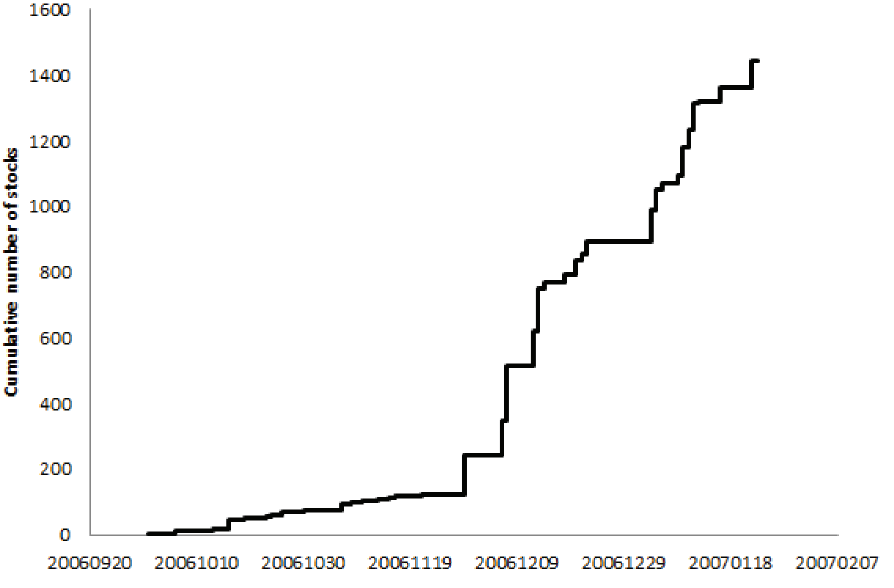
$$Y_{i,t} = \alpha_i + \gamma_t + AT_{i,t} + MCAP_{i,m-1} + 1/PRC_{i,m-1} + Turnover_{i,m-1} + Volatility_{i,m-1} + \epsilon_{i,t} \quad (3.4)$$

where $Y_{i,t}$ is either coefficients estimates from equation (3.3), or incremental adjusted R^2 from equation (3.3) for stock i on day t, and α_i and γ_t are stock and day fixed effects. $AT_{i,t}$ is a proxy for algorithmic trading activity for stock i on day t . In addition, I control for daily log of market capitalization in billions ($MCAP_{i,m-1}$), inverse of price ($1/P_{i,m-1}$), annualized turnover ($Turnover_{i,m-1}$), and square root of high minus low range ($Volatility_{i,m-1}$) averaged over the previous month, $m-1$. As a set of instruments, I use all explanatory variables with $AT_{i,t}$ replaced by $Hybrid_{i,t}$, a dummy variable that equals one if the stock i on day t is rolled-out to the NYSE Hybrid Market and 0 otherwise. In other words, I estimate equation (3.4) by means of 2SLS with an exclusion restriction on the Hybrid Market introduction dummy.

[11] The results are robust for using a different proxy for algorithmic trading activity: a daily number of limit order book updates relative to daily trading volume (in \$10,000). On the one hand, by construction this is a better proxy for algorithmic trading activity in the limit order book. On the other hand, my limit order book data is limited as it takes into account only first 10 levels of the limit order book (aggregated depth at first 10 price levels). In addition, I do not have order level data (submission, revision, cancellation). Therefore, the change in this measure due to NYSE Hybrid Market introduction is bounded from above due to data limitations. Results with this proxy are available from the author upon a request.

Figure 3.2. NYSE Hybrid Market introduction

This figure shows the staggered way of NYSE Hybrid Market introduction for the stocks included in the analysis from October 1, 2006 to January 31, 2007. To be included in the sample, a stock should have NYSE as its primary exchange and have CRSP daily data available for the period from June 2006 to May 2007. Data on common stocks and primary exchange code are obtained from CRSP database (PRIMEXCH=N, and SHRCD=10 or 11, EXCHCD =1 or 31). Data on Hybrid introduction are from Terrence Hendershott's website.



Unreported results of the first stage regression show that AT increases significantly with NYSE Hybrid Market introduction (an increase of 1.12 best bid-offer updates per \$10,000 of daily trading volume). The null hypothesis that instrument does not enter first-stage regression is strongly rejected.

The results for the second stage regression for AT are presented in Table 3.6. In particular, I estimate the effect of algorithmic trading on the coefficients (Panel A) and incremental adjusted R^2 (Panel B) from predictive regressions of one-minute mid-quote returns on lagged surprises in returns, $MOIB$, and LOB variables (see equation (3.3)). I test the efficient technology hypothesis (Hypothesis 3) against the competition hypothesis (Hypothesis 4).

Panel A of Table 3.6 shows that in line with the competition hypothesis, the coefficients of lagged $MOIB^U$ significantly increase in an absolute sense with an increase in algorithmic trading activity. However, there is also an increase in the $Bid\ Inner^U$ and $Ask\ Inner^U$ coefficients in line with the efficient technology hypothesis. This is consistent with slow traders, who are likely to be uninformed, moving away from the inner to outer levels, while fast and potentially informed traders continue operating at the inner levels of the bid and ask sides of the limit order book. The coefficients of the lagged returns also increase in an absolute sense, consistent with the fact that high-frequency traders (subset of algorithmic traders) are known to end their day with a flat inventory position. Therefore, inventory management concerns should generate a stronger return reversal in the presence of algorithmic traders.

Panel B of Table 3.6 reports the effect of algorithmic trading on the incremental adjusted R^2 from equation (3.3). Algorithmic trading participation increases the predictive power of all variables, although the increase in predictive power of the depth concentration at the outer levels of the bid and ask sides of the limit order book is marginal. In particular, a one standard deviation increase in AT leads to an increase of 8.2 basis points in the adjusted R^2 attributable to lagged surprises returns, a 5.3 basis points increase in the adjusted R^2 attributable to $MOIB^U$, a 2.6 (2.7) basis points increase in the adjusted R^2 attributable to $Bid\ Inner^U$ ($Ask\ Inner^U$), and a 0.7 (0.6) basis points increase in the adjusted R^2 attributable to $Bid\ Outer^U$

Table 3.6. Second stage regression: Impact of algorithmic trading activity on intraday return predictability

This table shows the impact of algorithmic trading activity on intraday return predictability from the limit order book, see equation (3.4). As proxy for algorithmic trading activity I use daily number of best bid-offer quote updates per \$10,000 of daily trading volume ($AT_{i,t}$). In order to identify causal effect of algorithmic trading activity, I instrument it with staggered introduction of NYSE Hybrid. The set of instruments include all explanatory variables with $AT_{i,t}$ substituted by $Hybrid_{i,t}$. I control for daily market capitalization, price, turnover, and volatility averaged over the previous calendar month. The specification includes stock (α_i) and day (γ_t) fixed effects. Standard errors are adjusted for double clustering. Panel A reports the effect of algorithmic trading on the coefficients from predictive regressions. Panel B reports the effect of algorithmic trading on the incremental adjustedR^2 from predictive regressions. The last two rows in each panel show the total number of stock-day observations and the average number of stocks per day. The estimation period is Jun-2006 untill May-2007. To be included in the sample, a stock should have NYSE as its primary exchange and have CRSP data available for all days under consideration. Data on common stocks and primary exchange code are obtained from CRSP database (PRIMEXCH=N, and SHRCD=10 or 11, EXCHCD =1 or 31). Data on consolidated trades, quotes, and 10 best levels of the limit order book are provided by TRTH. Data on NYSE Hybrid introduction comes from Terrence Hendershott's website. ***,**,* indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Effect of algorithmic trading on coefficients

	$Ret^U_{i,t-1}$	$MOIB^U_{i,t-1}$	$Bid\ Inner^U_{i,t-1}$	$Ask\ Inner^U_{i,t-1}$	$Bid\ Outer^U_{i,t-1}$	$Ask\ Outer^U_{i,t-1}$
$AT_{i,t}$	-0.004***	0.095**	0.337***	-0.275***	-0.006	0.001
	(-3.67)	(2.40)	(9.21)	(-6.66)	(-0.17)	(0.05)
$\ln(MCAP)_{i,m-1}$	-0.010***	0.162	1.158***	-1.026***	0.182*	-0.153
	(-2.78)	(1.13)	(8.28)	(-7.19)	(1.71)	(-1.30)
$1/PRC_{i,m-1}$	-0.321***	54.778***	8.288**	-7.782*	8.490**	-4.437
	(-3.28)	(11.14)	(1.97)	(-1.82)	(2.47)	(-1.32)
$Turnover_{i,m-1}$	-0.045***	-1.594***	1.373***	-0.952***	0.022	-0.029
	(-8.02)	(-7.68)	(6.70)	(-4.45)	(0.12)	(-0.18)
$Volatility_{i,m-1}$	0.748***	56.901***	-4.218	-0.311	-9.770**	9.048**
	(6.13)	(10.24)	(-0.80)	(-0.06)	(-2.45)	(2.31)
Stock FE	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES
Double clustering	YES	YES	YES	YES	YES	YES
R^2	0.14	0.35	0.06	0.06	0.01	0.01
# of stock-days	296,130	296,130	296,130	296,130	296,130	296,130
Average # of stocks	1,180	1,180	1,180	1,180	1,180	1,180

Table 3.6. Second stage regression (continued)

Panel B: Effect of algorithmic trading on incremental Adjusted R^2							
	$Adj\ R^2_{i,t}$	$Ret^U_{i,t-1}$	$MOIB^U_{i,t-1}$	$Bid\ Inner^U_{i,t-1}$	$Ask\ Inner^U_{i,t-1}$	$Bid\ Outer^U_{i,t-1}$	$Ask\ Outer^U_{i,t-1}$
$AT^{BBO}_{i,t}$	0.211***	0.080***	0.052***	0.025***	0.026***	0.007**	0.006*
	(9.11)	(7.93)	(8.20)	(5.64)	(5.26)	(2.17)	(1.76)
$\ln(MCAP)_{i,m-1}$	0.380***	0.234***	0.004	0.049***	0.079***	0.001	-0.005
	(6.31)	(7.88)	(0.24)	(4.05)	(5.72)	(0.09)	(-0.60)
$1/PRC_{i,m-1}$	-4.410**	-1.681**	-1.461***	-0.692**	-0.550	0.163	0.129
	(-2.53)	(-2.19)	(-2.83)	(-2.08)	(-1.41)	(0.68)	(0.50)
$Turnover_{i,m-1}$	1.209***	0.508***	0.260***	0.132***	0.140***	0.047***	0.037**
	(10.72)	(9.88)	(8.52)	(5.89)	(5.59)	(3.07)	(2.12)
$Volatility_{i,m-1}$	-27.372***	-11.065***	-4.728***	-3.027***	-3.631***	-1.430***	-1.355***
	(-12.40)	(-10.42)	(-7.97)	(-7.12)	(-7.55)	(-4.73)	(-4.34)
Stock FE	YES	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES	YES
Double clustering	YES	YES	YES	YES	YES	YES	YES
R^2	0.06	0.04	0.06	0.03	0.04	0.02	0.02
# of stock-days	296,130	296,130	296,130	296,130	296,130	296,130	296,130
Average # of stocks	1,180	1,180	1,180	1,180	1,180	1,180	1,180

(*Ask Outer*^{U}).[12] Put differently, I find evidence consistent with both the efficient technology (predictive power of limit orders increases) and the competition (predictive power of market orders increases) hypotheses. However, the effects of the competition hypothesis may dominate those of the efficient technology hypothesis.

Note that intraday return predictability (total as well as incremental) increases with the size and turnover, and decreases with the inverse of price and volatility. Size and turnover could be viewed as a proxies for stocks' liquidity. Lower transaction costs allow traders to benefit even from small pieces of information, on which they would not trade otherwise, which in turn increases the predictive power of limit and primarily market orders.

Overall, I contribute to the debate on whether algorithmic traders adversely select other market participants. I provide evidence that the increased degree of algorithmic trading participation is associated with an increase in the informational content of not only market orders, but also limit orders at the inner levels of the limit order book (with outer levels being only marginally affected). In other words, an increase in algorithmic trading activity leads to an increase in informed trading via both market (demanding liquidity) and limit orders (providing liquidity), with a relative shift from informed liquidity provision to informed liquidity consumption.

3.5.3 Realized volatility and informed trader's choice

I test the realized volatility hypothesis based on the theoretical predictions from Goettler, Parlour, and Rajan (2009), who argue that informed traders tend to use market orders for high volatility stocks and limit orders for low volatility stocks (Hypothesis 5). These effects should be mainly observed for the orders posted at the inner levels of the limit order book as these orders are more likely to be hit.

I estimate predictive regressions of one-minute mid-quote returns (see equation (3.3)) with one-minute lagged surprises in returns, one-minute lagged surprises in market order imbalance, and lagged surprises in depth concentration at the inner and outer levels of the ask and bid sides of the limit order book as explanatory variables on a stock-day basis. Then, I sort the stocks into four portfolios based on

[12] Recall from Section 3.5.1 that an average adjusted R^2 for the whole sample period is 1.50%.

one-day lagged realized volatility (realized volatility is computed from one-minute mid-quote returns during the day).

Ex ante, I expect a monotonic increase in the absolute coefficient of the surprises in market order imbalance and adjusted R^2 from the low volatility portfolio to the high volatility portfolio, while I expect the opposite for the surprises in depth concentration at the inner levels of the ask and bid sides of the limit order book. Table 3.7 reports the estimation results for the average coefficients and the average Newey-West t-statistics (Panel A) and adjusted R^2 decomposition (Panel B) for the whole sample period only.

Table 3.7 Panel A shows a monotonic increase for the coefficient of $MOIB^U$ from 1.18 to 11.63 while moving from the low realized volatility portfolio to the high realized volatility portfolio. In other words, coefficient of $MOIB^U$ is 9.86 times greater for high volatility stocks than for low volatility stocks. The coefficients of $Bid\ Inner^U (Ask\ Inner^U)$ also increase monotonically in absolute sense from the low volatility portfolio to the high volatility portfolio from 1.64 (-1.71) to 3.78 (-3.81), but this increase is very moderate compared to $MOIB^U$. The coefficients of $Bid\ Outer^U (Ask\ Outer\ ^U)$ are not significant.[13]

Table 3.7 Panel B shows adjusted R^2 decomposition for each of the explanatory variables for the four realized volatility portfolios. There is a monotonic increase in the adjusted R^2 attributable to $MOIB^U$ while moving from the low realized volatility portfolio to the high realized volatility portfolio from 0.29% to 0.38% in absolute terms (19.42% to 22.88% in relative terms). However, there is a slightly U-shaped pattern for the adjusted R^2 attributable to LOB variables in absolute terms and a monotonically decreasing pattern in relative terms (48.51% to 44.97%). The rest of predictive power comes from surprises in lagged returns.

All in all, I provide evidence that informed traders may prefer market orders to limit orders at the inner levels of the limit order book for high volatility stocks and limit orders at the inner levels of the limit order book to market orders for low volatility stocks. In other words, informed traders are more likely to consume

[13]The results are robust for the sub-period analysis.

Table 3.7. Realized volatility portfolios and intraday return predictability from surprises in *MOIB* and *LOB*

This table shows the average estimation results of predictive regressions of one-minute mid-quote returns on lagged surprises in returns, lagged surprises in market order imbalance ($MOIB$), and lagged surprises in depth concentration at the inner and outer levels of the ask and bid sides of the limit order book for NYSE-listed common stocks during the sample period (2002-2010), see equation (3.3). Surprises are computed as residual values from $VAR(k)$ regression per stock-day, number of lags, k, can take values from 1 to 5 and is selected by AIC criteria. Superscript U indicates that this a residual value from $VAR(k)$. I run this regression on the stock-day basis. For each day I sort all the stocks into four portfolios based on one-day lagged realized volatility (realized volatility is computed from one-minute mid-quote returns). The table reports average coefficients together with significance levels based on the average Newey-West t-statistics (Panel A), and adjusted R^2 decomposition (Panel B). Coefficient for order imbalance is scaled by 10^9. All other coefficients are scaled by 10^4. To compute average Newey-West t-statistic, I use a time-series of estimated coefficients for each stock to compute Newey-West t-statistics and average it across stocks. The ordering of the variables used to decompose the adjusted R^2 is identical to the order in which they appear in the table. The last two rows show the total number of stock-day observations and the average number of stocks per day. To be included in the sample, a stock should have NYSE as its primary exchange. Data on common stocks and primary exchange code are obtained from CRSP database (PRIMEXCH=N, and SHRCD=10 or 11, EXCHCD =1 or 31). Data on consolidated trades, quotes, and 10 best levels of the limit order book are provided by TRTH. ***,**,* indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Coefficient estimates (dependent variable: Ret_t)

	RV1 (low)	RV2	RV3	RV4 (high)
$Constant$	0.005	0.003	-0.001	-0.016
Ret^U_{t-1}	-0.011*	-0.009	-0.013**	-0.027***
$MOIB^U_{t-1}$	1.175***	2.397***	4.420***	11.630***
$Bid\ Inner^U_{t-1}$	1.642***	2.089***	2.585***	3.781***
$Ask\ Inner^U_{t-1}$	-1.714***	-2.135***	-2.625***	-3.813***
$Bid\ Outer^U_{t-1}$	-0.072	-0.075	-0.005	0.156
$Ask\ Outer^U_{t-1}$	0.102	0.094	0.063	-0.068
Adjusted R^2	1.50%	1.44%	1.47%	1.68%
# of stock-days	684,531	683,138	683,801	682,742
Average # of stocks	307	306	307	306

Table 3.7. Realized volatility portfolios and intraday return predictability from surprises in *MOIB* and *LOB* (continued)

Panel B: Adjusted R^2 decomposition (dependent variable: Ret_t)

	RV1 (low)		RV2		RV3		RV4 (high)	
	Adjusted R^2		Adjusted R^2		Adjusted R^2		Adjusted R^2	
	Absolute	Relative	Absolute	Relative	Absolute	Relative	Absolute	Relative
Constant								
Ret^U_{t-1}	0.48%	32.07%	0.45%	31.67%	0.46%	31.43%	0.54%	32.15%
$MOIB^U_{t-1}$	0.29%	19.42%	0.29%	20.08%	0.31%	21.01%	0.38%	22.88%
$Bid\ Inner^U{}_{t-1}$	0.21%	14.00%	0.20%	13.82%	0.20%	13.44%	0.21%	12.52%
$Ask\ Inner^U{}_{t-1}$	0.22%	14.95%	0.21%	14.78%	0.21%	14.35%	0.22%	13.17%
$Bid\ Outer^U{}_{t-1}$	0.15%	9.78%	0.14%	9.80%	0.15%	9.92%	0.16%	9.68%
$Ask\ Outer^U{}_{t-1}$	0.15%	9.78%	0.14%	9.86%	0.14%	9.84%	0.16%	9.60%
$Total\ Inner^U$	0.43%	28.95%	0.41%	28.60%	0.41%	27.79%	0.43%	25.69%
$Total\ Outer^U$	0.30%	19.56%	0.28%	19.66%	0.29%	19.76%	0.32%	19.28%
$Total\ LOB^U$	0.73%	48.51%	0.69%	48.26%	0.70%	47.55%	0.75%	44.97%
$Total^U$	1.50%	100.00%	1.44%	100.00%	1.47%	100.00%	1.68%	100.00%
# of stock-days	684,531		683,138		683,801		682,742	
Average # of stocks	307		306		307		306	

liquidity for high volatility stocks and to supply liquidity for low volatility stocks.

3.6 Conclusion

The recent public debates regarding algorithmic traders (and their subset — high-frequency traders) adversely selecting retail investors highlighted the importance of understanding how the informed trading is taking place and how it was affected by the emergence of algorithmic trading. Motivated by this, I investigate the intraday return predictability from informed market orders and informed limit orders to answer the questions of whether informed traders choose to act as liquidity suppliers or liquidity demanders and what are the determinants of their choice. In particular, I study one-minute mid-quote return predictability from the lagged informed market order flow (measured by surprises in market order imbalances) and lagged informed limit orders (measured by surprises in depth concentration at the inner and outer levels of the ask and bid sides of the limit order book).

To the best of my knowledge, I am the first to address this question with such a comprehensive data set, which includes one-minute observations for all NYSE-listed common stocks for the 2002-2010 period. I show that informed limit orders are predictive of intraday returns beyond the informed market orders. Moreover, the majority of informed trading occurs via limit orders (as measured by incremental adjusted R^2 from predictive regressions). This result holds for the whole period under consideration as well as for the sub-period analysis.

I also examine the effect of algorithmic trading activity on informed trader's choice between market and limit orders. Overall, there is a relative shift from informed liquidity provision (limit orders) to informed liquidity consumption (market orders) while moving from stocks with a low presence of algorithmic traders to stocks with a high presence of algorithmic traders.

In conclusion, informed traders actively use both market orders (consume liquidity) and limit orders (provide liquidity) with the largest chunk of the informed trading happening via limit orders. This fact should not be neglected while analyzing the adverse selection effects on financial markets.

Chapter 4

Low-Latency Trading and Price Discovery without Trading: Evidence from the Tokyo Stock Exchange in the Pre-Opening Period and the Opening Batch Auction*

4.1 Introduction

During the past decade, global equity markets have been fundamentally altered due to the vast improvements in the speed of trading and the consequent fragmentation of market activity. For example, on January 4, 2010, the Tokyo Stock Exchange (TSE) launched a new trading system named "Arrowhead", which has reduced the order submission response time to 2 milliseconds. 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. Among other changes, traditional market makers have been replaced by high-frequency traders (HFTs) in most markets.[1] This replacement has had a dramatic impact on the

*This chapter is based on Bellia, Pelizzon, Subrahmanyam, Uno, and Yuferova (2016) "Low-Latency Trading and Price Discovery without Trading: Evidence from the Tokyo Stock Exchange in the Pre-Opening Period and the Opening Batch Auction" (available at http://ssrn.com/abstract=2705962). We are grateful to Jonathan Brogaard, Björn Hagströmer, Frank Hatheway, 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, and the SAFE Microstructure workshop, Goethe University, for helpful suggestions. We also thank the Tokyo Stock Exchange for providing anonymous detailed account-level data, which form the basis of the research reported 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]See Brogaard (2010), Jovanovic and Menkveld (2015), Hendershott and Riordan (2013), and Raman and Yadav (2014), for evidence of this.

behavior of liquidity providers in financial markets. 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. Other types of regulations more directly target the types of behavior displayed by HFTs, such as the minimum display time for limit orders and the relative frequency of cancellations of trades. Recent theoretical work by Budish, Cramton, and Shim (2015) advocates frequent batch auctions instead of a continuous auction, while Fricke and Gerig (2015) analyze the optimal interval of auction cycle. These papers are theoretical justifications, but need empirical verification before any clear conclusion can be drawn about the relative merits of frequent batch auctions compared to the traditional continuous trading.

The existing empirical literature on HFTs focuses on trader behavior during the *continuous* trading session. This paper instead studies whether, in the pre-opening period without trading, low-latency traders (HFTs) still participate in the equity market, and how the presence of low-latency traders contributes to price discovery in the subsequent opening call auction. To our knowledge, there are no other papers that investigate the role of HFTs in the *pre-opening* period and shed light on the potential role of HFTs in periodic *batch auctions*. In this paper, we aim to contribute to the literature on low-latency trading, with a clear focus on price discovery in the opening batch auction period. Our motivation for filling this void in the literature is that the pre-opening period has very different characteristics to the continuous session. 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. Given the growing presence of low-latency traders in the market, the manner in which price discovery occurs during the pre-opening period is a crucial issue to investigate.

The main questions we address in this paper are related to the role of low-latency

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

traders (including HFTs) in the pre-opening period and the difference in trader behavior between the opening of the call auction and the continuous trading session that follows. More specifically, we investigate whether, in the absence of trading, low-latency traders (including HFTs) still participate in the market *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 session that follows, and (ii) how and precisely when they participate during the opening batch auction period. Finally, and more importantly, we investigate how the presence of low-latency traders contributes to price discovery in the opening batch auction period and the following continuous session, and compare the behavior of the low-latency traders that do and do not participate in the opening call auction during the continuous session that ensues. In order to empirically investigate these questions, we use a unique dataset provided by the 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)).

In the TSE, the execution of orders is not permitted during the pre-opening period, hence buy/sell schedules can be crossed. In fact, traders cannot seek immediacy in this period; hence, low-latency traders, 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 a potentially smaller presence of HFTs in the opening batch auction period, although this warrants empirical scrutiny. Therefore, it is interesting to investigate the incentives and behavior of low-latency traders during these periods.

There are potentially also several ways to settle an opening batch auction. In most markets, and in the TSE, there is no time priority for limit orders submitted during the pre-opening period. As long as the limit price is identical to other pending buy (or sell) orders, the time of order submission does not affect the execution of orders at the opening call auction. This feature may cause traders to delay order submission until just before market opening. For example, institutional investors that are interested in executing large orders at market opening may enter them into the order book at the very last moment (perhaps the last millisecond prior to opening). The early entry of large orders during the pre-opening period has clear disadvantages: large orders attract other participants and induce other investors to react sooner, causing a deterioration in the execution price of such orders. Hence, these large orders may have a significant impact on the opening price.[3] The issue of whether or not a low-latency trading environment amplifies this order placement behavior has not been investigated so far. Nor, indeed, have researchers looked into whether low-latency traders strategically decide upon the timing of their order submissions during the pre-opening period and how this might affect price discovery.

Further, the cancellation of existing orders is possible at *any* time prior to the opening time and is free of charge, so that a trader with access to a low-latency trading facility may wait until the very last moment before the opening time, if they wish to cancel. Some investors may enter "noisy" orders and cancel them right before execution occurs. The term "noisy" connotes a type of order that uses an aggressive limit price to send a signal to investors on the opposite side, to induce them to provide liquidity. Indeed, some investors may have an incentive to enter false orders with aggressive limit prices to elicit a favorable response from true orders on the opposite side of the limit order book. While this strategy does not always work to the advantage of the aggressive investor, it may serve to add noise to the pre-opening quotes. Since a low-latency environment allows traders to delay their final action until very close to market opening, the noise effects may prevail right until the final seconds of the pre-opening period. If that is so, it will be useful to investigate which *order type* causes a deterioration of the pre-opening quotes.

A low-latency trading environment influences not only the behavior of HFTs but also other types of low-latency trading, such as algorithmic trading, which motivates us to develop a more comprehensive classification of traders than in the prior literature, and to investigate the behavior of all the different categories of traders, based on their capability for low-latency trading. This is in contrast to the rapidly growing

[3] This empirical evidence is documented by Kraus and Stoll (1972), Chan and Lakonishok (1993), and Chiyachantana, Jain, Jiang, and Wood (2004) in earlier studies of the price impact of institutional trades.

empirical literature on HFTs, which is largely based on HFT datasets[4] that provide limited coverage of HFT activity and rarely provide account-level data; this prevents researchers from identifying the specific series of actions taken by individual HFTs. Even though account-level data have become available more recently, the identification of HFTs is, in most cases, based on screening using just a couple of metrics, such as the order-to-cancellation ratio. It goes without saying that the thresholds for the metrics used in such classifications are fairly arbitrary. Indeed, a report by the Securities and Exchange Commission (SEC (2014)) argues that the current metrics used to identify HFT activity (as in, e.g., Kirilenko, Kyle, Samadi, and Tuzun (2015)) can be too narrow to capture the true range of activity in a low-latency environment. In particular, the SEC (2014) emphasizes that not all low-latency and high-frequency trading activity should necessarily be classified as HFT activity; rather, HFT activity is a subset of a more general phenomenon of algorithmic trading, and should be studied as such. In this study, we take this broad criticism into account and undertake a more comprehensive analysis of trading strategies employed by various trading entities, avoiding referring to all of them as HFTs, given that we do not yet have a commonly accepted framework for defining and identifying HFTs. We adopt an entirely different methodology from those used by prior researchers to identify low-latency trading activity, 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. Such detailed data have not previously been used in the literature, to our knowledge.[5] The unique dataset used in this paper is one of the most comprehensive ones on HFT used in the literature, thus far. Hence, it offers several advantages for researchers that are worth highlighting.

First, our data relate to trading information at the disaggregated level of indi-

[4]HFT datasets are datasets provided by exchanges themselves, e.g. the NASDAQ dataset. Typically, these datasets include HFT/non-HFT flags for each order submission.

[5]The study that is closest to ours is by Brogaard, Hagströmer, Norden, and Riordan (2015) and uses 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 as we do.

vidual servers used for trading. Based on server usage, we are, therefore, able to infer account level trading, without resorting to arbitrary criteria as in prior studies, which define classifications based on arbitrary thresholds of latency of trading and inventory to identify the type of trader. Hitherto, no study has examined server configuration, which is a crucial determinant of the horse power of execution capability. Consequently, prior studies in the microstructure literature, which did not have access to such disaggregated data, were forced to rely on either a HFT/non-HFT flag or, when they did use account level information, or were able to cover only a small sample of the market and, even then, are typically focused on the continuous session.

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 session that ensues, based on the type of trader. In turn, our data allow us to measure the impact of different types of traders on price discovery and liquidity provisions. Thus, only with our data can one shed some light on the consequences of slowing the trading down, from continuous trading to batch auctions, as suggested by Budish, Cramton, and Shim (2015).

Third, our data permit a comprehensive classification scheme, which applies to the trading data on the stock-day 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 a HFT, forever a HFT."*

Fourth, a further advantage of our data set is the availability of account-level information during the pre-opening period, which allows us to investigate how price discovery takes place without trading and which trader type is responsible for it.

Using the granular data available to us, we classify traders into twelve subgroups based on latency and inventory behavior during the continuous session. In terms of speed, we identify three subgroups, namely FAST, MEDIUM, and SLOW, based on latency; 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; in contrast, we show that speed and inventory actually exhibit low correlation (with a Pearson correlation coefficient equal to 0.12). We also show that *both* FAST / SMALL traders (market makers) and FAST / LARGE traders (position takers) can be FAST traders. Thus, it is important to take *both* the speed and the inventory dimensions into account in order to identify low-latency (high-frequency) trading activity, which justifies our 3 x 4 classification into 12 groups for the detailed analysis.

Our novel database allows us to investigate and compare, in depth, the behavior of the different types of traders. Our analysis shows that traders generally exhibit different types of behavior across stocks and over time. This means that the usual characterization of a trader acting as an HFT, for all time and for all stocks, is likely to be invalid. In particular, we observe that, on average, only in 28% of cases do traders remain in the same group, among the 12 described above, from one active day to the next, for a particular stock. Moreover, FAST / SMALL and FAST / MEDIUM, as well as MEDIUM / SMALL and MEDIUM / MEDIUM, traders exhibit wide variation in their activity from stock to stock during the pre-opening period. This pattern is especially strong for FAST / SMALL traders (high-frequency market makers): their relative representation in the overall sample varies from 4.54% to 60.05%.

Our empirical results for the TSE show that FAST traders participate in the pre-opening period to a lesser extent than in the continuous session. Only 27.4% of FAST / SMALL traders, 33.7% of FAST / MEDIUM traders, and 16.8% of FAST / LARGE traders participate in the pre-opening period. These percentages are smaller than those for MEDIUM / SMALL (50.4%), MEDIUM / MEDIUM (50.0%), and MEDIUM / LARGE traders (18.6%). However, with respect to the total number of orders in the pre-opening period, FAST traders that participate play a dominant role in the pre-opening period, submitting 51% of them, while MEDIUM and SLOW traders submit 42% and 7%, respectively. Furthermore, FAST traders submit 36% out of their 51% of orders in the first 10 minutes of the pre-opening period, and 8% of their orders in the last 10 minutes. One reason for submitting orders as early as 8 am may be that traders, such as index arbitrageurs, seek a higher execution probability for their orders (time priority matters for orders with limit price equal to the opening price). In addition, 32.4% of aggressive orders, which influence the mid-quotes in the pre-opening period, are submitted by FAST / SMALL traders. This indicates that their order submission strategy contributes to the price discovery process through their seeking of a higher probability of order execution.

We quantify price discovery by means of the weighted price contribution (WPC) as in the previous literature.[6] The WPC is the weighted percentage amount by which an incoming aggressive order moves the prevailing mid-quotes closer to the opening price over the accumulated price discovery contribution during the pre-opening period. We analyze the price discovery contribution of the 12 groups described above (i) by order, (ii) in the cross-sectional analysis, and (iii) with a panel specification. We find that, both in the by-order and the cross-section of stocks, FAST / SMALL traders (high-frequency market makers) and FAST / MEDIUM traders, as well as MEDIUM / SMALL and MEDIUM / MEDIUM traders, are those that contribute the most to price discovery. Besides that, we show that these four groups of traders strategically choose the stocks in which to participate, by taking into account the stocks' characteristics, such as market capitalization, liquidity, and volatility. These results indicate that low-latency traders contribute to price discovery and lead the price formation process throughout the pre-opening period, in particular after the first 10 minutes. The by-order analysis shows that these 12 groups of traders largely contribute to price discovery with their intense activity in new limit orders and price revisions. Cancellation of limit orders deteriorates price discovery, but cancellation of market orders improves price discovery. These results are confirmed by the panel analysis in which both the time-series and cross-sectional dimensions are taken into consideration, in addition to the stock and time fixed effects. The role of low-latency traders in price discovery is also confirmed by a test for the unbiasedness of the pre-opening quotes.

Inspired by the active discussion on whether continuous trading or frequent batch

[6] See Barclay and Warner (1993), Cao, Ghysels, and Hatheway (2000), and Barclay and Hendershott (2003).

auctions constitute a better market design in the presence of low-latency traders, as suggested by Budish, Cramton, and Shim (2015), we investigate the difference between the behavior of low-latency traders that participate in the opening call auction, and that of those participating only in the continuous session that ensues. We acknowledge that frequent batch auctions are qualitatively different from the opening call auction in important ways, e.g. the degree of information dissemination and the ability to quickly unwind positions after the auction has taken place. However, the opening call auction is the closest approximation to the frequent batch auctions one sees today in developed (major) equity markets. We find that low-latency traders that are active in the call auction do not aid price discovery during the first 30 minutes of the continuous session but, if anything, slightly deteriorate it. However, they remain the main liquidity providers. Low-latency traders that are active only during the continuous session are the main contributors to the price discovery process and also the main consumers of liquidity.

The outline of the paper is as follows. In Section 4.2, we survey the literature on price discovery and HFTs, particularly relating to the pre-opening period. In Section 4.3, we provide a description of the TSE market architecture and the special features of our database. In Section 4.4, we present our empirical design and, in particular, our data-filtering procedures used to identify the 12 trader groups based on activity during the continuous session. Our empirical analysis and results are presented in Section 4.5. Section 4.6 concludes.

4.2 Literature review

The recent HFT-specific theoretical literature deals with the speed advantage of HFTs in terms of information processing and trading. Most of it focuses only on the continuous trading session. 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 (2015) 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 (2015), 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 low-latency traders. Our paper provides empirical insights on HFT behavior in the batch auction setting.

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. This 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 findings regarding the microstructure of trading activity in the market pre-opening period, while the second relates to the impact of HFTs on price discovery. 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)). However, 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 sessions. 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 low-latency trading environment, well known in the literature for encouraging HFT activity. Moreover, those authors do not distinguish between the different types of traders.

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,[7] although there was no low-latency trading in the period they consider.

The body of empirical studies on HFT trading activities is growing rapidly.[8] It should be noted, however, that the focus of most of the literature is the *continuous trading* session, 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 session. 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 session, starting with a zero-inventory position at the beginning of the trading day. Some strategies em-

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

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

ployed 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), Hendershott, Jones, and Menkveld (2011), Easley, de Prado, and O'Hara (2012), Hendershott and Riordan (2013), Malinova, Park, and Riordan (2013), Boehmer, Fong, and Wu (2015), 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 session, 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 in 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 4.4.2 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.

4.3 Institutional structure

4.3.1 Opening Call Auction and Pre-opening order submissions in the Tokyo Stock Exchange

The opening price of the TSE is determined by a single price auction ("Itayose" in Japanese) that kicks off at 9 am, based on buy and sell orders accumulated during the pre-opening period. There are two types of orders allowed on the TSE: limit orders and market orders.[9]

The principle for order matching is based on price and time priority in the continuous session. In the pre-opening period, however, time priority is ignored. That is, all orders placed before the determination of the opening price are regarded as simultaneous orders. The opening auction determines the price at which the largest amount of executions is possible. There are three conditions to be met: (1) All market orders must be executed at the opening price. (2) Orders with sell limit price higher than the opening price and buy limit price lower 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 means that, often, orders on either side whose limit price is equal to the opening price cannot be fully executed. When this happens, the TSE allocates the available shares to participating member firms on a pro-rated basis (often based on time priority).[10] When the buy/sell quantities at the best quotes do not satisfy the above three conditions for the opening price, the TSE disseminates special quotations immediately after 9 am. Special quotations are where the best ask and best bid are at the same price, while the amounts at the two quotes are different, indicating an order imbalance between buyers and sellers, inviting further new orders to bridge the gap. For our paper, cases of the opening price not having been determined at 9 am are excluded from our sample.

On the one hand, the feature of the opening call auction whereby there is no

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

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

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. On the other hand, a trader engaging in index arbitrage between cash and index futures contracts may enter a basket of orders as early as 8 am in order to enhance the execution probability. Member firms of the exchange often allocate filled limit orders, with limit price equal to the opening price, to their customers on a time-priority basis, which means that placing orders early can improve a trader's probability of execution, at least to some extent. Index arbitrageurs and institutional investors are well aware of this practice, and will take it into account in their order placement strategy. Thus, in the pre-opening period, preference over order placement timing diverges to the two extreme points: just after 8 am and just before 9 am.

Each trading day, the TSE starts receiving orders from brokers at 8 am, and does so until the single price auction for the market opening begins, at 9 am. As soon as it receives orders, the TSE disseminates the pre-opening quotes, not only the best ask and best bid, but the 10 quotes above and below the best quotes, to the market.[11] Every time it receives an order, the pre-opening quotes are refreshed. In Japan, the TSE is the exclusive venue hosting the pre-opening price formation. Two other private venues start their operations at 9 am. However, the Nikkei Stock Index Futures traded in Singapore start their trading at 8:45 am, Tokyo time, and may contribute to price discovery.

4.3.2 Server IDs and data

We use two sources of data for analysis. First, order data covering the complete history of an order (new entry, execution, revision of quantity or price, and cancellation in the pre-opening and continuous trading periods) is 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). Second, tick-by-tick quotes information in the pre-opening pe-

[11] In the pre-opening period, according to the TSE's definition of the best ask and the best bid, the amount of orders displayed at the best ask (bid) includes all limit sell (buy) orders below the best ask (above the best bid). A subscriber to the full quotes service can see information (price and quantity) on the entire book. However, the quantities for the best ask and the best bid are the same as for the standard service.

riod is obtained from the Thomson-Reuters Tick History (TRTH) database with a millisecond time stamp.[12]

The unique feature of this study is that we use the novel data provided by the TSE, which include the unique IDs of the VSs (Appendix C.1 describes a hypothetical setup of VSs). We find that 5,580 such servers were used in our sample period and we identify 3,021 groups, which we call traders.[13] Figure 4.1 depicts the sizes of the traders based on the number of VSs they employ. Among 3,021 traders, 329 utilize between 2 and 41 VSs, while the rest (2,692) use only a single VS.[14]

To determine the relationship between servers, we investigate the entire universe of stocks traded on the TSE's First Section (there were 1,702 stocks listed as of April 1, 2013).[15]

We also investigate the latency of the different traders. We measure latency as the minimum time that elapsed between two consecutive order submissions for the same stock. Table 4.1 presents the characteristics of the traders, based on their trading environment of 1,702 stocks. Traders with just a single server place orders on 605.8 stocks, on average, with a median latency of 2 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 number of stocks placed per server gets smaller, except between 30 and 39 servers. In general, although the number of stocks per server and the median latency are positively correlated, the median inventory varies considerably across traders, reflecting the variety of investment horizons among them.

In the TSE, some traders, such as HFTs, use multiple VSs exclusively because of a limitation on the number of messages submitted per second for each server.[16] Using

[12]We use TRTH only for the unbiasedness analysis (see Section 4.5.4).

[13]In Appendix C.1, we describe how we identify "traders".

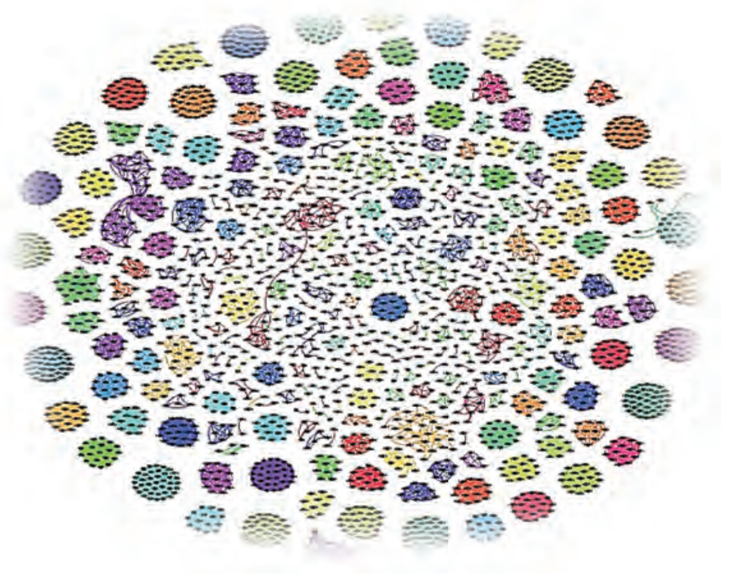
[14]In contrast to Brogaard, Hagströmer, Norden, and Riordan (2015), who use the grade of the co-location service as a categorizing device for measuring the speed requirements of traders, we focus instead on how traders configure their respective trading environments.

[15]Stocks listed in 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.

[16]The TSE provides three levels of service, with a maximum of 60, 40, and 20 messages per second,

Figure 4.1. 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-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.



5

Table 4.1. Traders' characteristics during the continuous session

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 session, identify 3,021 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, average number of servers used per trade, number of stocks traded (in total and per server), median latency (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 messages (in total and per stock), and average volume share per day (the proportion of the buy volume plus the sell volume per trading desk). These characteristics are based on the continuous session 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	2,692	213	81	19	11	5
Average # of servers	1.00	4.44	13.43	22.42	31.47	40.54
# of stocks traded in total	605.81	376.95	343.75	330.51	515.73	475.59
# of stocks traded per server	605.81	84.96	25.59	14.74	16.39	11.73
Median latency	2.024	0.214	0.012	0.002	0.005	0.001
Median inventory	100.00%	93.87%	64.89%	6.61%	49.09%	43.32%
# of messages per stock-day	8	14	48	163	138	492
Average volume share	98.54%	36.33%	27.92%	15.39%	10.99%	10.73%

multiple servers, each trader optimizes the performance of the trading operations for their subset of stocks. Some traders operate in 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 reduce their latency significantly.

4.4 Empirical design

4.4.1 Universe of stocks and sample period

We select our universe of stocks from the constituents of the TOPIX100 index, which is comprised of the stocks on the TSE's first section, with high liquidity and relatively large 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

respectively. According to a prominent 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 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.

is the trading system on this exchange.[17]

The sample period we select for our analysis lies between April 1 and May 31, 2013. In this period, the volatility of the stock market rose after the new governor of the Bank of Japan, Haruhiko Kuroda, announced the bank's new aggressive quantitative easing (QE) policy. A number of unexpected events occurred during this period, making the role of the pre-opening quotes more crucial than at any other time. 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.

Table 4.2 shows the relative frequencies of order types over the whole period and the relevant subperiods. In the entire pre-opening period, new limit orders make up about 85%, new market orders about 6% and cancellations and price revisions roughly 4% and 5%, respectively. In the last 10 minutes, and particularly the last minute of the pre-opening period, the share of new limit orders orders drops to less than 50%, and those of cancellations and price revisions of limit orders and new market orders increase accordingly.

4.4.2 HFT identification strategy

A useful guideline defining the features of HFTs has been presented by the SEC in the U.S. The 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 latency and inventory to classify

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

Table 4.2. Distribution of order flow during pre-opening period

This table shows the distribution of the order flow for 97 stocks from the TOPIX100 during the sample period of April-May 2013. We report the average number of orders across stock-days, the relative frequency of orders, and the average size of the orders, in terms of number of shares, submitted during the whole pre-opening period (8:00:00.000 - 8:59:59.999), during the last 10 minutes of the pre-opening period (8:50:00.000 - 8:59:59.999), and during the last minute of the pre-opening period (8:59:00.000 - 8:59:59.999). All orders are grouped according to their type: new orders, quantity revisions (changes in the order size), price revisions, and cancellations (withdrawals of orders) for limit and market orders, respectively. Order flow data with order IDs as well as virtual server IDs are provided by the Tokyo Stock Exchange.

	Limit Orders				Market Orders			
	New orders	Quantity Revisions	Price Revisions	Cancellations	New orders	Quantity Revisions	Price Revisions	Cancellations
Panel A: 8:00:00-8:59:59								
Average # of orders	3,402.14	20.41	186.97	122.82	236.37	3.97	6.26	18.34
Relative frequency of orders	85.11%	0.51%	4.68%	3.07%	5.91%	0.10%	0.16%	0.46%
Average size of orders in shares	1,409.39	1,263.86	1,763.68	2,000.33	2,543.83	2,421.39	1,913.40	3,029.43
Panel B: 8:50:00-8:59:59								
Average # of orders	261.64	13.89	93.32	58.23	75.53	3.43	3.98	9.66
Relative frequency of orders	50.35%	2.67%	17.96%	11.20%	14.53%	0.66%	0.77%	1.86%
Average size of orders in shares	3,125.26	1,329.18	2,079.26	2,215.51	3,186.81	2,732.71	2,183.99	3,180.41
Panel C: 8:59:00-8:59:59								
Average # of orders	74.67	7.75	29.60	22.32	17.64	2.20	2.12	3.07
Relative frequency of orders	46.85%	4.86%	18.57%	14.01%	11.07%	1.38%	1.33%	1.93%
Average size of orders in shares	3,462.24	1,945.92	2,559.50	2,565.09	3,661.92	2,989.56	2,722.44	4,218.96

traders. These two metrics are closely related to all five characteristics listed above: latency matches characteristics 1, 2, and 4 above, while inventory matches characteristics 3 and 5. Latency 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 characteristics 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 different categories are all 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 the liquidity-taking or the liquidity-making strategy to offer more profitable opportunities. Therefore, we assume that traders can engage in different types of trading strategies on a stock-by-stock and day-by-day basis.

To address this concern, we compute our metrics on a per-stock, per-day basis, for all trading desks. Our aim is to investigate how the behavior of a low-latency trader affects the pre-opening price. As far as we know, all the empirical studies in the literature except ASIC (2013) assume that HFTs behave in an identical manner on every day and for every stock.

4.4.2.1 Latency

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 of new orders, cancellations, and revisions during our sample period), as a measure of latency.[18] A realization of low latency has to be supported by the appropriate trader's trading infrastructure. Hence, the number of servers a trader uses is a crucial determinant of latency. As noted earlier, we observe

[18]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, unlike us, their data do not enable them to identify individual traders.

varying numbers of servers, ranging from 1 to 41, in our sample period. We also find that the number of stocks allocated to an individual server is associated with the latency of the trader and vice versa. Appendix C.2 provides a detailed analysis of the relationship between latency and messages per server.

4.4.2.2 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.

4.4.2.3 Classification

We classify all traders according to observed latency and inventory during the continuous trading session for each stock-day. We apply the following classification scheme: We divide all traders, based on their latency, into three groups: FAST, MEDIUM, and SLOW. For each stock-day, the SLOW group includes traders with a latency greater than 60 seconds. We then look at the remainder of the latency distribution and split it relative to the median. Therefore, the FAST group includes traders whose latency is smaller than the median, and the MEDIUM group includes traders whose latency is greater than the median but smaller than or equal to 60 seconds. Where we are unable to compute the latency due to the absence of multiple orders for the same stock on the same day, we treat the trader as a SLOW trader.

We divide all traders into four groups based on their inventory for each stock-day: LARGE, MEDIUM, SMALL, and NOTRADE. 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

Table 4.3. 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 session on the same day. First, we divide all traders into 3 groups based on their latency (minimum time elapsed between two consecutive orders for the same stock): FAST, MEDIUM, 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 latency below the median (excluding all trader-stock-days for which the minimum latency is higher than 60 seconds)
	MEDIUM	Traders with latency above the median (excluding all trader-stock-days for which the minimum latency is higher than 60 seconds)
	SLOW	Traders with latency 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)

we differentiate a trader who ends the day with a flat inventory as a result of buy and sell activity throughout a day from a NOTRADE agent. It should also be noted that NOTRADE agents include traders who submit orders, but whose orders are not filled. Table 4.3 briefly summarizes our scheme, while Table 4.4 Panel A shows the summary statistics for latency and inventory for each group under our classification procedure.

The average latency in the FAST group varies across different inventory subgroups from 0.02 seconds to 0.04 seconds. The MEDIUM speed group exhibits a much higher latency, ranging from 9.41 to 12.73 seconds. The SLOW group has an average latency above 2,000 seconds. By construction, the LARGE inventory subgroup always has a 100% inventory, meaning that, during the day, traders either purely buy or purely sell the stock. Traders from the MEDIUM inventory subgroup tend to end their

Table 4.4. Description of traders' characteristics

This table shows summary statistics for the classification of the traders during the continuous session according to the scheme proposed in Table 4.3 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 in the pre-opening period, but do not trade in the opening call auction (Active-w/o-Trade), and traders that participate in the pre-opening period and trade in the call auction (Active-w-Trade). Panel A describes characteristics of all traders; Panel B, Panel C, and Panel D describe characteristics of Non-Active, Active-w/o-Trade, and Active-w-Trade traders. We report the average number of traders per stock-day, average latency per trader-stock-day, inventory per trader-stock-day, average number of new orders per trader-stock-day, average trade-to-(new) order ratio (even partial execution of orders is included), cancellation ratios of new orders, proportion of activity during pre-opening period and continuous session (ratio of messages for each trader group divided by the total number of messages during the pre-opening or continuous period, excluding trade messages), proportion of total trading activity (ratio of trade messages for each trader group divided by the total number of trade messages during the pre-opening or continuous period), and the presence ratio (the proportion of traders that are active during both the pre-opening and continuous sessions). 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: Characteristics of all traders

Speed	Inventory	Average # of traders	Average latency	Average inventory	Average # of new orders	Average trade-to-order ratio	Average cancellation ratio	Activity during pre-opening period	Activity during continuous session	Trading activity	Average presence ratio	Presence ratio (P5%)	Presence ratio (P95%)
FAST	LARGE	78.44	0.04	100.0%	28.62	36.4%	58.6%	1.9%	8.8%	5.8%	16.8%	5.9%	34.4%
	MEDIUM	90.69	0.03	66.8%	86.13	46.3%	48.8%	15.7%	24.1%	26.1%	33.7%	17.9%	55.9%
	SMALL	93.85	0.02	16.4%	182.42	42.8%	49.8%	34.0%	48.6%	41.3%	27.4%	10.8%	56.4%
	NOTRADE	42.98	0.04		14.23		89.2%	0.2%	3.4%	0.0%	7.1%	0.0%	27.3%
MEDIUM	LARGE	97.98	11.63	100.0%	8.99	49.8%	41.3%	2.6%	2.4%	2.8%	18.6%	7.4%	34.1%
	MEDIUM	80.80	10.29	65.7%	28.57	55.1%	25.6%	16.6%	4.4%	8.2%	50.0%	29.6%	71.3%
	SMALL	75.86	9.41	17.0%	41.72	56.0%	22.7%	22.3%	5.7%	11.3%	50.4%	29.7%	70.3%
	NOTRADE	49.68	12.73		3.83		85.3%	0.1%	0.9%	0.0%	5.9%	0.0%	17.9%
SLOW	LARGE	214.47	4035.39	100.0%	2.11	83.8%	3.4%	2.1%	0.7%	2.3%	16.5%	8.9%	26.3%
	MEDIUM	43.00	2393.04	64.9%	6.49	76.7%	3.9%	2.1%	0.4%	1.3%	38.6%	14.4%	65.7%
	SMALL	34.65	2398.59	16.2%	6.15	76.7%	4.7%	1.5%	0.3%	1.0%	34.4%	10.9%	61.1%
	NOTRADE	37.55	2579.22		1.88		41.8%	0.7%	0.1%	0.0%	41.8%	17.9%	69.0%

Table 4.4. Description of traders' characteristics (continued)

Speed	Inventory	Average # of traders	Average latency	Average inventory	Average # of new orders	Average trade-to-order ratio	Average cancellation ratio	Activity during pre-opening period	Activity during continuous session	Trading activity
Panel B: Characteristics of Non-Active traders										
FAST	LARGE	65.34	0.04	100.0%	24.82	35.9%	59.4%		6.9%	3.7%
	MEDIUM	60.18	0.02	65.6%	51.27	42.8%	53.5%		10.7%	8.2%
	SMALL	67.51	0.02	16.0%	170.11	38.7%	55.5%		35.7%	19.5%
	NOTRADE	40.28	0.04		14.15		89.8%		3.3%	
MEDIUM	LARGE	80.25	11.23	100.0%	7.16	50.6%	42.6%		1.8%	1.7%
	MEDIUM	41.60	8.34	66.2%	17.25	56.3%	33.7%		1.8%	2.4%
	SMALL	37.47	8.35	16.2%	27.00	57.0%	30.2%		2.2%	3.0%
	NOTRADE	46.77	12.77		3.78		86.2%		0.8%	
SLOW	LARGE	179.71	4267.25	100.0%	1.72	87.5%	2.8%		0.6%	1.7%
	MEDIUM	27.90	2888.19	64.4%	4.17	85.2%	2.9%		0.2%	0.7%
	SMALL	23.43	2813.85	16.2%	4.18	84.7%	3.8%		0.2%	0.6%
	NOTRADE	21.76	2442.22		1.87		56.7%		0.1%	
Panel C: Characteristics of Active-w/o-Trade traders										
FAST	LARGE	8.02	0.04	100.0%	27.67	27.0%	66.7%	0.7%	0.8%	0.5%
	MEDIUM	9.54	0.03	67.4%	60.04	40.4%	55.2%	1.2%	1.9%	1.6%
	SMALL	8.66	0.03	16.6%	103.95	42.8%	53.7%	1.9%	2.6%	3.2%
	NOTRADE	2.70	0.04		15.42		81.0%	0.2%	0.2%	
MEDIUM	LARGE	10.49	13.71	100.0%	10.56	36.5%	42.9%	0.8%	0.3%	0.3%
	MEDIUM	13.12	14.96	65.4%	21.55	49.7%	22.7%	2.5%	0.5%	1.0%
	SMALL	11.62	13.63	17.2%	22.07	52.0%	23.7%	2.1%	0.5%	0.9%
	NOTRADE	2.90	12.03		4.68		70.4%	0.1%	0.1%	
SLOW	LARGE	18.06	3010.33	100.0%	4.83	48.0%	9.1%	1.2%	0.1%	0.3%
	MEDIUM	7.03	1213.36	64.7%	10.32	54.3%	6.9%	0.9%	0.1%	0.3%
	SMALL	5.50	1425.63	15.8%	9.42	53.8%	8.7%	0.6%	0.1%	0.2%
	NOTRADE	15.78	4001.93		1.89		21.2%	0.7%	0.0%	
Panel D: Characteristics of Active-w-Trade traders										
FAST	LARGE	5.08	0.03	100.0%	78.99	57.7%	36.2%	1.2%	1.1%	1.7%
	MEDIUM	20.97	0.02	69.9%	198.01	59.0%	32.4%	14.5%	11.5%	16.2%
	SMALL	17.68	0.03	17.6%	267.87	58.4%	26.0%	32.2%	10.4%	18.6%
	NOTRADE									
MEDIUM	LARGE	7.25	13.11	100.0%	27.00	59.4%	24.2%	1.7%	0.4%	0.8%
	MEDIUM	26.08	11.04	65.0%	50.16	55.8%	14.0%	14.1%	2.0%	4.9%
	SMALL	26.77	9.07	18.0%	70.84	56.3%	11.6%	20.2%	3.0%	7.3%
	NOTRADE									
SLOW	LARGE	16.70	2989.36	100.0%	3.38	82.4%	2.9%	0.9%	0.0%	0.3%
	MEDIUM	8.07	1588.21	66.7%	11.18	66.8%	4.5%	1.2%	0.1%	0.4%
	SMALL	5.71	1454.01	16.7%	11.08	66.2%	4.6%	0.9%	0.1%	0.2%
	NOTRADE									

trading day with an inventory around 66%, while traders from the SMALL inventory subgroup can end up with inventory as low as 16%.

Based on the speed and inventory classifications, one can consider FAST / SMALL traders as HFT market makers, while FAST / MEDIUM traders could be viewed as HFT position takers. These two groups tend to submit more new orders per stock-day, on average, than any other group, with the greatest amount of new order traffic coming from HFT market makers (182.42 new orders per stock-day). The highest cancellation ratios are, however, a distinctive feature of the NOTRADE inventory subgroups (more than 80% for FAST / NOTRADE and MEDIUM / NOTRADE traders, and more than 40% for SLOW / NOTRADE traders). As one would intuitively expect, these traders are active during both the pre-opening and continuous trading periods, although they cancel their orders before the opening call auction on that particular day. (The latter can also be observed from the trade-to-order ratio, which equals 0%). In order to avoid undesirable execution, these traders have to cancel their orders more often than any other group of traders. The trade-to-order ratio is the highest in the SLOW group of traders (above 75%) and the lowest in the FAST group of traders (around 40%), excluding those in the NOTRADE group.

4.4.2.4 FAST trader participation pattern

Table 4.4 allows us to answer the first question we aim to investigate in this paper: Do low-latency traders participate in the pre-opening period? If so, do they do so with the same intensity as in the continuous session? Table 4.4 Panel A shows that low-latency traders do indeed participate in the pre-opening period but that the participation rates of the three FAST trader classes are smaller in the pre-opening period than in the continuous sessions. For example, of FAST / SMALL traders that participate in the continuous session, on average only 27.4% also participate in the pre-opening period. This means that about three quarters of the low-latency traders do not participate in the pre-opening period, but do participate in the continuous trading regime. An examination of the stock-level presence ratio at the 95th percentile shows that 56.4% of FAST / SMALL traders are present, which is more than double the average. This indicates that these traders select stocks in which to participate for the day.

Next, we split traders into three categories: traders who do not participate in the pre-opening period (Non-Active), traders who participate during the pre-opening period, but do not trade in the opening call auction (Active-w/o-Trade), and traders who participate during the pre-opening period and also trade in the call auction (Active-w-Trade). Panels B, C, and D of Table 4.4 show the traders' characteristics for these three categories. We focus our attention on FAST / SMALL traders. First, the average number of Non-Active traders is higher than the average number of traders who are active during the pre-opening period. Second, the average latency, inventory, number of new orders submitted during the continuous session, and trading activity are comparable between FAST / SMALL Non-Active and FAST / SMALL Active-w-Trade traders. However, these two groups are different in terms of the trade-to-order and cancellation ratios. In particular, we observe a higher trade-to-order ratio and a lower cancellation ratio for FAST / SMALL (Active-w-Trade) traders than for FAST / SMALL (Non-Active) traders. These findings suggest that there is a difference between the trading strategies employed by low-latency traders who are active and by those who are not active during the pre-opening period.

Besides that, we also compare traders who always participate in the pre-opening period, traders who sometimes participate in the pre-opening period, and traders who never participate in the pre-opening period. The results are generally in line with the previous analysis. Compared to those traders within the same category who do not participate in the pre-opening period, "always"-participating FAST / SMALL traders have relatively low cancellation-to-order ratios and higher trade-to-order ratios. "Never"-participating and "sometimes"-participating FAST / SMALL traders are lower-latency traders with higher cancellation-to-order and lower trade-to-order ratios (see Appendix C.3 and in particular Table C.3.1).

We emphasize that we use information from the continuous session 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 4.5 for the transition frequency matrix of trader strategies). In particular, on average, only in 28.12% 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 (52.44%). Among FAST traders, the highest persistence is observed for the FAST / SMALL group (41.87%). 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 63.44% of the cases. Traders tend to remain in the same inventory group in 46.96% of the cases, on average, ignoring the speed dimension, with the largest contribution to this persistence coming from the LARGE inventory group.

For comparison purposes, we also 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 Appendix C.4, this classification does not identify low-latency traders and their activity during the pre-opening period.

4.5 Empirical Analysis

4.5.1 Pre-opening and opening batch auction order flow

As explained in Section 4.3.1, the pre-market-opening period of the TSE starts at 8 am. All member firms begin to send orders from their customers' and their own accounts to the exchange. Figure 4.2 Panel A shows all order submissions entered every second as a percentage of the total number of orders during the pre-opening period.

The results from the three different trading-speed groups are reported in Panel A. The green line represents orders from FAST traders, who play a dominant role during the whole pre-opening phase. FAST traders submit 50.5% of the total number of orders in the pre-opening period, with MEDIUM and SLOW traders submitting 42.5% and 7.0%, respectively. In the first 10 minutes of the pre-opening period, 73.7% of the total number of orders of the entire pre-opening period are submitted. FAST traders submit 36.0% out of their 50.5% of orders in the first 10 minutes, MEDIUM

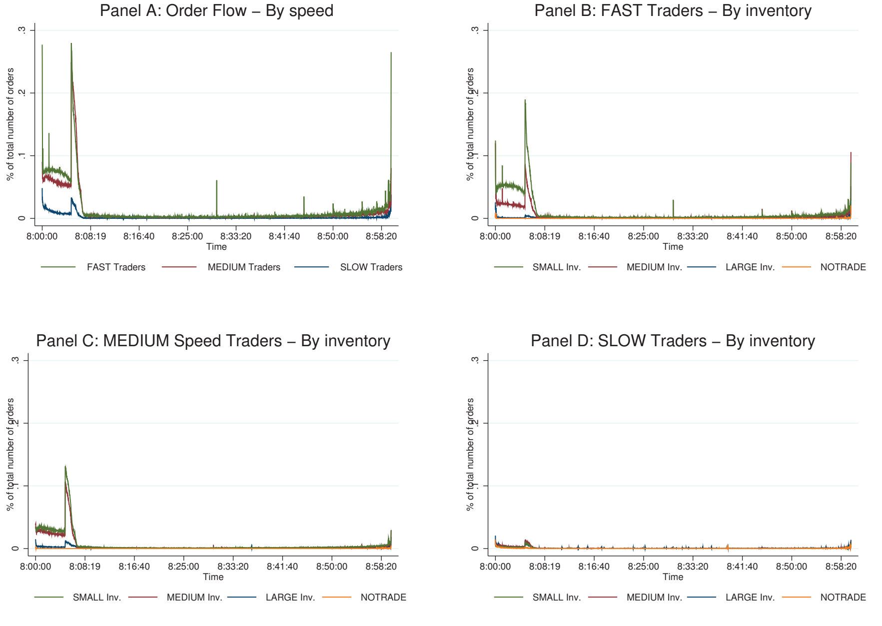
Table 4.5. Transition matrix for trader classification

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 4.3, using information about speed and inventory from the same day's continuous session. 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				MEDIUM				SLOW			
Date $t-1$		LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE
FAST	LARGE	24.83%	15.04%	10.15%	8.20%	12.91%	5.34%	3.60%	4.87%	10.63%	1.51%	1.20%	1.73%
	MEDIUM	12.82%	31.47%	23.84%	1.73%	6.75%	9.01%	6.76%	1.25%	3.85%	1.14%	0.89%	0.49%
	SMALL	8.29%	23.28%	41.87%	1.56%	4.11%	6.46%	8.69%	0.93%	2.60%	0.89%	0.92%	0.39%
	NOTRADE	15.14%	3.66%	3.51%	35.84%	7.28%	1.45%	1.24%	16.63%	9.74%	0.88%	0.84%	3.79%
MEDIUM	LARGE	10.24%	6.45%	4.08%	3.18%	24.33%	9.80%	6.38%	8.25%	18.73%	2.99%	2.32%	3.25%
	MEDIUM	4.96%	10.26%	7.52%	0.73%	11.38%	24.33%	20.46%	2.25%	7.96%	5.25%	3.82%	1.08%
	SMALL	3.68%	8.02%	10.52%	0.66%	8.02%	21.78%	29.29%	1.88%	6.53%	4.26%	4.19%	1.18%
	NOTRADE	7.88%	2.34%	1.84%	14.93%	16.55%	3.80%	2.97%	23.42%	16.52%	1.41%	1.39%	6.97%
SLOW	LARGE	4.11%	1.80%	1.28%	1.82%	9.16%	3.41%	2.58%	3.85%	52.44%	7.62%	5.84%	6.08%
	MEDIUM	2.73%	2.57%	2.10%	0.74%	6.99%	10.16%	7.66%	1.50%	36.27%	15.06%	10.75%	3.46%
	SMALL	2.77%	2.46%	2.74%	0.94%	6.78%	9.11%	9.49%	1.86%	34.55%	13.32%	11.87%	4.10%
	NOTRADE	3.85%	1.29%	1.03%	4.79%	9.19%	2.61%	2.56%	9.56%	34.25%	4.16%	4.00%	22.70%

Figure 4.2. Flow of total order submission in the pre-opening period

This figure depicts the second-by-second order flow for 97 stocks from the TOPIX100 during the sample period of April-May 2013. The Tokyo Stock Exchange starts receiving orders at 8 am and starts the call auction at 9 am. The average percentage of the total number of orders is the total number of orders in each second divided by the total number of orders submitted during the whole pre-opening period (8:00:00.000 - 8:59:59.999). The *Y*-axis represents the percentage of the total number of new orders in the pre-opening period, and the *X*-axis represents the time in seconds between 8 am and 9 am. Panel A depicts the average percentage of the total number of orders by speed group, as defined in Table 4.3 using information about speed and inventory from the same day's continuous session, per second. Panels B, C, and D report, for each speed group, the average percentage of the total number of orders according to level of inventory, as defined in Table 4.3. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.



and SLOW traders submit 32.8% and 5.0%, respectively. The order submission intensity slows down after the first 10 minutes, and is reactivated 10 minutes before the official opening time. The high level of order submissions in the first 10 minutes partly reflects the accumulation of orders overnight. Early investors also have a desire to lead price formation for the opening call auction. Figure 2 Panels B, C, and D present the pattern of order submission activity for the FAST, MEDIUM, and SLOW traders during the pre-opening period, classified according to level of inventory for each group. They clearly show a peak at the very beginning of the period for traders with SMALL and MEDIUM levels of inventory, and another very close to the opening time for FAST / SMALL traders, vastly exceeding the number of orders submitted by slower traders. FAST traders submit 7.5% of the total number of orders in the last 10 minutes of the pre-opening period, and MEDIUM and SLOW traders submit 4.4% and 1.0%, respectively. Traders with a LARGE inventory and those in the NOTRADE group submit most of their orders at 8 am. One of the reasons traders submit more orders at 8 am is to ensure a higher probability of execution of their orders due to the time-priority-based allocation most brokers employ, as explained in Section 4.3.1.

Figure 4.3 Panel A shows the new order submissions and cancellations as a percentage of all orders submitted by FAST, MEDIUM, and SLOW traders, in the last 10 minutes of the pre-opening period. While the magnitude of the order submission differs (as the scale of the y-axis differs between FAST, MEDIUM, and SLOW traders), the pattern is quite similar for all three groups. Traders accentuate their pattern of order submission during the last three minutes of the pre-opening period. A rise in order cancellations (indicated by the black line) happens suddenly, one second before 9 am, for all trader groups. For instance, the percentage of cancellation messages increases from less than 0.1% to 0.9% (of the total number of orders in the pre-opening period) per second for FAST traders, and from less than 0.01% to around 0.25% for MEDIUM-speed traders.

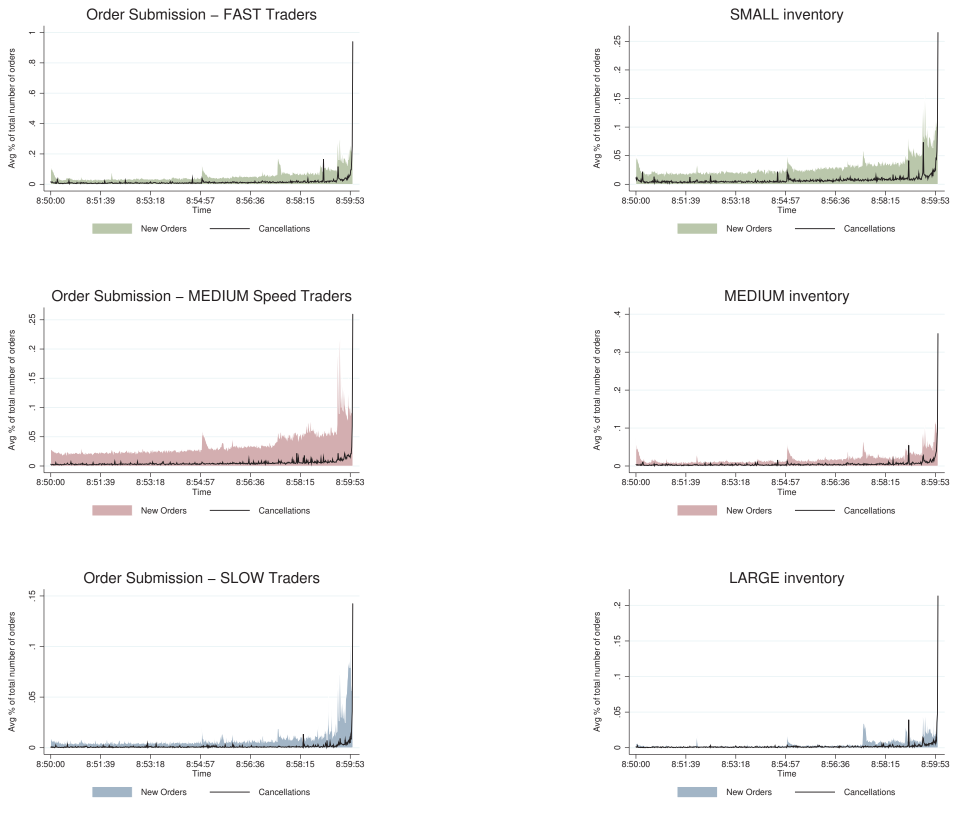
Figure 4.3 Panel B depicts order submissions and cancellations for the different inventory subgroups within the FAST group. It is interesting to note that cancella-

Figure 4.3. Flow of new orders and cancellations in the last 10 minutes of the pre-opening period

This figure depicts the second-by-second new orders and cancellations for 97 stocks from the TOPIX100 during the sample period of April-May 2013. The Tokyo Stock Exchange starts receiving orders at 8 am and starts the call auction at 9 am. New orders and cancellations are measured as the percentage of the total number of orders submitted in the last 10 minutes of the pre-opening period (8:50:00.000 - 8:59:59.999). The ***Y***-axis represents the percentage of the total number of orders in the last 10 minutes of the pre-opening period, and the ***X***-axis represents the time in seconds between 8:50 am and 9 am. Panel A depicts the average percentage of the total number of orders by speed group, as defined in Table 4.3 using information about speed and inventory from the same day's continuous session, during the sample period, per second. Panel B reports, for FAST traders, new orders and cancellations, according to the level of inventory. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: Submission of new orders and cancellations by speed category

Panel B: Submission of new orders and cancellations by FAST traders



tions from all trader groups reach their peak at the very last second. We investigate this in detail at the millisecond level and present the results in Figure 4.4 Panel A.

We confirm that the cancellations indeed occur less than one second before 9 am. As Figure 4.4 Panel B shows, the cancellation phenomenon starts at 500 milliseconds before 9 am and peaks at 130 milliseconds before 9 am. The peak is particularly pronounced for FAST traders and is not specifically related to inventory. The final action of limit price adjustment takes place just milliseconds before the opening time, which would not be possible in the absence of a low-latency trading environment.

4.5.2 Best quotes during the pre-opening period

4.5.2.1 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 session. During the continuous session, 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.

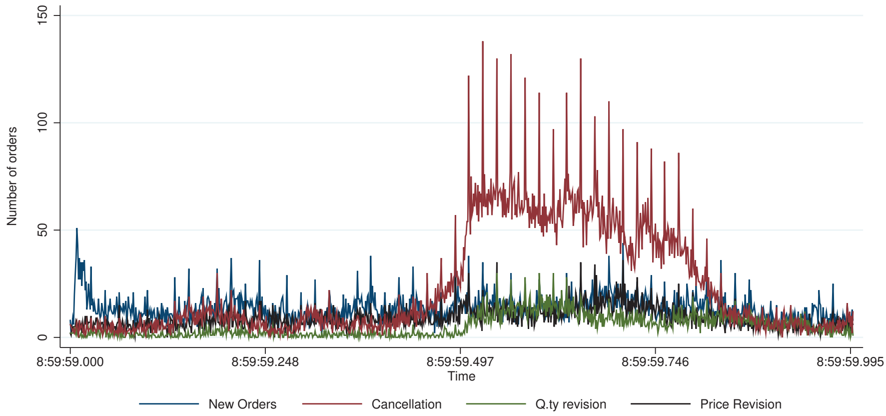
On the contrary, during the pre-opening period no execution is allowed before the opening auction, when all orders are executed at a single price. Therefore, the best bid and ask prices reported during the pre-opening period are the respective prices at which the bid (demand) and ask (supply) schedules intersect. For a detailed example, see Appendix C.5. 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. The best bid and ask prices during the pre-opening period indicate the range within which the opening price (auction price) will be determined. Therefore, we use the average of these two prices (the mid-quote) as a proxy for the single auction price.[19]

[19]We use two different sources for the best bid and ask prices in the pre-opening period. First,

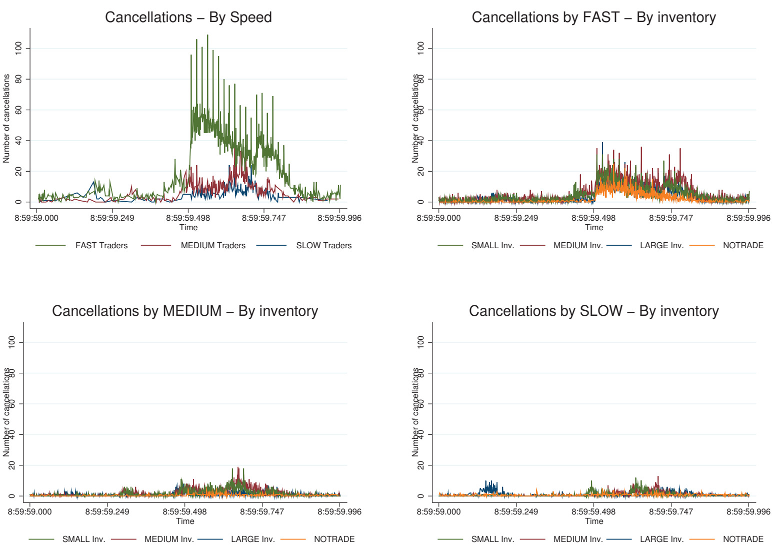
Figure 4.4. Order flow during the last second of the pre-opening period

This figure shows four types of order submission activity: new orders (blue line), cancellations (red line), quantity revisions (green line), and price revisions (black line), in the last second of the pre-opening period (8:59:59.000 - 8:59:59.999) at the millisecond level, for all 97 stocks from the TOPIX100 during the sample period of April and May 2013. Panel A reports traders' order submission activity. Panel B focuses on cancellations, distinguishing between the different groups of traders. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: Order flow for all traders



Panel B: Cancellations by speed and inventory group



4.5.2.2 Deviation of mid-quotes from the opening price

One of the questions we aim to answer with this paper concerns price discovery. We showed in the previous section that the number of order submissions rises right before the opening time. To explore how the order submissions by different traders contribute to price discovery, we look into the movements in the pre-opening-period quotes between 8 and 9 am to determine how quickly pre-opening quotes approach the opening price for the day. For this purpose, we compute the absolute value of the relative deviation of the mid-quotes from the opening price for each stock, on each day:

$$Deviation_{j,k,t} = \left| \frac{M_{j,k,t}}{O_{j,k}} - 1 \right| \times 100 \tag{4.1}$$

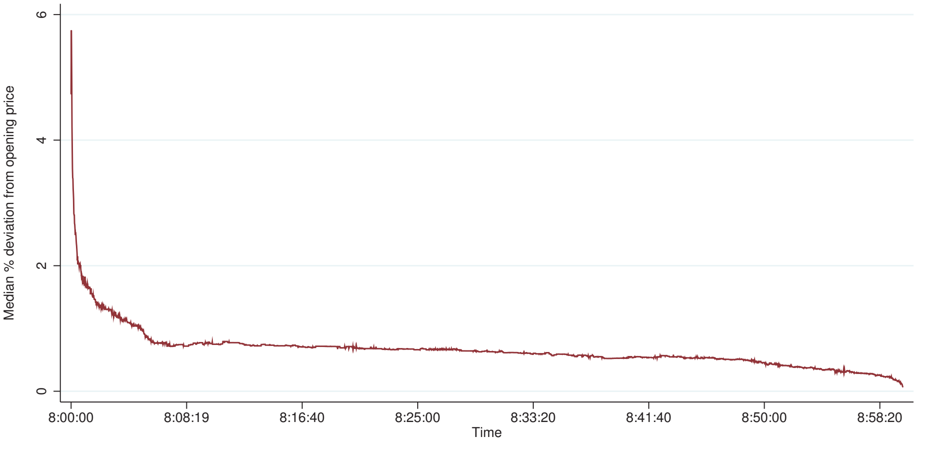
where $M_{j,k,t}$ is the mid-quote at time t for stock j on day k, and $O_{j,k}$ is the opening price for stock j on day k. First, we compute equation (4.1) second-by-second per stock per day. Then, we calculate the second-by-second medians.

Figure 4.5 shows the median of the second-by-second movements of the pre-opening quotes across the 97 stocks. During the first five minutes, the deviation declines rapidly from above 2% to between 0.6% and 0.7%. This means that significant amounts of order submissions contribute to price discovery during this period. However, after 8:05 am, the deviation becomes almost flat, with some spikes, and it then resumes its adjustment toward the opening price after 8:59 am. The deviation diminishes to 0.22% just before the opening time, which is still a little bit wider than a half-spread, on average, for the sample stocks during the trading session. This shows that lower latency does not attenuate the reduction of the deviation between the pre-opening quotes and the opening price. Hence, the orders submitted after 8:50

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.

Figure 4.5. 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 entire pre-opening period (8:00:00.000 - 8:59:59.999) for 97 stocks from the TOPIX100 during the sample period of April-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 (4.1). 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.



am play an important role in price discovery.

4.5.3 Price discovery contribution

During the pre-opening period, the accumulation of orders in general contributes to the reduction in the absolute deviation of the pre-opening quotes from the official opening price. However, the speed of convergence varies across stocks and throughout the day. We investigate which trader groups contribute to the price discovery process, and compare the extent of their contribution using order-by-order data and associated mid-quote changes. In this manner, we take advantage of our detailed data as we can pinpoint an order that moves the mid-quote and, thus, we can identify which trader group submits the order and the type of that order.

4.5.3.1 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 session and are briefly explained in Section 4.3.1. 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.[20]

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 4.6 Panel A shows the total number of orders from the 12 trader groups defined earlier. The largest proportion of aggressive orders comes from FAST / SMALL traders (HFT market makers). On average, they submit 248.4 aggressive orders (76.1 market orders and 172.3 limit orders). The next largest group of aggressive traders are the MEDIUM / SMALL traders who submit 174.7 aggressive orders (53.0 market orders and 121.8 limit orders). Note that our classification does not take into account trading share such as top quartile of volume, and only one quarter of FAST / SMALL traders participate in the pre-opening period, but their submission of aggressive orders is significantly greater than that of the other groups. The ratios of aggressive limit orders relative to the total number of limit orders from these two most aggressive groups of traders are 14.1% and 14.6%, respectively. Their aggressiveness ratios for limit orders are low in comparison to those of the other ten groups. The highest aggressiveness ratio is exhibited by FAST / NOTRADE traders, being 36.4%. This is an interesting contrast because FAST / NOTRADE traders place orders most aggressively, but their orders are not executed. However, the FAST / SMALL and MEDIUM / SMALL traders submit the largest portion of aggressive

[20] 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.

limit orders.

Table 4.6 Panel B shows similar statistics after the exclusion of the first 10 minutes of the pre-opening period, because, in the first 10 minutes, most of the orders entered are those waiting for the exchange's opening at 8 am. After 10 minutes past 8 am, most of the orders are submitted by traders who actively monitor the pre-opening quotes. In the remaining 50 minutes, the largest proportion of aggressive orders still comes from FAST / SMALL traders (HFT market makers), who submit 136.7 aggressive orders (46.2 market orders and 90.5 limit orders). The next-most-aggressive group of traders are the MEDIUM / SMALL traders, who submit 74.7 aggressive orders (30.1 market orders and 44.6 limit orders). The ratios of aggressive limit orders to total limit orders for the two most aggressive groups of traders rise to 31.0% and 26.3%, respectively. The highest aggressiveness ratio in this period is that of the FAST / LARGE traders, at 44.8%. This ratio indicates the trader's willingness to execute the order at the opening price. On the other hand, the FAST / SMALL group places the most aggressive number of limit orders in terms of the total number of aggressive orders, which indicates their interest in affecting the price. None of the NOTRADE traders in any of the three speed groups change their order aggressiveness during these 50 minutes. Put differently, they do not adjust their orders according to the changes in the prevailing quotes. This may be one of the reasons why their orders are not executed.

4.5.3.2 Price discovery contribution by order

We measure the amount of new information incorporated into stock prices during the pre-opening period using the weighted price contribution, $\boldsymbol{WPC}$ (e.g., Barclay and Warner (1993), Cao, Ghysels, and Hatheway (2000), and Barclay and Hendershott (2003)). First, we define the price discovery contribution as the amount by which an incoming order moves the prevailing mid-quote closer to the opening price. Thus, we compute the price discovery contribution ($\boldsymbol{PDC}$) on an order-by-order basis as

Table 4.6. Aggressive orders during pre-opening period

This table reports the summary statistics for order aggressiveness during the pre-opening period for the 12 trader groups. We split all traders into 12 groups on a stock-day basis, as described in Table 4.3, using information about speed and inventory from the same day's continuous session. 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. The total number of aggressive orders is the average number of aggressive orders made by the trader group across stock-days. The total number of market orders is the average number of aggressive market orders made by the trader group across stock-days. The total number of aggressive limit orders is the average number of aggressive limit orders made by the trader group across stock-days. The ratio of total order aggressiveness is the number of aggressive orders over the total number of orders. The ratio of limit order aggressiveness is the number of aggressive limit orders over the total number of orders. Panel A describes the order aggressiveness of each trader group during the entire pre-opening period, while Panel B describes that excluding the first 10 minutes of the pre-opening period for 97 stocks from the TOPIX100 during the sample period of April-May 2013. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: 8:00-8:59

Speed	Inventory	Total # of aggressive orders	Total # of market orders	Total # of aggressive limit orders	Ratio of total order aggressiveness	Ratio of limit order aggressiveness
FAST	LARGE	29.3	8.5	20.8	38.2%	30.5%
	MEDIUM	155.2	55.8	99.4	24.6%	17.3%
	SMALL	248.4	76.1	172.3	19.2%	14.1%
	NOTRADE	7.3	3.1	4.3	50.0%	36.4%
MEDIUM	LARGE	27.2	8.9	18.3	26.0%	19.2%
	MEDIUM	140.0	39.0	101.0	20.4%	15.6%
	SMALL	174.7	53.0	121.8	19.7%	14.6%
	NOTRADE	3.8	2.0	1.8	42.6%	25.4%
SLOW	LARGE	34.6	12.0	22.6	37.8%	28.5%
	MEDIUM	27.8	7.7	20.1	30.1%	23.7%
	SMALL	19.9	5.6	14.3	30.3%	23.7%
	NOTRADE	10.1	2.4	7.7	31.0%	25.6%

Table 4.6. Aggressive orders during pre-opening period (continued)

Panel B: 8:10-8:59

Speed	Inventory	Total # of aggressive orders	Total # of market orders	Total # of aggressive limit orders	Ratio of total order aggressiveness	Ratio of limit order aggressiveness
FAST	LARGE	24.2	7.5	16.8	54.1%	44.9%
	MEDIUM	94.3	40.7	53.6	48.6%	34.9%
	SMALL	136.7	46.2	90.5	40.4%	31.0%
	NOTRADE	6.3	3.0	3.3	52.5%	36.7%
MEDIUM	LARGE	16.7	7.1	9.6	43.0%	30.3%
	MEDIUM	55.6	23.2	32.4	38.5%	26.8%
	SMALL	74.7	30.1	44.6	37.5%	26.4%
	NOTRADE	3.4	2.0	1.5	45.5%	25.9%
SLOW	LARGE	18.4	9.1	9.4	56.4%	39.5%
	MEDIUM	10.5	4.9	5.6	49.1%	34.1%
	SMALL	8.2	3.7	4.5	49.7%	35.2%
	NOTRADE	5.6	2.2	3.4	36.4%	25.8%

follows:

$$PDC_{i,j,k} = Deviation_{i,j,k} - Deviation_{i-1,j,k} \tag{4.2}$$

$Deviation_{i,j,k}$ is the absolute deviation of the mid-quote from the opening price immediately after order i is entered for stock j on day k (see equation (4.1)). $Deviation_{i-1,j,k}$ is the absolute deviation of the mid-quote from the opening price immediately before order i is entered for stock j on day k. The difference between $Deviation_{i,j,k}$ and $Deviation_{i-1,j,k}$ is the contribution to price discovery made by order i. When $PDC_{i,j,k}$ is negative, the deviation is reduced and the mid-quote moves closer to the opening price. We define the WPC for stock j on day k and order i as

$$WPC_{i,j,k} = \frac{PDC_{j,k}}{\sum_{j=1}^{J} |PDC_{j,k}|} \times \frac{PDC_{i,j,k}}{PDC_{j,k}} \tag{4.3}$$

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 during the pre-opening period 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 the pre-opening 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. 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)). Table 4.7 shows the WPC for each trading group. It turns out that MEDIUM / SMALL traders make the largest contribution (WPC =-20.57%).

This means that, on average, 20.57% of the daily price discovery is contributed by this group. They are followed by MEDIUM / MEDIUM (-18.79%) and FAST / SMALL (-16.37%) traders (see Table 4.7 Panel A). Furthermore, if we distinguish between new limit orders and new market orders, the contribution of the latter is much smaller than that of the former.

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, we focus on the remaining 50 minutes, during which traders monitor pre-opening quotes and make order submission decisions accordingly. In this period (see Table 4.7 Panel B), the main contribution comes from the FAST / MEDIUM (-5.51%) traders, followed by the FAST / SMALL (-3.32%) and MEDIUM / MEDIUM (-2.96%) traders. This reflects the more intensive activity of FAST traders after the first 10 minutes, especially in the last 10 minutes of the pre-opening period.

Which types of orders contribute most to price discovery? According to Table 4.7 Panel A, the types of orders contributing most to WPC are new limit orders. Cancellations of market orders and price revisions of limit orders also contribute. Quantity revisions and cancellations of limit orders increase the mid-quote deviation from the opening price. Price discovery in the pre-opening period is achieved mainly through new limit orders and price revisions of limit orders, and the results indicate that the effects of cancellations are limited. Our overall results indicate that quote setting during the pre-opening period is conducted by the FAST / SMALL & MEDIUM and MEDIUM / SMALL & MEDIUM groups. Therefore, traders with low latency 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 low-latency traders participate in the pre-opening period.

4.5.3.3 Cross-sectional analysis

In this section, we aim to answer the question of whether stocks with a greater presence of one trader group relative to another trader group tend to exhibit different

Table 4.7. Contribution to weighted price discovery by type of order

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 (4.3). 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 4.3, using information about speed and inventory from the same day's continuous session. Panel A describes WPC during the pre-opening period, while Panel B describes WPC excluding the first 10 minutes of the pre-opening period for 97 stocks from the TOPIX100 during the sample period of April-May 2013. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: 8:00-8:59

		Total	Limit orders				Market orders				Zero imbalance
Speed	Inventory		New	Qty revision	Cancellation	Price revision	New	Qty revision	Cancellation	Price revision	
FAST	LARGE	-1.90%	-1.87%	0.09%	0.45%	-0.22%	-0.43%	0.04%	0.05%	0.00%	0.00%
	MEDIUM	-14.27%	-11.29%	0.09%	0.00%	-0.19%	-2.53%	-0.09%	-0.31%	0.06%	0.00%
	SMALL	-16.37%	-14.90%	0.08%	0.53%	-0.23%	-1.68%	-0.02%	-0.14%	0.01%	-0.02%
	NOTRADE	-0.25%	-0.30%	0.03%	0.06%	-0.01%	-0.02%	0.00%	0.00%	0.00%	-0.01%
MEDIUM	LARGE	-2.78%	-3.54%	0.01%	0.03%	-0.05%	0.85%	0.00%	-0.06%	-0.02%	0.00%
	MEDIUM	-18.79%	-17.50%	0.00%	0.11%	0.01%	-1.30%	-0.03%	-0.08%	0.00%	0.00%
	SMALL	-20.57%	-18.70%	0.01%	0.09%	-0.04%	-1.76%	0.00%	-0.11%	-0.04%	-0.01%
	NOTRADE	-0.23%	-0.26%	0.01%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SLOW	LARGE	-9.23%	-8.77%	0.01%	0.03%	0.02%	-0.46%	-0.01%	-0.03%	-0.01%	-0.02%
	MEDIUM	-6.86%	-6.11%	0.00%	-0.01%	-0.05%	-0.63%	0.00%	-0.04%	-0.01%	-0.01%
	SMALL	-4.34%	-4.16%	0.00%	0.06%	0.00%	-0.13%	0.00%	-0.11%	0.00%	0.00%
	NOTRADE	-4.43%	-4.46%	0.04%	0.04%	-0.01%	-0.03%	0.00%	-0.02%	0.00%	0.00%

Table 4.7. Weighted price discovery contribution (continued)

Panel B: 8:10-8:59

		Total	Limit orders				Market orders				Zero imbalance
Speed	Inventory		New	Qty revision	Cancellation	Price revision	New	Qty revision	Cancellation	Price revision	
FAST	LARGE	-1.42%	-1.24%	0.09%	0.42%	-0.23%	-0.56%	0.04%	0.05%	0.01%	0.00%
	MEDIUM	-5.51%	-2.32%	0.09%	-0.31%	-0.19%	-2.45%	-0.07%	-0.29%	0.03%	0.00%
	SMALL	-3.32%	-2.00%	0.07%	0.17%	-0.21%	-1.21%	-0.02%	-0.11%	-0.01%	0.00%
	NOTRADE	-0.09%	-0.16%	0.03%	0.06%	-0.01%	-0.01%	0.00%	0.00%	0.00%	0.00%
MEDIUM	LARGE	-1.21%	-0.66%	0.01%	0.00%	-0.05%	-0.43%	-0.01%	-0.07%	-0.01%	0.00%
	MEDIUM	-2.96%	-1.47%	0.00%	0.04%	0.01%	-1.42%	-0.03%	-0.09%	-0.01%	0.00%
	SMALL	-2.71%	-1.73%	0.01%	0.02%	-0.04%	-0.84%	0.00%	-0.09%	-0.04%	0.00%
	NOTRADE	-0.09%	-0.13%	0.01%	0.03%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
SLOW	LARGE	-1.60%	-0.95%	0.01%	0.02%	0.01%	-0.65%	-0.01%	-0.03%	0.00%	0.00%
	MEDIUM	-1.07%	-0.45%	0.00%	-0.02%	-0.03%	-0.54%	0.00%	-0.01%	-0.01%	0.00%
	SMALL	-0.41%	-0.33%	0.00%	0.04%	0.01%	-0.02%	0.00%	-0.11%	0.00%	0.00%
	NOTRADE	-0.45%	-0.51%	0.04%	0.05%	-0.02%	0.01%	0.00%	-0.01%	0.00%	0.00%

patterns of mid-quote convergence to the opening price. We conduct this analysis in two steps. First, we investigate whether we observe 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 messages (messages that could potentially have an impact on the mid-quote) from each trader group relative to the number of aggressive messages from all trader groups during the whole pre-opening period, and during the pre-opening period excluding the first 10 minutes, aggregated across stocks and days (see Table 4.8). FAST / SMALL and FAST / MEDIUM traders, as well as MEDIUM / SMALL and MEDIUM / MEDIUM traders, exhibit wide variation in their activity from stock to stock, for both the whole pre-opening period and for the pre-opening period excluding the first 10 minutes. This pattern is especially strong for FAST / SMALL traders (high-frequency market makers): their relative activity varies from 4.54% to 60.05% (5.80% to 58.65%) for the whole pre-opening period (for the pre-opening period excluding the first 10 minutes).

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 any of the four groups of traders (FAST / SMALL, FAST / MEDIUM, MEDIUM / SMALL, or MEDIUM / MEDIUM) during the whole pre-opening period crosses a threshold of 30% (18 stocks), and all other stocks (79 stocks). Figure 4.6 presents the median absolute deviation of the mid-quote from the opening price per second of the pre-opening period, and separately for the first and 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 gap between the two series is largest at the beginning of the pre-opening period (with a maximum of 1.08%). During the last 10 minutes of the pre-opening period, the gap size varies 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. All in

Table 4.8. 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 4.3, using information about speed and inventory from the same day's continuous session. 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 entire pre-opening period (Panel A) and for the pre-opening period, excluding the first 10 minutes (Panel B) 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
Panel A: 8:00-8:59								
FAST	LARGE	1.03%	1.45%	2.26%	3.40%	4.61%	8.15%	10.29%
	MEDIUM	8.90%	11.91%	14.38%	16.64%	19.40%	22.10%	34.52%
	SMALL	4.54%	6.53%	12.24%	16.67%	27.99%	44.10%	60.05%
	NOTRADE	0.01%	0.03%	0.12%	0.27%	0.51%	1.48%	3.59%
MEDIUM	LARGE	1.08%	1.77%	2.49%	3.67%	4.55%	6.59%	7.70%
	MEDIUM	4.62%	10.02%	15.92%	18.84%	21.12%	25.31%	27.51%
	SMALL	8.18%	11.40%	17.30%	20.37%	24.55%	28.10%	31.21%
	NOTRADE	0.00%	0.03%	0.08%	0.17%	0.30%	0.73%	2.11%
SLOW	LARGE	0.63%	1.36%	2.75%	4.87%	8.02%	15.00%	20.47%
	MEDIUM	0.28%	0.83%	2.03%	4.66%	7.07%	9.79%	11.40%
	SMALL	0.26%	0.48%	1.35%	3.17%	4.91%	6.62%	8.20%
	NOTRADE	0.05%	0.18%	0.48%	1.13%	1.81%	4.56%	7.87%
Panel B: 8:10-8:59								
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	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%
MEDIUM	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%

all, to sum up, the presence of the FAST / SMALL, FAST / MEDIUM, MEDIUM / SMALL, and MEDIUM / MEDIUM traders improves the price discovery process.

Next, we examine whether the same stocks attract the activity of each of the four trader groups. Table 4.9 shows the correlation coefficients between the relative activity levels of different trader groups during the whole pre-opening period (Panel A) and the pre-opening period excluding the first 10 minutes (Panel B). In particular, Panel A of Table 4.9 shows that the relative activity levels of the FAST / SMALL and FAST / MEDIUM groups are positively correlated (correlation coefficient 22%), as are the relative activity levels of the MEDIUM / SMALL and MEDIUM / MEDIUM groups (correlation coefficient 45%). However, across the speed groups, only FAST / SMALL and MEDIUM / SMALL are positively correlated, with the other trader groups exhibiting strong negative correlation, reaching -66% between the FAST / SMALL and MEDIUM / MEDIUM trader groups. Results for the pre-opening period excluding the first 10 minutes are qualitatively similar, with one exception of FAST / SMALL and FAST / MEDIUM activity being negatively correlated. All in all, different stocks attract the activity of the FAST / SMALL & MEDIUM and MEDIUM / SMALL & MEDIUM traders, who are the main contributors to the price discovery process, as based on the WPC analysis (see Section 4.5.3.2).

In order to examine which stocks attract more of the activity of the four above-mentioned groups of traders, we run a cross-sectional regression using the relative activity of the trader groups as the dependent variable and stock characteristics as explanatory variables:

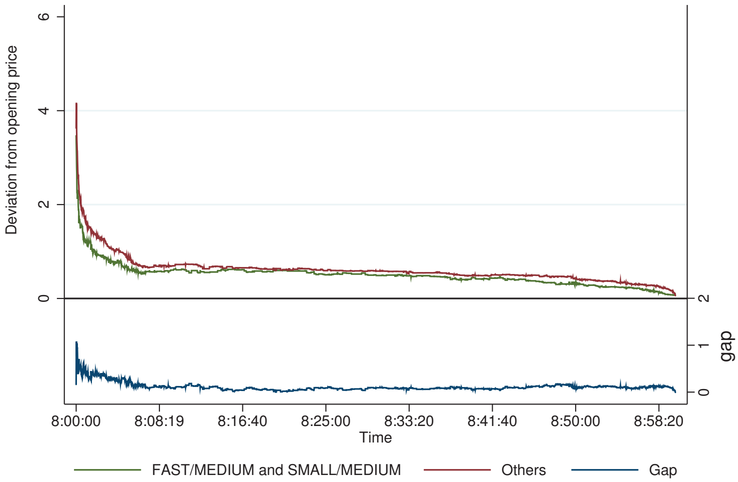
$$\begin{aligned} Activity_{j,l} = \alpha + \beta_1 Deviation_j + \beta_2 MCAP_j + \beta_3 PQSPR_j + \\ \beta_4 Range_j + \beta_5 Industry_j + \beta_6 ADR_j + \epsilon_j \end{aligned} \tag{4.4}$$

where $Activity_{j,l}$ is the ratio of the aggressive orders of trader group l for stock j to the total number of aggressive orders for stock j; $Deviation_j$ is the median of the absolute deviation of the mid-quote from the opening price during the first second of the pre-opening period (or of the first second of the pre-opening period excluding

Figure 4.6. Comparison of the deviation from the opening price between stocks for which low-latency 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 at each second of the entire pre-opening period (8:00:00.000 - 8:59:59.999) for 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&MEDIUM / 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 first 10 minutes and 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 first and last 10 minutes for the two groups of traders

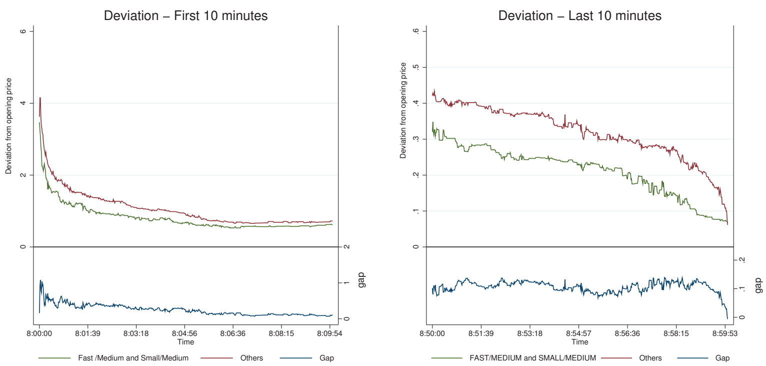


Table 4.9. Correlation of order aggressiveness across stocks for different group of traders

This table presents the correlation analysis for the aggressive orders across stocks from different trader groups. We divide all traders into 12 groups on a stock-day basis, as described in Table 4.3, using information about speed and inventory from the same day's continuous session. For each stock, we compute the correlation between 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 entire pre-opening period (Panel A) and for the pre-opening period, excluding the first 10 minutes (Panel B) 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.

Panel A: 8:00-8:59

		FAST				MEDIUM				SLOW			
		LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE
FAST	LARGE	1.00											
	MEDIUM	0.06	1.00										
	SMALL	-0.45	0.22	1.00									
	NOTRADE	0.67	0.00	-0.20	1.00								
MEDIUM	LARGE	0.59	-0.30	-0.80	0.32	1.00							
	MEDIUM	-0.17	-0.20	-0.66	-0.24	0.25	1.00						
	SMALL	-0.63	-0.14	0.14	-0.47	-0.52	0.45	1.00					
	NOTRADE	0.61	-0.29	-0.53	0.52	0.77	-0.09	-0.64	1.00				
SLOW	LARGE	0.62	-0.38	-0.72	0.39	0.86	0.06	-0.62	0.87	1.00			
	MEDIUM	0.45	-0.43	-0.81	0.23	0.83	0.31	-0.46	0.64	0.87	1.00		
	SMALL	0.44	-0.36	-0.79	0.21	0.80	0.28	-0.47	0.59	0.84	0.96	1.00	
	NOTRADE	0.61	-0.37	-0.63	0.43	0.80	0.00	-0.60	0.93	0.94	0.73	0.66	1.00

Panel B: 8:10-8:59

		FAST				MEDIUM				SLOW			
		LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE
FAST	LARGE	1.00											
	MEDIUM	0.16	1.00										
	SMALL	-0.71	-0.23	1.00									
	NOTRADE	0.68	-0.07	-0.45	1.00								
MEDIUM	LARGE	0.82	0.07	-0.80	0.57	1.00							
	MEDIUM	-0.13	0.00	-0.48	-0.22	0.05	1.00						
	SMALL	-0.73	-0.30	0.30	-0.57	-0.63	0.48	1.00					
	NOTRADE	0.74	-0.07	-0.60	0.71	0.79	-0.15	-0.65	1.00				
SLOW	LARGE	0.81	-0.09	-0.75	0.64	0.87	-0.04	-0.64	0.83	1.00			
	MEDIUM	0.76	-0.06	-0.79	0.55	0.84	0.10	-0.59	0.72	0.89	1.00		
	SMALL	0.77	0.06	-0.78	0.53	0.82	0.05	-0.63	0.68	0.87	0.93	1.00	
	NOTRADE	0.80	-0.03	-0.69	0.70	0.84	-0.12	-0.65	0.91	0.89	0.80	0.76	1.00

the first 10 minutes) (see equation (4.1)); $MCAP_j$ is the log of the average daily market capitalization of stock j; $PQSPR_j$ is the average of the daily proportional quoted spread of stock j; $Range_j$ is the square root of the daily average high minus low range for stock j; $Industry_j$ is a dummy variable equalling 1 if the stock is in the Machinery and Business Equipment industry and 0 otherwise; ADR_j is a dummy variable that equals 1 if the stock has an American Depositary Receipt (ADR) and 0 otherwise. $MCAP$, $PQSPR$, and $Range$ are measured over March 2013, which is before the start of the period for which data are provided by the TSE. Data on stock characteristics come from Datastream. All the variables are winsorized at the 1% and 99% levels. Table 4.10 presents the estimates of the cross-sectional regression for the whole pre-opening period (Panel A) and for the pre-opening period, excluding the first 10 minutes (Panel B). We consider only those effects that are robust to the exclusion of the first 10 minutes of the pre-opening period.

Specifically, Table 4.10 shows that large stocks are more attractive for FAST / SMALL & MEDIUM and MEDIUM & SMALL traders, while the relative activity of SLOW traders is more pronounced in small stocks. Liquid stocks attract more activity from MEDIUM / SMALL & MEDIUM traders. FAST and MEDIUM-speed traders with SMALL inventories are more active in high-volatility stocks, while other trader groups prefer low-volatility stocks. The smaller the size of the absolute deviation of the first mid-quote from the opening price, the greater is the activity of FAST / SMALL traders. On the contrary, FAST / MEDIUM traders prefer stocks with larger absolute deviation. The activity of the FAST / SMALL traders is also greater if the stock has an ADR.

To sum up, FAST&MEDIUM / SMALL&MEDIUM traders have preferences for a certain type of stocks.

4.5.3.4 Panel Analysis

We extend our analysis of price discovery during the pre-opening period using a panel dataset at 100-millisecond intervals for the 97 stocks of the TOPIX100 index. We focus our analysis on the relation between a trader's aggregated aggressiveness and the change in the absolute deviation of the mid-quote from the opening price

Table 4.10. Cross-sectional regression for the traders stock preferences

This table shows the estimation results of the cross-sectional regression of aggressive activity of different trader groups as defined in equation (4.4). We divide all traders into 12 groups on a stock-day basis, as described in Table 4.3, using information about speed and inventory from the same day's continuous session. As the dependent variable we use a 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 entire pre-opening period (Panel A) and for the pre-opening period, excluding the first 10 minutes (Panel B) for April and May 2013 across 97 stocks from TOPIX100. As explanatory variables we use stock characteristics as median of absolute deviation of the mid-quote from the opening price (***Deviation***) during the first second of the pre-opening period (of the first second of the pre-opening period, excluding the first 10 minutes), log of market capitalization, proportional quoted spread (***PQSPR***), the square root of the daily average high minus low range (***Range***); a dummy variable which equals 1 if the stock is in Machinery and Business Equipment industry and 0 otherwise (***Industry***); a dummy variable which equals 1 if the stock has an ADR and 0 otherwise (***ADR***). ***, **, * indicate significance at 1%, 5%, and 10% levels. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange. Data on stock characteristics are obtained from Datastream.

Panel A: 8:00-8:59

	FAST				MEDIUM				SLOW			
	LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE
Deviation	0.059	0.226	-1.073*	-0.041	0.092	0.013	0.062	-0.002	0.286	0.201	0.167*	0.066
	(0.52)	(1.24)	(-1.90)	(-1.52)	(1.08)	(0.05)	(0.19)	(-0.11)	(1.34)	(1.28)	(1.71)	(0.86)
MCAP	-1.178***	2.101***	8.709***	-0.197***	-1.077***	-1.795**	1.450*	-0.132***	-3.003***	-2.545***	-1.799***	-0.805***
	(-5.51)	(4.62)	(5.18)	(-3.40)	(-6.40)	(-2.44)	(1.76)	(-5.24)	(-6.27)	(-7.60)	(-7.31)	(-4.89)
PQSPR	0.066***	0.123***	0.090	0.025***	0.013	-0.172***	-0.224***	0.003	0.036	0.003	0.031	0.004
	(2.79)	(3.23)	(0.62)	(3.43)	(0.87)	(-2.66)	(-3.16)	(0.86)	(0.74)	(0.09)	(1.10)	(0.25)
RANGE	-0.448	-0.244	6.376**	-0.196**	-0.484	-1.352	2.900***	-0.098**	-2.622***	-1.896***	-1.600***	-0.517*
	(-1.34)	(-0.30)	(2.26)	(-2.01)	(-1.53)	(-1.16)	(2.76)	(-2.15)	(-3.40)	(-3.05)	(-3.80)	(-1.97)
Industry	0.596	-0.719	-2.293	0.300	0.375	0.714	0.463	0.108	0.213	-0.079	-0.077	0.255
	(0.93)	(-1.03)	(-0.94)	(1.45)	(0.89)	(0.50)	(0.30)	(1.00)	(0.17)	(-0.15)	(-0.23)	(0.46)
ADR	-0.968**	-0.533	6.431***	-0.079	-0.896***	-0.687	1.682	-0.075	-1.946***	-1.165**	-0.915**	-0.545**
	(-2.60)	(-0.81)	(2.72)	(-0.68)	(-3.66)	(-0.65)	(1.61)	(-1.66)	(-2.86)	(-2.30)	(-2.55)	(-2.17)
Constant	19.614***	-16.458**	-112.446***	3.055***	19.391***	50.233***	0.154	2.220***	51.046***	43.251***	30.173***	13.489***
	(6.37)	(-2.30)	(-4.87)	(3.60)	(7.95)	(4.88)	(0.01)	(5.54)	(7.05)	(8.61)	(8.37)	(5.33)
Obs.	97	97	97	97	97	97	97	97	97	97	97	97
Adjusted R-squared	0.192	0.224	0.419	0.156	0.373	0.157	0.114	0.124	0.310	0.422	0.434	0.161

Table 4.10. Cross-sectional regression for the traders stock preferences (continued)

Panel B: 8:10-8:59

	FAST				MEDIUM				SLOW			
	LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE	LARGE	MEDIUM	SMALL	NOTRADE
Deviation	1.909**	3.956**	-6.086**	0.150	1.097**	-0.531	-2.984	0.356**	0.457	0.254	0.585	0.842**
	(2.16)	(2.43)	(-2.14)	(0.74)	(2.08)	(-0.40)	(-1.52)	(2.23)	(0.47)	(0.46)	(1.42)	(2.30)
MCAP	-2.452***	1.061*	7.714***	-0.379***	-1.230***	-0.603	2.239***	-0.218***	-2.724***	-1.640***	-1.227***	-0.640***
	(-7.01)	(1.83)	(5.28)	(-4.58)	(-6.76)	(-1.08)	(2.88)	(-5.48)	(-7.34)	(-6.98)	(-7.45)	(-5.45)
PQSPR	0.050	0.016	0.154	0.029***	0.006	-0.127***	-0.176***	0.001	0.025	0.000	0.018	0.003
	(1.54)	(0.39)	(1.20)	(3.93)	(0.37)	(-2.87)	(-2.69)	(0.34)	(0.68)	(0.00)	(0.97)	(0.31)
RANGE	-1.814***	-1.258	6.439**	-0.392**	-1.041***	-0.129	4.181***	-0.276***	-2.525***	-1.243***	-1.127***	-0.739***
	(-3.13)	(-1.18)	(2.22)	(-2.30)	(-2.74)	(-0.14)	(3.91)	(-3.34)	(-3.58)	(-2.83)	(-3.61)	(-3.45)
Industry	0.746	-0.476	-3.498	0.262	0.400	0.663	1.210	0.106	0.164	0.093	0.152	0.071
	(1.09)	(-0.59)	(-1.65)	(1.18)	(0.99)	(0.72)	(1.05)	(0.83)	(0.21)	(0.27)	(0.67)	(0.25)
ADR	-1.297**	-1.442*	4.289*	-0.173	-0.781***	0.359	1.950*	-0.038	-1.258**	-0.645**	-0.582**	-0.316*
	(-2.40)	(-1.81)	(1.96)	(-1.24)	(-2.80)	(0.45)	(1.98)	(-0.56)	(-2.40)	(-2.10)	(-2.53)	(-1.89)
Constant	42.028***	5.386	-97.256***	5.901***	22.893***	26.132***	-15.957	3.647***	47.872***	28.620***	21.035***	10.934***
	(8.32)	(0.62)	(-4.70)	(4.50)	(8.75)	(3.23)	(-1.44)	(5.83)	(8.77)	(8.39)	(9.09)	(6.03)
Obs.	97	97	97	97	97	97	97	97	97	97	97	97
Adjusted R-squared	0.386	0.101	0.390	0.180	0.388	0.063	0.215	0.204	0.369	0.416	0.446	0.275

every 100 milliseconds. To compute the change in the absolute deviation when there are several mid-quote updates in a particular 100-millisecond interval, we take the last value of the mid-quote during that interval. Afterwards, we examine how the aggregated aggressive orders of each group of traders affect the convergence of the mid-quote to the opening price. We winsorize the change in the absolute deviation at the 0.1% and 99.9% levels.

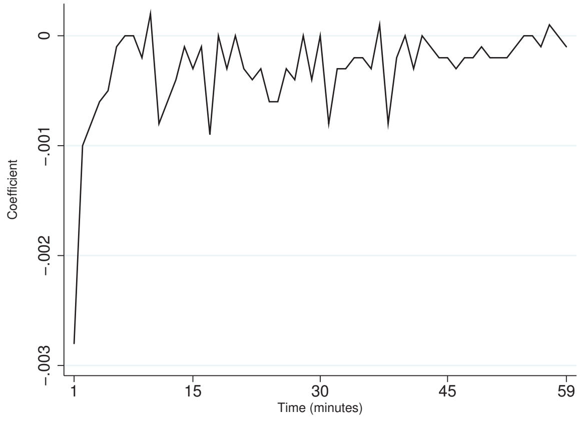
In particular, for each group of traders, we aggregate the number of new orders, cancelled orders, and revised orders, separately for limit and market orders, for each 100-millisecond interval, and scale it by the total number of orders for each stock-day. We also use the number of shares in each order as the dependent variable. When we aggregate orders, they must satisfy the conditions for aggressive orders defined in Section 4.5.3.1. We do not distinguish between buy and sell orders because our dependent variable does not represent the direction of the price movement. Both buy and sell orders can equally narrow or widen the deviation. We do not take into account orders categorized as non-aggressive orders, because these orders do not affect the prevailing quotes and are not visible to market participants. Therefore, traders cannot speculate on other traders' behavior based on non-aggressive order flow. We employ a stock and time (minute) fixed effects panel regression to conduct the abovementioned analysis:

$$\begin{aligned}
Change\ in\ Deviation_{j,k,t} = \alpha + &\\
&\sum_{l=1}^{12}(\beta_{1,l} New\ Limit_{j,k,t,l} + \beta_{2,l} New\ Market_{j,k,t,l} +\\
&\beta_{3,l} Cancel\ Limit_{j,k,t,l} + \beta_{4,l} Cancel\ Market_{j,k,t,l} +\\
&\beta_{5,l} Qty\ Rev\ Limit_{j,k,t,l} + \beta_{6,l} Qty\ Rev\ Market_{j,k,t,l} +\\
&\beta_{7,l} Price\ Rev\ Limit_{j,k,t,l} + \beta_{8,l} Price\ Rev\ Market_{j,k,t,l} +\\
&\beta_{9,l} Zero\ Imbalance_{j,k,t,l}) + \epsilon_{j,k,t}
\end{aligned} \tag{4.5}$$

where $Change\ in\ Deviation_{j,k,t}$ is the change in the deviation of the mid-quote from the opening price for stock j on date k, t is the 100-millisecond interval, and l refers

Figure 4.7. Intra-day patterns of pre-opening mid-quotes

This figure shows the coefficients of time series dummies from the panel regression of the change in the deviation of the mid-quote from the opening price, per stock, on the trading activity of the 12 trader groups, for the 97 stocks from the TOPIX100 during the sample period of April-May 2013, as defined in equation (4.5). Please refer to Table 4.11 for more details. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.



to a particular group of traders. *Price Rev Market* means the change of the order from market to limit or vice versa.

We run panel regressions with stock fixed effects because the 97 stocks in our sample differ by minimum tick size and price level, both of which have significant effects on the minimum percentage change in the dependent variable. Time fixed effects take into account the intra-day pattern in the pre-opening quotes (see Figure 4.7).

We run these regressions for four different time periods: the entire period (8:00-8:59), the period excluding the first 10 minutes (8:10-8:59), the last 10 minutes (8:50-8:59), and the last minute (8:59:00-8:59:99). We report only the results for the entire period (8:00-8:59) and the period excluding the first 10 minutes (8:10-8:59). Table 4.11 presents the results of the panel regressions. We discuss each time period separately below.

Table 4.11 Panel A shows the results for the entire pre-opening period. During the pre-opening period, we observe statistically significant negative coefficients for

Table 4.11. Panel regression for the determinants of the absolute deviation of the mid-quote from the opening price

This table shows the estimation results of the panel regressions of the change in the deviation of the mid-quote from the opening price, per stock-day, on the trading activity of the 12 trader groups, for the 97 stocks from the TOPIX100 during the sample period, April-May 2013, as defined in equation (4.5). We report coefficients and corresponding standard error with significance levels denoted by ***, **, and * for 1%, 5%, and 10%, respectively. The activity of the different trader groups for each 100-millisecond-stock-day is measured as the number of a certain type of messages from each trader group during a particular 100-millisecond interval relative to the total number of messages from all categories on a particular stock-day. We include in the sample only those 100-millisecond intervals for which we observe a change in the absolute deviation. All regressions include stock fixed effects and time fixed effects per minute. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Panel A: Orders from 8:00 to 8:59

		Limit Orders				Market Orders				Zero Imbalance
Speed	Inventory	New	Qty. Rev	Cancellation	Price Rev	New	Qty. Rev	Cancellation	Price Rev	
FAST	LARGE	-0.0080***	0.0017	-0.0001	-0.00887***	-0.00677***	0.00055	-0.00231	-0.00022	0.00007
		0.001	0.0018	0.0008	0.00165	0.00068	0.00186	0.00176	0.00210	0.00025
	MEDIUM	-0.0118***	0.0018	0.0011	-0.00189**	-0.00587***	-0.00343*	-0.00504***	-0.00069	-0.00275***
		0.001	0.0011	0.0009	0.00076	0.00057	0.00178	0.00093	0.00188	0.00042
	SMALL	-0.0194***	0.0039**	-0.0005	-0.00485***	-0.00621***	-0.00846***	-0.00760***	-0.00355**	-0.00295***
		0.001	0.0019	0.0013	0.00063	0.00045	0.00239	0.00159	0.00148	0.00044
	NOTRADE	-0.0087***	0.0094**	0.0005	-0.00687***	-0.01414**	-0.24398***	0.00211		-0.00066
		0.003	0.0041	0.0022	0.00161	0.00591	0.00099	0.00479		0.00097
MEDIUM	LARGE	-0.0134***	0.0060	0.0001	-0.00503***	-0.00299***	-0.00694	-0.00810***	-0.00775**	-0.00143***
		0.001	0.0038	0.0014	0.00106	0.00070	0.00613	0.00223	0.00373	0.00032
	MEDIUM	-0.0186***	-0.0134***	-0.0013	-0.00262***	-0.00611***	-0.01000**	-0.00631***	-0.00376***	-0.00301***
		0.001	0.0040	0.0011	0.00048	0.00039	0.00395	0.00128	0.00127	0.00022
	SMALL	-0.0202***	-0.0004	-0.0028**	-0.00347***	-0.00619***	-0.01037*	-0.00761***	-0.00231*	-0.00359***
		0.001	0.0038	0.0013	0.00053	0.00041	0.00533	0.00116	0.00123	0.00026
	NOTRADE	-0.0159***	0.0127***	0.0071***	0.00388	-0.00868***		0.00061	0.00726***	0.00223***
		0.002	0.0048	0.0027	0.00295	0.00249		0.00405	0.00107	0.00070
SLOW	LARGE	-0.0129***	0.0003	0.0009	-0.00350***	-0.00499***	-0.00329	-0.00524***	-0.00315	-0.00276***
		0.001	0.0058	0.0019	0.00119	0.00044	0.00524	0.00175	0.00206	0.00024
	MEDIUM	-0.0189***	-0.0022	-0.0042***	-0.00487***	-0.00593***	-0.00077	-0.00716***	-0.00556**	-0.00305***
		0.001	0.0100	0.0013	0.00125	0.00060	0.00468	0.00154	0.00236	0.00030
	SMALL	-0.0187***	0.0153**	-0.0007	-0.00301***	-0.00530***	-0.00006	-0.01430***	-0.00180	-0.00305***
		0.001	0.0073	0.0017	0.00103	0.00064	0.01257	0.00227	0.00331	0.00034
	NOTRADE	-0.0265***	0.0080*	0.0067**	-0.00039	-0.00476***	-0.08032**	-0.00330*	-0.00033	-0.00270***
		0.003	0.0041	0.0029	0.00140	0.00174	0.03435	0.00201	0.00221	0.00027
	Constant	0.0003*		Observations	6090450	R-Squared	0.0421	Stock FE	YES	Std. Err. adjusted for 97 cluster
		(0.000)		N. of Groups	97			Time FE	YES	

Table 4.11. Panel regression for the determinants of the absolute deviation of the mid-quote from the opening price (continued)

Panel B: Orders from 8:10 to 8:59

		Limit Orders				Market Orders				Zero Imbalance
Speed	Inventory	New	Qty. Rev	Cancellation	Price Rev	New	Qty. Rev	Cancellation	Price Rev	
FAST	LARGE	-0.0090***	0.0019	0.0001	-0.00879***	-0.00634***	0.00107	-0.00221	0.00007	-0.00032
		0.001	0.0017	0.0008	0.00173	0.00068	0.00185	0.00172	0.00206	0.00033
	MEDIUM	-0.0087***	0.0020*	-0.0006	-0.00201***	-0.00562***	-0.00233	-0.00425***	-0.00109	-0.00177***
		0.001	0.0011	0.0007	0.00071	0.00060	0.00182	0.00094	0.00188	0.00057
	SMALL	-0.0144***	0.0033*	-0.0038***	-0.00483***	-0.00604***	-0.00803***	-0.00694***	-0.00411***	-0.00115*
		0.001	0.0017	0.0012	0.00062	0.00053	0.00251	0.00166	0.00148	0.00070
	NOTRADE	-0.0114***	0.0092**	0.0014	-0.00684***	-0.01269**	-0.24373***	0.00279		-0.00075
		0.003	0.0041	0.0020	0.00167	0.00582	0.00128	0.00480		0.00088
MEDIUM	LARGE	-0.0087***	0.0048	-0.0014	-0.00591***	-0.00535***	-0.00601	-0.00758***	-0.00517	-0.00061
		0.001	0.0034	0.0013	0.00099	0.00078	0.00609	0.00218	0.00352	0.00051
	MEDIUM	-0.0093***	-0.0137***	-0.0021*	-0.00230***	-0.00644***	-0.00994**	-0.00589***	-0.00298**	-0.00259***
		0.001	0.0042	0.0011	0.00050	0.00043	0.00409	0.00134	0.00121	0.00050
	SMALL	-0.0099***	-0.0005	-0.0034***	-0.00303***	-0.00536***	-0.00908*	-0.00659***	-0.00195*	-0.00295***
		0.001	0.0036	0.0013	0.00055	0.00042	0.00525	0.00123	0.00114	0.00042
	NOTRADE	-0.0143***	0.0121***	0.0067***	0.00332	-0.00741***		0.00097	0.00823***	0.00157**
		0.003	0.0047	0.0026	0.00350	0.00256		0.00391	0.00114	0.00072
SLOW	LARGE	-0.0058***	0.0002	0.0005	-0.00398***	-0.00544***	-0.00501	-0.00534***	-0.00272	-0.00206***
		0.001	0.0062	0.0018	0.00113	0.00051	0.00448	0.00163	0.00221	0.00043
	MEDIUM	-0.0080***	-0.0015	-0.0043***	-0.00475***	-0.00581***	0.00006	-0.00617***	-0.00480**	-0.00296***
		0.001	0.0101	0.0013	0.00133	0.00092	0.00477	0.00168	0.00217	0.00114
	SMALL	-0.0090***	0.0113**	-0.0011	-0.00269***	-0.00422***	0.00097	-0.01564***	0.00013	-0.00344***
		0.001	0.0054	0.0016	0.00098	0.00064	0.01312	0.00250	0.00318	0.00083
	NOTRADE	-0.0141***	0.0059	0.0048*	-0.00133	-0.00343*	-0.08073**	-0.00253	0.00045	-0.00191***
		0.002	0.0039	0.0026	0.00162	0.00183	0.03490	0.00201	0.00218	0.00054
	Constant	-0.0012***		Observations	2587686	R-Squared	0.0262	Stock FE	YES	Std. Err. adjusted for 97 clusters
		(0.000)		N. of Groups	97			Time FE	YES	

Table 4.11. Panel regression for the determinants of the absolute deviation of the mid-quote from the opening price (continued)

Panel C: Volume of shares from 8:00 to 8:59

		Limit Orders				Market Orders				Zero Imbalance
Speed	Inventory	New	Qty. Rev	Cancellation	Price Rev	New	Qty. Rev	Cancellation	Price Rev	
FAST	LARGE	-0.0016 ***	-0.0022 **	-0.0003	-0.00245 ***	-0.00118 ***	-0.00051	-0.00026	0.00044	0.00000
		0.000	0.0010	0.0003	0.00051	0.00019	0.00106	0.00038	0.00041	0.00005
	MEDIUM	-0.0010 ***	-0.0004	-0.0006 **	-0.00077 ***	-0.00102 ***	-0.00108	-0.00094 **	-0.00017	-0.00020 ***
		0.000	0.0003	0.0003	0.00025	0.00013	0.00100	0.00047	0.00050	0.00007
	SMALL	-0.0012 **	-0.0002	-0.0002	-0.00089 ***	-0.00095 ***	-0.00162	-0.00063	-0.00099 **	-0.00046 ***
		0.000	0.0007	0.0001	0.00028	0.00015	0.00111	0.00051	0.00048	0.00011
	NOTRADE	-0.0001	0.0014	0.0002	-0.00059	-0.00203 ***	-0.00449 ***	0.00052		0.00006
		0.001	0.0036	0.0004	0.00093	0.00046	0.00003	0.00038		0.00007
MEDIUM	LARGE	-0.0021 ***	0.0030 ***	-0.0009 **	-0.00088 **	0.00028 *	0.00016	-0.00169 ***	0.00020	-0.00009
		0.000	0.0008	0.0004	0.00044	0.00016	0.00250	0.00061	0.00101	0.00008
	MEDIUM	-0.0018 ***	-0.0038 ***	-0.0007 ***	-0.00092 ***	-0.00126 ***	-0.00571 ***	-0.00197 ***	-0.00047	-0.00019 **
		0.000	0.0011	0.0002	0.00023	0.00015	0.00144	0.00036	0.00078	0.00009
	SMALL	-0.0031 ***	-0.0003	-0.0010 **	-0.00080 ***	-0.00154 ***	-0.00096	-0.00223 ***	-0.00031	-0.00037 ***
		0.000	0.0013	0.0004	0.00023	0.00029	0.00095	0.00031	0.00033	0.00009
	NOTRADE	-0.0023 *	0.0059 **	0.0001	-0.00089	-0.00053		-0.00018	0.00081 ***	-0.00004
		0.001	0.0027	0.0010	0.00163	0.00152		0.00179	0.00011	0.00012
SLOW	LARGE	-0.0023 ***	0.0034 ***	0.0005	-0.00043 **	-0.00047 ***	-0.00134	-0.00086 ***	-0.00058 **	-0.00008
		0.000	0.0008	0.0006	0.00020	0.00011	0.00144	0.00032	0.00024	0.00008
	MEDIUM	-0.0027 ***	-0.0022	-0.0010 *	-0.00123 ***	-0.00052 ***	0.00102	-0.00068 ***	-0.00120 ***	-0.00043 ***
		0.000	0.0061	0.0006	0.00042	0.00012	0.00176	0.00023	0.00026	0.00011
	SMALL	-0.0029 ***	0.0031	0.0001	-0.00057 *	-0.00091 ***	-0.01025	-0.00209 ***	-0.00047	-0.00053 ***
		0.001	0.0024	0.0003	0.00029	0.00022	0.01831	0.00080	0.00317	0.00013
	NOTRADE	-0.0049 ***	0.0033 **	0.0016 **	0.00051	-0.00066	-0.32769 ***	-0.00134	-0.00061	-0.00013 *
		0.001	0.0014	0.0008	0.00073	0.00077	0.04257	0.00110	0.00312	0.00007

Constant	-0.0015 ***	Observations	6090450	R-Squared	0.0233	Stock FE	YES	Std. Err. adjusted for 97 clusters
	(0.000)	N. of Groups	97			Time FE	YES	

Table 4.11. Panel regression for the determinants of the absolute deviation of the mid-quote from the opening price (continued)

Panel D: Volume of shares from 8:10 to 8:59

		Limit Orders				Market Orders				Zero Imbalance
Speed	Inventory	New	Qty. Rev	Cancellation	Price Rev	New	Qty. Rev	Cancellation	Price Rev	
FAST	LARGE	-0.0017***	-0.0023**	-0.0003	-0.00254***	-0.00126***	-0.00051	-0.00027	0.00041	0.00001
		0.000	0.0010	0.0003	0.00051	0.00020	0.00106	0.00038	0.00039	0.00005
	MEDIUM	-0.0010***	-0.0004	-0.0006**	-0.00078***	-0.00109***	-0.00095	-0.00094**	-0.00007	-0.00018*
		0.000	0.0003	0.0003	0.00024	0.00014	0.00103	0.00048	0.00048	0.00009
	SMALL	-0.0011***	-0.0003	-0.0002	-0.00089***	-0.00106***	-0.00161	-0.00059	-0.00085*	-0.00040***
		0.000	0.0007	0.0002	0.00029	0.00016	0.00110	0.00050	0.00044	0.00015
	NOTRADE	-0.0019***	0.0022	0.0004	-0.00052	-0.00199***	-0.00449***	0.00051		0.00010
		0.001	0.0041	0.0005	0.00093	0.00046	0.00004	0.00039		0.00007
MEDIUM	LARGE	-0.0018***	0.0031***	-0.0009**	-0.00100**	-0.00175***	0.00018	-0.00164***	0.00039	-0.00015
		0.000	0.0008	0.0005	0.00041	0.00036	0.00248	0.00062	0.00096	0.00010
	MEDIUM	-0.0012***	-0.0040***	-0.0007***	-0.00104***	-0.00137***	-0.00573***	-0.00197***	-0.00041	-0.00003
		0.000	0.0011	0.0002	0.00023	0.00016	0.00146	0.00037	0.00076	0.00023
	SMALL	-0.0019***	-0.0002	-0.0012***	-0.00086***	-0.00141***	-0.00091	-0.00217***	0.00004	-0.00034*
		0.000	0.0013	0.0004	0.00025	0.00030	0.00093	0.00031	0.00031	0.00019
	NOTRADE	-0.0035***	0.0061**	0.0003	-0.00095	-0.00047		-0.00023	0.00048***	-0.00002
		0.001	0.0026	0.0011	0.00159	0.00155		0.00178	0.00014	0.00012
SLOW	LARGE	-0.0015***	0.0033***	-0.0007	-0.00055**	-0.00065***	-0.00098	-0.00086***	-0.00053**	0.00000
		0.000	0.0009	0.0005	0.00022	0.00014	0.00120	0.00032	0.00024	0.00017
	MEDIUM	-0.0012***	-0.0025	-0.0012**	-0.00124***	-0.00055***	0.00106	-0.00065***	-0.00119***	-0.00118***
		0.000	0.0062	0.0005	0.00044	0.00014	0.00177	0.00020	0.00026	0.00035
	SMALL	-0.0020***	0.0020	0.0001	-0.00042*	-0.00076***	-0.01032	-0.00217***	0.00099	-0.00067**
		0.000	0.0016	0.0003	0.00025	0.00028	0.01858	0.00084	0.00270	0.00030
	NOTRADE	-0.0036***	0.0028**	0.0014*	0.00047	-0.00048	-0.32599***	-0.00135	-0.00129	0.00000
		0.001	0.0014	0.0007	0.00074	0.00083	0.04291	0.00111	0.00316	0.00008

Constant	-0.0024***	Observations	2587686	R-Squared	0.0297	Stock FE	YES	Std. Err. adjusted for 97 clusters
	(0.000)	N. of Groups	97			Time FE	YES	

new limit and market orders, from all traders, indicating their contribution to price discovery. However, the coefficients for new limit orders are larger than those for new market orders except in the case of FAST / NOTRADE traders, indicating the larger role new limit orders play in price discovery. Quantity revisions from most of the groups are positive, indicating a deterioration of price discovery. Cancellations for limit orders are mixed, and mostly insignificant. After the exclusion of the first 10 minutes, new limit and market orders from each group still contribute to the price discovery (Table 4.11 Panel B). New market orders from FAST / SMALL and FAST / LARGE traders show statistically significant contributions. The results for the last 10 minutes and the very last minute (unreported results, which are available upon request from the authors) are similar to those from the analysis excluding the first 10 minutes. The most stable contribution comes from new limit and market orders.

Table 4.11 Panel C shows the results obtained by using the number of shares instead of the number of orders from each group. The negative coefficient for new limit and market orders remains unchanged. The positive coefficient for new market orders is only seen for the MEDIUM / LARGE group, and is marginally significant. The sizes of the coefficients for new limit and market orders are more similar across the groups than are those in the case of the number of orders shown in Panel A. Overall, the results are consistent with Table 4.7. They confirm that new limit orders contribute consistently towards price discovery throughout the pre-opening period and across traders.

4.5.4 Tests of unbiasedness of the pre-opening quotes

We next repeat the test for price efficiency on the pre-opening quotes using an unbiasedness regression that has been widely used in the literature.[21] Specifically, the first to use it are Biais, Hillion, and Spatt (1999), who use it to characterize the extent to which there is learning and price discovery in the pre-opening period. They use the closing price of the day as a proxy for the equilibrium price v. We modify

[21] 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).

their framework for our purposes and estimate equation (4.6) as follows:

$$\nu - E(\nu|I_0) = \alpha_t + \beta_t [P_t - E(\nu|I_0)] + Z_t \tag{4.6}$$

where ν is the opening price (instead of the closing price used in Biais, Hillion, and Spatt (1999)), P_t is the pre-opening mid-quote, and $E(\nu|I_0)$ is the previous day's closing price. The distribution of the change in price, from the previous da 's close to the mid-quote, varies over time as the opening time approaches. The amount of noise in the mid-quote is also likely to vary with time. In this spirit, we estimate the unbiasedness regression using the specification shown in equation (4.6), for each 100-millisecond interval and for each stock in our sample period. If the pre-opening mid-quote 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 is equal to 1, over the pre-opening period.

This section is structured as follows. First, we analyze the cross-sectional patterns in the estimation results of the unbiasedness regression. Second, we compare the results of the unbiasedness regression for three different time periods (November-December 2009, January-March 2010, and April-May 2013) to exploit a quasi-natural experiment of the "Arrowhead" introduction.

4.5.4.1 Cross-sectional analysis of the unbiasedness of the pre-opening quotes

We follow the same approach as for the cross-sectional analysis of the absolute deviation of the mid-quote from the opening price (see Section 4.5.3.3). In particular, we split stocks into two groups based on the activity of FAST / SMALL, FAST / MEDIUM, MEDIUM / SMALL, and MEDIUM / MEDIUM traders. The activity of each trader group is measured by the proportion of aggressive messages (messages that have the potential to change the prevailing mid-quote) for each stock across all days. We separate stocks for which the activity of any of the trader groups (FAST / SMALL, FAST / MEDIUM, MEDIUM / SMALL, or MEDIUM / MEDIUM) exceeds 30% (18 stocks), from all other stocks (79 stocks).

Figure 4.8 shows the β estimates and t-statistics under the null hypothesis that β is equal to 1 for every 100-millisecond interval in the last 200 seconds of the pre-opening period, for these two groups of stocks, for April and May 2013. Remarkably, the β for stocks subject to high activity from the FAST / MEDIUM and SMALL/MEDIUM trader groups differs insignificantly from 1 during the last 200 seconds. On the contrary, the β for stocks subject to low activity from the FAST / SMALL, FAST / MEDIUM, MEDIUM / SMALL, and MEDIUM / MEDIUM traders increases slowly from 0.7 to 1. Even during the last 100 milliseconds, the β for this group of stocks is still significantly different from 1. Overall, these results are consistent with FAST / SMALL, FAST / MEDIUM, MEDIUM / SMALL, and MEDIUM / MEDIUM traders improving price discovery during the pre-opening period.

4.5.4.2 Unbiasedness of the pre-opening quotes and "Arrowhead" introduction

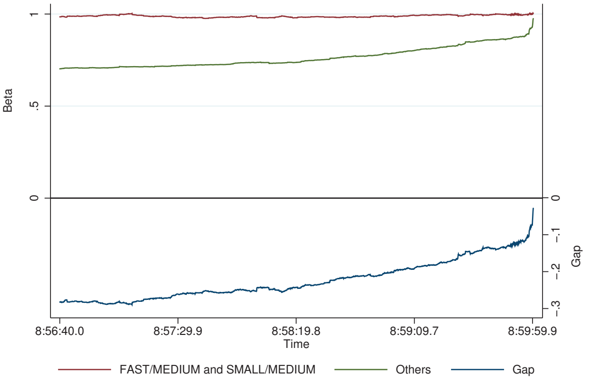
On January 4, 2010, the TSE introduced the "Arrowhead" system, which substantially reduced the latency in the Japanese stock market. For benchmarking purposes, we refer to the period from November 2009 through March 2010 as the comparative (control) period. In particular, the initial three months of January 2010 give us the opportunity to examine the turning point of the TS 's platform change and its effect on order submission behavior, with the other months being used for robustness checks to capture the effect of the exogenous event — the introduction of the "Arrowhead" system.

Figure 4.9 shows the average of the coefficients, β_t, and the bands of $+/-2\sigma$ of the cross-sectional standard errors over time, for three different time periods (November-December 2009, January-March 2010, and April-May 2013). In order to investigate price discovery at the millisecond level, we run the same regression for the three different periods, every 100 milliseconds of the last 200 seconds (Figure 4.9 Panel A) and every 10 milliseconds in the last 20 seconds (Figure 4.9 Panel B). The inclusion

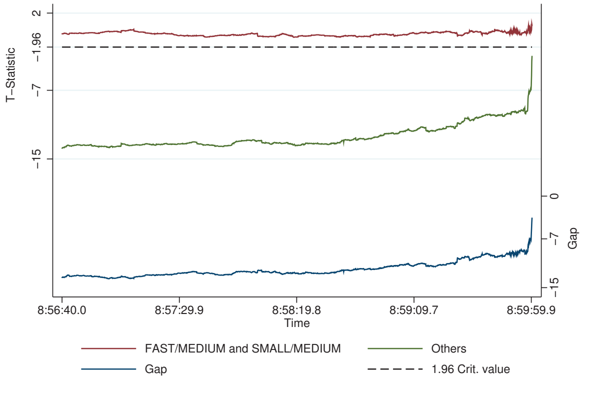
Figure 4.8. Comparison of the test of unbiasedness regressions between stocks with different levels of low-latency trader participation

Using mid-quotes, at each 100-millisecond interval, we estimate equation (4.6) for every 100 milliseconds of the last 200 seconds of the pre-opening period (8:56:40.000 - 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 the aggressive activity of FAST&MEDIUM / 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 in the last 200 seconds



Panel B: T-statistic in the last 200 seconds



of the two additional periods allows us to test changes in the price discovery process due to the introduction of the "Arrowhead" low-latency trading platform and the implementation of several other institutional changes, such as the co-location service (see Uno and Shibata (2012)).

The implementation of the new trading platform that changed the latency caused a shift in the behavior of all traders. This structural change created room for the HFTs to exploit the breakthrough in the latency. Thus, this natural experiment is ideal for assessing the effect of the latency regime on price informativeness: reducing the latency potentially increases the speed of order flow, which, in turn, may lead to more accurate prices, better liquidity, and faster price discovery.

To test these hypotheses, we investigate whether the time when β becomes insignificantly different from 1 is the same or different across the three regimes. This analysis shows whether there was a structural change due to the introduction of the "Arrowhead" system.

Figure 4.9 Panel B shows that the β becomes insignificantly different from 1, at a time of 550 milliseconds before 9 am, in November-December 2009. However, β never reaches 1 in either April-May 2013 or November-December 2009: the average β in April-May 2013 in the last 10 milliseconds before 9 am is around 0.9, while the corresponding average β in November-December 2009 is only around 0.7. The comparison between 2013 and 2010 suggests that the introduction of "Arrowhead" and its increased usage by HFTs delayed price discovery by 550 milliseconds. From 2010 to 2013, the proportion of orders coming through co-location servers more than tripled, from 10%-15% to more than 50% (Hosaka (2014)). Although the moment at which the β becomes 1 is delayed in 2013, it does reach 0.9 much earlier than in 2010. The convergence path for 2010 shows a stepwise trend, a symptom of caution in the quote submissions from HFTs. The fact that β does not reach 1 at all in 2009 is indicative of slow price discovery and inaccurate opening prices. This may partially be due to the fact that 32 stocks out of 97 in our sample experienced a tick-size change, which became effective in January 2010. The larger tick size may also have contributed to the amplification of the difference between the opening price and the mid-quote.

Overall, the results indicate that price efficiency improved in the low-latency regime following the introduction of "Arrowhead". The new latency regime created a different trading environment for all players, but the learning process required for traders to exploit the improved speed efficiently will require time and a careful calibration of the algorithms.

HFTs were not present in the TSE before 2010, because of the three-second matching interval used in the continuous session (see Uno and Shibata (2012)). The natural experiment that we analyze here shows that the introduction of the "Arrowhead" system was an exogenous event that triggered several consequences: changed price accuracy, the need for adaptation by HFTs, a reduction of price dispersion, and an improvement of liquidity. However, we caution that, given the design of the experiment and the absence of a control group, we cannot say anything conclusive about causality. We can only conclude that our findings are consistent with the hypothesis that high-frequency quote updates contribute to price discovery.

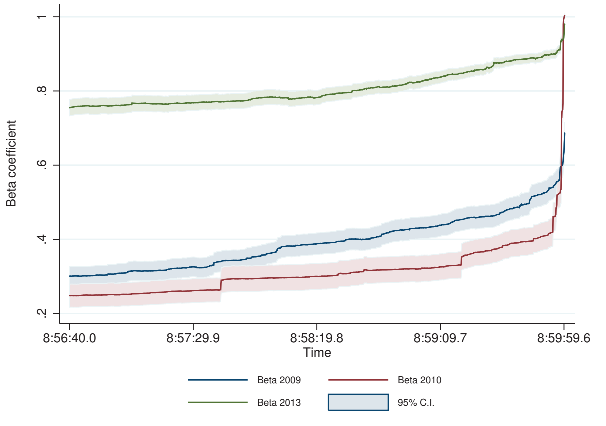
4.5.5 Trading activity during the call auction and first 30 minutes of the continuous session

In this section, we discuss the trading patterns of different trader groups in the opening call auction as well as the first 30 minutes of the continuous session. In particular, we examine how orders from each trader group are filled at the opening call auction and how different trader groups behave during the first 30 minutes of the continuous session. We focus on the first 30 minutes of the continuous auction as we want to analyze the difference in trader behavior, based on the same information, but in the pre-opening call and continuous trading session settings. If we extended the sample to the full continuous session, we would contaminate our analysis with new information arriving in the market later in the trading day. The pre-opening call auction is the closest approximation in the equity markets to frequent batch auctions, as suggested by the theoretical analysis of Budish, Cramton, and Shim (2015), with the major difference between the two being the information dissemination before the auction takes place: no information in the case of frequent batch auctions versus

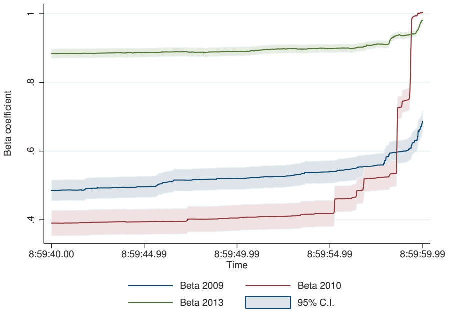
Figure 4.9. Tests of unbiasedness regressions of the pre-opening mid-quotes

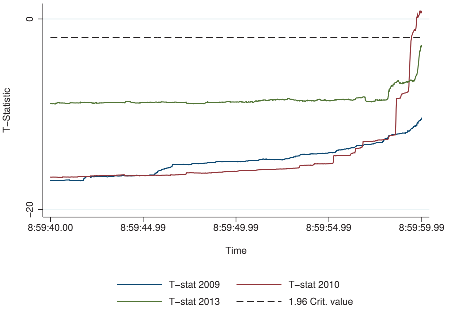
Using mid-quotes, we estimate equation (4.6) for Panel A every 100 milliseconds in the last 200 seconds of the pre-opening period (8:56:40.000 - 8:59:59.999) and for Panel B every 10 milliseconds in the last 20 seconds (8:59:40.000 - 8:59:59.999), for each of the 97 stocks from the TOPIX100, in three different periods (Nov-Dec 2009, Jan-Mar 2010, and Apr-May 2013). The figures show the averages of the β coefficients and the t-statistics under the null hypothesis that β is equal to 1. The tick-by-tick data, time stamped to the millisecond, are obtained from the Thomson-Reuters Tick History Database.

Panel A: Beta coefficient estimated each second of the last 200 seconds



Panel B: Beta coefficient and t-statistic estimated every 10 milliseconds of the last 20 seconds





dissemination of the pre-opening order flow in the case of the TSE pre-opening period. An additional argument for our choice of trading periods for our analysis is that a significant fraction of the daily trading volume is executed during the opening call auction (around 5%) and the first 30 minutes of the continuous session (around 15%).

4.5.5.1 Liquidity provision

We start by investigating the role of different trader groups in the liquidity provision during the call auction and the first 30 minutes of the continuous trading session. In the case of the opening call auction, trading activity is said to provide (consume) liquidity if traders trade in the opposite (same) direction to the price movement.

Table 4.12 presents the trading activity during the call auction. We report the liquidity-demanding and liquidity-supplying trading volumes relative to the total trading volume during the pre-opening call, averaged across stock-days. The most active trader groups during the call auction are FAST&MEDIUM / SMALL&MEDIUM. These traders are jointly responsible for roughly 70% of the volume executed during the opening call auction and are present on both sides of the market. We also conduct a *t*-test of whether the imbalances between liquidity demand and liquidity supply are significantly different from 0. We show that the FAST / SMALL, MEDIUM / SMALL, and MEDIUM / MEDIUM trader groups have negative imbalances between liquidity demand and liquidity supply, which are significant at the 1% level: -0.98%, -1.22%, and -2.58%, respectively. In other words, these trader groups act as net liquidity providers during the opening call auction. Then, we investigate the behavior of the traders during the first 30 minutes of the continuous trading session. In this case, we refer to orders that initiate the transaction as liquidity-consuming and those on the opposite side of the transaction as liquidity-providing. 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.[22] We discard those transactions of the continuous trading session for which we cannot identify the initiating order.[23]

[22] 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)).

[23] 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

Table 4.12. Liquidity consumption and liquidity provision at the opening call auction and during the first 30 minutes of the continuous trading session

This table shows liquidity consumption and supply for 9 trader groups (the NOTRADE category is omitted as these traders do not trade during the stock-day) and the imbalance between liquidity consumption and liquidity supply 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 4.3 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 in the pre-opening period, but do not trade at the opening call auction (Active-w/o-Trade), and traders that participate in 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 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 to 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 liquidity-demanding and liquidity-supplying trading volumes 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 the t-test for whether the imbalance is significantly different from 0. ***, **, and * indicate the 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.

		Call auction			First 30 minutes of continuous session								
Speed	Inventory	Consume	Supply	Imbalance	Consume			Supply			Imbalance		
					Non-Active	Active-w/o-Trade	Active-w-Trade	Non-Active	Active-w/o-Trade	Active-w-Trade	Non-Active	Active-w/o-Trade	Active-w-Trade
FAST	LARGE	6.85%	-5.18%	1.52%***	2.61%	0.68%	1.41%	-4.17%	-1.23%	-2.71%	-1.58%***	-0.79%***	-1.64%***
	MEDIUM	28.50%	-25.29%	3.16%***	8.13%	1.39%	7.78%	-7.72%	-2.53%	-15.42%	0.41%***	-1.28%***	-7.75%***
	SMALL	18.01%	-18.88%	-0.98%***	29.62%	1.34%	7.60%	-14.85%	-3.80%	-12.42%	14.76%***	-2.79%***	-4.97%***
MEDIUM	LARGE	5.31%	-5.17%	0.04%	2.31%	0.74%	0.75%	-2.96%	-1.32%	-1.06%	-0.66%***	-0.66%***	-0.34%***
	MEDIUM	14.03%	-16.60%	-2.58%***	5.02%	1.41%	3.86%	-2.65%	-2.71%	-6.03%	2.39%***	-1.35%***	-2.23%***
	SMALL	12.44%	-13.62%	-1.22%***	8.29%	1.35%	4.72%	-2.84%	-2.44%	-6.73%	5.46%***	-1.13%***	-2.09%***
SLOW	LARGE	9.61%	-9.11%	0.43%*	6.67%	0.61%	0.42%	-1.85%	-1.38%	-0.57%	4.83%***	-0.83%***	-0.14%***
	MEDIUM	4.91%	-4.98%	-0.18%	2.28%	0.63%	0.57%	-0.68%	-1.16%	-0.87%	1.62%***	-0.55%***	-0.32%***
	SMALL	2.82%	-2.93%	-0.18%**	1.95%	0.45%	0.43%	-0.86%	-0.80%	-0.54%	1.15%***	-0.38%***	-0.12%***

Table 4.12 presents the trading activity during the first 30 minutes of the continuous trading session. As the table shows, FAST / SMALL traders that do not participate at all during the pre-opening period (Non-Active) show the highest amount of trading activity, with around 30% of cases consuming liquidity with their trades, and with roughly 15% of cases supplying liquidity. This group of traders are the main liquidity consumers in the market, based on the imbalance between liquidity consumption and liquidity provision (14.8%). The main liquidity providers are FAST / SMALL & MEDIUM traders that are active during the pre-opening period and trade at the call auction (Active-w-Trade). In fact, in this case, their imbalances between liquidity supply (7.6% and 7.8%) and liquidity demand (-12.4% and -15.4%) are around -5.0% and -7.8%, which are significantly different from 0 at the 1% level. The next most important net liquidity providers are the FAST / SMALL (Active-w/o-Trade) and MEDIUM / SMALL & MEDIUM (Active-w-Trade) groups, with imbalances ranging between -2.1% and -2.8%. To sum up, the FAST&MEDIUM / SMALL&MEDIUM (Active-w-Trade) traders are the main liquidity providers on the market for both the call auction and the first 30 minutes of the continuous session. On the contrary, the FAST / SMALL (Non-Active) traders are the main liquidity consumers during the first 30 minutes of the continuous session.

4.5.5.2 Price discovery

We now move on to the analysis of the role of different trader groups in the price discovery process during the first 30 minutes of the continuous trading session. In order to estimate the price discovery contribution, we follow the methodology developed for the pre-opening period with slight modifications. First, we compute the deviation of the trading prices from the price at 9:30 am:

$$Deviation930_{j,k,n} = \left| \frac{P_{j,k,n}}{P930_{j,k}} - 1 \right| \times 100 \tag{4.7}$$

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 $P_{j,k,n}$ is the trading price at the time of the n-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 return during the first 30 minutes of the continuous trading, we use the average trading price between 9:30 and 9:35 to avoid the bid-ask bounce problem. Then we define the price discovery contribution as follows:

$$PDC930_{j,k,n} = Deviation930_{j,k,n} - Deviation930_{j,k,n-1} \tag{4.8}$$

$Deviation930_{j,k,n}$ is the absolute deviation of the trading price from the price at 9:30 am at the time of the n-th trade for stock j on day k (see equation (4.7)). $Deviation930_{j,k,n-1}$ is the absolute deviation of the trading price from the price at 9:30 am at the time of the $(n-1)$-th trade for stock j on day k. The difference between $Deviation930_{j,k,n}$ and $Deviation930_{j,k,n-1}$ is the contribution to price discovery made by the order that initiates the n-th trade. We define the WPC for stock j on day k and trade n as

$$WPC_{j,k,n} = \frac{PDC930_{j,k}}{\sum_{j=1}^{J} |PDC930_{j,k}|} \times \frac{PDC930_{j,k,n}}{PDC930_{j,k}} \tag{4.9}$$

where $PDC930_{j,k,n}$ is the price discovery contribution of the order that initiates the n-th trade for stock j on day k, and $PDC930_{j,k}$ is the accumulated price discovery contribution during the pre-opening period for stock j on day k. We winsorize $PDC930_{j,k,n}$ at the 0.1% and 99.9% levels.

Table 4.13 shows the results of the WPC analysis for the first 30 minutes of the continuous trading session for the continuation and reversal regimes, separately. We also distinguish between the different order types initiating the transaction: new and revised limit orders that cross or lock in the prevailing bid-ask spread, and new and revised market orders. The largest contributions to price discovery are made by two groups of traders: FAST / SMALL (Non-Active) (-39.99%) and FAST / MEDIUM (Non-Active) (-18.25%). Put differently, the FAST / SMALL & MEDIUM (Non-

Table 4.13. Contribution to weighted price discovery by type of order during first 30 minutes of continuous session

This table presents the summary statistics for the weighted price discovery contribution (WPC) during the first 30 minutes of the continuous session, the percentage amount by which each new transaction moves the trading price closer to the trading price at 9:30 am divided by the accumulated price discovery contribution during the first 30 minutes of the continuous session, as defined in equation (4.9). WPC is attributable to 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 orders with a limit price equal to the best ask (bid) price ("Lock"). We distinguish between WPC for each of the 6 different types of order. We divide all traders into 12 groups on a stock-day basis, as described in Table 4.3, 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 in the pre-opening period but do not trade at the opening call auction (Active-w/o-Trade), and traders that participate in the pre-opening period and trade at the call auction (Active-w-Trade). Panel A describes WPC for Non-Active traders, Panel B for Active-w/o-Trade traders, and Panel C for Active-w-Trade traders, for 97 stocks from the TOPIX100 during the sample period of April-May 2013. Order flow data, with order IDs as well as virtual server IDs, are provided by the Tokyo Stock Exchange.

Speed	Inventory	Total	New limit orders		Revised limit orders		Market orders	
			Cross	Lock	Cross	Lock	New	Price revision
Panel A: Weighted price discovery contribution of Non-Active traders								
FAST	LARGE	-5.37%	0.17%	-5.12%	0.04%	-0.04%	-0.29%	-0.13%
	MEDIUM	-18.25%	-0.26%	-18.68%	-0.08%	0.01%	0.77%	-0.01%
	SMALL	-37.99%	1.74%	-39.96%	0.00%	-0.38%	0.53%	0.08%
MEDIUM	LARGE	-2.48%	0.28%	-3.93%	0.11%	0.32%	0.60%	0.14%
	MEDIUM	-8.38%	-0.79%	-8.94%	0.08%	0.05%	1.16%	0.05%
	SMALL	-8.05%	0.15%	-8.76%	-0.03%	0.09%	0.41%	0.08%
SLOW	LARGE	-7.33%	-0.52%	-6.30%	0.01%	0.14%	-0.75%	0.09%
	MEDIUM	-1.76%	0.19%	-2.10%	0.01%	0.00%	0.10%	0.04%
	SMALL	-2.61%	-0.07%	-2.61%	0.00%	-0.02%	0.07%	0.01%
Panel B: Weighted price discovery contribution of Active-w/o-Trade traders								
FAST	LARGE	-0.64%	-0.08%	-0.90%	-0.04%	0.15%	0.21%	0.02%
	MEDIUM	-4.35%	-0.60%	-3.69%	-0.06%	-0.25%	0.17%	0.07%
	SMALL	-1.33%	-0.39%	-1.07%	-0.02%	0.06%	0.17%	-0.08%
MEDIUM	LARGE	1.58%	0.23%	-0.01%	0.08%	0.10%	1.06%	0.12%
	MEDIUM	1.74%	0.33%	-0.34%	-0.04%	0.19%	1.18%	0.42%
	SMALL	-1.63%	0.16%	-1.10%	-0.03%	-0.20%	-0.20%	-0.25%
SLOW	LARGE	1.04%	-0.08%	0.09%	0.10%	0.05%	0.77%	0.10%
	MEDIUM	-0.88%	-0.40%	-0.50%	-0.03%	-0.02%	0.09%	-0.01%
	SMALL	-1.23%	-0.24%	-0.26%	-0.09%	-0.12%	-0.60%	0.08%
Panel C: Weighted price discovery contribution of Active-w-Trade traders								
FAST	LARGE	-1.87%	1.01%	-1.30%	0.03%	-0.56%	-0.94%	-0.12%
	MEDIUM	-10.53%	-0.59%	-14.46%	0.39%	-1.16%	5.06%	0.23%
	SMALL	3.85%	2.10%	2.63%	0.38%	-1.53%	0.11%	0.17%
MEDIUM	LARGE	0.82%	0.40%	0.04%	-0.07%	-0.24%	0.98%	-0.30%
	MEDIUM	7.77%	-0.28%	0.93%	0.09%	0.35%	6.14%	0.54%
	SMALL	-3.08%	-0.91%	-2.60%	0.27%	-0.73%	0.75%	0.14%
SLOW	LARGE	0.11%	0.22%	-0.02%	-0.01%	-0.06%	0.02%	-0.03%
	MEDIUM	0.89%	0.12%	-0.06%	-0.04%	0.06%	0.72%	0.10%
	SMALL	-0.02%	-0.19%	0.17%	0.00%	0.02%	0.01%	-0.03%
TOTAL		-100.00%	1.71%	-118.85%	1.04%	-3.74%	18.33%	1.52%

Active) traders that are the main consumers of liquidity are, at the same time, responsible for more than 50% of the price discovery process. On the contrary, among the traders that are actively supplying liquidity, that is, FAST&MEDIUM / SMALL&MEDIUM (Active-w-Trade), only FAST / MEDIUM traders contribute toward improving the price discovery (-10.53%, the third largest contribution), while the other traders deteriorate price discovery. Breaking down the WPC by order type, we show that the majority of the price discovery occurs via locked limit orders (-118.85%), while the majority of the price deterioration occurs via new market orders.

Based on the analysis of liquidity provision and weighted price discovery contribution, we show that FAST&MEDIUM / SMALL&MEDIUM traders that are active during the pre-opening period (Active-w-Trade) are different from the FAST&MEDIUM / SMALL&MEDIUM traders that are active only during the continuous session (Non-Active). The trading behavior of the former group is very close to the behavior of low-latency market makers, while the behavior of the latter group represents the behavior of low-latency informed traders.

The price formation process between the opening call auction and the following continuous session has been studied extensively in the literature. In particular, it has been shown that stock prices exhibit non-trivial reversals in the first half hour of the continuous trading session relative to the overnight price movement (e.g., Stoll and Whaley (1990) and Amihud and Mendelson (1991)). Therefore, as a robustness check, we separate stock-days into two regimes (following Brogaard, Riordan, Shkilko, and Sokolov (2015)). The continuation (reversal) regime represents cases in which the overnight return is of the same (opposite) sign as the return during the first 30 minutes of continuous trading. We find that the results are qualitatively similar to the results obtained without such separation (the results are available from the authors upon request).

4.6 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 that has come to dominate global equity markets. Key questions we ask in this research are whether, in the absence of trading, low-latency traders (including HFTs) still participate in the market, and how the presence of low-latency 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 low-latency trading.

We classify traders into three speed and four inventory groups (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 batch auction to a lesser extent than in the continuous session. 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 MEDIUM 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 low-latency traders 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. Cancellation of limit orders deteriorates price discovery, but cancellation of market orders improves it.

It is important to note that, due to the lack of immediacy in execution, the presence of FAST traders in the pre-opening period is smaller than in the continuous session. 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 strategically select stocks in which they are more active, based on the stocks' characteristics.

Our results suggest that three quarters of FAST / SMALL traders do not participate in the opening call auction. These traders are the most active players in the first 30 minutes of the continuous session (they are responsible for initiating around 30% of the trades). This suggests that the majority of low-latency traders prefer an environment in which immediate execution is possible. Our findings also suggest that FAST / SMALL traders who are active only during the continuous session (Non-Active) are responsible for the majority of the price discovery process and the majority of liquidity consumption. On the contrary, FAST / SMALL traders 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. Based on these results, we conclude that low-latency traders that are active only during the continuous session may be viewed as informed low-latency traders, while low-latency traders that are active both at the opening call auction and the continuous session are low-latency market makers.

However, our results cannot be considered as direct evidence concerning trader behavior in the periodic batch auction. The opening call auction and the periodic batch (call) auction differ from each other in two important ways. First, the opening call auction is not a sealed auction, while frequent batch auctions are (as suggested by Budish, Cramton, and Shim (2015)). Put differently, information about the pre-opening order flow is disseminated to the market in the case of the opening call auction, while there is no information dissemination in the case of the frequent batch auctions proposed by Budish, Cramton, and Shim (2015). Second, although immediate execution is not possible in either auction type, the opening call auction is followed by the continuous trading session, which allows market participants to unwind the positions taken during the opening call auction almost immediately, if necessary. However, in the case of frequent batch auctions, additional waiting time is introduced in between auctions, therefore increasing the risk of holding undesired inventory (see, e.g., Garbade and Silber (1979) and Kehr, Krahnen, and Theissen (2001)). These two key differences may lead to different participation rates being exhibited by low-latency traders in the opening call auction versus the frequent batch auction.

To sum up, our results suggest that HFTs that participate in the pre-opening session are different from those that only participate in the continuous session. 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 low-latency traders. Our findings offer some preliminary evidence in the context of the debate on the relative merits of periodic batch versus continuous auctions.

Chapter 5

Summary and Concluding Remarks

This dissertation consists of three empirical papers in the field of market microstructure. These papers investigate the impact of increased interconnectedness of the financial markets and the vast trading speed improvements on two important functions of financial markets: price discovery and liquidity provision.

Chapter 2 examines the role of liquidity and trading activity in the origination and propagation of shocks to prices across international equity markets. The findings show that equity markets are strongly interconnected with respect to the transmission of shocks to prices and trading activity, while liquidity dry-ups seem to be isolated events. The findings suggest that shocks to prices are the result of information rather than liquidity as prices do not revert after the occurrence of shock and are strongly associated with macroeconomic news announcements. For investors, the findings have implications with respect to international portfolio diversification. From a regulatory perspective, the findings suggest that in order to reduce the vulnerability of financial system cross-country policies have to be implemented that address shock transmission at the intraday level.

Chapter 3 analyzes the choice that an informed trader makes between market (consuming liquidity) and limit (supplying liquidity) orders and how this choice is affected by the rise of algorithmic trading. Contrary to the traditional view in the literature, the findings show that the majority of informed trading takes place via limit (not market) orders. Moreover, informed algorithmic traders actively use limit orders for informed trading. Combined together, these findings suggest that price informativeness does not necessarily come at the expense of low liquidity, even in the presence of algorithmic traders. The findings suggest that measures of informed trading activity should be re-evaluated in order to incorporate informed trading via limit orders. Furthermore, I show that there is a large cross-sectional variation in the order types used for informed trading. Therefore, investors may want to adjust their strategies to decrease the likelihood of having an informed counterparty. The findings also suggest that regulators should distinguish between different types of algorithmic traders when imposing restrictions/taxes on their activity.

Chapter 4 examines whether low-latency traders are improving or deteriorating price discovery in the pre-opening period. The findings show that low-latency traders actively participate in the pre-opening period despite the absence of immediate execution, although to a lesser extent than during the continuous trading session. Furthermore, low-latency traders lead price discovery during the pre-opening period. The findings contribute to the ongoing debate on whether continuous trading or frequent batch auctions are an appropriate market design in the presence of low-latency traders. In other words, whether continuous trading leads to an endless arms race between traders in terms of speed, while the speed advantage of the low-latency traders is negligible in the call auction setup. The findings also show that if the pre-opening order flow is disseminated to the public, then low-latency traders still can take advantage of their speed by delaying their actions to the very last moment before the call auction takes place.

Overall, my current and future research agenda is related to the design of the well-functioning financial market with specific focus on the regulation of the low-latency traders.

Nederlandse samenvatting (Summary in Dutch)

Deze dissertatie bestaat uit drie empirische artikelen binnen de literatuur die zich bezig houdt met de microstructuur van de markt. Deze artikelen onderzoeken het effect van de toegenomen verwevenheid van financiële markten en de grote handelssnelheid-verbeteringen op twee belangrijke functies van financiële markten: prijsontdekking en liquiditeitsvoorziening.

Hoofdstuk 2 onderzoekt de rol van liquiditeit en handelsactiviteit in het ontstaan en verspreiden van prijsschokken over internationale aandelenmarkten. We vinden dat aandelenmarkten inderdaad sterk verweven zijn met betrekking tot de overbrenging van prijsschokken en handelsactiviteit, terwijl liquiditeitsopdrogingen geïsoleerde gebeurtenissen lijken te zijn. Onze bevindingen suggereren dat prijsschokken voortvloeien uit informatie in plaats van liquiditeit omdat prijzen zich niet herstellen na een schok en ze sterk geassocieerd zijn met macro-economisch nieuws. Vanuit het oogpunt van een investeerder hebben onze bevindingen implicaties met betrekking tot internationale portfolio diversificatie. Vanuit het perspectief van de autoriteiten suggereren onze bevindingen dat cross-country beleid dat zich richt op het reduceren van de intraday transmissie van schokken geïmplementeerd dient te worden met het doel de kwetsbaarheid van het financiële systeem te verminderen.

Hoofdstuk 3 analyseert de keuze die een geïnformeerde handelaar maakt tussen market orders (die liquiditeit consumeren) en limit orders (die liquiditeit aanbieden) en hoe deze wordt beïnvloed door de opkomst van algoritmisch handelsverkeer. In contrast met het perspectief vanuit de literatuur toon ik aan dat het grootste deel van de geïnformeerde handel plaats vindt via limit (geen market) orders. Daarnaast gebruiken handelaren die algoritmisch handelen limit orders regelmatig voor geïnformeerde handel. Samen suggereren deze bevindingen dat het informatiegehalte van de prijs niet noodzakelijkerwijs ten koste gaat van liquiditeit, zelfs in het bijzijn van handelaren die algoritmisch handelen. Mijn bevindingen impliceren dat het nodig is om maatstaven van geïnformeerde handel opnieuw te bekijken en ervoor te zorgen dat ze geïnformeerde handel via limit orders ook meenemen. Daarnaast laat ik zien dat er grote variatie is in de cross-sectie van order types die gebruikt worden in geïnformeerde handel. Daarom zouden investeerders hun strategieën kunnen aanpassen om de waarschijnlijkheid te verminderen met een geïnformeerde tegenpartij te handelen. Mijn bevindingen suggereren ook dat autoriteiten onderscheid zouden moeten maken tussen verschillende types van handelaren die algoritmisch handelen bij het opleggen van belastingen of beperkingen aan de activiteiten van dergelijke handelaren.

Hoofdstuk 4 onderzoekt of snelle handelaren prijsontdekking bevorderen of tegenwerken in de periode voor het opengaan van de beurs. Wij laten zien dat snelle handelaren zich actief bezighouden in deze periode ondanks dat orders niet onmiddellijk worden uitgevoerd, hoewel in mindere mate dan gedurende de continue sessie. Daarnaast tonen wij aan dat snelle handelaren het proces van prijsontdekking leiden tijdens de periode voor het opengaan van de beurs. Onze bevindingen dragen bij aan het huidige debat of continue handel of frequente veilingen een passend markt ontwerp zijn in het bijzijn van snelle handelaren. Met andere woorden, of continue handel leidt tot een eindeloze wapenwedloop tussen handelaren met betrekking tot snelheid, terwijl het snelheidsvoordeel voor een snelle handelaar verwaarloosbaar is in de veiling set-up. We laten zien dat de snelle handelaren nog steeds gebruik kunnen maken van hun snelheid door hun acties uit te stellen tot het allerlaatste moment voordat de veiling plaats vindt, als de orders die ingediend zijn voor het opengaan van de beurs openbaar zijn.

Globaal gezien zijn mijn huidige en toekomstige onderzoeksagenda's gerelateerd aan het ontwerp van goed functionerende financiële markten met een specifieke focus op de regulering van snelle handelaren.

Appendices

Appendix A

The Propagation of Shocks Across International Equity Markets: A Microstructure Perspective

A.1 Sample selection and data screens

This appendix describes the sample and data filters used in the paper. We start with a detailed description of the data sources and sample selection, subsequently discuss our data screens, and conclude with a discussion of potential limitations in our sample construction.

A.1.1 Data sources and sample selection

We use two databases to build our sample: Datastream and Thomson Reuters Tick History (TRTH). From the former, we obtain Reuters Instrument Codes (RICs) for all common stocks that are traded on 12 exchanges around the world. Then, we identify common stocks that were ever part of the major local equity index for each of these exchanges from 1996 till 2011 through the TRTH Speedguide. We obtain tick-by-tick data on trades and quotes for these stocks from TRTH. The exchanges in our sample can be classified into three regions based on time zones: America, Asia, and Europe/Africa. The American region includes the following countries (the major equity index used is in parentheses): Brazil (BOVESPA), Canada (TSX COMPOSITE), Mexico (IPC), and the U.S. (S&P100). The Asian region includes Hong Kong (HSI), India (NIFTY50), Japan (NIKKEI225), and Malaysia (KLCI). The European/African region includes France (CAC40), Germany (DAX), South Africa (JALSH), and the U.K. (FTSE100). Data for these exchanges are generally available over 1996-2011, with a few exceptions. In particular, data availability for Germany and South Africa starts in 1997, for Mexico in 1998, for India in 2000, and for Brazil in 2004.

We obtain the historical opening hours for each of the exchanges from several sources: the TRTH Speedguide, Skeete (2004), exchanges' websites, and the Federation of European Securities Exchanges. We cross-check these opening hours by examining the trading activity patterns observed in the data and select the shortest opening hours when in doubt. Since we cannot clearly distinguish between auctions and continuous trading sessions, we disregard the first and the last 15 minutes of each trading day.

A.1.2 Data screens

We filter the data following Rösch, Subrahmanyam, and van Dijk (2015). We use two sets of screens: one set for trade data and another set for quote data. We discard trades when they occur outside the opening hours of the exchange; the trade price is not positive; the trade size is more than 10,000 shares (to exclude block trades from our sample); the trade price differs from the prices of the 10 surrounding ticks by more than 10% since these are likely to be erroneous entries. We discard quotes when quotes occur outside the opening hours of the exchange; the bid and ask prices are not positive; the bid price is higher than the ask price; the bid or ask price differs from the bid or ask price of the 10 surrounding ticks by more than 10% since these are likely to be erroneous entries; the proportional bid-ask spread exceeds 25%. In addition, we discard stock-days if a stock is traded fewer than ten 5-minute intervals per day. When aggregating stock level data to the market-level, we discard 5-minute intervals in which fewer than 10 stocks are traded.

A.1.3 Sample construction limitations

There are several potential limitations in our sample construction. First, we use RICs that ever refer to the stock that was part of the index during our sample period (1996-2011). However, RICs can change through time and TRTH does not provide information on re-used RICs. Therefore, some of the data in our sample could stem from different stocks than the index constituents. Second, for the same reason linking TRTH data to data on the market capitalization of individual stocks (for example, from Datastream) is challenging. All of our analyses are therefore based on equally-weighted averages of the variables across stocks only. We believe that these limitations are not severe due to the trading activity filters we apply: stocks should trade at least ten 5-minute intervals per day. Hereby, we avoid many small and illiquid stocks that could definitely not be part of the index in the time interval under consideration. Because the stocks in our sample are relatively large and liquid, analyzing equally-weighted averages seems an appropriate choice. Using an equally-weighted average also reduces the problem of one stock dominating the whole market (e.g., Nokia in Finland).

A.2 Jump measure (BNS)

This appendix describes the BNS jump measure (Barndorff-Nielsen and Shephard (2006)) computation together with the algorithm that we use to determine the exact 5-minute interval during which a jump occurs. Following Pukthuanthong and Roll (2015), we use jump measures to identify extreme events on financial markets. A jump measure is a statistical non-parametric way to test for jumps in a time-series. In this paper, we use the BNS ratio measure:

$$H_t = \frac{\sqrt{T}\left(\frac{\pi}{2}\frac{B_t}{S_t} - 1\right)}{\sqrt{\upsilon \frac{Q_t}{B_t^2}}} \tag{A.2.1}$$

$$S_t = \Sigma_{k=2}^{T}(V_{k,t})^2 \tag{A.2.2}$$

$$B_t = \Sigma_{k=2}^{T}|V_{k,t}||V_{k-1,t}| \tag{A.2.3}$$

$$Q_t = T \cdot \Sigma_{k=4}^{T}|V_{k,t}||V_{k-1,t}||V_{k-2,t}||V_{k-3,t}| \tag{A.2.4}$$

$$\upsilon = \left(\frac{\pi}{2}\right)^2 + \pi - 5 \tag{A.2.5}$$

where H_t is the BNS ratio measure on day t, S_t is the squared variation on day

t based on 5-minute observations within the day, B_t is the bipower variation on day t based on 5-minute observations within the day, Q_t is the "quarticity" of the process (which is part of the scaling factor for statistics to follow a standard normal distribution), V_{kt} is the variable of interest (returns, changes in proportional quoted or effective spreads, turnover, or order imbalance) at k-th 5-minute interval during day t, T is the total number of valid 5-minute intervals within day t. Under the null hypothesis of no jumps, H_t follows a standard normal distribution.

The BNS jump statistic is based on the assumption that V_{kt} follows a Brownian motion with zero drift and some diffusion plus a Poisson jump process. The bipower variation is the variation of the continuous part of process (the Brownian motion itself) that is free of any jumps, while the squared variation is the variation of the process including the jumps. Thus, without jumps, the squared variation should be approximately the same as the scaled bipower variation. But in case there is a jump, the squared variation exceeds the bipower variation. Hence, the ratio of these two variables gives an indication of whether a jump occurred. If there is a jump on day t, then H_t should be negative and large in absolute terms. In addition to the assumption that our variables follow a Brownian motion with zero drift plus a Poisson jump process, there are several other important assumptions underlying the formulas above. First, we assume that variation is constant over day t. We acknowledge that volatility exhibits intraday patterns, but we circumvent this issue to a large extent by discarding the first and last 15 minutes of the trading session. Second, we also assume that T is large enough ($T \sim T-1 \sim T-3$).

The BNS measure indicates whether there was a jump on a given trading day, but does not pinpoint the exact 5-minute interval when the jump occurs. To determine the exact time of the jump, we propose the following algorithm. We first compute H_t for any day with at least 25 5-minute observations within the day. Then, we check whether we can reject the null hypothesis of no jumps (based on a threshold of the 0.1% percentile of the standard normal distribution). If the null hypothesis is rejected, we search for the most influential observation within day t. In other words, we identify the observation that has the maximum effect on the jump measure and is greater in absolute terms than 1.96 jump-free standard deviations (that is, the square root of the scaled bipower variation). We mark this 5-minute interval as a jump interval. We repeat the procedure (temporarily discarding 5-minute intervals that have been identified as jump observations) until we no longer reject the null hypothesis of no jumps or until there are fewer than 10 observations left. In our sample, the latter of these two conditions never becomes binding.

Appendix B

Intraday Return Predictability, Informed Limit Orders, and Algorithmic Trading

B.1 Sample selection

In this paper, I use two databases to construct my sample: TRTH and CRSP. From the TRTH database, I obtain trade data, best bid-offer data, and limit order book data for the U.S. consolidated limit order book for NYSE-listed securities. I use Reuters Instrumental Codes (RICs), which are identical to TICKERs, to obtain data on common stocks and primary exchange code from CRSP database (PRIMEXCH=N, and SHRCD=10 or 11, EXCHCD =1 or 31). Thus, I focus on all NYSE-listed common stocks that have NYSE as their primary exchange from 2002 untill 2010. These filters leave me with 2,047 unique TICKERs in total.

B.2 Data Screens

I filter the data following Rösch, Subrahmanyam, and van Dijk (2015). First, I discard trades, quotes, and limit order book data that are not part of the continuous trading session. Continuous trading session hours for NYSE are 9:30-16:00 ET and they remain unchanged during the sample period.

Second, I discard block trades, i.e., trades with a trade size greater than 10,000 shares, as these trades are likely to receive a special treatment.

Third, I discard data entries that are likely to be faulty. Faulty entries include entries with negative or zero prices or quotes, entries with negative bid-ask spread, entries with proportional bid-ask spread bigger than 25%, entries that have trade price, bid price, or ask price which deviates from the 10 surrounding ticks by more than 10%.

In addition, I require that at least five levels of the limit order book are available in the end of each one-minute interval. For a stock-day to enter my sample, at least 100 valid one-minute intervals with at least one trade are required.

Appendix C

Low-Latency Trading and Price Discovery without Trading: Evidence from the Tokyo Stock Exchange in the Pre-Opening Period and the Opening Batch Auction

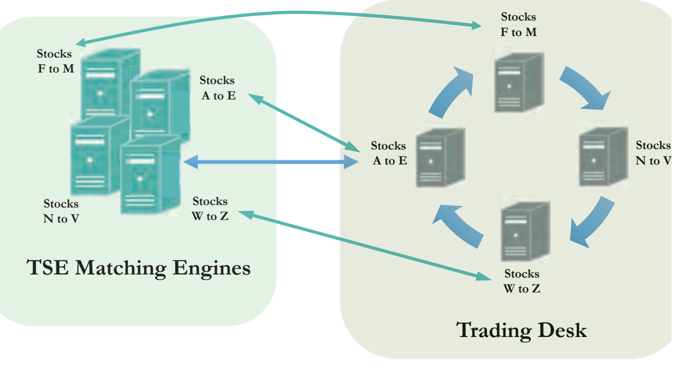
C.1 Configuration of multiple virtual servers (VSs) used by one trader

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 *(i)* accelerated computer-processing speeds, *(ii)* 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 *(iii)* 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 in order 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 4.1 shows, by optimizing the number of stocks per server she can reduce her latency significantly. Figure C.1.1 illustrates one example of a server configuration.

Figure C.1.1. 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 her 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.



C.2 Latency model estimation

Due to the limitation on the number of messages per second per server, the coverage of stocks and intensity of messages of a trader determines the size of their operation. Our novel data server ID allows us to estimate the relation between latency, server configuration, and message intensity with the following equation:

$$\begin{aligned} Latency_{j,k,l} = a + b\ln(Message_{j,k,l}) + \\ c\ln(Nstock_{k,l}/Nserver_l) + d\ln(MaxMessage_{k,l}) + \epsilon_{j,k,l} \end{aligned} \tag{C.2.1}$$

$Latency_{j,k,l}$ is the latency measure for stock j, day k, and trader l. $Message_{j,k,l}$ is the number of messages 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). $MaxMessage_{k,l}$ is the maximum number of messages per second sent by trader l on day k.

Table C.2.1. Latency model estimation

Estimation, using Tobit regression, of the model in equation (C.2.1). $Latency_{j,k,l}$ is the latency measure for stock j, day k, and trader l. $Message_{j,k,l}$ is the number of messages 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). $MaxMessage_{k,l}$ is the maximum number of messages 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: $Latency_{j,t,k}$		
	Coef	t-stat
$Constant$	5.44	571.65
$\ln(Message_{j,t,k})$	-2.08	-1555.60
$\ln(Nstock_{t,k}/Nserver_k)$	0.41	263.60
$\ln(MaxMessage_{t,k})$	-1.35	-489.31
Left-censored obs	73,011	
Right-censored obs	0	
Uncensored obs	3,120,836	
Total obs	3,193,847	

The daily number of stocks per server indicates the trader's speed requirement. The number of messages is used by other HFT studies to identify HFTs who engage in market making. The maximum number of messages 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 latency 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.2.1).

Table C.2.1 shows a strong relation between the number of stocks per server, the total number of messages, and the maximum number of messages per second. The smaller the number of stocks per server, and the larger the number of messages (maximum number of messages per second), the lower is the latency. This result suggests that latency-based classification is equivalent to classification based on the total number of messages.

C.3 Comparative summary statistics for active and non-active trader groups

In order to understand how our four most active groups of traders (FAST / SMALL, FAST / MEDIUM, MEDIUM / SMALL, and MEDIUM / MEDIUM) participate during the pre-opening period, we split traders belonging to each group into three subgroups: those who are always, sometimes, and never active in the pre-opening period, respectively. In Table C.3.1, Panel A shows the results for FAST / SMALL traders, Panel B for FAST / MEDIUM traders, Panel C for MEDIUM / SMALL traders, and Panel D for MEDIUM / MEDIUM traders. We report the total number of observations, the average latency and inventory, the average number of new orders, cancellations, and trades per stock-day, the average trade-to-order and cancellation ratios, the number of messages during the pre-opening period and the continuous session.

C.4 Comparison with an alternative classification scheme

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: HFTs and non-HFTs. In particular, in 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, a trader satisfies the following criteria. First, the trader's end-of-day inventory is no greater than 10% of her trading volume for that stock on that day. Second, the trader's inventory at the end of each minute is no greater than 15% of her trading volume for that stock on that day. Third, the trader's 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 April 2013 only, as there was a change in the definition of server IDs at the beginning of May 2013.

Table C.4.1 presents a summary of trader characteristics based on this classification scheme. In particular, HFTs are characterized by a 4% net inventory at the

Table C.3.1. Traders active during the pre-opening period

The following table shows summary statistics for subgroups of the 12 trader groups based on their activity during pre-opening period. We split all traders into 12 groups on a stock-day basis, as described in Table 4.3, using information about speed and inventory from the same day's continuous session. Afterwards, we split each of the 12 trader groups into 3 subgroups based on their participation in the pre-opening period. We report the total number of observations, the average latency and inventory, the average number of new orders, cancellations, and trades per stock-day, the average trade-to-order and cancellation ratios, the number of messages during the pre-opening period and the continuous session 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/Inventory	Participation in pre-opening	Total stock-days	Latency	Inventory	# of new orders	# of trades	# of cancel	Cancel ratio	Trade-to-order	# of pre-open messages	# of cont. messages
Panel A: Active and non-active traders in FAST / SMALL group											
FAST / SMALL	Total	373,255	0.023	16.4%	182.4	60.7	88.1	49.8%	42.8%	56.1	293.3
	Always	27,249	0.045	17.9%	415.4	219.7	38.1	10.8%	52.8%	180.0	374.9
	Sometimes	201,876	0.025	17.5%	105.6	41.8	43.4	50.8%	46.4%	9.2	171.8
	Never	144,130	0.015	14.5%	246.0	57.1	160.2	55.7%	35.9%	.	448.1
Panel B: Active and non-active traders in FAST / MEDIUM group											
Speed/Inventory	Participation in pre-opening	Total stock-days	Latency	Inventory	# of new orders	# of trades	# of cancel	Cancel ratio	Trade-to-order	# of pre-open messages	# of cont. messages
FAST / MEDIUM	Total	360,667	0.025	66.8%	86.1	39.6	38.3	48.8%	46.3%	22.1	150.4
	Always	13,765	0.068	59.4%	295.8	148.1	21.5	9.2%	50.0%	140.1	237.2
	Sometimes	270,984	0.024	69.0%	82.3	40.7	40.2	49.6%	48.1%	6.3	152.2
	Never	75,918	0.021	60.5%	61.9	16.1	34.3	53.1%	39.1%	.	128.1
Panel C: Active and non-active traders in MEDIUM / SMALL group											
Speed/Inventory	Participation in pre-opening	Total stock-days	Latency	Inventory	# of new orders	# of trades	# of cancel	Cancel ratio	Trade-to-order	# of pre-open messages	# of cont. messages
MEDIUM / SMALL	Total	301,683	9.412	17.0%	41.7	20.5	5.2	22.7%	56.0%	24.2	42.5
	Always	12,116	8.770	18.6%	107.3	51.0	5.9	5.8%	48.9%	54.8	74.6
	Sometimes	237,402	9.681	17.0%	37.2	19.8	5.9	24.7%	57.1%	21.6	39.3
	Never	52,165	8.336	16.6%	47.2	16.6	2.1	17.2%	52.6%	.	49.4
Panel D: Active and non-active traders in MEDIUM / MEDIUM group											
Speed/Inventory	Participation in pre-opening	Total stock-days	Latency	Inventory	# of new orders	# of trades	# of cancel	Cancel ratio	Trade-to-order	# of pre-open messages	# of cont. messages
MEDIUM / MEDIUM	Total	321,336	10.285	65.7%	28.6	14.0	5.1	25.6%	55.1%	17.8	30.9
	Always	6,636	9.985	63.5%	98.5	45.2	4.5	5.2%	47.9%	52.8	63.2
	Sometimes	275,613	10.254	66.2%	27.3	14.1	5.6	27.4%	55.9%	16.3	30.5
	Never	39,087	10.556	62.7%	26.0	8.0	1.8	16.0%	50.4%	.	27.9

end of the day, in contrast to the 16% net inventory of our FAST / SMALL traders. Trade-to-order and cancellation ratios are around 50% and 30% respectively for both the HFT and non-HFT groups. The average latency for HFTs is 2.34 seconds, which is not that different to the typical human reaction time. Based on this classification scheme, we identify 59 traders as HFTs. Each of these traders is active in 10 stocks on average (with a maximum coverage of 73 stocks and a minimum coverage of just 1 stock). However, these HFTs are responsible for only 12% of the total activity during the continuous session, and for 1.5% of the total activity during the pre-opening period. Note that most of the observations are marked as non-HFT, suggesting that the Kirilenko, Kyle, Samadi, and Tuzun (2015) scheme is a stricter (narrower) classification of HFTs than the classification proposed in our paper. We believe that the diversity of market participants in the TSE better suits our more comprehensive approach than the narrower alternative scheme. The low participation rate of HFTs in the case of the Kirilenko, Kyle, Samadi, and Tuzun (2015) scheme suggests that there are other active traders who do not meet the three conditions above.[1]

The Table C.4.2 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 the FAST / SMALL or MEDIUM / SMALL groups. 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. Overall, we believe that the Kirilenko, Kyle, Samadi, and Tuzun (2015) scheme is not appropriate for the TSE market, at least with the current thresholds in place, as it does not properly capture the distinctive features of HFTs as discussed in the SEC (2014) report.

C.5 Best bid and best ask price during pre-opening period

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

[1] In the case of the TSE, the number of listed stocks (1,702 stocks as of April 1, 2013) is much larger than the number in the NASDAQ OMX market studied by Brogaard, Hagströmer, Norden, and Riordan (2015). Moreover, the activity of foreign investors, including some foreign-based HFTs, accounts for about 60% of the total trading volume, according to TSE statistics.

Table C.4.1. 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. We report the total number of observations, the average number of observations per stock-day, the average latency and inventory, the average number of new orders per stock-day, the average trade-to-order and cancellation ratios, the proportion of activity during the pre-opening period and the continuous session, the proportion of total trading activity, and the presence ratio (the proportion of traders that are active during both the pre-opening and continuous sessions). These characteristics are per group on a stock-day basis 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.

	# of obs	Average # of servers	Latency	Inventory	# of orders	Trade to-order ratio	Cancellation ratio	Activity during pre-opening period	Activity during continuous session	Trading activity	Presence ratio
HFT	11,593	5.98	2.34	3.94%	677.76	57.7%	39.7%	1.5%	12.2%	11.5%	14.77%
Non-HFT	1,774,943	914.92	777.06	70.97%	31.69	51.5%	33.4%	98.5%	87.8%	88.5%	26.49%

Table C.4.2. 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 4.3. 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.

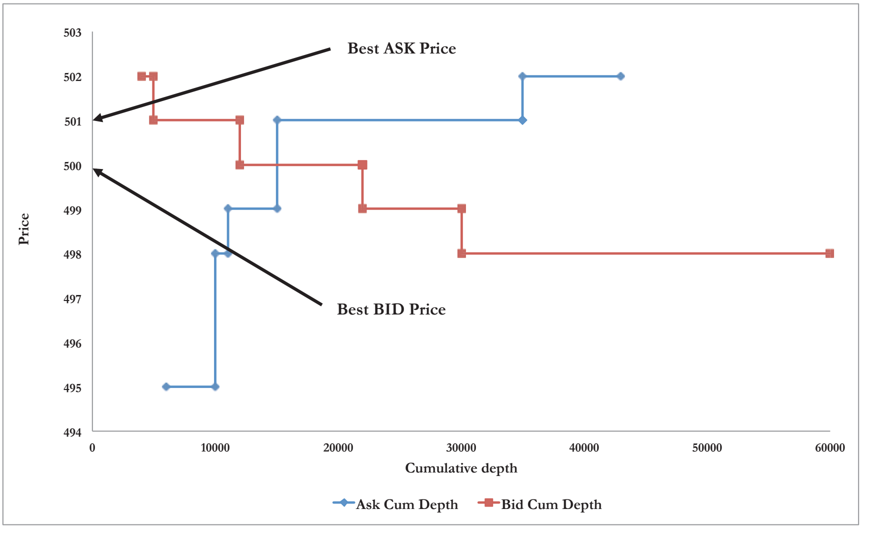
Speed	Inventory	HFT	Non-HFT
FAST	LARGE	13	149,609
	MEDIUM	82	171,029
	SMALL	9,695	166,734
	NOTRADE	1	80,901
MEDIUM	LARGE	4	186,921
	MEDIUM	28	151,315
	SMALL	1,726	141,448
	NOTRADE	6	94,061
SLOW	LARGE	17	409,866
	MEDIUM	0	80,372
	SMALL	17	65,482
	NOTRADE	4	77,205

of eligible buy and sell orders at each price (depth). Usually, more buy orders are accumulated 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 4.3.1. 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

Figure C.5.1. Determination of best bid and ask prices during the pre-opening period

scriptsize
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.



these special situations, the TSE has another rule to determine 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.

References

Acharya, V. V. and L. H. Pedersen (2005). Asset pricing with liquidity risk. *Journal of Financial Economics 77*(2), 375–410.

Agarwal, A. (2012). High frequency trading: Evolution and the future. *Capgemini report*.

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

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics 17*(2), 223–249.

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

Anand, A., S. Chakravarty, and T. Martell (2005). Empirical evidence on the evolution of liquidity: Choice of market versus limit orders by informed and uninformed traders. *Journal of Financial Markets 8*(3), 288–308.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review 93*(1), 36–62.

Andrade, S. C., C. Chang, and M. S. Seasholes (2008). Trading imbalances, predictable reversals, and cross-stock price pressure. *Journal of Financial Economics 88*(2), 406–423.

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

Bae, K.-H., G. A. Karolyi, and R. M. Stulz (2003). A new approach to measuring financial contagion. *Review of Financial Studies 16*(3), 717–763.

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.

Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics 4*(1), 1–30.

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

Baruch, S., M. Panayides, and K. Venkataraman (2016). Informed trading and price discovery before corporate events. *Working paper*.

Bekaert, G., M. Ehrmann, M. Fratzscher, and A. Mehl (2014). The global crisis and equity market contagion. *Journal of Finance 69*(6), 2597–2649.

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*(4), 715–741.

Bernardo, A. E. and I. Welch (2004). Liquidity and financial market runs. *Quarterly Journal of Economics 119*(1), 135–158.

Bharath, S. T., P. Pasquariello, and G. Wu (2009). Does asymmetric information drive capital structure decisions? *Review of Financial Studies 22*(8), 3211–3243.

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.

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

Boehmer, E., K. Fong, and J. Wu (2012). Algorithmic trading and changes in firms' equity capital. *Working paper*.

Boehmer, E., K. Fong, and J. Wu (2015). International evidence on algorithmic trading. *Working paper*.

Bollerslev, T., T. H. Law, and G. Tauchen (2008). Risk, jumps, and diversification. *Journal of Econometrics 144*(1), 234–256.

Bongaerts, D., L. Kong, and M. Van Achter (2016). Trading speed competition: Can the arms race go too far? *Working paper*.

Bongaerts, D. and M. Van Achter (2016). High-frequency trading and market stability. *Working Paper*.

Boudt, K. and M. Petitjean (2014). Intraday liquidity dynamics and news releases around price jumps: Evidence from the DJIA stocks. *Journal of Financial Markets 17*, 121–149.

Brockman, P., D. Y. Chung, and C. Pérignon (2009). Commonality in liquidity: A global perspective. *Journal of Financial and Quantitative Analysis 44*(4), 851–882.

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.

Brogaard, J., T. Hendershott, and R. Riordan (2015). Price discovery without trading: Evidence from limit orders. *Working paper*.

Brogaard, J., R. Riordan, A. Shkilko, and K. Sokolov (2015). High-frequency trading and extreme price movements. *Working paper*.

Brunnermeier, M. K. (2008). Deciphering the liquidity and credit crunch 2007-08. *Journal of Economic Perspectives 23*(1), 77–100.

Brunnermeier, M. K., A. Crockett, C. A. Goodhart, A. Persaud, and H. S. Shin (2009). The fundamental principles of financial regulation. *Geneva Reports on the World Economy 11*.

Brunnermeier, M. K. and L. H. Pedersen (2009). Market liquidity and funding liquidity. *Review of Financial Studies 22*(6), 2201–2238.

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. *Journal of Finance 55*(3), 1339–1365.

Cao, C., O. Hansch, and X. Wang (2009). The information content of an open limit-order book. *Journal of Futures Markets 29*(1), 16–41.

Cenesizoglu, T., G. Dionne, and X. Zhou (2014). Effects of the limit order book on price dynamics. *Working paper*.

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

Cespa, G. and T. Foucault (2014). Illiquidity contagion and liquidity crashes. *Review of Financial Studies 27*(6), 1615–1660.

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.

Chakravarty, S. and C. W. Holden (1995). An integrated model of market and limit orders. *Journal of Financial Intermediation 4*(3), 213–241.

Chan, L. K. and J. Lakonishok (1993). Institutional trades and intraday stock price behavior. *Journal of Financial Economics 33*(2), 173–199.

Chen, Q., I. Goldstein, and W. Jiang (2007). Price informativeness and investment sensitivity to stock price. *Review of Financial Studies 20*(3), 619–650.

Chiang, T. C., B. N. Jeon, and H. Li (2007). Dynamic correlation analysis of financial contagion: Evidence from Asian markets. *Journal of International Money and Finance 26*(7), 1206–1228.

Chiyachantana, C. N., P. K. Jain, C. Jiang, and R. A. Wood (2004). International evidence on institutional trading behavior and price impact. *Journal of Finance 59*(2), 869–898.

Chordia, T., R. Roll, and A. Subrahmanyam (2000). Commonality in liquidity. *Journal of Financial Economics 56*(1), 3–28.

Chordia, T., R. Roll, and A. Subrahmanyam (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics 76*(2), 271–292.

Chordia, T., R. Roll, and A. Subrahmanyam (2008). Liquidity and market efficiency. *Journal of Financial Economics 87*(2), 249–268.

Chordia, T., R. Roll, and A. Subrahmanyam (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics 101*(2), 243–263.

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

Cohen, K. J., S. F. Maier, R. A. Schwartz, and D. K. Whitcomb (1981). Transaction costs, order placement strategy, and existence of the bid-ask spread. *Journal of Political Economy 89*(2), 287–305.

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.

Connolly, R. A. and F. A. Wang (2003). International equity market comovements: Economic fundamentals or contagion? *Pacific-Basin Finance Journal 11*(1), 23–43.

Cont, R., A. Kukanov, and S. Stoikov (2013). The price impact of order book events. *Journal of Financial Econometrics 12*(1), 47–88.

Copeland, T. E. and D. Galai (1983). Information effects on the bid-ask spread. *Journal of Finance 38*(5), 1457–1469.

Cremers, K. M. and J. Mei (2007). Turning over turnover. *Review of Financial Studies 20*(6), 1749–1782.

Degryse, H., F. de Jong, and V. van Kervel (2014). Does order splitting signal uninformed order flow? *Working paper*.

Dempster, G., J. M. Wells, and D. T. Wills (2000). A common-features analysis of Amsterdam and London financial markets during the eighteenth century. *Economic Inquiry 38*(1), 19–33.

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.

Easley, D., S. Hvidkjaer, and M. O'Hara (2002). Is information risk a determinant of asset returns? *Journal of Finance 57*(5), 2185–2221.

Easley, D., N. M. Kiefer, M. O'Hara, and J. B. Paperman (1996). Liquidity, information, and infrequently traded stocks. *Journal of Finance 51*(4), 1405–1436.

Eichengreen, B., A. Mody, M. Nedeljkovic, and L. Sarno (2012). How the subprime crisis went global: Evidence from bank credit default swap spreads. *Journal of International Money and Finance 31*(5), 1299–1318.

Ellul, A., C. W. Holden, P. Jain, and R. Jennings (2003). Determinants of order choice on the New York Stock Exchange. *Working paper*.

Eun, C. S. and S. Shim (1989). International transmission of stock market movements. *Journal of Financial and Quantitative Analysis 24*(2), 241–256.

Fong, K., C. Holden, and C. Trzcinka (2014). What are the best liquidity proxies for global research? *Working paper*.

Forbes, K. J. and R. Rigobon (2002). No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance 57*(5), 2223–2261.

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial markets 2*(2), 99–134.

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

Foucault, T., O. Kadan, and E. Kandel (2005). Limit order book as a market for liquidity. *Review of Financial Studies 18*(4), 1171–1217.

Foucault, T., R. Kozhan, and W. W. Tham (2015). Toxic arbitrage. *Working paper*.

Fricke, D. and A. Gerig (2015). Too fast or too slow? Determining the optimal speed of financial markets. *Working paper*.

Frino, A., V. Mollica, and Z. Zhou (2014). Commonality in liquidity across international borders: Evidence from futures markets. *Journal of Futures Markets 34*(8), 807–818.

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

Gârleanu, N. B. and L. H. Pedersen (2007). Liquidity and risk management. *American Economic Review 97*(2), 193–197.

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

Getter, D. E., M. Jickling, M. Labonte, and E. V. Murphy (2007). Financial crisis? The liquidity crunch of august 2007. *CRS report RL34182*.

Glosten, L. R. (1994). Is the electronic open limit order book inevitable? *Journal of Finance 49*(4), 1127–1161.

Glosten, L. R. and L. E. Harris (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics 21*(1), 123–142.

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 14(1), 71–100.

Goettler, R. L., C. A. Parlour, and U. Rajan (2005). Equilibrium in a dynamic limit order market. *Journal of Finance 60*(5), 2149–2192.

Goettler, R. L., C. A. Parlour, and U. Rajan (2009). Informed traders and limit order markets. *Journal of Financial Economics 93*(1), 67–87.

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

Gorton, G. B. (2009a). Information, liquidity, and the (ongoing) panic of 2007. *American Economic Review 99*(2), 567–572.

Gorton, G. B. (2009b). The subprime panic. *European Financial Management 15*(1), 10–46.

Gromb, D. and D. Vayanos (2002). Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics 66*(2), 361–407.

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

Hamao, Y., R. W. Masulis, and V. Ng (1990). Correlations in price changes and volatility across international stock markets. *Review of Financial Studies 3*(2), 281–307.

Handa, P. and R. A. Schwartz (1996). Limit order trading. *Journal of Finance 51*(5), 1835–1861.

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

Hartmann, P., S. Straetmans, and C. G. De Vries (2004). Asset market linkages in crisis periods. *Review of Economics and Statistics 86*(1), 313–326.

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

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

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 A. J. Menkveld (2014). Price pressures. *Journal of Financial Economics 114*(3), 405–423.

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*.

Ho, T. and H. R. Stoll (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics 9*(1), 47–73.

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

Huang, J. and J. Wang (2009). Liquidity and market crashes. *Review of Financial Studies 22*(7), 2607–2643.

Huang, R. D. and H. R. Stoll (1997). The components of the bid-ask spread: A general approach. *Review of Financial Studies 10*(4), 995–1034.

Irvine, P. J., G. J. Benston, and E. Kandel (2000). Liquidity beyond the inside spread: Measuring and using information in the limit order book. *Working paper*.

Jacod, J. and V. Todorov (2009). Testing for common arrivals of jumps for discretely observed multidimensional processes. *The Annals of Statistics 37*(4), 1792–1838.

Jarnecic, E. and T. H. McInish (1997). An empirical investigation of the option value of the limit order book on the Australian Stock Exchange. *Working paper*.

Jiang, G., I. Lo, and G. Valente (2014). High-frequency trading around macroeconomic news announcements: Evidence from the US treasury market. *Bank of Canada Working Paper 2014-56*.

Jiang, G. J., I. Lo, and A. Verdelhan (2011). Information shocks, liquidity shocks, jumps, and price discovery: Evidence from the US treasury market. *Journal of*

Financial and Quantitative Analysis 46(2), 527–551.

Jiang, G. J. and R. C. Oomen (2008). Testing for jumps when asset prices are observed with noise – a "swap variance" approach. *Journal of Econometrics 144*(2), 352–370.

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*.

Karolyi, G. A. (2003). Does international financial contagion really exist? *International Finance 6*(2), 179–199.

Karolyi, G. A., K.-H. Lee, and M. A. Van Dijk (2012). Understanding commonality in liquidity around the world. *Journal of Financial Economics 105*(1), 82–112.

Kavajecz, K. A. and E. R. Odders-White (2004). Technical analysis and liquidity provision. *Review of Financial Studies 17*(4), 1043–1071.

Kehr, C. H., J. P. Krahnen, and E. Theissen (2001). The anatomy of a call market. *Journal of Financial Intermediation 10*(3), 249–270.

King, M. A. and S. Wadhwani (1990). Transmission of volatility between stock markets. *Review of Financial studies 3*(1), 5–33.

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*.

Kodres, L. E. and M. Pritsker (2002). A rational expectations model of financial contagion. *Journal of Finance 57*(2), 769–799.

Korajczyk, R. A. and R. Sadka (2008). Pricing the commonality across alternative measures of liquidity. *Journal of Financial Economics 87*(1), 45–72.

Kozhan, R., M. Moore, and R. Payne (2014). Market order flows, limit order flows and exchange rate dynamics. *Working paper*.

Kozhan, R. and M. Salmon (2012). The information content of a limit order book: The case of an FX market. *Journal of Financial Markets 15*(1), 1–28.

Kraus, A. and H. R. Stoll (1972). Price impacts of block trading on the New York Stock Exchange. *Journal of Finance 27*(3), 569–588.

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

Kyle, A. S. and W. Xiong (2001). Contagion as a wealth effect. *Journal of Finance 56*(4), 1401–1440.

Lai, S., L. Ng, and B. Zhang (2014). Does PIN affect equity prices around the world? *Journal of Financial Economics 114*(1), 178–195.

Lau, S. T., L. Ng, and B. Zhang (2012). Information environment and equity risk premium volatility around the world. *Management Science 58*(7), 1322–1340.

Lee, C. and M. J. Ready (1991). Inferring trade direction from intraday data. *Journal of Finance 46*(2), 733–746.

Lee, S. S. (2012). Jumps and information flow in financial markets. *Review of Financial Studies 25*(2), 439–479.

Lee, S. S. and P. A. Mykland (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies 21*(6), 2535–2563.

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

Lin, W.-L., R. F. Engle, and T. Ito (1994). Do bulls and bears move across borders? International transmission of stock returns and volatility. *Review of Financial Studies 7*(3), 507–538.

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

Lo, A. W. and J. Wang (2000). Trading volume: Definitions, data analysis, and implications of portfolio theory. *Review of Financial Studies 13*(2), 257–300.

Longstaff, F. A. (2010). The subprime credit crisis and contagion in financial markets. *Journal of Financial Economics 97*(3), 436–450.

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.

Madhavan, A., M. Richardson, and M. Roomans (1997). Why do security prices change? A transaction-level analysis of NYSE stocks. *Review of Financial Studies 10*(4), 1035–1064.

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

Marshall, B. R., N. H. Nguyen, and N. Visaltanachoti (2012). Commodity liquidity measurement and transaction costs. *Review of Financial Studies 25*(2), 599–638.

Marshall, B. R., N. H. Nguyen, and N. Visaltanachoti (2013a). Liquidity commonality in commodities. *Journal of Banking and Finance 37*(1), 11–20.

Marshall, B. R., N. H. Nguyen, and N. Visaltanachoti (2013b). Liquidity measurement in frontier markets. *Journal of International Financial Markets, Institutions and Money 27*, 1–12.

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

Menkhoff, L., C. L. Osler, and M. Schmeling (2010). Limit-order submission strategies under asymmetric information. *Journal of Banking and Finance 34*(11), 2665–2677.

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? Exchange latency and liquidity. *Working paper*.

Morris, S. and H. S. Shin (2004). Liquidity black holes. *Review of Finance 8*(1), 1–18.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica 55*(3), 703–708.

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

Opschoor, A., N. Taylor, M. Van der Wel, and D. van Dijk (2014). Order flow and volatility: An empirical investigation. *Journal of Empirical Finance 28*, 185–201.

Pagnotta, E. and T. Philippon (2015). Competing on speed. *Working paper*.

Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies 11*(4), 789–816.

Pasquariello, P. (2007). Imperfect competition, information heterogeneity, and financial contagion. *Review of Financial Studies 20*(2), 391–426.

Pastor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies 22*(1), 435–480.

Pukthuanthong, K. and R. Roll (2015). Internationally correlated jumps. *Review of Asset Pricing Studies 5*(1), 92–111.

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

Ramchand, L. and R. Susmel (1998). Volatility and cross correlation across major stock markets. *Journal of Empirical Finance 5*(4), 397–416.

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

Reinhart, C. and S. Calvo (1996). Capital flows to Latin America: Is there evidence of contagion effects? *Working paper*.

Roşu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies 22*(11), 4601–4641.

Roşu, I. (2016). Liquidity and information in order driven markets. *Working paper*.

Rodriguez, J. C. (2007). Measuring financial contagion: A copula approach. *Journal of Empirical Finance 14*(3), 401–423.

Roll, R. (1989). Price volatility, international market links, and their implications for regulatory policies. *Journal of Financial Services Research 3*, 211–246.

Rösch, D. M., A. Subrahmanyam, and M. A. van Dijk (2015). An empirical analysis of co-movement in market efficiency measures. *Working paper*.

Rose, A. K. and M. M. Spiegel (2010). Cross-country causes and consequences of the 2008 crisis: International linkages and American exposure. *Pacific Economic*

Review 15(3), 340–363.

Sadka, R. (2006). Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk. *Journal of Financial Economics 80*(2), 309–349.

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.

Seppi, D. J. (1997). Liquidity provision with limit orders and a strategic specialist. *Review of Financial Studies 10*(1), 103–150.

Skeete, H. (Ed.) (2004). *The Handbook of World Stock, Derivative and Commodity Exchanges*. Mondovisione.

Stoll, H. R. (1978). The supply of dealer services in securities markets. *Journal of Finance 33*(4), 1133–1151.

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

Susmel, R. and R. F. Engle (1994). Hourly volatility spillovers between international equity markets. *Journal of International Money and Finance 13*(1), 3–25.

Swan, P. L. and P. J. Westerholm (2006). Market architecture and global exchange efficiency: One design need not fit all stock sizes. *Working paper*.

Tauchen, G. and H. Zhou (2011). Realized jumps on financial markets and predicting credit spreads. *Journal of Econometrics 160*(1), 102–118.

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

Uno, J. and M. Shibata (2012). Speed of trade and liquidity. *Working paper*.

Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics 82*(1), 103–133.

Wuyts, G. (2011). The impact of liquidity shocks through the limit order book. *Working paper*.

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.

Ye, M., C. Yao, and J. Gai (2013). The externalities of high frequency trading. *Working paper*.

Yuan, K. (2005). Asymmetric price movements and borrowing constraints: A rational expectations equilibrium model of crises, contagion, and confusion. *Journal of Finance 60*(1), 379–411.

Zhang, Z., J. Cai, and Y. L. Cheung (2009). Explaining country and cross-border liquidity commonality in international equity markets. *Journal of Futures Markets 29*(7), 630–652.

About the author

Darya Yuferova was born on February 28, 1990 in Novosibirsk, Russia. She obtained her Bachelor degree in Economics from Novosibirsk State University, majoring in Mathematical Methods in Economics. During 2010–2011, Darya did her Master studies in Finance at Duisenberg School of Finance / Free University Amsterdam, majoring in Risk Management. During her master studies, Darya did a "Super Quant" internship in Robeco Asset Management.

In 2011, Darya has started her work on the PhD project "Financial market liquidity: a broad perspective" at Rotterdam School of Management, Erasmus University under supervision of Mathijs van Dijk and Dion Bongaerts. Her main area of research interest is interplay of asset pricing and market microstructure.

During her PhD studies, Darya went on a research visit to NYU Stern School of Business (host: Marti Subrahmanyam). Darya has taken various courses from leading scholars in the field of finance, such as Ekkehart Boehmer, Robert Engle, Thierry Foucault, Harrison Hong, Anthony Lynch, Albert Menkveld, Stijn van Nieuwerburgh, Marco Pagano, Dimitri Vayanos, and Wei Xiong. Her work was presented at several international conferences, such as the Financial Risks International Forum on Scenarios, Stress, and Forecasts in Finance; the Emerging Markets Finance Conference; the Annual Meeting of the German Finance Association; the International Conference on the Industrial Organisation of Securities and Derivatives Markets: High Frequency Trading.

As of August 15, 2016, Darya will be working as an Assistant Professor of Finance in Norwegian School of Economics (NHH), Bergen, Norway.

Portfolio

PhD Courses (2011–2012)

Advanced Econometrics III, Tinbergen Institute

Computational Econometrics, Tinbergen Institute

Corporate Finance Theory, Rotterdam School of Management

Empirical Asset Pricing, Rotterdam School of Management

Interaction Performance Training, Rotterdam School of Management

Market Microstructure, Tinbergen Institute

Portfolio Management, Erasmus School of Economics

Publishing Strategy, Rotterdam School of Management

Seminar Asset Pricing, Rotterdam School of Management

Topics in Philosophy of Science, Rotterdam School of Management

Working Papers

Intraday Return Predictability, Informed Limit Orders, and Algorithmic Trading (single authored)

The Propagation of Shocks Across International Equity Markets: A Microstructure Perspective (with Dion Bongaerts, Richard Roll, Dominik Rösch, and Mathijs van Dijk)

Low-Latency Trading and Price Discovery without Trading: Evidence from the Tokyo Stock Exchange in the Pre-Opening Period and the Opening Batch Auction (with Mario Bellia, Loriana Pelizzon, Marti G. Subrahmanyam, and Jun Uno)

Conferences and Seminars (* presented by a co-author)

2016 (including future presentations): 19th SGF Conference; CONSOB/BAFFI - CAREFIN conference (*); Job Market Seminar at Cornerstone Research; Job Market Seminar at Norwegian School of Economics; Job Market Seminar at SAFE, Goethe University

2015: PhD seminar, Tinbergen Institute; FMA Doctoral Student Consortium; PhD course "Market liquidity" by Thierry Foucault and Marco Pagano; PhD seminar, Rotterdam School of Management, Erasmus University; 8th Financial Risks International Forum on Scenarios, Stress, and Forecasts in Finance (Paris, France); SAFE Microstructure workshop, Goethe University (*); 4th International Conference on the Industrial Organisation of Securities and Derivatives Markets: High Frequency Trading (*); FMA European Conference (*)

2014: PhD seminar, NYU Stern Business School; PhD seminar, Rotterdam School of Management, Erasmus University; 5th Emerging Markets Finance Conference (Mumbai, India); joint conference of the 21st Annual Meeting of the German Finance Association (DGF) and 13th Symposium on Finance, Banking, and Insurance (Karlsruhe, Germany) (*); Extreme Events in Finance (Royaumont, France) (*); INFER workshop on Financial Globalization, International Trade, and Development (Bordeaux, France); PhD seminar Rotterdam School of Management, Erasmus University

Teaching Experience (Rotterdam School of Management, Erasmus University)

2015: Lectures for Alternative investments (Bachelor course)

2013 and 2015: Workshops for Investments (Master course)

2012 – 2015: Master thesis supervision and co-readerships

2012 – 2013: Bachelor thesis supervision

Prizes, Awards, and Scholarships

Grant from EUROFIDAI (EUR 40,000): research proposal "Strategic behavior of high frequency traders during pre-opening period" (with Mario Bellia, Loriana Pelizzon, Marti G. Subrahmanyam, and Jun Uno), 2014

Vereniging Trustfonds Erasmus Universiteit Rotterdam Research Visit Grant (EUR 1,000), 2014

AFA Student Travel Grant (USD 1,500), 2014

Best Student Award, 2011 MSc in Risk Management program, Duisenberg School of Finance

Scholarship, Duisenberg School of Finance (EUR 19,500; 75% of tuition fee)

1st place, International Scientific Students Conference XLVIII, 2010, Novosibirsk, Russia

MDM Bank Scholarship (Sep-2009 – Jun-2010)

Ernst & Young Scholarship (Sep-2009 – Jun-2010)

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

ERIM PH.D. SERIES RESEARCH IN MANAGEMENT

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://repub.eur.nl/pub. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics at the Erasmus University Rotterdam (EUR).

DISSERTATIONS LAST FIVE YEARS

Abbink, E.J., *Crew Management in Passenger Rail Transport*, Promotor(s): Prof.dr. L.G. Kroon & Prof.dr. A.P.M. Wagelmans, EPS-2014-325-LIS, http://repub.eur.nl/ pub/76927

Acar, O.A., *Crowdsourcing for Innovation: Unpacking Motivational, Knowledge and Relational Mechanisms of Innovative Behavior in Crowdsourcing Platforms*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2014-321-LIS, http://repub.eur.nl/pub/76076

Akin Ates, M., *Purchasing and Supply Management at the Purchase Category Level: strategy, structure and performance*, Promotor(s): Prof.dr. J.Y.F. Wynstra & Dr. E.M. van Raaij, EPS-2014-300-LIS, http://repub.eur.nl/pub/50283

Akpinar, E., *Consumer Information Sharing*, Promotor(s): Prof.dr.ir. A. Smidts, EPS- 2013-297-MKT, http://repub.eur.nl/pub/50140

Alexander, L., *People, Politics, and Innovation: A Process Perspective*, Promotor(s): Prof.dr. H.G. Barkema & Prof.dr. D.L. van Knippenberg, EPS-2014-331-S&E, http: //repub.eur.nl/pub/77209

Almeida e Santos Nogueira, R.J. de, *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*, Promotor(s): Prof.dr.ir. U. Kaymak & Prof.dr. J.M.C. Sousa, EPS-2014-310-LIS, http://repub.eur.nl/pub/51560

Bannouh, K., *Measuring and Forecasting Financial Market Volatility using High-frequency Data*, Promotor(s): Prof.dr. D.J.C. van Dijk, EPS-2013-273-F&A, http://repub.eur.nl/pub/38240

Ben-Menahem, S.M., *Strategic Timing and Proactiveness of Organizations*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2013-278-S&E, http://repub.eur.nl/pub/39128

Benning, T.M., *A Consumer Perspective on Flexibility in Health Care: Priority Access Pricing and Customized Care*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2011-241-MKT, http://repub.eur.nl/pub/23670

Benschop, N, *Biases in Project Escalation: Names, frames & construal levels*, Promotors: Prof.dr. K.I.M. Rhode, Prof.dr. H.R. Commandeur, Prof.dr. M.Keil & Dr. A.L.P. Nuijten, EPS-2015-375-S&E, hdl.handle.net/1765/79408

Berg, W.E. van den, *Understanding Salesforce Behavior using Genetic Association Studies*, Promotor(s): Prof.dr. W.J.M.I. Verbeke, EPS-2014-311-MKT, http://repub.eur.nl/pub/51440

Betancourt, N.E., *Typical Atypicality: Formal and Informal Institutional Conformity, Deviance, and Dynamics*, Promotor(s): Prof.dr. B. Krug, EPS-2012-262-ORG, http://repub.eur.nl/pub/32345

Bliek, R. de, *Empirical Studies on the Economic Impact of Trust*, Promotor(s): Prof.dr. J. Veenman & Prof.dr. Ph.H.B.F. Franses, EPS-2015-324-ORG, http://repub.eur.nl/pub/78159

Blitz, D.C., *Benchmarking Benchmarks*, Promotor(s): Prof.dr. A.G.Z. Kemna & Prof.dr. W.F.C. Verschoor, EPS-2011-225-F&A, http://repub.eur.nl/pub/22624

Boons, M., *Working Together Alone in the Online Crowd: The Effects of Social Motivationsand Individual Knowledge Backgrounds on the Participation and Performance of Members of Online Crowdsourcing Platforms*, Promotor(s): Prof.dr. H.G. Barkema & Dr. D.A. Stam, EPS-2014-306-S&E, http://repub.eur.nl/pub/50711

Brazys, J., *Aggregated Marcoeconomic News and Price Discovery*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2015-351-F&A, http://repub.eur.nl/pub/78243

Burger, M.J., *Structure and Cooptition in Urban Networks*, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.R. Commandeur, EPS-2011-243-ORG, http://repub.eur. nl/pub/26178

Byington, E., *Exploring Coworker Relationships: Antecedents and Dimensions of Interpersonal Fit,Coworker Satisfaction, and Relational Models*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-292-ORG, http://repub.eur.nl/pub/41508

Camacho, N.M., *Health and Marketing: Essays on Physician and Patient Decision- Making*, Promotor(s): Prof.dr. S. Stremersch, EPS-2011-237-MKT, http://repub.eur.nl/pub/23604

Cancurtaran, P., *Essays on Accelerated Product Development*, Promotor(s): Prof.dr. F. Langerak & Prof.dr.ir. G.H. van Bruggen, EPS-2014-317-MKT, http://repub.eur.nl/pub/76074

Caron, E.A.M., *Explanation of Exceptional Values in Multi-dimensional Business Databases*, Promotor(s): Prof.dr.ir. H.A.M. Daniels & Prof.dr. G.W.J. Hendrikse, EPS-2013-296-LIS, http://repub.eur.nl/pub/50005

Carvalho, L. de, *Knowledge Locations in Cities: Emergence and Development Dynamics*, Promotor(s): Prof.dr. L. Berg, EPS-2013-274-S&E, http://repub.eur.nl/pub/38449

Cox, R.H.G.M., *To Own, To Finance, and To Insure - Residential Real Estate Revealed*, Promotor(s): Prof.dr. D. Brounen, EPS-2013-290-F&A, http://repub.eur.nl/pub/40964

Deichmann, D., *Idea Management: Perspectives from Leadership, Learning, and Network Theory*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2012-255-ORG, http://repub.eur.nl/pub/31174

Deng, W., *Social Capital and Diversification of Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2015-341-ORG, http://repub.eur.nl/pub/77449

Desmet, P.T.M., *In Money we Trust? Trust Repair and the Psychology of Financial Compensations*, Promotor(s): Prof.dr. D. de Cremer, EPS-2011-232-ORG, http://repub.eur.nl/pub/23268

Dollevoet, T.A.B., *Delay Management and Dispatching in Railways*, Promotor(s): Prof.dr. A.P.M. Wagelmans, EPS-2013-272-LIS, http://repub.eur.nl/pub/38241

Doorn, S. van, *Managing Entrepreneurial Orientation*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch, & Prof.dr. H.W. Volberda, EPS-2012-258- STR, http://repub.eur.nl/pub/32166

Douwens-Zonneveld, M.G., *Animal Spirits and Extreme Confidence: No Guts, No Glory?* Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2012-257-F&A, http://repub.eur.nl/pub/31914

Duca, E., *The Impact of Investor Demand on Security Offerings*, Promotor(s): Prof.dr. A. de Jong, EPS-2011-240-F&A, http://repub.eur.nl/pub/26041

Duyvesteyn, J.G., *Empirical Studies on Sovereign Fixed Income Markets*, Promotor(s): Prof.dr P.Verwijmeren & Prof.dr. M.P.E. Martens, EPS-2015-361-F&A, hdl.handle.net/1765/79033

Duursema, H., *Strategic Leadership: Moving Beyond the Leader-Follower Dyad*, Promotor(s): Prof.dr. R.J.M. van Tulder, EPS-2013-279-ORG, http://repub.eur.nl/pub/39129

Eck, N.J. van, *Methodological Advances in Bibliometric Mapping of Science*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2011-247-LIS, http://repub.eur.nl/pub/26509

Elemes, A., *Studies on Determinants and Consequences of Financial Reporting Quality*, Promotor: Prof.dr. E.Peek, EPS-2015-354-F&A, http://hdl.handle.net/1765/79037

Ellen, S. ter, *Measurement, Dynamics, and Implications of Heterogeneous Beliefs in Financial Markets*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2015-343-F&A, http://repub.eur.nl/pub/78191

Eskenazi, P.I., *The Accountable Animal*, Promotor(s): Prof.dr. F.G.H. Hartmann, EPS- 2015-355-F&A, http://repub.eur.nl/pub/78300

Essen, M. van, *An Institution-Based View of Ownership*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. G.M.H. Mertens, EPS-2011-226-ORG, http://repub.eur.nl/pub/22643

Evangelidis, I., *Preference Construction under Prominence*, Promotor(s): Prof.dr. S.M.J. van Osselaer, EPS-2015-340-MKT, http://repub.eur.nl/pub/78202

Faber, N., *Structuring Warehouse Management*, Promotor(s): Prof.dr. MB.M. de Koster, Prof.dr. Ale Smidts, EPS-2015-336-LIS, http://repub.eur.nl/pub/78603

Fernald, K., *The Waves of Biotechnological Innovation in Medicine: Interfirm Cooperation Effects and a Venture Capital Perspective*, Promotor(s): Prof.dr. E.Claassen, Prof.dr. H.P.G.Pennings & Prof.dr. H.R. Commandeur, EPS-2015-371-S&E, http://hdl.handle.net/1765/79120

Fourne, S.P., *Managing Organizational Tensions: A Multi-Level Perspective on Exploration, Exploitation and Ambidexterity*, Promotor(s): Prof.dr. J.J.P. Jansen & Prof.dr. S.J. Magala, EPS-2014-318-S&E, http://repub.eur.nl/pub/76075

Gharehgozli, A.H., *Developing New Methods for Efficient Container Stacking Operations*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2012-269-LIS, http://repub.eur.nl/pub/37779

Gils, S. van, *Morality in Interactions: On the Display of Moral Behavior by Leaders and Employees*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2012-270-ORG, http://repub.eur.nl/pub/38027

Ginkel-Bieshaar, M.N.G. van, *The Impact of Abstract versus Concrete Product Communications on Consumer Decision-making Processes*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2012-256-MKT, http://repub.eur.nl/pub/31913

Gkougkousi, X., *Empirical Studies in Financial Accounting*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. E. Peek, EPS-2012-264-F&A, http://repub.eur.nl/pub/37170

Glorie, K.M., *Clearing Barter Exchange Markets: Kidney Exchange and Beyond*, Promotor(s): Prof.dr. A.P.M. Wagelmans & Prof.dr. J.J. van de Klundert, EPS-2014-329-LIS, http://repub.eur.nl/pub/77183

Hekimoglu, M., *Spare Parts Management of Aging Capital Products*, Promotor: Prof.dr.ir. R. Dekker, EPS-2015-368-LIS, http://hdl.handle.net/1765/79092

Heij, C.V., *Innovating beyond Technology. Studies on how management innovation, co-creation and business model innovation contribute to firm's (innovation) performance*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-370-STR, http://repub.eur.nl/pub/78651

Heyde Fernandes, D. von der, *The Functions and Dysfunctions of Reminders*, Promotor(s): Prof.dr. S.M.J. van Osselaer, EPS-2013-295-MKT, http://repub.eur.nl/pub/41514

Heyden, M.L.M., *Essays on Upper Echelons & Strategic Renewal: A Multilevel Contingency Approach*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-259-STR, http://repub.eur.nl/pub/32167

Hoever, I.J., *Diversity and Creativity*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2012-267-ORG, http://repub.eur.nl/pub/37392

Hogenboom, A.C., *Sentiment Analysis of Text Guided by Semantics and Structure*, Promotor(s):Prof.dr.ir. U.Kaymak & Prof.dr. F.M.G. de Jong, EPS-2015-369-LIS, http://hdl.handle.net/1765/79034

Hogenboom, F.P., *Automated Detection of Financial Events in News Text*, Promotor(s): Prof.dr.ir. U. Kaymak & Prof.dr. F.M.G. de Jong, EPS-2014-326-LIS, http://repub.eur.nl/pub/77237

Hollen, R.M.A., *Exploratory Studies into Strategies to Enhance Innovation-Driven International Competitiveness in a Port Context: Toward Ambidextrous Ports*, Promotor(s) Prof.dr.ing. F.A.J. Van Den Bosch & Prof.dr. H.W.Volberda, EPS-2015-372-S&E, hdl.handle.net/1765/78881

Hoogendoorn, B., *Social Entrepreneurship in the Modern Economy: Warm Glow, Cold Feet*, Promotor(s): Prof.dr. H.P.G. Pennings & Prof.dr. A.R. Thurik, EPS-2011-246-STR, http://repub.eur.nl/pub/26447

Hoogervorst, N., *On The Psychology of Displaying Ethical Leadership: A Behavioral Ethics Approach*, Promotor(s): Prof.dr. D. de Cremer & Dr. M. van Dijke, EPS-2011- 244-ORG, http://repub.eur.nl/pub/26228

Hout, D.H. van, *Measuring Meaningful Differences: Sensory Testing Based Decision Making in an Industrial Context; Applications of Signal Detection Theory and Thurstonian Modelling*, Promotor(s): Prof.dr. P.J.F. Groenen & Prof.dr. G.B. Dijksterhuis, EPS- 2014-304-MKT, http://repub.eur.nl/pub/50387

Houwelingen, G.G. van, *Something To Rely On*, Promotor(s): Prof.dr. D. de Cremer & Prof.dr. M.H. van Dijke, EPS-2014-335-ORG, http://repub.eur.nl/pub/77320

Hurk, E. van der, *Passengers, Information, and Disruptions*, Promotor(s): Prof.dr. L.G. Kroon & Prof.mr.dr. P.H.M. Vervest, EPS-2015-345-LIS, http://repub.eur.nl/pub/78275

Hytonen, K.A., *Context Effects in Valuation, Judgment and Choice: A Neuroscientific Approach*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2011-252-MKT, http://repub.eur.nl/pub/30668

Iseger, P. den, *Fourier and Laplace Transform Inversion with Applications in Finance*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2014-322-LIS, http://repub.eur.nl/pub/76954

Jaarsveld, W.L. van, *Maintenance Centered Service Parts Inventory Control*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2013-288-LIS, http://repub.eur.nl/pub/39933

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2011-222-LIS, http://repub.eur.nl/pub/22156

Kappe, E.R., *The Effectiveness of Pharmaceutical Marketing*, Promotor(s): Prof.dr. S. Stremersch, EPS-2011-239-MKT, http://repub.eur.nl/pub/23610

Karreman, B., *Financial Services and Emerging Markets*, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.P.G. Pennings, EPS-2011-223-ORG, http://repub.eur.nl/pub/22280

Khanagha, S., *Dynamic Capabilities for Managing Emerging Technologies*, Promotor(s): Prof.dr. H.W. Volberda, EPS-2014-339-S&E, http://repub.eur.nl/pub/77319

Kil, J., *Acquisitions Through a Behavioral and Real Options Lens*, Promotor(s): Prof.dr. H.T.J. Smit, EPS-2013-298-F&A, http://repub.eur.nl/pub/50142

Klooster, E. van 't, *Travel to Learn: the Influence of Cultural Distance on Competence Development in Educational Travel*, Promotor(s): Prof.dr. F.M. Go & Prof.dr. P.J. van Baalen, EPS-2014-312-MKT, http://repub.eur.nl/pub/51462

Koendjbiharie, S.R., *The Information-Based View on Business Network Performance: Revealing the Performance of Interorganizational Networks*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.mr.dr. P.H.M. Vervest, EPS-2014-315-LIS, http://repub.eur.nl/pub/51751

Koning, M., *The Financial Reporting Environment: The Role of the Media, Regulators and Auditors*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. P.G.J. Roosenboom, EPS-2014-330-F&A, http://repub.eur.nl/pub/77154

Konter, D.J., *Crossing Borders with HRM: An Inquiry of the Influence of Contextual Differences in the Adoption and Effectiveness of HRM*, Promotor(s): Prof.dr. J. Paauwe & Dr. L.H. Hoeksema, EPS-2014-305-ORG, http://repub.eur.nl/pub/50388

Korkmaz, E., *Bridging Models and Business: Understanding Heterogeneity in Hidden Drivers of Customer Purchase Behavior*, Promotor(s): Prof.dr. S.L. van de Velde & Prof.dr. D. Fok, EPS-2014-316-LIS, http://repub.eur.nl/pub/76008

Kroezen, J.J., *The Renewal of Mature Industries: An Examination of the Revival of the Dutch Beer Brewing Industry*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens, EPS-2014- 333-S&E, http://repub.eur.nl/pub/77042

Kysucky, V., *Access to Finance in a Cros-Country Context*, Promotor(s): Prof.dr. L. Norden, EPS-2015-350-F&A, http://repub.eur.nl/pub/78225

Lam, K.Y., *Reliability and Rankings*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS- 2011-230-MKT, http://repub.eur.nl/pub/22977

Lander, M.W., *Profits or Professionalism? On Designing Professional Service Firms*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2012-253- ORG, http://repub.eur.nl/pub/30682

Langhe, B. de, *Contingencies: Learning Numerical and Emotional Associations in an Uncertain World*, Promotor(s): Prof.dr.ir. B. Wierenga & Prof.dr. S.M.J. van Osselaer, EPS-2011-236-MKT, http://repub.eur.nl/pub/23504

Legault-Tremblay, P.O., *Corporate Governance During Market Transition: Heterogeneous responses to Institution Tensions in China*, Promotor: Prof.dr. B. Krug, EPS-2015-362-ORG, http://repub.eur.nl/pub/78649

Lenoir, A.S. *Are You Talking to Me? Addressing Consumers in a Globalised World*, Promotor(s) Prof.dr. S. Puntoni & Prof.dr. S.M.J. van Osselaer, EPS-2015-363-MKT, , http://hdl.handle.net/1765/79036

Leunissen, J.M.,*All Apologies: On the Willingness of Perpetrators to Apologize*, Promotor(s): Prof.dr. D. de Cremer & Dr. M. van Dijke, EPS-2014-301-ORG, http://repub.eur.nl/pub/50318

Li, D., *Supply Chain Contracting for After-sales Service and Product Support*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2015-347-LIS, http://repub.eur.nl/pub/78526

Li, Z., *Irrationality: What, Why and How*, Promotor(s): Prof.dr. H. Bleichrodt, Prof.dr. P.P. Wakker, & Prof.dr. K.I.M. Rohde, EPS-2014-338-MKT, http://repub.eur.nl/pub/77205

Liang, Q.X., *Governance, CEO Identity, and Quality Provision of Farmer Cooperatives*, Promotor(s): Prof.dr. G.W.J. Hendrikse, EPS-2013-281-ORG, http://repub.eur.nl/pub/39253

Liket, K., *Why 'Doing Good' is not Good Enough: Essays on Social Impact Measurement*, Promotor(s): Prof.dr. H.R. Commandeur & Dr. K.E.H. Maas, EPS-2014-307-STR, http://repub.eur.nl/pub/51130

Loos, M.J.H.M. van der, *Molecular Genetics and Hormones: New Frontiers in Entrepreneurship Research*, Promotor(s): Prof.dr. A.R. Thurik, Prof.dr. P.J.F. Groenen, & Prof.dr. A. Hofman, EPS-2013-287-S&E, http://repub.eur.nl/pub/40081

Lovric, M., *Behavioral Finance and Agent-Based Artificial Markets*, Promotor(s): Prof.dr. J. Spronk & Prof.dr.ir. U. Kaymak, EPS-2011-229-F&A, http://repub.eur.nl/pub/22814

Lu, Y., *Data-Driven Decision Making in Auction Markets*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.dr. W. Ketter, EPS-2014-314-LIS, http://repub.eur.nl/pub/51543

Manders, B., *Implementation and Impact of ISO 9001*, Promotor(s): Prof.dr. K. Blind, EPS-2014-337-LIS, http://repub.eur.nl/pub/77412

Markwat, T.D., *Extreme Dependence in Asset Markets Around the Globe*, Promotor(s): Prof.dr. D.J.C. van Dijk, EPS-2011-227-F&A, http://repub.eur.nl/pub/22744

Mees, H., *Changing Fortunes: How China's Boom Caused the Financial Crisis*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2012-266-MKT, http://repub.eur.nl/pub/34930

Mell, J.N., *Connecting Minds: On The Role of Metaknowledge in Knowledge Coordination*, Promotor: Prof.dr.D.L. van Knippenberg, EPS-2015-359-ORG, http://hdl.handle.net/1765/78951

Meuer, J., *Configurations of Inter-firm Relations in Management Innovation: A Study in China's Biopharmaceutical Industry*, Promotor(s): Prof.dr. B. Krug, EPS-2011-228-ORG, http://repub.eur.nl/pub/22745

Micheli, M.R., *Business Model Innovation: A Journey across Managers' Attention and Inter-Organizational Networks*, Promotor(s): Prof.dr. J.J.P. Jansen, EPS-2015-344-S&E, http://repub.eur.nl/pub/78241

Mihalache, O.R., *Stimulating Firm Innovativeness: Probing the Interrelations between Managerial and Organizational Determinants*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch, & Prof.dr. H.W. Volberda, EPS-2012-260-S&E, http://repub.eur.nl/pub/32343

Milea, V., *News Analytics for Financial Decision Support*, Promotor(s): Prof.dr.ir. U. Kaymak, EPS-2013-275-LIS, http://repub.eur.nl/pub/38673

Naumovska, I., *Socially Situated Financial Markets: A Neo-Behavioral Perspective on Firms, Investors and Practices*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. A. de Jong, EPS-2014-319-S&E, http://repub.eur.nl/pub/76084

Nielsen, L.K., *Rolling Stock Rescheduling in Passenger Railways: Applications in short term planning and in disruption management*, Promotor(s): Prof.dr. L.G. Kroon, EPS- 2011-224-LIS, http://repub.eur.nl/pub/22444

Nuijten, A.L.P., *Deaf Effect for Risk Warnings: A Causal Examination applied to Information Systems Projects*, Promotor(s): Prof.dr. G.J. van der Pijl, Prof.dr. H.R. Commandeur & Prof.dr. M. Keil, EPS-2012-263-S&E, http://repub.eur.nl/pub/34928

Osadchiy, S.E., *The Dynamics of Formal Organization: Essays on bureaucracy and formal rules*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens, EPS-2011-231-ORG, http://repub.eur.nl/pub/23250

Ozdemir, M.N., *Project-level Governance, Monetary Incentives, and Performance in Strategic R&D Alliances*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende, EPS-2011-235-LIS, http://repub.eur.nl/pub/23550

Peers, Y., *Econometric Advances in Diffusion Models*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-251-MKT, http://repub.eur.nl/pub/30586

Peters, M., *Machine Learning Algorithms for Smart Electricity Markets*, Promotor(s): Prof.dr. W. Ketter, EPS-2014-332-LIS, http://repub.eur.nl/pub/77413

Porck, J., *No Team is an Island: An Integrative View of Strategic Consensus between Groups*, Promotor(s): Prof.dr. P.J.F. Groenen & Prof.dr. D.L. van Knippenberg, EPS- 2013-299-ORG, http://repub.eur.nl/pub/50141

Porras Prado, M., *The Long and Short Side of Real Estate, Real Estate Stocks, and Equity*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2012-254-F&A, http://repub.eur.nl/pub/30848

Poruthiyil, P.V., *Steering Through: How organizations negotiate permanent uncertainty and unresolvable choices*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. S.J. Magala, EPS-2011-245-ORG, http://repub.eur.nl/pub/26392

Pourakbar, M., *End-of-Life Inventory Decisions of Service Parts*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2011-249-LIS, http://repub.eur.nl/pub/30584

Pronker, E.S., *Innovation Paradox in Vaccine Target Selection*, Promotor(s): Prof.dr. H.J.H.M. Claassen & Prof.dr. H.R. Commandeur, EPS-2013-282-S&E, http://repub.eur.nl/pub/39654

Protzner, S., *Mind the gap between demand and supply: A behavioral perspective on demand forecasting*, Promotor(s): Prof.dr. S.L. van de Velde & Dr. L. Rook, EPS-2015-364-LIS, http://repub.eur.nl/pub/79355

Pruijssers, J.K., *An Organizational Perspective on Auditor Conduct*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2015-342-S&E, http://repub.eur.nl/pub/78192

Retel Helmrich, M.J., *Green Lot-Sizing*, Promotor(s): Prof.dr. A.P.M. Wagelmans, EPS- 2013-291-LIS, http://repub.eur.nl/pub/41330

Rietveld, N., *Essays on the Intersection of Economics and Biology*, Promotor(s): Prof.dr. A.R. Thurik, Prof.dr. Ph.D. Koellinger, Prof.dr. P.J.F. Groenen, & Prof.dr. A. Hofman, EPS-2014-320-S&E, http://repub.eur.nl/pub/76907

Rijsenbilt, J.A., *CEO Narcissism: Measurement and Impact*, Promotor(s): Prof.dr. A.G.Z. Kemna & Prof.dr. H.R. Commandeur, EPS-2011-238-STR, http://repub.eur.nl/pub/23554

Rösch, D., *Market Efficiency and Liquidity*, Promotor: Prof.dr. M.A. van Dijk, EPS-2015-353-F&A, http://hdl.handle.net/1765/79121

Roza-van Vuren, M.W., *The Relationship between Offshoring Strategies and Firm Performance: Impact of innovation, absorptive capacity and firm size*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2011-214-STR, http://repub.eur.nl/pub/22155

Rubbaniy, G., *Investment Behaviour of Institutional Investors*, Promotor(s): Prof.dr. W.F.C. Verschoor, EPS-2013-284-F&A, http://repub.eur.nl/pub/40068

Schoonees, P. *Methods for Modelling Response Styles*, Promotor: Prof.dr P.J.F. Groenen, EPS-2015-348-MKT, http://repub.eur.nl/pub/79327

Shahzad, K., *Credit Rating Agencies, Financial Regulations and the Capital Markets*, Promotor(s): Prof.dr. G.M.H. Mertens, EPS-2013-283-F&A, http://repub.eur.nl/pub/39655

Sousa, M.J.C. de, *Servant Leadership to the Test: New Perspectives and Insights*, Promotor(s): Prof.dr. D.L. van Knippenberg & Dr. D. van Dierendonck, EPS-2014-313-ORG, http://repub.eur.nl/pub/51537

Spliet, R., *Vehicle Routing with Uncertain Demand*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2013-293-LIS, http://repub.eur.nl/pub/41513

Staadt, J.L., *Leading Public Housing Organisation in a Problematic Situation: A Critical Soft Systems Methodology Approach*, Promotor(s): Prof.dr. S.J. Magala, EPS-2014-308- ORG, http://repub.eur.nl/pub/50712

Stallen, M., *Social Context Effects on Decision-Making: A Neurobiological Approach*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2013-285-MKT, http://repub.eur.nl/pub/39931

Tarakci, M., *Behavioral Strategy: Strategic Consensus, Power and Networks*, Promotor(s): Prof.dr. D.L. van Knippenberg & Prof.dr. P.J.F. Groenen, EPS-2013-280-ORG, http://repub.eur.nl/pub/39130

Teixeira de Vasconcelos, M., *Agency Costs, Firm Value, and Corporate Investment*, Promotor(s): Prof.dr. P.G.J. Roosenboom, EPS-2012-265-F&A, http://repub.eur.nl/pub/37265

Troster, C., *Nationality Heterogeneity and Interpersonal Relationships at Work*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2011-233-ORG, http://repub.eur.nl/pub/23298

Tsekouras, D., *No Pain No Gain: The Beneficial Role of Consumer Effort in Decision- Making*, Promotor(s): Prof.dr.ir. B.G.C. Dellaert, EPS-2012-268-MKT, http://repub.eur.nl/pub/37542

Tuijl, E. van, *Upgrading across Organisational and Geographical Configurations*, Promotor(s): Prof.dr. L. van den Berg, EPS-2015-349-S&E, http://repub.eur.nl/pub/78224

Tuncdogan, A., *Decision Making and Behavioral Strategy: The Role of Regulatory Focus in Corporate Innovation Processes*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch, Prof.dr. H.W. Volberda, & Prof.dr. T.J.M. Mom, EPS-2014-334-S&E, http://repub.eur.nl/pub/76978

Uijl, S. den, *The Emergence of De-facto Standards*, Promotor(s): Prof.dr. K. Blind, EPS-2014-328-LIS, http://repub.eur.nl/pub/77382

Vagias, D., Liquidity, *Investors and International Capital Markets*, Promotor(s): Prof.dr. M.A. van Dijk, EPS-2013-294-F&A, http://repub.eur.nl/pub/41511

Veelenturf, L.P., *Disruption Management in Passenger Railways: Models for Timetable, Rolling Stock and Crew Rescheduling*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2014-327- LIS, http://repub.eur.nl/pub/77155

Venus, M., *Demystifying Visionary Leadership: In search of the essence of effective vision communication*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013-289- ORG, http://repub.eur.nl/pub/40079

Vermeer, W., *Propagation in Networks:The impact of information processing at the actor level on system-wide propagation dynamics*, Promotor: Prof.mr.dr. P.H.M.Vervest, EPS-2015-373-LIS, http://repub.eur.nl/pub/79325

Visser, V.A., *Leader Affect and Leadership Effectiveness:How leader affective displays influence follower outcomes*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2013- 286-ORG, http://repub.eur.nl/pub/40076

Vlam, A.J., *Customer First? The Relationship between Advisors and Consumers of Financial Products*, Promotor(s): Prof.dr. Ph.H.B.F. Franses, EPS-2011-250-MKT, http://repub.eur.nl/pub/30585

Waltman, L., *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*, Promotor(s): Prof.dr.ir. R. Dekker & Prof.dr.ir. U. Kaymak, EPS-2011-248- LIS, http://repub.eur.nl/pub/26564

Wang, T., *Essays in Banking and Corporate Finance*, Promotor(s): Prof.dr. L. Norden & Prof.dr. P.G.J. Roosenboom, EPS-2015-352-F&A, http://repub.eur.nl/pub/78301

Wang, Y., *Information Content of Mutual Fund Portfolio Disclosure*, Promotor(s): Prof.dr. M.J.C.M. Verbeek, EPS-2011-242-F&A, http://repub.eur.nl/pub/26066

Wang, Y., *Corporate Reputation Management: Reaching Out to Financial Stakeholders*, Promotor(s): Prof.dr. C.B.M. van Riel, EPS-2013-271-ORG, http://repub.eur.nl/pub/38675

Weenen, T.C., *On the Origin and Development of the Medical Nutrition Industry*, Promotor(s): Prof.dr. H.R. Commandeur & Prof.dr. H.J.H.M. Claassen, EPS-2014-309-S&E, http://repub.eur.nl/pub/51134

Wolfswinkel, M., *Corporate Governance, Firm Risk and Shareholder Value*, Promotor(s): Prof.dr. A. de Jong, EPS-2013-277-F&A, http://repub.eur.nl/pub/39127

Yang, S., *Information Aggregation Efficiency of Prediction Markets*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck, EPS-2014-323-LIS, http://repub.eur.nl/pub/77184

Zaerpour, N., *Efficient Management of Compact Storage Systems*, Promotor(s): Prof.dr.ir. M.B.M. de Koster, EPS-2013-276-LIS, http://repub.eur.nl/pub/38766

Zhang, D., *Essays in Executive Compensation*, Promotor(s): Prof.dr. I. Dittmann, EPS- 2012-261-F&A, http://repub.eur.nl/pub/32344

Zwan, P.W. van der, *The Entrepreneurial Process: An International Analysis of Entry and Exit*, Promotor(s): Prof.dr. A.R. Thurik & Prof.dr. P.J.F. Groenen, EPS-2011-234-ORG, http://repub.eur.nl/pub/23422

Erasmus Research Institute of Management - ERiM

PRICE DISCOVERY, LIQUIDITY PROVISION, AND LOW-LATENCY TRADING

This dissertation consists of three empirical papers in the field of market microstructure. These papers investigate the impact of increased interconnectedness of the financial markets and the vast trading speed improvements on two important functions of financial markets: price discovery and liquidity provision.

Chapter 2 examines the role of liquidity and trading activity in the origination and propagation of shocks to prices across international equity markets. The findings show that equity markets are strongly interconnected with respect to the transmission of shocks to prices and trading activity, while liquidity dry-ups seem to be isolated events.

Chapter 3 analyzes the choice that an informed trader makes between market (consuming liquidity) and limit (supplying liquidity) orders and how this choice is affected by the rise of algorithmic trading. My findings suggest that price informativeness does not necessarily come at the expense of low liquidity, even in the presence of algorithmic traders.

Chapter 4 examines whether low-latency traders are improving or deteriorating price discovery in the pre-opening period. The findings show that low-latency traders actively participate in the pre-opening period despite the absence of immediate execution, although to a lesser extent than during the continuous trading session. Furthermore, low-latency traders lead price discovery during the pre-opening period. Overall, my current and future research agenda is related to the design of the well-functioning financial market with specific focus on the regulation of the low-latency traders.

Design & layout: B&T Ontwerp en advies (www.b-en-t.nl) Print: Haveka (www.haveka.nl)

ERiM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERIM PhD Series

Research in Management

Erasmus Research Institute of Management - ERiM
Rotterdam School of Management (RSM)
Erasmus School of Economics (ESE)
Erasmus University Rotterdam (EUR)
P.O. Box 1738, 3000 DR Rotterdam,
The Netherlands

Tel. +31 10 408 11 82
Fax +31 10 408 96 40
E-mail info@erim.eur.nl
Internet www.erim.eur.nl