TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Milla Siikanen

Investors, Information Arrivals, and Market Liquidity

Empirical Evidence from Financial Markets

Julkaisu 1581 • Publication 1581

Tampere 2018

Tampereen teknillinen yliopisto. Julkaisu 1581
Tampere University of Technology. Publication 1581

Milla Siikanen

Investors, Information Arrivals, and Market Liquidity

Empirical Evidence from Financial Markets

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Rakennustalo Building, Auditorium RG202, at Tampere University of Technology, on the 19th of October 2018, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2018

Doctoral candidate: Milla Siikanen
Industrial and Information Management
Faculty of Business and Built Environment
Tampere University of Technology
Finland

Supervisor: Juho Kanniainen, Professor
Industrial and Information Management
Faculty of Business and Built Environment
Tampere University of Technology
Finland

Pre-examiners: Michael McAleer, University Research Chair Professor
Department of Finance
Asia University
Taiwan

Seppo Pynnönen, Professor
Department of Mathematics and Statistics
University of Vaasa
Finland

Opponent: Pekka Malo, Assistant Professor
Department of Information and Service Economy
Aalto University
Finland

ISBN 978-952-15-4211-4 (printed)
ISBN 978-952-15-4231-2 (PDF)
ISSN 1459-2045

Abstract

Tampere University of Technology

Siikanen Milla: INVESTORS, INFORMATION ARRIVALS, AND MARKET LIQUIDITY: Empirical Evidence from Financial Markets

Key words: Stock market, FX market, Investor behavior, Investor sophistication, Decision making, Social media, Company announcement, Liquidity, Limit order book, Liquidity aggregator, Transaction costs, High-frequency data

Well-functioning financial markets can be argued to benefit society widely. Investors, information arrivals, and market liquidity are all key aspects of financial markets. Without investors who trade, there would be no markets to begin with. Furthermore, information arrivals are important because information drives prices: new information may affect the valuation of assets traded. Finally, for prices to adjust efficiently to new information, the market needs to be sufficiently liquid, meaning that investors can trade when they want at a low transaction cost.

Earlier research on these topics exists, but the interrelations between these factors have not been studied in depth. The objective of this thesis is to improve our knowledge of the interrelations between investors, information arrivals, and liquidity in the context of financial markets. By addressing several research gaps related to these themes, this thesis aims to provide new empirical evidence in order to help the scientific community develop more reliable and robust models that describe the markets; in general, a better understanding of these topics and such interrelations may help improve regulations, exchange organizations, and investment management.

This thesis consists of an introductory part and four research papers (Articles I−IV). Article I uses logistic regression to study how Nokia's Facebooks posts and related activities are associated with investors' decisions to buy versus sell Nokia stock. In Article II, a framework from event studies is combined with high-frequency limit order book data to examine how liquidity in stock limit order books evolves around scheduled and non-scheduled company announcements. Article III applies regression analysis to identify the factors affecting the magnitude of order book liquidity shocks that company announcement releases cause in the limit order books. Finally, Article IV uses a unique data set to study the proportion of liquidity streams that a trader observes in a foreign exchange (FX) liquidity aggregator, as well as quantifies a trader's theoretical improvements in the observed spread and the cost savings when comparing the current situation with the optimal combination of streams; the optimal combinations are solved using a genetic algorithm (GA).

Earlier literature has studied how news articles affect the trading of different investors, and this thesis contributes by providing evidence that the (potentially biased) information a company releases on social media affects the behaviors of different investors in the stock market differently. While the decisions of arguably less sophisticated investors—passive households and non-profit organizations—are associated with Facebook data, those of more sophisticated investors—financial institutions—seem to be independent of Facebook data.

Moreover, company announcements are found to cause significant changes in the stock limit order book liquidity, which is inconsistent with the finding of an earlier study using news data. In particular, scheduled announcement releases may improve liquidity to an abnormally high level, indicating that scheduled announcement releases resolve asymmetric information problems in the market, whereas the order book liquidity remains relatively low in many cases still an hour after the non-scheduled announcement releases. The immediate liquidity shocks following the announcement releases are amplified by order book asymmetry prior to the announcements releases. Moreover, a fast reaction is a strong reaction (the faster the illiquidity peak is reached after the announcement release, the larger the peak usually is), and in case of non-scheduled announcement releases, recent losses amplify the liquidity shocks. The findings also indicate that liquidity measured over multiple order book price levels behaves quite differently compared to the conventional bid−ask spread calculated using data from the best order book levels, indicating that measuring liquidity just using top-of-the-book data may lead to misleading inferences.

Finally, the results show that in a liquidity aggregator, traders observe only a small proportion of liquidity streams available: on average, a trader observes 5.4 streams out of the total 165 streams provided by 42 liquidity providers (the maximum is 23 and the minimum is 1). However, traders observe relatively tight spreads already with four or five streams, and traders with more streams observe only marginal improvements in spread, if any. In theory, most traders could cut their observed spread by more than a half and save up to $0.18 basis points per €1 traded with the optimal combination of liquidity streams; in practice, however, traders may not be able to exploit the improvements because they are not free to choose just any streams in the aggregator, and if they would change the streams they observe, the liquidity providers would likely change their quoting behavior. Nevertheless, the novel empirical results can be used to assess the efficiency of the aggregator as a trading technology and the liquidity provision in the FX market, in general.

Tiivistelmä
(Abstract in Finnish)

Tampereen teknillinen yliopisto

Siikanen Milla: SIJOITTAJAT, INFORMAATION SAAPUMINEN JA MARKKINALIKVIDITEETTI: Empiirisiä tuloksia finanssimarkkoilta

Avainsanat: Osakemarkkinat, Valuuttamarkkinat, Sijoittajakäyttäytyminen, Sijoittajian sofistikoituneisuus, Päätöksenteko, Sosiaalinen media, Yritystiedotteet, Likviditeetti, Tarjouskirja, Likviditeettiaggregaattori, Kaupankäyntikustannukset, Tiheästi poimittu data

Hyvin toimivat rahoitusmarkkinat ovat yhteiskunnalle monin tavoin hyödylliset. Sijoittajat, informaation saapuminen ja markkinalikviditeetti puolestaan ovat kaikki avain asemassa rahoitusmarkkinoilla. Ilman sijoittajia jotka käyvät kauppaa, markkinoita ei olisi. Uuden informaation saapuminen markkinoille on puolestaan tärkeää, sillä informaatio voi vaikuttaa kaupankäynninkohteiden arvostuksiin ja siten hintoihin. Toisaalta, jotta hinnat voivat päivittyä heijastamaan uutta informaatiota tehokkaasti, markkinoiden tulee olla riittävän likvidit: sijoittajilla tulee olla mahdollisuus käydä kauppaa silloin kun he haluavat ja kaupankäyntikustannusten tulee olla matalat.

Kaikkia kolmea aihetta on tutkittu aiemmin, mutta niiden keskinäisiin suhteisiin ei ole perehdytty syvällisesti aiemmassa kirjallisuudessa. Tässä väitöskirjassa tavoitteena on parantaa tietämystämme sijoittajien, informaation saapumisen ja likviditeetin keskinäisistä suhteista rahoitusmarkkinoiden kontekstissa. Perehtymällä aiemmasta kirjallisuudesta tunnistettuihin tutkimusaukkoihin, tämä tutkimus pyrkii tarjoaman uutta empiiristä tietoa, jota voidaan hyödyntää tieteellisessä tutkimuksessa luotettavampien rahoitusmarkkinoita käsittelevien mallien luomiseen; yleisellä tasolla luotu parempi ymmärrys näistä aiheista voi auttaa parantamaan markkinoiden sääntelyä, pörssien toimintaa ja sijoitusten hallintaa.

Tämä väitöskirja koostuu kahdesta osasta: yhteenveto-osuudesta ja neljästä tutkimusartikkelista (Artikkelit I−IV). Artikkeli I tutkii logistisella regressiolla sitä, miten Nokian Facebook julkaisut ja niihin liittyvät aktiviteetit liittyvät sijoittajien päätöksiin ostaa vai myydä Nokian osaketta. Artikkeli II hyödyntää tapahtumatutkimuksille tyypillistä rakennetta ja tiheästi poimittua dataa tarjouskirjamarkkinoilta sen tutkimiseen, miten osakkeiden tarjouskirjat kehittyvät aikataulutettujen ja ei-aikataulutettujen yritystiedotteiden julkaisuajankohtien ympärillä. Tähän liittyen, Artikkeli III tutkii regressioanalyysillä mitkä tekijät vaikuttavat yritystiedotteiden tarjouskirjoihin aiheuttamien likviditeettishokkien suuruuteen. Artikkeli IV puolestaan tutkii uniikkia datasettiä, ja artikkelin tavoitteena on määrittää kuinka suuri osa kokonaislikviditeetistä yksittäisellä sijoittajalla on käytössään valuuttamarkkinoilla likviditeettiaggregaattorissa ja kuinka paljon sijoittajan havaitsema osto- ja myyntikurssin erotus voisi pienentyä, jos sijoittaja voisi valita

optimaalisen yhdistelmän likviditeettivirroista sijoittajan nykyisen yhdistelmän sijaan. Optimaaliset yhdistelmät selvitetään tutkimuksessa geneettistä algoritmia hyödyntäen.

Aiempi tutkimus on selvittänyt, miten uutiset vaikuttavat eri sijoittajaryhmien kaupankäyntiin, ja yksi tämän tutkimuksen kontribuutioista on osoittaa, että (mahdollisesti puolueellinen) informaatio, jonka yritys julkaisee sosiaalisessa mediassa vaikuttaa eri sijoittajaryhmiin eri tavalla. Oletettavasti vähemmän sofistikoituneiden sijoittajien—passiiviset kotitaloudet ja voittoa tavoittelemattomat organisaatiot—päätösten ja Facebook datan väliltä löydetään yhteys, kun taas sofistikoituneemmat sijoittajat—finassiinsituutiot—vaikuttavat käyttäytyvän Facebook datasta riippumattomasti.

Yritystiedotteiden puolestaan havaitaan aiheuttavan merkittäviä muutoksia osakkeiden tarjouskirjoissa, mikä eroaa aiemmasta tutkimuksesta, jossa ei löydetty näyttöä siitä, että uutiset vaikuttaisivat tarjouskirjojen likviditeettiin. Aikataulutetut yritystiedotteet voivat nostaa tarjouskirjojen likviditeettiä normaalia tasoa paremmaksi, mikä antaa viitteitä siitä, että aikatalutettujen tiedotteiden julkaisu voi helpottaa markkinoilla vallitsevaa informaation asymmetriaa. Ei-aikataulutettujen yritystiedotteiden julkaisun jälkeen tarjouskirjojen likviditeetti on sen sijaan melko matala monessa tapauksessa vielä tunti tiedotteen julkaisun jälkeen. Tiedotteiden aiheuttamat välittömät likviditeettishokit ovat sitä voimakkaampia, mitä suurempi tarjouskirjojen asymmetria on ennen uutista. Lisäksi nopea reaktio on vahva reaktio (mitä nopeammin epälikvidisyyshuippu saavutetaan tiedotteen julkaisemisen jälkeen, sitä suurempi se yleensä on) ja ei-aikataulutettujen yritystiedotteiden tapauksessa hinnan lasku ennen uutista indikoi suurempaa likviditeettishokkia. Tulokset myös osoittavat, että likviditeetti mitattuna usean tarjouskirjan hintatason yli käyttäytyy melko eri lailla kuin tavallisesti käytetty likviditeetti mittari, joka perustuu parhaiden osto-ja myynti hintojen erotukseen. Tämä osoittaa, että likviditeetin mittaaminen käyttäen tietoa vain parhaista hintatasoista voi johtaa harhaanjohtaviin päätelmiin.

Tulokset osoittavat myös, että sijoittajilla on käytössään suhteellisen pieni osuus kokonaislikviditeetistä likviditeettiaggregaattorissa valuuttamarkkinoilla: sijoittajalla on käytössään keskimäärin 5.4 likviditeettivirtaa (maksimi on 23 ja minimi 1), vaikka aggregaattorin 42 likviditeetin tarjoajaa tarjoavat yhteensä 165 likviditeettivirtaa. Sijoittajat saavuttavat kuitenkin suhteellisen tiukan eron osto- ja myyntikurssien välillä jo neljällä tai viidellä likviditeettivirralla, ja sijoittajat joilla on tätä enemmän likviditeettivirtoja käytössään havaitsevat vain pienen parannuksen osto- ja myyntikurssien välisessä erotuksessa, jos sitäkään. Teoriassa, suurin osa sijoittajista voisi havaita yli puolet pienemmän osto- ja myyntikurssien välisen erotuksen ja säästää jopa $0.18 korkopistettä per ostettu tai myyty euro, mutta käytännössä sijoittajat eivät välttämättä pysty hyödyntämään näitä parannuksia ja säästöjä täysimääräisesti. Tämä johtuu siitä, että sijoittajat eivät voi valita likviditeetin tarjoajien tarjoamia likviditeettivirtoja täysin vapaasti, ja toisaalta siitä, että jos sijoittajat muuttaisivat käytössään olevien likviditeettivirtojen yhdistelmiä, likviditeetin tarjoajat todennäköisesti myös muuttaisivat tarjoamaansa likviditeettiä. Tästä huolimatta, esitettyjä uusia empiirisiä havaintoja voidaan käyttää arvioimaan aggregaattorin tehokkuutta kaupankäyntiteknologiana ja likviditeetin tarjoamista valuuttamarkkinoilla yleisesti ottaen.

Preface

Little did I know when I started working on my Master's thesis at the Department of Industrial Management under the supervision of Professor Juho Kanniainen in September 2013 that I would still be here at TUT five years later. To be honest, I had no plans on staying or even pursuing a Ph.D.: I did not even really know what it would mean in practice. That being said, I am glad that I stayed. This project has taught me so much on many things—and not least on myself. I would not be here now without the help, support, and encouragement of so many people. Although I must give some credit also to my own ambitiousness, it would not have been possible to get here without the support I received, and I will forever be grateful to all of you (including all those I unintentionally forgot to name here).

As per the most official part, I want to thank TUT's doctoral school for believing in me and for providing me funding for an extensive period of four years, thus allowing me to focus on my research instead of worrying if I have funding for the next year. I am also grateful to all the foundations that supported my work with travel and encouragement grants during all these years: Marcus Wallenberg's Foundation, the Finnish Foundation for Technology Promotion, the Foundation for Economic Education, Savings Banks' Research Foundation, OP Group Research Foundation, Waldemar von Frenckells stiftelse, KAUTE Foundation, Nordea Bank Foundation sr., and Finnish Concordia Fund. They provided not only financial support but also trust and confidence that I am doing something interesting and valuable. I would also like to express my gratitude to the amazing people at the Graduate School of Finance, who allowed me to participate in their events and provided me with valuable feedback, which helped me improve my research considerably.

I am most indebted to my supervisor, Professor Juho Kanniainen, who was the one who encouraged me to be involved in all of this in the first place. Juho always had time for me and my concerns whenever it was needed the most, especially at the beginning of this project. Later on, he moved the responsibility more to me, ensuring that I was growing throughout the process. His encouragement, enthusiasm, and positive attitude toward problem solving have been indispensable to me.

I am also grateful to the guys at big xyt AG, especially Dr. Ulrich Nögel. Ulrich's support in helping me understand how things work "in real life" has been truly important and useful. I have learned so much more during our discussions than I could have ever learned from any book. I appreciate all of his efforts in acquiring the unique data for our FX market research and I also appreciate the efforts and support I received from the people working in the company providing the data. Moreover, I owe my sincerest thanks to all the great people at the Department of Financial Mathematics at Fraunhofer ITWM (and their IT support), especially Dr. Andreas Wagner, for having me there during my research visit, and, again, to Ulrich, for helping me make this visit happen.

Importantly, I wish to thank my ingenious colleagues, current and former, from the DARE Business Data Research Group (formerly, our part of the group was known as the Financial Engineering Research Group). Especially the ones combating financial data—Kestutis, Martin, Margarita, Sindhuja, Jaakko, Ye, Jun, Hanxue, Binghuan, Jimmy, Adam, Perttu, and Teemu, as well as all the research assistants we had—all the discussions and other (not-so-official) activities we had have been amazing and helped improve the quality of my life during these past years. You are the most ingenious of the geniuses I know! To all my other colleagues in the department—especially Tero, Natalia, Lauri V., Johanna, Sanna, Hanna-Riikka, Toni, Matias, Olli, Rami, Jesse, and Jukka, among many others—your unwavering (peer) support was irreplaceable; thank you for this! Because of you, my dear colleagues, I never felt alone in my struggles. My special thanks go to Eija; the amount of tea and sympathy (read: whining, although we also shared many moments of happiness) we shared during the past four years were truly enormous (as one cup of tea is close to half a liter and on our best days, we probably had five cups).

To my amazing friends outside the academia: although you perhaps did not fully understand what I have been doing, you managed to give me just the support that I needed. Please know that I could not have done this without you! For example, by going to aerial yoga and playing football with me (in addition to so many other things), you all contributed to keeping my head together during this enlightening, although sometimes kind of stressful and exhausting, process.

Finally, to my dear family—my sisters and my parents, along with all other relatives and in-laws of mine—I am truly grateful for having you in my life, and I would not be me and would not be here now without you. I will forever be grateful for the support I received from you! To my dear husband, Ville, all the exaggerated words have probably been used already, but honestly, there are no words great enough to describe how much you mean to me and how big a role you have played in all of what I have accomplished thus far and will still accomplish in the future. I trust that you already know that your love and support mean the world to me.[†]

Tampere, August 2018

Milla Aleksandra Siikanen

[†]Ville, please also accept my sincerest apologies for making you wear a suit and participate in the formal dissertation celebration party, but I think this is one of those once-in-a-lifetime things.

Olisihan sitä voinut lähteä jo aamiaisen jälkeen,
mutta käsitäthän, että tämä on sellainen tapaus,
että meidän on odotettava auringonlaskua.
Suuri lähtö on yhtä tärkeä kuin ensimmäiset
luvut kirjassa. Ne määräävät kaiken.

Muumipappa (Muumipappa ja meri)[‡]

[‡]Loosely translated: *One could have left already after breakfast, but on an occasion like this we must wait for sunset. The grand departure is just as important as the first chapters in a book. They determine everything.* Moominpappa, in the book "Moominpappa at Sea."

Contents

Acronyms

EURUSD	euro to US dollar exchange rate
FX	foreign exchange
GA	genetic algorithm
LP	liquidity provider
LOB	limit order book
OTC	over-the-counter
XLM	Exchange Liquidity Measure

List of Publications

This thesis is based on the following original publications and a manuscript, which are referred to in the text as Articles I–IV. These original articles have been reproduced with the kind permission of the publishers.

I Siikanen M., Baltakys, K., Kanniainen, J., Vatrapu, R., Mukkamala, R., Hussain, A. “Facebook drives behavior of passive households in stock markets”, *Finance Research Letters*, forthcoming, March 2018.

II Siikanen M., Kanniainen, J., Valli, J. “Limit order books and liquidity around scheduled and non-scheduled announcements: Empirical evidence from NASDAQ Nordic”, *Finance Research Letters*, vol 21, pp. 264-271, May 2017.

III Siikanen M., Kanniainen, J., Luoma, A. “What drives the sensitivity of limit order books to company announcement arrivals?”, *Economics Letters*, vol 159, pp. 65–68, Oct. 2017.

IV Siikanen M., Nögel, U., Kanniainen, J. Liquidity in the FX market: empirical evidence from an aggregator”, *Unpublished manuscript; submitted to a journal*, July 2018.

Author’s contributions to the co-authored publications

The following describes my personal contributions to the research process underpinning Articles I–IV, which involved multiple authors. I was the first author in all of these articles.

In Article I, in collaboration with my co-authors, I planned the research design and methodology and researched the literature. In addition, I conducted the analysis and wrote most parts of the original draft, and I participated in and guided the review and editing of the paper. The comments and suggestions of anonymous reviewers and editors have also helped improve the article. The article will also be a part of the compendium dissertation of my colleague, Kęstutis Baltakys.

In Article II, in collaboration with my co-authors, I planned the research design and methodology and pre-processed the data. I conducted the analysis, researched the literature, and wrote the original draft. I also reviewed and edited the paper with the help of my co-authors. The comments and suggestions of anonymous reviewers and editors have likewise helped improve the article.

In Article III, in collaboration with my co-authors, I planned the research design and methodology and pre-processed the data. I conducted the analysis, researched the literature, and wrote the original draft. I also reviewed and edited the paper with the support of my co-authors. The comments and suggestions of anonymous reviewers and editors have likewise helped improve the article.

In Article IV, in collaboration with my co-authors, I planned the research design and methodology and prepared the data. Additionally, I conducted the analysis, researched the literature, and wrote most parts of the original draft. I was also responsible for reviewing and editing the paper with my co-authors.

1 Introduction

This chapter introduces the reader to the topics and research questions of this dissertation. The first section gives the general motivation for the topics, whereas the second section discusses the research gaps and frames the research questions. The last section outlines the structure of this dissertation and provides short summaries of Articles I−IV.

1.1 Background and motivation

In financial markets, various *investors* trade financial assets, such as stocks.[1] Society can be argued to benefit from well-functioning financial markets: for example, from economics perspective, financial markets can improve the allocation of capital within an economy and facilitate specialization by enabling companies to hedge (i.e. transfer the risk), leading to more efficient production processes (Harris, 2003; Wurgler, 2000). According to Harris (2003), the public benefits of having well-functioning markets can be categorized into two groups: those arising from having markets that produce *informative prices*, and those coming from having *liquid markets*. Accordingly, there are two widely recognized motives of investors to trade, *information* and *liquidity*: informed investors trade on their private information with the aim to profit, whereas the needs of liquidity investors to trade emerge from outside the market and are not directly linked to the future payoffs of the asset (see e.g. Admati and Pfleiderer, 1988; Harris, 2003). This dissertation is dedicated to the interrelations between these three key aspects of financial markets—investors, information arrivals, and liquidity—as illustrated in Figure 1.1.

Intuitively, in the core of financial markets are the investors who trade: without them, there would be no trading and hence no markets. Investors differ in their motives to trade, their risk profile, the information they have access to, the regulatory constraints they face, and so on (see e.g. Harris, 2003; Lillo et al., 2015). All these qualities potentially affect the trading behaviors and the decisions made by investors, and the field of behavioral finance and theoretical models could "clearly [...] benefit from a more complete picture of how investors actually behave and how they differ from one another in the way they react to the same information." (Grinblatt and Keloharju, 2000, see also Grinblatt and Keloharju 2001).

Furthermore, information arrivals are important because they affect the information available for investors, and they potentially influence the valuations of assets, thus affecting stock prices (Cochrane, 2005; Fama, 1970; Fama et al., 1969). Nowadays, financial markets are characterized by an almost continuous flow of information from various sources (Foucault et al., 2016). In addition to official (mandatory) company announcement releases, social media sites create new opportunities for companies to

[1] In this dissertation, I use the term "investor" for all market participants trading in financial markets.

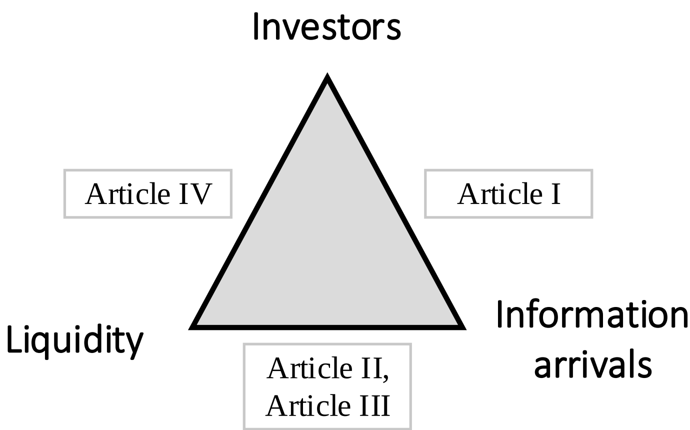

Figure 1.1: The key topics of this dissertation and the positions of Articles I – IV with respect to the topics.

improve their communication. In fact, as of January 2013, around 45% of S&P1500 companies used Facebook and Twitter to communicate about their businesses (Jung et al., 2017).

However, in order for stock markets to be efficient in the sense that prices adjust rapidly to new information (Fama, 1970), the markets need to be sufficiently liquid (see also Chordia et al., 2008). In fact, liquidity has been widely recognized as an important characteristic of any well-functioning market (see e.g. Chordia et al., 2001; Harris, 2003; O'Hara, 1995).[2] Although defining liquidity poses a challenging task (see e.g. O'Hara, 1995), roughly speaking, in liquid markets, an investor has the ability to trade when he/she wants at a low transaction cost (Harris, 2003). Understanding liquidity is important from many perspectives, as it is directly linked to the corporates cost of capital, and understanding liquidity dynamics can help in formulating effective hedging and trading strategies and aid in preventing disruptive events, such as the flash crash in 2010[3]; more generally, a better understanding of liquidity and trading activity could improve exchange organization, regulation, and investment management (Amihud and Mendelson, 1991; Chordia et al., 2001; Cumming et al., 2011; Rosa, 2016)

Generally speaking, investors provide and consume (make and take) liquidity in financial markets, although how this process is organized depends on the market structure (Bloomfield et al., 2005; Harris, 2003). In turn, the market structure also affects the way liquidity can be measured in the markets. In limit order books, which are typical in stock markets, all investors can choose to make liquidity by submitting a limit order (indicate that they are willing to trade a certain quantity with a certain price), or take liquidity by submitting a market order (trade immediately a certain quantity with the best price available). All the outstanding (unexecuted and uncancelled) limit orders constitute the limit order book, and buy orders are referred to as bids and sell orders as asks. The data based on the best bid and ask prices and the respective quantities (bid orders with the

[2] In this dissertation, the term "liquidity" is used to refer to "market liquidity," which is not the same as "funding liquidity" (see e.g. Brunnermeier and Pedersen, 2008).

[3] See e.g. "FINDINGS REGARDING THE MARKET EVENTS OF MAY 6, 2010" by CFTC and SEC, available at https://www.sec.gov/news/studies/2010/marketevents-report.pdf.

highest price and ask orders with the lowest price) are referred to as Level I data, and data on the orders beyond the best price levels are referred to as Level II order book data (see e.g. Cao et al., 2009). On the other hand, in some market structures, such as in liquidity aggregators, which are typical in the foreign exchange (FX) market (see Oomen, 2017), there are designated liquidity providers that provide all the liquidity available, and the traders then only take liquidity.

1.2 Research objectives and research questions

As shown in the previous section, investors, information arrivals, and liquidity are all key elements in the financial markets and are inevitably linked to one another. However, the interrelations between them have not been studied in depth in earlier literature. This dissertation aims to provide empirical evidence that will fill several gaps related to the interrelations between investors, information arrivals, and liquidity in the context of financial markets. The **overall objective** of this dissertation is to:

Enhance the understanding of and contribute to the empirical knowledge related to the interrelations between investors, information arrivals, and liquidity.

One of the important questions is how information arrivals affect the behavior of investors. Companies have to communicate information, for example, on their operations and financial situation through exchange routed company announcements (see NASDAQ Helsinki Ltd, 2018), but nowadays, they may choose to communicate also via less official channels, such as social media. In the field of finance, there exists substantial research on how data from social media affect stock prices (see e.g. Bollen et al., 2011; Chen et al., 2014; Karabulut, 2013; Nofer and Hinz, 2015; Siganos et al., 2014; You et al., 2017; Zhang et al., 2011, 2017; Zheludev et al., 2014), but research on how social media affects the decisions of (individual) investors is scant. Lillo et al. (2015) study how Thomson Reuters news articles affect the trading behavior of different investor groups, but the effects of information released via social media on the trading of individual investors are still largely an unexplored area. Article I aims to fill this gap by answering the following research question:

RQ I: *Does the information that companies release on social media affect investors' decisions to buy versus sell?*

By combining unique investor-level shareholding registration data that include the trading of all Finnish investors with a data set on Facebook posts and activities on Nokia's Facebook wall, Article I examines the extent to which investors' decisions are driven by Facebook. In particular, for different investor classes, given that an investor trades, Article I studies how Facebook data are related to the investor's decision to increase versus decrease his/her position. Studying investors from different classes separately is important because an investor's access to data sources and his/her sophistication may affect the extent to which the (potentially biased, see Jung et al., 2017) information communicated via social media drives this investor's behavior.

While information arrivals affect the trading decisions of investors, theory and earlier literature suggest that information arrivals also influence the liquidity provision (see e.g. Baruch et al., 2017; Graham et al., 2006; Lee et al., 1993; Riordan et al., 2013). On the other hand, sufficient liquidity is especially important around information arrivals, when the prices need to adjust to the new information, and, thus, immediacy may be required. In particular, in limit order markets, liquidity beyond the best levels is important around information releases, as informed investors may cause a large, immediate demand for liquidity over multiple price levels. However, the majority of earlier studies examining liquidity around information arrivals focus on the bid−ask spread, i.e. the difference between the best ask and bid prices (Level I data) (see e.g. Graham et al., 2006; Groß-Klußmann and Hautsch, 2011; Krinsky and Lee, 1996; Lee et al., 1993; Neuhierl et al., 2013), although some important exceptions exist (see e.g. Gomber et al., 2015; Riordan et al., 2013). Moreover, scheduled and non-scheduled announcement releases may affect liquidity differently, so it is important to make a distinction between them (see e.g. Graham et al., 2006), which most of earlier research does not do. Article II addresses this gap by answering the following research question:

RQ II: *How is order book liquidity, i.e. liquidity beyond the best levels in an order book, affected by scheduled and non-scheduled company announcement releases?*

Furthermore, there are studies (including Article II) showing that information arrivals may cause liquidity shocks to the markets (Engle et al., 2012; Erenburg and Lasser, 2009; Riordan et al., 2013; Rosa, 2016, see). However, the question on the different factors affecting the magnitude of order book liquidity shocks following announcement releases remains unanswered. As liquidity is especially important around information arrivals, the research question in Article III is as follows:

RQ III: *What factors affect the magnitude of the order book liquidity shocks caused by scheduled and non-scheduled company announcement releases?*

The analysis in Articles II and III utilize high-frequency limit order book data for 75 actively traded stocks on NASDAQ Nordic around official exchange routed company announcements to study how order book liquidity evolves around scheduled and non-scheduled information arrivals (Article II), as well as how different variables calculated from the order book data affect the magnitude of the order book liquidity shocks following the announcement releases (Article III). Order book liquidity is measured by capturing the shape of the order book through an estimation of the slopes of the order book curves (see Deuskar and Johnson, 2011; Härdle et al., 2012; Malo and Pennanen, 2012; Næs and Skjeltorp, 2006). Article II also compares the evolution of order book liquidity with the evolution of the spread, as earlier research has shown that liquidity measured using just top-of-the-book data (i.e. the spread calculated from Level I data) may lead to a conclusion that is different from that when information from multiple order book levels is used (see Rosa, 2016; Sensoy, 2016).

In limit order books, which are typical in stock markets, liquidity making and taking are relatively straightforward: all investors can make and take liquidity, all investors observe the order book (or a certain part of it, i.e. n best bid and ask price levels)[4], and all investors can trade with one another. By contrast, in liquidity aggregators, which are typical in FX markets and other over-the-counter (OTC) markets (Oomen, 2017), traders observe only a subset of liquidity streams quoted in the total aggregator and can trade only with a limited subset of the liquidity providers. Liquidity in FX markets is needed for international trade, and it is also highly important because of the market's huge size and the crucial role in guaranteeing efficiency and arbitrage conditions in many other markets; nevertheless, the liquidity there is not understood well enough (Karnaukh et al., 2015; King et al., 2012; Mancini et al., 2013). In particular, although Oomen (2017) provides a profound theoretical discussion of liquidity in an aggregator, to my knowledge and the

[4]Excluding hidden orders, which no one observes (see e.g. Moro et al., 2009).

knowledge of my co-authors in Article IV, there exists no empirical research on liquidity aggregators. To address this gap, Article IV explores the following research question:

RQ IV: *What is the proportion of liquidity traders observed in a liquidity aggregator, and how much could a trader's observed spread be improved by obtaining an optimal combination of liquidity streams?*

The unique data set used in Article IV includes detailed information on all the streamed quotes for individual liquidity providers and individual traders for EURUSD; such detailed data are usually not available for academic research.[5] One reason for the lack of available and representative data may be the overall fragmentation of the market (Karnaukh et al., 2015; Mancini et al., 2013). Furthermore, given its complexity and opaque nature, the FX market is currently under pressure and is being criticized for its lack of transparency and its recent scandals.[6] Moreover, Gould et al. (2017) write that studying the subset of liquidity that traders observe in the FX market poses an interesting challenge for future research. All this makes the insights provided in Article IV unique and highly interesting.

1.3 Dissertation structure and outline of the original articles

This dissertation is divided into two sections: (i) the introductory part with the introduction and summary, and (ii) the original publications (Articles I−IV). There is also an Appendix section at the end of the dissertation presenting the appendix to this introductory part (Appendix A) and the online appendices of the original publications I−III (Appendix B, Appendix C, and Appendix D).

The remainder of this introductory part is structured as follows. Chapter 2 introduces the reader to the key concepts of this dissertation and some related research. Chapter 3 presents the data sets used in Articles I−IV, and Chapter 4 outlines the methods used. Chapter 5 summarizes the findings of Articles I−IV. Finally, Chapter 6 discusses and concludes this dissertation.

Article I studies how investors' trading decisions are related to Facebook posts and related activities. Article I utilizes data on Finnish investors' transactions on Nokia stock from mid-2010 until the end of 2016. The results of logistic regressions indicate that especially passive household investors' and non-profit organizations' decisions to buy versus sell are associated with Facebook data. At the same time, arguably more sophisticated investors—financial and insurance institutions—seem to behave independently from Facebook activities.

Articles II and III use high-frequency limit order book data for 75 frequently traded stocks from NASDAQ Nordic over the four-year period of 2006−2009. Exploiting the framework from event study analysis, Article II explores order book liquidity during a two-hour event window around scheduled and non-scheduled company announcement releases. Article II finds significant intra-day changes in order book liquidity around company announcement arrivals: the announcements are followed by immediate liquidity shocks (within a few

[5] The data set is highly sensitive, and I have obtained permission to publish the academic results as long as it is not possible to identify the individual liquidity providers or traders. To guarantee their anonymity, we are not allowed to publish the name or any other background information of the data provider.

[6] See, for example, "Third Barclays Trader Faces U.S. Charges in FX Scandal" (Bloomberg; Jan 17$^{\text{th}}$ 2018) and "The global FX rigging scandal" (Reuters; Jan 11$^{\text{th}}$ 2017). The concerns were also partially addressed in the recently published "FX Code of Conduct," available at `https://www.globalfxc.org/docs/fx_global.pdf`.

minutes after the announcement release). The order book liquidity is exceptionally low before scheduled announcement releases, but it improves to an exceptionally good level after them, indicating that investors react to scheduled announcement releases already at least an hour before the announcements, and the release of scheduled announcements may significantly reduce information asymmetries between the investors and adverse selection costs. By contrast, after non-scheduled announcement releases, the aggregated order book liquidity is still abnormally low an hour after the announcement release in most cases. Additionally, the analysis shows that the spread (Level I liquidity measure) behaves quite differently compared to the order book liquidity (Level II liquidity measure).

Article III focuses on identifying the factors related to the magnitude of the order book liquidity shocks caused by scheduled and non-scheduled company announcement releases. The results of linear regression using within-transformation indicate that recent losses amplify the illiquidity shocks caused by non-scheduled announcements. Additionally, the faster the maximum illiquidity is reached, the more illiquid the order book becomes (i.e. a fast reaction is a strong reaction), and the asymmetry observed in the book before an announcement arrival is positively associated with the magnitude of the illiquidity shock.

Article IV studies liquidity in an FX aggregator with a unique, detailed data set on all the streamed quotes for individual liquidity providers and individual traders for EURUSD over a 10-trading day period from 26 September 2016 to 7 October 2016. The results show that traders observe, on average, 5.4 streams out of the total 165 active streams quoted by 42 liquidity providers, while the minimum is 1 and the maximum is 23. Moreover, traders obtain a relatively tight spread already with four or five observed liquidity streams; the use of more streams yields only a marginal benefit, if any. The optimal combinations of liquidity streams are solved using a genetic algorithm (GA), and the comparisons show that most of the traders could—at least in theory—reduce the average spread they observe by more than half with the optimal combination of streams, and a trader could save up to $0.18 basis points per €1 traded. In practice, however, the traders may not be able to fully exploit improvements in spreads because they are not completely free to choose just any liquidity streams quoted by the liquidity providers in the aggregator; significant changes in the sets of streams that the traders observe could also affect the quoting behavior of the liquidity providers.

2 Key concepts and related research

This chapter introduces readers to the key concepts of this dissertation and discusses briefly the related literature. It provides the background for Articles I−IV, partially complementing the articles, although some of the topics are discussed in greater detail in the articles. Sections 2.1, 2.2, and 2.4 present the key concepts of this dissertation, whereas the rest of the sections are dedicated to the interrelations between the concepts.

2.1 Investors in financial markets

In financial markets, investors trade financial instruments, for example stocks, and without investors, there would be no trades and hence no markets. In this dissertation, the term "investor" is used for all market participants trading in financial markets. These investors can be divided into different groups based on their qualities and the reasons why they trade (Harris, 2003). For example, investors can differ in their risk profile, the size of their holdings, the information they have access to, and the regulatory constraints they face (Lillo et al., 2015).

Classifying investors based on their category considers many of the differences between investors (Lillo et al., 2015). Baltakys et al. (2018); Grinblatt and Keloharju (2000, 2001); Lillo et al. (2015); Tumminello et al. (2012) use the following investor categories arising from their data sets: financial and insurance corporations, non-financial corporations, general governmental organizations, non-profit institutions, households, and foreign investors. Article I utilizes this categorization for domestic investors. Grinblatt and Keloharju (2000) argue that institutions (i.e. financial and insurance corporations and non-financial corporations) are likely to be the most sophisticated domestic investor groups, as they tend to take larger positions, have more resources to spend on research, and often view investment as a full-time career. Governmental and non-profit institutions seem to be less sophisticated than the other institutions but are still more sophisticated than households, whereas larger households are presumably more sophisticated than the smaller ones (Grinblatt and Keloharju, 2000, see also Grinblatt and Keloharju, 2001).

Grinblatt and Keloharju (2000) study how past returns are related to investors' propensity to buy and sell, and they find that domestic investors, particularly households, tend to be contrarian traders (selling past winning stocks and buying past losers), whereas foreign investors exhibit momentum. Consistently, Grinblatt and Keloharju (2001) show that when trading, sophisticated investor classes place less weight on past returns when deciding whether to buy or sell, whereas less sophisticated investors (households, general governmental organizations, and nonprofit institutions) are more likely to sell rather than buy stocks with large past returns. Lillo et al. (2015) show that investors in different categories react to exogenous factors (number of news articles and their sentiment) and endogenous factors (returns and volatility) differently: households and non-financial companies are very sensitive to both factors, whereas governmental and non-profit organizations are only weakly sensitive, and financial institutions are intermediate between these two cases. Tumminello et al. (2012) cluster investors based on their trading and find that several clusters show over-expression of a certain investor group, providing evidence that there is homogeneity between the investors in a category.

However, such detailed data on investor categories are usually not available, and much research focuses on the trading of individuals or institutions or on the differences between the two (see e.g. Gompers and Metrick, 2001; Griffin et al., 2003; Kaniel et al., 2008, 2012; Nofsinger and Sias, 1999, and references therein). For example, Griffin et al. (2003) document a strong contemporaneous positive (negative) relation between institutional (individual) trading activity and daily stock returns. Kaniel et al. (2008) find that intense buying by individuals leads to positive excess returns in the following month, and selling by individuals leads to negative excess returns.

Kaniel et al. (2008) write that "For a variety of reasons, financial economists tend to view individuals and institutions differently." Usually, institutions are seen as informed investors, whereas individuals are perceived as uninformed noise traders with psychological biases (Kaniel et al., 2008). The division between informed and uninformed investors is likely to stem from theoretical models (see e.g Admati and Pfleiderer, 1988; Black, 1986; Glosten and Milgrom, 1985; Kyle, 1985). The biases related to individual investors' behavior include individual investors' tendency to misinterpret new information (sell stocks announcing positive news and buy stocks announcing negative news), show poor stock-picking skills (the stocks that individuals buy underperform those that they sell), exhibit the disposition effect (sell winners and hold on to losers), and make contrarian trades (be net buyers (sellers) in a stock as the stock price falls (rises)) (see Linnainmaa, 2010, and references therein). However, Linnainmaa (2010) shows that many of the documented biases can actually be consequences of individual investors' tendency to use limit orders (for more information on limit orders and limit order markets, see Section 2.5).

2.2 Information arrivals

According to Cochrane (2005), all asset pricing theories start from one simple concept: the price of an asset equals this asset's expected discounted payoff. The expectation is conditional on the investor's (current) information (Cochrane, 2005). Information arrivals may change investors' expectations on the future payoffs, and, consequently, the price changes, i.e. the information arrivals drive the prices. If the markets are efficient, the new information is reflected in the prices immediately after the information arrival (Fama, 1970). An early work by Fama et al. (1969) shows that stock prices adjust to new information, namely, announcements of stock splits. Nowadays, financial markets are characterized by an almost continuous flow of information: investors rely increasingly on machine-readable texts, e.g. tweets, Facebook pages, blogs, newswires, economic and corporate reports, and company websites (Foucault et al., 2016). Thus, there are various channels through which companies can release information.

Company announcements filed with the stock exchange represent mandatory information releases. For listed companies, disclosure rules[1] regulate the procedures on releasing

[1] Disclosure rules are based on the Market Abuse Regulation EU 2014/596, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32014R0596.

information: in essence, all information that is likely to influence the valuation of a listed company must be published "as soon as possible in a manner that information is available in a non-discriminatory way enabling fast access and complete, correct and timely assessment of the information by the public. The listed company shall provide the inside information to major media as well as to Financial Supervision Authority and the Exchange." (NASDAQ Helsinki Ltd, 2018)[2] Articles II and III utilize these company announcements.

On the other hand, social media sites, such as Facebook and Twitter, create various opportunities for companies to improve their internal and external communication. As of January 2013, about 45% of S&P1500 companies utilize Facebook and Twitter to communicate externally about their business (Jung et al., 2017). Zhou et al. (2015) study a sample of almost 10,000 companies and report that only about 7% and 3.5% of the Facebook and Twitter messages that companies post are related to corporate disclosures. While it is mandatory for the listed companies to release certain information, they can strategically choose to further disseminate, discuss, and promote the same (and other) information via social media (Jung et al., 2017). Hence, the information disseminated via social media may be biased. Article I utilizes Facebook data.

The information arrivals can also be divided into two groups based on whether the timing of the announcement is known in advance (i.e. scheduled announcement) or if it comes as a surprise (non-scheduled announcement) (Chae, 2005; Graham et al., 2006; Lei and Wang, 2014; Zheng, 2018). Graham et al. (2006) argue that making this distinction is especially important when investigating information processing in financial markets: otherwise, it is impossible to isolate the effect of the event timing from the effect of the content. Articles II and III make this distinction. However, following this division mostly makes sense for official company announcements, as there is no regulation on information released via social media, and, hence, it would be difficult to argue why investors would know beforehand the timing of some of the information releases.

2.3 Investors and information arrivals

Graham et al. (2006) provide an excellent discussion of how scheduled and non-scheduled information arrivals affect the behavior of investors according to theoretic asymmetric information models. According to them, traditional models (e.g. Glosten and Milgrom, 1985; Kyle, 1985) assume that informed traders may have information about the timing and content of an upcoming announcement release, whereas uninformed traders may only know the timing of a scheduled announcement. Before announcement releases, informed investors aim to profit on their private information, and uninformed investors are reluctant to trade if they think that the probability of trading with an informed investor is high (i.e. before scheduled announcements; before non-scheduled announcements, informed trading may be difficult to detect). Informational asymmetries are assumed to be reduced by the announcement releases. According to Graham et al. (2006), in contrast to traditional models, Kim and Verrecchia (1994) assume that informed investors have an advantage in interpreting public news announcements, especially non-scheduled ones, leading to increased information asymmetry after the announcement releases.

Lillo et al. (2015) examine the trading and investment decisions of single investors and how these are affected by Thomson Reuters news articles. They find that governmental

[2] See also http://business.nasdaq.com/list/Rules-and-Regulations/European-rules/common/index.html#tcm:5044-19496.

and nonprofit organizations are weakly sensitive to the news, whereas households and non-financial companies are very sensitive to both the news articles and their sentiment, and financial institutions are intermediate between these two cases. Griffin et al. (2003) study the relation between the trading of institutional and individual investors and stock returns, and they find many interesting patterns; they write that "These patterns could be driven by institutional and individual investors trading on different information and/or perceiving past stock return moves differently." These points highlight why it is interesting to study the behavior of investors from different categories separately when examining the effects of information arrivals.

Kaniel et al. (2012) look at the trading of individual investors around earnings announcements and find evidence of informed or skillful trading by individuals (at least on an aggregated level). Brennan et al. (2018) find evidence of informed trading both before and after scheduled and non-scheduled company announcement releases. Chae (2005) and Zheng (2018) find a decreased trading volume before scheduled announcement releases, which is postulated to be caused by the postponement of trading by uninformed (liquidity) traders, whereas prior to non-scheduled (unexpected) announcements, an increased trading volume found is associated with increased trading by informed traders. Lei and Wang (2014) study the trading of informed investors (insiders) around scheduled and non-scheduled announcements, and they find that the trading volume of informed investors increases with the amount of trading of uninformed traders before both announcement types.

While the discussion above shows that there is research on the trading of investors around some types of information releases, to my knowledge, there exists no such studies on social media data, particularly with such detailed data as those used by Lillo et al. (2015) to explore the effects of news. This is perhaps due to the lack of availability of investor account-level data. In financial market research, data from social media have been used mainly to predict stock returns (see e.g. Bollen et al., 2011; Chen et al., 2014; Karabulut, 2013; Nofer and Hinz, 2015; Siganos et al., 2014; You et al., 2017; Zhang et al., 2011, 2017; Zheludev et al., 2014).[3] However, the fact that data from social media affect the stock returns gives an indication that social media could also affect the behavior of investors (because it affects the prices).

Using survey data, Yang et al. (2017) find evidence that social media, and mass media in general, influences investors' trading decisions. Snow and Rasso (2017) argue that less sophisticated investors potentially benefit the most from disclosures communicated via social media because the information is essentially "pushed" to them on social media platforms, which makes this information easier to access, compared with the traditional practice of "pulling" information, in which investors actively seek information to facilitate their decision making. In addition, in an experimental setting, Snow and Rasso (2017) show that less sophisticated investors process the financial information received from social media differently from information received via a company's investor relations website, leading to different judgments about the information.

One reason why news and social media posts may affect particularly the behavior of household investors (in addition to the information released, which potentially affects the valuation of the company) is the attention grabbing behavior of households (see Barber and Odean, 2007). Barber and Odean (2007) argue that when buying stocks,

[3] See also Bukovina (2016) for an overview of research linking social media data to data from financial markets.

investors have literally thousands of possibilities (by contrast, when selling, individual investors, in particular, mostly consider only stocks they already own, which are typically only a few). As the search problem for stocks to buy is huge, many investors may solve this by considering only those stocks that have recently caught their attention. In line with this thought, Barber and Odean (2007) find that individual investors display attention-driven buying behavior—for example, they are net buyers on days when stocks are in news. At the same time, professional (institutional) investors have more time and resources to monitor continuously a wider range of stocks and are thus less prone to make attention-driven trades.

2.4 Liquidity

Liquidity is a crucial element of any well-functioning market (see e.g. Chordia et al., 2001; Harris, 2003; O'Hara, 1995). However, it is not easy to define liquidity, potentially because of its many dimensions (see e.g. Harris, 2003). In fact, Harris (2003) argues that liquidity means different things to different people, and while investors and regulators talk about liquidity all the time, they are rarely clear about what they mean.

In a perfectly liquid market, an investor can buy and sell any amount of shares immediately at the same price (see e.g. Chacko et al., 2011)—by contrast, in illiquid (real-world) markets, an investor always loses money if he/she first buys shares and then immediately sells them (i.e. makes a round-trip transaction, see e.g. Amihud and Mendelson, 1991). In general, liquid markets have the following characteristics (see e.g. Amihud and Mendelson, 1991; Black, 1971; Harris, 2003; Hasbrouck, 2007; Kyle, 1985; O'Hara, 1995, the terms in square brackets are from Kyle, 1985):

- The trading costs are low [tightness].
- The trades (including large ones) can be executed (almost) immediately, and the impact of trade size on the price is small [depth].
- Prices and market liquidity recover fast after large transactions [resiliency].

To put it simply, "liquidity is the ability to trade when you want to trade" (Harris, 2003).[4]

As defining liquidity is not an easy task, one could guess that measuring it is also not straightforward. Many liquidity measures capturing the different aspects of liquidity have been suggested and applied in the literature (see e.g. Aitken and Comerton-Forde, 2003; Gomber et al., 2015; Goyenko et al., 2009; Malo and Pennanen, 2012; Rakowski and Beardsley, 2008). This study focuses mainly on static measures of liquidity, without considering the time dimension (resiliency). The most common (static) liquidity measures are the bid−ask spread, i.e. the difference between the best ask and bid prices, which measures the transaction cost (for relatively small transactions), and depth, which

[4]The European Commission defines liquidity in *Glossary: Useful terms linked to markets in financial instruments* in the following way: "Liquidity is a complex concept that is used to qualify the markets and the instruments traded on these markets. It aims at reflecting how easy or difficult it is to buy or sell an asset, usually without affecting the price significantly. Liquidity is a function of both volume and volatility. Liquidity is positively correlated to volume and negatively correlated to volatility. A stock is said to be liquid if an investor can move a high volume in or out of the market without materially moving the price of that stock. If the stock price moves in response to investment or disinvestments, the stock becomes more volatile." (Available at https://ec.europa.eu/info/system/files/glossary_en.pdf.)

measures the quantity available for trading (see e.g. Harris, 2003). As the market structure also affects the measurement of liquidity, Section 2.5 discusses the measurement of liquidity in limit order books, and Section 2.7 discusses the same in liquidity aggregators.

2.5 Limit order books

Nowadays, most of the modern stock exchanges are limit order-based electronic markets (see e.g. Biais et al., 1995; Bloomfield et al., 2005; Cao et al., 2009; Gould et al., 2013). In limit order markets, all investors can make and take liquidity by submitting limit and market orders, respectively. This is in contrast to quote-driven markets, where there are designated dealers (market makers, liquidity providers) who supply all the liquidity (Harris, 2003). When an investor submits a limit order, he/she commits to buy (a buy limit order, bid) or sell (a sell limit order, ask) a certain quantity of shares for a certain price per share. The limit order is valid until it is executed or until the investor cancels it. All the outstanding (unexecuted) limit orders form the limit order book. Limit order execution takes place when another investor submits a market order either to buy or sell a certain quantity of shares for the best price available. The limit orders in the order book follow price and time priorities: an incoming market order is first executed against the limit order with the best price (the highest bid or the lowest ask), and if there exist many orders with the same price, the priority is given to the oldest limit order.

Figure 2.1 illustrates the working mechanisms of a limit order book: $t = 1$ is the initial state of the limit order book, and at $t = 2$, a market order to buy three shares arrives (i.e. a trade takes place on the ask side). Next, at time $t = 3$, a new limit order (to buy) with price 3.6 and a quantity of 2 is submitted to the order book. Then, at time $t = 4$, a limit order (price 3.1 and quantity 2) is cancelled from the bid side. For more information on limit order books, readers are referred to (Biais et al., 1995; Gould et al., 2013; Harris, 2003) for examples.

From an investor's perspective, a market order guarantees the execution (given that there are outstanding limit orders with a total quantity large enough to match the market order), but the price is uncertain. A limit order, on the other hand, guarantees the price, but the execution remains uncertain (both the time of the execution and whether the order will be executed at all). Furthermore, limit orders can also be seen as free options, and they run a risk of being adversely picked off at an undesirable price if the expected value of the stock changes (e.g. because of information arrival) (see e.g. O'Hara, 1995).

2.5.1 Measuring order book liquidity

Many empirical studies use top-of-the-book Level I data to measure liquidity, such as the bid−ask spread and depth at the best price levels (see e.g. Chordia et al., 2008; Graham et al., 2006; Groß-Klußmann and Hautsch, 2011; Neuhierl et al., 2013, among others). Hence, one could ask why there is a need to use multi-level order book data (Level II data) instead of just the bid−ask spread or Level I depth. Scholars already discussed the problems with measuring liquidity using only the spread in the 1990s, with Lee et al. (1993) asserting that the spread between the best bid and ask prices is just one dimension of liquidity and that one should consider both the price dimension—spread—and the quantity dimension—depth—when studying liquidity. O'Hara (1995) also points out a problem related to the spread: the spread for large trades may be significantly larger than that for small trades. According to Rakowski and Beardsley (2008), the shift to electronic limit order books and the introduction of decimal pricing have reduced the

$t = 1$

bid side		ask side	
price	quantity	price	quantity
3.5	4	3.7	2
3.2	3	3.8	4
3.1	2	4.0	3
2.9	4	4.4	5

$t = 2$

bid side		ask side	
price	quantity	price	quantity
3.5	4	3.8	3
3.2	3	4.0	3
3.1	2	4.4	5
2.9	4		

$t = 3$

bid side		ask side	
price	quantity	price	quantity
3.6	2	3.8	3
3.5	4	4.0	3
3.2	3	4.4	5
3.1	2		
2.9	4		

$t = 4$

bid side		ask side	
price	quantity	price	quantity
3.6	2	3.8	3
3.5	4	4.0	3
3.2	3	4.4	5
2.9	4		

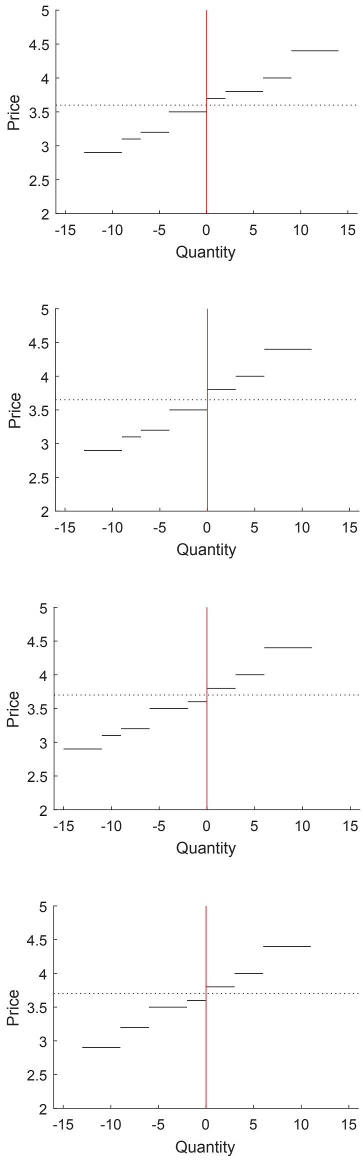

Figure 2.1: Example of a sequence of snapshots of a limit order book. The left side of the red line at zero quantity contains the bids (i.e. bid orders are presented as a negative quantity), referred to as the bid side of the book, and the right side contains the asks (i.e. positive quantity), referred to as the ask side. The black dotted lines represent the mid-prices, i.e. the mid-point between the best bid and ask prices.

overall relevance of liquidity at the inside quotes and steadily increased the importance of limit orders beyond the inside quotes. In addition, as Degryse et al. (2015) argue, “The depth beyond the best price levels matters to investors because it reflects the quantity immediately available for trading and therefore the price of immediacy.” The multi-level depth is important especially around information arrivals when informed investors require immediacy.

In order to appropriately measure the order book's liquidity across multiple price levels, the measure should capture dimensions with respect to both quantity over multiple levels (depth) and distances between the price levels (tightness). Estimating the order book slope by fitting a linear curve to the order book data to measure how price changes as a function of quantity (Malo and Pennanen, 2012) or how quantity changes as a function of price (Deuskar and Johnson, 2011; Härdle et al., 2012; Næs and Skjeltorp, 2006) is a popular approach to achieve this. Technically, Articles II and III follow Malo and Pennanen (2012).[5] One important aspect of the slope is that it can be made invariant for splits and comparable between different stocks and over time, enabling an aggregated analysis. Furthermore, one can estimate slopes separately on both bid and ask sides, enabling the study of the book's asymmetry.

Some other liquidity measures that incorporate Level II data also exist, such as the Exchange Liquidity Measure (XLM) that Gomber et al. (2015) use, which is based on the cost of a round trip. Both order book slope and XLM have two important advantages: liquidity is measured over multiple price levels, and liquidity can be measured on the bid side and ask side separately. However, XLM is determined for a specific trade size and the order book may not always be deep enough for calculating the XLM for a given order size because the total cumulative depths on the bid and ask sides vary in time. Especially around information arrivals the order books may be thin, because investors may be reluctant to keep their orders in the book because of the adverse selection risk. Because the liquidity measure used in Articles II and III should be available just around the announcement times, the order book slope is preferred as a multi-level liquidity measure because it is always possible to calculate it as long as the book is not empty. One other option would be to follow Engle et al. (2012); Erenburg and Lasser (2009); Riordan et al. (2013) and consider separately the depth at different price levels in the book; however, by considering the depth at different levels separately, the number of variables increases quickly, and the more that it does when asymmetry is also considered. Moreover, this methodology does not take into account the tightness dimension.

2.6 Information arrivals and (order book) liquidity

The use of multi-level order book data to measure liquidity is important because, as Gomber et al. (2015) point out, in liquid markets, restoring a small spread after a liquidity shock can be easy, although the book would remain thin, and, on the other hand, the depth available at the best levels can collapse momentarily, although the depth available at other levels would remain high. Especially around news arrivals, (informed) investors may want to take advantage of stale limit orders, and to do this they require immediacy; in this case, it is not possible to gradually execute a large block of trades to mitigate price impact. Consequently, investors may make single large trades or a bunch of simultaneous smaller trades that walk up the book through many price levels, necessitating the use of Level II data to capture the real announcement effects. To further motivate this point, Figure 2.2 plots the order book of Nokia around two scheduled announcement releases. It demonstrates that around information releases, investors may consume liquidity by trading with market orders and sweeping over multiple price levels, or they may cancel their limit orders or simply just not submit limit orders before the release of scheduled announcements to avoid adverse selection.

[5] See Malo and Pennanen (2012) and Articles II and III for details on order book slope estimation.

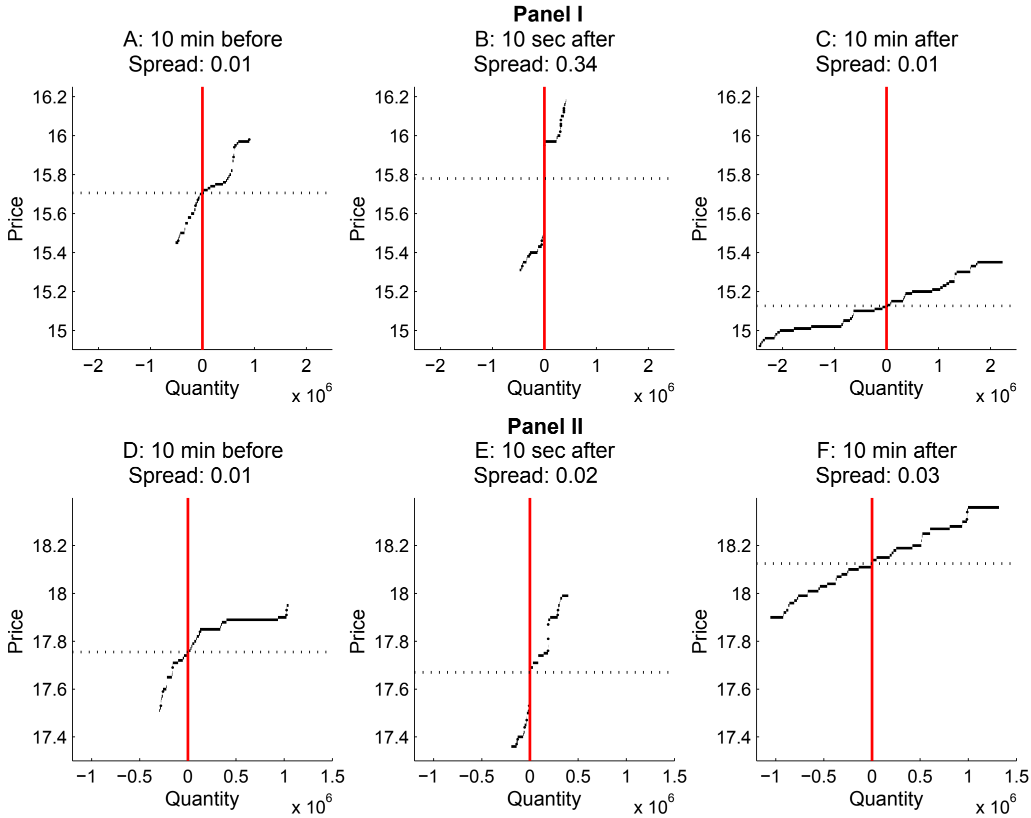

Figure 2.2: Limit order book around earnings announcements. Snapshots presenting the 20 best bid and ask levels in Nokia's limit order book around (scheduled) earnings announcements on 19.10.2006 at 12:01:40 in Panel I and on 19.4.2007 at 12:00:00 in Panel II. In Panel I (Panel II), sub-figure A (D) shows the situation 10 minutes before the announcement release; sub-figure B (E), 10 seconds after; and C (F), 10 minutes after the event. The left side of the red line at zero quantity contains the bids (i.e. bid orders are presented as a negative quantity), referred to as the bid side of the book, and the right side contains the asks (i.e. positive quantity), referred to as the ask side. The black dotted lines represent the mid-prices. The corresponding spreads appear above the sub-figures.

Graham et al. (2006) summarize the predictions of theoretical asymmetric information models on the market reactions before and after scheduled and non-scheduled information releases. According to them, traditional models (Glosten and Milgrom, 1985; Kyle, 1985) predict that the liquidity is low before scheduled announcement releases, and it reverts back to normal either quickly or gradually after the release. Before non-scheduled announcement releases, the liquidity is at a normal level (given that no informed trading is detected), and after the announcement, the liquidity is low before it returns to normal (unless the information processing is immediate).

Not too many studies have used Level II data to examine how information releases affect stocks' order book liquidity, but there are some important exceptions. Gomber et al. (2015) study Bloomberg ticker news and do not find a clear effect on order book liquidity; they write that this is perhaps because Bloomberg may not always be the first channel through which new information is distributed and the information content of the news may be limited. Riordan et al. (2013) study the impact of Thomson Reuters newswire messages on intra-day price discovery, liquidity, and trading intensity and use depth at multiple price levels (i.e. Level II data) as one liquidity proxy. They categorize the news as positive, negative, and neutral according to the tone of the news and find evidence of asymmetric reactions to the news. Riordan et al. (2013) find that liquidity measured by the spread increases around positive and neutral messages, whereas it decreases around negative messages; the depth, however, consistently increases around all types of news. Baruch et al. (2017) study how informed investors provide liquidity before positive and negative non-scheduled company announcements and show that when short selling is costly, when the magnitude of the news is small, or when the investor base is not broad, informed sellers do not expect competition and hence use limit orders before negative events, but in all other cases (positive announcements; high competition and negative announcements) informed traders use market orders prior to announcement releases.

Apart from these studies, Erenburg and Lasser (2009), Engle et al. (2012), and Rosa (2016) use Level II data with macro announcements and with data from the equity-index-linked securities market, the U.S. Treasury market, and futures market, respectively. All three studies use depth at multiple price levels, i.e. Level II information. Erenburg and Lasser (2009) observe that the bid−ask spread is higher than normal from around a minute before the announcement until three to four minutes after the announcement, and the depth beyond the best levels decreases within four minutes before and increases to the original levels within ten minutes after the release of announcements published during market hours. Erenburg and Lasser (2009) do not observe a significant decrease in depth at the best levels. Engle et al. (2012) find that the depths on multiple levels decline significantly within five minutes before the announcements, but they recover fast after the announcements. Rosa (2016) find that the depth beyond the best levels is abnormally low during the whole announcement day, hitting the intra-day low right before the announcement, whereas the spread increases in the minutes before the announcement release but returns to normal right after the announcement.

The empirical literature on liquidity around scheduled announcements using Level I data includes Graham et al. (2006); Krinsky and Lee (1996); Lee et al. (1993), among others. Most of the findings in the literature are in line with the predictions of traditional asymmetric information models concerning liquidity around scheduled announcements—in other words, before the release of scheduled announcements, the liquidity reduces (as the probability of informed trading increases), and after the announcements, the liquidity reverts back to normal either quickly or gradually (see Graham et al., 2006, Table 1).

In their empirical analysis, Graham et al. (2006) find wider spreads and lower depths after the release of non-scheduled news, indicating a transition period during which information is processed and incorporated into prices. Groß-Klußmann and Hautsch (2011) use news data from the Reuters NewsScope Sentiment Engine and exclude earnings announcements from their sample. Using data sampled every 20 seconds, they find that a high-frequency trading activity reacts significantly to company-specific news in a limit order market. They also conclude that the bid−ask spreads increase around the announcement times, but the depth at the best levels is not necessarily affected by the release. In their daily level analysis, Neuhierl et al. (2013) study different kinds of company releases and find that the bid−ask spread decreases after the announcement release for almost all news categories.

2.7 The FX market and liquidity aggregators

The FX market, in which investors buy and sell (i.e. exchange) currencies, is among the largest financial markets in the world, with an average daily traded volume of over $5 trillion (Bank for International Settlements, 2016). Currencies are traded on OTC markets, meaning that there is no formal exchange (see Harris, 2003; Hull, 2006; King et al., 2012). The FX market shows some unique characteristics because when compared with stock markets, the FX market has a high degree of decentralization, and compared with other OTC markets, it is extremely liquid with a high degree of electronic trading and a complex ecosystem of both bilateral and multilateral electronic trading platforms (Karnaukh et al., 2015; King et al., 2012; Mancini et al., 2013). Currently, FX markets are evolving rapidly in response to new trading technologies (King et al., 2013), and recent regulatory requirements, such as MiFID II[6], are expected to further push FX trading on platforms.

A typical modern trading technology used in the FX market is a liquidity aggregator (Oomen, 2017). An aggregator is an electronic trading tool that connects liquidity seekers (traders) with liquidity providers. The idea is to facilitate best-price execution: traders receive continuous streams of bid and ask quotes from selected (predefined) liquidity providers and can choose to trade with the best price. The logic of an aggregator is between the traditional RFQ protocol, in which a trader actively sends out requests for quotes for a number of liquidity providers asking prices for a transaction with a given quantity for a given currency pair (e.g. 1 Mio EURUSD), and a fully transparent all-to-all limit order book-style trading.

One of the key metrics in an aggregator is (a trader's) observed inside spread (Oomen, 2017). The observed inside spread is the difference between the best bid and ask prices that a trader observes at a specific moment (Oomen, 2017). An inside spread, as spread in general, is a measure of liquidity measuring the transaction costs. Table 2.1 shows a snapshot of what an investor with four streams in his/her aggregator setting could observe at a specific moment.

An important question for a trader is to determine how many streams and which streams to include in his/her aggregator setting because this also affects directly the trader's observed inside spread. As the liquidity provision in FX markets is highly bespoke and as it can be assumed that liquidity providers' behavior is fairly heterogeneous, the problem is highly non-trivial (see also Oomen, 2017). Liquidity providers usually provide different streams for certain traders (or subsets of traders), and, thus, while traders can choose with which liquidity providers they trade, they are not free to choose just any streams. Additionally, the selection of liquidity providers with whom a trader trades is further complicated by the fact that given the OTC nature of FX markets, two parties trading need a bilateral trading agreement and sufficient bilateral credit in order to trade (see also Gould et al., 2017).

Oomen (2017) provides a thorough discussion of liquidity aggregators and develops a theoretical model for liquidity dynamics in an aggregator. However, to my knowledge, there exists no prior research studying an aggregator empirically (which is the focus of Article IV). Other research focusing on liquidity in FX markets (but not in liquidity aggregators) includes those of Banti et al. (2012); Danielsson and Payne (2012); Gould et al. (2017); Karnaukh et al. (2015); Mancini et al. (2013); Payne (2003).

[6] See https://www.esma.europa.eu/policy-rules/mifid-ii-and-mifir.

Table 2.1: Snapshot of the most recent quotes observed by a trader with four streams (A, B, C, D). Each stream can have an arbitrary number of price and quantity pairs for both the bid and ask sides. The best bid and ask prices are highlighted with **bold**, and the trader's observed inside spread is calculated based on these: in this case, the observed inside spread equals $1.12545 - 1.12542 = 0.00003 = 0.3$ basis points (bp) (This table is from Article IV).

	Bid quotes		**Ask quotes**	
Stream	price	quantity [M]	price	quantity [M]
A	1.12513	1	1.12587	1
A	1.12512	2	1.12588	2
B	1.12536	0.5	**1.12545**	0.5
B	1.12535	1	1.12546	1
B	1.12533	3	1.12549	3
C	1.12536	0.5	1.12547	0.5
C	1.12534	1	1.12550	1
D	**1.12542**	1	1.12549	1
D	1.12539	2	1.12551	2
D	1.12536	3	1.12557	3
D	1.12532	5	—	—

3 Data

This chapter introduces the dataset used in Articles I–IV. Table 3.1 provides a short overview of the data sets used. Article I combines the shareholding registration record data described in Section 3.1 and the data collected from Nokia's Facebook page, described in Section 3.2. Articles II and III utilize stock limit order book data (Section 3.3) and data on company announcement releases (Section 3.4). Paper IV uses data from an FX liquidity aggregator, described in Section 3.5. In addition to these main data sets, some of the articles use supporting data sets, e.g. for the control variables, and these are described in the articles. Section 3.6 describes the processing of the data.

Table 3.1: Data sets used in Articles I–IV.

Data set	Time period	Asset(s)	Market	Article
Shareholding registration records	6/2010 – 12/2016	Nokia (stock)	NASDAQ Helsinki	I
Facebook	6/2010 – 12/2016	Nokia	—	I
Limit order book	1/2006 – 12/2010	75 stocks	NASDAQ Nordic	II, III
Company announcements	1/2006 – 12/2010	75 stocks	NASDAQ Nordic	II, III
FX aggregator	26.9.2016 – 7.10.2016	EURUSD	FX market	IV

3.1 Shareholding registration record data

Article I uses shareholding registration record data including all Finnish household and institutional investors from June 7, 2010 to the end of 2016, obtained from Euroclear Ltd.[1] The data set covers all the daily changes in over 280,000 Finnish investors' shareholdings in Nokia stock. Each record in the data contains information about the type, registration date, and volume related to the change in holdings, among other attributes. It also contains meta-data about the investor. In the first part of the analysis, Article I uses investors' sector codes to divide the investors into the following five groups: non-financial corporations, financial and insurance corporations, general governmental organizations, nonprofit institutions, and households (the categorization arises from the data).[2] In addition, to take a closer look at the trading of households, household investors are divided into activity groups based on their trading during the past two months.

[1] Data starting from April 12, 2010 are used to determine the activity levels of different investors.

[2] Grinblatt and Keloharju (2000, 2001); Lillo et al. (2015); Tumminello et al. (2012) use the same categories, with the exception that they also include the group of foreign investors. We exclude foreign investors from our analysis because a large percentage of them choose to use nominee registration (see also Grinblatt and Keloharju, 2000; Lillo et al., 2015), and for these investors, we can only observe aggregate behavior and cannot distinguish between unique traders.

Baltakys et al. (2018); Grinblatt and Keloharju (2000, 2001); Lillo et al. (2015); Tumminello et al. (2012) use the data sets from the same source, and they provide detailed descriptions of the data. However, one should note that they use data from before 2009, when all transactions were reported separately, while after moving to Central Counterparty Clearing in late 2009, the Euroclear research data set contains only aggregated daily trades without specifying the actual trading dates: instead, a registration date is associated with every record.

Thus, the data set reports a registration date for each change in shareholdings, which is not the same as the actual trading date. To analyze the contemporaneous relationship between investors' trading activities and the activities on Nokia's Facebook page, the trading dates are reverse engineered from the registration dates. The official T+3 settlement convention is used for shareholdings registered before and on October 8, 2014, and T+2 is used afterwards (see Euroclear, 2014). Using the derived trading dates, transactions are aggregated on a weekly basis. Weekly aggregation will reduce the possible noise of inaccurate trading date derivation.

3.2 Facebook data

The Facebook data include all posts and related comments, likes, and shares from Nokia's Facebook wall[3] between June 5, 2010 and December 31, 2016. The data were collected using the Social Data Analytics Tool (SODATO) (see Hussain and Vatrapu, 2014a,b; Hussain et al., 2014). Aggregated at the daily level, the data show how many posts Nokia made on its Facebook wall on a given day, as well as how many comments, likes, and shares these posts received. Note that comments, likes, and shares are always related to a specific post—in other words, the post is the main action. Therefore, the numbers of comments, likes, and shares are assigned to the date of the original post—that is, not the date when the actual comment, like, or share was made. In effect, the numbers of comments, likes, and shares are indicative of the importance of the post released on a particular day in the data.

The daily Facebook data are aggregated to weekly by summing the numbers of posts, comments, likes, and shares during a week. In total, the sample comprises 342 weekly observations for posts, comments, likes, and shares. The week is taken to begin on Saturday and end on Friday because trading does not occur on weekends. This way, the Facebook activity on weekends is related to the week in which it can actually affect investors' trading decisions.

3.3 Limit order book data

The Level II limit order book data in Articles II and III are from NASDAQ OMX Nordic, which are continuous limit order-based markets.[4] The limit order book data consist of

[3] https://www.facebook.com/nokia.

[4] We use data from Nordic markets instead of US markets (the most liquid in the world) because the former are less fragmented compared with the latter. In the US, fragmentation is clearly an important feature of equity markets (O'Hara and Ye, 2011); the limit orders for a given asset are spread between several exchanges, thus posing a problem for empirical research—in particular, matching rules and transaction costs complicate the comparisons between different limit order books for the same asset (Gould et al., 2013). Furthermore, as Butt and Virk (2015) argue, another advantage of using data from less liquid Nordic markets is that "it is more appropriate to test liquidity-related models in markets that are sufficiently illiquid to diagnose the level and strength of bearing [...] risks".

snapshots of the limit order books representing quotes on the 20 best bid and ask price levels, and the data are sampled every 10 seconds. Articles II and III use the Level II order book data from the beginning of 2006 until the end of 2010.

The data sample consists of 75 frequently traded stocks from the Helsinki, Stockholm, and Copenhagen Stock Exchanges. The stocks included have been involved in OMX Helsinki 25, OMX Stockholm 30, or OMX Copenhagen 20 stock indexes at some point. Out of the total 75 stocks, 27 are traded in Helsinki, 28 in Stockholm, and 20 in Copenhagen Stock Exchange. The list of stocks included is available in Appendix A.

3.4 Company announcements

The news data used in Articles II and III come from NASDAQ OMX Nordic's website.[5] The announcements included are from the beginning of 2006 until the end of 2010, and the respective companies filed them with NASDAQ OMX. The announcement times are reported at one second precision, but they are rounded to the nearest 10 seconds because the order book data are sampled every 10 seconds.

Articles II and III do not restrict the study to any specific news class, such as earnings announcements, as many other studies do. Rather, the announcements are re-categorized into two specific groups: scheduled and non-scheduled announcements. An announcement is classified as scheduled if the exact publishing date is known to the public beforehand. This happens if the date is given in advance in earlier stock exchange releases (e.g. in the financial calendar). For some of the scheduled announcements, the exact (intra-day) time of the announcement is known beforehand, but these announcements are not analyzed separately because distinguishing between the two cases reliably is not always possible.

An announcement is classified as non-scheduled if one may assume that external stakeholders do not know that the announcement is going to be released or when it will be released. In particular, a release is considered non-scheduled if it is irregular, its publishing schedule is not given and cannot be reliably estimated, or the release is obviously unexpected. Announcements whose publishing timespan is given non-specifically in earlier stock exchange releases or that are somewhat regular by nature, such as proposals to annual general meetings by the board or nomination committee, are excluded from the sample to be on the safe side. Most of the excluded announcements are notices to convene annual general meetings, notices of the publication of annual reports or summaries, and invitations to press conferences related to publishing financial reports.

The final sample includes 408 (329) scheduled announcements and 2,629 (2,102) non-scheduled announcements with 30 (60)-minute pre- and post-event windows. Just over 35% of the announcements originate from NASDAQ OMX Helsinki, around 45% are from NASDAQ OMX Stockholm, and a bit under 20% come from NASDAQ OMX Copenhagen. In the final sample, over 70% of the scheduled announcements are financial announcements. Appendix A gives the number of announcements per company and additional information on filtering the news included in the study.

[5] http://www.nasdaqomxnordic.com/news/companynews, see this page also for detailed information.

3.5 FX aggregator data

Paper IV analyzes quote data for EURUSD from an FX liquidity aggregator.[6] The sample period covers 10 trading days from September 26, 2016 to October 7, 2016. The data set includes information on all quote updates for all liquidity streams by the liquidity providers operating in the aggregator: there are, on average, more than 30 million quote updates per day. Additionally, the data set includes information on all transactions in the aggregator, including information on both counterparties: (anonymised) trader ID and stream ID. Constructing a snapshot of the quotes that a specific trader observes at any point in time (a trader's "personal order book") is possible using this information.

Article IV uses information on the best bid and ask prices quoted by each liquidity stream, sampled every minute at 6:00 − 16:00.[7] Additionally, for each trader in the aggregator, the trader's observed inside spread is determined every minute by using the information on the liquidity streams that the trader observes and the best bid and ask prices quoted by these streams. The set of streams a trader observes is approximated with the set of streams that the trader has traded with; in reality, therefore, the number of observed streams may be higher, as there may be streams from which a trader receives quotes but with which he/she did not trade during our sample period.

3.6 Data processing

A substantial proportion of the time I used in conducting the research related to this dissertation project was spent on processing the data. For Article I, my co-authors provided me with the weekly observations for the shareholding registration data and the daily observations for the Facebook data. The shareholding registration data were created by filtering and aggregating the numbers of investors changing their positions. The Facebook data were obtained by scraping all the posts, comments, likes, and shares from Nokia's Facebook wall and transforming the data into tabular format by summing the daily numbers of the Facebook activities (using SODATO, see Hussain and Vatrapu, 2014a,b; Hussain et al., 2014). Then, I aggregated the Facebook data to the weekly level and calculated the supporting data for the control variables. Finally, I converted and combined the data sets to the format required by the analysis.

I already calculated the order book liquidity measure from the limit order book data as a part of my Master's thesis research (see Salo, 2014). First , the order book state data were sampled every 10 seconds, and second, the slopes for the bid and ask sides were calculated using linear regression. This was a huge task, as the sampling frequency was high, the sample period was four years long, and there were 75 stocks included in the study. In addition, for Articles II and III, I calculated various other measures needed for the analysis. The company announcement data for Articles II and III were scraped from NASDAQ Nordic's webpage by my co-author. The announcements were sorted out into scheduled and non-scheduled ones by a research assistant according to my instructions and under my close supervision. Then, I combined the data sets and converted the data to a suitable format for running the analysis.

For Article IV, I handled the data in collaboration with big xyt AG. In particular, the data provider gave the data to my co-author at big xyt AG, and they imported the data

[6] We are not allowed to disclose any detailed information on the data provider.

[7] The analysis is restricted to data from 6:00 − 16:00 because around 90% of the trading (measured both by the traded volume and the number of trades) takes place during these hours.

to their database. Then, I checked and confirmed the data quality and wrote the codes to access the data before transforming them into a suitable format for calculating the variables needed. Next, I wrote the codes to extract and calculate the variables for the actual analysis. During this process, I also received occasional support from big xyt AG in executing the codes as efficiently as possible. I likewise had many discussions with the company providing the data to ensure that I understand the data correctly.

4 Methods

This chapter describes the research methods used in Articles I−IV. Articles I and III use regression analysis: Article I uses logistic regression, and Article III utilizes linear regression, both discussed in Section 4.1. Article II utilizes the framework from the event analysis method, presented in Section 4.2. Article IV exploits a GA, described in Section 4.3, to solve an optimization problem.

4.1 Regression analysis

Multiple linear regression analysis is a general statistical technique used to study the relationship between a dependent (explained) variable and several independent (explanatory) variables (Greene, 2012; Hair et al., 2010). Regression analysis can be used to study a wide variety of research problems (Hair et al., 2010), and it is widely applied in finance literature (see e.g. Degryse et al., 2015; Graham et al., 2006; Grinblatt and Keloharju, 2001; Kaniel et al., 2008).

In Article III, linear regression is used to investigate the factors related to the magnitude of the illiquidity shock following announcement releases. The data are pooled on a cross-section of stocks. To remove the stock-level effects, within-transformation (also called fixed effects transformation) is used (see Baltagi, 2013). In within-transformation, each variable is demeaned by subtracting its within-individual average over time (Baltagi, 2013).[1]

In Article I, the dependent variable is binary: there are two possible outcomes, either an investor (who changed his/her holdings on Nokia stock) increased or decreased his/her holdings. The logistic regression model is the most frequently used regression model for this kind of discrete data (Hosmer Jr et al., 2013), and, hence, it is applied in Article I. Logistic regression utilizes the logit link function, but two other common choices when working with binary data exist, which are the probit and complementary log-log link functions (Faraway, 2005). Logit link is preferred in Article I because it provides a simple interpretation of the regression results with odds-ratios (see e.g. Faraway, 2005).

4.2 Event study

In finance research, event study analysis is used to measure the effect of some firm-specific or economy-wide event on stock returns (see e.g. Campbell et al., 1997; Kothari

[1] Additionally, for the data in Article III, random effects models are fitted and compared with the corresponding fixed-effect models by using the Hausman test (see e.g. Baltagi, 2013). As the test indicates that in some cases (with non-scheduled announcements), the estimated parameters may be biased, Article III reports the results using within-transformation to be on the safe side. However, the results seem consistent with those obtained using random effects models (not reported in Article III).

and Warner, 2007). The method is widely applied, Kothari and Warner (2007) report that between the years 1974 and 2000, there were 565 event studies published in five leading finance journals. For example, researchers have studied mergers and acquisitions (see e.g. Mentz and Schiereck, 2008; Moeller and Schlingemann, 2005), as well as layoff announcements (see e.g. Chalos and Chen, 2002; Velásquez et al., 2018). The basic idea is to determine the effect of an event on returns by comparing the observations in an event window (potentially affected by the event) with the observations in an estimation window (assumed to represent normal times without the event) (see, e.g. Campbell et al., 1997): Figure 4.1 illustrates this. When aggregating over the events, the analysis shows whether the event leads to abnormal returns, i.e. whether the returns in the event windows are substantially higher or lower than those in the estimation windows.

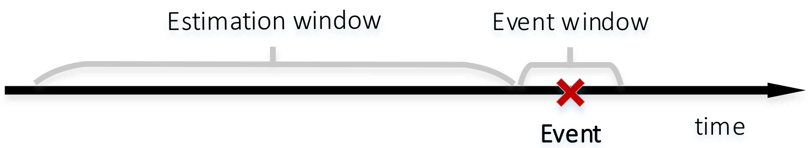

Figure 4.1: Timeline of an event study (adapted from Campbell et al., 1997)

However, Article II focuses on the effect of announcement releases on order book liquidity and not on the effect on returns as traditional event studies. Nevertheless, event study methodology provides a natural framework for the study, as order book liquidity around announcement releases (i.e. in the event window) is compared with normal order book liquidity (i.e. order book liquidity in the estimation window). Furthermore, Gomber et al. (2015) utilize the event study approach to examine liquidity. Another difference when compared with more traditional event studies is that Article II uses high-frequency intra-day data, whereas event studies often utilize daily-level data (see e.g. Chalos and Chen, 2002; Mentz and Schiereck, 2008; Moeller and Schlingemann, 2005, and discussion in Velásquez et al., 2018).

4.3 GA for optimization

The problem in Article IV involves selecting the set of liquidity streams that minimizes the trader's average observed inside spread, given the number of liquidity streams in the trader's aggregator setting. This optimization problem is combinatorial and non-linear, and solving it exactly using brute force, i.e. computing all the possible solutions and choosing the best one, is not feasible. A GA is an optimization method that can be applied to a variety of problems, including discrete combinatorial problems (Sivanandam and Deepa, 2008). Hence, it is a natural choice for solving the complex optimization problem in Article IV.

A GA is a metaheuristic optimization method[2] inspired by natural selection and survival of the fittest (see Holland, 1992). It is the most popular technique in a broader category of

[2] A *deterministic* algorithm finds an exact solution to a problem, but in practice, many problems are too complex to be solved exactly within a reasonable computing time; a *(meta)heuristic* method, on the other hand, provides a sufficiently good solution within a reasonable computing time (Schlottmann and Seese, 2004). Because a GA is a metaheuristic method, there is no guarantee that the global optimum is reached, but the analysis in Article IV shows that the best combinations found are indeed sufficiently good for the purposes of the analysis.

evolutionary computation methods (Sivanandam and Deepa, 2008). Nowadays, GAs are used to solve many optimization problems because they are easy to apply, and they are robust in finding good solutions to difficult problems (Contreras-Bolton and Parada, 2015; Sivanandam and Deepa, 2008). In the field of economics and finance, GAs have been used to solve problems related, for example, to portfolio selection and risk management (see e.g. Schlottmann and Seese, 2004, and references therein).

The basic idea of GA is intuitive—Figure 4.2 describes this. At the beginning, an initial population is generated as a random sample of possible solutions, and a fitness value is calculated for each individual solution in the population. Then, a new population is generated through crossover (combining two parent solutions; the higher fitness value makes an individual more likely to be a parent) and mutation (applying a random change in an individual solution). Next, a fitness value is calculated for each individual in the new population. Then, the situation is checked against predefined stopping criteria, for example, if the average improvement in the best fitness value is less than a given limit or a maximum number of generations or a time limit is reached. If at least one of the stopping criteria is satisfied, the algorithm halts; otherwise a new generation is created (for more information, see e.g. Sivanandam and Deepa, 2008).

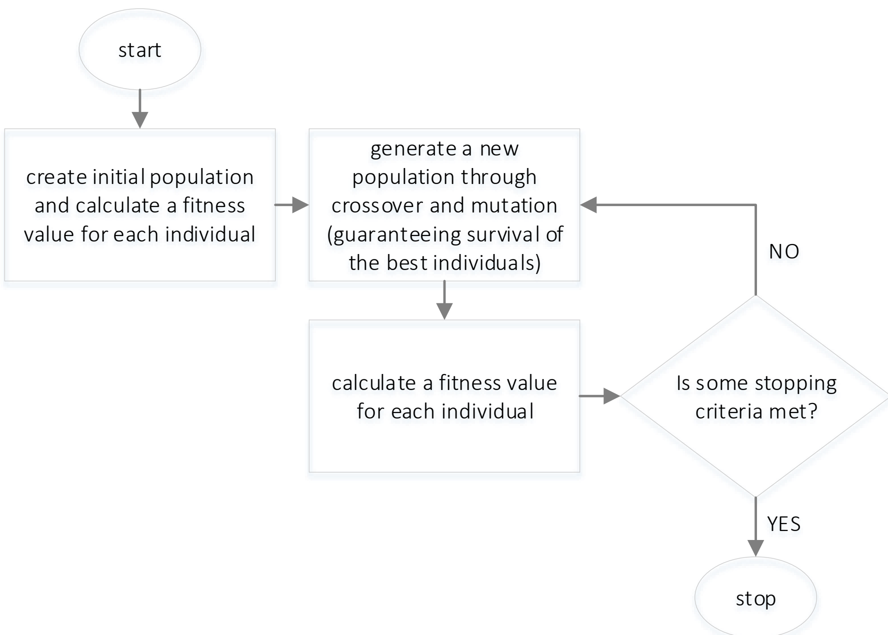

Figure 4.2: Flowchart of a genetic algorithm (adapted from Sivanandam and Deepa, 2008).

5 Findings

This chapter summarizes the findings of Articles I−IV. In addition, Section 5.5 recaps the research questions of this dissertation and highlights the answers that Articles I−IV find. More detailed findings can be found in the articles attached to this dissertation.

5.1 Facebook drives behavior of passive households in stock markets

Siikanen et al. (2018a) (Article I of this thesis) examine the relation between investors' decisions and Facebook data. In particular, given that an investor trades, Siikanen et al. (2018a) study whether Facebook data are related to an investor's decision to increase or decrease his/her position on Nokia stock. In the first part of the analysis, investors are divided into the following five categories: companies, financial (and insurance) institutions, governmental organizations, non-profit organizations, and households. In the second part, households are divided into four categories based on their trading activity.

Siikanen et al. (2018a) give the first empirical evidence that Facebook activities affect the trading of different investors differently. The analysis shows a clear association between Facebook data and the decisions of (especially passive) household investors and non-profit organizations. By contrast, there is clearly no association between the decisions of financial institutions and Facebook data. Given that finance and insurance institutions can be viewed as the most sophisticated domestic investor group (see Grinblatt and Keloharju, 2000, and the discussion in Section 2.1), the findings indicate that more sophisticated investors are more independent of Facebook activities, whereas less sophisticated investors may rely more on the information communicated via Facebook. Assuming that an investor's activeness is related to his/her sophistication, the findings on household activity groups support this result (Grinblatt and Keloharju, 2000, show that larger household investors are more sophisticated than smaller ones).

Facebook is not a regulated information channel, and, hence, companies are likely to strategically select the information disseminated via this platform (Jung et al., 2017). Thus, the findings of Siikanen et al. (2018a) suggest that the decisions of less sophisticated investors may be driven by biased information. In line with this view, Ammann and Schaub (2017) find that postings on a social trading platform, which do not contain value-relevant information, affect the trading decisions of unsophisticated investors.

5.2 Limit order books and liquidity around scheduled and non-scheduled announcements

Siikanen et al. (2017a) (Article II in this thesis) studies the dynamics of order book liquidity around scheduled and non-scheduled company announcement releases. The aggregated results depict significant changes in order book liquidity during a two-hour event window around a company announcement arrival. The announcements are followed by an immediate liquidity shock (within a few minutes after the announcement release). Before scheduled announcement releases, the order book liquidity is abnormally low on both bid and ask sides already an hour before the announcement release. This indicates that the order books (investors) adjust to the scheduled announcement releases already (at least) an hour before the release: many investors may either cancel their orders or choose not to submit new limit orders. Therefore, informed investors have limited possibilities of taking advantage of stale limit orders prior to scheduled information releases. After scheduled announcement releases, the aggregated order book liquidity recovers to an exceptionally good level, which may indicate that the release of scheduled announcements reduces information asymmetries and adverse selection costs significantly. The finding on reduced liquidity prior to scheduled announcement releases is in line with the predictions of the traditional asymmetric information models of liquidity (see Table 1 in Graham et al., 2006, and the discussion in Section 2.6); however, the models predict that the liquidity returns back to normal after the announcement release, whereas the analysis of Siikanen et al. (2017a) indicates that liquidity can revert to an exceptionally good level.

Additionally, the results show some statistically significant pre-reactions to non-scheduled announcement releases, especially on the ask side, although the pre-reactions are clearly weaker than those for scheduled announcements. According to traditional models, these pre-reactions can indicate information leakage (see Graham et al., 2006). In contrast to that after scheduled announcement releases, the order book liquidity is still abnormally low an hour after the non-scheduled announcement releases in most of the cases, indicating that it takes a relatively long time for the markets to adjust to unanticipated information arrivals (see Graham et al., 2006).

Additionally, the analysis in Siikanen et al. (2017a) shows that the order book liquidity (Level II liquidity measure) behaves quite differently around company announcement releases when compared to the spread (Level I liquidity measure). Furthermore, the analysis of order book asymmetry indicates that the news-momentum (news-contrarian) trading strategy through market orders is relatively expensive (inexpensive) after non-scheduled announcements, whereas scheduled announcements decrease the transaction costs for both large buy and sell trades: the book remains symmetric, meaning that after scheduled company announcements, both news-contrarian and news-momentum trading strategies are symmetrically inexpensive.

5.3 What drives the sensitivity of limit order books to company announcements?

Siikanen et al. (2017b) (Article III in this thesis) analyze the factors affecting the magnitude of the illiquidity shocks caused by scheduled and non-scheduled announcement releases. In particular, the effects of several variables related to pre-reactions in the order book, the sign (positiveness/negativeness) of the announcement releases, and the market's reaction times are studied.

The results show that recent losses amplify the illiquidity shocks following non-scheduled announcement releases. This seems consistent with the results of Hameed et al. (2010), who document that negative market returns decrease liquidity. At the same time, there is no such association with scheduled announcements; therefore, this result may help clarify investors' reactions to the arrival of unexpected news: recent losses can make investors' liquidity provision more sensitive to surprising news arrivals.

Moreover, the results show that a fast reaction is a strong reaction: the faster the limit order book illiquidity reaches its maximum after both scheduled and non-scheduled announcement releases, the more illiquid the order book becomes. This could indicate that if there is no immediate illiquidity shock following the announcement, it is likely that there is no (strong) liquidity shock. Siikanen et al. (2017b) also provide evidence that asymmetry between the bid and ask sides of an order book before announcement releases is positively associated with the magnitude of illiquidity shocks, especially for bid side liquidity. This seems to be in line with the finding of Chordia et al. (2002) that order imbalances in either direction reduce the liquidity of the aggregate market. Additionally, positive announcement releases cause larger illiquidity shocks on the ask side and smaller shocks on the bid side and vice versa for negative announcements. This seems reasonable because informed investors may buy (sell) shares by picking off stale sell (buy) limit orders just after the arrival of new positive (negative) information.

5.4 Liquidity in the FX market

Siikanen et al. (2018b) (Article IV in this thesis) provides empirical evidence of the liquidity observable by individual investors in an FX liquidity aggregator. The analysis shows that investors observe, on average, 5.4 streams out of the total 165 active streams quoted by the 42 liquidity providers in the aggregator (the maximum is 23 and the minimum is 1). Additionally, investors observe relatively tight inside spreads already with four or five streams in their aggregator setting, and having more streams leads to just marginal benefits, if any.

Given the number of streams in an investor's aggregator setting, an optimal combination of streams is determined to minimize the investor's observed inside spread. Comparisons of the observed spreads with the optimal ones show that most of the investors could cut their observed spread in less than half while keeping the number of streams the same, given that they could obtain the optimal combination of streams. Moreover, if investors could obtain the optimal combination of streams, they would save \$0.13–\$0.18 basis points per €1 traded. However, fully exploiting the cost savings and reductions in observed spreads may not be possible for traders because they are not free to choose just any streams quoted by the liquidity providers, and changes in the traders' aggregator settings would likely change the quoting behaviors of the liquidity providers.

Finally, Siikanen et al. (2018b) calibrate the model by (Oomen, 2017) to the empirical data. The analysis shows that although the model is studied under a quite simplistic assumption of homogeneous liquidity providers, the model can be accurately fitted to real-world data, and, hence, it can be used to describe the liquidity in aggregators.

5.5 Summary of the findings and review of the research questions

RQ I of this dissertation asks whether the information that companies release on social media affects the trading decisions of investors. To answer this question, Article I studied data collected from Nokia's Facebook wall, combined with data on the holdings of all Finnish investors on Nokia stock, and conclude that there is an association between the Facebook data and the buy versus sell decisions of some investors, but not all. In particular, the decisions of arguably less sophisticated investors—passive households and non-profit institutions—are associated with Facebook data, whereas more sophisticated investors—financial institutions—seem to behave independently of the Facebook data.

RQ II concerns how liquidity in limit order books evolves around scheduled and non-scheduled company announcement releases. Article II answers this question by studying the limit order books of 75 stocks from NASDAQ Nordic around scheduled and non-scheduled company announcement releases. The order book liquidity is abnormally low before scheduled announcement releases, and it is somewhat lower than normal before non-scheduled announcements. Both announcement types are followed by immediate illiquidity shocks. After scheduled announcement releases, the aggregated liquidity recovers to an abnormally high level, whereas the liquidity after non-scheduled announcements remains relatively low at least an hour after the announcement release in most cases.

RQ III explores the factors that affect the magnitude of order book liquidity shocks caused by scheduled and non-scheduled announcements. Using the same data as those in Article II, Article III finds that recent losses amplify the shocks after non-scheduled announcements, a larger asymmetry right before the announcements leads to larger shocks, and a fast reaction is a strong reaction.

RQ IV focuses on liquidity in aggregators, exploring the proportion of liquidity observed by investors and how much the spreads could be improved by selecting the liquidity streams optimally. By using a unique data set from an FX aggregator, Article IV finds that investors observe, on average, 5.4 streams out of the total 165 streams, and that most of the investors could cut their observed spread by more than half with the same number of streams they currently observe, given that they would be able to observe the optimal combination of streams.

6 Conclusion

This chapter provides the concluding remarks for this dissertation. First, the contribution of the dissertation is discussed, and, second, the reliability and validity of the research are assessed. Finally, some limitations and suggestions for future research are presented.

6.1 Contribution

This dissertation contributes to the empirical literature on financial markets related to three key elements of these markets: investors, information arrivals, and liquidity. Combined, Articles I−IV address several research gaps on the interrelations of these topics. A better understanding of these topics can help researchers develop more reliable and robust models related to the market's micro structure and investor behavior; on a more general level, exchange organizations, regulations, and investment management can all be benefitted by this better understanding (see Amihud and Mendelson, 1991; Chordia et al., 2001; Grinblatt and Keloharju, 2000, 2001; O'Hara, 1995). Additionally, the new empirical findings could aid researchers in planning new empirical studies by indicating aspects that should be considered, for instance, highlighting the importance of separating investors into categories, making the distinction between scheduled and non-scheduled announcement releases, and measuring liquidity also deeper in the order book.

While Lillo et al. (2015) study how Reuters news articles affect the trading decisions of different types of investors, Article I contributes by providing the first evidence that the information companies release on social media affects the behavior of different investors in stock markets differently. In particular, the findings indicate that the decisions of passive household investors and non-profit institutions to buy versus sell are associated with Facebook data, whereas the behavior of more sophisticated investors—financial institutions—seems to be independent of Facebook data. One implication of the result could be as follows: as information communicated via social media affects the behavior of certain investors, companies should be careful in releasing information on social media, especially because there is no detailed research yet on how social media activities affect trading behavior. On the other hand, as Jung et al. (2017) show that companies disseminate information on social media strategically, such information may be biased, and, thus, less sophisticated investors should be cautious when relying on this information in investment decision making.

Article II contributes by using data on scheduled and non-scheduled first-hand official company announcements and by finding clear effects of announcement releases on limit order book liquidity in stock markets. This is a novel finding given its strong contrast to that in the work of Gomber et al. (2015), who find no evidence that Bloomberg ticker news would cause liquidity shocks to order books. The findings provide evidence related to the asymmetric information models of liquidity (see Graham et al., 2006): scheduled announcement releases seem to resolve the asymmetric information problem (the order book liquidity is abnormally high an hour after scheduled announcement releases), but an hour after non-scheduled announcement releases, in many cases, the liquidity remains relatively low, indicating relatively long information processing times. As liquidity measurement is also an important topic from the regulatory perspective (Amihud and Mendelson, 1991; Chordia et al., 2001; Cumming et al., 2011)[1], the evidence shown in Article II that liquidity measured over multiple price levels (liquidity for larger transactions) may differ from top-of-the-book liquidity (liquidity for relatively small trades) is valuable.

The contribution of Article III is the identification of the factors that magnify the liquidity shocks following scheduled and non-scheduled announcement releases. Article II and some other studies (Engle et al., 2012; Erenburg and Lasser, 2009; Riordan et al., 2013; Rosa, 2016) show that announcement releases may cause liquidity shocks to the markets, but the factors affecting the magnitude of the liquidity shocks related to information arrivals in stocks markets have not been studied in depth. A better understanding of these factors may help liquidity traders avoid trading during periods of low liquidity and could potentially help exchange operators and regulators plan policies to prevent disruptions in order book liquidity.

The novelty of Article IV is its use of a unique and detailed data set and the insights derived from it. Although Oomen (2017) provides a thorough discussion liquidity aggregators and develops a model for liquidity provision in an aggregator, to my and my co-authors' knowledge, Article IV is the first study to provide empirical evidence of liquidity in an FX aggregator. As FX market liquidity is important from many perspectives, for instance, it is needed for international trade and it is crucial in guaranteeing the efficiency of other markets, too (Karnaukh et al., 2015; King et al., 2012; Mancini et al., 2013), the empirical evidence that Article IV provides is valuable for all researchers, practitioners, and regulators. Article IV also contributes to the literature by answering the call of Gould et al. (2017) for literature studying the subset of liquidity that individual traders observe in the FX market. In particular, the findings imply that traders in aggregators could try to improve the observed spread by changing the aggregator setting, and regulators could perhaps consider the optimality and efficiency of this kind of trading technology (see also Black, 1971).

6.2 Reliability and validity of the research

Assessing the reliability and the validity of a study is important in estimating the rigor of the research process and the trustworthiness of the findings. Reliability relates to consistency, in a sense that it asks whether we did things correctly; validity, on the other hand, relates to that we did correct things, in the first place, to answer the research questions (see e.g. Heale and Twycross, 2015; Roberts et al., 2006).

For the research to be reliable and valid, the data used must be of high quality. In this dissertation, the data quality is ensured by obtaining the data from official sources. For Article I, the shareholding registration data are acquired from Euroclear Ltd.; the data should be highly reliable, as electronic records represent official certificates of ownership (see Grinblatt and Keloharju, 2000). For Articles II and III, the limit order book data are

[1] For the MiFID II perspective, see e.g. https://www.emissions-euets.com/internal-electricity-market-glossary/805-liquidity.

acquired directly from NASDAQ Nordic, and the liquidity aggregator data used in Article IV are obtained directly from the market place operator. All data handling processes are done carefully to ensure and maintain the high quality of the data.

Furthermore, this research is conducted using widely established statistical methods. Regarding the reliability of the results, multiple robustness checks are run to ensure that the results are not sensitive to small modifications in the test settings. Articles II and III repeat the analysis using both 30- and 60-minute pre- and post-event windows, and Article III additionally runs the analysis for mean values instead of medians. Article II uses both statistical and visual analytics to study the behavior of order book liquidity. The additional analysis in Article I on households' activity groups can also be viewed as a robustness check—on the other hand, analyzing the numbers of comments, likes, and shares that the posts receive, in addition to the plain numbers of posts, serves as a robustness check, too. Moreover, in Article IV, the GA is run 10 times for each aggregator setting with different numbers of liquidity streams to ensure that the results do not rely on the result of just one optimization run.

Furthermore, Articles I−III are published in high-quality journals, further confirming the validity of the research: the articles have been peer reviewed and revised before their publication. The measures used in the articles have also been selected carefully after consultation of prior literature to ensure that they measure the subjects of interest.

6.3 Limitations and suggestions for future research

Many of the limitations of this research are related to the data. Article I studies only one company, Nokia, and this may pose a clientele bias: investors interested in Nokia may generally be more social media and technology savvy, so they likely follow Facebook. It is unclear if the results are generalizable to other companies and other markets. In future research, it would be interesting to include more companies from different industries in the analysis. However, extending such a detailed analysis to other markets may be complicated, as detailed data are not available for most of the markets. In addition, extending the analysis to the semantics of the posts and comments, in addition to the numbers, could shed more light on the question of how information communicated via social media affects the decisions of different investors.

For Articles II and III, the sample time period overlaps with the financial crisis of 2008, and this may have had an impact on the results, although Chordia et al. (2001) argue that while liquidity levels may vary with market trends, the day-to-day changes in liquidity are probably the same in most environments. Furthermore, the proportion of (high-frequency) algo trading started to rise in NASDAQ Nordic around 2010 (NASDAQ OMX, 2011), and it would be interesting to see if this has changed the liquidity dynamics around information arrivals. Another limitation is that Articles II and III utilize multi-level order book state data but not the complete order flow data (ITCH feed), which would allow a detailed analysis of trades, order submissions, and cancellations. Future research using ITCH feed data could give a more complete picture of liquidity making and taking around the announcement releases and provide evidence of what causes the changes in order books around the announcement releases (trades, cancellations, or low limit order submission rates).

The results in Article II show that the spread may behave quite differently from the order book liquidity over multiple price levels; correspondingly, the order book liquidity over 20 price levels (which is used in Articles II and III) may differ from the order book liquidity over, for example, 5 or 10 order book levels. Future research could study how the number of order book levels studied affects the results. Furthermore, Articles II and III focus on the liquid stocks of relatively large companies, and the results for less liquid stocks and smaller companies may differ substantially, for example, because of analyst coverage (see e.g. Irvine, 2003). Additionally, **RQ3** seeks to determine the factors that affect the magnitude of order book liquidity shocks, but Article III focuses on studying only a limited set of variables calculated from order book data; naturally, there are probably many other factors affecting the magnitude of liquidity shocks. Another related research topic would be to examine the factors affecting the time it takes for the order book to return to the normal levels, i.e. the factors related to the resiliency of the order books.

As for Article IV, the sample period is relatively short and covers only one currency pair. After these initial efforts, more research covering longer time periods and other currency pairs is definitely needed to enhance the understanding of liquidity provision in aggregators, although the availability of data may hinder the work of researchers pursuing this. Comparisons of liquidity in different market places (as the characteristics of traders and liquidity providers may have an effect) and with different trading technologies are also needed to understand liquidity in FX markets and to ensure efficient trading and regulation that supports it.

Bibliography

Admati, A. R. and Pfleiderer, P., “A theory of intraday patterns: Volume and price variability,” *The Review of Financial Studies*, vol. 1, no. 1, pp. 3–40, 1988.

Aitken, M. and Comerton-Forde, C., “How should liquidity be measured?” *Pacific-Basin Finance Journal*, vol. 11, no. 1, pp. 45–59, 2003.

Amihud, Y. and Mendelson, H., “Liquidity, asset prices and financial policy,” *Financial Analysts Journal*, vol. 47, no. 6, pp. 56–66, 1991.

Ammann, M. and Schaub, N., “The impact of internet postings on individual investors,” in *proceedings American Finance Association 2018 Annual Meeting*, 2017.

Baltagi, B., *Econometric analysis of panel data*, 5th ed. John Wiley & Sons, 2013.

Baltakys, K., Kanniainen, J., and Emmert-Streib, F., “Multilayer aggregation with statistical validation: Application to investor networks,” *Scientific reports*, vol. 8, no. 1, p. 8198, 2018.

Bank for International Settlements, “Foreign exchange turnover in 2016, Triennial Central Bank Survey,” 2016.

Banti, C., Phylaktis, K., and Sarno, L., “Global liquidity risk in the foreign exchange market,” *Journal of International Money and Finance*, vol. 31, no. 2, pp. 267–291, 2012.

Barber, B. M. and Odean, T., “All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors,” *The Review of Financial Studies*, vol. 21, no. 2, pp. 785–818, 2007.

Baruch, S., Panayides, M., and Venkataraman, K., “Informed trading and price discovery before corporate events,” *Journal of Financial Economics*, vol. 125, no. 3, pp. 561–588, 2017.

Biais, B., Hillion, P., and Spatt, C., “An empirical analysis of the limit order book and the order flow in the paris bourse,” *the Journal of Finance*, vol. 50, no. 5, pp. 1655–1689, 1995.

Black, F., “Toward a fully automated stock exchange, part i,” *Financial Analysts Journal*, vol. 27, no. 4, pp. 28–35, 1971.

———, “Noise,” *The journal of finance*, vol. 41, no. 3, pp. 528–543, 1986.

Bloomfield, R., O'Hara, M., and Saar, G., "The "make or take" decision in an electronic market: Evidence on the evolution of liquidity," *Journal of Financial Economics*, vol. 75, no. 1, pp. 165–199, 2005.

Bollen, J., Mao, H., and Zeng, X., "Twitter mood predicts the stock market," *Journal of computational science*, vol. 2, no. 1, pp. 1–8, 2011.

Brennan, M. J., Huh, S.-W., and Subrahmanyam, A., "High-frequency measures of informed trading and corporate announcements," *The Review of Financial Studies*, vol. 31, no. 6, pp. 2326–2376, 2018.

Brunnermeier, M. K. and Pedersen, L. H., "Market liquidity and funding liquidity," *The review of financial studies*, vol. 22, no. 6, pp. 2201–2238, 2008.

Bukovina, J., "Social media big data and capital markets—an overview," *Journal of Behavioral and Experimental Finance*, vol. 11, pp. 18–26, 2016.

Butt, H. A. and Virk, N. S., "Liquidity and asset prices: an empirical investigation of the nordic stock markets," *European Financial Management*, vol. 21, no. 4, pp. 672–705, 2015.

Campbell, J. Y., Lo, A. W.-C., MacKinlay, A. C. *et al.*, *The econometrics of financial markets.* princeton University press Princeton, NJ, 1997, vol. 2.

Cao, C., Hansch, O., and Wang, X., "The information content of an open limit-order book," *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, vol. 29, no. 1, pp. 16–41, 2009.

Chacko, G., Evans, C. L., Gunawan, H., and Sjoman, A. L., *The Global Economic System: How Liquidity Shocks Affect Financial Institutions and Lead to Economic Crises (paperback).* FT Press, 2011.

Chae, J., "Trading volume, information asymmetry, and timing information," *The journal of finance*, vol. 60, no. 1, pp. 413–442, 2005.

Chalos, P. and Chen, C. J., "Employee downsizing strategies: Market reaction and post announcement financial performance," *Journal of Business Finance & Accounting*, vol. 29, no. 5-6, pp. 847–870, 2002.

Chen, H., De, P., Hu, Y., and Hwang, B.-H., "Wisdom of crowds: The value of stock opinions transmitted through social media," *The Review of Financial Studies*, vol. 27, no. 5, pp. 1367–1403, 2014.

Chordia, T., Roll, R., and Subrahmanyam, A., "Market liquidity and trading activity," *The journal of finance*, vol. 56, no. 2, pp. 501–530, 2001.

———, "Order imbalance, liquidity, and market returns," *Journal of Financial economics*, vol. 65, no. 1, pp. 111–130, 2002.

———, "Liquidity and market efficiency," *Journal of Financial Economics*, vol. 87, no. 2, pp. 249–268, 2008.

Cochrane, J. H., *Asset pricing: Revised edition.* Princeton university press, 2005.

Contreras-Bolton, C. and Parada, V., "Automatic combination of operators in a genetic algorithm to solve the traveling salesman problem," *PloS one*, vol. 10, no. 9, p. e0137724, 2015.

Cumming, D., Johan, S., and Li, D., "Exchange trading rules and stock market liquidity," *Journal of Financial Economics*, vol. 99, no. 3, pp. 651–671, 2011.

Danielsson, J. and Payne, R., "Liquidity determination in an order-driven market," *The European Journal of Finance*, vol. 18, no. 9, pp. 799–821, 2012.

Degryse, H., De Jong, F., and Kervel, V. v., "The impact of dark trading and visible fragmentation on market quality," *Review of Finance*, vol. 19, no. 4, pp. 1587–1622, 2015.

Deuskar, P. and Johnson, T. C., "Market liquidity and flow-driven risk," *The Review of Financial Studies*, vol. 24, no. 3, pp. 721–753, 2011.

Engle, R., Fleming, M., Ghysels, E., and Nguyen, G., "Liquidity, volatility, and flights to safety in the us treasury market: Evidence from a new class of dynamic order book models," 2012, staff Report, Federal Reserve Bank of New York, No. 590, Federal Reserve Bank of New York, New York, NY.

Erenburg, G. and Lasser, D., "Electronic limit order book and order submission choice around macroeconomic news," *Review of Financial Economics*, vol. 18, no. 4, pp. 172–182, 2009.

Euroclear, "T+2 implementation questions & answers," *https://www.euroclear.com/dam/EFi/Campaigns/T-2_cycleQuestionsAndAnswers.pdf*, 2014.

Fama, E. F., "Efficient capital markets: A review of theory and empirical work," *The journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R., "The adjustment of stock prices to new information," *International economic review*, vol. 10, no. 1, pp. 1–21, 1969.

Faraway, J. J., *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models.* CRC press, 2005.

Foucault, T., Hombert, J., and Roşu, I., "News trading and speed," *The Journal of Finance*, vol. 71, no. 1, pp. 335–382, 2016.

Glosten, L. R. and Milgrom, P. R., "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of financial economics*, vol. 14, no. 1, pp. 71–100, 1985.

Gomber, P., Schweickert, U., and Theissen, E., "Liquidity dynamics in an electronic open limit order book: An event study approach," *European Financial Management*, vol. 21, no. 1, pp. 52–78, 2015.

Gompers, P. A. and Metrick, A., "Institutional investors and equity prices," *The quarterly journal of Economics*, vol. 116, no. 1, pp. 229–259, 2001.

Gould, M. D., Porter, M. A., Williams, S., McDonald, M., Fenn, D. J., and Howison, S. D., "Limit order books," *Quantitative Finance*, vol. 13, no. 11, pp. 1709–1742, 2013.

Gould, M. D., Porter, M. A., and Howison, S. D., “Quasi-centralized limit order books,” *Quantitative Finance*, vol. 17, no. 6, pp. 831–853, 2017.

Goyenko, R. Y., Holden, C. W., and Trzcinka, C. A., “Do liquidity measures measure liquidity?” *Journal of financial Economics*, vol. 92, no. 2, pp. 153–181, 2009.

Graham, J. R., Koski, J. L., and Loewenstein, U., “Information flow and liquidity around anticipated and unanticipated dividend announcements,” *The Journal of Business*, vol. 79, no. 5, pp. 2301–2336, 2006.

Greene, W. H., *Econometric analysis*, 7th ed. Pearson Education, 2012.

Griffin, J. M., Harris, J. H., and Topaloglu, S., “The dynamics of institutional and individual trading,” *The Journal of Finance*, vol. 58, no. 6, pp. 2285–2320, 2003.

Grinblatt, M. and Keloharju, M., “The investment behavior and performance of various investor types: a study of finland’s unique data set,” *Journal of financial economics*, vol. 55, no. 1, pp. 43–67, 2000.

———, “What makes investors trade?” *The Journal of Finance*, vol. 56, no. 2, pp. 589–616, 2001.

Groß-Klußmann, A. and Hautsch, N., “When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions,” *Journal of Empirical Finance*, vol. 18, no. 2, pp. 321–340, 2011.

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, *Multivariate data analysis*, 7th ed. Pearson Prentice Hall, 2010.

Hameed, A., Kang, W., and Viswanathan, S., “Stock market declines and liquidity,” *The Journal of Finance*, vol. 65, no. 1, pp. 257–293, 2010.

Härdle, W. K., Hautsch, N., and Mihoci, A., “Modelling and forecasting liquidity supply using semiparametric factor dynamics,” *Journal of Empirical Finance*, vol. 19, no. 4, pp. 610–625, 2012.

Harris, L., *Trading and exchanges: Market microstructure for practitioners.* Oxford University Press, USA, 2003.

Hasbrouck, J., *Empirical market microstructure: The institutions, economics, and econometrics of securities trading.* Oxford University Press, 2007.

Heale, R. and Twycross, A., “Validity and reliability in quantitative studies,” *Evidence-based nursing*, pp. ebnurs–2015, 2015.

Holland, J. H., *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.* MIT press, 1992.

Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X., *Applied logistic regression*, 3rd ed. John Wiley & Sons, 2013, vol. 398.

Hull, J., *Options, futures, and other derivatives.* Pearson/Prentice Hall, 2006.

Hussain, A. and Vatrapu, R., “Social data analytics tool: Design, development, and demonstrative case studies,” in *Enterprise Distributed Object Computing Conference Workshops and Demonstrations (EDOCW), 2014 IEEE 18th International.* IEEE, 2014, pp. 414–417.

———, "Social data analytics tool (sodato)," in *International Conference on Design Science Research in Information Systems.* Springer, 2014, pp. 368–372.

Hussain, A., Vatrapu, R., Hardt, D., and Jaffari, Z. A., "Social data analytics tool: A demonstrative case study of methodology and software," in *Analyzing Social Media Data and Web Networks.* Springer, 2014, pp. 99–118.

Irvine, P. J., "The incremental impact of analyst initiation of coverage," *Journal of Corporate Finance*, vol. 9, no. 4, pp. 431–451, 2003.

Jung, M. J., Naughton, J. P., Tahoun, A., and Wang, C., "Do firms strategically disseminate? evidence from corporate use of social media," *The Accounting Review*, 2017.

Kaniel, R., Saar, G., and Titman, S., "Individual investor trading and stock returns," *The Journal of Finance*, vol. 63, no. 1, pp. 273–310, 2008.

Kaniel, R., Liu, S., Saar, G., and Titman, S., "Individual investor trading and return patterns around earnings announcements," *The Journal of Finance*, vol. 67, no. 2, pp. 639–680, 2012.

Karabulut, Y., "Can facebook predict stock market activity?" *AFA 2013 San Diego Meetings Paper, available at SSRN: https://ssrn.com/abstract=2017099*, 2013.

Karnaukh, N., Ranaldo, A., and Söderlind, P., "Understanding fx liquidity," *The Review of Financial Studies*, vol. 28, no. 11, pp. 3073–3108, 2015.

Kim, O. and Verrecchia, R. E., "Market liquidity and volume around earnings announcements," *Journal of accounting and economics*, vol. 17, no. 1-2, pp. 41–67, 1994.

King, M. R., Osler, C., and Rime, D., "Foreign exchange market structure, players, and evolution," *Handbook of Exchange Rates*, pp. 1–44, 2012.

King, M. R., Osler, C. L., and Rime, D., "The market microstructure approach to foreign exchange: Looking back and looking forward," *Journal of International Money and Finance*, vol. 38, pp. 95–119, 2013.

Kothari, S. and Warner, J. B., "Econometrics of event studies," *Handbook of empirical corporate finance*, vol. 1, pp. 3–36, 2007.

Krinsky, I. and Lee, J., "Earnings announcements and the components of the bid-ask spread," *The Journal of Finance*, vol. 51, no. 4, pp. 1523–1535, 1996.

Kyle, A. S., "Continuous auctions and insider trading," *Econometrica: Journal of the Econometric Society*, pp. 1315–1335, 1985.

Lee, C. M., Mucklow, B., and Ready, M. J., "Spreads, depths, and the impact of earnings information: An intraday analysis," *The Review of Financial Studies*, vol. 6, no. 2, pp. 345–374, 1993.

Lei, Q. and Wang, X., "Time-varying liquidity trading, private information and insider trading," *European Financial Management*, vol. 20, no. 2, pp. 321–351, 2014.

Lillo, F., Miccichè, S., Tumminello, M., Piilo, J., and Mantegna, R. N., "How news affects the trading behaviour of different categories of investors in a financial market," *Quantitative Finance*, vol. 15, no. 2, pp. 213–229, 2015.

Linnainmaa, J. T., “Do limit orders alter inferences about investor performance and behavior?” *The Journal of Finance*, vol. 65, no. 4, pp. 1473–1506, 2010.

Malo, P. and Pennanen, T., “Reduced form modeling of limit order markets,” *Quantitative Finance*, vol. 12, no. 7, pp. 1025–1036, 2012.

Mancini, L., Ranaldo, A., and Wrampelmeyer, J., “Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums,” *The Journal of Finance*, vol. 68, no. 5, pp. 1805–1841, 2013.

Mentz, M. and Schiereck, D., “Cross-border mergers and the cross-border effect: the case of the automotive supply industry,” *Review of Managerial Science*, vol. 2, no. 3, pp. 199–218, 2008.

Moeller, S. B. and Schlingemann, F. P., “Global diversification and bidder gains: A comparison between cross-border and domestic acquisitions,” *Journal of Banking & Finance*, vol. 29, no. 3, pp. 533–564, 2005.

Moro, E., Vicente, J., Moyano, L. G., Gerig, A., Farmer, J. D., Vaglica, G., Lillo, F., and Mantegna, R. N., “Market impact and trading profile of hidden orders in stock markets,” *Physical Review E*, vol. 80, no. 6, p. 066102, 2009.

Næs, R. and Skjeltorp, J. A., “Order book characteristics and the volume–volatility relation: Empirical evidence from a limit order market,” *Journal of Financial Markets*, vol. 9, no. 4, pp. 408–432, 2006.

NASDAQ Helsinki Ltd, *Rules of the Exchange*, 2018, unofficial transalation version, available at http://business.nasdaq.com/media/Rules_of_the_Exchange_3_Jan_2018_tcm5044-25490.pdf.

NASDAQ OMX, *Computer Trading in the Nordics*, 2011, slides from Computer Trading Seminar, available at https://slidex.tips/download/copyright-2011-the-nasdaq-omx-group-inc-all-rights-reserved-computer-trading-in.

Neuhierl, A., Scherbina, A., and Schlusche, B., “Market reaction to corporate press releases,” *Journal of Financial and Quantitative Analysis*, vol. 48, no. 4, pp. 1207–1240, 2013.

Nofer, M. and Hinz, O., “Using twitter to predict the stock market,” *Business & Information Systems Engineering*, vol. 57, no. 4, p. 229, 2015.

Nofsinger, J. R. and Sias, R. W., “Herding and feedback trading by institutional and individual investors,” *The Journal of finance*, vol. 54, no. 6, pp. 2263–2295, 1999.

O’Hara, M., *Market microstructure theory.* Blackwell Publishers Cambridge, MA, 1995, vol. 108.

O’Hara, M. and Ye, M., “Is market fragmentation harming market quality?” *Journal of Financial Economics*, vol. 100, no. 3, pp. 459–474, 2011.

Oomen, R., “Execution in an aggregator,” *Quantitative Finance*, vol. 17, no. 3, pp. 383–404, 2017.

Payne, R., “Informed trade in spot foreign exchange markets: an empirical investigation,” *Journal of International Economics*, vol. 61, no. 2, pp. 307–329, 2003.

Rakowski, D. and Beardsley, X. W., "Decomposing liquidity along the limit order book," *Journal of Banking & Finance*, vol. 32, no. 8, pp. 1687–1698, 2008.

Riordan, R., Storkenmaier, A., Wagener, M., and Zhang, S. S., "Public information arrival: Price discovery and liquidity in electronic limit order markets," *Journal of Banking & Finance*, vol. 37, no. 4, pp. 1148–1159, 2013.

Roberts, P., Priest, H., and Traynor, M., "Reliability and validity in research." *Nursing standard*, vol. 20, no. 44, 2006.

Rosa, C., "Walking on thin ice: Market quality around fomc announcements," *Economics Letters*, vol. 138, pp. 5–8, 2016.

Salo, M., "Liquidity effects of earnings announcements in stock markets," 2014, Master's thesis, Tampere University of Technology.

Schlottmann, F. and Seese, D., "Financial applications of multi-objective evolutionary algorithms: Recent developments and future research directions," in *Applications of multi-objective evolutionary algorithms.* World Scientific, 2004, pp. 627–652.

Sensoy, A., "Commonality in liquidity: Effects of monetary policy and macroeconomic announcements," *Finance Research Letters*, vol. 16, pp. 125–131, 2016.

Siganos, A., Vagenas-Nanos, E., and Verwijmeren, P., "Facebook's daily sentiment and international stock markets," *Journal of Economic Behavior & Organization*, vol. 107, pp. 730–743, 2014.

Siikanen, M., Kanniainen, J., and Luoma, A., "What drives the sensitivity of limit order books to company announcement arrivals?" *Economics Letters*, vol. 159, pp. 65–68, 2017.

Siikanen, M., Kanniainen, J., and Valli, J., "Limit order books and liquidity around scheduled and non-scheduled announcements: Empirical evidence from NASDAQ Nordic," *Finance Research Letters*, vol. 21, pp. 264–271, 2017.

Siikanen, M., Baltakys, K., Kanniainen, J., Vatrapu, R., Mukkamala, R., and Hussain, A., "Facebook drives behavior of passive households in stock markets," *Finance Research Letters*, 2018.

Siikanen, M., Nögel, U., and Kanniainen, J., "Liquidity in the fx market: Empirical evidence from an aggregator," 2018, working paper. Available at SSRN: https://ssrn.com/abstract=3208229.

Sivanandam, S. and Deepa, S., "Genetic algorithm optimization problems," in *Introduction to Genetic Algorithms.* Springer, 2008, pp. 165–209.

Snow, N. M. and Rasso, J., "If the tweet fits: How investors process financial information received via social media," 2017, working paper. Available at SSRN: url-https://ssrn.com/abstract=2945554.

Tumminello, M., Lillo, F., Piilo, J., and Mantegna, R. N., "Identification of clusters of investors from their real trading activity in a financial market," *New Journal of Physics*, vol. 14, no. 1, p. 013041, 2012.

Velásquez, S., Kanniainen, J., Mäkinen, S., and Valli, J., “Layoff announcements and intra-day market reactions,” *Review of Managerial Science*, vol. 12, no. 1, pp. 203–228, 2018.

Wurgler, J., “Financial markets and the allocation of capital,” *Journal of financial economics*, vol. 58, no. 1-2, pp. 187–214, 2000.

Yang, W., Lin, D., and Yi, Z., “Impacts of the mass media effect on investor sentiment,” *Finance Research Letters*, vol. 22, pp. 1–4, 2017.

You, W., Guo, Y., and Peng, C., “Twitter’s daily happiness sentiment and the predictability of stock returns,” *Finance Research Letters*, vol. 23, pp. 58–64, 2017.

Zhang, X., Fuehres, H., and Gloor, P. A., “Predicting stock market indicators through twitter "I hope it is not as bad as I fear",” *Procedia-Social and Behavioral Sciences*, vol. 26, pp. 55–62, 2011.

Zhang, Y., An, Y., Feng, X., and Jin, X., “Celebrities and ordinaries in social networks: Who knows more information?” *Finance Research Letters*, vol. 20, pp. 153–161, 2017.

Zheludev, I., Smith, R., and Aste, T., “When can social media lead financial markets?” *Scientific reports*, vol. 4, p. 4213, 2014.

Zheng, L., “The type of corporate announcements and its implication on trading behaviour,” *Accounting & Finance*, 2018.

Zhou, M., Lei, L., Wang, J., Fan, W., and Wang, A. G., “Social media adoption and corporate disclosure,” *Journal of information systems*, vol. 29, no. 2, pp. 23–50, 2015.

Publications

Publication I

Siikanen M., Baltakys, K., Kanniainen, J., Vatrapu, R., Mukkamala, R., Hussain, A. "Facebook drives behavior of passive households in stock markets", *Finance Research Letters*.

Available at: https://doi.org/10.1016/j.frl.2018.03.020

Finance Research Letters xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Facebook drives behavior of passive households in stock markets

Milla Siikanen[*,a], Kęstutis Baltakys[a], Juho Kanniainen[a], Ravi Vatrapu[b,c], Raghava Mukkamala[b,c], Abid Hussain[b]

[a] DARE Business Data Research Group, Laboratory of Industrial and Information Management, Tampere University of Technology, Finland
[b] Centre for Business Data Analytics, Copenhagen Business School, Denmark
[c] Westerdals Oslo School of Arts, Communication and Technology, Norway

ARTICLE INFO

Keywords:
Investor behavior
Social media
Stock markets
Investor sophistication
Decision making

JEL classification:
G10
G11

ABSTRACT

Recent studies using data on social media and stock markets have mainly focused on predicting stock returns. Instead of predicting stock price movements, we examine the relation between Facebook data and investors' decision making in stock markets with a unique data on investors' transactions on Nokia. We find that the decisions to buy versus sell are associated with Facebook data especially for passive households and for nonprofit organizations. At the same time, it seems that more sophisticated investors—financial and insurance institutions—are behaving independently from Facebook activities.

1. Introduction

Social media sites, such as Facebook and Twitter, create various opportunities for companies to improve their internal and external communications and to collaborate and communicate with their customers, partners, and other stakeholders, such as investors. Given the importance of social media in external communications, it is not surprising that social media data have been used recently to predict real-world outcomes (see e.g. Asur and Huberman, 2010). In the financial market research, numerous scholars have used Facebook data (Karabulut, 2013; Siganos et al., 2014; Bukovina et al., 2015) and data from other social media sites (Bollen et al., 2011; Zhang et al., 2011; Zheludev et al., 2014; Chen et al., 2014; Nofer and Hinz, 2015; Zhang et al., 2017; You et al., 2017).[1] The primary aim of such research has been to predict market-wide stock movements, yet there is scant research on how social media data relate to the behavior of *individual investors*, perhaps because of the lack of availability of investor account level data.

In this paper, we examine the extent to which investors' trading decisions are driven by Facebook posts and activity. To this end, we use a unique investor-level shareholding registration data set that includes the trading of all Finnish investors over multiple years. In particular, given that an investor trades, we study how Facebook data relate to investors' decisions to increase or decrease their positions. This question is addressed for different investor groups, including financial institutions, nonprofit organizations, and households, and their trades in Nokia stock. As Nokia was one of the most liquid stocks on the Finnish stock market, this unique data has been studied in several articles,[2] and here we combine it with social media data. Paper by Lillo et al. (2015) is the most closely related study to ours. It also investigates the trading behavior of different investor groups with Nokia stock, but with Thomson Reuters news articles—which are not social media data per se.

* Corresponding author.
E-mail address: milla.siikanen@tut.fi (M. Siikanen).

[1] See also Bukovina (2016) for an overview of research related to a link between social media and capital markets.
[2] See for example Westerholm (2009); Tumminello et al. (2012); Lillo et al. (2015) and Ranganathan et al. (2017).

https://doi.org/10.1016/j.frl.2018.03.020
Received 5 March 2018; Accepted 16 March 2018
1544-6123/

Currently, Facebook is clearly the most widely used social media platform, with 2.2 billion monthly active users worldwide (Statista, 2018). As of January 2013, social media sites such as Facebook and Twitter are used by about 45% of S&P1500 firms to communicate externally formal and informal information about their business (Jung et al., 2017). Specifically, companies communicate both corporate disclosures and other information via social media (Zhou et al., 2014). Yang et al. (2017) show that social media, and mass media in general, influences investor's trading decisions. Snow and Rasso (2017) argue that less sophisticated investors potentially benefit most from disclosures communicated via social media, because, on social media platforms, the information is essentially "pushed" to them, which makes this information easier to access. In addition, Snow and Rasso (2017) show that less sophisticated investors process financial information received from social media differently from information received via company's investor relations website.

It is important to remember that typically companies use official exchange-routed company announcements as a primary communication channel (see e.g. Jung et al., 2017), followed by other channels, including newspapers and social media.[3] Additionally, communicating information via social media is voluntary, while some company announcement releases are mandatory. Furthermore, Jung et al. (2017) show that companies disseminate strategically, i.e. companies are less likely to disseminate information in Twitter when the news is bad. In this regard, we wish to determine how the investment decisions of, for example, less sophisticated and professional investors, among other investor groups, correlate with potentially biased Facebook information. We note that the relationship between Facebook data and trading can also be related to the attention grabbing behavior of investors, especially households (see Barber and Odean, 2007).

2. Data

2.1. Shareholding registration record data

To identify the trading of different investor categories, we use shareholding registration record data including all domestic investors from June 7, 2010 to the end of 2016, obtained from Euroclear Ltd.[4] Each record in the data contains detailed information about the investor and the change in his/her holdings. During our analysis period, 282,269 distinct Finnish investors traded Nokia stock. We divide them into five groups according to their sector codes: nonfinancial corporations, financial and insurance corporations, general governmental organizations, nonprofit organizations, and households. Household investors are further divided into four investor activity groups. Investor's activity group is defined based on the number of days the investor traded during the past eight weeks, including the analyzed week. If the number of active days in the past 8 weeks is equal to 1, the investor is considered inactive; if it is between 2 and 5, the investor is passive; 6–20 means moderate; and 21–40 means active. Notably, this is a dynamic group, as one investor might appear in several groups throughout the analysis period.

For the purposes of our analysis, we calculate the number of investors in each group who changed their holdings during a week and the number of investors who increased their holdings (bought more than sold) during that week. Table 1 gives the descriptive statistics of the investor groups and their weekly trading in our data sample. We see that financial and governmental institutions are on average most active sector groups, where as households and nonprofit organizations are least active.

2.2. Facebook data

We collect daily numbers of posts and related comments, likes, and shares from Nokia's Facebook wall[5] between June 2010 and December 2016 using the Social Data Analytics Tool (SODATO) (see Hussain et al., 2014; Hussain and Vatrapu, 2014a, 2014b). The comments, likes, and shares are always related to a specific post, i.e. the post is the main action. Therefore, we assign the numbers of comments, likes, and shares to the date of the original post—that is, not the date when the actual comment, like, or share was made. In effect, the numbers of comments, likes, and shares quantify the attention the posts released on a particular day received.

We aggregate the daily Facebook data to weekly by summing the numbers of posts, comments, likes, and shares during a week. We take the week beginning on Saturday and ending on Friday, since trading does not occur on weekends. This way, we relate the Facebook activity on weekends to the week in which they can actually affect investors' trading decisions. In total, our sample comprises of 342 weekly observations for posts, comments, likes, and shares. Table 2 gives descriptive statistics of these time series. We can see that on average, there is more than one post made per day, and calculate that one post got on average 274 comments, 4379 likes, and 7 shares.

[3] See Siikanen et al. (2017b,a), and references therein, for effects of company announcements in stock markets.

[4] Grinblatt and Keloharju (2000, 2001); Tumminello et al. (2012); Lillo et al. (2015) and Baltakys et al. (2018) use data sets from the same source, and provide descriptions of the data. However, they use data from before 2009, when all transactions were reported separately with exact trading dates. After moving to Central Counterparty Clearing in late 2009, the Euroclear research data set contains only *aggregated* daily trades *without specifying the actual trading dates*—instead a registration date is reported for each record. Thus, we reverse engineer the trading dates from the registration dates. We use the official $T+3$ settlement convention for data before and on October 8, 2014 and $T+2$ afterwards (see Euroclear, 2014). Using the derived trading dates, we aggregate transactions on a weekly basis, and this reduced the possible noise of inaccurate trading date derivation.

[5] https://www.facebook.com/nokia.

Table 1
Descriptive statistics on investor groups. *N* gives the total number of investors per group. Mean, median and standard deviation (st.Dev) relate to the weekly observations on numbers of investors in each group that changed their net holdings during a week. In Panel B, household investors are categorized into activeness groups on the basis of their trading in the past eight weeks (40 trading days).

Panel A: Investor categories

Sector	*N*	Mean (%of all)	Median	st.Dev
Companies	12,213	271 (2.2%)	230	166
Financial	427	28 (6.6%)	27	9
Governmental	89	7 (7.9%)	7	4
Nonprofit	1177	18 (1.5%)	16	12
Households	268,363	4640 (1.7%)	3694	3179
Total	282,269			

Panel B: Activity groups of household investors

Activeness; # of active days	*N*	Mean (%of all)	Median	st.Dev
Active; (20, 40]	1228	54 (4.4%)	51	22
Moderate; (5, 20]	16,019	502 (3.1%)	450	227
Passive; (1, 5]	120,906	1856 (1.5%)	1402	1422
Inactive; 1	264,942	2228 (0.8%)	1670	1897

Table 2
Descriptive statistics on Facebook data. *N* gives the total number of each Facebook activity in our sample. Mean, median and standard deviation (st.Dev) relate to the weekly observations on numbers of each Facebook activity.

Activity	*N*	Mean	Median	st.Dev
Post	2906	8	8	6
Comment	797,586	2332	1585	2808
Like	12,725,171	37,208	11,977	43,500
Share	919,380	2688	461	4525

2.3. *Company announcement data*

The announcement data is collected from NASDAQ OMX Nordic's website.[6] The data set includes all the announcements that Nokia filed with Nasdaq between June 2010 and December 2016. Altogether, we have 507 company announcements in the sample. We aggregate the announcement data into weekly by summing the number of announcements from Saturday to Friday, i.e. in similar way as the Facebook data. In the regressions, we use a dummy variable to indicate whether there was at least one announcement release during a week. Our sample includes 187 weeks with at least one announcement release (out of total 342 weeks).

2.4. *Weekly return data*

The daily adjusted closing price data used to calculate the returns is collected from NASDAQ OMX Nordic's website.[7] For each week, we calculate the log return as $\mathrm{Ret}_t = \ln[P_t/P_{t-1}]$, where P_t is the closing price from the last trading day on the week (usually Friday), and P_{t-1} is the closing price from last trading day on the previous week $t-1$ (usually previous week's Friday). The average weekly return for Nokia during the sample period was -0.16%.

3. Framework of the empirical analysis

Our analysis is based on logistic regressions to explain how Facebook activity relates to an increase versus a decrease in Nokia shares in investors' portfolios.[8] To identify the groups of investors whose trading behavior is related to Facebook data, we run separate regressions for each investor group with each Facebook variable.

[6] http://www.nasdaqomxnordic.com/news/companynews, see the page also for detailed information.
[7] http://www.nasdaqomxnordic.com/shares/microsite?Instrument=HEX24311.
[8] Another option would be (instead of restricting the analysis to a binary outcome) to use linear regressions with continuous dependent variable (i.e. how much an investor changed the position). However, in order to use continuous dependent variable, proportional changes in investors' positions would have to be calculated, which, in turn, requires information on investors' holdings. In contrast to changes in holdings, the levels of holdings, however, were not accurately available. The use of "changes in holdings" as a non-proportional variable is problematic, because investors are trading by very different amounts of shares. These problems are addressed by using logistic regression.

The dependent variable in our regressions is a dummy variable with value 1 if an investor increased his/her holdings in Nokia stock during a given week (bought more than sold) and 0 if the investor decreased the holdings ($D_t^{\text{increased}}$). In a given week, only investors whose net position for Nokia changed are included. The explanatory variable of main interest is the number of posts, comments, likes, or shares depending on the regression (FB). We control for company announcement releases with company announcement dummy (NEWS_t), which is 1 if there was an announcement released during week t and 0 otherwise. Additionally, we use the number of investors in the group who increased their holdings during the previous week scaled by the total number of investors who changed their holdings during the previous week. This is depicted as follows:

$$\text{scD}_{t-1}^{\text{increased}} = \frac{1}{n_{t-1}} \sum_{i=1}^{n_{t-1}} D_{i,t-1}^{\text{increased}}$$

where n_{t-1} is the number of investors who changed (increased or decreased) their holdings in Nokia during week $t-1$. We also add control variables for the return on present week (Ret_t) and the previous week (Ret_{t-1}). Lastly, we include monthly (M_t) and yearly (Y_t) dummy variables. The monthly dummies control for the potential yearly seasonality in the trading (for example, realizing the losses in December for tax purposes, see e.g. Grinblatt and Keloharju, 2001), and the yearly dummies accommodate the analysis for example to possible changes due to the abandonment of Nokia's mobile business (in 2014, Nokia's mobile business was acquired by Microsoft, changing the focus of the company to a telecommunications infrastructure business). To summarize, the regressions we run are of the following form:

$$\begin{aligned} g(D_t^{\text{increased}}) &= \alpha_1 + \alpha_2 \cdot \text{FB}_t + \alpha_3 \cdot \text{NEWS}_t + \alpha_4 \cdot \text{scD}_{t-1}^{\text{increased}} \\ &\quad + \alpha_5 \cdot \text{Ret}_t + \alpha_6 \cdot \text{Ret}_{t-1} + \sum_{j=1}^{11} \alpha_{j+6} \cdot M_j + \sum_{j=1}^{6} \alpha_{j+17} \cdot Y_j \end{aligned} \tag{1}$$

where g is the logit function.

4. Results

Panel A in Table 3 shows that for households and nonprofit institutions, all the regression estimates are statistically significant. The results indicate that the decisions of investors in these groups to buy vs. sell have a clear association with the Facebook data. For nonprofit institutions, the economic significance is relatively high: the odds of a nonprofit institution buying rather than selling range from 1.111 to 1.212 when the amount of Facebook activity increases by one standard deviation. For financial institutions, Panel A in Table 3 shows no association between the buy vs. sell decisions and the Facebook data. The results for companies and governmental institutions are something between those of financial institutions and households and nonprofit institutions, as half or less of the estimates are statistically significant.

To take a closer look at the effect of Facebook on the trading of households, Panel B in Table 3 presents the estimated regression results for individual investors in different activity groups. We observe that, in general, the more active a household is, the weaker is the association between Facebook data and buying/selling behavior. The odds ratios for passive and inactive investors are more modest than those of nonprofit institutions, though for posts they are still relatively high (1.088 and 1.072). For brevity, we do not report the regression estimates for interception and control variables here, but they are available in Online appendix. In general, most of the estimates for control variables are statistically significant.

Grinblatt and Keloharju (2000) argue that, roughly speaking, finance and insurance institutions, as well as companies, can be viewed as the most sophisticated investor groups, as they generally take larger positions, have more resources to spend on research, and in many cases view investment as full-time career. In light of this, our findings indicate that more sophisticated investors are more independent of Facebook activities, as there is clearly no association between Facebook activities and decisions of financial institutions. Assuming that an investor's activeness is related to his/her sophistication, our findings on household activity groups supports the result that more sophisticated investors behave more independently of Facebook data.

Facebook can be seen as a secondary information channel compared to first-hand official company announcements published on the exchange, and companies are likely to strategically select information disseminated in Facebook (Jung et al., 2017). Nonprofit organizations and households, as arguably less sophisticated investors (Grinblatt and Keloharju, 2000), may allow their trading decision to be affected by Facebook posts and activity, especially if they have no access to professional data sources. In line with this view, Ammann and Schaub (2017) find that the trading decisions of unsophisticated investors are affected by postings that do not contain value-relevant information on a social trading platform.

As our question is if the decisions of different investors are associated with the Facebook data, we are mostly interested in whether the regression estimates for the Facebook variables are statistically and economically significant, while the signs of the coefficients are not in the main focus.[9] However, a couple of words about the signs of the estimates in Table 3. In Panel A, the signs for posts, comments, and likes are consistently positive, except comments for households. The signs for shares are both positive (governmental and nonprofit) and negative (companies, financial, households), though not all of them are statistically significant, which can explain

[9] The number of data points in the regression analysis is 332–341, which does not automatically lead to significant estimates as very large data samples do.

Table 3

Regression estimates: Trading of investor groups and Facebook data. The estimates related to Facebook variables of logistic regressions described in Section 3 (Eq. (1)) for all the investor categories. The dependent variable is a dummy variable getting value of 1 if an investor increased his/her holdings during the week, and 0 if the investor decreased the holdings. In addition to the Facebook related variables (for which we report the estimates here), we control for company announcement releases, number of investors who changed their position during previous week (scaled), current and previous weeks returns, and in addition we have monthly and yearly dummies. The regression estimates for control variables (omitted here) are available in Online appendix. In Panel B, household investors are categorized into activeness groups on the basis of their trading in the past eight weeks (40 trading days). Number of observations (weeks in the analysis) is 341 for all the other regressions, except 332 for group governmental. p-values are given in parentheses (), and odds ratios (ORs) are given in curly brackets {}. ORs are calculated on the basis of one standard deviation change in the explanatory variable.

	Posts	Comments	Likes	Shares
Panel A: Investor categories				
Companies	0.011***	5.55E−06	7.61E−07**	−6.40E−07
	(3.71E−12)	(0.098)	(6.44E−03)	(0.775)
	{1.064}	{1.016}	{1.034}	{0.997}
Financial	5.92E−03	1.25E−05	9.42E−07	−1.94E−06
	(0.185)	(0.197)	(0.226)	(0.758)
	{1.034}	{1.036}	{1.042}	{0.991}
Governmental	0.015	5.45E−05**	2.62E−06	2.19E−05
	(0.091)	(7.20E−03)	(0.112)	(0.087)
	{1.086}	{1.165}	{1.121}	{1.104}
Nonprofit	0.033***	3.76E−05**	4.43E−06***	3.03E−05**
	(2.27E−07)	(4.79E−03)	(2.23E−04)	(1.84E−03)
	{1.203}	{1.111}	{1.212}	{1.147}
Households	0.011***	−4.46E−06***	1.92E−07**	−9.04E−06***
	(5.74E−83)	(7.64E−07)	(6.89E−03)	(1.23E−41)
	{1.064}	{0.988}	{1.008}	{0.960}
Panel B: Activity groups of household investors				
Active	2.41E−04	−5.82E−06	−9.74E−07	−1.36E−05**
	(0.942)	(0.386)	(0.065)	(1.32E−03)
	{1.001}	{0.984}	{0.959}	{0.940}
Moderate	1.85E−03	4.11E−06	2.67E−07	−1.71E−06
	(0.083)	(0.071)	(0.140)	(0.225)
	{1.011}	{1.012}	{1.012}	{0.992}
Passive	0.012***	−9.33E−06***	−7.33E−07***	−1.60E−05***
	(8.55E−55)	(2.37E−10)	(4.53E−10)	(1.73E−49)
	{1.072}	{0.974}	{0.969}	{0.930}
Inactive	0.015***	1.01E−05***	1.84E−06***	1.09E−07
	(5.12E−67)	(5.19E−12)	(1.24E−42)	(0.911)
	{1.088}	{1.029}	{1.083}	{1.000}

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

the variation. Panel B with activity groups reports positive estimates for posts, but there is more variation for comments, likes, and shares as passive and inactive investors have negative estimates. Looking deeper into the reasons of these findings is out of the scope of this paper and left for the future research, as it would require semantic analysis.[10]

5. Summary and conclusion

This paper gives the first empirical evidence that Facebook activities affect the trading of different investors differently. We provide evidence that the decisions of arguably less sophisticated investors—that is, households and nonprofit organizations—to increase or decrease shareholdings are clearly associated with Facebook data. At the same time, the decisions of financial institutions, which are likely to be among the most sophisticated investors in the market, are not associated with Facebook activity. Moreover, less active households' decisions are related to Facebook, while the decisions of more active ones are not, which gives additional evidence that the less sophisticated the investor, the more closely related the behavior is to Facebook. Given that Facebook is not a regulated information channel compared to first-hand official exchange releases, companies are likely to strategically select what information to disseminate in Facebook (Jung et al., 2017). This suggests that less sophisticated investors, who may not have access to professional sources for financial data and news, may be driven by biased information.

In the future research we are planning to do sentimental analysis on the posts and comments to give a more comprehensive

[10] Additionally, one could consider if observed associations can represent a reverse causality so that investors are not reacting to social media posts but companies are posting on Facebook in response to changes in investment behavior. However, the reverse causality seems unlikely, because the information about numbers of traders changing their position is not public.

picture of the reactions of different investors to Facebook activities. Concentrating only on Nokia may introduce some investor clientele bias, since the investors interested in Nokia may in general be more social media and technology savvy and follow the posts because of their inclination towards technology. At this point, we were only able to collect the data for Nokia, but in the future research we are planning to extend the sample to a wider variety of stocks.

Acknowledgments

We want to thank Hannu Kärkkäinen and Jari Jussila for their valuable efforts to enable this project. The research leading to these results received funding from Tampere University of Technology through Networked Big Data Science and Engineering Center and TUT's doctoral school. This project also received funding from the European Union's Horizon 2020 research and innovation program under Marie Sklodowska-Curie grant agreement no. 675044. First author is grateful for the grants received from KAUTE Foundation and Nordea Bank Foundation sr. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplementary material

Online appendix associated with this article can be found, in the online version, at doi:10.1016/j.frl.2018.03.020.

References

Ammann, M., Schaub, N., 2017. The impact of internet postings on individual investors. Proceedings American Finance Association 2018 Annual Meeting.
Asur, S., Huberman, B.A., 2010. Predicting the future with social media. Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on. 1. IEEE, pp. 492–499.
Baltakys, K., Kanniainen, J., Emmert-Streib, F., 2018. Multilayer aggregation with statistical validation: application to investor networks. arXiv preprint arXiv:1708.09850.
Barber, B.M., Odean, T., 2007. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. Rev. Financ. Stud. 21 (2), 785–818.
Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. J. Comput. Sci. 2 (1), 1–8.
Bukovina, J., 2016. Social media big data and capital marketsan overview. J. Behav. Exp. Finance 11, 18–26.
Bukovina, J., et al., 2015. Sentiment and blue-chip returns. Firm level evidence from a dynamic threshold model. Technical Report. Mendel University in Brno, Faculty of Business and Economics.
Chen, H., De, P., Hu, Y., Hwang, B.-H., 2014. Wisdom of crowds: the value of stock opinions transmitted through social media. Rev. Financ. Stud. 27 (5), 1367–1403.
Euroclear, 2014. T+2 Implementation Questions & Answers. https://www.euroclear.com/dam/EFi/Campaigns/T-2_cycleQuestionsAndAnswers.pdf.
Grinblatt, M., Keloharju, M., 2000. The investment behavior and performance of various investor types: a study of finland's unique data set. J. Financ. Econ. 55 (1), 43–67.
Grinblatt, M., Keloharju, M., 2001. What makes investors trade? J. Finance 56 (2), 589–616.
Hussain, A., Vatrapu, R., 2014a. Social data analytics tool: design, development, and demonstrative case studies. Enterprise Distributed Object Computing Conference Workshops and Demonstrations (EDOCW), 2014 IEEE 18th International. IEEE, pp. 414–417.
Hussain, A., Vatrapu, R., 2014b. Social data analytics tool (sodato). International Conference on Design Science Research in Information Systems. Springer, pp. 368–372.
Hussain, A., Vatrapu, R., Hardt, D., Jaffari, Z.A., 2014. Social Data Analytics Tool: A Demonstrative Case Study of Methodology and Software. Analyzing Social Media Data and Web Networks. Springer, pp. 99–118.
Jung, M.J., Naughton, J.P., Tahoun, A., Wang, C., 2017. Do firms strategically disseminate? evidence from corporate use of social media. Acc. Rev. in press https://doi.org/10.2308/accr-51906.
Karabulut, Y., 2013. Can facebook predict stock market activity? AFA 2013 San Diego Meetings Paper. available at SSRN.
Lillo, F., Miccichè, S., Tumminello, M., Piilo, J., Mantegna, R.N., 2015. How news affects the trading behaviour of different categories of investors in a financial market. Quant. Finance 15 (2), 213–229.
Nofer, M., Hinz, O., 2015. Using twitter to predict the stock market. Bus. Inf. Syst. Eng. 57 (4), 229.
Ranganathan, S., Kivelä, M., Kanniainen, J., 2017. Dynamics of investor spanning trees around dot-com bubble. arXiv preprint arXiv:1708.04430.
Siganos, A., Vagenas-Nanos, E., Verwijmeren, P., 2014. Facebook's daily sentiment and international stock markets. J. Econ. Behav. Organ. 107, 730–743.
Siikanen, M., Kanniainen, J., Luoma, A., 2017a. What drives the sensitivity of limit order books to company announcement arrivals? Econ. Lett. 159, 65–68.
Siikanen, M., Kanniainen, J., Valli, J., 2017b. Limit order books and liquidity around scheduled and non-scheduled announcements: empirical evidence from nasdaq nordic. Finance Res. Lett. 21, 264–271.
Snow, N.M., Rasso, J., 2017. If the Tweet Fits: How Investors Process Financial Information Received via Social Media. SSRN. working paper
Statista, 2018. Most Famous Social Network Sites Worldwide as of January 2018. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.
Tumminello, M., Lillo, F., Piilo, J., Mantegna, R.N., 2012. Identification of clusters of investors from their real trading activity in a financial market. New J. Phys. 14 (1), 013041.
Westerholm, P.J., 2009. Do uninformed crossed and internalized trades tap into unexpressed liquidity? The case of nokia. Acc. Finance 49 (2), 407–424.
Yang, W., Lin, D., Yi, Z., 2017. Impacts of the mass media effect on investor sentiment. Finance Res. Lett. 22, 1–4.
You, W., Guo, Y., Peng, C., 2017. Twitter's daily happiness sentiment and the predictability of stock returns. Finance Res. Lett. 23, 58–64.
Zhang, X., Fuehres, H., Gloor, P.A., 2011. Predicting stock market indicators through twitter "I hope it is not as bad as I fear". Procedia-Soc. Behav. Sci. 26, 55–62.
Zhang, Y., An, Y., Feng, X., Jin, X., 2017. Celebrities and ordinaries in social networks: who knows more information? Finance Res. Lett. 20, 153–161.
Zheludev, I., Smith, R., Aste, T., 2014. When can social media lead financial markets? Sci. Rep. 4, 4213.
Zhou, M., Lei, L., Wang, J., Fan, W., Wang, A.G., 2014. Social media adoption and corporate disclosure. J. Inf. Syst. 29 (2), 23–50.

Publication II

Siikanen M., Kanniainen, J., Valli, J. "Limit order books and liquidity around scheduled and non-scheduled announcements: Empirical evidence from NASDAQ Nordic", *Finance Research Letters*.

Available at: https://doi.org/10.1016/j.frl.2016.12.016

Finance Research Letters 21 (2017) 264–271

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Limit order books and liquidity around scheduled and non-scheduled announcements: Empirical evidence from NASDAQ Nordic

Milla Siikanen*, Juho Kanniainen, Jaakko Valli

Department of Industrial Management, Tampere University of Technology, P.O. Box 541, FI-33101 Tampere, Finland

ARTICLE INFO

Article history:
Received 6 October 2016
Accepted 28 December 2016
Available online 29 December 2016

JEL classification:
G10
G14

Keywords:
Limit order book
Liquidity
Company announcement
High-frequency data

ABSTRACT

Information arrivals may drive investors to require immediacy, generating sudden liquidity demand across multiple price levels in limit order books. We document significant intraday changes in stock limit order book characteristics and liquidity beyond the best levels around scheduled and non-scheduled company announcements. At aggregated level, liquidity beyond the best levels behaves quite differently from the bid–ask spread around scheduled announcements. Moreover, scheduled announcements improve multi-level liquidity to an exceptionally good level. We also provide evidence for pre-reactions in order books before non-scheduled announcements, which suggest the possibility of information leakage.

1. Introduction

Multi-level high-frequency data from stock order book markets has allowed researchers to empirically examine order book characteristics and dynamics (see e.g. Erenburg and Lasser, 2009; Engle et al., 2012; Deuskar and Johnson, 2011; Malo and Pennanen, 2012; Härdle et al., 2012; Gomber et al., 2015; Sensoy, 2016). As Degryse et al. (2015) argue, data beyond the best levels is intriguing, because it matters to investors as it reflects the quantity immediately available for trading and therefore the price of immediacy. It is especially important around information arrivals when informed investors may require immediacy and walk through the limit order book, taking advantage of stale limit orders, thus causing a large, immediate demand for liquidity on multiple levels.

Despite its importance, we know surprisingly little about how order book characteristics and liquidity available across multiple levels evolve around new information releases. In this paper, we aim to fill this gap with an extensive high-frequency multi-level order book data for 75 frequently traded and liquid stocks on Nasdaq Nordic markets (Helsinki, Stockholm, and Copenhagen) around first-hand stock exchange company announcements. We make the important distinction between scheduled and non-scheduled announcements, justified in Graham et al. (2006). Methodologically we follow (Malo and Pennanen, 2012; Deuskar and Johnson, 2011; Härdle et al., 2012) to capture the shape of the order book by estimating the slopes of the order book curves. The order book slopes can be used not only to characterize the order book but also to measure the order book liquidity across multiple order book levels (Malo and Pennanen, 2012). Our research question concerns how the order book characteristics and liquidity measured by order book slope across multiple levels (hereafter

* Corresponding author.
E-mail addresses: milla.siikanen@tut.fi (M. Siikanen), juho.kanniainen@tut.fi (J. Kanniainen), jaakko.s.valli@gmail.com (J. Valli).

http://dx.doi.org/10.1016/j.frl.2016.12.016
1544-6123/

referred to as "order book liquidity") reacts to scheduled and non-scheduled announcements, with an emphasis on comparison to the conventional spread between the best bid and ask levels.

Our research differs from the earlier studies in many important ways. In contrast to papers analysing liquidity around announcement releases with data on best levels only (see, e.g. Lee et al., 1993; Graham et al., 2006; Gross-Klussmann and Hautsch, 2011; Neuhierl et al., 2013), we use multi-level data. To our knowledge, not too many papers have employed multi-level data to study the effects of information releases on stocks' order book liquidity, but some important exceptions exist. Gomber et al. (2015) find in their study that the information content of Bloomberg ticker news is rather limited, causing no shocks to the order books, perhaps because Bloomberg is not always the first channel through which new information is distributed. We, however, document clear effects, especially with scheduled company announcements. The difference between our paper and theirs lies mainly in the type and source of the information. Additionally, Sensoy (2016) examines the impact of specific macro-announcements on liquidity commonality in Turkey. Though he addresses a different question than we do in this paper, his findings are in line with ours; order book liquidity is significantly affected beyond the best price quotes and it can be misleading to use only top-of-the-book data. Our research also relates to the study of Riordan et al. (2013), who explore the impact of Thomson Reuters newswire messages on intra-day price discovery, liquidity and trading intensity using depth at multiple price levels as one liquidity proxy. In addition to differences in methodologies and data, they do not divide the news into scheduled and non-scheduled ones. Apart from these studies, Erenburg and Lasser (2009) and Engle et al. (2012) combine multi-level data with macro announcements, but with data from the equity-index-linked securities market and the U.S. Treasury market, whereas our study focuses on equity markets with company announcements. Overall, in comparison to existing papers that use multi-level order book data, this research uses exceptionally extensive and frequently sampled data sets, examines impacts of stock exchange company announcements (instead of macro-announcements, ticker news, or newswire messages), and additionally, makes the important distinction between scheduled and non-scheduled announcements. Importantly, by these settings we provide novel results on pre- and post reactions in order books.[1]

This paper is constructed as follows. First, Section 2 introduces the order book parametrisation used in this study. Section 3 describes the data sets. Then, Section 4 demonstrates the framework of our empirical analsys, Section 5 reports the results and, finally, Section 6 summarizes and concludes the paper.

2. Order book parametrisation

An appropriate approach to characterize the order book and to measure the book's liquidity across multiple price levels should capture aspects with respect to both quantity (depth) over multiple levels and distances between the price levels. A popular approach is to estimate order book slope, which is obtained by fitting a linear curve to the order book data to measure how quantity changes as a function of price (Deuskar and Johnson, 2011; Härdle et al., 2012) or inversely, how price changes as a function of quantity (Malo and Pennanen, 2012). In this paper we technically follow (Malo and Pennanen, 2012). Importantly, the slope can be made invariant for splits and comparable between different stocks and over time, which enables an aggregated analysis.[2] In this paper, order book slope is called Order Book Illiquidity (OBI) as it measures liquidity across multiple price levels.

By interpreting buy orders as sell orders of negative quantity (see Malo and Pennanen, 2012), we can describe the state of a limit order book with marginal price curve $s(x)$, which is a piecewise constant and non-decreasing function of order quantity. We use a simple monetary measure $r(h)$, introduced by Malo and Pennanen (2012),

$$r(h) := \ln(s(h/\bar{s})) - \ln(\bar{s}),$$

where $\bar{s}$ refers to mid-price, and $h = \bar{s}x$ is the mark-to-market value of a market order of x shares—i.e. the value of the market order if we were to pay the mid-price for all the x shares. Hence, $r(h)$ gives the log change in the marginal price $s(x) = s(h/\bar{s})$ relative to the mid-price $\bar{s}$, as a function of h, the mark-to-market value of an order of x shares. Therefore, given that one buys stocks with a certain amount of money h, $r(h)$, also called the relative price impact curve, reveals how large a price impact (in monetary units) the trade has per share bought. The shape of a limit order book can be captured with a linear approximation of the relative price impact curve, which is of the form $r(h) = \text{OBI} \cdot h$. Here OBI is positive and is considered to measure order book liquidity (see Malo and Pennanen, 2012). Obviously, the smaller the value of OBI, the

[1] Siikanen et al. (2016) takes a step further and studies the factors affecting the sensitivity of order book liquidity to scheduled and non-scheduled company announcements. They also use the same datasets as we use in this study.

[2] Some other liquidity measures exist that incorporate multi-level data, too, such as the Exchange Liquidity Measure (XLM), which is based on cost of round-trip (see Gomber et al., 2015; Sensoy, 2016). However, an issue arises with XLM in our research, because it is determined for a specific trade size, yet the total multi-level depths on the bid and ask sides vary in time and consequently, the order book is not always deep enough to allow us to calculate XLM for a given order size. This is an issue especially just before announcements, when informed investors may cancel existing orders or choose not to submit new ones to avoid adverse selection, and also right after the announcements, when informed investors may take advantage of stale limit orders, both resulting in thin order book. Because the measure should be available especially around announcement times in this paper, we prefer to use order book slope as a multi-level measure—it is always possible to calculate as long as the order book is not totally empty. Additionally, Riordan et al. (2013); Engle et al. (2012)), and Erenburg and Lasser (2009) consider the depth on different price levels in the book in addition to the bid–ask spread. However, by this methodology the number of variables increases quickly when considering the depth at different levels separately, and the other dimension, distance between the price levels, is not taken into account.

more liquid the stock is. OBI is an ex-ante liquidity measure and captures information from multiple price levels in the limit order book. Whereas OBI measures the costs arising from order book illiquidity especially for large trades or a large bunch of simultaneous small trades sweeping over many price levels, spread measures the liquidity and costs of trading for small transactions that take place only at the best quotes. For visual illustration of the measure, see (Malo and Pennanen, 2012).

To eliminate the effects of the pre- and post-trading sessions, we follow Malo and Pennanen (2012) and exclude the first and last trading hours from the data. Moreover, we exclude an additional half-hour from the end of the trading day for stocks traded on OMX Helsinki and OMX Stockholm in order to get the same length of the daily periods with OMX Copenhagen. In addition, we de-seasonalise the observations of OBI, asymmetry of the two sides of the book, and relative spread. The de-seasonalisation is done by first estimating an average value for each 10-second observation moment from the estimation window for each event. We then de-seasonalise all the observations of each variable in the estimation and event windows by subtracting the average value of that moment from the observation and artificially convert them to noon, since most of the events take place around 12:00 ECT.

3. Data

3.1. Limit order book data

We use Level II order book data from NASDAQ OMX Nordic (Helsinki, Stockholm, Copenhagen), which are continuous limit order based markets.[3] Our sample includes 75 frequently traded stocks listed on NASDAQ OMX Helsinki, NASDAQ OMX Stockholm, and NASDAQ OMX Copenhagen. The stocks in our sample have been involved in the following stock indexes: OMX Helsinki 25, OMX Stockholm 30, and OMX Copenhagen 20. Out of the 75 stocks, 27 are traded in Helsinki, 28 are traded in Stockholm, and 20 are traded in Copenhagen.

We use the multi-level order book data from 1.1.2006 to 1.1.2010 and calculate the values of OBI based on snapshots of the order book taken every 10 seconds with data on the 20 best ask and bid price levels using linear regression. If there are data on less than 20 levels available, we use as many levels as possible. We distinguish the moments when trading halts occur and exclude them from this study, since normal trading and auto matching of the orders are halted during those moments. In particular, we exclude all the events from our sample for which there occurs a trading halt within the event window around the announcement release, and exclude the observations during trading halts in estimation windows when aggregating over time or events. However, the proportion of announcement releases excluded because of trading halts is relatively small. Additionally, we exclude trading days when technical errors at Nasdaq occurred and non-corrupted data was not available.

3.2. News data

The news data in this study come from NASDAQ OMX Nordic's website.[4] The announcements included in this study are from 1.1.2006 to 1.1.2010, and the respective companies filed them with NASDAQ OMX. The announcement times are given at one second precision in the data, but because we sample the order book data every 10 seconds, the times of the announcements are rounded to the nearest 10 seconds. In this paper, an announcement is considered positive (negative) if the mid-price increases (decreases) between the observation moment preceding the release and the last observation moment of the event window.

We re-categorise the announcements into two specific groups: scheduled and non-scheduled announcements. An announcement is scheduled if its exact publishing date is known to the public beforehand and non-scheduled if it is irregular, its publishing schedule is not given and cannot be reliably estimated, or the release is obviously unexpected. To be on the safe side, we exclude announcements whose publishing timespan is given non-specifically in earlier stock exchange releases or that are somewhat regular by nature.

Moreover, we exclude announcements that clearly contain no new information. These mostly include announcements published in multiple languages, in which case only the first one is involved. We also remove announcements for which we do not have enough data to form the 27-day estimation window and that have been published during non-trading hours. Additionally, if several announcements are published at the same second on the same stock, then only one of them is involved. Finally, we do not consider the cases where there has been a trading halt near the announcement time (within the pre- or post-window) ceasing continuous trading. With these restrictions we usually end up with less than 10 scheduled announcements and some dozens of non-scheduled announcements per company.

The final sample contains 329 (408) scheduled announcements and 2,102 (2,629) non-scheduled announcements with 60 (30) minute pre- and post-event windows. Just over 35% of them come from NASDAQ OMX Helsinki, around 45% orig-

[3] Data from NASDAQ Nordic has some advantages over U.S. limit order book markets. First, NASDAQ Nordic markets are little fragmented in comparison to the U.S. markets, where the limit orders for a given asset are spread between several exchanges, which poses a problem for empirical research (O'Hara and Ye, 2011), and where matching rules and transaction costs complicate comparisons between different limit order books for the same asset (Gould et al., 2013). Another advantage of using Nordic data from less liquid markets is that, as Butt and Virk (2015) argue, "it is more appropriate to test liquidity-related models in markets that are sufficiently illiquid to diagnose the level and strength of bearing [...] risks."

[4] http://www.nasdaqomxnordic.com/news/companynews , see the page also for detailed information.

inate from NASDAQ OMX Stockholm, and just under 20% are from NASDAQ OMX Copenhagen. Over 70% of the scheduled announcements in the final sample are financial announcements.

4. Framework of the empirical analysis

In our analyses, we utilise a framework from prior event studies (see, e.g. Campbell, 1997; Velásquez et al., 2016) by comparing observations in an event window around the event—in this case, announcement release—to observations in an estimation window. An estimation window comprises observations with 10-second frequency from 27 days preceding the day of an event. An event window consists of two sub-windows: a 60-minute pre-window and another 60-minute post-window. As a robustness check, we run the analyses using 30-minute pre- and post-windows and get similar results, but we report results on the 60-minute windows to demonstrate a more comprehensive data. We denote the set of observations from the estimation window (27 days preceding the announcement) by $\mathcal{E}$, from the pre-window (60 or 30 minutes preceding the announcement) by $\mathcal{A}^-$, and from the post-window by $\mathcal{A}^+$ (60 or 30 minutes following the announcement), and from the whole event window by $\mathcal{A}$ (120 or 60 minutes around the announcement).

In our aggregated analysis, to give equal weight to all the announcements and to make them comparable with each other regardless of differences in liquidity level and currency, we standardize all the values in estimation and event windows using the means and standard deviations calculated from the estimation windows.[5] In addition, to make the plots more readable and independent of number of observations N, we standardize the variables by multiplying the mean values taken over the events at each moment by $\sqrt{N}$ and consequently the variables are distributed with mean 0 and standard deviation of 1 in the estimation window.[6] We also sub-sample non-scheduled announcements from the original sample by choosing the same number of announcements as with the corresponding scheduled announcement sample with largest relative price impact (log-change in mid-price right before the event till the end of event window) to make the two event types more comparable with each other. We denote the aggregated, scaled values obtained by taking the mean over the events and multiplying it by $\sqrt{N}$ with over-line as $\overline{\text{variable}}$. Return makes an exception—in order to get an idea about the actual returns around announcements, we just apply the average values without standardization and scaling.

5. Results

5.1. Order book liquidity

To demonstrate announcement effects on the order book dynamics and the evolution of the liquidity at an aggregated level, Table 1 and Fig. 1 present $\overline{\text{OBI}}$ for the ask and bid sides of the order book separately around scheduled and non-scheduled announcement and positive and negative releases. Additionally, Fig. 1 plots the aggregated mid-price return in the event window, $\mathcal{A}$, calculated as follows:

$$r_{i,t} = \ln[m_{i,t}] - \ln[m_{i,t_0}], \quad t \in \mathcal{A},$$

where t_0 refers to the first moment in the event window $\mathcal{A}$ and $m_{i,\,t}$ denotes the mid-price for event i at time t.

Scheduled Announcements

For the scheduled announcements, the dynamics around the announcement times for all four cases (Table 1, first panel, and Fig. 1, plots A, C, E, G) are more or less the same. Notably, for the scheduled announcements, we observe that the order book liquidity during the 60-minute period before the event is much lower (i.e. OBI is higher) than in the estimation window; all the observations of $\overline{\text{OBI}}$ before the event in Table 1 are outside of the 99.9% confidence interval and in Fig. 1, $\overline{\text{OBI}}$ is constantly above the long-term maximum (obtained from the estimation window) on both ask and bid sides for positive announcements and on ask side for negative announcements. Moreover, on bid side for negative announcements $\overline{\text{OBI}}$ is well above the 95% confidence level and also reaches the long term maximum. As can be observed from Fig. 1, maximum and minimum values are sensitive to single observations and therefore it is better to analyse the results against confidence levels. We also verify the existence of statistically significant pre-reactions in the order book in all four cases by conducting a t-test comparing the mean values of OBI calculated from pre- and estimation windows (the results are available in the Online Appendix).

The results indicate that limit order books adjust to scheduled announcements before their release. The volume of limit orders standing in the book is low, indicating that many investors may have either cancelled their orders prior to the announcements or chosen not to submit new limit orders, and therefore informed investors have limited possibilities to take advantage of stale limit orders. This could indicate that investors tend not to leave limit orders standing in the order book, to avoid adverse selection.

[5] In our analysis, unit of h is 1 million EUR for stocks traded on the Helsinki Stock Exchange and 1 million SEK and 1 million DKK for stocks traded on the Stockholm and Copenhagen Stock Exchanges, respectively. While this has an impact on the values of OBI, it is not interfering our analysis since in our aggregated analysis we use standardized values.

[6] The variance of the mean taken over the events (assuming that the observations over the events are independent of each other) is $1/N$, i.e. the standard deviation is $1/\sqrt{N}$, and by multiplying the mean with $\sqrt{N}$, the scaled variable is distributed with standard deviation (and variance) of 1 in the estimation window, regardless of N.

Table 1
Order book illiquidity around announcements. This table presents observations of $\overline{\text{OBI}}$, standardised and aggregated order book illiquidity, on ask and bid sides around scheduled and non-scheduled announcements separately for positive and negative announcements. The statistical significance is calculated based on the empirical distribution of $\overline{\text{OBI}}$ from the estimation window (i.e. 27 days preceding the announcement) and indicated by asterisks: ***, **, and * indicate that the observations is outside of two sided 99.9%, 99%, and 95% confidence intervals, respectively. The non-scheduled announcement sample is extracted from the original sample by choosing the same number of announcements as with the corresponding scheduled announcement sample with largest relative price impact (log-change in mid-price right before the event till the end of event window). The sample size N is 151 for positive and 165 for negative announcements.

Scheduled announcements							
		Before					
Announcement	Variable	−10 s	−1 min	−5 min	−15 min	−30 min	−60 min
Positive	$\overline{\text{OBI}}_{Ask}$	16.59 ***	16.10 ***	14.21 ***	11.80 ***	10.83 ***	9.53 ***
	$\overline{\text{OBI}}_{Bid}$	27.52 ***	22.17 ***	17.75 ***	16.89 ***	16.29 ***	13.50 ***
Negative	$\overline{\text{OBI}}_{Ask}$	17.18 ***	17.32 ***	14.46 ***	10.45 ***	10.10 ***	8.72 ***
	$\overline{\text{OBI}}_{Bid}$	25.20 ***	25.19 ***	20.61 ***	15.75 ***	16.97 ***	9.66 ***
		After					
Announcement	Variable	+10 s	+1 min	+5 min	+15 min	+30 min	+60 min
Positive	$\overline{\text{OBI}}_{Ask}$	19.37 ***	22.89 ***	8.24 ***	0.16	−3.16 **	−4.17 ***
	$\overline{\text{OBI}}_{Bid}$	21.66 ***	23.17 ***	11.35 ***	0.62	−3.13 ***	−2.94 ***
Negative	$\overline{\text{OBI}}_{Ask}$	19.15 ***	22.83 ***	12.85 ***	2.49	−1.71	−1.68
	$\overline{\text{OBI}}_{Bid}$	28.25 ***	31.82 ***	15.05 ***	2.96	0.60	−2.20
Non-scheduled announcements							
		Before					
Announcement	Variable	−10 s	−1 min	−5 min	−15 min	−30 min	−60 min
Positive	$\overline{\text{OBI}}_{Ask}$	6.73 ***	6.49 ***	6.47 ***	6.28 ***	6.99 ***	6.46 ***
	$\overline{\text{OBI}}_{Bid}$	5.65 ***	2.66 *	6.16 ***	2.27	2.03	2.66 *
Negative	$\overline{\text{OBI}}_{Ask}$	3.05 **	3.35 **	2.99 **	4.32 ***	3.76 ***	5.01 ***
	$\overline{\text{OBI}}_{Bid}$	3.06 *	2.66 *	2.36 *	2.04	1.69	2.25
		After					
Announcement	Variable	+10 s	+1 min	+5 min	+15 min	+30 min	+60 min
Positive	$\overline{\text{OBI}}_{Ask}$	7.69 ***	9.94 ***	10.51 ***	7.34 ***	5.14 **	4.62 **
	$\overline{\text{OBI}}_{Bid}$	4.57 ***	4.78 ***	2.01	−1.50	−1.21	−1.51
Negative	$\overline{\text{OBI}}_{Ask}$	2.88 **	7.19 ***	6.56 ***	3.73 ***	3.80 ***	3.61 ***
	$\overline{\text{OBI}}_{Bid}$	5.18 ***	9.06 ***	5.32 ***	3.80 **	5.57 ***	4.68 ***

This result on the markets' pre-reaction in terms of liquidity is in line with the predictions of the traditional asymmetric information models of liquidity (see Graham et al., 2006, Table 1). Empirically, Graham et al. (2006) obtain consistent results for spread (they do not find any pattern for depth), Lee et al. (1993) with best-level-liquidity measures, and Krinsky and Lee (1996) with the adverse selection component of spread. In addition, the findings of Engle et al. (2012) on the Treasury markets and Erenburg and Lasser (2009) on the equity-index-linked securities market, both using depth at multiple price levels, show a decrease in depth before announcements, though they find that the reaction starts within five minutes before the announcements, which is in contrast to our finding that order book liquidity is at a significantly low level already an hour before an announcement release.

Next, Table 1 and Fig. 1 demonstrate that at the aggregate level, OBI starts to decline within a few minutes after a scheduled announcement and returns to a normal level within 10 to 20 minutes after the announcement, again, consistent with the predictions of the traditional asymmetric information models (Graham et al., 2006). Lee et al. (1993) find that spread and depth return to a normal level substantially slower (one day and three hours, respectively), but we argue that this is likely due to the market design and other technical developments since. The recovery time we find is consistent with the time observed in Gomber et al. (2015) using multi-level data and Degryse et al. (2005) using data on best levels to study limit order market recovery around illiquidity shocks. Our results are also in line with the findings of Engle et al. (2012) on the Treasury markets, as they find that depths at multiple levels recover fast after an announcement, and Erenburg and Lasser (2009) on the equity-index-linked securities market, as they observe that depths rise to original levels within 10 minutes after an announcement.

One of the main results of the paper is that after scheduled announcements, aggregated order book liquidity recovers to a level better than the long-term average. Particularly, for scheduled positive announcements, the level of $\overline{\text{OBI}}$ falls even below the long-term minimum, indicating that the supply of limit orders on both sides of the book is high and the transaction costs for large trades sweeping over multiple price levels are low. This may indicate that the release of scheduled announcements reduces information asymmetries and adverse selection costs significantly.

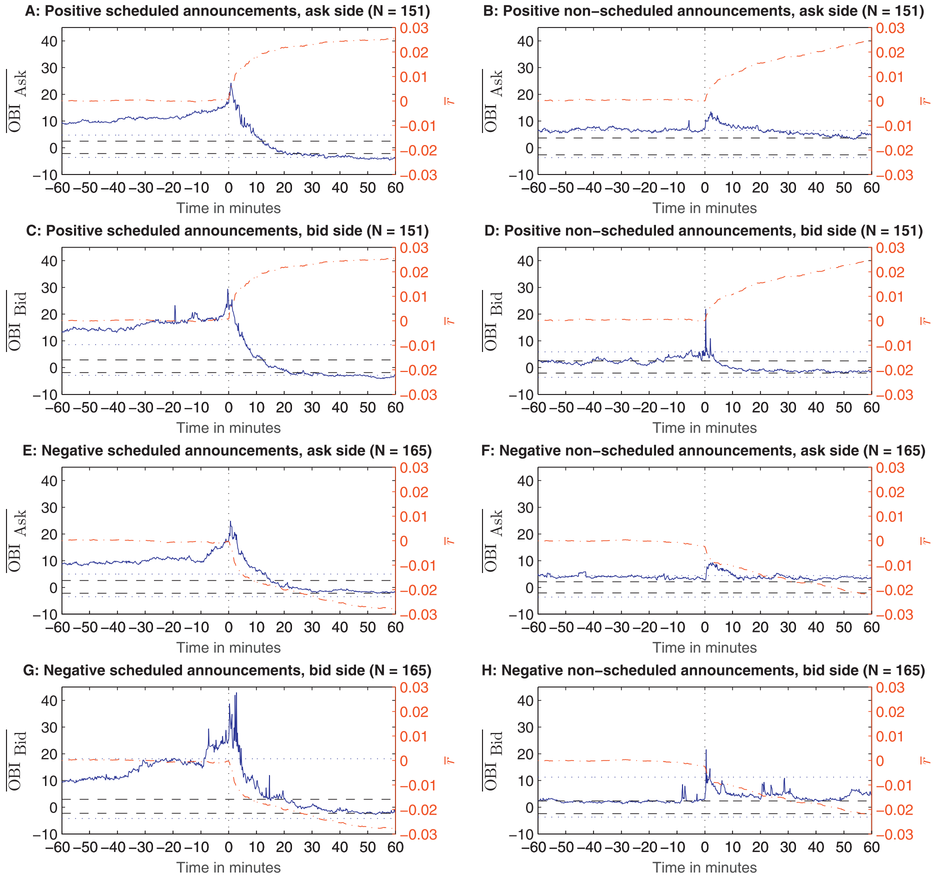

Fig. 1. Order book illiquidity around announcements. $\overline{\text{OBI}}$, standardised and aggregated order book illiquidity, on ask and bid sides around scheduled and non-scheduled announcements separately for positive and negative announcements and $\bar{r}$, average return on mid-price. The blue solid line corresponds to $\overline{\text{OBI}}$, the black dashed lines correspond to the 95% confidence level of $\overline{\text{OBI}}$ based on the empirical distribution from the estimation window (i.e. 27 days preceding the announcement) and the blue dotted lines correspond to the maximum and minimum values of $\overline{\text{OBI}}$ in the estimation window, all with the scale on the left-hand side. The red dash-dot line corresponds to $\bar{r}$ with the scale on the right-hand side in red. The black dotted vertical line at time zero corresponds to the time of the announcement. The non-scheduled announcement sample is extracted from the original sample by choosing the same number of announcements as with the corresponding scheduled announcement sample with largest relative price impact (log-change in mid-price right before the event till the end of event window). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Non-Scheduled Announcements

For non-scheduled announcements, Table 1, and Fig. 1 show statistically significant pre-reactions, especially on ask side, though the pre-reactions are clearly weaker than those for scheduled announcements. According to the predictions of traditional models, these pre-reactions can indicate information leakage (see Graham et al., 2006). In their study, Gross-Klussmann and Hautsch (2011) find an increase in spread already before the announcements.

Moreover, there is variation in the liquidity dynamics after non-scheduled announcements. On the bid side, liquidity improves to a normal level due to positive non-scheduled news (plot D), whereas the bid side liquidity remains at a relatively low level (for 60 minutes at least) after negative non-scheduled news (plot H). On the ask side, $\overline{\text{OBI}}$ peaks slightly within a few minutes after both positive and negative non-scheduled announcement releases and the order book remains relatively illiquid (plots B and F). Overall, positive non-scheduled announcements clearly improve liquidity on the bid side, but liquidity remains low in other cases (positive announcements/ask side, negative announcements/ask side, negative announcements/bid side) in comparison to its long-term mean. Riordan et al. (2013) find that liquidity increases (spread tightens and depth increases) before and after positive announcements, whereas it decreases around negative announcements (in their paper, positivity/negativity is based on the tone of the news). Graham et al. (2006) observe wider spreads and lower depth at the best levels after non-scheduled news, whereas Gross-Klussmann and Hautsch (2011) find an increased bid–ask spread around announcement releases, but find no significant effects on depth at the best levels.

5.2. Asymmetry

Next, we examine the asymmetry between the bid and ask sides, defined as

$$\mathrm{OBI}^{\mathrm{Asymmetry}} = \mathrm{OBI}_{\mathrm{Bid}} - \mathrm{OBI}_{\mathrm{Ask}}.$$

We run similar analysis on $\mathrm{OBI}^{\mathrm{Asymmetry}}$ as presented for $\overline{\mathrm{OBI}}$, and the table and the figure are available in the Online Appendix.

We find that the order book is abnormally asymmetric before scheduled announcements. In particular, ask side is relatively more liquid, meaning that it is easier to buy and harder to sell with market orders before scheduled announcements. After the scheduled announcements $\overline{\mathrm{OBI}^{\mathrm{Asymmetry}}}$ returns to normal level within 10 to 30 minutes. In contrast, asymmetry of the order book is more or less at normal levels before non-scheduled announcements. The arrival of positive non-scheduled announcements makes the book asymmetric (the ask side is relatively more illiquid) after the release, and this effect seems to be quite long lasting (at least 60 minutes). For negative announcements we observe weaker reactions, but the reaction seems to be of the opposite sign: negative non-scheduled announcements seem to make the bid side relatively illiquid.

Intuitively, when the non-scheduled news is positive, informed investors want to buy immediately using market orders that are executed against stale limit sell orders available in the book (i.e. OBI on the ask side increases). Moreover, there can exist investors who want to buy using limit orders and thus provide liquidity on the bid side (though facing uncertainty of the execution). Therefore, the imbalance between the bid and ask sides can be explained by a decrease in the provision of sell (buy) limit orders or an increase in the provision of buy (sell) limit orders, or both. The effect is exceptionally strong for positive non-scheduled events. This result implies that the news-momentum (news-contrarian) trading strategy through market orders after non-scheduled announcements is relatively expensive (inexpensive). In contrast to this, scheduled announcements decrease the transaction costs for both large buy and sell trades, keeping the book symmetric, meaning that both news-contrarian and news-momentum trading strategies are symmetrically inexpensive after scheduled company announcements.

5.3. Spread

We also compare spread (liquidity measure using best offers) and OBI (liquidity measure with multi-level data) around announcement arrivals. The relative spread—i.e. the spread between the best bid and ask divided by the mid-price (hereafter just spread)—is calculated every 10 seconds around each event (for estimation and event windows). We run similar analysis on $\overline{\mathrm{SPREAD}}$ as presented for $\overline{\mathrm{OBI}}$, and the table and the figure are available in the Online Appendix.

One should note that from the practical point of view, spread and order book liquidity measure the transaction costs for different types of trades. Around news arrivals, informed investors may require immediacy to take advantage of stale limit orders, and in this situation, it is not necessarily possible to gradually execute a large block of orders to mitigate price impact. Hence, there can be single large trades or a bunch of simultaneous smaller trades that walk up the book through many levels, increasing the practical relevance of liquidity in the book beyond the best levels and necessitating the use of multi-level data to capture the real announcement effects.

We observe a quick and strong reaction in spread around scheduled announcements: the spread starts to rise slightly above the 95% confidence level around 10 to 30 minutes before the event and peaks immediately after the announcement release, after which it starts to gradually decline, reaching normal level within around 30 to 40 minutes. This indicates that around scheduled announcements, there is a relatively short time period during which the trading cost of trades at the best levels is above the normal level.

For non-scheduled announcements, $\overline{\mathrm{SPREAD}}$ is slightly above the normal level already before releases especially for negative announcements. After the non-scheduled announcements release, $\overline{\mathrm{SPREAD}}$ peaks to well above the long-term maximum value (obtained from the estimation window). Aggregated spread seems to stay above the long-term level during the whole 60 minute post event window.

Importantly, the observed effect of scheduled announcements on order books with data on multiple price levels is significantly different from what the spread between the best prices indicates. Whereas (i) $\overline{\mathrm{SPREAD}}$ rises significantly just 10 to 30 minutes before a scheduled announcement and (ii) new information does not improve it beyond its long-term level, $\overline{\mathrm{OBI}}$ (i) rises to an abnormally high level well (at least 60 minutes) before the scheduled news arrives and (ii) improves even beyond its one-month minimum in 20–30 minutes. Given that the order book slope reflects differences in investors views on the stock price and hence is related to information asymmetry, multi-level order book data provide evidence that scheduled news releases (especially positive ones) *reduce* the information asymmetry in stock markets in comparison to the long-term level, which cannot observed from the best levels.

6. Summary and conclusion

This paper examines the stock limit order book characteristics and liquidity around scheduled and non-scheduled company announcements using high frequency multi-level limit order book data of 75 frequently traded stocks listed on exchanges belonging to NASDAQ Nordic for the years 2006 to 2009. Parameterising the order book data with an order book slope enables us to measure the liquidity available over multiple price levels of the order books.

In the aggregated analysis, we find quite contradictory results for the multi-level order book liquidity measure (order book slope) and conventional bid–ask spread: whereas multi-level order book liquidity is exceptionally low at least an hour before the release of a scheduled announcement, the spread widens significantly only 10–30 minutes before the announcement comes. Additionally, unlike the spread, the multi-level liquidity measure improves to an exceptionally good level, even beyond its one-month record after the release of a scheduled announcement, which may indicate that scheduled announcements reduce information asymmetry. Hence, our results provide evidence that order book liquidity is significantly affected beyond the best price quotes and it can be misleading to use only top-of-the-book data, which is in line with the findings of Sensoy (2016). Additionally, we find that before the release of *non-scheduled* announcements, aggregated order book liquidity is above the normal level with statistical significance (especially on the ask side), which can indicate information leakage (see Graham et al., 2006).

We also find that the asymmetry of the order book is at an abnormally high level after the release of a non-scheduled announcement so that the ask (bid) side is abnormally illiquid in comparison to the bid (ask) side after positive (negative) news. This indicates that a news-momentum (news-contrarian) trading strategy is relatively expensive (cheap) after the release of non-scheduled announcements. Interestingly, scheduled announcements do not demonstrate such a phenomenon; they improve the liquidity on both sides of the order book equally.

There are some limitations in our study. Most importantly, we use multi-level data, but not order flow data (ITCH feed), which would be the most comprehensive data on order book markets. In future research, it would be interesting to use order flow data to study how order book liquidity is provided and consumed around announcements. For example, does the liquidity decrease due to trades (executions) or order cancellations? The statistical properties of message arrivals (submissions, cancellations, executions) around news releases would shed more light on the liquidity provision in the limit order markets around information releases.

Acknowledgments

The research leading to these results received funding from Tampere University of Technology through Networked Big Data Science and Engineering Center and TUT's doctoral school. The corresponding author is grateful for the grants received from Marcus Wallenberg's Foundation, Finnish Foundation for Technology Promotion, Foundation for Economic Education, and Savings Banks' Research Foundation.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.frl.2016.12.016

References

Butt, H., Virk, N., 2015. Liquidity and asset prices: an empirical investigation of the nordic stock markets. Eur. Financ. Manage. 21 (4), 672–705.
Campbell, J., 1997. The Econometrics of Financial Markets. Princeton University Press.
Degryse, H., De Jong, F., Van Kervel, V., 2015. The impact of dark trading and visible fragmentation on market quality. Rev. Finance 19, 1587–1622.
Degryse, H., De Jong, F., Van Ravenswaaij, M., Wuyts, G., 2005. Aggressive orders and the resiliency of a limit order market. Rev. Finance 9 (2), 201–242.
Deuskar, P., Johnson, T.C., 2011. Market liquidity and flow-driven risk. Rev. Financ. Stud. 24, 721–753.
Engle, R., Fleming, M., Ghysels, E., Nguyen, G., 2012. Liquidity, volatility, and flights to safety in the U.S. treasury market: Evidence from a new class of dynamic order book models. Technical Report. Staff Report, Federal Reserve Bank of New York.
Erenburg, G., Lasser, D., 2009. Electronic limit order book and order submission choice around macroeconomic news. Rev. Financ. Econ. 18, 172–182.
Gomber, P., Schweickert, U., Theissen, E., 2015. Liquidity dynamics in an electronic open limit order book: an event study approach. Eur. Financ. Manage. 21 (1), 52–78.
Gould, M., Porter, M., Williams, S., McDonald, M., Fenn, D., Howison, S., 2013. Limit order books. Quant. Finance 13 (11), 1709–1742.
Graham, J., Koski, J., Loewenstein, U., 2006. Information flow and liquidity around anticipated and unanticipated dividend announcements. J. Bus. 79 (5), 2301–2336.
Gross-Klussmann, A., Hautsch, N., 2011. When machines read the news: using automated text analytics to quantify high frequency news-implied market reactions. J. Empirical Finance 18 (2), 321–340.
Härdle, W., Hautsch, N., Mihoci, A., 2012. Modelling and forecasting liquidity supply using semiparametric factor dynamics. J. Empirical Finance 19, 610–615.
Krinsky, I., Lee, J., 1996. Earnings announcements and the components of the bid-ask spread. J Finance 51 (4), 1523–1535.
Lee, C., Mucklow, B., Ready, M., 1993. Spreads, depths, and the impact of earnings information: an intraday analysis. Rev. Financ. Stud. 6 (2), 345–374.
Malo, P., Pennanen, T., 2012. Reduced form modeling of limit order markets. Quant. Finance 12, 1025–1036.
Neuhierl, A., Scherbina, A., Schlusche, B., 2013. Market reaction to corporate press releases. J. Financ. Quant. Anal. 48 (4), 1207–1240.
O'Hara, M., Ye, M., 2011. Is market fragmentation harming market quality? J. Financ. Econ. 100, 459–474.
Riordan, R., Storkenmaier, A., Wagener, M., Zhang, S., 2013. Public information arrival: price discovery and liquidity in electronic limit order markets. J. Banking Finance 37 (4), 1148–1159.
Sensoy, A., 2016. Commonality in liquidity: effects of monetary policy and macroeconomic announcements. Finance Res. Lett. 16, 125–131.
Siikanen, M., Kanniainen, J., Luoma, A., 2016. What drives the sensitivity of limit order books to company announcement arrivals? Available at SSRN: https://ssrn.com/abstract=2891262.
Velásquez, S., Kanniainen, J., Mäkinen, S., Valli, J., 2016. Layoff announcements and intra-day market reactions doi:10.1007/s11846-016-0219-7, forthcoming in Review of Managerial Science.

Publication III

Siikanen M., Kanniainen, J., Luoma, A. “What drives the sensitivity of limit order books to company announcement arrivals?”, *Economics Letters*.

Available at: https://doi.org/10.1016/j.econlet.2017.07.018

Economics Letters 159 (2017) 65–68

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

What drives the sensitivity of limit order books to company announcement arrivals?

Milla Siikanen [a,*], Juho Kanniainen [a], Arto Luoma [b,1]

[a] Department of Industrial Management, Tampere University of Technology, P.O. Box 541, FI-33101 Tampere, Finland
[b] School of Management, FI-33014 University of Tampere, Finland

HIGHLIGHTS

- We study illiquidity shocks following scheduled and non-scheduled announcements.
- We measure liquidity over multiple limit order book levels from high-frequency data.
- Recent losses amplify illiquidity shocks following non-scheduled announcements.
- Faster market reactions in terms of order book illiquidity lead to larger shocks.
- Larger asymmetry in the book before announcements is associated with larger shocks.

ARTICLE INFO

Article history:
Received 21 December 2016
Received in revised form 13 July 2017
Accepted 17 July 2017
Available online 20 July 2017

JEL classification:
G10
G14

Keywords:
Limit order book
Liquidity
Company announcement
High-frequency data

ABSTRACT

We provide evidence that recent losses amplify order book illiquidity shocks caused by non-scheduled news. Moreover, the faster markets' reaction to scheduled and non-scheduled news arrivals is in terms of order book illiquidity, the more illiquid the order book becomes; that is, a fast reaction is a strong reaction. Additionally, order book asymmetry observed before announcement arrivals is positively associated with the magnitude of illiquidity shocks.

1. Introduction

Many studies show that information arrivals can cause liquidity shocks (see e.g. Erenburg and Lasser, 2009; Engle et al., 2012; Riordan et al., 2013; Rosa, 2016; Siikanen et al., 2017). However, to our knowledge there are no earlier studies investigating the factors which affect the magnitude of liquidity shocks in limit order books (LOB) caused by announcement releases. In this paper, we aim to explain the sensitivity of LOB liquidity to information arrivals using high-frequency LOB data for 75 companies from NASDAQ Nordic combined with set of scheduled and non-scheduled company announcements, for four-year period of 2006–2009.

LOB characteristics and the liquidity dynamics beyond the best levels are intriguing, especially around information arrivals, because high trading activity and investors' impatience may generate a sudden liquidity demand across multiple price levels. Thus, using the conventional bid–ask spread might lead to misleading results (Rosa, 2016; Sensoy, 2016; Siikanen et al., 2017). An appropriate method to characterise the LOB and to measure the LOB liquidity across multiple price levels should capture aspects with respect to both quantity (depth) over multiple levels and distances between the price levels. A popular approach is to estimate order book slope (see e.g. Deuskar and Johnson, 2011; Härdle et al., 2012; Malo and Pennanen, 2012; Siikanen et al., 2017), which in this paper is called Order Book Illiquidity (OBI).[2]

* Corresponding author.
E-mail address: milla.siikanen@tut.fi (M. Siikanen).
[1] Present address: Department of Mathematics and Statistics, P.O. Box 35, FI-40014 University of Jyväskylä, Finland.

[2] Some other liquidity measures exist that incorporate data from multiple LOB levels, too, such as the Exchange Liquidity Measure (XLM). However, an issue arises with XLM, because this measure is determined for a specific trade size, yet the total

http://dx.doi.org/10.1016/j.econlet.2017.07.018

Siikanen et al. (2017) find that after the immediate illiquidity shock, scheduled announcements can improve LOB liquidity to exceptionally good level and provide evidence for pre-reaction in LOBs before scheduled announcements, which suggests the possibility of information leakage (see also Graham et al., 2006).[3] Additionally, Riordan et al. (2013) and Gomber et al. (2015) study liquidity over multiple LOB levels in equity markets around information arrivals. Apart from these studies, Erenburg and Lasser (2009), Engle et al. (2012), and Rosa (2016) combine multi-level LOB data with macro announcements, but with data from equity-index-linked securities market, the U.S. Treasury market, and futures market, respectively. However, none of these studies looks extensively into the factors driving the LOB sensitivity, which is the focus of this paper.

We use 20 order book levels to calculate OBI, and one should note that this may affect the results presented here. Specifically, Siikanen et al. (2017) show that spread behaves quite differently around announcement releases when compared to OBI, so it is also likely that OBI calculated for example over 5 or 10 levels behaves differently from OBI over 20 levels. Additionally, we restrict our analysis to liquid stocks, and the results for illiquid stocks may differ considerably.[4]

2. LOB parametrisation and liquidity measure

To parametrise the LOB, we follow Malo and Pennanen (2012). The shape of a LOB is linearly captured as follows:

$$r(h) = \mathrm{OBI} \cdot h,$$

where

$$r(h) := \ln(s(h/\bar{s})) - \ln(\bar{s}),$$

where $\bar{s}$ refers to mid-price, and $h = \bar{s}x$ is the mark-to-market value of a market order of x shares. Here OBI is positive and is considered to measure LOB liquidity (see Malo and Pennanen, 2012).[5] Obviously, the smaller the value of OBI, the more liquid the stock is.

We use simple linear regression to calculate the values of OBI based on snapshots of the LOB taken every 10 s including data from 20 best ask and bid price levels. In case there are not 20 different price levels available on one side of the book, we use as many as are available. We also eliminate the effects of the pre- and post-trading sessions and exclude the first and last trading hours from the data.[6] In addition, we de-seasonalise the observations of ln(OBI).

3. Data

We use LOB data from 1.1.2006 to 1.1.2010 for 75 frequently traded stocks listed on NASDAQ OMX Helsinki, Stockholm, and Copenhagen, which are continuous limit order based markets. The stocks in our sample have been involved at some point in OMX Helsinki 25, OMX Stockholm 30, or OMX Copenhagen 20. Out of the 75 stocks, 27 are traded in Helsinki, 28 are traded in Stockholm, and 20 are traded in Copenhagen.[7]

The news data in this study come from NASDAQ OMX Nordic's website.[8] The announcement times are given at one second precision in the data, but because we sample the LOB data every 10 s, the times of the announcements are rounded to the nearest 10 s. We do not restrict our study to any specific news class, such as earnings announcements, as many other studies do. Rather, we re-categorise the announcements into two specific groups: scheduled and non-scheduled announcements (see Siikanen et al., 2017 for the categorisation). The final sample contains 408 scheduled and 2,629 non-scheduled announcements: 35%, 45%, and 20% originate from NASDAQ OMX Helsinki, Stockholm, Copenhagen, respectively. Over 70% of the scheduled announcements in the final sample are financial announcements.

4. Empirical analysis

4.1. Framework of the empirical analysis

An estimation window comprises observations with 10-second frequency from 27 days preceding the day of an event. An event window consists of two sub-windows: a 30-minute pre-window and another 30-minute post-window. We denote the set of observation times from the estimation window by $\mathcal{E}$, from the pre-window by $\mathcal{A}^-$, and from the post-window by $\mathcal{A}^+$, and from the whole event window by $\mathcal{A}$.

4.2. Regression variables

The dependent variable measures the relative magnitude of LOB illiquidity shock due to the release of information:

$$\Delta \ln(\mathrm{OBI})^{\mathrm{Max}}_{\mathcal{E},\mathcal{A}+} = \max_{t \in \mathcal{A}^+} [\ln(\mathrm{OBI})_t] - \ln(\mathrm{OBI})^{\mathrm{Med}}_{\mathcal{E}},$$

where

$$\ln(\mathrm{OBI})^{\mathrm{Med}}_{\mathcal{E}} = \underset{t \in \mathcal{E}}{\mathrm{Median}}[\ln(\mathrm{OBI})_t],$$

is a median observation from the estimation window.

We choose the explanatory variables to capture pre-reactions in the LOB, the sign (positiveness/negativeness) of new information, and the markets' reaction times. The first explanatory variable, $\ln(\mathrm{OBI})^{\mathrm{Med}}_{\mathcal{E}}$, is used to control for the preceding level of ln(OBI) (for the ask and bid sides separately).

Second, the expectations of the effects of the announcements may be visible in the pre-announcement window, which can indicate information leakage (see e.g. Lee, 1992; Graham et al., 2006; Siikanen et al., 2017). Intuitively, if the liquidity available on the ask side is exceptionally low in comparison to the bid side, it might be that the markets are expecting a positive announcement and vice versa. So, our second explanatory variable is the maximum

multi-level depths on the bid and ask sides vary in time and consequently, the LOB is not always deep enough to allow us to calculate XLM for a given order size. This is an issue especially just before scheduled announcements.

3 Siikanen et al. (2017) use largely the same data sets as we use in this study.

4 We thank an anonymous referee for pointing out these important observations.

5 Malo and Pennanen (2012) refer to LOB illiquidity as β, but for clarity, we refer to it as OBI, since in the finance literature β usually refers to CAPM β.

6 For stocks traded on OMX Helsinki and OMX Stockholm, we remove an additional half-hour from the end of the trading day because of the different length of the trading day in comparison to OMX Copenhagen.

7 We use data from Nordic markets instead of U.S. markets (the most liquid in the world) because the former are little fragmented in comparison to the latter. In the United States, fragmentation is clearly an important feature of equity markets (O'Hara and Ye, 2011). Another advantage of using Nordic data from less liquid markets is that, as Butt and Virk (2015) argue, "it is more appropriate to test liquidity-related models in markets that are sufficiently illiquid to diagnose the level and strength of bearing [...] risks".

8 http://www.nasdaqomxnordic.com/news/companynews, see the page also for detailed information.

Table 1
Association between LOB illiquidity shock and LOB related factors. The robust standard errors appear in parentheses. Economic significance levels are reported in braces and they give the change in the maximum relative price impact due to announcements: for α_1 – α_4 they are based on one standard deviation increase in explanatory variables and for α_5 on 10 min increase in τ^* (α_6 is an estimate of a dummy variable).

Parameter	Variable	Scheduled announcements		Non-scheduled announcements	
		ask	bid	ask	bid
α_1	$\ln(\text{OBI})_{\mathcal{E}}^{\text{Med}}$	-0.098^{*} (0.046) {−0.184}	-0.229^{**} (0.073) {−0.373}	-0.065^{**} (0.021) {−0.123}	−0.046 (0.024)
α_2	$\ln(\text{OBI})_{\mathcal{A}-}^{\text{Asymmetry}}$	0.121 (0.074)	0.445^{***} (0.107) {0.284}	0.072^{*} (0.028) {0.041}	0.291^{***} (0.045) {0.178}
α_3	$r_{\mathcal{E},\mathcal{A}-}$	−0.339 (0.543)	−0.407 (0.534)	-1.655^{***} (0.260) {−0.142}	-1.121^{***} (0.266) {−0.099}
α_4	$\text{SPREAD}_{\mathcal{A}-}^{\text{Med}}$	70.176^{***} (10.788) {0.188}	61.635^{***} (12.983) {0.163}	29.972^{***} (3.421) {0.086}	12.117 (9.610)
α_5	τ^*	-0.017^{***} (0.005) {−0.171}	-0.023^{**} (0.008) {−0.228}	-0.005^{***} (0.001) {−0.051}	-0.003^{*} (0.001) {−0.032}
α_6	D_+	0.150^{*} (0.075)	-0.261^{*} (0.107)	0.146^{***} (0.032)	-0.122^{***} (0.029)
Number of observations		408	408	2,629	2,629
R^2		0.081	0.166	0.098	0.095

*** $p < 0.001$.
** $p < 0.01$.
* $p < 0.05$.

asymmetry, calculated as

$$\ln(\text{OBI})_{\mathcal{A}-}^{\text{Asymmetry}} = \max_{t\in\mathcal{A}^-}\left[\left|\ln(\text{OBI})_t^{\text{Bid}} - \ln(\text{OBI})_t^{\text{Ask}}\right|\right].$$

To assess the effects of price pre-reactions to illiquidity shock, the third explanatory variable is

$$r_{\mathcal{E},\mathcal{A}-} = \ln\left[\underset{t\in\mathcal{A}^-}{\text{Median}}(m_t)\right] - \ln\left[\underset{s\in\mathcal{E}}{\text{Median}}(m_s)\right],$$

where m_t denotes the mid-price at time t.

The fourth explanatory variable is the median relative spread in the pre-window, denoted by $\text{SPREAD}_{\mathcal{A}-}^{\text{Med}}$. Here we use observations from the pre-event window to avoid a correlation with the other liquidity measure, $\ln(\text{OBI})_{\mathcal{E}}^{\text{Med}}$.

The fifth explanatory variable in the regression is time in minutes from the release of an announcement to the maximum value of ln(OBI) within the post-window. Formally, by setting $\min[\mathcal{A}^+] = 0$,

$$\tau^* = \underset{t\in\mathcal{A}^+}{\text{argmax}}\,[\ln(\text{OBI})_t]\,.$$

With this variable, we investigate how the length of "reaction time" from a release to an illiquidity shock is associated with the magnitude of the shock. One can also think of this variable from the perspective of aggressiveness: is there an association between how fast (i.e. aggressively) liquidity is consumed and what is the amount of liquidity consumed?

The last explanatory variable in our regression is a dummy variable for positive events, D_+, which is 1 if the news is positive—that is, the mid-price at the end of the post-window is larger than the mid-price right before the announcement. Importantly, we calculate τ^* and D_+ from post-window whereas other variables are calculated from pre-window. Means, medians, and standard deviations of the regression variables are available in the Online Appendix.

Overall, our stock-specific fixed effect regression is of the form:[9]

$$\Delta\ln(\text{OBI})_{\mathcal{E},\mathcal{A}+}^{\text{Max}} = \alpha_1\cdot\ln(\text{OBI})_{\mathcal{E}}^{\text{Med}} + \alpha_2\cdot\ln(\text{OBI})_{\mathcal{A}-}^{\text{Asymmetry}} + \alpha_3\cdot r_{\mathcal{E},\mathcal{A}-} + \alpha_4\cdot\text{SPREAD}_{\mathcal{A}-}^{\text{Med}} + \alpha_5\cdot\tau^* + \alpha_6\cdot D_+. \quad (1)$$

4.3. Regression results

Table 1 presents the regression results. The estimated regression coefficients for $\ln(\text{OBI})_{\mathcal{E}}^{\text{Med}}$ indicate that the more liquid the stock, the larger the illiquidity shock due to the announcement with both statistical and economic significance.

For $\ln(\text{OBI})_{\mathcal{A}-}^{\text{Asymmetry}}$, the estimated regression coefficients indicate that the larger the imbalance between the ask and bid sides before the event, the larger the maximum illiquidity cost after the announcement. The asymmetry seems not only be statistically but also economically significant, especially with the bid side illiquidity shock. This result supports the result of Chordia et al. (2002) that order imbalances in either direction reduce liquidity.

The estimates of $r_{\mathcal{E},\mathcal{A}-}$ for non-scheduled announcements suggest with statistical and economic significance that the more the stock price decreases (increases) before the announcement arrival, the larger (smaller) the impact of the announcement is on the magnitude of illiquidity. This finding seems consistent with the study of Hameed et al. (2010), who document that negative market returns decrease liquidity.[10] It is interesting that there is no apparent statistical association with scheduled announcements. Overall, this result can be used to understand investor reactions

---

9 Alternatively, Eq. (1) can be expressed as

$$\max_{t\in\mathcal{A}^+}[\text{OBI}_t] = (1+\alpha_1)\cdot\text{OBI}_{\mathcal{E}}^{\text{Med}} + \alpha_2\cdot\text{OBI}_{\mathcal{A}-}^{\text{Asymmetry}} + \alpha_3\cdot r_{\mathcal{E},\mathcal{A}-} + \alpha_4\cdot\text{SPREAD}_{\mathcal{A}-}^{\text{Med}} + \alpha_5\cdot\tau^* + \alpha_6\cdot D_+.$$

.

10 We perform an additional analysis to check if squared returns between the estimation window and illiquidity shocks at the post-event window, $r^2_{\mathcal{E},\mathcal{A}+}$, are associated with realised returns, $r_{\mathcal{E},\mathcal{A}-}$, but find no statistically significant associations (see Online Appendix).

to unexpected news releases: recent losses can make investors' liquidity provision more sensitive to announcements whose arrival is unexpected.

The regression coefficients estimated for $\text{SPREAD}^{\text{Med}}_{\mathcal{A}-}$ indicate with statistical and economic significance that the larger the relative spread before the announcement, the larger the illiquidity shock the announcement causes. Interestingly, this is contradictory with respect to other liquidity measure, $\ln(\text{OBI})^{\text{Med}}_{\mathcal{E}}$. However, they are calculated from different windows, which can partially explain the difference. Also, as demonstrated in Siikanen et al. (2017), multi-level liquidity and spread can behave differently around announcement releases (see also Sensoy, 2016; Rosa, 2016).

For τ^*, the estimated regression coefficient show that the faster the illiquidity shock occurs after the announcement, i.e. the faster the LOB illiquidity reaches its maximum after a news release, the larger the illiquidity shock is. As the table demonstrates, if the shock occurs 10 min later, its magnitude decreases by approximately 17%–23% for scheduled announcements and 3%–5% for non-scheduled announcements.

The parameter estimates for dummy variable D_+ show that, in comparison to negative news releases, positive news releases cause larger illiquidity shocks on the ask side and smaller shocks on the bid side and vice versa. This is reasonable because informed investors may buy (sell) shares by picking off stale sell (buy) limit orders just after the arrival of new positive (negative) information.

As a robustness check, we run the regressions using 60-minute pre- and post-event windows, and get similar results (the results are available in the Online Appendix). As an additional robustness check, we run the regressions using mean values instead of median values and the results are essentially the same as those reported for the median values and are available upon request.

5. Summary and conclusion

We perform regression analysis to explain the magnitude of the illiquidity shock that follows scheduled and non-scheduled company announcement releases, and find several associations with both statistical and economic significance. Most importantly, recent losses make the illiquidity shock following a non-scheduled announcement larger. Moreover, a fast reaction is a strong reaction; that is, the faster the LOB illiquidity reaches its maximum after a news release, the more illiquid the LOB becomes. We also provide evidence that the LOB asymmetry before both scheduled and non-scheduled announcements is positively associated with the magnitude of illiquidity shocks.

The results may be sensitive to the number of LOB levels used to determine multi-level liquidity, and future research should consider using different numbers of levels to see how this affects the findings. Additional analysis with different liquidity measures and on less liquid stocks could also provide new valuable insights on the topic. In the future research, it would also be interesting to use order flow data to study how different factors affect directly the order submission and cancellation rates (liquidity provision) around company announcements.

Acknowledgements

The research leading to these results received funding from Tampere University of Technology through Networked Big Data Science and Engineering Center and TUT's doctoral school. Additionally, this project has received funding from the European Union's Horizon 2020 research and innovation program under Marie Sklodowska-Curie grant agreement No. 675044. The corresponding author is grateful for the grants received from Marcus Wallenberg's Foundation, Finnish Foundation for Technology Promotion, Foundation for Economic Education and Savings Banks' Research Foundation. We highly appreciate the feedback of participants at Graduate School of Finance's Summer Workshop in Finance 2015. We also thank Henri Palonkoski and Jaakko Valli for the excellent research assistance. We are responsible for any remaining errors.

Appendix A. Online Appendix

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.econlet.2017.07.018.

References

Butt, H., Virk, N., 2015. Liquidity and asset prices: An empirical investigation of the nordic stock markets. European Financ. Manage. 21 (4), 672–705.
Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order imbalance, liquidity, and market returns. J. Financ. Econ. 65, 111–130.
Deuskar, P., Johnson, T.C., 2011. Market liquidity and flow-driven risk. Rev. Financ. Stud. 24, 721–753.
Engle, R., Fleming, M., Ghysels, E., Nguyen, G., 2012. Liquidity, Volatility, and Flights to Safety in the U.S. Treasury Market: Evidence from a New Class of Dynamic Order Book Models. Technical Report. Staff Report, Federal Reserve Bank of New York.
Erenburg, G., Lasser, D., 2009. Electronic limit order book and order submission choice around macroeconomic news. Rev. Financ. Econ. 18, 172–182.
Gomber, P., Schweickert, U., Theissen, E., 2015. Liquidity dynamics in an electronic open limit order book: An event study approach. European Financ. Manage. 21 (1), 52–78.
Graham, J., Koski, J., Loewenstein, U., 2006. Information flow and liquidity around anticipated and unanticipated dividend announcements. J. Bus. 79 (5), 2301–2336.
Hameed, A., Kang, W., Viswanathan, S., 2010. Stock market declines and liquidity. J. Finance 65 (1), 257–293.
Härdle, W., Hautsch, N., Mihoci, A., 2012. Modelling and forecasting liquidity supply using semiparametric factor dynamics. J. Empir. Finance 19, 610–615.
Lee, C., 1992. Earnings news and small traders: An intraday analysis. J. Account. Econ. 15 (2), 265–302.
Malo, P., Pennanen, T., 2012. Reduced form modeling of limit order markets. Quant. Finance 12, 1025–1036.
O'Hara, M., Ye, M., 2011. Is market fragmentation harming market quality? J. Financ. Econ. 100, 459–474.
Riordan, R., Storkenmaier, A., Wagener, M., Zhang, S., 2013. Public information arrival: price discovery and liquidity in electronic limit order markets. J. Bank. Finance 37 (4), 1148–1159.
Rosa, C., 2016. Walking on thin ice: Market quality around FOMC announcements. Econom. Lett. 138, 5–8.
Sensoy, A., 2016. Commonality in liquidity: effects of monetary policy and macroeconomic announcements. Finance Res. Lett. 16, 125–131.
Siikanen, M., Kanniainen, J., Valli, J., 2017. Limit order books and liquidity around scheduled and non-scheduled announcements: Empirical evidence from NASDAQ Nordic. Finance Res. Lett. 21, 264–271.

Publication IV

Siikanen M., Nögel, U., Kanniainen, J. “Liquidity in the FX market: empirical evidence from an aggregator”, *unpublished manuscript, submitted to a journal.*

Available at SSRN: https://ssrn.com/abstract=3208229

To appear in *SSRN*, Vol. 00, No. 00, Month 20XX, 1–15

Liquidity in the FX market: empirical evidence from an aggregator

MILLA SIIKANEN*†, ULRICH NÖGEL‡ and JUHO KANNIAINEN†

†DARE Business Data Research Group, Laboratory of Industrial and Information Management,
Tampere University of Technology, Finland
‡big xyt AG, Frankfurt am Main, Germany

(*released July 2018*)

In foreign exchange (FX) trading, an aggregator is used to connect traders with liquidity providers (LPs). In an aggregator, a trader receives a continuous stream of bid and ask quotes from a predefined set of LPs, and the difference between the best bid and ask prices over a set of liquidity streams is called an inside spread. In this paper, we empirically study liquidity in an FX aggregator and show that, on average, traders obtain a relatively tight spread already with four or five streams; the use of more streams yields a marginal benefit only. For given numbers of liquidity streams, we determine the optimal combinations of streams minimizing the spread. Our findings indicate that most of the traders could—at least in theory—reduce the average spread by more than half with the optimal combination of streams, and a trader could save up to \$0.18 basis points per €1 traded. However, traders may not be able to fully exploit improvements in spreads because, in practice, the liquidity streams are set by LPs and not by the trader. In addition, we find that Oomen's [Quantitative Finance, **17**, 3, (2017)] model fits our empirical data accurately, even under simplistic assumptions.

Keywords: FX market; Liquidity; Aggregation; Transaction costs

JEL Classification: G15

1. Introduction

The foreign exchange (FX) market is one of the largest financial markets in the world with an average daily traded volume of over five trillion U.S. dollars (Bank for International Settlements 2016). The FX market is characterized by a high degree of decentralization and fragmentation, and though it is commonly regarded as highly liquid, FX liquidity is not well understood (see e.g. Mancini *et al.* 2013, Karnaukh *et al.* 2015). FX market shows a high level of electronic trading with a complex ecosystem of both bilateral and multilateral electronic trading platforms with varying trading protocols. For example, there are platforms operated by large single dealer banks (single dealer platforms) and independent platform operators that combine liquidity with more than one bank as the liquidity provider (LP) (multi-dealer platforms) (see e.g. King *et al.* 2012). In general, FX markets are evolving rapidly in response to new trading technologies (King *et al.* 2013).

The current regulatory requirements, like MiFID II, which came into force this year, are expected to further push FX trading on platforms, even if the exact shift in the market share and further regulatory-driven transformation of the FX market are still in the early stages (see e.g. Sherif 2018). In summary, the FX market shows the very unique characteristic of being still an over-the-counter

*Corresponding author. Email: milla.siikanen@tut.fi

(OTC) market, but an extremely liquid one with a very high degree of electronic trading. Further on, due to its complexity and opaque nature, the FX market is currently under pressure due to its criticized lack of transparency and recent scandals[1], which were partially addressed in the recently published 'FX Code of Conduct'.[2]

During recent years, the FX market has been undergoing a transition from pure OTC to electronic trading with a structure and intrinsic trading dynamics partially similar to central limit order books. A typical modern trading technology, especially in FX spot trading, but also typical in other OTC markets, is an aggregator (Oomen 2017a). An aggregator is an electronic trading tool that connects liquidity seekers (traders) with LPs. The idea is to facilitate best-price execution: traders receive continuous streams of bid and ask quotes from selected (predefined) LPs and can choose to trade with the best price. One of the key metrics characterizing a liquidity aggregator is the trader's observed inside spread (Oomen 2017a). The observed inside spread measures the transaction costs and is defined as the difference between the best bid and ask prices over a set of liquidity streams disclosed for a trader.

According to Mancini *et al.* (2013) and Karnaukh *et al.* (2015), liquidity in FX markets has not been studied in depth, despite of FX market's the high importance due to its huge size and the crucial role in guaranteeing efficiency and arbitrage conditions also in many other markets. One reason for this could be the overall fragmentation of the market which makes the available and representative data fairly sparse. Regarding aggregator trading technology, which plays an important role in the FX markets, a welcomed theoretical contribution is Oomen (2017a), who has introduced a promising stochastic model for liquidity dynamics in an aggregator. However, to our knowledge, there exists no empirical research utilizing real market data from a liquidity aggregator. This paper aims to fill this gap by using a unique data set for FX spot trading from an FX aggregator. We not only aim to reflect our empirical results in light of Oomen's model but also, by looking at the liquidity streams that individual traders access in an aggregator, respond to Gould *et al.* (2017), who recently suggested (in the context of quasi-centralized limit order books for FX trading) that 'An interesting challenge for future research will be to gain a deeper understanding of the subset of liquidity [...] that individual institutions can access in their local LOBs [limit order books]'.[1]

The data set that we use includes detailed information on all the streamed quotes for individual LPs and individual traders on EURUSD over a 10-day trading period. Such detailed data are usually not available for academic research. To our knowledge, this is the first paper to study empirical data from an FX aggregator. By studying the empirical properties of an FX aggregator with a disclosed streaming protocol, we aim to obtain deeper insight into the new and exciting ground of the FX OTC market. In particular, since traders may access different numbers of liquidity streams in their aggregator settings, we determine the optimal combinations of streams to minimize the inside spreads. We do the optimization for different numbers of liquidity streams (corresponding to traders observing different numbers of streams) and compare the optimal spreads to the inside spreads that the traders in the aggregator actually observe. Additionally, we quantify the theoretical

[1] See, for example, 'Third Barclays Trader Faces U.S. Charges in FX Scandal' (Bloomberg 2018) and 'The global FX rigging scandal' (Reuters 2017).

[2] FX GLOBAL CODE: A set of global principles of good practice in the FX market, developed in a partnership between central banks and market participants from 16 jurisdictions around the globe, available at `https://www.globalfxc.org/docs/fx_global.pdf`.

[1] Besides Oomen (2017a,b), earlier research on FX market liquidity in an aggregator is scant, but there exists literature focusing on liquidity in FX markets in general. Mancini *et al.* (2013) study systemically liquidity in the FX market and find significant variation in liquidity across exchange rates, substantial illiquidity costs, and strong commonality in liquidity across currencies and with equity and bond markets. Karnaukh *et al.* (2015) provide a comprehensive study on liquidity in the FX market and show that liquidity declines with funding constraints and global risk, and that liquidity can be measured accurately using daily data. Payne (2003) studies information asymmetries in inter-dealer FX markets and finds that asymmetric information accounts for around 60% of the average spreads. Banti *et al.* (2012) construct a measure of global liquidity risk in the FX market and find that liquidity risk is priced in the cross-section of currency returns. Danielsson and Payne (2012) study order level data from an FX broking system and analyze several liquidity measures, and they find strong predictability in the arrival of liquidity supply and demand events.

cost savings that traders could obtain with the optimal combination of streams (compared to the present setting). Moreover, using these data, we show that the stochastic model formulated by Oomen (2017a) can yield empirically reasonable solutions.

Our empirical analysis shows that the more liquidity streams are observed, the lower the observed inside spread tends to be, as suggested by the model introduced by Oomen (2017a)[1]. However, there exists variation among traders: in some cases, a trader with fewer streams observes a lower average spread than a trader with more streams. According to our data, traders, on one hand, obtain a relatively tight spread already, on average, with four or five streams, meaning that those with more than five streams obtain only a marginal benefit, if any. On the other hand, if traders can obtain the optimal combination of liquidity streams, most of them could cut their observed inside spread to less than half. Moreover, an optimal combination of just four streams would give a better spread than any trader observes at the moment (the average number of streams that a trader observes is 5.4 and the maximum is 23). Oomen (2017a) shows in his model that the trader's expected observed inside spread can become negative, and we show with our data that indeed, the optimal combination of streams leads to a negative average spread when the number of streams in the aggregator setting is seven or more. Additionally, our results indicate that the most active trader would save over $1 million per year in transaction costs if he/she could obtain the optimal combination of streams. Finally, we provide empirical evidence for the model introduced by Oomen (2017a): even if we assume homogenous streams, which is quite simplistic, Oomen's model provides a good fit for the real world data and can be used to describe the observed spread quite accurately.

It should be noted that the cost savings and improvements in average spreads are only theoretical, since (i) traders are not completely free to choose just any liquidity streams provided by the LPs, and (ii) due to the last look trading protocol[2], streamed prices are indicative in nature. Moreover, changing the trading more towards all-to-all type of trading would be likely to change the LPs' behaviour. However, the unique insights into the trading behaviour in an aggregator with disclosed streaming that we present in this paper indicate that there is room for improvement both for optimizing the selection of liquidity streams in a trader's aggregator setting and for the evolution of the FX market in general. We leave the realization of the theoretical cost savings to be explored by practitioners.

The remainder of the paper is structured as follows. Section 2 describes the logic of the liquidity aggregator in detail. Then Section 3 describes the data and Section 4 outlines the optimization method used. Section 5 depicts the empirical results and Section 6 describes the calibration of the model (Proposition 1) proposed by Oomen (2017a). Finally, Section 7 discusses and concludes this study.

2. Liquidity aggregator and the selection of liquidity streams

In general, an aggregator can be defined as an electronic trading tool that collects streaming price quotes from different sources, such as FX dealers, electronic brokers and multibank trading systems (King *et al.* 2012). Aggregators are not only used in spot FX markets; the trading protocol in one form or another is taking an increasingly prominent role in many other OTC markets (Oomen 2017a). In the following, we only consider an aggregator connecting liquidity seekers (traders) with LPs without further connecting the trader to other external sources of liquidity.

While in regulated equity markets, such as in classical stock exchanges, centralized limit order books (CLOBs) are typical[3], FX trading is still OTC and the trading venues usually allow a variety

[1]However, Oomen (2017a) notes that nominal spread (as opposed to observed inside spread), which takes into account the slippage costs caused by possible trade rejections, may not decrease as the number of liquidity streams increases. Nevertheless, studying nominal spreads is beyond the scope of this paper.
[2]Last look trading protocol gives an LP a chance choose to either accept or reject the deal request within a certain time period, typically some milliseconds, after a trader requests the deal (see Oomen 2017a,b, Cartea *et al.* 2018)
[3]See, e.g. Malo and Pennanen (2012), Siikanen *et al.* (2017a,b) and references therein for research on stock markets with

of trading protocols (see e.g. King *et al.* 2012). Typical trading protocols in FX markets include, for example, request for quote (RFQ) and disclosed streaming (DS). While in CLOBs, by default, a trader always trades with the best price available in the order book, the situation is more complex in the OTC FX markets and their typical trading venues.

The logic of an aggregator with DS protocol is between the traditional RFQ protocol, which mimics the classical voice broking in an electronic form and a fully transparent all-to-all (CLOB) style trading. In an RFQ model, the trader actively sends out 'requests for quotes' for a number of dealers (i.e. LPs)—that is, asking prices for a transaction with a given quantity for a given currency pair (e.g. 1Mio EURUSD). Then LPs can (but are not obliged to) respond by quoting their bid and ask prices. The trader then can choose to enter the trade with the LP quoting the most favourable price.[1]

In an aggregator, the trader receives a continuous stream of quotes from a set of LPs with whom he/she can trade. Thus, the LPs are constantly providing bid and ask prices for certain quantities and currency pairs without receiving an explicit request for a certain quote. In a sense, a trader observes his/her 'personal order book', though in contrast to CLOBs, the trader cannot submit quotes (limit orders) himself/herself. Another important difference when compared to CLOBs is that in an aggregator, the LPs streaming quotes are unable to observe the quotes of other LPs (Oomen 2017a). CLOBs are also usually anonymous, whereas in an aggregator, both parties know the identity of the counterparty.

Another interesting trading protocol used in FX markets is the quasi-centralized limit order book (QCLOB) (see Gould *et al.* 2017), which can be thought to be between an aggregator and CLOB. QCLOBs are otherwise like CLOBS, but in QCLOBs, investors can access the trading opportunities offered only by counterparties with whom they possess sufficient bilateral credit. Consequently, in QCLOBs (as also in aggregators), investors observe only part of the total liquidity, as they observe their 'local order books'. However, in QCLOBs, all market participants can submit both limit and market orders, whereas in aggregators, LPs provide liquidity (in a sense, they 'submit limit orders') and traders take liquidity (by 'submitting market orders') (for more information, see Gould *et al.* 2017).

One of the key metrics in an aggregator is (trader's) observed inside spread (Oomen 2017a). A (half-) inside spread measures the transaction costs that the trader pays at the execution of a trade and, in general, spread can be seen as a measure of liquidity. Formally, the observed inside spread is defined as the difference between the best bid and ask prices that the trader observes at a specific moment. To illustrate this, Table 1 shows a snapshot of the active quotations that a trader with four streams observes at a specific point in time. A stream can provide an arbitrary number of price and quantity pairs for both the bid and ask sides. The observed inside spread is the difference between the lowest ask price and the highest bid price: in this case, the observed inside spread equals $1.12545 - 1.12542 = 0.00003 = 0.3$ basis points (bp).

Taking into account the OTC nature of the aggregator in question, it should be clearly pointed out that an LP can (and usually will) provide different streams of quotes for certain clients (or subsets of clients). The decision of an LP usually depends on certain criteria, like the client's credit rating or their general business relationship; in other words, the LP might provide his/her 'favourite' clients with better prices. In fact, it has been reported that 'FX trading is becoming increasingly relationship-driven' (Moore *et al.* 2016). Further on, the streamed quotes are highly dynamic on a microstructure scale and depend on the movement of the FX market as a whole and on the LP's current personal internal FX flow, exposure and general trading behaviour. For instance, some LPs in general internalize trades more often (offset against the LP's internal trade flow) or tend to externalize (offload the trade on the platform itself) (see also Butz and Oomen

CLOBs.

[1]Nevertheless, it should be noted that the trader can still decide to trade with an LP not quoting the best possible price, but the trader may add other criteria to his decision, e.g. his/her general relationship to the LP, where he/she might just choose his/her 'favourite' LP with whom he/she is in a general business relationship.

Table 1. **Snapshot of most recent quotes observed by a trader with four streams (A, B, C, D).** Each stream can have an arbitrary number of price and quantity pairs for both the bid and ask sides. The best bid and ask prices are highlighted with **bold**, and the trader's observed inside spread is calculated based on these.

	Bid quotes		Ask quotes	
Stream	price	quantity [M]	price	quantity [M]
A	1.12513	1	1.12587	1
A	1.12512	2	1.12588	2
B	1.12536	0.5	**1.12545**	0.5
B	1.12535	1	1.12546	1
B	1.12533	3	1.12549	3
C	1.12536	0.5	1.12547	0.5
C	1.12534	1	1.12550	1
D	**1.12542**	1	1.12549	1
D	1.12539	2	1.12551	2
D	1.12536	3	1.12557	3
D	1.12532	5	—	—

2017, King *et al.* 2012).[2]

In addition, because of the last look trade acceptance process adopted in the aggregator, the streamed quotes are indicative in nature. 'Last look' gives an LP a chance to accept or reject a deal request within a fixed period (usually some milliseconds) after the trader has initiated the deal request (Oomen 2017a,b, Cartea *et al.* 2018). This means that even though a trader would observe a negative spread in his/her aggregator setting, the arbitrage opportunity is only theoretical, since LPs may reject trade requests.

As the liquidity provision to a trader is highly bespoke, and as it can be assumed that LPs' behaviour is generally fairly heterogeneous, the problem of a trader deciding on a suitable aggregation setup is highly non-trivial (see also Oomen 2017a). For instance, how many and which LPs should the trader include in his/her aggregator setup and which ones should he/she choose? Moreover, because of the OTC nature of the trading in FX markets, two parties trading need a bilateral trading agreement and sufficient bilateral credit to trade (see also Gould *et al.* 2017), which may further restrict the selection of LPs.

In the following analysis, based on unique empirical microstructure data, we focus on the question of how many quoted liquidity streams a trader should include to obtain the 'best execution' in the sense of minimal bid-ask spreads, and compare the inside spreads observed by the traders with their current setups to the optimal combination of liquidity streams. However, it is important to keep in mind that it is the LP who makes a decision about the liquidity streams that are provided to a trader (i.e. a trader can choose the LP, but the LP chooses the streams). Hence, the optimal combinations of liquidity streams that we find in Section 5 are not necessarily achievable by traders in practice, but they should be considered as the most efficient set of streams, and thus serve as a yardstick for the actual situation.

3. Streaming data and descriptive statistics

In this paper, we use a detailed and unique data set on all the streamed quotes from an FX aggregator. The data set is highly sensitive, and we have permission to publish our academic results as long as it is not possible to identify individual LPs or traders. To guarantee the anonymity of LPs and traders, we are not allowed to publish the name or any other background information of the data provider.

[2]Lyons (1995) studies the trading of a spot FX dealer and finds support for both inventory control and asymmetric information channels, through which the order flow affects prices according to microstructure theory. Menkhoff and Schmeling (2010) study information heterogeneity across FX traders and show that information about the counterparty of a trade affects the future trading decisions of individual traders. Moreover, they argue that 'microstructure research has shown in several ways that foreign exchange markets are populated by heterogenous participants'.

We analyse quote data for EURUSD, from an FX liquidity aggregator. EURUSD was the most traded currency pair in 2016 with 23% of the traded volume (Bank for International Settlements 2016). Our sample covers 10 trading days from 26 September 2016 to 7 October 2016. Table 2 provides descriptive statistics on the data sample. During this two-week period, the traded volume is over €10 billion and there are almost 12,000 trades. Figure 1 plots the intra-day pattern of how the trading is distributed over a trading day. Around 90% of the trading (measured both by traded volume and number of trades) takes place between 6:00 and 16:00, and hence we restrict our analysis to data from these hours.

Table 2. **Descriptive statistics on the data sample.** The sample covers a period of 10 trading days on EURUSD. '# of streams with > 95% of obs' provides the number of streams that have active quotes for over 95% of the minute-by-minute observations, and values in Panel B are calculated over these 165 'active' streams. AQS stands for average quoted spread (of a stream), and AOIS stands for (trader's) average observed inside spread in the aggregator. When the presented number is an average value, the median value is presented in parentheses ().

Panel A: Aggregator		
Traded volume [M€]	10,541	
# of trades	11,448	
Average (median) trade size [€]	920,782	(500,000)
# of liquidity providers	42	
# of quote updates [M]	306	
# of traders	105	
# of streams	190	
# of streams with > 95% of obs	165	
Panel B: Average stream		
Average (median) traded volume [M€]	86	(24)
Average (median) number of trades	89	(22)
Average (median) number of quote updates [M]	1.62	(1.32)
Average (median) AQS [bp $]	1.3083	(0.8656)
Panel C: Average trader		
Average (median) traded volume [M€]	100	(21)
Average (median) number of trades	109	(24)
Average (median) number of streams	5.4	(5)
Average (median) AOIS [bp $]	0.3796	(0.3373)

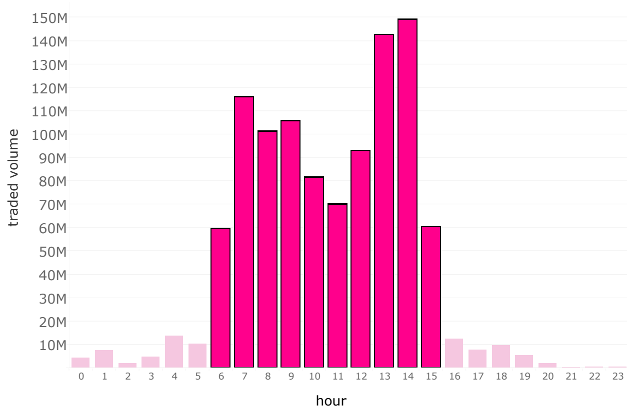

Figure 1. **Hourly average volume traded per day.** The unit of traded volume is €. The darker bars present data from 6:00–16:00, constituting 92% of the traded volume.

Altogether, there are 190 streams provided by 42 LPs. This means that one LP discloses on average 4.5 streams, highlighting the point discussed in Section 2 that an LP provides different streams for different (subsets of) customers. Over all the streams, there were over 300 million quote updates during the sample period. We sample the quote data for each stream each minute to obtain the available bid and ask prices and respective quantities. From these, for every observation time point and stream, we determine the highest bid price and lowest ask price. Then we subtract the best bid prices from the best ask prices to get the time series of quoted spread for the streams. We multiply the absolute spread values by 10,000 to have basis points as units. From these, we calculate average quoted spreads (AQSs) for each stream. Some streams are not quoting actively during our sample period, and hence we restrict our analysis to streams for which we have at least 95% of the minute-by-minute observations, yielding a sample of 165 streams. Each of these streams has, on average, 89 trades with a total traded volume of € 68 million, and over 1.5 million quote updates, with median AQS of 0.8656.

In the final sample, we have 105 traders with at least one trade. The average traded volume (number of trades) per trader over the period of 10 trading days is just over € 100 million (109), while the median is around 21 million (24). We approximate the set of streams that a trader observes with the set of streams that the trader trades with; hence, in reality, the number of observed streams may be higher, as there may be streams from which a trader receives quotes, but with which he/she does not trade during our sample period. On average, a trader trades with 5.4 streams, while the minimum is 1 and the maximum is 23. Given the set of streams that an investor observes, we determine the best bid and ask prices to form the time series of the observed inside spread. We get the average observed inside spread (AOIS) as the average of this time series and see that a trader's average (median) AOIS is 0.3796 bp (0.3373).

4. Optimization of the combination of liquidity streams

Given a number of streams in a trader's aggregation setting, our optimization problem is to select the set of streams with the minimum AOIS. Formally, we can formulate our combinatorial optimization problem as follows. We denote the set of all possible streams as $\mathbb{M} = \{1, 2, ..., M\}$, where M, the total number of streams, equals 165. Additionally, let $\{b_1^{(s)}, b_2^{(s)}, ..., b_T^{(s)}\}$ be the time series of the minute-by-minute observations of the highest bid prices and, correspondingly, $\{a_1^{(s)}, a_2^{(s)}, ..., a_T^{(s)}\}$ for the lowest ask prices for stream $s \in \mathbb{M}$, where $T = 6,010$ is the number of minute-by-minute observations per stream. Let n be the number of streams in a trader's aggregation setting and $\mathbf{n} \subset \mathbb{M}$ be the subset of $\mathbb{M}$ including n selected streams. Our problem is to find the optimal combination of streams $\mathbf{n}^* \subset \mathbb{M}$ that minimizes the AOIS over the set of all streams:

$$\mathbf{n}^* = \arg\min_{\mathbf{n}\subset\mathbb{M}} \sum_{t=1}^{T} [\min_{p\in\mathbf{n}} a_t^{(p)} - \max_{q\in\mathbf{n}} b_t^{(q)}]/T.$$

As the total number of streams is 165, calculating the AOIS for all possible combinations (i.e. to use brute force) is not feasible: for example, in a setting of 10 streams, the number of possible combinations is 3.12E+15. The problem is also non-linear: given two sets of streams, adding the same stream to both sets does not improve the average inside spread by the same amount.

To solve the optimization problem, we use a genetic algorithm (GA). GA is a metaheuristic optimization method inspired by natural selection (see e.g. Holland 1992, Contreras-Bolton and Parada 2015).[1] The basic idea is to

(i) create a population of possible solutions (i.e. multiple sets of n streams)
(ii) calculate fitness values (i.e. AOISs) for the solutions in the population

[1] GA was also adapted in previous research in the field of economics and finance; see e.g. Schlottmann and Seese (2004), Acosta-González *et al.* (2012), Chen (2013).

(iii) generate randomly the next population of possible solutions through crossover and mutation while guaranteeing that the best solutions found so far (sets of streams with lowest AOISs) move to the next population
(iv) continue steps (ii)–(iii) until some predefined stopping criteria are met, for example, if the average improvement in best fitness value is less than a given limit or a maximum number of generations is reached

For each aggregator set-up with number of streams $n = \{2, 3, \ldots, 15\}$, we run GA 10 times and select the best solution.[1] As GA is a metaheuristic method, there is no guarantee that we will find the global optimum; instead, we find a sufficiently good solution. To be exact, the actual solution (global optima) is not worse than the optimized AOIS for a given number of streams. However, for the purposes of this analysis, this is sufficient, as the analysis in the following Section 5 shows that the AOISs which we find are still substantially lower than what the traders observe and, when n increases, we get relatively close to the minimum AOIS over all 165 streams in the aggregator. Additionally, for cases where $n \leq 5$, we calculate the true optima with brute force and find that the genetic algorithm works quite well: when we repeat the GA 10 times for each $n = 2, 3, 4, 5$, the best solution over the 10 runs is always the global optima.

5. Optimal and observed inside spreads

Figure 2 shows the optimal inside spread as a function of the number of streams in an aggregator set-up, n. In addition, the figure depicts the variation of AOISs over the traders. We see that increasing the number of streams typically reduces the inside spread, although there is variation among investors and sometimes a trader observing a smaller number of streams observes a better spread than another trader with more streams. Interestingly, on average, a trader with four or five streams obtains a relatively tight spread compared to other traders having more streams. Therefore, increasing the number of streams to more than five yields a quite marginal benefit, if any. We also see that for all n, the optimal (minimum) average inside spreads is clearly lower than the AOISs that traders observe. In general, most of the traders would cut their observed spread by less than half given that they could obtain the optimal combination of streams. In fact, with the optimal combination of just four streams, an investor could observe a lower inside spread than any trader observes currently with any number of streams.

In general, the shape of the curve formed by the optimal spread points resembles the theoretical prediction of Oomen (2017a) (see Figure 2. Panel A in Oomen 2017a, we also get back to this in Section 6). The model proposed by Oomen (2017a) also suggests that in an aggregator, the expected observed inside spread may become negative as n increases. One reason why negative spreads may arise quite naturally is that LPs cannot observe the prices streamed by other LPs. In the case of our data, the optimal average inside spread is negative when n is larger or equal to seven. The minimum spread over all 165 streams is –0.0968 bp. A negative observed inside spread would introduce an arbitrage opportunity with firm quotes, but the last look practice, which makes the quoted prices indicative, may prevent an investor from exploiting the arbitrage (see also Oomen 2017a,b, Cartea *et al.* 2018).[2] Moreover, even if the traders can choose the LPs, they cannot choose the streams disclosed by the LPs; therefore, it may not be possible in practice to select this optimal combination of liquidity streams in the aggregator.

[1] We do not run the analysis for $n > 15$ because there is only one customer observing more than 15 streams. For more details on parameter specifications, see Appendix A.
[2] Gould *et al.* (2017) study FX market liquidity in quasi-centralized limit order books and identify periods with negative global spread. In a sense, this corresponds to negative inside spread in our aggregator setting when combining quotes from all streams. Akram *et al.* (2008) study arbitrage opportunities in FX markets and document short-lived, economically significant deviations from the covered interest rate parity. Foucault *et al.* (2017) study short-lived toxic arbitrage opportunities, where traders take advantage of stale quotes. They show with FX market data that with more frequent toxic arbitrage opportunities and faster responses by trades to these opportunities, liquidity is affected negatively.

Additionally, we can see from Figure 2 that there is considerable variation in AOIS values between different investors with the same n. In many cases, for example, when the number of streams is 1, 2, 4, or 7, the AOIS of the trader with the lowest value is less than half that of the one with the highest value. This means that even though two traders would have the same number of streams in their aggregator setting, the inside spreads that they observe can differ quite markedly.

Table 3 presents the differences in average trader's AOIS with the optimal inside spreads for different values of n. The half-difference in spread indicates the amount saved in trading costs per one € traded when trading with the optimal spread instead of AOIS. We see that depending on n, the (theoretical) saved amount is between $0.1329 bp and $0.1801 bp $ per one € traded. The table also shows that the (theoretical) amount that an average trader would save in transaction costs per year varies from around $800 to over $1 million.

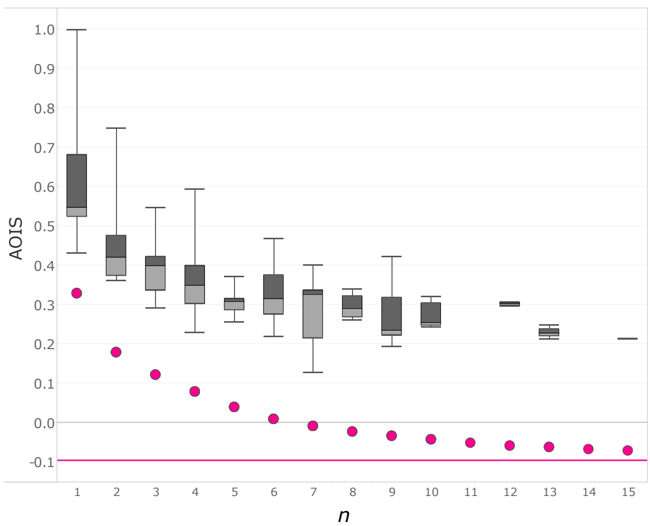

Figure 2. **Optimal and observed inside spreads.** n gives the number of streams in the aggregator setting and the red circles give the optimal inside spreads. The box plot gives the median, maximum and minimum average observed inside spread (AOIS) values with the 25^{th} and 75^{th} percentiles for investors with given numbers of streams. Values for inside spread are depicted in basis points (bp). The pink line gives the inside spread over all 165 streams. In addition, there is one investor with 23 observed streams, and his/her AOIS is 0.2058 bp (omitted in the figure).

Table 3. **Comparison of optimal inside spread and average trader's AOIS.** Values presented in the table are scaled to yearly level. The 'GA optimum' panel gives the optimal average inside spread obtained using a genetic algorithm. The 'Average trader' panel gives the average observed inside spread (AOIS), number of trades and traded volume for an average trader with n streams. The 'Comparison' panel gives half of the difference between average trader's AOIS and optimal spread—that is, the (theoretical) amount saved in transaction costs per €1 traded, and the average saved amount per year if a trader would be able to trade with the optimal spread instead of AOIS. *The values for a trader with 23 streams are calculated according to the optimal inside spread with 15 streams.

		GA optimum	**Average trader**			**Comparison**	
n	# of investors	optimal average inside spread (bp)	AOIS (bp)	# of trades per trader per year	traded volume per trader per year (M €)	difference in half-spread (bp)	avg saved amount per trader per year ($)
1	20	0.3279	0.6059	182	81	0.1390	1,119
2	10	0.1774	0.4507	78	55	0.1367	756
3	7	0.1210	0.3932	847	650	0.1361	8,846
4	11	0.0771	0.3639	1,324	523	0.1434	7,507
5	9	0.0392	0.3050	3,605	1,514	0.1329	20,118
6	12	0.0090	0.3283	1,469	1,528	0.1596	24,396
7	9	–0.0093	0.2889	1,517	3,168	0.1491	47,247
8	4	–0.0247	0.2949	1,905	1,891	0.1598	30,223
9	9	–0.0345	0.2654	8,233	7,302	0.1499	109,466
10	7	–0.0435	0.2742	3,146	2,474	0.1589	39,305
11		–0.0528					
12	2	–0.0591	0.3010	9,633	3,693	0.1801	66,494
13	3	–0.0636	0.2290	7,037	6,600	0.1463	96,571
14		–0.0687					
15	1	–0.0719	0.2122	9,100	4,722	0.1420	67,070
23	1		0.2058	55,562	78,429	0.1388*	1,088,943*

6. Calibration to model for expected observed spread

As stated in Section 5, the shape of the curve formed by the optimal spread points in Figure 2 resembles the curve for expected observed spread in (Oomen 2017a, Figure 2. Panel A). To further test Oomen's model against our empirical data, we calibrate his model to our data in a special case where LPs are assumed to be homogeneous (see Oomen 2017a, Proposition 1). Oomen (2017a) assumes that the true (logarithmic) price process follows a random walk, and the mid-price of a stream[1] i deviates from the true price by $m_t^{(i)}$. Then $m_t^{(i)}$ follows an AR(1) process (without a constant), and the error term of that process is depicted with $\eta_t^{(i)} \sim$ i.i.d. $\mathcal{N}(0, (1-\beta^2)\omega^2)$, where $0 \leq \beta < 1$ is the autocorrelation coefficient of the AR(1) process, and $\rho_{i,j} = \text{corr}(\eta_t^{(i)}, \eta_t^{(j)})$. According to Proposition 1 by Oomen (2017a), for a panel of n homogeneous streams quoting a spread of s, the expected observed spread in the aggregator is

$$S = s - 2\omega\sqrt{1-\rho}\psi_n, \tag{1}$$

where $\psi_n = \mathbb{E}\left(max_i\{u_i\}_{i=1}^n\right)$ for $u_i \sim$ i.i.d. $\mathcal{N}(0,1)$ (Eq 5 in Oomen 2017a). To sum up, the parameters of the model to be calibrated are $s, s > 0$ (spread of individual stream), ω ('dispersion') and $\rho, -1 \leq \rho \leq 1$ (correlation between the error terms of two streams). The form of Eq 1 (Eq 5 in Oomen 2017a) prevents us from estimating ω and ρ without fixing one of them because otherwise, an increase in value of ρ can be counterbalanced with an increase of ω, and vice versa, and thus there would be an arbitrary number of ρ and ω pair combinations that lead to the same fit.

We calibrate the model by minimizing the mean squared error between model-implied values of S (Eq 1) and the average AOIS stemming from the data. In particular, we use average AOIS values over actual traders given the number of streams in the aggregator setting to calibrate the model for the data. Since there are only a few stray traders observing more than 10 streams, we use values of n up to 10. Additionally, we use several different fixed values for both ρ and ω, corresponding to low and high correlation and dispersion values. Table 4 shows the parameter estimates and Figure 3 plots the fitted model with the lowest mean squared error and data points (calibrated models (A)–(D) and (H)–(J) in Table 4 all lead to the same fitted values up to the precision of six decimal points).

Figure 3 shows that the mapping based on Proposition 1 in (Oomen 2017a) is reasonable: the model fits well when we calibrate s and ω (or ρ) freely. Table 4 shows that panels (A)–(D) and (H)–(J) provide the best fit (lowest MSE), and regardless of the fixed value for ρ or ω, the parameter estimate for inside spread is 0.587. This is quite close to the median spread that the streams quote in the aggregator, which is 0.866 (see Table 2). When ρ (ω) increases, the value of ω (ρ) increases as well. Panels (F) and (G) show that fixing the value of ω at a very low value leads to a poor model fit, since ρ (correlation) is naturally limited to –1, and hence it cannot counterbalance values of ω lower than 0.076 (see panel (A) in Table 4). Calibrating parameters for panel (J) in Table 4 leads to numerical issues with a matrix being singular to working precision when setting $\rho = 1$, and hence the fit is not as good as for panels (A)–(D) and (H)–(J).

7. Discussion and conclusion

In this paper, we study liquidity in an FX aggregator. With our unique, detailed data set, we show that a trader with a larger number of streams typically observes a lower inside spread, thought there is considerable variation among the traders, and in some cases traders with fewer

[1] We refer to 'stream' where Oomen (2017a) refers to 'LP', since in our aggregator, an LP may be streaming multiple liquidity streams.

Table 4. **Calibrated parameter estimates.** The table gives the estimated parameter values. The fixed values are depicted with *italics*: we calibrate the model using several values for correlation ρ and ω, i.e. 'dispersion'. p-values are reported in parentheses (). MSE gives the calibrated model's mean squared error.

	(A)	(B)	(C)	(D)	(E)
Fixed ρ:	*–1*	*–0.5*	*0*	*0.5*	*1*
s	0.587	0.587	0.587	0.587	0.357
	(1.02E–13)	(1.02E–13)	(1.02E–13)	(1.02E–13)	(0.500)
ω	0.076	8.79E-02	0.108	1.52E-01	0.549
	(1.27E–10)	(1.27E–10)	(1.27E–10)	(1.27E–10)	(0.500)
MSE	2.51E–04	2.51E–04	2.51E–04	2.51E–04	9.94E–03

	(F)	(G)	(H)	(I)	(J)
Fixed ω:	*0.01*	*0.05*	*0.1*	*0.35*	*0.5*
s	0.387	0.508	0.587	0.587	0.587
	(3.29E–09)	(4.46E–13)	(1.02E–13)	(1.02E–13)	(1.02E–13)
ρ	–1.000	–1.000	–0.158	0.905	0.954
	(1.000)	(1.000)	(0.984)	(4.44E-16)	(0.00E+00)
MSE	7.56E–03	1.39E–03	2.51E–04	2.51E–04	2.51E–04

streams still observe better spreads than traders with more streams. On average, traders obtain a relatively tight spread already with four or five streams, and traders with more streams obtain a marginal benefit only, if any. Additionally, we show that if traders could choose the optimal combination of streams, most of them could cut their observed spread by more than half while maintaining the same number of liquidity streams. Moreover, with seven or more streams, the

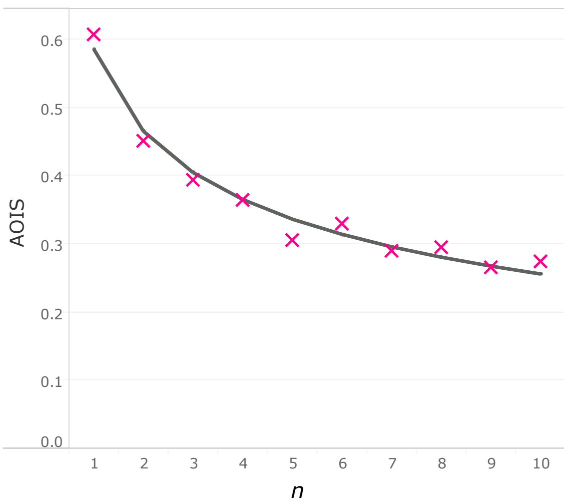

Figure 3. **Calibrated model fit.** n gives the number of streams in the aggregator setting and the red crosses give the average inside spreads that a trader observes with n streams in his/her aggregator setting. The grey line depicts the calibrated model fit. The parameter values of the model correspond to models (A)–(D) and (H)–(J) in Table 4, which all lead to the same fitted values for up to 6 decimal points. Values for inside spread are depicted in basis points (bp).

optimal combination leads to negative average spreads. We also calibrate the model (Proposition 1) proposed by Oomen (2017a) to our empirical data and find that though we study the model under a simplistic assumption of homogeneous liquidity streams, the model can be fitted to the real world data accurately and be used to describe the markets.

For completeness, it should be mentioned that the analysis only captures part of the 'real world'. The aggregator (with DS protocol) is in the market evolution between the classical OTC with the electronic version of a platform-based RFQ trading model and CLOBs. Therefore, a trader sees a constant stream of quotes, but in a disclosed fashion, such as only the streamed quotes that he/she is receiving from his/her LPs. Respectively, the LP quotes the stream for this particular trader or a subset of traders and cannot see streamed prices disclosed by other LPs. Moreover, because of the OTC nature of the trading, a trader may only have a trading relationship with a limited number of LPs. Furthermore, even if a trader can select an LP, it is the LP that selects the streams that are observable to the trader.

Additionally, the LP's quoting behaviour usually depends on multiple factors, like the client's credit rating, trading behaviour (e.g. algo trader vs institutional investor driven by physical FX demand) and even more soft criteria, such as the general business relationship of the LP with the client (trader). Therefore, it is also not completely straightforward that in a totally transparent market (e.g. CLOB trading protocol), the quoting behaviour of LPs would be exactly the same as in a DS model (aggregator), as LPs would have to provide quotes which fit all their clients. Moreover, unlike in an aggregator, in a CLOB, LPs would observe the quotes of the other LPs. Therefore, this enhanced transparency of a CLOB might affect the behaviour of the LPs.

Nevertheless, this analysis provides valuable and unique insights into the trading behaviour in the DS (aggregator) market model and the FX market as an OTC market in general. This analysis shows that there is still significant room for improvement in optimizing the combination of liquidity streams and in the evolution of the FX market in general. The ongoing discussion and regulatory pressure (e.g. enhanced transparency requirements with MiFID II) may further improve the efficiency of the market to benefit the end clients.

Acknowledgements

We thank big xyt AG, and especially Wieslaw Nosal, for all of their support in handling the data. We highly appreciate the comments and feedback of participants at the Graduate School of Finance's summer workshop in Finance 2018. The first author is grateful for the grants received from the OP Group Research Foundation, Marcus Wallenberg's Foundation and Waldemar von Frenckells stiftelse.

Disclaimer

The views and opinions expressed in this paper reflect the authors' personal views on the subject and do not necessarily represent the views of big xyt AG. This article is necessarily general and is not intended to be comprehensive, nor does it constitute legal or financial advice in relation to any particular situation.

References

Acosta-González, E., Fernández-Rodríguez, F. and Sosvilla-Rivero, S., On factors explaining the 2008 financial crisis. *Economics Letters*, 2012, **115**, 215–217.

Akram, Q.F., Rime, D. and Sarno, L., Arbitrage in the foreign exchange market: Turning on the microscope. *Journal of International Economics*, 2008, **76**, 237–253.

Bank for International Settlements, Foreign exchange turnover in 2016, Triennial Central Bank Survey. , 2016.
Banti, C., Phylaktis, K. and Sarno, L., Global liquidity risk in the foreign exchange market. *Journal of International Money and Finance*, 2012, **31**, 267–291.
Butz, M. and Oomen, R., Internalisation by Electronic FX Spot Dealers. , 2017.
Cartea, Á., Jaimungal, S. and Walton, J., Foreign exchange markets with last look. *Mathematics and Financial Economics*, 2018.
Chen, S.H., *Evolutionary computation in economics and finance*, Vol. 100, , 2013, Physica.
Contreras-Bolton, C. and Parada, V., Automatic combination of operators in a genetic algorithm to solve the traveling salesman problem. *PloS one*, 2015, **10**, e0137724.
Danielsson, J. and Payne, R., Liquidity determination in an order-driven market. *The European Journal of Finance*, 2012, **18**, 799–821.
Foucault, T., Kozhan, R. and Tham, W.W., Toxic arbitrage. *The Review of Financial Studies*, 2017, **30**, 1053–1094.
Gould, M.D., Porter, M.A. and Howison, S.D., Quasi-centralized limit order books. *Quantitative Finance*, 2017, **17**, 831–853.
Holland, J.H., *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, 1992, MIT press.
Karnaukh, N., Ranaldo, A. and Söderlind, P., Understanding FX liquidity. *The Review of Financial Studies*, 2015, **28**, 3073–3108.
King, M.R., Osler, C. and Rime, D., Foreign Exchange Market Structure, Players, and Evolution. *Handbook of Exchange Rates*, 2012, pp. 1–44.
King, M.R., Osler, C.L. and Rime, D., The market microstructure approach to foreign exchange: Looking back and looking forward. *Journal of International Money and Finance*, 2013, **38**, 95–119.
Lyons, R.K., Tests of microstructural hypotheses in the foreign exchange market. *Journal of Financial Economics*, 1995, **39**, 321–351.
Malo, P. and Pennanen, T., Reduced form modeling of limit order markets. *Quantitative Finance*, 2012, **12**, 1025–1036.
Mancini, L., Ranaldo, A. and Wrampelmeyer, J., Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *The Journal of Finance*, 2013, **68**, 1805–1841.
Menkhoff, L. and Schmeling, M., Trader see, trader do: How do (small) FX traders react to large counterparties trades?. *Journal of International Money and Finance*, 2010, **29**, 1283–1302.
Moore, M., Schrimpf, A. and Sushko, V., Downsized FX markets: causes and implications. *BIS Quarterly Review*, 2016.
Oomen, R., Execution in an aggregator. *Quantitative Finance*, 2017a, **17**, 383–404.
Oomen, R., Last look. *Quantitative Finance*, 2017b, **17**, 1057–1070.
Payne, R., Informed trade in spot foreign exchange markets: an empirical investigation. *Journal of International Economics*, 2003, **61**, 307–329.
Schlottmann, F. and Seese, D., Modern heuristics for finance problems: A survey of selected methods and applications. In *Handbook of Computational and Numerical Methods in Finance*, pp. 331–359, 2004, Springer.
Sherif, N., Single-dealer platforms win in Mifid forex shake-up. *Risk.net*, 2018.
Siikanen, M., Kanniainen, J. and Luoma, A., What drives the sensitivity of limit order books to company announcement arrivals?. *Economics Letters*, 2017a, **159**, 65–68.
Siikanen, M., Kanniainen, J. and Valli, J., Limit order books and liquidity around scheduled and non-scheduled announcements: Empirical evidence from NASDAQ Nordic. *Finance Research Letters*, 2017b, **21**, 264–271.

Appendix A: Details on genetic algorithm set up

The genetic algorithm optimization was run using MATLAB's `ga`-function with R2017b version and using parallel computing with eight parallel workers. Table A1 presents the option values specified for the `ga`-function (for other options, MATLAB's default values were used).

The option values were tested with $n = \{2, 3, 4, 5\}$, and `FunctionTolerance` was set to a low enough value so that the algorithm found the true optimum at least with a few runs out of the 10 total runs. `PopulationSize`, `MaxTime`, `EliteCount`, `MaxStallGenerations` and `MaxGenerations` were set to increase as n increases, although they increase with a lower rate than the number of possible combinations. `CrossoverFraction` was set to a slightly lower value when compared to the default value to add more randomization to the algorithm. The optimal combination of $n-1$ streams with an added random stream is seeded to the initial population for n using `InitialPopulationMatrix` to ensure that the optimum found for n is not lower than for $n-1$ (the spread is strictly non-decreasing when a new stream is added to the set).

Table A2 presents statistics on the optimization runs (we run the optimization algorithm 10 times for each $n = \{2, 3, \ldots, 15\}$). For all runs with $n < 6$, and some runs with $n = 6$, the algorithm converges. For larger values of n, the optimization is terminated because the time limit is reached. For $n > 9$, some of the runs have a substantially smaller number of generations because the computer momentarily slowed down (not evident in Table A2). However, the average number of evaluations of the fitness function (i.e. how many stream combinations were evaluated during the optimization) still increases with n.

Table A1. **Option values for GA.** The table presents the option values set for MATLAB's `ga`-function. Other values are MATLAB's default values and can be checked from MATLAB's `ga`-function documentation.

Option	Value	Description (from MATLAB's `ga`-function documentation)
`FunctionTolerance`	1E–100,000,000	'The algorithm stops if the average relative change in the best fitness function value over `MaxStallGenerations` generations is less than or equal to `FunctionTolerance`.'
`PopulationSize`	`ceil`(`max`(n^3.8, 50))	'Size of the population'
`MaxTime`	60*20*n	'The algorithm stops after running after `MaxTime` seconds, as measured by `tic` and `toc`. This limit is enforced after each iteration, so ga can exceed the limit when an iteration takes substantial time.'
`CrossoverFraction`	0.75	'The fraction of the population at the next generation, not including elite children, that the crossover function creates'
`EliteCount`	`min`(`ceil`(`max`(n^4, 50)*0.05), 20)	'Positive integer specifying how many individuals in the current generation are guaranteed to survive to the next generation'
`MaxStallGenerations`	`max`(n^5, 100)	'The algorithm stops if the average relative change in the best fitness function value over `MaxStallGenerations` generations is less than or equal to `FunctionTolerance`.'
`UseParallel`	`TRUE`	'Compute fitness and nonlinear constraint functions in parallel'
`MaxGenerations`	`max`(n^6, 300)	'Maximum number of iterations before the algorithm halts.'
`InitialPopulationMatrix`	The best combination found with $n-1$ streams, with a random stream added	'Initial population used to seed the genetic algorithm. Has up to `PopulationSize` rows and N columns, where N is the number of variables. You can pass a partial population, meaning one with fewer than `PopulationSize` rows. In that case, the genetic algorithm uses `CreationFcn` to generate the remaining population members.'

Table A2. **Satistics on GA runs.** There are 10 runs for each n. Stopping criteria equal to 1 means that the algorithm converged ('Optimization terminated: average change in the penalty fitness value less than options. FunctionTolerance and constraint violation are less than options. ConstraintTolerance.') and stopping criteria equal to 2 means that the time limit was exceeded ('Optimization terminated: time limit exceeded.'). The function count is the number of evaluations of the fitness function.

n	stopping criteria	avg # generations	avg function count
2	1	144	7,231
3	1	294	19,451
4	1	1,300	253,677
5	1	3,773	1,709,714
6	1 & 2	10,063	9,117,532
7	2	14,727	23,962,457
8	2	13,093	35,393,083
9	2	14,250	60,252,806
10	2	9,622	60,720,500
11	2	7,243	65,663,243
12	2	5,650	71,291,755
13	2	4,371	74,768,041
14	2	3,307	74,968,163
15	2	2,846	83,855,441

Appendix

Appendix A

This Appendix describes filtering of the company announcement data set. In addition, Tables 1, 2, 3, 4, 5, and 6 give the lists of stocks included in Articles II and III from Helsinki, Stockholm, and Copenhagen stock exchanges and the numbers of scheduled and non-scheduled announcements per company in each stage of the filtering.

The following rules are used to filter the company announcements:

- Announcements that clearly contain no new information are excluded. These mostly include announcements published in multiple languages, in which case only the first one is involved.[2]
- Announcements for which there are not enough data to from the 27-day estimation window are removed from the sample.
- To study the immediate reactions, the sample is restricted to announcements published during trading hours, with enough data from that day to get the pre- and post-event samples (i.e. announcement release in the middle of the trading day).
- If several announcements are published at the same second on the same stock, then only one of them is involved and the others are excluded.
- The cases where there has been a trading halt ceasing continuous trading near the announcement time (within the 30 or 60 minute pre- or post-window around the announcement release) are not considered.

[2]For NASDAQ OMX Stockholm (Copenhagen), the announcements are commonly published both in English and Swedish (Danish), and for NASDAQ OMX Helsinki, many companies publish the announcements in English, Finnish, and Swedish.

Table 1: Numbers of scheduled company announcement releases for Helsinki stock exchange.

						30 minute window		60 minute window	
ISIN	Company	first release & new information	with 27 previous days in data	inside trading hours	same time excluded	middle of trading day	no trading halts	middle of trading day	no trading halts
FI0009000681	Nokia	12	10	8	8	7	7	7	7
SE0000667925	TeliaSonera	32	22	3	3	2	2	1	1
FI0009005987	UPM	49	35	22	20	18	18	6	6
FI0009007132	Fortum	29	21	4	4	1	0	0	0
FI0009013403	Kone	30	18	17	17	17	16	16	16
FI0009000285	Amer Sports	30	24	22	22	20	20	19	19
FI0009013429	Cargotec	35	21	18	18	15	14	9	9
FI0009007884	Elisa	34	21	5	5	1	1	1	1
FI0009000459	Huhtamäki	27	17	1	1	1	1	1	1
FI0009005870	Konecranes	29	19	9	9	1	1	1	1
FI0009000202	Kesko	30	21	11	11	3	3	2	2
FI0009004824	Kemira	28	21	3	3	3	3	1	1
FI0009007835	Metso	32	23	21	21	21	20	9	9
FI0009902530	Nordea	17	13	1	1	0	0	0	0
FI0009013296	Neste	26	19	3	3	2	2	1	1
FI0009005318	Nokian Renkaat	25	18	3	3	1	1	1	1
FI0009014377	Orion	31	17	14	14	9	9	7	7
FI0009014575	Outotec	20	14	1	1	1	1	0	0
FI0009003222	Pohjola	67	45	9	9	5	5	4	4
FI0009003552	Rautaruukki	31	22	4	4	4	4	1	1
FI0009003305	Sampo	23	17	0	0	0	0	0	0
FI0009005961	Stora Enso	31	20	11	11	8	7	5	5
FI0009003727	Wärtsilä	29	22	6	6	2	2	2	2
FI0009800643	YIT	36	26	6	6	4	4	4	4
FI0009002422	Outokumpu	29	20	17	17	15	14	13	13
FI0009007694	Sanoma	23	16	11	11	3	3	0	0
FI0009000277	Tieto	26	17	2	2	2	2	2	2
	Total	811	559	232	230	166	160	113	113

Table 2: Numbers of scheduled company announcement releases for Stockholm stock exchange.

ISIN	Company	first release & new info	with 27 previous days in data	inside trading hours	same time excluded	30 minute window: middle of trading day	30 minute window: no trading halts	60 minute window: middle of trading day	60 minute window: no trading halts
CH0012221716	ABB	25	17	4	4	4	4	4	4
SE0000695876	Alfa	32	23	14	14	10	10	10	10
SE0000255648	Assa Abloy	23	15	3	3	3	3	3	3
SE0000101032	Atlas Copco	56	35	29	29	25	23	23	23
GB0009895292	Astra Zeneca	25	16	15	15	15	15	14	13
SE0000869646	Boliden	26	16	14	13	12	9	6	6
SE0000103814	Electrolux	44	30	3	3	1	1	1	1
SE0000108649	Ericsson	53	33	1	1	0	0	0	0
SE0000202624	Getinge	29	19	13	13	11	11	9	9
SE0000106270	Hennes & Mauritz	24	14	1	1	0	0	0	0
SE0000107419	Investor	27	17	1	1	0	0	0	0
SE0000825820	Lundin Petroleum	28	18	1	1	0	0	0	0
SE0000412371	Modern Times Group	26	17	12	12	12	12	12	12
SE0000427361	Nordea Bank	51	34	20	20	5	5	4	4
SE0000667891	Sandvik	28	18	4	4	4	4	4	4
SE0000171886	SCA	26	19	12	12	9	8	8	8
SE0000308280	SCANIA	26	17	13	13	8	8	6	6
SE0000148884	SEB	50	30	14	14	5	5	5	5
SE0000163594	Securitas	28	17	1	1	0	0	0	0
SE0000193120	Svenska Handelsbanken	26	18	10	10	10	10	8	8
SE0000113250	Skanska	28	19	1	1	0	0	0	0
SE0000108227	SKF	23	14	8	8	8	8	8	8
SE0000171100	SSAB	52	37	21	17	14	14	10	10
SE0000310336	Swedish Match	22	14	2	2	2	2	0	0
SE0000314312	Tele2	24	16	4	4	3	3	3	3
SE0000667925	TeliaSonera	31	21	2	2	1	1	1	1
SE0000115446	Volvo	28	18	2	2	2	2	1	1
SE0000242455	Swedbank	41	28	15	18	14	14	13	13
	Total	902	590	240	238	178	172	153	152

Table 3: Numbers of scheduled company announcement releases for Copenhagen stock exchange.

ISIN	Company	first release & new information	with 27 previous days in data	inside trading hours	same time excluded	30 minute window: middle of trading day	30 minute window: no trading halts	60 minute window: middle of trading day	60 minute window: no trading halts
DK0060534915	Novo Nordisk	19	14	0	0	0	0	0	0
DK0060336014	Novozymes	20	15	5	5	5	0	0	0
DK0060228559	TDC	24	19	4	4	2	1	0	0
DK0060477503	Topdanmark	17	13	12	12	12	8	8	8
DK0060013274	Tryg	23	18	4	4	3	1	1	1
DK0010268606	Vestas Wind Systems	15	12	0	0	0	0	0	0
DK0010268440	William Demant Holding	24	18	9	9	9	4	4	4
DK0060448595	Coloplast	26	20	13	13	10	6	6	6
DK0010274414	Danske Bank	25	21	12	12	12	5	3	1
DK0010234467	FLSmidth	16	12	10	10	10	7	7	7
DK0010272202	Genmab	41	32	0	0	0	0	0	0
DK0010272632	GN Store Nord	20	17	13	13	11	7	7	7
DK0010307958	Jyske Bank	21	16	13	13	3	2	0	0
DK0010244508	AP Moeller Maersk	16	12	10	10	8	5	5	5
SE0000427361	Nordea Bank	23	19	7	7	1	1	1	1
DK0010287234	H Lundbeck	20	16	0	0	0	0	0	0
DK0010207497	Danisco	24	19	18	18	16	11	8	8
DK0010287663	NKT Holding	15	12	9	9	6	5	4	3
DK0010311471	Sydbank	20	16	16	16	15	10	10	10
DK0010181676	Carlsberg	25	20	4	4	4	3	3	3
	Total	434	341	159	159	127	76	67	64

Table 4: Numbers of non-scheduled company announcement releases for Helsinki stock exchange.

ISIN	Company	first release & new information	with 27 previous	inside trading hours	same time excluded	30 minute window: middle of trading day	30 minute window: no trading halts	60 minute window: middle of trading day	60 minute window: no trading halts
FI0009000681	Nokia	98	81	43	43	21	20	17	17
SE0000667925	TeliaSonera	215	160	102	102	59	59	51	51
FI0009005987	UPM	146	106	66	66	45	45	35	35
FI0009007132	Fortum	121	86	50	50	29	29	25	25
FI0009013403	Kone	94	56	41	41	29	29	28	28
FI0009000285	Amer Sports	118	89	59	59	34	34	31	31
FI0009013429	Cargotec	147	99	56	56	37	37	35	35
FI0009007884	Elisa	105	70	40	40	27	27	23	23
FI0009000459	Huhtamäki	86	58	21	21	13	13	11	11
FI0009005870	Konecranes	137	108	94	94	59	59	47	47
FI0009000202	Kesko	103	72	43	43	29	29	22	22
FI0009004824	Kemira	84	58	33	33	25	25	22	22
FI0009007835	Metso	237	189	146	146	96	96	88	88
FI0009902530	Nordea	175	106	95	71	47	47	17	17
FI0009013296	Neste	75	51	34	34	26	26	24	24
FI0009005318	Nokian Renkaat	76	50	37	37	19	19	16	16
FI0009014377	Orion	130	57	29	29	12	12	9	9
FI0009014575	Outotec	59	38	22	22	15	14	13	13
FI0009003222	Pohjola	222	141	105	105	55	55	44	44
FI0009003552	Rautaruukki	146	117	79	79	51	51	39	39
FI0009003305	Sampo	112	78	52	52	31	30	19	19
FI0009005961	Stora Enso	267	180	105	105	62	62	56	56
FI0009003727	Wärtsilä	83	55	28	28	15	15	14	14
FI0009800643	YIT	122	94	64	64	40	39	34	34
FI0009002422	Outokumpu	96	57	39	39	27	27	19	19
FI0009007694	Sanoma	99	61	44	44	23	23	20	20
FI0009000277	Tieto	186	114	74	74	58	58	47	47
	Total	3,539	2,431	1,601	1,577	984	980	806	806

Table 5: Numbers of non-scheduled company announcement releases for Stockholm stock exchange.

						30 minute window		60 minute window	
ISIN	Company	first release & new info	with 27 previous days in data	inside trading hours	same time excluded	middle of trading day	no trading halts	middle of trading day	no trading halts
CH0012221716	ABB	239	150	126	126	90	90	83	83
SE0000695876	Alfa	131	92	35	35	20	20	18	18
SE0000255648	Assa Abloy	59	40	20	20	13	13	10	10
SE0000101032	Atlas Copco	187	121	72	72	39	38	26	26
GB0009895292	Astra Zeneca	324	226	123	123	72	72	58	58
SE0000869646	Boliden	86	52	38	38	27	27	24	24
SE0000103814	Electrolux	42	25	17	17	10	10	8	8
SE0000108649	Ericsson	449	242	181	176	62	61	45	45
SE0000202624	Getinge	51	40	22	22	14	14	11	11
SE0000106270	Hennes & Mauritz	73	54	7	7	1	1	1	1
SE0000107419	Investor	59	46	21	21	9	9	5	5
SE0000825820	Lundin Petroleum	208	152	16	16	10	10	7	7
SE0000412371	Modern Times Group	150	90	39	39	25	25	22	22
SE0000427361	Nordea Bank	152	91	80	78	49	49	33	33
SE0000667891	Sandvik	74	53	34	34	17	17	9	9
SE0000171886	SCA	99	77	67	66	39	39	30	30
SE0000308280	SCANIA	216	151	123	123	77	77	66	66
SE0000148884	SEB	239	142	121	120	57	57	35	34
SE0000163594	Securitas	96	68	42	42	18	18	14	14
SE0000193120	Svenska Handelsbanken	369	323	298	298	153	153	123	123
SE0000113250	Skanska	435	286	97	97	57	57	43	43
SE0000108227	SKF	130	79	63	63	46	46	36	36
SE0000171100	SSAB	55	46	29	29	15	12	11	11
SE0000310336	Swedish Match	31	17	9	9	6	6	4	4
SE0000314312	Tele2	146	75	34	34	21	21	15	15
SE0000667925	TeliaSonera	244	158	99	99	55	54	46	46
SE0000115446	Volvo	417	275	205	205	115	115	88	88
SE0000242455	Swedbank	430	307	171	172	94	94	78	78
	Total	5,191	3,478	2,189	2,181	1,211	1,205	949	948

Table 6: Numbers of non-scheduled company announcement releases for Copenhagen stock exchange.

ISIN	Company	first release & new information	with 27 previous days in data	inside trading hours	same time excluded	30 minute window: middle of trading day	30 minute window: no trading halts	60 minute window: middle of trading day	60 minute window: no trading halts
DK0060534915	Novo Nordisk	41	35	18	18	9	9	8	8
DK0060336014	Novozymes	21	17	12	12	10	7	4	3
DK0060228559	TDC	113	70	43	43	27	19	17	17
DK0060477503	Topdanmark	36	25	20	20	14	13	10	10
DK0060013274	Tryg	40	28	18	18	14	13	10	10
DK0010268606	Vestas Wind Systems	170	131	80	80	49	42	30	30
DK0010268440	William Demant Holding	22	16	14	14	6	5	4	4
DK0060448595	Coloplast	80	75	61	61	38	33	27	27
DK0010274414	Danske Bank	109	88	74	74	36	35	23	23
DK0010234467	FLSmidth	146	116	86	86	50	36	32	32
DK0010272202	Genmab	234	195	87	87	47	36	30	30
DK0010272632	GN Store Nord	62	43	33	33	18	11	11	11
DK0010307958	Jyske Bank	43	38	32	32	20	15	11	11
DK0010244508	AP Moeller Maersk	38	28	20	20	14	14	11	11
SE0000427361	Nordea Bank	60	47	39	39	30	30	21	21
DK0010287234	H Lundbeck	150	123	49	49	30	24	19	19
DK0010207497	Danisco	40	35	24	24	16	14	13	13
DK0010287663	NKT Holding	59	38	35	35	24	19	14	14
DK0010311471	Sydbank	34	28	27	27	24	19	15	14
DK0010181676	Carlsberg	130	117	77	77	52	50	41	40
	Total	1,628	1,293	849	849	528	444	351	348

Appendix B

Online appendix of Article I:

Siikanen M., Baltakys, K., Kanniainen, J., Vatrapu, R., Mukkamala, R., Hussain, A. "Facebook drives behavior of passive households in stock markets", *Finance Research Letters*.

Available at: https://doi.org/10.1016/j.frl.2018.03.020

Facebook drives behavior of passive households in stock markets

Milla Siikanen[*,1], Kęstutis Baltakys[1], Juho Kanniainen[1], Ravi Vatrapu[2,3], Raghava Mukkamala[2,3], and Abid Hussain[2]

[1] DARE Business Data Research Group, Laboratory of Industrial and Information Management, Tampere University of Technology, Finland
[2] Centre for Business Data Analytics, Copenhagen Business School, Denmark
[3] Westerdals Oslo School of Arts, Communication and Technology, Norway

Online Appendix: Complete regression tables

For descriptions of the variables and regression equations, see Section *3. Framework of empirical analysis* in the paper

Investor sectors

Nonfinancial companies

1. Companies, posts

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.34255	0.055783	-6.1408	318	2.4477e-09	-0.4523	-0.2328
FB_posts	0.010967	0.0015177	7.226	318	3.7113e-12	0.0079811	0.013953
scD(t-1)	1.8742	0.057405	32.649	318	1.4251e-103	1.7613	1.9871
NEWS(T)	-0.10647	0.016281	-6.5397	318	2.4586e-10	-0.13851	-0.074441
Ret(t)	-5.5532	0.10637	-52.207	318	5.0366e-158	-5.7624	-5.3439
Ret(t-1)	0.17885	0.13126	1.3626	318	0.17396	-0.079387	0.43709
Y(2011)	-0.30798	0.030747	-10.017	318	1.0467e-20	-0.36847	-0.24749
Y(2012)	-0.32176	0.031833	-10.108	318	5.1894e-21	-0.38439	-0.25913
Y(2013)	-0.73415	0.036955	-19.866	318	1.1378e-57	-0.80686	-0.66144
Y(2014)	-0.64433	0.03684	-17.49	318	1.8486e-48	-0.71681	-0.57185
Y(2015)	-0.61576	0.038191	-16.123	318	3.6959e-43	-0.69089	-0.54062
Y(2016)	-0.025703	0.032439	-0.79233	318	0.42876	-0.089525	0.03812
M(Feb)	0.18572	0.034169	5.4354	318	1.0905e-07	0.11849	0.25294
M(Mar)	0.13361	0.036486	3.662	318	0.00029294	0.061829	0.2054
M(Apr)	0.27877	0.036435	7.6511	318	2.3957e-13	0.20709	0.35046
M(May)	-0.15644	0.035718	-4.3798	318	1.6145e-05	-0.22671	-0.086166
M(Jun)	0.15413	0.036492	4.2237	318	3.1417e-05	0.082336	0.22593
M(Jul)	0.12814	0.036428	3.5177	318	0.00049872	0.056472	0.19981
M(Aug)	-0.083501	0.034778	-2.401	318	0.016925	-0.15192	-0.015077
M(Sep)	-0.079554	0.035298	-2.2538	318	0.024892	-0.149	-0.010106
M(Oct)	0.22586	0.033608	6.7204	318	8.388e-11	0.15974	0.29198
M(Nov)	0.065511	0.034227	1.914	318	0.056512	-0.0018281	0.13285
M(Dec)	-0.14215	0.034438	-4.1277	318	4.6851e-05	-0.20991	-0.074397

2. Companies, comments

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.30743	0.05575	-5.5144	318	7.253e-08	-0.41711	-0.19774
FB_comments	5.5534e-06	3.3498e-06	1.6578	318	0.098342	-1.0373e-06	1.2144e-05
scD(t-1)	1.9194	0.057063	33.636	318	9.1555e-107	1.8071	2.0317
NEWS(T)	-0.11705	0.016207	-7.2219	318	3.8087e-12	-0.14893	-0.085159
Ret(t)	-5.5641	0.10647	-52.259	318	3.7858e-158	-5.7735	-5.3546
Ret(t-1)	0.31624	0.12987	2.435	318	0.015441	0.060722	0.57176
Y(2011)	-0.28513	0.030771	-9.266	318	2.9999e-18	-0.34567	-0.22459
Y(2012)	-0.27476	0.032301	-8.5065	318	7.135e-16	-0.33831	-0.21122
Y(2013)	-0.67357	0.036362	-18.524	318	1.7792e-52	-0.74511	-0.60203

[*] Corresponding author, milla.siikanen@tut.fi

Y(2014)	-0.56256	0.034976	-16.084	318	5.2165e-43	-0.63137	-0.49374
Y(2015)	-0.6157	0.038325	-16.065	318	6.1832e-43	-0.6911	-0.54029
Y(2016)	0.018798	0.032304	0.5819	318	0.56105	-0.044759	0.082355
M(Feb)	0.13989	0.033566	4.1677	318	3.9711e-05	0.073851	0.20593
M(Mar)	0.1065	0.036373	2.928	318	0.0036579	0.034937	0.17806
M(Apr)	0.23245	0.036001	6.4568	318	3.9981e-10	0.16162	0.30328
M(May)	-0.17939	0.035568	-5.0436	318	7.6917e-07	-0.24937	-0.10941
M(Jun)	0.13576	0.036441	3.7255	318	0.00023063	0.064064	0.20746
M(Jul)	0.12108	0.036377	3.3285	318	0.00097549	0.049512	0.19265
M(Aug)	-0.095455	0.0347	-2.7509	318	0.0062838	-0.16372	-0.027185
M(Sep)	-0.083545	0.035273	-2.3685	318	0.018458	-0.15294	-0.014146
M(Oct)	0.2267	0.033603	6.7466	318	7.1619e-11	0.16059	0.29281
M(Nov)	0.071577	0.034108	2.0985	318	0.036647	0.0044706	0.13868
M(Dec)	-0.14686	0.034319	-4.2791	318	2.4857e-05	-0.21438	-0.079335

3. Companies, likes

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.28819	0.055549	-5.1881	318	3.7924e-07	-0.39748	-0.1789
FB_likes	7.606e-07	2.7732e-07	2.7427	318	0.0064393	2.1499e-07	1.3062e-06
scD(t-1)	1.9119	0.057135	33.463	318	3.289e-106	1.7995	2.0243
NEWS(T)	-0.11035	0.016403	-6.7272	318	8.0522e-11	-0.14262	-0.078074
Ret(t)	-5.5392	0.10702	-51.761	318	5.8002e-157	-5.7498	-5.3287
Ret(t-1)	0.30788	0.12997	2.3688	318	0.018442	0.052166	0.56359
Y(2011)	-0.28845	0.030686	-9.4001	318	1.1111e-18	-0.34882	-0.22807
Y(2012)	-0.30968	0.035972	-8.6088	318	3.4696e-16	-0.38045	-0.2389
Y(2013)	-0.73109	0.04381	-16.688	318	2.4001e-45	-0.81728	-0.64489
Y(2014)	-0.61224	0.039751	-15.402	318	2.2342e-40	-0.69045	-0.53404
Y(2015)	-0.63232	0.038399	-16.467	318	1.7183e-44	-0.70787	-0.55677
Y(2016)	0.0041003	0.032173	0.12745	318	0.89867	-0.059198	0.067399
M(Feb)	0.13855	0.033481	4.1383	318	4.4855e-05	0.072681	0.20442
M(Mar)	0.10029	0.036225	2.7687	318	0.0059587	0.029024	0.17157
M(Apr)	0.23194	0.035749	6.488	318	3.33e-10	0.1616	0.30227
M(May)	-0.18324	0.035476	-5.1651	318	4.248e-07	-0.25303	-0.11344
M(Jun)	0.13145	0.03636	3.6153	318	0.00034863	0.059917	0.20299
M(Jul)	0.10745	0.036616	2.9344	318	0.003585	0.035406	0.17949
M(Aug)	-0.10812	0.035	-3.089	318	0.0021853	-0.17698	-0.039256
M(Sep)	-0.093253	0.035574	-2.6214	318	0.0091787	-0.16324	-0.023263
M(Oct)	0.21122	0.034196	6.1766	318	1.9995e-09	0.14394	0.2785
M(Nov)	0.063598	0.034218	1.8586	318	0.064005	-0.0037239	0.13092
M(Dec)	-0.15842	0.034489	-4.5934	318	6.2939e-06	-0.22628	-0.090566

4. Companies, shares

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.29924	0.055651	-5.3771	318	1.4686e-07	-0.40873	-0.18975
FB_shares	-6.4e-07	2.2416e-06	-0.28551	318	0.77544	-5.0502e-06	3.7702e-06
scD(t-1)	1.9216	0.057164	33.615	318	1.0717e-106	1.8091	2.034
NEWS(T)	-0.11814	0.016425	-7.1925	318	4.584e-12	-0.15045	-0.085823
Ret(t)	-5.5749	0.1073	-51.957	318	1.9808e-157	-5.786	-5.3638
Ret(t-1)	0.31521	0.12987	2.4271	318	0.015774	0.059697	0.57072
Y(2011)	-0.27668	0.030626	-9.0342	318	1.6401e-17	-0.33693	-0.21642
Y(2012)	-0.25257	0.035525	-7.1097	318	7.7103e-12	-0.32247	-0.18268
Y(2013)	-0.65575	0.039594	-16.562	318	7.3876e-45	-0.73365	-0.57785
Y(2014)	-0.55774	0.036032	-15.479	318	1.1299e-40	-0.62863	-0.48684
Y(2015)	-0.61977	0.038476	-16.108	318	4.2226e-43	-0.69546	-0.54407
Y(2016)	0.012991	0.03221	0.40332	318	0.68698	-0.05038	0.076362
M(Feb)	0.13545	0.033473	4.0466	318	6.5306e-05	0.069596	0.20131
M(Mar)	0.10118	0.036242	2.7918	318	0.005558	0.029878	0.17249
M(Apr)	0.2233	0.035751	6.2459	318	1.3495e-09	0.15296	0.29364
M(May)	-0.1834	0.0355	-5.1661	318	4.2267e-07	-0.25325	-0.11355
M(Jun)	0.1317	0.036376	3.6206	318	0.00034191	0.060134	0.20327
M(Jul)	0.11993	0.036404	3.2943	318	0.0010979	0.048303	0.19155
M(Aug)	-0.094141	0.035099	-2.6821	318	0.0076977	-0.1632	-0.025085
M(Sep)	-0.079099	0.035439	-2.232	318	0.026313	-0.14882	-0.0093743
M(Oct)	0.23129	0.034139	6.775	318	6.0335e-11	0.16413	0.29846
M(Nov)	0.071537	0.034159	2.0942	318	0.037031	0.0043306	0.13874
M(Dec)	-0.14772	0.034442	-4.2889	318	2.3848e-05	-0.21548	-0.079954

Financial and insurance institutions

5. Financial, posts

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.27359	0.14378	-1.9028	318	0.05797	-0.55647	0.0092961
FB_posts	0.0059153	0.0044514	1.3289	318	0.18485	-0.0028426	0.014673
scD(t-1)	1.4857	0.14501	10.246	318	1.7822e-21	1.2004	1.771
NEWS(T)	-0.10078	0.046347	-2.1744	318	0.03041	-0.19196	-0.0095925
Ret(t)	-1.741	0.32126	-5.4192	318	1.1845e-07	-2.373	-1.1089
Ret(t-1)	-1.014	0.34216	-2.9636	318	0.0032702	-1.6872	-0.34084
Y(2011)	-0.26161	0.090425	-2.8931	318	0.004078	-0.43952	-0.083701
Y(2012)	-0.37656	0.094684	-3.977	318	8.6461e-05	-0.56284	-0.19027
Y(2013)	-0.63504	0.10044	-6.3226	318	8.7001e-10	-0.83265	-0.43743
Y(2014)	-0.45838	0.098697	-4.6443	318	5.0016e-06	-0.65256	-0.2642
Y(2015)	-0.37838	0.096714	-3.9123	318	0.00011182	-0.56866	-0.1881
Y(2016)	-0.086654	0.096335	-0.89951	318	0.36906	-0.27619	0.10288
M(Feb)	0.040755	0.1021	0.39918	318	0.69003	-0.16012	0.24163
M(Mar)	-0.021464	0.10422	-0.20594	318	0.83697	-0.22651	0.18358
M(Apr)	0.27561	0.10545	2.6137	318	0.0093828	0.068146	0.48308
M(May)	-0.084239	0.10172	-0.82814	318	0.40821	-0.28437	0.11589
M(Jun)	-0.11166	0.10319	-1.0821	318	0.28001	-0.31467	0.091351
M(Jul)	0.085172	0.10749	0.79238	318	0.42873	-0.12631	0.29665
M(Aug)	-0.14365	0.10012	-1.4347	318	0.15235	-0.34063	0.05334
M(Sep)	-0.22464	0.098681	-2.2764	318	0.023486	-0.41879	-0.030488
M(Oct)	0.20171	0.098066	2.0568	318	0.040516	0.0087666	0.39465
M(Nov)	-0.0461	0.099805	-0.4619	318	0.64447	-0.24246	0.15026
M(Dec)	0.0050582	0.10184	0.049668	318	0.96042	-0.19531	0.20542

6. Financial, comments

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.25559	0.14224	-1.7969	318	0.073298	-0.53543	0.024257
FB_comments	1.2522e-05	9.6864e-06	1.2927	318	0.19705	-6.5359e-06	3.1579e-05
scD(t-1)	1.4851	0.14514	10.232	318	1.9797e-21	1.1995	1.7706
NEWS(T)	-0.10288	0.046279	-2.223	318	0.026917	-0.19393	-0.011828
Ret(t)	-1.7315	0.3215	-5.3857	318	1.4056e-07	-2.364	-1.099
Ret(t-1)	-0.96583	0.34042	-2.8371	318	0.0048441	-1.6356	-0.29606
Y(2011)	-0.2659	0.090877	-2.9259	318	0.0036817	-0.4447	-0.087102
Y(2012)	-0.3836	0.096472	-3.9763	318	8.6685e-05	-0.57341	-0.1938
Y(2013)	-0.63089	0.099859	-6.3178	318	8.9449e-10	-0.82735	-0.43442
Y(2014)	-0.42411	0.094102	-4.5069	318	9.2562e-06	-0.60925	-0.23897
Y(2015)	-0.37767	0.096808	-3.9012	318	0.00011683	-0.56814	-0.1872
Y(2016)	-0.054148	0.096408	-0.56165	318	0.57475	-0.24382	0.13553
M(Feb)	0.023447	0.10081	0.23258	318	0.81624	-0.1749	0.22179
M(Mar)	-0.026627	0.10389	-0.2563	318	0.79789	-0.23103	0.17778
M(Apr)	0.26566	0.10421	2.5492	318	0.011267	0.060625	0.4707
M(May)	-0.089213	0.10142	-0.87963	318	0.37972	-0.28875	0.11033
M(Jun)	-0.11135	0.10323	-1.0787	318	0.28152	-0.31444	0.091738
M(Jul)	0.08613	0.10748	0.80137	318	0.42352	-0.12533	0.29759
M(Aug)	-0.1475	0.10002	-1.4746	318	0.1413	-0.34429	0.049295
M(Sep)	-0.23391	0.098936	-2.3643	318	0.018666	-0.42856	-0.03926
M(Oct)	0.1947	0.098145	1.9838	318	0.048138	0.0016043	0.3878
M(Nov)	-0.038349	0.099645	-0.38485	318	0.7006	-0.2344	0.1577
M(Dec)	0.0061782	0.10184	0.060664	318	0.95166	-0.19419	0.20655

7. Financial, likes

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.2329	0.14207	-1.6394	318	0.10212	-0.51241	0.046609
FB_likes	9.4247e-07	7.7724e-07	1.2126	318	0.22619	-5.8672e-07	2.4717e-06
scD(t-1)	1.4932	0.14464	10.323	318	9.737e-22	1.2086	1.7777
NEWS(T)	-0.095804	0.04683	-2.0458	318	0.041602	-0.18794	-0.0036675
Ret(t)	-1.7111	0.32258	-5.3044	318	2.1219e-07	-2.3457	-1.0764
Ret(t-1)	-0.96512	0.34064	-2.8333	318	0.0049018	-1.6353	-0.29493
Y(2011)	-0.25983	0.090342	-2.876	318	0.0042988	-0.43757	-0.082084
Y(2012)	-0.40819	0.10564	-3.8639	318	0.00013526	-0.61604	-0.20034
Y(2013)	-0.68726	0.12076	-5.6912	318	2.8659e-08	-0.92485	-0.44967
Y(2014)	-0.48133	0.1074	-4.4819	318	1.0339e-05	-0.69262	-0.27003
Y(2015)	-0.39996	0.096603	-4.1403	318	4.4486e-05	-0.59002	-0.2099
Y(2016)	-0.07959	0.095878	-0.83012	318	0.40709	-0.26823	0.10905
M(Feb)	0.021093	0.10076	0.20934	318	0.83432	-0.17715	0.21934
M(Mar)	-0.040063	0.10364	-0.38654	318	0.69935	-0.24398	0.16385
M(Apr)	0.2576	0.10352	2.4883	318	0.013347	0.053919	0.46127
M(May)	-0.098709	0.10124	-0.97502	318	0.33029	-0.29789	0.10047
M(Jun)	-0.12185	0.10286	-1.1846	318	0.23708	-0.32422	0.080529
M(Jul)	0.068195	0.10801	0.63137	318	0.52825	-0.14431	0.2807
M(Aug)	-0.1627	0.1007	-1.6158	318	0.10714	-0.36082	0.035414
M(Sep)	-0.23992	0.099478	-2.4118	318	0.016441	-0.43564	-0.044202
M(Oct)	0.17868	0.099592	1.7941	318	0.073741	-0.017261	0.37462
M(Nov)	-0.048569	0.099917	-0.48609	318	0.62724	-0.24515	0.14801
M(Dec)	-0.0063766	0.10175	-0.062669	318	0.95007	-0.20657	0.19381

8. Financial, shares

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.24623	0.14238	-1.7294	318	0.084709	-0.52636	0.033895
FB_shares	-1.9448e-06	6.3045e-06	-0.30847	318	0.75792	-1.4348e-05	1.0459e-05
scD(t-1)	1.5077	0.14473	10.418	318	4.6489e-22	1.223	1.7925
NEWS(T)	-0.10724	0.046981	-2.2826	318	0.023115	-0.19967	-0.014805
Ret(t)	-1.7555	0.32294	-5.4359	318	1.0876e-07	-2.3908	-1.1201
Ret(t-1)	-0.96966	0.34029	-2.8495	318	0.0046643	-1.6392	-0.30016
Y(2011)	-0.24577	0.09027	-2.7226	318	0.0068347	-0.42337	-0.068166
Y(2012)	-0.32812	0.10514	-3.1209	318	0.0019686	-0.53496	-0.12127
Y(2013)	-0.58476	0.10875	-5.3769	318	1.4702e-07	-0.79873	-0.37079
Y(2014)	-0.41073	0.097118	-4.2292	318	3.0698e-05	-0.60181	-0.21966
Y(2015)	-0.38722	0.096884	-3.9967	318	7.9894e-05	-0.57783	-0.1966
Y(2016)	-0.067442	0.096061	-0.70207	318	0.48315	-0.25644	0.12155
M(Feb)	0.018691	0.10077	0.18549	318	0.85297	-0.17956	0.21694
M(Mar)	-0.036118	0.10365	-0.34846	318	0.72773	-0.24005	0.16781
M(Apr)	0.24376	0.10344	2.3564	318	0.019059	0.040235	0.44728
M(May)	-0.096697	0.10124	-0.95513	318	0.34024	-0.29588	0.10249
M(Jun)	-0.12341	0.10293	-1.199	318	0.23142	-0.32593	0.079099
M(Jul)	0.083551	0.10756	0.77675	318	0.43788	-0.12808	0.29518
M(Aug)	-0.14459	0.1009	-1.433	318	0.15283	-0.3431	0.05392
M(Sep)	-0.22135	0.099184	-2.2317	318	0.026331	-0.41649	-0.02621
M(Oct)	0.20587	0.099729	2.0643	318	0.039805	0.0096537	0.40208
M(Nov)	-0.037633	0.099786	-0.37714	318	0.70632	-0.23396	0.15869
M(Dec)	0.00043414	0.10175	0.0042666	318	0.9966	-0.19976	0.20063

General governmental organizations

9. Governmental, posts

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.052584	0.24366	-0.21581	309	0.82928	-0.53202	0.42685
FB_posts	0.014641	0.0086466	1.6933	309	0.091412	-0.0023726	0.031655
scD(t-1)	0.70828	0.18245	3.882	309	0.00012669	0.34927	1.0673
NEWS(T)	0.20171	0.095884	2.1037	309	0.036215	0.013041	0.39038
Ret(t)	-2.0893	0.67428	-3.0986	309	0.0021229	-3.4161	-0.76259
Ret(t-1)	-1.9302	0.69156	-2.7911	309	0.0055802	-3.2909	-0.56943
Y(2011)	-0.50268	0.16858	-2.9819	309	0.0030928	-0.83439	-0.17097
Y(2012)	-0.48266	0.18034	-2.6765	309	0.0078376	-0.8375	-0.12782
Y(2013)	-0.97142	0.19832	-4.8982	309	1.563e-06	-1.3617	-0.58118
Y(2014)	-0.8833	0.19908	-4.4369	309	1.2709e-05	-1.275	-0.49158
Y(2015)	-1.0279	0.20705	-4.9644	309	1.1414e-06	-1.4353	-0.62048
Y(2016)	-0.45068	0.19325	-2.3322	309	0.020335	-0.83093	-0.070435
M(Feb)	0.062391	0.20193	0.30897	309	0.75755	-0.33494	0.45973
M(Mar)	0.26514	0.21088	1.2573	309	0.20958	-0.14979	0.68008
M(Apr)	0.46221	0.20898	2.2117	309	0.027718	0.051002	0.87343
M(May)	0.28204	0.21426	1.3164	309	0.18903	-0.13955	0.70363
M(Jun)	0.1684	0.21964	0.7667	309	0.44384	-0.26378	0.60058
M(Jul)	0.24541	0.22395	1.0958	309	0.27401	-0.19525	0.68607
M(Aug)	-0.19373	0.20465	-0.94663	309	0.34457	-0.59641	0.20896
M(Sep)	0.085418	0.20959	0.40756	309	0.68388	-0.32698	0.49781
M(Oct)	0.5011	0.20099	2.4932	309	0.013184	0.10562	0.89659
M(Nov)	0.28888	0.20528	1.4072	309	0.16036	-0.11504	0.6928
M(Dec)	0.074541	0.21514	0.34648	309	0.72922	-0.34879	0.49787

10. Governmental, comments

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.039507	0.23797	-0.16602	309	0.86825	-0.50774	0.42873
FB_comments	5.4546e-05	2.0163e-05	2.7053	309	0.0072031	1.4872e-05	9.4219e-05
scD(t-1)	0.67547	0.18322	3.6867	309	0.00026823	0.31496	1.036
NEWS(T)	0.19591	0.095952	2.0417	309	0.042029	0.0071042	0.38471
Ret(t)	-2.0751	0.67443	-3.0768	309	0.0022793	-3.4022	-0.74806
Ret(t-1)	-1.7414	0.68892	-2.5278	309	0.011977	-3.097	-0.38586
Y(2011)	-0.55124	0.1702	-3.2387	309	0.0013316	-0.88614	-0.21633
Y(2012)	-0.5752	0.1853	-3.1041	309	0.0020855	-0.93982	-0.21058
Y(2013)	-1.0181	0.19868	-5.124	309	5.2799e-07	-1.409	-0.62711
Y(2014)	-0.81809	0.19193	-4.2624	309	2.6895e-05	-1.1958	-0.44043
Y(2015)	-1.0116	0.20691	-4.8892	309	1.6304e-06	-1.4188	-0.60449
Y(2016)	-0.35213	0.19359	-1.8189	309	0.069895	-0.73305	0.028801
M(Feb)	0.042501	0.20016	0.21234	309	0.83198	-0.35135	0.43635
M(Mar)	0.29009	0.21112	1.374	309	0.17043	-0.12533	0.7055
M(Apr)	0.48291	0.20768	2.3252	309	0.020707	0.074261	0.89157
M(May)	0.30657	0.2145	1.4292	309	0.15395	-0.11549	0.72863
M(Jun)	0.18274	0.21982	0.83128	309	0.40646	-0.2498	0.61527
M(Jul)	0.26122	0.22415	1.1654	309	0.24477	-0.17984	0.70228
M(Aug)	-0.18473	0.20467	-0.90255	309	0.36747	-0.58746	0.218
M(Sep)	0.039457	0.21048	0.18746	309	0.85142	-0.3747	0.45361
M(Oct)	0.48516	0.20125	2.4108	309	0.016503	0.089169	0.88115
M(Nov)	0.3232	0.20481	1.578	309	0.11559	-0.079809	0.72621
M(Dec)	0.1013	0.21568	0.46968	309	0.63891	-0.32309	0.5257

11. Governmental, likes

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.079835	0.2354	0.33914	309	0.73473	-0.38336	0.54303
FB_likes	2.6212e-06	1.643e-06	1.5953	309	0.11166	-6.1174e-07	5.8542e-06
scD(t-1)	0.6898	0.18289	3.7717	309	0.00019427	0.32994	1.0497
NEWS(T)	0.21615	0.096781	2.2334	309	0.026235	0.025722	0.40659
Ret(t)	-2.0272	0.67495	-3.0035	309	0.002887	-3.3553	-0.69914
Ret(t-1)	-1.8004	0.68907	-2.6128	309	0.0094199	-3.1562	-0.44453
Y(2011)	-0.50681	0.16892	-3.0004	309	0.002916	-0.83918	-0.17444
Y(2012)	-0.57798	0.20504	-2.8189	309	0.0051301	-0.98142	-0.17453
Y(2013)	-1.1363	0.24587	-4.6216	309	5.5995e-06	-1.6201	-0.65254
Y(2014)	-0.97138	0.22205	-4.3745	309	1.6659e-05	-1.4083	-0.53445
Y(2015)	-1.0995	0.20723	-5.3056	309	2.1465e-07	-1.5072	-0.69172
Y(2016)	-0.44867	0.19323	-2.322	309	0.020885	-0.82888	-0.068459
M(Feb)	0.010471	0.19905	0.052606	309	0.95808	-0.38119	0.40213
M(Mar)	0.23358	0.20918	1.1166	309	0.26502	-0.17802	0.64518
M(Apr)	0.41772	0.20497	2.0379	309	0.042408	0.014402	0.82103
M(May)	0.25218	0.21259	1.1862	309	0.23645	-0.16613	0.67048
M(Jun)	0.12845	0.21834	0.5883	309	0.55676	-0.30117	0.55807
M(Jul)	0.18769	0.22485	0.83476	309	0.4045	-0.25473	0.63012
M(Aug)	-0.26145	0.20544	-1.2726	309	0.20411	-0.66569	0.14279
M(Sep)	0.026845	0.21177	0.12676	309	0.89921	-0.38985	0.44354
M(Oct)	0.43846	0.2035	2.1546	309	0.031963	0.038047	0.83887
M(Nov)	0.28485	0.20495	1.3898	309	0.16558	-0.11843	0.68813
M(Dec)	0.043126	0.21542	0.20019	309	0.84146	-0.38076	0.46701

12. Governmental, shares

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.097605	0.23615	0.41332	309	0.67966	-0.36706	0.56227
FB_shares	2.1855e-05	1.2734e-05	1.7163	309	0.087118	-3.2015e-06	4.6911e-05
scD(t-1)	0.68168	0.18309	3.7232	309	0.00023377	0.32141	1.0419
NEWS(T)	0.21933	0.096932	2.2627	309	0.024349	0.028597	0.41006
Ret(t)	-1.982	0.67637	-2.9303	309	0.0036389	-3.3128	-0.65108
Ret(t-1)	-1.8061	0.68895	-2.6215	309	0.0091879	-3.1617	-0.45045
Y(2011)	-0.50449	0.16857	-2.9927	309	0.0029879	-0.83619	-0.1728
Y(2012)	-0.57892	0.20125	-2.8766	309	0.0042994	-0.97491	-0.18292
Y(2013)	-1.0697	0.21872	-4.8907	309	1.6188e-06	-1.5001	-0.63933
Y(2014)	-0.88097	0.19843	-4.4396	309	1.2558e-05	-1.2714	-0.49052
Y(2015)	-1.1096	0.20778	-5.3405	309	1.8001e-07	-1.5185	-0.7008
Y(2016)	-0.45485	0.19347	-2.351	309	0.019349	-0.83554	-0.07417
M(Feb)	0.0039606	0.199	0.019903	309	0.98413	-0.3876	0.39552
M(Mar)	0.23816	0.20917	1.1386	309	0.25575	-0.17342	0.64974
M(Apr)	0.41816	0.2048	2.0418	309	0.042022	0.015178	0.82113
M(May)	0.25279	0.21266	1.1887	309	0.23547	-0.16565	0.67123
M(Jun)	0.13848	0.21841	0.63404	309	0.52652	-0.29128	0.56824
M(Jul)	0.21092	0.22373	0.94274	309	0.34655	-0.22931	0.65115
M(Aug)	-0.2736	0.20633	-1.326	309	0.18581	-0.67958	0.13239
M(Sep)	0.029716	0.21121	0.1407	309	0.8882	-0.38587	0.4453
M(Oct)	0.43169	0.20364	2.1199	309	0.034812	0.030997	0.83238
M(Nov)	0.28837	0.20468	1.4089	309	0.15988	-0.11438	0.69112
M(Dec)	0.042815	0.21519	0.19896	309	0.84242	-0.3806	0.46623

Nonprofit institutions

13. Nonprofit, posts

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.41785	0.19479	2.1451	318	0.032701	0.034608	0.8011
FB_posts	0.032679	0.0061768	5.2907	318	2.2734e-07	0.020527	0.044832
scD(t-1)	1.3724	0.13864	9.899	318	2.5805e-20	1.0997	1.6452
NEWS(T)	0.041756	0.062358	0.66961	318	0.50359	-0.080931	0.16444
Ret(t)	-0.63331	0.40943	-1.5468	318	0.1229	-1.4388	0.17222
Ret(t-1)	-1.6034	0.42745	-3.751	318	0.00020925	-2.4444	-0.76239
Y(2011)	-0.83456	0.12535	-6.6577	318	1.2212e-10	-1.0812	-0.58793
Y(2012)	-0.94372	0.13529	-6.9758	318	1.7704e-11	-1.2099	-0.67756
Y(2013)	-1.7647	0.14827	-11.902	318	2.8925e-27	-2.0565	-1.473
Y(2014)	-1.4929	0.14882	-10.032	318	9.3177e-21	-1.7857	-1.2001
Y(2015)	-1.3249	0.14122	-9.3822	318	1.2692e-18	-1.6028	-1.0471
Y(2016)	-0.45549	0.13052	-3.4899	318	0.00055141	-0.71227	-0.1987
M(Feb)	0.0066968	0.12964	0.051658	318	0.95883	-0.24836	0.26175
M(Mar)	-0.2825	0.13962	-2.0233	318	0.043875	-0.5572	-0.0078023
M(Apr)	0.31919	0.13637	2.3406	318	0.019873	0.050882	0.58749
M(May)	-0.28778	0.12792	-2.2496	318	0.025157	-0.53946	-0.036095
M(Jun)	-0.34654	0.14148	-2.4494	318	0.014848	-0.62489	-0.068189
M(Jul)	-0.45161	0.15709	-2.8749	318	0.0043139	-0.76067	-0.14255
M(Aug)	0.0015287	0.14423	0.010599	318	0.99155	-0.28224	0.2853
M(Sep)	-0.34669	0.13489	-2.5702	318	0.010619	-0.61207	-0.081301
M(Oct)	0.18621	0.13462	1.3832	318	0.16756	-0.078645	0.45106
M(Nov)	-0.11959	0.12943	-0.924	318	0.35618	-0.37424	0.13505
M(Dec)	-0.15748	0.13742	-1.146	318	0.25267	-0.42784	0.11289

14. Nonprofit, comments

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.50168	0.19403	2.5856	318	0.010167	0.11993	0.88342
FB_comments	3.7564e-05	1.3222e-05	2.841	318	0.0047877	1.155e-05	6.3578e-05
scD(t-1)	1.4567	0.13752	10.593	318	1.1745e-22	1.1862	1.7273
NEWS(T)	0.018372	0.062107	0.29582	318	0.76756	-0.10382	0.14056
Ret(t)	-0.56132	0.40944	-1.3709	318	0.17137	-1.3669	0.24424
Ret(t-1)	-1.3599	0.42627	-3.1901	318	0.0015639	-2.1985	-0.52119
Y(2011)	-0.81209	0.12569	-6.461	318	3.8995e-10	-1.0594	-0.5648
Y(2012)	-0.87963	0.13787	-6.3801	318	6.2433e-10	-1.1509	-0.60837
Y(2013)	-1.6334	0.14549	-11.227	318	7.3382e-25	-1.9196	-1.3471
Y(2014)	-1.2501	0.14	-8.9296	318	3.5002e-17	-1.5256	-0.97469
Y(2015)	-1.3221	0.14157	-9.3393	318	1.7449e-18	-1.6006	-1.0436
Y(2016)	-0.31492	0.13006	-2.4213	318	0.016022	-0.57081	-0.059034
M(Feb)	-0.08647	0.12838	-0.67356	318	0.50108	-0.33905	0.16611
M(Mar)	-0.31869	0.1397	-2.2812	318	0.023195	-0.59354	-0.043836
M(Apr)	0.24295	0.13544	1.7938	318	0.073787	-0.023512	0.50941
M(May)	-0.32371	0.12677	-2.5536	318	0.011129	-0.57312	-0.074301
M(Jun)	-0.35988	0.14142	-2.5448	318	0.011406	-0.63812	-0.081648
M(Jul)	-0.43714	0.15673	-2.7892	318	0.0056022	-0.7455	-0.12879
M(Aug)	-0.0097459	0.14385	-0.067749	318	0.94603	-0.29277	0.27328
M(Sep)	-0.37549	0.13432	-2.7955	318	0.0054963	-0.63976	-0.11123
M(Oct)	0.17766	0.13405	1.3253	318	0.18603	-0.086085	0.44141
M(Nov)	-0.074263	0.12806	-0.57989	318	0.5624	-0.32622	0.1777
M(Dec)	-0.13406	0.13649	-0.98217	318	0.32676	-0.40259	0.13448

15. Nonprofit, likes

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.61566	0.1935	3.1817	318	0.0016086	0.23496	0.99637
FB_likes	4.4277e-06	1.1856e-06	3.7345	318	0.00022288	2.095e-06	6.7603e-06
scD(t-1)	1.4074	0.13848	10.163	318	3.3903e-21	1.1349	1.6798
NEWS(T)	0.053152	0.062844	0.84578	318	0.39831	-0.070491	0.1768
Ret(t)	-0.44153	0.41229	-1.0709	318	0.28501	-1.2527	0.36962
Ret(t-1)	-1.35	0.42895	-3.1471	318	0.0018048	-2.1939	-0.50603
Y(2011)	-0.81176	0.12476	-6.5063	318	2.9909e-10	-1.0572	-0.56629
Y(2012)	-1.0462	0.15171	-6.8961	318	2.8892e-11	-1.3447	-0.74774
Y(2013)	-1.949	0.17783	-10.96	318	6.3236e-24	-2.2989	-1.5991
Y(2014)	-1.5478	0.16373	-9.4532	318	7.4784e-19	-1.87	-1.2257
Y(2015)	-1.4197	0.14184	-10.009	318	1.1066e-20	-1.6988	-1.1407
Y(2016)	-0.40512	0.12968	-3.124	318	0.0019482	-0.66025	-0.14998
M(Feb)	-0.098622	0.12755	-0.77321	318	0.43997	-0.34957	0.15232
M(Mar)	-0.33273	0.13869	-2.3991	318	0.01701	-0.60559	-0.059867
M(Apr)	0.23834	0.13466	1.7699	318	0.077697	-0.026599	0.50328
M(May)	-0.34163	0.12644	-2.7019	318	0.0072652	-0.59041	-0.092861
M(Jun)	-0.38387	0.14074	-2.7275	318	0.006736	-0.66077	-0.10697
M(Jul)	-0.51121	0.15695	-3.2572	318	0.0012467	-0.82	-0.20242
M(Aug)	-0.078559	0.14468	-0.543	318	0.58751	-0.36321	0.20609
M(Sep)	-0.40045	0.13514	-2.9632	318	0.0032745	-0.66634	-0.13457
M(Oct)	0.10546	0.13552	0.77817	318	0.43705	-0.16118	0.3721
M(Nov)	-0.10993	0.12838	-0.85626	318	0.3925	-0.3625	0.14265
M(Dec)	-0.18347	0.13684	-1.3408	318	0.18095	-0.4527	0.085753

16. Nonprofit, shares

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.63207	0.19427	3.2535	318	0.0012624	0.24985	1.0143
FB_shares	3.0261e-05	9.6311e-06	3.142	318	0.001836	1.1312e-05	4.921e-05
scD(t-1)	1.4068	0.13881	10.135	318	4.2169e-21	1.1337	1.6799
NEWS(T)	0.049813	0.062948	0.79133	318	0.42934	-0.074035	0.17366
Ret(t)	-0.40964	0.41327	-0.9912	318	0.32234	-1.2227	0.40346
Ret(t-1)	-1.371	0.42754	-3.2068	318	0.0014788	-2.2122	-0.52986
Y(2011)	-0.79589	0.12451	-6.3924	318	5.8153e-10	-1.0408	-0.55093
Y(2012)	-0.98982	0.1498	-6.6076	318	1.6453e-10	-1.2846	-0.6951
Y(2013)	-1.7965	0.16222	-11.075	318	2.51e-24	-2.1157	-1.4774
Y(2014)	-1.3584	0.14566	-9.3263	318	1.921e-18	-1.645	-1.0719
Y(2015)	-1.4226	0.1423	-9.9974	318	1.2146e-20	-1.7026	-1.1427
Y(2016)	-0.40508	0.1299	-3.1184	318	0.0019849	-0.66066	-0.14951
M(Feb)	-0.11333	0.12745	-0.88919	318	0.37457	-0.36409	0.13743
M(Mar)	-0.33647	0.13888	-2.4228	318	0.01596	-0.60971	-0.063235
M(Apr)	0.23024	0.13454	1.7113	318	0.087994	-0.034457	0.49494
M(May)	-0.33348	0.12646	-2.637	318	0.0087739	-0.58228	-0.084676
M(Jun)	-0.37494	0.14092	-2.6606	318	0.0081952	-0.6522	-0.097684
M(Jul)	-0.47441	0.15667	-3.028	318	0.0026628	-0.78266	-0.16617
M(Aug)	-0.076577	0.14507	-0.52786	318	0.59796	-0.36199	0.20884
M(Sep)	-0.39276	0.13482	-2.9131	318	0.0038314	-0.65802	-0.1275
M(Oct)	0.11811	0.13554	0.87137	318	0.38421	-0.14856	0.38478
M(Nov)	-0.09836	0.12812	-0.7677	318	0.44324	-0.35044	0.15372
M(Dec)	-0.17922	0.13638	-1.3142	318	0.18974	-0.44754	0.089095

Households

17. Households, posts

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.036811	0.014795	2.4881	318	0.013354	0.007703	0.065918
FB_posts	0.010896	0.00040914	26.632	318	5.7387e-83	0.010091	0.011701
scD(t-1)	1.5001	0.014255	105.24	318	3.4599e-249	1.4721	1.5281
NEWS(T)	-0.039957	0.0041561	-9.6142	318	2.2392e-19	-0.048134	-0.03178
Ret(t)	-5.2764	0.026473	-199.31	318	0	-5.3285	-5.2243
Ret(t-1)	-0.48546	0.031587	-15.369	318	2.9861e-40	-0.54761	-0.42332
Y(2011)	-0.21663	0.0081581	-26.554	318	1.0907e-82	-0.23268	-0.20058
Y(2012)	-0.20003	0.0083398	-23.984	318	2.6671e-73	-0.21643	-0.18362
Y(2013)	-0.75546	0.0094403	-80.025	318	9.414e-213	-0.77404	-0.73689
Y(2014)	-0.7962	0.0096552	-82.463	318	1.0453e-216	-0.8152	-0.77721
Y(2015)	-0.63079	0.0098027	-64.349	318	2.1219e-184	-0.65008	-0.6115
Y(2016)	0.07737	0.0084833	9.1202	318	8.7552e-18	0.060679	0.09406
M(Feb)	0.2758	0.0090565	30.454	318	2.7852e-96	0.25799	0.29362
M(Mar)	0.21536	0.0096964	22.21	318	1.214e-66	0.19628	0.23444
M(Apr)	0.22415	0.009544	23.486	318	1.9248e-71	0.20537	0.24292
M(May)	0.070463	0.0095385	7.3872	318	1.329e-12	0.051697	0.08923
M(Jun)	0.2067	0.0095061	21.744	318	7.1728e-65	0.188	0.2254
M(Jul)	0.19668	0.0092698	21.217	318	7.3224e-63	0.17844	0.21491
M(Aug)	0.014564	0.0089101	1.6346	318	0.10312	-0.0029658	0.032095
M(Sep)	-0.0020314	0.0090309	-0.22494	318	0.82217	-0.019799	0.015736
M(Oct)	0.17061	0.0087869	19.417	318	6.198e-56	0.15332	0.1879
M(Nov)	0.066463	0.0089276	7.4446	318	9.1892e-13	0.048898	0.084027
M(Dec)	-0.18137	0.0087435	-20.743	318	4.7857e-61	-0.19857	-0.16416

18. Households, comments

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.11745	0.014779	7.9473	318	3.342e-14	0.088375	0.14653
FB_comments	-4.4579e-06	8.8363e-07	-5.045	318	7.6393e-07	-6.1964e-06	-2.7194e-06
scD(t-1)	1.5072	0.014279	105.55	318	1.3585e-249	1.4791	1.5353
NEWS(T)	-0.050044	0.0041388	-12.091	318	5.9927e-28	-0.058186	-0.041901
Ret(t)	-5.3052	0.026515	-200.08	318	0	-5.3573	-5.253
Ret(t-1)	-0.40271	0.031479	-12.793	318	1.6273e-30	-0.46465	-0.34078
Y(2011)	-0.18399	0.0081646	-22.535	318	7.1688e-68	-0.20005	-0.16793
Y(2012)	-0.13096	0.0084913	-15.423	318	1.8518e-40	-0.14767	-0.11426
Y(2013)	-0.682	0.0093033	-73.308	318	2.8931e-201	-0.7003	-0.6637
Y(2014)	-0.71917	0.0092151	-78.042	318	1.8711e-209	-0.7373	-0.70104
Y(2015)	-0.65546	0.0098601	-66.477	318	1.3941e-188	-0.67486	-0.63606
Y(2016)	0.10514	0.0084697	12.413	318	4.0421e-29	0.088474	0.1218
M(Feb)	0.22511	0.0089298	25.209	318	8.1937e-78	0.20754	0.24268
M(Mar)	0.181	0.0096687	18.72	318	3.1017e-53	0.16197	0.20002
M(Apr)	0.1657	0.0094557	17.524	318	1.3638e-48	0.1471	0.1843
M(May)	0.047583	0.0095291	4.9934	318	9.7965e-07	0.028835	0.066331
M(Jun)	0.18251	0.0094991	19.214	318	3.7718e-55	0.16383	0.2012
M(Jul)	0.19061	0.0092574	20.59	318	1.847e-60	0.1724	0.20882
M(Aug)	0.0051563	0.0088948	0.5797	318	0.56253	-0.012344	0.022656
M(Sep)	-0.0022091	0.0090052	-0.24531	318	0.80637	-0.019926	0.015508
M(Oct)	0.17591	0.0087766	20.043	318	2.3599e-58	0.15864	0.19318
M(Nov)	0.069467	0.0089007	7.8046	318	8.6843e-14	0.051955	0.086979
M(Dec)	-0.19131	0.0087185	-21.943	318	1.2489e-65	-0.20847	-0.17416

19. Households, likes

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.1051	0.014589	7.2044	318	4.2541e-12	0.076402	0.13381
FB_likes	1.9186e-07	7.0535e-08	2.7201	318	0.0068852	5.3088e-08	3.3064e-07
scD(t-1)	1.5121	0.014262	106.02	318	3.4962e-250	1.484	1.5401
NEWS(T)	-0.047872	0.0041922	-11.419	318	1.5368e-25	-0.05612	-0.039624
Ret(t)	-5.2895	0.026641	-198.55	318	0	-5.3419	-5.2371
Ret(t-1)	-0.39393	0.031462	-12.521	318	1.6351e-29	-0.45583	-0.33203
Y(2011)	-0.19163	0.0081284	-23.575	318	8.8996e-72	-0.20762	-0.17564
Y(2012)	-0.15729	0.0094328	-16.675	318	2.6956e-45	-0.17585	-0.13873
Y(2013)	-0.70775	0.010964	-64.55	318	8.4517e-185	-0.72933	-0.68618
Y(2014)	-0.73225	0.010262	-71.355	318	9.4631e-198	-0.75244	-0.71206
Y(2015)	-0.65063	0.0097949	-66.426	318	1.7502e-188	-0.6699	-0.63136
Y(2016)	0.10926	0.0084132	12.987	318	3.1326e-31	0.092707	0.12581
M(Feb)	0.22968	0.0088952	25.821	318	4.8215e-80	0.21218	0.24718
M(Mar)	0.18462	0.009634	19.164	318	5.8991e-55	0.16567	0.20358
M(Apr)	0.17505	0.0093885	18.645	318	6.0312e-53	0.15658	0.19352
M(May)	0.050653	0.009505	5.3292	318	1.8729e-07	0.031953	0.069354
M(Jun)	0.18617	0.009474	19.651	318	7.7064e-57	0.16753	0.20481
M(Jul)	0.18845	0.0093244	20.21	318	5.3366e-59	0.17011	0.2068
M(Aug)	0.0025782	0.0089501	0.28807	318	0.77348	-0.015031	0.020187
M(Sep)	-0.0055631	0.0090473	-0.61489	318	0.53907	-0.023363	0.012237
M(Oct)	0.17034	0.0089165	19.104	318	1.0087e-54	0.15279	0.18788
M(Nov)	0.06879	0.0089202	7.7118	318	1.6071e-13	0.05124	0.08634
M(Dec)	-0.19057	0.0087307	-21.827	318	3.4476e-65	-0.20774	-0.17339

20. Households, shares

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.10485	0.014605	7.1791	318	4.9886e-12	0.076114	0.13358
FB_shares	-9.0402e-06	5.7473e-07	-15.729	318	1.2251e-41	-1.0171e-05	-7.9094e-06
scD(t-1)	1.5051	0.014274	105.44	318	1.902e-249	1.477	1.5331
NEWS(T)	-0.060349	0.0041961	-14.382	318	1.787e-36	-0.068604	-0.052093
Ret(t)	-5.3519	0.026697	-200.47	318	0	-5.4044	-5.2994
Ret(t-1)	-0.41371	0.031478	-13.143	318	8.2429e-32	-0.47564	-0.35178
Y(2011)	-0.18012	0.0081142	-22.198	318	1.3478e-66	-0.19609	-0.16416
Y(2012)	-0.071317	0.0093383	-7.6371	318	2.6275e-13	-0.08969	-0.052944
Y(2013)	-0.62961	0.0099336	-63.381	318	1.8538e-182	-0.64915	-0.61006
Y(2014)	-0.69029	0.0094072	-73.379	318	2.1576e-201	-0.7088	-0.67179
Y(2015)	-0.63949	0.0098074	-65.205	318	4.2695e-186	-0.65879	-0.62019
Y(2016)	0.1204	0.0084168	14.305	318	3.4979e-36	0.10384	0.13696
M(Feb)	0.22823	0.0089021	25.638	318	2.2229e-79	0.21072	0.24575
M(Mar)	0.1856	0.0096429	19.247	318	2.7954e-55	0.16663	0.20457
M(Apr)	0.16046	0.0093935	17.082	318	7.1174e-47	0.14197	0.17894
M(May)	0.049881	0.0095182	5.2406	318	2.921e-07	0.031155	0.068608
M(Jun)	0.18145	0.0094837	19.133	318	7.7371e-55	0.1628	0.20011
M(Jul)	0.19695	0.009266	21.256	318	5.2127e-63	0.17872	0.21519
M(Aug)	0.022949	0.0089673	2.5592	318	0.010953	0.0053065	0.040592
M(Sep)	0.008407	0.0090222	0.93182	318	0.35214	-0.0093437	0.026158
M(Oct)	0.19793	0.0089067	22.222	318	1.0912e-66	0.18041	0.21545
M(Nov)	0.076644	0.0089086	8.6033	318	3.6065e-16	0.059117	0.094171
M(Dec)	-0.17974	0.0087284	-20.593	318	1.8101e-60	-0.19691	-0.16257

Household activity groups

Active investors

21. Active investors, posts

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.60538	0.092679	6.532	318	2.5715e-10	0.42304	0.78772
FB_posts	0.00024134	0.0033095	0.072923	318	0.94191	-0.00627	0.0067527
scD(t-1)	-0.82969	0.097284	-8.5285	318	6.1119e-16	-1.0211	-0.63829
NEWS(T)	-0.13305	0.033665	-3.9523	318	9.5428e-05	-0.19929	-0.066819
Ret(t)	-7.9946	0.27513	-29.058	318	1.6452e-91	-8.5359	-7.4533
Ret(t-1)	1.6007	0.30539	5.2413	318	2.911e-07	0.99981	2.2015
Y(2011)	-0.12886	0.062632	-2.0574	318	0.040459	-0.25209	-0.0056358
Y(2012)	-0.088595	0.059998	-1.4766	318	0.14077	-0.20664	0.029449
Y(2013)	-0.079748	0.06202	-1.2859	318	0.19943	-0.20177	0.042273
Y(2014)	-0.020625	0.066816	-0.30868	318	0.75777	-0.15208	0.11083
Y(2015)	-0.028919	0.068288	-0.42349	318	0.67223	-0.16327	0.10543
Y(2016)	0.10083	0.063944	1.5768	318	0.11584	-0.024982	0.22663
M(Feb)	0.046021	0.078498	0.58627	318	0.55811	-0.10842	0.20046
M(Mar)	0.087959	0.077226	1.139	318	0.25557	-0.06398	0.2399
M(Apr)	-0.098081	0.083243	-1.1783	318	0.23958	-0.26186	0.065695
M(May)	-0.11265	0.084266	-1.3368	318	0.18224	-0.27844	0.053142
M(Jun)	-0.17391	0.083183	-2.0906	318	0.037354	-0.33757	-0.010247
M(Jul)	-0.032485	0.080163	-0.40523	318	0.68558	-0.1902	0.12523
M(Aug)	-0.16885	0.080643	-2.0938	318	0.037068	-0.32751	-0.010191
M(Sep)	-0.11495	0.078531	-1.4637	318	0.14425	-0.26946	0.039557
M(Oct)	0.032025	0.075289	0.42536	318	0.67086	-0.1161	0.18015
M(Nov)	-0.03916	0.075414	-0.51927	318	0.60393	-0.18753	0.10921
M(Dec)	-0.050813	0.07676	-0.66197	318	0.50847	-0.20183	0.10021

22. Active investors, comments

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.61757	0.091376	6.7586	318	6.6641e-11	0.43779	0.79735
FB_comments	-5.8237e-06	6.7028e-06	-0.86884	318	0.38559	-1.9011e-05	7.3637e-06
scD(t-1)	-0.82853	0.096677	-8.5701	318	4.5609e-16	-1.0187	-0.63833
NEWS(T)	-0.13361	0.033577	-3.9791	318	8.5729e-05	-0.19967	-0.067545
Ret(t)	-8.0021	0.27526	-29.071	318	1.4796e-91	-8.5436	-7.4605
Ret(t-1)	1.6028	0.30324	5.2856	318	2.3318e-07	1.0062	2.1994
Y(2011)	-0.1195	0.062149	-1.9228	318	0.055399	-0.24178	0.0027757
Y(2012)	-0.066811	0.061938	-1.0787	318	0.28155	-0.18867	0.05505
Y(2013)	-0.063958	0.061507	-1.0399	318	0.29919	-0.18497	0.057053
Y(2014)	-0.017393	0.063602	-0.27347	318	0.78467	-0.14253	0.10774
Y(2015)	-0.036331	0.068405	-0.53112	318	0.59571	-0.17092	0.098253
Y(2016)	0.094549	0.063514	1.4886	318	0.13758	-0.030412	0.21951
M(Feb)	0.044578	0.077387	0.57604	318	0.56499	-0.10768	0.19683
M(Mar)	0.08213	0.076383	1.0752	318	0.28308	-0.06815	0.23241
M(Apr)	-0.10733	0.081747	-1.3129	318	0.19016	-0.26816	0.053507
M(May)	-0.11733	0.083998	-1.3968	318	0.16345	-0.28259	0.047935
M(Jun)	-0.17977	0.083093	-2.1634	318	0.031252	-0.34325	-0.016285
M(Jul)	-0.034668	0.080139	-0.4326	318	0.6656	-0.19234	0.123
M(Aug)	-0.16769	0.080547	-2.0819	318	0.038151	-0.32616	-0.0092181
M(Sep)	-0.10968	0.078731	-1.3931	318	0.16457	-0.26458	0.045222
M(Oct)	0.035482	0.07534	0.47097	318	0.63799	-0.11274	0.18371
M(Nov)	-0.039115	0.075418	-0.51865	318	0.60437	-0.1875	0.10927
M(Dec)	-0.05407	0.076615	-0.70573	318	0.48087	-0.20481	0.096667

23. Active investors, likes

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.5954	0.090741	6.5616	318	2.16e-10	0.41688	0.77393
FB_likes	-9.7367e-07	5.253e-07	-1.8536	318	0.064728	-2.0072e-06	5.9827e-08
scD(t-1)	-0.82842	0.096672	-8.5693	318	4.5855e-16	-1.0186	-0.63822
NEWS(T)	-0.14222	0.033928	-4.1918	318	3.5905e-05	-0.20897	-0.075468
Ret(t)	-8.0472	0.277	-29.052	318	1.7262e-91	-8.5922	-7.5022
Ret(t-1)	1.6012	0.303	5.2843	318	2.3473e-07	1.005	2.1973
Y(2011)	-0.11403	0.061849	-1.8437	318	0.066153	-0.23572	0.007652
Y(2012)	-0.011414	0.070421	-0.16209	318	0.87134	-0.14996	0.12714
Y(2013)	0.016191	0.078214	0.20701	318	0.83613	-0.13769	0.17007
Y(2014)	0.043365	0.07196	0.60262	318	0.54719	-0.098213	0.18494
Y(2015)	-0.018897	0.068181	-0.27716	318	0.78183	-0.15304	0.11524
Y(2016)	0.11316	0.063309	1.7874	318	0.074824	-0.011398	0.23772
M(Feb)	0.046537	0.077373	0.60146	318	0.54796	-0.10569	0.19876
M(Mar)	0.096434	0.076282	1.2642	318	0.20709	-0.053647	0.24651
M(Apr)	-0.10167	0.081279	-1.2508	318	0.21191	-0.26158	0.058246
M(May)	-0.10686	0.083961	-1.2728	318	0.20402	-0.27205	0.058325
M(Jun)	-0.1692	0.082921	-2.0405	318	0.042129	-0.33234	-0.0060549
M(Jul)	-0.014418	0.080717	-0.17862	318	0.85835	-0.17322	0.14439
M(Aug)	-0.14715	0.081367	-1.8084	318	0.071481	-0.30723	0.012938
M(Sep)	-0.083781	0.080274	-1.0437	318	0.29742	-0.24172	0.074155
M(Oct)	0.062521	0.077039	0.81155	318	0.41766	-0.08905	0.21409
M(Nov)	-0.023606	0.075872	-0.31113	318	0.75591	-0.17288	0.12567
M(Dec)	-0.040007	0.076815	-0.52083	318	0.60285	-0.19114	0.11112

24. Active investors, shares

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.5885	0.0907	6.4884	318	3.3229e-10	0.41005	0.76695
FB_shares	-1.3561e-05	4.1848e-06	-3.2406	318	0.0013189	-2.1795e-05	-5.328e-06
scD(t-1)	-0.83397	0.096712	-8.6233	318	3.1319e-16	-1.0242	-0.6437
NEWS(T)	-0.15116	0.03404	-4.4407	318	1.2387e-05	-0.21813	-0.084189
Ret(t)	-8.1049	0.27788	-29.167	318	6.9221e-92	-8.6517	-7.5582
Ret(t-1)	1.5762	0.30292	5.2034	318	3.5148e-07	0.98025	2.1722
Y(2011)	-0.10616	0.061774	-1.7185	318	0.086672	-0.2277	0.015376
Y(2012)	0.040188	0.069535	0.57795	318	0.56371	-0.096619	0.17699
Y(2013)	0.034905	0.068783	0.50746	318	0.61219	-0.10042	0.17023
Y(2014)	0.033569	0.065631	0.51148	318	0.60937	-0.095558	0.1627
Y(2015)	-0.0053148	0.068365	-0.077742	318	0.93808	-0.13982	0.12919
Y(2016)	0.12596	0.06347	1.9846	318	0.048044	0.0010908	0.25084
M(Feb)	0.048186	0.0774	0.62255	318	0.53403	-0.1041	0.20047
M(Mar)	0.090362	0.076126	1.187	318	0.23611	-0.059412	0.24014
M(Apr)	-0.1087	0.081363	-1.336	318	0.1825	-0.26878	0.051375
M(May)	-0.11027	0.083908	-1.3142	318	0.18974	-0.27535	0.054816
M(Jun)	-0.17829	0.082859	-2.1517	318	0.03217	-0.34131	-0.01527
M(Jul)	-0.020393	0.080197	-0.25428	318	0.79944	-0.17818	0.13739
M(Aug)	-0.13075	0.081349	-1.6073	318	0.10898	-0.2908	0.029297
M(Sep)	-0.070921	0.079628	-0.89066	318	0.37379	-0.22758	0.085742
M(Oct)	0.081037	0.076761	1.0557	318	0.29191	-0.069987	0.23206
M(Nov)	-0.02133	0.075604	-0.28213	318	0.77803	-0.17008	0.12742
M(Dec)	-0.035848	0.076734	-0.46718	318	0.64069	-0.18682	0.11512

Moderate investors

25. Moderate investors, posts

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.54365	0.03572	15.22	318	1.1216e-39	0.47337	0.61392
FB_posts	0.0018524	0.0010636	1.7416	318	0.082539	-0.00024018	0.003945
scD(t-1)	-0.024271	0.04043	-0.60033	318	0.54872	-0.10382	0.055273
NEWS(T)	-0.14771	0.011301	-13.071	318	1.5195e-31	-0.16995	-0.12548
Ret(t)	-7.6544	0.084024	-91.098	318	6.6606e-230	-7.8198	-7.4891
Ret(t-1)	0.47688	0.10513	4.5361	318	8.1314e-06	0.27004	0.68372
Y(2011)	-0.27199	0.022059	-12.33	318	8.1202e-29	-0.31539	-0.22859
Y(2012)	-0.18958	0.02179	-8.7002	318	1.8137e-16	-0.23245	-0.14671
Y(2013)	-0.36673	0.023078	-15.891	318	2.9146e-42	-0.41214	-0.32133
Y(2014)	-0.39681	0.024365	-16.287	318	8.6073e-44	-0.44475	-0.34888
Y(2015)	-0.44763	0.02471	-18.116	318	6.8641e-51	-0.49625	-0.39902
Y(2016)	-0.039814	0.022435	-1.7746	318	0.076917	-0.083955	0.0043262
M(Feb)	0.075762	0.0242	3.1307	318	0.001906	0.02815	0.12337
M(Mar)	0.32328	0.025054	12.903	318	6.3674e-31	0.27399	0.37258
M(Apr)	0.068406	0.025804	2.651	318	0.0084269	0.017638	0.11917
M(May)	0.080219	0.025941	3.0923	318	0.0021618	0.029181	0.13126
M(Jun)	0.12228	0.026217	4.6642	318	4.5678e-06	0.070702	0.17386
M(Jul)	0.15454	0.025664	6.0217	318	4.7619e-09	0.10405	0.20504
M(Aug)	-0.037919	0.024942	-1.5203	318	0.12943	-0.086991	0.011153
M(Sep)	-0.036723	0.024634	-1.4908	318	0.13702	-0.08519	0.011743
M(Oct)	0.17037	0.023294	7.3141	318	2.1218e-12	0.12454	0.2162
M(Nov)	0.17819	0.023731	7.5086	318	6.0725e-13	0.1315	0.22488
M(Dec)	0.094594	0.02377	3.9796	318	8.5561e-05	0.047828	0.14136

26. Moderate investors, comments

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.54767	0.035417	15.464	318	1.2935e-40	0.47799	0.61735
FB_comments	4.1075e-06	2.2706e-06	1.809	318	0.071392	-3.5973e-07	8.5747e-06
scD(t-1)	-0.020608	0.040281	-0.51161	318	0.60928	-0.099859	0.058643
NEWS(T)	-0.14904	0.011266	-13.229	318	3.9456e-32	-0.1712	-0.12687
Ret(t)	-7.655	0.084026	-91.103	318	6.5597e-230	-7.8203	-7.4897
Ret(t-1)	0.49747	0.10429	4.77	318	2.8112e-06	0.29228	0.70266
Y(2011)	-0.27176	0.022024	-12.34	318	7.5044e-29	-0.31509	-0.22843
Y(2012)	-0.19268	0.022251	-8.6595	318	2.4223e-16	-0.23646	-0.14891
Y(2013)	-0.36604	0.022918	-15.972	318	1.4158e-42	-0.41113	-0.32095
Y(2014)	-0.38596	0.023311	-16.557	318	7.7032e-45	-0.43182	-0.34009
Y(2015)	-0.44628	0.024765	-18.02	318	1.6074e-50	-0.495	-0.39755
Y(2016)	-0.028937	0.022358	-1.2943	318	0.19651	-0.072925	0.015051
M(Feb)	0.069433	0.023792	2.9183	318	0.0037706	0.022622	0.11624
M(Mar)	0.32006	0.02482	12.895	318	6.83e-31	0.27123	0.36889
M(Apr)	0.064499	0.025392	2.5401	318	0.011556	0.014542	0.11446
M(May)	0.079295	0.025894	3.0623	318	0.0023838	0.02835	0.13024
M(Jun)	0.12126	0.026166	4.6342	318	5.2348e-06	0.069778	0.17274
M(Jul)	0.15392	0.025646	6.0015	318	5.3233e-09	0.10346	0.20438
M(Aug)	-0.041399	0.024898	-1.6627	318	0.09735	-0.090386	0.0075869
M(Sep)	-0.042049	0.024727	-1.7005	318	0.090006	-0.090698	0.0065997
M(Oct)	0.1669	0.023311	7.1599	318	5.6277e-12	0.12104	0.21276
M(Nov)	0.17874	0.023727	7.5334	318	5.1717e-13	0.13206	0.22543
M(Dec)	0.093743	0.023739	3.9489	318	9.6704e-05	0.047038	0.14045

27. Moderate investors, likes

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.55861	0.03544	15.762	318	9.1807e-42	0.48888	0.62833
FB_likes	2.6701e-07	1.807e-07	1.4776	318	0.14049	-8.8506e-08	6.2252e-07
scD(t-1)	-0.022048	0.040364	-0.54624	318	0.58528	-0.10146	0.057365
NEWS(T)	-0.14671	0.011396	-12.874	318	8.1571e-31	-0.16913	-0.12429
Ret(t)	-7.6488	0.084253	-90.784	318	1.9214e-229	-7.8146	-7.4831
Ret(t-1)	0.49417	0.10439	4.7339	318	3.3207e-06	0.28879	0.69954
Y(2011)	-0.27044	0.021993	-12.296	318	1.0796e-28	-0.31371	-0.22717
Y(2012)	-0.19955	0.025112	-7.9465	318	3.3594e-14	-0.24896	-0.15015
Y(2013)	-0.38131	0.028115	-13.563	318	2.2289e-33	-0.43663	-0.326
Y(2014)	-0.40231	0.026268	-15.316	318	4.7991e-40	-0.454	-0.35063
Y(2015)	-0.45439	0.024799	-18.323	318	1.0757e-51	-0.50318	-0.4056
Y(2016)	-0.037196	0.02229	-1.6687	318	0.096154	-0.081051	0.0066585
M(Feb)	0.067934	0.023782	2.8566	318	0.0045644	0.021145	0.11472
M(Mar)	0.31548	0.024739	12.752	318	2.304e-30	0.26681	0.36415
M(Apr)	0.060401	0.025212	2.3957	318	0.017166	0.010797	0.11001
M(May)	0.075376	0.025827	2.9185	318	0.0037675	0.024563	0.12619
M(Jun)	0.11705	0.026116	4.4819	318	1.0336e-05	0.065669	0.16843
M(Jul)	0.14806	0.025807	5.7374	318	2.2398e-08	0.097291	0.19884
M(Aug)	-0.046126	0.025193	-1.8309	318	0.068048	-0.095692	0.0034395
M(Sep)	-0.045323	0.025121	-1.8042	318	0.072154	-0.094748	0.0041023
M(Oct)	0.16146	0.023835	6.774	318	6.0714e-11	0.11456	0.20835
M(Nov)	0.17508	0.023824	7.3489	318	1.699e-12	0.12821	0.22196
M(Dec)	0.088485	0.023845	3.7108	318	0.0002438	0.041571	0.1354

28. Moderate investors, shares

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.54964	0.035417	15.519	318	7.9077e-41	0.47996	0.61932
FB_shares	-1.7066e-06	1.4039e-06	-1.2157	318	0.22502	-4.4686e-06	1.0554e-06
scD(t-1)	-0.015233	0.040303	-0.37795	318	0.70572	-0.094527	0.064062
NEWS(T)	-0.15146	0.011409	-13.275	318	2.6545e-32	-0.17391	-0.12901
Ret(t)	-7.6699	0.084504	-90.764	318	2.0554e-229	-7.8361	-7.5036
Ret(t-1)	0.50227	0.10429	4.816	318	2.2694e-06	0.29708	0.70745
Y(2011)	-0.26343	0.021934	-12.01	318	1.1817e-27	-0.30658	-0.22027
Y(2012)	-0.16346	0.024558	-6.6562	318	1.2321e-10	-0.21178	-0.11515
Y(2013)	-0.34208	0.024928	-13.723	318	5.5599e-34	-0.39112	-0.29304
Y(2014)	-0.37771	0.02393	-15.784	318	7.5416e-42	-0.42479	-0.33063
Y(2015)	-0.44708	0.024814	-18.017	318	1.6524e-50	-0.4959	-0.39826
Y(2016)	-0.031188	0.022303	-1.3984	318	0.16298	-0.075068	0.012692
M(Feb)	0.068283	0.023785	2.8708	318	0.0043688	0.021487	0.11508
M(Mar)	0.3162	0.024729	12.787	318	1.7152e-30	0.26755	0.36486
M(Apr)	0.057004	0.025229	2.2595	318	0.024532	0.0073669	0.10664
M(May)	0.075899	0.025831	2.9383	318	0.0035418	0.025077	0.12672
M(Jun)	0.1179	0.026107	4.5159	318	8.8939e-06	0.066533	0.16926
M(Jul)	0.15394	0.025659	5.9994	318	5.3871e-09	0.10345	0.20442
M(Aug)	-0.035038	0.025287	-1.3856	318	0.16684	-0.084789	0.014713
M(Sep)	-0.033083	0.024937	-1.3267	318	0.18557	-0.082146	0.015979
M(Oct)	0.17466	0.023734	7.3591	318	1.5921e-12	0.12796	0.22135
M(Nov)	0.18011	0.023773	7.5764	318	3.9067e-13	0.13334	0.22689
M(Dec)	0.094619	0.02381	3.974	318	8.7497e-05	0.047775	0.14146

Passive investors

29. Passive investors, posts

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.37889	0.022039	17.191	318	2.667e-47	0.33553	0.42225
FB_posts	0.012314	0.00064394	19.122	318	8.5505e-55	0.011047	0.01358
scD(t-1)	0.68562	0.020412	33.589	318	1.2973e-106	0.64546	0.72578
NEWS(T)	-0.053805	0.006542	-8.2246	318	5.0651e-15	-0.066676	-0.040934
Ret(t)	-5.2018	0.043553	-119.43	318	3.0976e-266	-5.2875	-5.1161
Ret(t-1)	-1.5922	0.049179	-32.375	318	1.1223e-102	-1.6889	-1.4954
Y(2011)	-0.27167	0.012836	-21.165	318	1.1575e-62	-0.29692	-0.24641
Y(2012)	-0.24963	0.013122	-19.023	318	2.0704e-54	-0.27544	-0.22381
Y(2013)	-0.5247	0.014359	-36.542	318	7.3271e-116	-0.55295	-0.49645
Y(2014)	-0.55782	0.014777	-37.75	318	1.6201e-119	-0.5869	-0.52875
Y(2015)	-0.31908	0.01499	-21.286	318	3.9728e-63	-0.34857	-0.28959
Y(2016)	0.29242	0.013184	22.18	318	1.5779e-66	0.26648	0.31836
M(Feb)	0.31605	0.014244	22.188	318	1.4704e-66	0.28803	0.34408
M(Mar)	0.50699	0.015252	33.24	318	1.7231e-105	0.47698	0.537
M(Apr)	0.31124	0.01526	20.396	318	1.034e-59	0.28122	0.34126
M(May)	0.26069	0.01494	17.449	318	2.6701e-48	0.2313	0.29008
M(Jun)	0.29087	0.014876	19.552	318	1.8533e-56	0.2616	0.32013
M(Jul)	0.15161	0.014626	10.365	318	7.0052e-22	0.12283	0.18039
M(Aug)	0.16516	0.014004	11.794	318	7.078e-27	0.13761	0.19271
M(Sep)	0.27598	0.014267	19.344	318	1.1847e-55	0.24791	0.30405
M(Oct)	0.35994	0.013815	26.055	318	6.828e-81	0.33277	0.38712
M(Nov)	0.4006	0.014293	28.028	318	6.3925e-88	0.37248	0.42872
M(Dec)	0.22895	0.013904	16.466	318	1.7299e-44	0.20159	0.2563

30. Passive investors, comments

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.49717	0.021758	22.85	318	4.6447e-69	0.45436	0.53998
FB_comments	-9.3322e-06	1.4257e-06	-6.5458	318	2.3706e-10	-1.2137e-05	-6.5273e-06
scD(t-1)	0.67761	0.02045	33.134	318	3.7865e-105	0.63737	0.71784
NEWS(T)	-0.063599	0.0065187	-9.7563	318	7.6502e-20	-0.076424	-0.050773
Ret(t)	-5.2286	0.043597	-119.93	318	8.5882e-267	-5.3143	-5.1428
Ret(t-1)	-1.5282	0.049135	-31.102	318	1.8423e-98	-1.6249	-1.4315
Y(2011)	-0.23317	0.012855	-18.139	318	5.549e-51	-0.25846	-0.20788
Y(2012)	-0.1569	0.013342	-11.76	318	9.3498e-27	-0.18315	-0.13065
Y(2013)	-0.4354	0.014158	-30.752	318	2.7348e-97	-0.46326	-0.40755
Y(2014)	-0.46554	0.014024	-33.195	318	2.4023e-105	-0.49314	-0.43795
Y(2015)	-0.35915	0.01505	-23.863	318	7.5196e-73	-0.38876	-0.32954
Y(2016)	0.31943	0.013154	24.283	318	2.0766e-74	0.29355	0.34531
M(Feb)	0.24945	0.013989	17.832	318	8.6512e-50	0.22193	0.27697
M(Mar)	0.45547	0.015167	30.029	318	7.6659e-95	0.42562	0.48531
M(Apr)	0.22674	0.015019	15.096	318	3.3329e-39	0.19719	0.25629
M(May)	0.22648	0.014936	15.163	318	1.8462e-39	0.1971	0.25587
M(Jun)	0.25065	0.014826	16.906	318	3.4171e-46	0.22148	0.27982
M(Jul)	0.13114	0.014593	8.986	318	2.3266e-17	0.10242	0.15985
M(Aug)	0.14713	0.013962	10.538	318	1.8098e-22	0.11966	0.1746
M(Sep)	0.26898	0.014231	18.901	318	6.1528e-54	0.24098	0.29698
M(Oct)	0.3561	0.013813	25.78	318	6.818e-80	0.32892	0.38328
M(Nov)	0.39636	0.014268	27.779	318	4.7742e-87	0.36829	0.42443
M(Dec)	0.20496	0.013859	14.789	318	5.0282e-38	0.17769	0.23223

31. Passive investors, likes

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.47594	0.021504	22.133	318	2.3752e-66	0.43364	0.51825
FB_likes	-7.3294e-07	1.1389e-07	-6.4354	318	4.5294e-10	-9.5701e-07	-5.0886e-07
scD(t-1)	0.67793	0.02045	33.15	318	3.3615e-105	0.63769	0.71816
NEWS(T)	-0.070091	0.0066076	-10.608	318	1.0435e-22	-0.083091	-0.057091
Ret(t)	-5.24	0.043677	-119.97	318	7.6843e-267	-5.326	-5.1541
Ret(t-1)	-1.5343	0.04917	-31.204	318	8.4075e-99	-1.631	-1.4375
Y(2011)	-0.23608	0.012809	-18.431	318	4.1062e-52	-0.26128	-0.21088
Y(2012)	-0.13478	0.014819	-9.0948	318	1.0542e-17	-0.16394	-0.10562
Y(2013)	-0.3915	0.016962	-23.081	318	6.3343e-70	-0.42487	-0.35813
Y(2014)	-0.41947	0.015979	-26.251	318	1.3434e-81	-0.45091	-0.38803
Y(2015)	-0.34191	0.014943	-22.88	318	3.5876e-69	-0.37131	-0.31251
Y(2016)	0.33763	0.013066	25.841	318	4.0721e-80	0.31192	0.36334
M(Feb)	0.25578	0.013934	18.357	318	7.9445e-52	0.22837	0.2832
M(Mar)	0.46588	0.015102	30.849	318	1.2965e-97	0.43616	0.49559
M(Apr)	0.23341	0.014898	15.667	318	2.1255e-41	0.2041	0.26272
M(May)	0.23511	0.014891	15.789	318	7.2128e-42	0.20581	0.26441
M(Jun)	0.25832	0.01478	17.478	318	2.0569e-48	0.22924	0.2874
M(Jul)	0.14593	0.014676	9.943	318	1.8436e-20	0.11705	0.1748
M(Aug)	0.15996	0.014084	11.358	318	2.5405e-25	0.13225	0.18767
M(Sep)	0.27357	0.014266	19.176	318	5.3046e-55	0.2455	0.30164
M(Oct)	0.37173	0.014097	26.369	318	5.0368e-82	0.344	0.39947
M(Nov)	0.40597	0.014305	28.379	318	3.7642e-89	0.37782	0.43411
M(Dec)	0.21751	0.013881	15.67	318	2.0738e-41	0.1902	0.24482

32. Passive investors, shares

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	0.47085	0.021534	21.866	318	2.4579e-65	0.42849	0.51322
FB_shares	-1.6005e-05	9.0144e-07	-17.755	318	1.7318e-49	-1.7778e-05	-1.4231e-05
scD(t-1)	0.6733	0.020476	32.883	318	2.4733e-104	0.63302	0.71359
NEWS(T)	-0.081345	0.0066067	-12.312	318	9.4425e-29	-0.094343	-0.068346
Ret(t)	-5.2983	0.043818	-120.92	318	6.6858e-268	-5.3845	-5.2121
Ret(t-1)	-1.5621	0.04918	-31.762	318	1.1702e-100	-1.6588	-1.4653
Y(2011)	-0.22793	0.012782	-17.833	318	8.6269e-50	-0.25308	-0.20278
Y(2012)	-0.05932	0.014556	-4.0753	318	5.8097e-05	-0.087958	-0.030682
Y(2013)	-0.34785	0.015125	-22.997	318	1.2994e-69	-0.37761	-0.31809
Y(2014)	-0.41334	0.014371	-28.762	318	1.7361e-90	-0.44162	-0.38507
Y(2015)	-0.33096	0.014962	-22.12	318	2.6645e-66	-0.3604	-0.30152
Y(2016)	0.3475	0.013068	26.592	318	7.9519e-83	0.32179	0.37321
M(Feb)	0.25625	0.013939	18.384	318	6.2507e-52	0.22883	0.28368
M(Mar)	0.46576	0.015109	30.826	318	1.5485e-97	0.43603	0.49549
M(Apr)	0.22142	0.014896	14.864	318	2.5817e-38	0.19211	0.25073
M(May)	0.23267	0.014903	15.613	318	3.4544e-41	0.20335	0.26199
M(Jun)	0.25306	0.014788	17.112	318	5.44e-47	0.22396	0.28215
M(Jul)	0.14507	0.014595	9.9396	318	1.8913e-20	0.11635	0.17379
M(Aug)	0.1835	0.014116	12.999	318	2.814e-31	0.15573	0.21127
M(Sep)	0.2856	0.014242	20.053	318	2.1558e-58	0.25758	0.31362
M(Oct)	0.39828	0.014062	28.324	318	5.8487e-89	0.37061	0.42594
M(Nov)	0.41146	0.01428	28.814	318	1.1454e-90	0.38336	0.43955
M(Dec)	0.22877	0.013883	16.478	318	1.5644e-44	0.20145	0.25608

Inactive investors

33. Inactive investors, posts

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.53309	0.023454	-22.729	318	1.3256e-68	-0.57923	-0.48694
FB_posts	0.014977	0.00067134	22.309	318	5.118e-67	0.013656	0.016298
scD(t-1)	2.7061	0.021064	128.47	318	4.2012e-276	2.6647	2.7476
NEWS(T)	0.014338	0.0066059	2.1705	318	0.03071	0.0013412	0.027335
Ret(t)	-4.6532	0.039041	-119.19	318	5.8857e-266	-4.7301	-4.5764
Ret(t-1)	-0.44413	0.047307	-9.3882	318	1.2133e-18	-0.5372	-0.35105
Y(2011)	-0.2492	0.012984	-19.193	318	4.5602e-55	-0.27474	-0.22365
Y(2012)	-0.15783	0.01348	-11.708	318	1.4353e-26	-0.18435	-0.13131
Y(2013)	-0.89546	0.016024	-55.883	318	1.6839e-166	-0.92699	-0.86393
Y(2014)	-0.95598	0.01612	-59.305	318	5.4177e-174	-0.9877	-0.92427
Y(2015)	-0.80527	0.016245	-49.57	318	1.1995e-151	-0.83723	-0.77331
Y(2016)	-0.093399	0.01375	-6.7925	318	5.4283e-11	-0.12045	-0.066346
M(Feb)	0.28113	0.014101	19.936	318	6.0939e-58	0.25338	0.30887
M(Mar)	0.0071384	0.015488	0.4609	318	0.64519	-0.023333	0.03761
M(Apr)	0.22834	0.01472	15.513	318	8.3886e-41	0.19938	0.2573
M(May)	-0.14367	0.014649	-9.8077	318	5.1775e-20	-0.17249	-0.11485
M(Jun)	0.17724	0.014879	11.912	318	2.672e-27	0.14796	0.20651
M(Jul)	0.25704	0.014413	17.834	318	8.5142e-50	0.22869	0.2854
M(Aug)	-0.13043	0.01374	-9.4925	318	5.5763e-19	-0.15746	-0.10339
M(Sep)	-0.22186	0.014273	-15.544	318	6.354e-41	-0.24994	-0.19378
M(Oct)	-0.0094852	0.013844	-0.68515	318	0.49375	-0.036723	0.017752
M(Nov)	-0.1521	0.01382	-11.006	318	4.3775e-24	-0.17929	-0.12491
M(Dec)	-0.58371	0.013784	-42.346	318	9.3673e-133	-0.61083	-0.55659

34. Inactive investors, comments

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.49684	0.023623	-21.032	318	3.7415e-62	-0.54332	-0.45036
FB_comments	1.0111e-05	1.4096e-06	7.1729	318	5.1878e-12	7.3373e-06	1.2884e-05
scD(t-1)	2.7456	0.021074	130.29	318	5.2664e-278	2.7041	2.787
NEWS(T)	0.002696	0.0065825	0.40957	318	0.6824	-0.010255	0.015647
Ret(t)	-4.6551	0.03919	-118.78	318	1.7018e-265	-4.7322	-4.5779
Ret(t-1)	-0.29915	0.047047	-6.3585	318	7.0753e-10	-0.39171	-0.20659
Y(2011)	-0.22246	0.012996	-17.118	318	5.1191e-47	-0.24803	-0.1969
Y(2012)	-0.11503	0.013715	-8.387	318	1.646e-15	-0.14202	-0.088047
Y(2013)	-0.82229	0.01575	-52.21	318	4.9675e-158	-0.85328	-0.79131
Y(2014)	-0.84671	0.015325	-55.249	318	4.5234e-165	-0.87687	-0.81656
Y(2015)	-0.80669	0.016353	-49.33	318	4.7025e-151	-0.83886	-0.77451
Y(2016)	-0.034296	0.013727	-2.4984	318	0.012979	-0.061303	-0.0072887
M(Feb)	0.24096	0.014016	17.191	318	2.6781e-47	0.21338	0.26853
M(Mar)	-0.0068534	0.015519	-0.4416	318	0.65908	-0.037387	0.02368
M(Apr)	0.19463	0.014702	13.238	318	3.6338e-32	0.1657	0.22355
M(May)	-0.15241	0.014634	-10.415	318	4.754e-22	-0.1812	-0.12362
M(Jun)	0.17608	0.014915	11.805	318	6.4442e-27	0.14673	0.20542
M(Jul)	0.27194	0.01437	18.924	318	5.0125e-54	0.24366	0.30021
M(Aug)	-0.12676	0.013718	-9.2408	318	3.6138e-18	-0.15375	-0.099774
M(Sep)	-0.20609	0.014215	-14.498	318	6.4797e-37	-0.23406	-0.17812
M(Oct)	0.015157	0.013759	1.1017	318	0.27144	-0.011912	0.042227
M(Nov)	-0.12793	0.013718	-9.3261	318	1.9235e-18	-0.15492	-0.10094
M(Dec)	-0.56432	0.0137	-41.192	318	1.5938e-129	-0.59128	-0.53737

35. Inactive investors, likes

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.48393	0.023321	-20.751	318	4.4679e-61	-0.52981	-0.43805
FB_likes	1.8368e-06	1.149e-07	15.987	318	1.2431e-42	1.6108e-06	2.0629e-06
scD(t-1)	2.7522	0.021051	130.74	318	1.7823e-278	2.7108	2.7936
NEWS(T)	0.0173	0.0066498	2.6016	318	0.0097143	0.0042166	0.030383
Ret(t)	-4.5906	0.0395	-116.22	318	1.5019e-262	-4.6683	-4.5129
Ret(t-1)	-0.30593	0.047069	-6.4996	318	3.1122e-10	-0.39853	-0.21332
Y(2011)	-0.22639	0.012898	-17.552	318	1.062e-48	-0.25176	-0.20101
Y(2012)	-0.21376	0.015283	-13.987	318	5.602e-35	-0.24383	-0.18369
Y(2013)	-0.95545	0.018229	-52.415	318	1.6231e-158	-0.99131	-0.91958
Y(2014)	-0.9579	0.016858	-56.822	318	1.3592e-168	-0.99107	-0.92474
Y(2015)	-0.82965	0.016244	-51.076	318	2.5489e-155	-0.86161	-0.7977
Y(2016)	-0.059337	0.013624	-4.3555	318	1.7934e-05	-0.086141	-0.032533
M(Feb)	0.23785	0.013914	17.095	318	6.3287e-47	0.21047	0.26522
M(Mar)	-0.020017	0.015438	-1.2967	318	0.19568	-0.05039	0.010355
M(Apr)	0.20427	0.014622	13.97	318	6.5037e-35	0.1755	0.23304
M(May)	-0.15798	0.014578	-10.837	318	1.6882e-23	-0.18666	-0.1293
M(Jun)	0.17349	0.014865	11.671	318	1.9495e-26	0.14425	0.20274
M(Jul)	0.24147	0.014482	16.674	318	2.7129e-45	0.21297	0.26996
M(Aug)	-0.14518	0.013767	-10.546	318	1.7006e-22	-0.17227	-0.1181
M(Sep)	-0.21678	0.01429	-15.17	318	1.7371e-39	-0.24489	-0.18866
M(Oct)	-0.014625	0.013887	-1.0531	318	0.2931	-0.041948	0.012698
M(Nov)	-0.14076	0.013736	-10.248	318	1.754e-21	-0.16778	-0.11373
M(Dec)	-0.5834	0.013735	-42.475	318	4.1205e-133	-0.61042	-0.55637

36. Inactive investors, shares

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
Intercept	-0.46994	0.023328	-20.145	318	9.5339e-59	-0.51584	-0.42405
FB_shares	1.0911e-07	9.7565e-07	0.11184	318	0.91102	-1.8104e-06	2.0286e-06
scD(t-1)	2.737	0.021045	130.06	318	9.0966e-278	2.6956	2.7784
NEWS(T)	0.0015177	0.0066619	0.22781	318	0.81994	-0.011589	0.014625
Ret(t)	-4.6787	0.039507	-118.43	318	4.3212e-265	-4.7565	-4.601
Ret(t-1)	-0.30962	0.047036	-6.5825	318	1.9086e-10	-0.40216	-0.21708
Y(2011)	-0.20882	0.012887	-16.204	318	1.7989e-43	-0.23417	-0.18346
Y(2012)	-0.086963	0.015318	-5.6772	318	3.087e-08	-0.1171	-0.056826
Y(2013)	-0.80129	0.016866	-47.508	318	1.7795e-146	-0.83448	-0.76811
Y(2014)	-0.84592	0.01565	-54.053	318	2.4467e-162	-0.87671	-0.81513
Y(2015)	-0.82065	0.016284	-50.396	318	1.132e-153	-0.85269	-0.78861
Y(2016)	-0.047446	0.013637	-3.4791	318	0.00057329	-0.074276	-0.020615
M(Feb)	0.22938	0.013925	16.473	318	1.6367e-44	0.20198	0.25677
M(Mar)	-0.016097	0.015474	-1.0403	318	0.29901	-0.04654	0.014347
M(Apr)	0.18046	0.014639	12.327	318	8.3442e-29	0.15165	0.20926
M(May)	-0.15891	0.014614	-10.874	318	1.2599e-23	-0.18766	-0.13016
M(Jun)	0.16915	0.014893	11.358	318	2.5379e-25	0.13985	0.19846
M(Jul)	0.27327	0.014384	18.999	318	2.5698e-54	0.24497	0.30157
M(Aug)	-0.12712	0.013776	-9.2275	318	3.9857e-18	-0.15422	-0.10001
M(Sep)	-0.20558	0.014211	-14.467	318	8.4953e-37	-0.23354	-0.17762
M(Oct)	0.017963	0.013893	1.2929	318	0.19697	-0.0093709	0.045296
M(Nov)	-0.13051	0.013712	-9.5176	318	4.6213e-19	-0.15748	-0.10353
M(Dec)	-0.56859	0.0137	-41.502	318	2.1318e-130	-0.59554	-0.54163

Appendix C

Online appendix of Article II:

Siikanen M., Kanniainen, J., Valli, J. "Limit order books and liquidity around scheduled and non-scheduled announcements: Empirical evidence from NASDAQ Nordic", *Finance Research Letters.*

Available at: https://doi.org/10.1016/j.frl.2016.12.016

Online Appendix: Limit Order Books and Liquidity around Scheduled and Non-Scheduled Announcements: Empirical Evidence from NASDAQ Nordic

Milla Siikanen[a,*], Juho Kanniainen[a], Jaakko Valli[a]

[a] *Department of Industrial Management, Tampere University of Technology, P.O. Box 541, FI-33101 Tampere, Finland*

Appendix

This appendix provides additional tables and figures: i) t-test, ii) the evolution of order book asymmetry, and iii) the evolution of spread between the best levels.

Table 1: **Pre-reaction in order book illiquidity (OBI).** This table tests if there is a liquidity reaction in the book before the announcement using the one-sample t-test. Given that $\Delta\ln(\text{OBI})_{\mathcal{E},\mathcal{A}-} = \ln(\text{OBI})^{\text{Med}}_{\mathcal{A}-} - \ln(\text{OBI})^{\text{Med}}_{\mathcal{E}}$, where $\ln(\text{OBI})^{\text{Med}}_{\mathcal{A}-}$ and $\ln(\text{OBI})^{\text{Med}}_{\mathcal{E}}$ are median values of ln(OBI) from pre-event and estimation windows, respectively, the null hypothesis is that $\Delta\ln(\text{OBI})_{\mathcal{E},\mathcal{A}-}$ comes from a normal distribution with mean zero. $\Delta\ln(\text{OBI})_{\mathcal{E},\mathcal{A}-}$ is calculated separately for the bid and ask sides for positive and negative, scheduled and non-scheduled announcements, using 60-minute pre-event window. We use logarithmic values to better fulfill the the normality assumption. For the robustness check, we run a Wilcoxon signed rank test to test the null hypothesis that $\Delta\ln(\text{OBI})_{\mathcal{E},\mathcal{A}-}$ comes from a distribution whose median is zero (the test assumes that observations come from a continuous distribution, symmetric about its median) and get similar results. We also repeat the analysis using mean values from pre- and estimation windows instead of median to calculate $\Delta\ln(\text{OBI})_{\mathcal{E},\mathcal{A}-}$, and the results remain the same. The non-scheduled announcement sample is extracted from the original sample by choosing the same number of announcements as with the corresponding scheduled announcement sample with largest relative price impact (log-change in mid-price right before the event till the end of event window). The sample size N is 151 for positive and 165 for negative announcements.

Scheduled	Positive		Negative	
	ask	bid	ask	bid
MEAN ($\Delta\text{OBI}_{\mathcal{E},\mathcal{A}-}$)	0.753 ***	0.923 ***	0.562 ***	0.685 ***
t-statistic	(8.560)	(8.952)	(5.738)	(6.315)

Non-scheduled	Positive		Negative	
	ask	bid	ask	bid
MEAN ($\Delta\text{OBI}_{\mathcal{E},\mathcal{A}-}$)	0.457 ***	0.086	0.173 *	0.149
t-statistic	(5.169)	(0.919)	(1.977)	(1.931)

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

*Corresponding author: tel. +358 50 353 1130

Email addresses: milla.siikanen@tut.fi (Milla Siikanen), juho.kanniainen@tut.fi (Juho Kanniainen), jaakko.s.valli@gmail.com (Jaakko Valli)

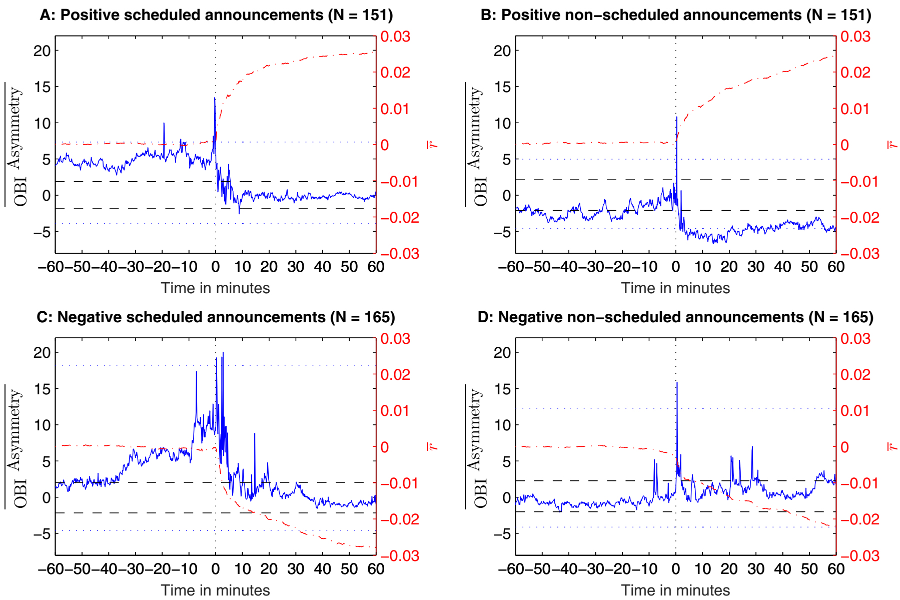

Figure 1: **Order book asymmetry around announcements.** $\overline{\mathrm{OBI}^{\mathrm{Asymmetry}}}$, standardised and aggregated of bid side minus ask side order book illiquidity, around scheduled and non-scheduled announcements, separately for positive and negative announcements and $\bar{r}$, average return on mid-price. The blue solid line corresponds to $\overline{\mathrm{OBI}^{\mathrm{Asymmetry}}}$, the black dashed lines correspond to the 95% confidence level of $\overline{\mathrm{OBI}^{\mathrm{Asymmetry}}}$ based on the empirical distribution from the estimation window (i.e. 27 days preceding the announcement) and the blue dotted lines correspond to the maximum and minimum values of $\overline{\mathrm{OBI}^{\mathrm{Asymmetry}}}$ in the estimation window, all with the scale on the left-hand side. The red dash-dot line corresponds to $\bar{r}$ with the scale on the right-hand side in red. The black dotted vertical line at time zero corresponds to the time of the announcement. The non-scheduled announcement sample is extracted from the original sample by choosing the same number of announcements as with the corresponding scheduled announcement sample with largest relative price impact (log-change in mid-price right before the event till the end of event window).

Table 2: **Order book asymmetry around announcements.** This table presents observations of $\overline{\text{OBI}^{\text{Asymmetry}}}$, standardised and aggregated of bid side minus ask side order book illiquidity, around scheduled and non-scheduled announcements, separately for positive and negative announcements. The statistical significance is calculated based on the empirical distribution from the estimation window (i.e. 27 days preceding the announcement) and indicated by asterisks: ***, **, and * indicate that the observations is outside of two sided 99.9%, 99%, and 95% confidence intervals, respectively. The non-scheduled announcement sample is extracted from the original sample by choosing the same number of announcements as with the corresponding scheduled announcement sample with largest relative price impact (log-change in mid-price right before the event till the end of event window). The sample size N is 151 for positive and 165 for negative announcements.

Scheduled announcements

Announcement	Variable	Before -10 sec	-1 min	-5 min	-15 min	-30 min	-60 min
Positive	$\overline{\text{OBI}^{Asymmetry}}$	11.88 ***	6.26 ***	4.17 **	5.25 ***	5.16 ***	4.03 **
Negative	$\overline{\text{OBI}^{Asymmetry}}$	9.60 ***	9.84 ***	7.45 ***	5.27 ***	5.95 ***	1.52

Announcement	Variable	After +10 sec	+1 min	+5 min	+15 min	+30 min	+60 min
Positive	$\overline{\text{OBI}^{Asymmetry}}$	3.59 **	0.92	2.93 *	-0.01	-0.51	0.36
Negative	$\overline{\text{OBI}^{Asymmetry}}$	10.83 ***	11.67 ***	3.41 ***	0.63	1.52	-0.99

Non-scheduled announcements

Announcement	Variable	Before -10 sec	-1 min	-5 min	-15 min	-30 min	-60 min
Positive	$\overline{\text{OBI}^{Asymmetry}}$	0.30	-2.11 *	0.30	-2.12 *	-2.60 *	-1.80
Negative	$\overline{\text{OBI}^{Asymmetry}}$	0.73	0.20	0.12	-1.10	-1.01	-1.07

Announcement	Variable	After +10 sec	+1 min	+5 min	+15 min	+30 min	+60 min
Positive	$\overline{\text{OBI}^{Asymmetry}}$	-1.34	-2.73 *	-5.45 ***	-6.45 ***	-4.36 ***	-4.37 ***
Negative	$\overline{\text{OBI}^{Asymmetry}}$	2.69 *	2.93 **	0.06	0.49	2.36 *	1.85

Table 3: **Relative spread around announcements.** This table presents observations of $\overline{\text{SPREAD}}$, standardised and aggregated relative spread around scheduled and non-scheduled announcements, separately for positive and negative announcements. The statistical significance is calculated based on the empirical distribution from the estimation window (i.e. 27 days preceding the announcement) and indicated by asterisks. ***, **, and * indicated by asterisks: ***, **, and * indicate that the observations is outside of two sided 99.9%, 99%, and 95% confidence intervals, respectively. The non-scheduled announcement sample is extracted from the original sample by choosing the same number of announcements as with the corresponding scheduled announcement sample with largest relative price impact (log-change in mid-price right before the event till the end of event window). The sample size N is 151 for positive and 165 for negative announcements.

Scheduled announcements

Announcement	Variable	Before -10 sec	-1 min	-5 min	-15 min	-30 min	-60 min
Positive	$\overline{\text{SPREAD}}$	8.09 ***	6.12 ***	1.14	2.76 *	2.45 *	-0.23
Negative	$\overline{\text{SPREAD}}$	5.41 ***	5.35 ***	7.61 ***	1.43	2.29 *	-0.04

Announcement	Variable	After +10 sec	+1 min	+5 min	+15 min	+30 min	+60 min
Positive	$\overline{\text{SPREAD}}$	13.02 ***	29.19 ***	13.61 ***	4.22 ***	1.25	2.11
Negative	$\overline{\text{SPREAD}}$	11.20 ***	24.68 ***	21.53 ***	7.16 ***	3.73 **	2.17

Non-scheduled announcements

Announcement	Variable	Before -10 sec	-1 min	-5 min	-15 min	-30 min	-60 min
Positive	$\overline{\text{SPREAD}}$	3.04 **	2.44 *	2.02	1.76	0.34	2.19
Negative	$\overline{\text{SPREAD}}$	3.79 **	3.79 **	2.66 *	3.54 **	1.39	4.00 ***

Announcement	Variable	After +10 sec	+1 min	+5 min	+15 min	+30 min	+60 min
Positive	$\overline{\text{SPREAD}}$	3.41 **	10.99 ***	10.15 ***	3.86 **	3.67 **	1.42
Negative	$\overline{\text{SPREAD}}$	7.55 ***	14.64 ***	8.34 ***	4.18 ***	3.78 **	5.50 ***

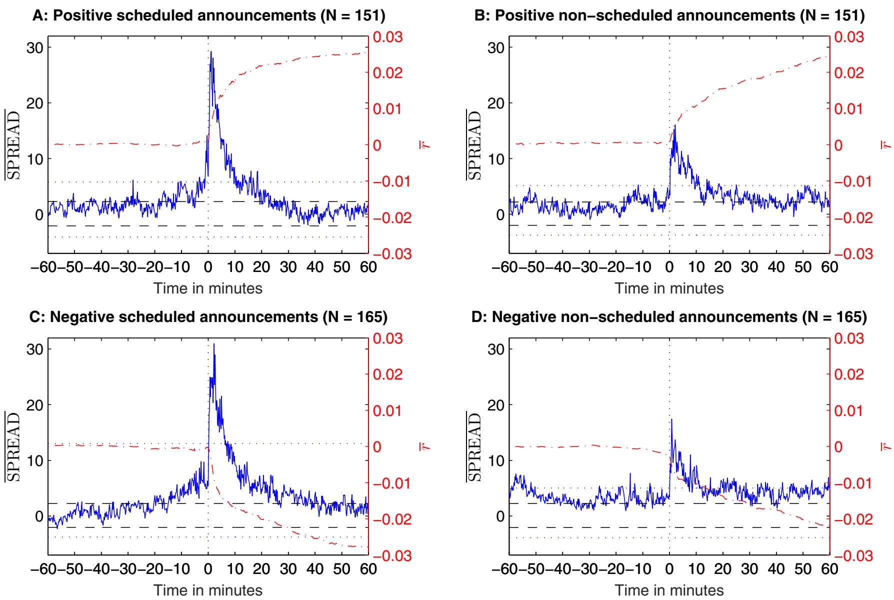

Figure 2: **Relative spread around announcements.** $\overline{\text{SPREAD}}$, standardised and aggregated relative spread, and $\overline{r}$, average return on mid-price, around scheduled and non-scheduled announcements, separately for positive and negative announcements. The blue solid line corresponds to $\overline{\text{SPREAD}}$, the black dashed lines correspond to the 95% confidence level of $\overline{\text{SPREAD}}$ based on the empirical distribution from the estimation window (i.e. 27 days preceding the announcement), and the blue dotted lines correspond to the maximum and minimum values of $\overline{\text{SPREAD}}$ in the estimation window, all with the scale on the left-hand side. The red dash-dot line corresponds to $\overline{r}$ with the scale on the right-hand side in red. The black line at time zero corresponds to the time of the announcement. The black dotted vertical line at time zero corresponds to the time of the announcement. The non-scheduled announcement sample is extracted from the original sample by choosing the same number of announcements as with the corresponding scheduled announcement sample with largest relative price impact (log-change in mid-price right before the event till the end of event window).

Appendix D

Online appendix of Article III:

Siikanen M., Kanniainen, J., Luoma, A. “What drives the sensitivity of limit order books to company announcement arrivals?”, *Economics Letters*.

Available at: https://doi.org/10.1016/j.econlet.2017.07.018

Online Appendix: What drives the sensitivity of limit order books to company announcement arrivals?

Milla Siikanen[a,*], Juho Kanniainen[a], Arto Luoma[b,1]

[a]*Department of Industrial Management, Tampere University of Technology, P.O. Box 541, FI-33101 Tampere, Finland*

[b]*School of Management, FI-33014 University of Tampere*

Appendix A. Means, medians and standard deviations of regression variables

Table 1 presents the means, medians, and standard deviations of the variables in our regression. We can see that the size of the relative illiquidity shock after scheduled announcement releases ($\Delta\ln(\mathrm{OBI})^{\mathrm{Max}}_{\mathcal{E},\mathcal{A}+}$) is, on average, larger than the one after nonscheduled announcements.[2] The average liquidity on bid side is lower than on the ask side ($\ln(\mathrm{OBI})^{\mathrm{Med}}_{\mathcal{E}}$ is higher), which is consistent with the observation of Malo and Pennanen (2012). The illiquidity peak takes place most commonly around 3 minutes after the scheduled announcement releases. For non-scheduled announcement releases, the peak happens usually around 13 (25) minutes after the announcement when we use the sample with 30 (60) minute pre- and post event windows. The maximum asymmetry of the book ($\ln(\mathrm{OBI})^{\mathrm{Asymmetry}}_{\mathcal{A}-}$) seems to be slightly larger before scheduled announcement releases, whereas relative spread remains the same. The average price change from the estimation window to the pre-event window ($r_{\mathcal{E},\mathcal{A}-}$)is positive for scheduled and negative for non-scheduled announcement releases in our sample. Moreover, scheduled announcement sample includes relatively more announcements with positive price impact after the release than the non-scheduled announcement sample.

*Corresponding author

Email address: `milla.siikanen@tut.fi` (Milla Siikanen)

[1]Present address: Department of Mathematics and Statistics, P.O.Box 35, FI-40014 University of Jyväskylä, Finland.

[2]See also Siikanen et al. (2017) for illustrations on evolution of OBI around scheduled and non-scheduled announcements.

Table 1: **Descriptive statistics of the regression variables.** Means, medians, and standard deviations of the regression variables. Side indicates for which side of the LOB, bid or ask side the variable is calculated. Some of the variables, such as mid-price returns are common for both sides.

Scheduled

Side	Variable	30 min Mean	30 min Median	30 min StDev	60 min Mean	60 min Median	60 min StDev
Ask							
	$\Delta\ln(\text{OBI})^{\text{Max}}_{\mathcal{E},\mathcal{A}+}$	0.889	0.892	0.813	0.981	0.940	0.746
	$\ln(\text{OBI})^{\text{Med}}_{\mathcal{E}}$	-4.926	-4.796	2.082	-5.049	-5.028	2.094
	τ^*	6.696	2.833	8.275	11.431	2.833	17.474
Bid							
	$\Delta\ln(\text{OBI})^{\text{Max}}_{\mathcal{E},\mathcal{A}+}$	1.094	1.028	0.976	1.190	1.122	0.929
	$\ln(\text{OBI})^{\text{Med}}_{\mathcal{E}}$	-4.810	-4.632	2.045	-4.931	-4.906	2.067
	τ^*	7.429	3.333	8.570	11.932	3.500	16.919
common							
	$\ln(\text{OBI})^{\text{Asymmetry}}_{\mathcal{A}-}$	1.110	1.034	0.561	1.248	1.103	0.582
	$r_{\mathcal{E},\mathcal{A}-}$	0.014	0.019	0.083	0.015	0.020	0.079
	$\text{spread}^{\text{Med}}_{\mathcal{A}-}$	0.002	0.002	0.002	0.002	0.002	0.002
	D_+	0.507	1.000	0.501	0.459	0.000	0.499
Number of observations			408			329	

Non-Scheduled

Side	Variable	30 min Mean	30 min Median	30 min StDev	60 min Mean	60 min Median	60 min StDev
Ask							
	$\Delta\ln(\text{OBI})^{\text{Max}}_{\mathcal{E},\mathcal{A}+}$	0.340	0.289	0.627	0.417	0.363	0.625
	$\ln(\text{OBI})^{\text{Med}}_{\mathcal{E}}$	-5.323	-5.429	2.033	-5.267	-5.376	2.043
	τ^*	13.210	11.833	10.019	26.591	24.667	20.219
Bid							
	$\Delta\ln(\text{OBI})^{\text{Max}}_{\mathcal{E},\mathcal{A}+}$	0.355	0.324	0.659	0.452	0.411	0.664
	$\ln(\text{OBI})^{\text{Med}}_{\mathcal{E}}$	-5.205	-5.335	2.036	-5.148	-5.288	2.048
	τ^*	13.587	12.667	9.933	26.791	24.500	20.083
common							
	$\ln(\text{OBI})^{\text{Asymmetry}}_{\mathcal{A}-}$	0.961	0.840	0.565	1.080	0.948	0.578
	$r_{\mathcal{E},\mathcal{A}-}$	-0.009	0.003	0.093	-0.008	0.002	0.092
	$\text{spread}^{\text{Med}}_{\mathcal{A}-}$	0.002	0.002	0.003	0.002	0.002	0.003
	D_+	0.436	0.000	0.496	0.447	0.000	0.497
Number of observations			2,629			2,102	

Appendix B. Regression results with 60-minute event windows

Table 2 presents the regression results using 60 minute pre-and post event windows.

Table 2: **Association between LOB illiquidity shock and LOB related factors using 60 minute pre- and post event windows.** The robust standard errors appear in parentheses. The regression results using the 30- and 60-minute pre- and post-event windows are mostly consistent, though some small variation exists. A potential reason for this is that an increase in the length of the window decreases the sample size, as some news releases occur too close to the beginning or end of the trading day to form the pre- and post-event windows. The use of the 60-minute window leads to around 20% decrease in the sample size when compared to the 30-minute window.

Parameter	Variable	Scheduled announcements ask	Scheduled announcements bid	Non-scheduled announcements ask	Non-scheduled announcements bid
α_1	$\ln(\text{OBI})_{\mathcal{E}}^{\text{Med}}$	-0.080	-0.180*	-0.036	-0.036
		(0.049)	(0.082)	(0.022)	(0.027)
α_2	$\ln(\text{OBI})_{\mathcal{A}-}^{\text{Asymmetry}}$	0.199**	0.422***	0.087**	0.270***
		(0.073)	(0.121)	(0.031)	(0.054)
α_3	$r_{\mathcal{E},\mathcal{A}-}$	-0.012	0.006	-1.736***	-1.040***
		(0.597)	(0.575)	(0.337)	(0.297)
α_4	$\text{SPREAD}_{\mathcal{A}-}^{\text{Med}}$	86.849**	73.779**	26.331***	21.096*
		(30.803)	(22.181)	(7.767)	(10.412)
α_5	τ^*	-0.006*	-0.009***	-0.002**	-0.001
		(0.002)	(0.003)	(0.001)	(0.001)
α_6	D_+	0.146	-0.263*	0.144***	-0.142***
		(0.075)	(0.106)	(0.034)	(0.040)
Number of observations		329	329	2,102	2,102
R^2		0.100	0.166	0.100	0.090

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Appendix C. Squared Return Regression

We run a linear regression to explain the squared returns from the estimation window to the illiquidity shock for non-scheduled announcements. In particular, the dependent variable is

$$r^2_{\mathcal{E},\mathcal{A}+} = \left(\ln\left[m^{\text{max, ln(OBI)}}\right] - \ln\left[\underset{t\in\mathcal{E}}{\text{Median}}(m_t)\right]\right)^2,$$

where $m^{\text{max, ln(OBI)}}$ is the mid-price at the moment when ln(OBI) reaches its maximum value in the post-event window. Note that $r^2_{\mathcal{E},\mathcal{A}+}$ is *not* the log return to the maximum mid-price in the post event window, but for the moment, ln(OBI) reaches its maximum value after the event. For a robustness check, we also use the squared periodic return instead of log return, i.e.

$$r^2_{\mathcal{E},\mathcal{A}+} = \left(\frac{m^{\text{max, ln(OBI)}}}{\text{Median}_{t\in\mathcal{E}}(m_t)} - 1\right)^2,$$

An explanatory variable is $r_{\mathcal{E},\mathcal{A}-}$. We run the regression separately for the ask and bid sides (the maximum ln(OBI) can be reached at different times for the ask and bid sides), 30- and 60-minute event windows, and both log-return and periodic return versions of $r^2_{\mathcal{E},\mathcal{A}+}$ for non-scheduled announcements. We use the within-transformation as in the regressions of the original paper.

Only one out of eight regressions gives a statistically significant regression estimate (bid side, 30-minute window, log-return version: $\hat{\alpha} = 0.234^*$ (robust standard error = 0.119)), while the rest of the regression estimates are insignificant and are not provided here, but are available upon request.

References

Malo, P., Pennanen, T., 2012. Reduced form modeling of limit order markets. Quantitative Finance 12, 1025–1036.

Siikanen, M., Kanniainen, J., Valli, J., 2017. Limit order books and liquidity around scheduled and non-scheduled announcements: Empirical evidence from nasdaq nordic. Finance Research Letters 21, 264–271.

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-4211-4
ISSN 1459-2045